DISSERTATION

ENGINEERED CO-CRYSTALS AS SCAFFOLDS FOR STRUCTURAL BIOLOGY

Submitted by

Abigail R. Orun

Department of Chemistry

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2022

Doctoral Committee:

Advisor: Christopher D. Snow

Christopher Ackerson
Seonah Kim
P. Shing Ho

ABSTRACT

ENGINEERED CO-CRYSTALS AS SCAFFOLDS FOR STRUCTURAL BIOLOGY

Biomolecules, like protein and DNA, serve as the foundation of life. The structure of biomolecules can give insight to their functions. X-ray crystallography is a cornerstone of structural biology, revealing atomic-level details of macromolecular structures. Even with advances in X-ray diffraction technology, haphazard and tedious crystal preparation remains the bottleneck of routine structure determination. An alternative to the crystal growth challenge is a scaffold crystal. Hypothetically, if one had a high-quality crystal already prepared with large enough pores for diffusion of a macromolecule, a biomolecule of interest could join the scaffold crystal for scaffold-assisted X-ray diffraction. An ideal scaffold crystal must be highly porous for guest addition, modular for installation of various guest molecules, and robust in changing solution conditions. A crystal with guest anchoring sites for *post*-crystallization guest addition may provide a high-throughput technique for guest DNA-binding protein structure determination.

The overarching goal of this work is to design a novel scaffold crystal capable of scaffold-assisted X-ray crystallography. The scaffold crystals we designed are co-crystals of DNA and DNA-binding protein. In the co-crystal, the DNA serves as the anchoring point for guest DNA-binding guest targets while the protein acts as connective tissue to hold the DNA structure together. The scaffold co-crystal we engineered, Co-Crystal 1 (CC1), is the first example of a porous host crystal for DNA-

binding guests. Ultimately, the expanded co-crystals may serve as a revolutionary figurative “lens” for routine structure determination.

In addition to scaffold crystal development, we advanced methods to enhance scaffold stability and solution-independence, thereby augmenting the bioconjugation toolkit for crystals containing stacking DNA-DNA junctions. Specifically, we optimized a known bioconjugation technique, carbodiimide chemical DNA ligation, templated by crystals with stacking DNA junctions. Furthermore, crystal crosslinking chemistries were optimized to provide crystal strength at both the nanoscale and the macroscale. Post-crosslinking, co-crystal nanostructures were preserved as assessed using X-ray diffraction and co-crystal macrostructures were bolstered in harsh solution conditions. The crosslinking chemistry and protocol guidelines may advance the progress of DNA crystals and protein-DNA co-crystals utility in biomedical applications and structural biology.

We are on the cusp of using designed co-crystals to host guest DNA-binding proteins for structural biology, bio-sensing, and bio-therapeutic delivery. Successful engineering of a designed porous co-crystal will open numerous application possibilities and scientific questions. For example, a future study could focus on quantifying guest protein diffusion rates and adsorption strength inside the porous scaffold crystals. The technology presented here may advance the study of DNA-binding proteins and advance our understanding of key proteins for cancer and disease.

ACKNOWLEDGEMENTS

“If I have seen further, it is by standing on the shoulders of Giants.” – Isaac Newton

No words can express the tremendous support and knowledge I gained at Colorado State University in the last six years. The work in my dissertation was made possible by the “giants” in my life.

I thank my advisor, Dr. Christopher Snow, the utmost advocate, teacher, and mentor in my PhD career. He believed in me from the beginning and always with a smile, pushed me to be my best. His never-ending enthusiasm for science is contagious and inspiring. I will always appreciate his patience and understanding, especially during the COVID-19 pandemic. Finally, his dedication to education created an inclusive learning environment fostering mentor-mentee relationships I will cherish forever.

I thank my committee members Dr. Christopher Ackerson, Dr. Seonah Kim, and Dr. P. Shing Ho for their dedication to my graduate education. A special thanks to Dr. Christopher Ackerson for his guidance in grant writing and inspiring conversations about science. I thank Dr. P. Shing Ho for his advice in DNA crystallography and methodology, and his motivating investigation for scientific truth. I also thank my previous committee member, Dr. Martin McCullagh, for his support in my professional development. I am also grateful to the Chemistry Department, Dr. Tony Rappé, and Dr. Harmony Tucker for giving me opportunity to develop chemistry teaching materials and learn from your teaching practices. I thank the Chemistry Department for providing travel funds to disseminate co-crystal research.

My gratitude extends to the Snow Lab, both alumni and current members. Thank you, Dr. Thaddaus Huber, Dr. Ning Zhou, and Dr. Ann Kowalski for companionship and mentorship, teaching me foundational skills. I especially thank Julius Stuart for his inspiring work ethic and friendship during the past 5 years. I thank colleagues Yaya Fan, Alec Jones, Jacob DeRoo, Ashlyn Chen, Rojina Shrestha, Moe Masri and Dr. Gayani Perera for companionship and exploring science together.

My fondest memories in the Snow Lab include the support and scientific endeavors of my faithful Co-Crystal Team. Sara Dmytriw, thank you for your positive attitude, challenging scientific ideas, and dedication to pursue scientific knowledge. Ananya Vajapayajula, thank you for your contagious scientific curiosity, encouragement, and people personality. Ethan Shields, thank you for your relentless scientific pursuit, positivity, and comradery.

My earliest days in lab included the BIOMOD team, a team of undergraduate researchers. I thank all members from 2017 and 2018 for the earliest exploration of expanded co-crystals and learning together. I especially thank Vincent Braud, Sehoo Park, Ramsey Smith, Xandria Amash, and Alex Frickenstein. I will always remember the BIOMOD jamboree in San Francisco and happy times learning DNA nanotechnology together. Similarly, I was fortunate to work with summer research interns including, REU student, Amanda Roley and high school students: Sasha Chappell, Audrey Graham, Larry Chen, Ryan Applebee, Alex Claiborne, Stefanie Pfahnl, and Nick Chappell. You challenged me to understand crystallography at the molecular level and you motivated me with your enthusiasm for research. I also thank Kelly Kappel with the Poudre School District for coordinating the high school interns.

I am thankful to be a part of ISP Meitnerium, the Colorado chapter of women in chemistry. The amazing women in STEM that taught me the importance of advocates and challenged me to be a better scientist and self: Dr. Nancy Levinger, Zarina Munshi, Victoria Combs, Mj Riches, Maggi Brasch-Turi and all members of ISP. I also thank my high school chemistry teacher, Mrs. Blackwood, who inspired me to continue in science and encouraged me to pursue a graduate degree.

I thank CSU Writes and Dr. Kristina Quynn for a supportive writing environment. Also, I want to thank Dr. Nancy Levinger for teaching the grant writing course and giving me an encouraging writing space, especially during the COVID pandemic.

In Chapter 1, the authors thank Dr. Paul Paukstelis for contributing the model for the expanded DNA crystal.

In Chapters 2, we thank Hataichanok (Mam) Scherman, Ph.D., Director of the Histone Source at Colorado State University for the expression and purification of the RepE54 initiation protein; Jay Nix, Ph.D. at the ALS Beamline 4.2.2 for extensive support of the XRD data collection; P. Shing Ho, Ph.D. for crystallography advice and guest DNA-binding molecule suggestions; Thaddaus Huber, Ph.D. for cloning expertise and PSB3 plasmid; Ashlyn Chen and Moe Masri for assistance setting up Gryphon screens; and Callie Slaughter for crystal pictures.

In Chapter 3, the authors thank Hataichanok (Mam) Scherman, Ph.D., Director of the Histone Source at Colorado State University for the expression and purification of the RepE54 Initiation Protein; Jay Nix, Ph.D. at the ALS Beamline 4.2.2 for extensive support of the XRD data collection; P. Shing Ho, Ph.D. for crystallography advice and guest DNA-binding molecule suggestions; Thaddaus Huber, Ph.D. for cloning expertise and PSB3 plasmid; and Julius Stuart for advice on confocal microscopy and TAMRA-labelled oligomer handling.

In Chapter 4, we thank Hataichanok (Mam) Scherman, Director of the Histone Source at Colorado State University for the expression and purification of the RepE54 transcription factor and the purification of E2F8 transcription factor; Shing Ho for advice and identifying a flaw in the interpretation of an earlier version of our analysis; Mark Stenglein and Mikaela Samsel at the Next Generation Sequencing Facility at Colorado State University for TapeStation analysis; Jay Nix at the ALS Beamline 4.2.2 for extensive support of the XRD data collection; The Taipale Lab for their CC2 protein plasmid donation; and Thaddaus Huber for cloning expertise and PSB3 plasmid.

In Chapter 5, we thank Hataichanok (Mam) Scherman, Director of the Histone Source at Colorado State University for the expression and purification of the RepE54 transcription factor and the purification of E2F8 transcription factor; Jay Nix at the ALS Beamline 4.2.2 for extensive support of the XRD data collection; and Alex Fenton and Rachel Cohen for crystal plate set up.

Finally, I am grateful for my family and their support, love and encouragement. I thank my parents for teaching me diligence and faithfulness. I thank my husband Onur Örün for being my rock during graduate school, a pandemic, and life. I would not be here without you.

Thank you all.

-Abby Örün (Ward)

DEDICATION

For my NaNa, Velma Ward, who told me to "Go get it, girl!"

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. POROUS CRYSTALS AS SCAFFOLDS FOR STRUCTURAL BIOLOGY[1]

1.1. Overview

Molecular scaffolds provide routes to otherwise inaccessible organized states of matter. Scaffolds that are crystalline can be observed in atomic detail using diffraction, along with any guest molecules that have adopted coherent structures therein. This approach, scaffold-assisted structure determination, is not yet routine. However, with varying degrees of guest immobilization, porous crystal scaffolds have recently been decorated with guest molecules. In this chapter, we analyze recent milestones, compare the relative advantages and challenges of different types of scaffold crystals, and weigh the merits of diverse guest installation strategies.

1.2. Introduction

X-ray crystallography and cryo-EM, mainstays of contemporary biomolecular structure determination, require failure-prone and tedious sample preparation before data collection. Obtaining single crystals that diffract well, or a suitable vitreous frozen monolayer is often a bottleneck that greatly limits the rate of structure acquisition. Structural biologists would therefore benefit from a robust, routine method for ordering biomolecules – a scaffold structure. In this chapter, the focus will be on the use of crystalline scaffolds intended to assist with structure determination *via* diffraction.

---

[1] The work in this chapter was published in 2020 in *Current Opinion in Structural Biology*. Conception, writing and visualization, A.R.W. and C.D.S.
Ward, A. R., Snow, C. D. Porous crystals as scaffolds for structural biology. *Curr. Opin. Struct. Biol.* 2020, 60: 85-92.

Crystalline scaffold-assisted structure determination (SASD) was formally proposed by Nadrian Seeman in terms of a three-dimensional DNA crystal lattice with modular DNA for a specific DNA-binding guest to join the crystal.[1] Rather than relying on chance crystal nucleation and growth processes (*i.e.,* highly solution-dependent traditional crystallography), high-affinity designed interactions can drive self-assembly to a pre-determined lattice. Elimination of sample preparation variability is the paramount advantage of SASD. Other theoretical advantages include less stringent guest purity requirements, lower guest quantity requirements, and a straightforward solution to the phasing problem.

The time is right to assess SASD prospects. Porous scaffold crystals are being explored for diverse applications beyond structural biology.[2-7] Moreover, Maita successfully resolved the structure of guest ubiquitin within a porous protein crystal scaffold using single crystal X-ray diffraction (XRD).[8] We will review Maita's work and others that offer insight into the challenges that remain to achieve routine SASD. In light of Seeman's proposition, scaffold structures may bridge the way to high throughput structure determination of traditionally difficult biomolecules, including membrane proteins, highly dynamic proteins, or weakly DNA-binding proteins.

1.3. Properties of an Ideal Crystalline Scaffold Material

1.3.1. Scaffold Symmetry and Porosity

An ideal XRD scaffold is general (easy to apply to diverse guests), reliable (low failure rate and minimal required troubleshooting), and environment independent (diffracting well in a variety of solvents and temperatures). The scaffold also needs space for high occupancy macromolecular guest installation. Several highly porous scaffolds are shown in Figure 1.1 and Table 1.1. Most crystals in the Protein Data Bank (PDB) lack suitable symmetry, porosity, and installation site spacing to serve as scaffolds. All else being equal, lower symmetry crystals have fewer asymmetric units (ASU) *per* unit cell (UC) and are therefore preferable to maximize volume-*per*-guest. For example, the R1EN crystals (*P*321 space group) used by Maita match the volume-*per*-guest available in the major pores (~55 nm^3) of the CJ crystals (*P*622) used in the Snow lab despite the smaller pore diameter and shorter unit cell in the former.

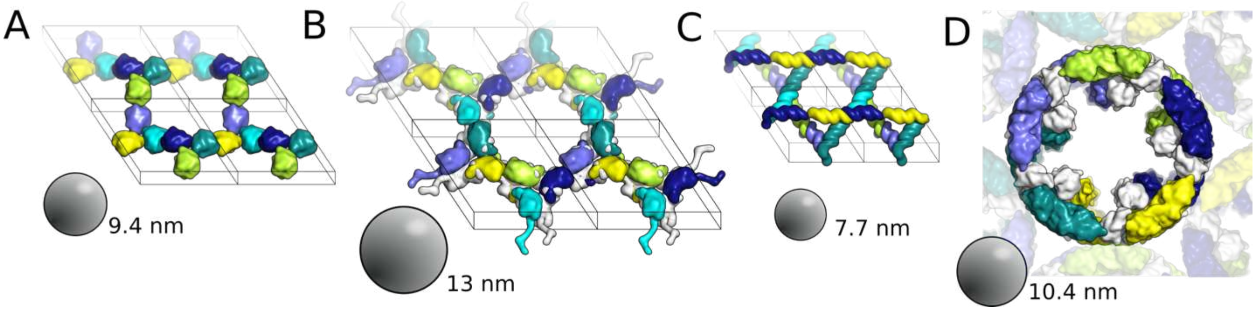


Figure 1.1. Selected porous crystal scaffolds. **(A)** R1EN, a R1Bm endonuclease domain from *B. mori*. **(B)** CJ, a YCEI protein from *C. jejuni*, with a domain swapped N-terminal beta strand **(C)** A model for a synthetic expanded DNA crystal in *P*6$_4$ (created on the basis of low-resolution diffraction and crystal surface AFM), provided courtesy of Dr. Paukstelis. **(D)** Molluscan hemocyanin decameric assembly. MAP_CHANNELS was used to estimate the maximum diameter for a probe sphere for 1-D diffusion through these crystals.[31] Within the top of each unit cell, building blocks (or hemocyanin protamers) are given distinct colors. We additionally compare these scaffolds on the basis of the periodic spacing between adjacent nanopores, as well as other unit cell (UC) and asymmetric unit (ASU) parameters.

Table 1.1. Features of porous scaffold crystals in Figure 1.1.

	R1EN [8]	**CJ** [13]	**DNA scaffold**† [20]	**Hemocyanin** [14]
Symmetry	$P321$	$P622$	$P6_4$	$\underline{D}5$ and $P12_11$
#ASU *per* UC	6	12	6	20
UC volume	648.6 nm^3	1394.0 nm^3	571.0 nm^3	27,837.6 nm^3
UC height	3.75 nm	5.0 nm	5.6 nm	17.1 nm
Pore-pore spacing	14.1 nm	17.9 nm	10.9 nm	31.1 nm
Max. guest diameter‡	9.4 nm	13.0 nm	7.7 nm	10.4 nm
ASU volume	108.1 nm^3	116.2 nm^3	95.2 nm^3	13,918.8 nm^3
Solvent content (%)	69.7 %	78.3%	87%*	67.6%
Shown PDB Entry	2ie9	5w17	NA	4yd9
Best Resolution	1.7 Å	2.5 Å	~5 Å	3 Å

* Estimated from MAP_CHANNELS[31]

† A model for a synthetic expanded DNA crystal (created on the basis of low-resolution diffraction and crystal surface AFM), provided courtesy of Dr. Paukstelis.

‡ Calculated using MAP_CHANNELS with a grid size of 5.0 Å for hemocyanin and 2.0Å for all other entries.

1.3.2. Asynchronous and Synchronous Guest Installation

On the basis of guest installation timing, we divide SASD into two categories (Figure 1.2). For synchronous guest installation (SGI), guests are attached to scaffold building blocks before crystallization. In contrast, for asynchronous guest installation (AGI), guest molecules are attached to pre-existing crystals. While AGI builds on a long-standing conventional crystallography tradition of soaking guest molecules into pre-existing crystals, SGI builds on a classical tradition of chaperone-assisted crystallography.[9,10] Pre-crystallization genetic fusion of a guest protein to a scaffold protein ensures 100% guest occupancy. Better yet, SGI has been successfully used to resolve a guest micro-peroxidase and ubiquitin.[8,11]

AGI also offers compelling advantages: (A) The traditional bottleneck problem, the search for reliable crystal growth conditions, must only be solved once. This compares favorably with SGI, where Maita fused seven guest domains to the R1EN scaffold and obtained crystals for two fused guests.[8] (B) Rather than a haphazard nucleation and growth process, guest addition is an adsorption process driven by affinity and mass action. (C) As noted with the metal organic framework (MOF) crystal sponge method,[12] adding a target molecule to a pre-existing crystal requires much less target molecule than a conventional crystallization campaign. (D) Purity requirements for the guest molecule could be much lower than traditional crystallography if the crystal binding motifs selectively adsorb the target molecule in the presence of impurities.

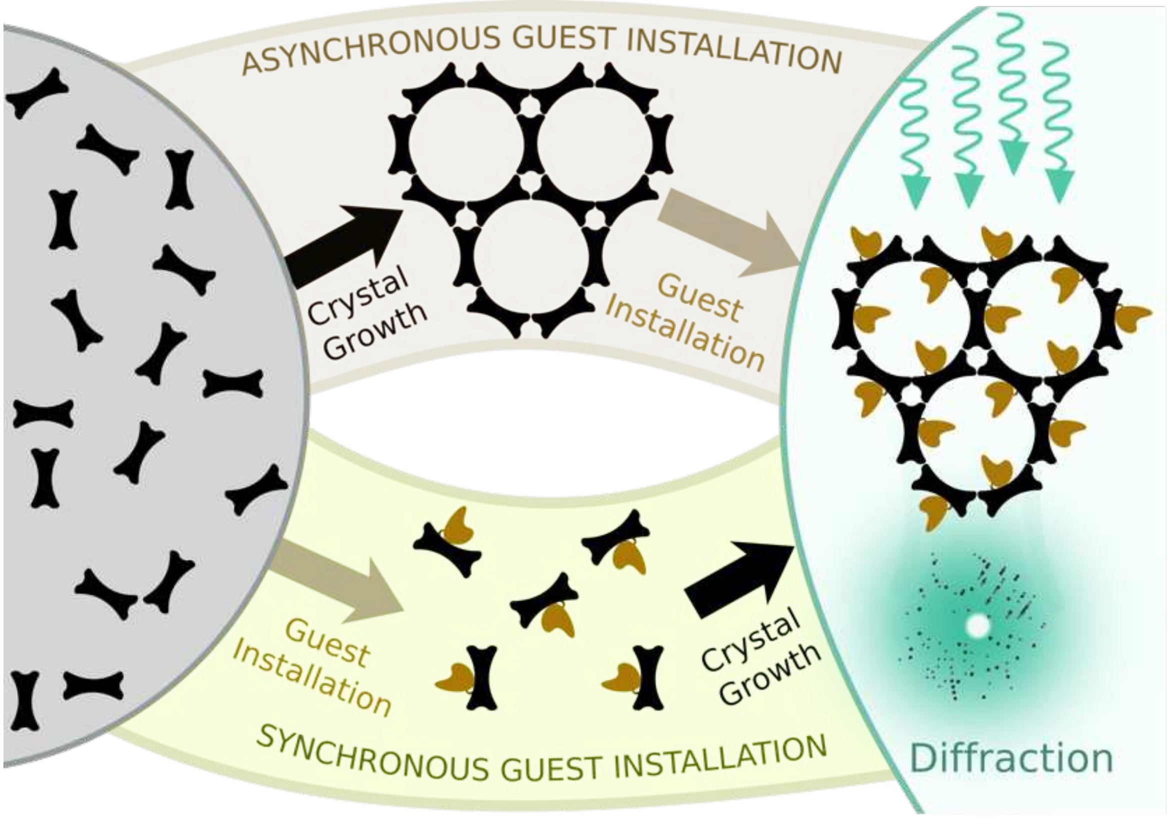


Figure 1.2. Guests can be installed before crystal growth (SGI) or after crystal growth (AGI).

On the other hand, AGI does create new challenges. Post-crystallization guest installation requires large solvent channels to enable intra-crystal diffusion. Also, it is not trivial to determine the yield of installed guests. From XRD alone, it is difficult to distinguish between a crystal in which 100% of the installation sites are occupied by disordered guests and a second crystal in which 10% of the sites are occupied by well-ordered guests.

1.3.3. Scaffold Stability

One less obvious benefit of AGI is the ability to stabilize the scaffold using reactive chemistry without also subjecting guest domains to the same reactions. Ultimately, stabilized scaffolds could dramatically increase the rate of structure acquisition, rapidly resolving the guest structure in a variety of solutions, including solvents that mimic native conditions. Rather than obtaining a single snapshot of the structure, researchers could conceivably observe hundreds of low-energy conformations. To realize this vision, scaffolds must maintain a precise structure (diffracting well) in a variety of solvents that differ from the idiosyncratic conditions used for crystal growth.

1.4. Existing Scaffold Crystals

1.4.1. Metal Organic Framework Scaffolds

Metal organic frameworks (MOFs) are three-dimensional crystalline materials generated by diverse combinations of metals and multivalent/symmetric organic ligands. As a structure determination scaffold, the MOF crystal ‘sponge’ method,[16] promoted by Fujita *et al.*, provides SASD for small molecule guests. Lacking designed attachment sites, can nonetheless join the MOF sponge in high occupancy under appropriate conditions, with sufficient binding of specific orientations to allow structure determination of chiral molecules.[12] Despite the existence of large-pore MOFs,[17] and MOFs that offer sites for guest attachment,[18] the prognosis for using MOFs enabling macromolecular SASD is unclear. MOF cages have entrapped small proteins, but the first example (an ubiquitin covalently tethered (cysteine to maleimide) to 1 of 24 organic ligands that forms a 63.5 nm^3 shell when complexed with Pd (II)) lacked interpretable electron density of the guest protein.[19] A yet-to-be-discovered MOF structure that offers large pores, reliable growth of single crystals, site-specific guest attachment sites, and stability in aqueous solvents may prove to be a compelling alternative to the biomolecular scaffold crystals described below.

1.4.2. DNA Crystal Scaffolds

Decades after the initial proposal of DNA crystal SASD, Paukstelis *et al.* discovered a notable 3D DNA crystal form.[20] The expanded version of these hexagonal crystals (Figure 1C), featuring 9 nm diameter solvent channels, were used to host small molecule fluorophores, act as a molecular sieve, and host enzymes.[21-23] Another important DNA crystal was successfully designed to self-assemble from triangular tiles.[24] These ‘tensegrity’ triangles combat the innate flexibility of DNA junctions by setting four-armed junctions at the triangle vertices.[25] Despite attempts to optimize junction and tile geometric details,[26,27] or co-crystallize two complementary tiles,[28] no reported variant on these scaffolds has simultaneously diffracted well (<3.5 Å) and had solvent pore structure suitable for guest macromolecule capture. New designed crystal forms from Yan *et al.*[29,30] may break this impasse, if they can be expanded while maintaining diffraction of ~3 Å or better. *Per* the MAP_CHANNELS software, the Zhang *et al.* scaffold may already allow diffusion of 3.8 nm diameter guests.[31] With regard to guest installation, DNA oligos can offer diverse non-natural functional groups for covalent guest conjugation. Inspiration may also be drawn from diverse groups who are using other polyhedral or low-resolution crystalline DNA assemblies to arrange third party biomolecules.[32-34]

1.4.3. Protein Crystal Scaffolds

While the hunt for improved DNA scaffolds continues, the PDB already contains diverse protein crystals that diffract reasonably well (<3.5 Å) despite having large solvent channels. Protein crystals can also be rationally engineered for improved stability and diffraction.[35] High specificity building blocks with a strong driving force to adopt a pre-determined lattice may be obtained by re-engineering existing crystal forms. Alternatively, 'promiscuous assembly' blocks that readily adopt new lattices, or a multitude of pre-determined lattices, could also prove useful for SGI. An exemplary instance of this latter strategy is the conversion of GFP to a scaffold protein (capable of adopting 33 different crystal forms) by introducing surface cysteines and partial metal-binding sites.[36]

In an early protein crystal SASD success story, a micro-peroxidase (a 9-amino acid heme-conjugated peptide, a proteolytic digest fragment of cytochrome cb562) was co-crystallized, bound to an engineered, exposed histidine *via* the heme iron.[11] The 1.9Å resolution dataset was sufficient to unambiguously place the heme and all 9 amino acids, despite the intrinsic peptide flexibility and the lack of predetermined scaffold design. The histidine-heme iron interaction, and interactions with a bed of hydrophobic residues (including five engineered sites) were likely necessary, though perhaps not sufficient to immobilize the guest. Multiple adventitious hydrogen bonds were likely important for full guest immobilization. The same scaffold is not suitable for AGI since macromolecules cannot diffuse through this crystal form nor access the inside of the protein cages.

Next-generation scaffolds may benefit from recent breakthroughs in protein cage engineering through symmetry, strong surface interaction motifs (*i.e.,* disulfides and metals), and computational design.[37,38] However, guest installation sites on the cage interior will often face space limitations. Hemocyanin crystals (Figure 1.1D) face the same challenge. Cage exterior installation sites could work in some cases, if the guests are added after crystallization (AGI), or if installation sites avoid crystallographic interfaces.

1.4.4. Hybrid Scaffolds

The search for a scaffold with improved functionality may lead researchers to embrace the greater complexity of co-crystalline scaffolds so as to combine favorable aspects of multiple types of building blocks. Early examples include protein/DNA assemblies,[39] protein/DNA/metal assemblies[40] or modular proteins/metal/organic linker assemblies.[41,42]

1.5. Guest Attachment Schemes

A literal and figurative weak point of the strategies above is the connection between the scaffold and the target molecule. How often will covalent guest fusion to a building block disrupt self-assembly? How should the installation site environment be crafted so as to create a single free energy minimum for the guest conformation? What guest installation strategy requires the least guest modification, and could be confidently applied to diverse guests of unknown structure? Current experimental examples do not definitively answer these questions.

Perhaps the most obvious strategy for attaching a guest protein to a protein crystal scaffold is to fuse the two proteins at the genetic level (necessitating SGI). This approach (Figure 1.3A) was used successfully by Maita to resolve the structure of guest ubiquitin fused to the C- termini of the R1EN scaffold protein.[8] No modification of the guest protein (beyond fusion to the scaffold) was required. As shown by Maita, scaffold to guest linker length and guest-guest contacts can play a profound role in guest immobilization. A preferred guest orientation was visible for linkers of 3, 5, or 7 aa but not 11 aa. However, the favored conformation clashes with one of two guest symmetry neighbors (Figure 1.4B), leading to static disorder and partial occupancy (50%). For the intermediate linker constructs (5 or 7 aa), a second guest conformation was visible. It is also notable that Maita's linkers were flexible (scaffold GG plus GGS, GTTGS, or GTTGEGS). The odds of assembling a rigid scaffold guest fusion might be increased by concatenating scaffold and guest alpha helices,[43] though this could prevent the extensive adventitious scaffold interactions (Figure 1.4C) that helped immobilize guest ubiquitin. Maita proposed that comparable adventitious interactions were lacking for fused guest SUMO (small ubiquitin-like modifier) domains, preventing successful observation thereof.[8]

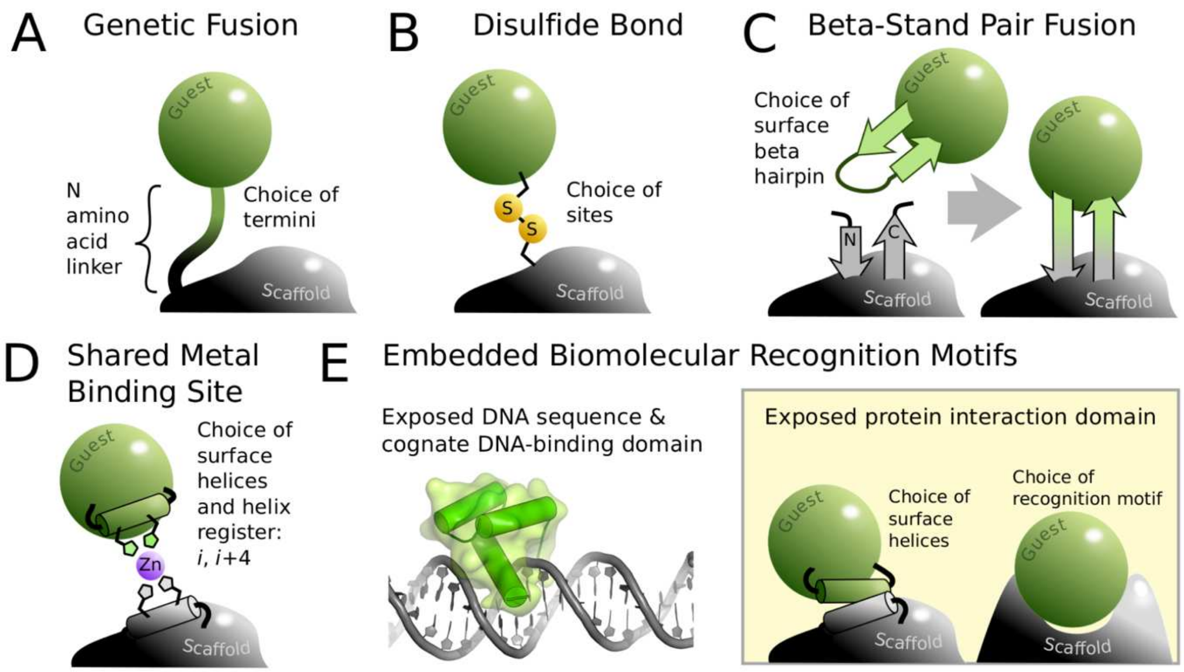


Figure 1.3. Existing and candidate guest installation strategies differ with respect to guest attachment affinity, the degree of required guest modification, and the number of easily rotated bonds. **(A)** Genetic fusion covalently links a target protein to the scaffold protein. Theoretically, linkers might be rigid, or of minimal length. Alternately, as shown by Maita, flexible linkers with varying length can be tested.[8] **(B)** For greater installation site flexibility, the scaffold and guest surfaces can be functionalized with individual exposed cysteine sidechains. Notably, thiols offer conjugation possibilities beyond the illustrated disulfide bond. **(C)** One possible future multi-anchor guest installation method is to genetically insert a suitable scaffold into a surface exposed beta hairpin on the guest.[48] **(D)** Incorporation of partial metal binding sites into both the scaffold and guest might provide a high affinity installation site with strong orientation constraints.[49] **(E)** Finally, future scaffolds might be engineered to rigidly embed known biomolecular binding domains with tunable specificity.[1]

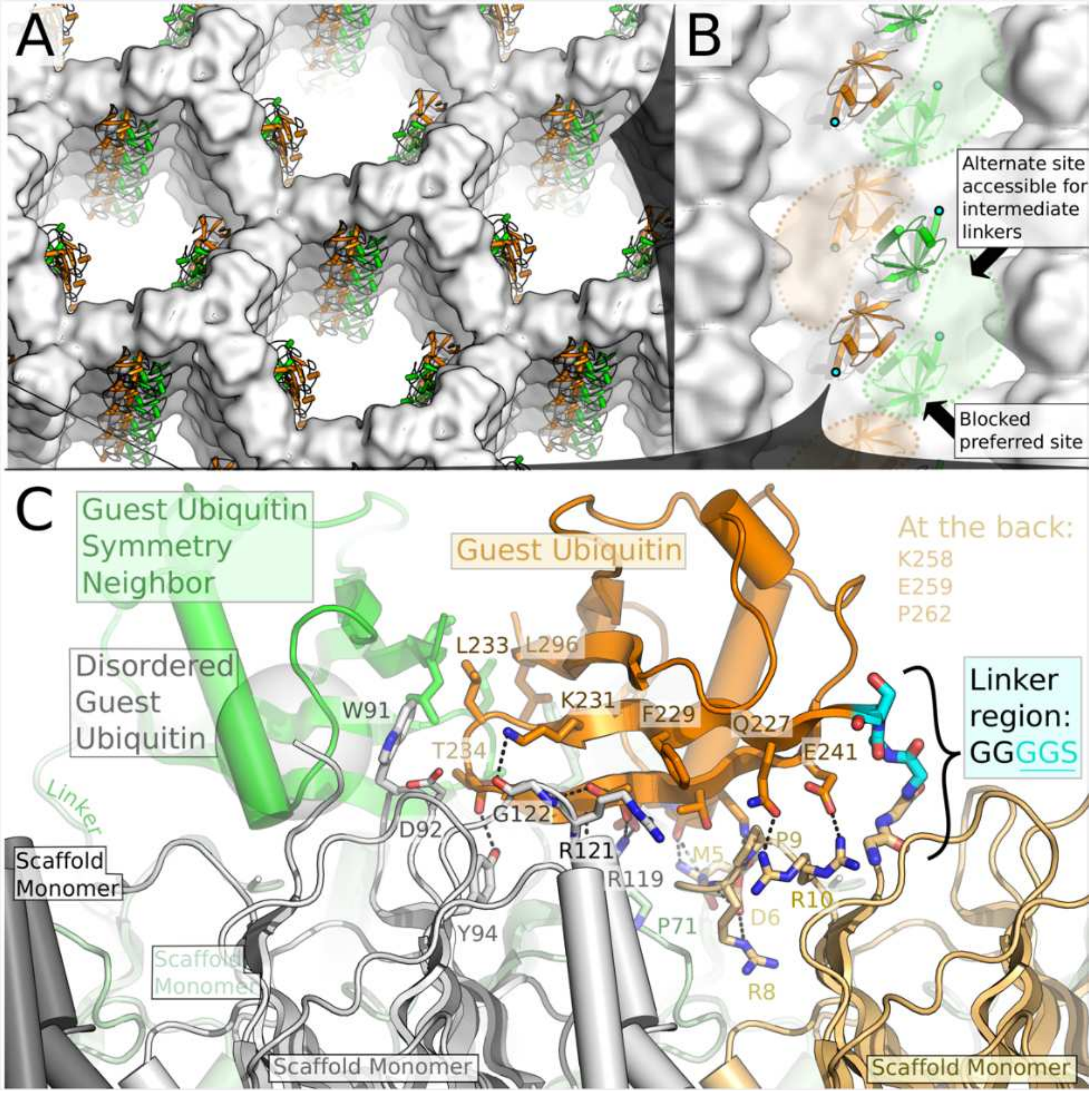


Figure 1.4. Synchronous guest installation within a porous protein crystal by Maita (PDB entry 6a42).[8] **(A)** Ordered binding sites for tethered ubiquitin domains occupy a small portion of the 10.5 nm diameter pores **(B)** A model for the partial occupancy and static disorder in the crystal induced by the inability for all guests to simultaneously adopt the preferred binding site (steric occlusion). **(C)** A highly flexible linker (cyan) allows the tethered ubiquitin to make numerous adventitious interactions (salt bridges, hydrogen bonds, hydrophobic) with scaffold monomers (tan, white, pale green) and a symmetry related neighbor (bright green).

Another covalent attachment strategy is to capture guest molecules using disulfide bonds (Figure 1.3B). An obvious advantage is the ability to create families of scaffold crystals with varying anchor sites. Guest modification will typically be required for conjugation; a cysteine could be placed at exposed N-termini or C-termini, or at sites predicted to be solvent exposed. While disulfide bonds have dihedral angle preferences,[44] it is still critical to avoid guest flexibility. When Huber *et al.* attempted structure determination for three scaffold cysteine sites with 4 small molecule conjugates, interpretable guest electron density was apparent for only about half of the cases.[45] Multiple conformations were also seen by Ueno *et al.* when using maleimide chemistry to conjugate small molecules (*e.g.,* fluorescein) to a cysteine facing the 4 nm diameter pores formed in *P*6 myoglobin.[46] Not surprisingly, fewer rotatable bonds and placement of bulky rings closer to the scaffold appeared to improve the odds of guest immobilization.

How might we improve the odds of immobilizing covalently linked guests? Surrounding scaffold sites could be mutated to create binding patches with reduced entropy,[47] charge, or hydrophobicity. Partial metal binding sites near covalent installation sites could create strong adventitious contacts with guest histidines or acids. Perhaps guest proteins may be immobilized by a combination of covalent attachment and steric confinement within scaffold cages of varying shape and size.

Yet another approach, connecting the guest to the scaffold at multiple points, is simple in conception but challenging to implement without foreknowledge of the guest structure. Predicted guest features such as solvent accessible beta hairpins and alpha helices could provide a route forward. Fusing a guest beta hairpin to a scaffold beta strand pair (Figure 1.3C) has precedent in the work of Yeates and Waldo.[48] Alternately, a guest surface alpha helix could host the i, i+4 dual histidine motif often used to good effect by Tezcan *et al.*[49] Guests might thereby be anchored *via* two sidechains, completing a rigid metal site with the scaffold (Figure 1.3D). Strong interactions between metals and scaffold sidechains (histidine and cysteine) have already been used by Ueno *et al.* for the detailed structure determination of gold nanocluster formation inside apo-ferritin crystals.[50] Rigidity is certainly critical; Capture of gold nanoparticles *via* flexible linkers did not lead to observable guest nanoparticle electron density.[51]

Lastly, we consider scaffold motifs for specific non-covalent capture of guests (Figure 1.3E). A key theoretical advantage of a scaffold with isolated DNA struts is modularity, freely varying the DNA sequence to match the binding specificity of target guest molecules. Protein crystals might be similarly equipped with exposed modular protein binding motifs (*e.g.,* coiled-coil motifs). A future crystalline scaffold with enough free space might rigidly incorporate domains (*e.g.,* immunoglobins or DARPins) capable of binding a variety of guests. Along these lines, Ernst *et al.* very recently engineered a generic crystalline lattice by fusing an N-terminal endo-a-N-acetylgalactosaminidase domain to a C-terminal designed ankyrin repeat protein domain. Modular target-binding of the latter domain allowed target guests (two 15-mer peptides and sfGFP) to be recruited into the lattice.[52]

1.6. Final Outlook

Further development is clearly needed in several key areas if SASD is to become a popular approach across the structural biology community. It is most critical to identify strategies that more *reliably* induce installed guests to adopt a uniform orientation with respect to the host crystal. Also, next-generation scaffolds that diffract well despite larger guest capacity are needed, as none of the currently reported crystalline scaffolds are suitable to host large proteins or complexes at isolated installation sites. Nevertheless, rapid SASD progress may be expected given the recent successful proof of concept work. Investments in the discovery, design, or optimization of scaffolds, or superior methods for guest immobilization are more justified than ever, given the promise of a new route for high-resolution, high-throughput structure determination.

CHAPTER 2. ENGINEERED PROTEIN-DNA CO-CRYSTALS AS SCAFFOLDS FOR DNA-BINDING MOLECULES[2]

2.1. Overview

X-ray crystallography of biomolecules is challenging, especially for DNA-binding molecules. Scaffold-assisted structure determination has potential to circumvent the crystallization challenge. Here, we engineered protein-DNA co-crystals to serve as scaffolds for DNA-binding molecules. The designed crystals, Isoreticular Co-Crystals, contain DNA-binding protein and cognate DNA blocks where the DNA-DNA junctions stack end-to-end. Furthermore, the crystal symmetry allows topology preserving (isoreticular) expansion of the DNA stack without breaking protein-protein contacts, hence providing larger solvent channels for guest diffusion. Experimentally, upon optimizing the crystal growth condition, the designed porous lattice was achieved. The Isoreticular Co-Crystal design offers a unique approach to a *post*-crystallization programmable scaffold for DNA-binding molecule and scaffold-assisted X-ray crystallography.

2.2. Introduction

Given a suitable biomolecular crystal, X-ray diffraction (XRD) remains the most routine and detailed method for elucidation of atomic structural details. Unfortunately, the

---

[2] Parts of the work in this chapter are formatted for submission as a research article. Writing-drafting, visualization A.R.O., C.D.S; Conceptualization and formal analysis and writing-editing, A.R.O., E.T.S., C.D.S.; data curation, acquisition, investigation, and validation; A.R.O., E.T.S., S.D., A.V., C.D.S.
Orun, A. R., Shields, E. T., Dmytriw, S., Vajapayajula, A., Snow, C. D. Engineered protein-DNA co-crystals as scaffolds for X-ray diffraction of DNA-binding molecules. *Pending submission.*

task of growing a highly ordered biomolecular crystal requires tedious sample preparation and haphazard crystal formation.[52] Crystallization is especially daunting in the case of DNA-binding entities, where crystal growth may be hindered by the dynamic DNA-binding event[53] and successful crystal growth can depend on the length and junction details for the DNA blocks.[54,55] Scaffold-assisted crystallography is an alternative, potentially high throughput strategy for co-crystallization of DNA-binding molecules. A scaffold crystal avoids the crystallization problem by providing a robust, high-quality single crystal that is self-assembled consistently and independently of the target biomolecule. In principle, guest molecules that adopt a uniform conformation with respect to the crystal lattice can also be revealed with X-ray diffraction. We had previously reviewed this concept of scaffold-assisted structure determination.[56] Here, we show a designed, expanded co-crystal scaffold for DNA-binding guests.

Engineered crystalline scaffolds capable of post-crystallization macromolecule installation will have diverse applications.[3-7,13,21-22,57-58] Several research groups have sought to design DNA crystals with utility as a programmable scaffold.[1,20,24,29,59-60] One ongoing challenge of DNA crystals is the limited X-ray diffraction resolution for DNA crystals that have solvent pores of sufficient diameter for guest macromolecule transport. We hypothesize that the diffraction resolution may be limited by the background noise produced by additional solvent content and perhaps the counterion requirements for close DNA packing. Other groups have evaluated highly porous protein crystals for scaffold-assisted X-ray diffraction of guest molecules.[8,45,61] Fusing guest proteins has proven successful, yet haphazard.[8] It is a challenge to optimize the right linker length for genetic fusion of the guest to lock the guest in place for XRD.[61] A larger group of researchers have championed the use of crystallographic chaperones, where a protein of interest is bound or fused to a protein domain that is crystallization prone or provides a strong lattice contact.[62-64] Even DNA can serve as a chaperone driven by sticky end cohesion.[65-66] Furthermore, Cote *et al.* complexed a reverse transcriptase N-terminal fragment to nucleic acid duplexes for chaperoned structure determination of unique DNA duplexes.[67] Here, we will use a narrow definition of scaffold-assisted structure determination that excludes the chaperone approach by insisting that the scaffold lattice does not vary with the guest. If the scaffold component alone is responsible for the final lattice the processing of the X-ray diffraction data is simplified, the scaffold can be optimized independently of guests, and the scaffold crystal may even be assembled and engineered prior to guest installation.

Protein-DNA co-crystals offer unique design opportunities by taking advantage of the distinct properties of each component. DNA is relatively inexpensive and quick to synthesize, making it possible to test numerous designed crystal variants. Furthermore, under the right conditions, the structure of B-DNA and the canonical bases is predictable.[68] Modeling and computational design tools for DNA are well developed such as 3DNA[69] and NUPACK,[70] allowing detailed structure prediction. Another advantage of employing DNA building blocks is the ease of incorporating non-canonical nucleotides including dyes and other unique chemical moieties. Meanwhile, the protein components may offer additional advantages, such as a reduced sensitivity to counterions. In designed protein-DNA co-crystals, we hypothesize that using protein components will remove the necessity of incorporating Holliday junctions into the DNA and remove the necessity of close packing DNA. Instead, our target lattice will consist of contiguous DNA double helices protein “insulated” by scaffold proteins. The protein portion of the scaffold also offers additional guest installation sites (*e.g., via* accessible thiols, fusion to the N- or C-termini, or insertion into loops) and greater control over the physicochemical composition of the scaffold crystal interior (given the larger variety of amino acids compared to nucleotides).

Our designed scaffold co-crystal has modular DNA sequence, designable DNA-DNA junctions, and large solvent channels to diffuse DNA binding molecules that range from small molecules to proteins in size. There are numerous examples of co-crystals already present in the PDB. Existing co-crystals are often stabilized *via* DNA-DNA stacking interactions at crystallographic interfaces. Sometimes these end-to-end base pair stacking interactions are aligned with the crystallographic symmetry axes so as to permit re-design of the DNA-DNA junction without perturbing the protein-protein interfaces. There are copious options for re-engineering the co-crystal lattice by simply re-engineering the DNA-DNA junctions. Tunable interactions include sticky base overhangs, sequence, and terminal phosphorylation. For example, the Mao group showed terminal 5’ phosphates in pure DNA crystals produce sticky end cohesion and drive crystal growth.[71] Critically, when the target co-crystal lattice allows the DNA sequence to become a design variable, a variety of engineered scaffold crystals might be synthesized, each incorporating accessible DNA sequences to anchor guest DNA-binding molecules.

2.3. Results and Discussion

Appendix I contains corresponding supplemental figures to Chapter 2.

2.3.1. The Isoreticular Co-crystal Design

Our designed protein-DNA co-crystals consist of a scaffold DNA-binding protein and cognate DNA sequence (Figure 2.1). The unique DNA junctions enable insertion of DNA at the junctions for a periodic, isoreticular expansion of existing co-crystals. To design a porous isoreticular co-crystal (ICC) we first sought a DNA-protein co-crystal lattice where (1) coaxial dsDNA:dsDNA contacts are the primary contacts enabling growth of the crystal in at least one dimension, (2) these DNA:DNA junctions align with the crystallographic symmetry so that isoreticular (topology preserving) expansion of the

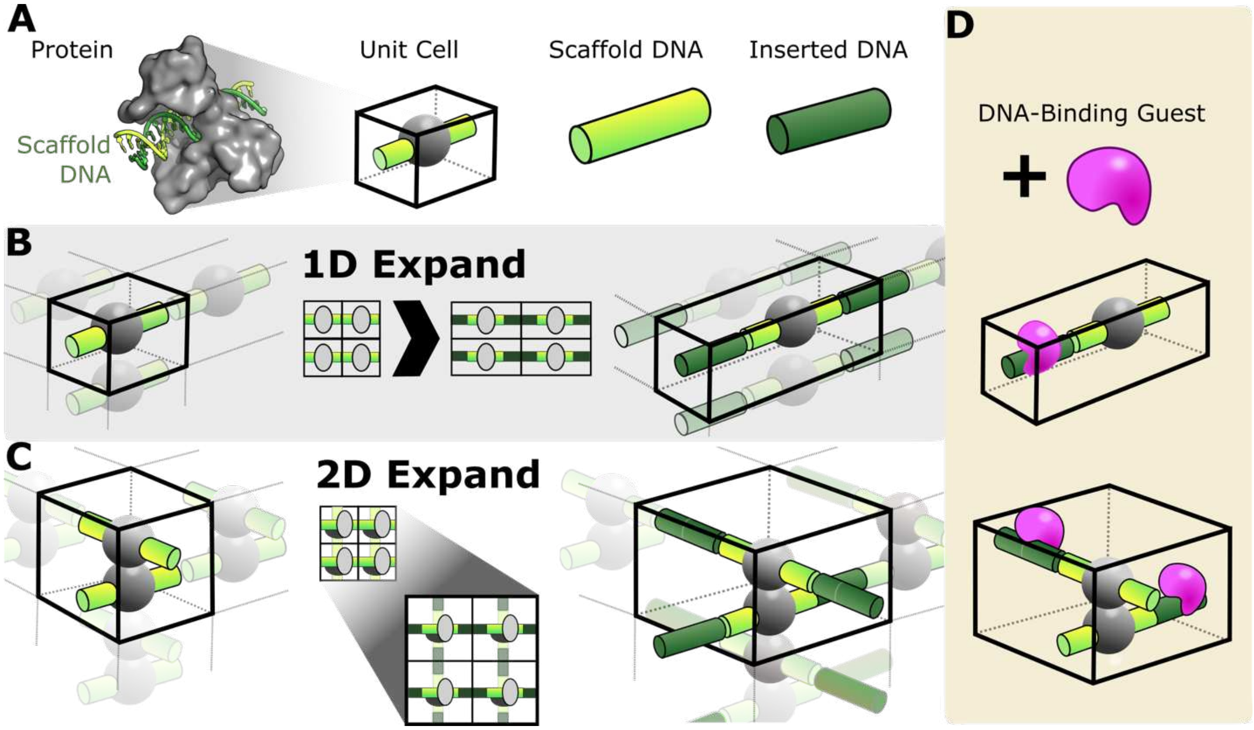


Figure 2.1. The Isoreticular Co-Crystal is built of a scaffold protein and DNA, (A) represented with a grey sphere and green rod. Inserted DNA (dark green rod) can be added at the current DNA-DNA junctions in the crystal. Insertion lengths that respect the DNA helical twist may allow junctions to be stabilized by coding sticky overhangs. (B) The simplest example of crystal expansion would occur in a 1D expansion, with parallel DNA stacks throughout the crystal. (C) 2D expansion may occur when DNA-DNA junctions mediate growth in two dimensions. (D) The inserted DNA provides an anchoring point for a DNA-binding guest of interest, thereby enabling scaffold-assisted X-ray diffraction structure determination.

DNA blocks could occur without altering other important crystallographic contacts, and (3) any important protein-protein contacts are orthogonal to the DNA-DNA contacts and might therefore be preserved during assembly of the resulting ICC.

We identified candidate co-crystal building blocks by filtering Protein Data Bank (PDB) entries (as of 01/2017) derived from X-ray crystallography. To this end, we used the PDB advanced search and custom Python scripts. We excluded entries that were very large (> 1 Mb) or lacked sufficient DNA content (less than twenty O5' atoms) or lacked sufficient protein content (less than 19 amino acids). We also excluded candidates with significant RNA content (more than 2 O2' atoms) or candidates with unknown atoms (UNK). After these content filters, we proceeded to filter the list using geometry, rejecting candidates that lacked stacked junctions between the DNA in the asymmetric unit and symmetry related DNA through inspection of pairwise distances between the ribose C1' atoms in terminal nucleotides. Candidate crystals were discarded unless there was a terminal C1' in the asymmetric unit and a nearby (3-7 Å) terminal C1' in a symmetry mate. Next, a more sophisticated script was used to identify DNA-DNA junctions spanning crystallographic interfaces in which both sides feature base paired dsDNA. First, we determined if the junctions were DNA stacks (end-to-end). We generated symmetry copies of the C1' positions, calculated distances between asymmetric unit C1' and the symmetry copy C1' and flagged possible stacking C1' pairs as those between 3-7 Å. Further, we analyzed termini base pair status. Base paired termini included structures in which C1' distances between terminal bases within the same asymmetric unit were within 9-12 Å. The last part of this script computes the step rotations at the termini to further assess the junctions and output a structure with highlighted C1’ features (Figure S2.1). Finally, we collected the structures in which there were stacked junctions containing duplexes that were base paired at these junctions.

The resulting list of 609 PDB entries was visually assessed. Specifically, we used PyMOL and the supercell script[72] to rapidly inspect the lattice contacts for each candidate and to flag candidates where expansion of the DNA stacks might occur without breaking protein-protein contacts. Most of the candidates that appeared to be suitable for isoreticular expansion were cases in which all the DNA blocks in the crystal were aligned in the same direction (1D expansion candidates). A minority of the candidates had symmetry such that DNA block insertions might expand the crystal in two dimensions, thereby rapidly increasing the diameter of the crystal solvent channels. We have yet to find an expansion candidate expandable in three dimensions, although this should be possible.

2.3.2. Co-crystal 1 Expansion

In this chapter, we highlight the 2D expansion candidate composed of RepE54 replication initiation protein and cognate DNA sequence (21-mer) (Figure 2.2).[73] For convenience we name this system co-crystal 1 (CC1). We recently reported an updated PDB model for CC1 with improved resolution to 1.89Å (PDB code: 7rva), while optimizing a protocol to dramatically stabilize such crystals using chemical ligation.[74] The asymmetric unit for the original, non-expanded CC1 system contains only one copy of the protein bound to the cognate dsDNA. In accord with the *C1*21 space group, both ends of the dsDNA form DNA-DNA junctions that span crystallographic interfaces between the asymmetric unit and its symmetry mates (Figure 2.2F). Notably, the blunt end DNA-DNA junctions stack closely, resembling contiguous DNA. One important difference from contiguous DNA is that the best non-expanded CC1 structure has dangling 3’ thymine residues. The same crystal lattice grew with blunt ended DNA-DNA junctions if we removed the dangling T bases (albeit with slightly lower X-ray diffraction resolution). There are two protein-protein interfaces in the CC1 unit cell (Figure 2.2C). During expansion of the DNA-DNA junctions, we aimed to preserve the existing protein-protein contacts.

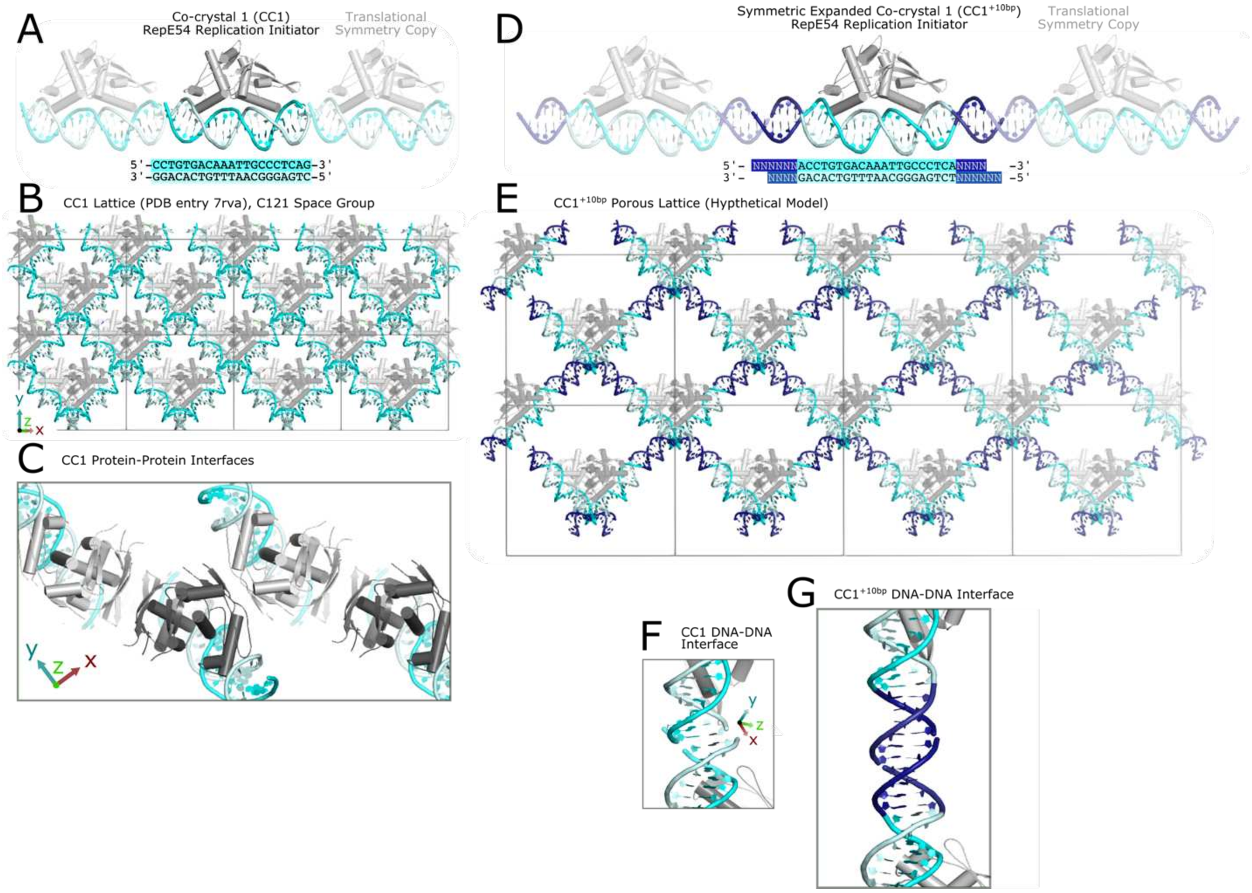


Figure 2.2. The asymmetric unit of Co-crystal 1 (A) is the complex of RepE54 replication initiator protein and a cognate 21-mer binding sequence with dangling T’s (PDB code: 7rva). The CC1 lattice (B) is in a C121 space group and there are two protein-protein interfaces (C) conserved throughout the lattice. Upon expansion of the DNA with 10 additional base pairs (D) the lattice is highly porous (E) and the DNA-DNA junctions (G) are modular and free for DNA-binding guest targets.

To achieve an isoreticular expanded co-crystal and insert another twist in the DNA at the junction, we truncated the dangling Ts that serve to anchor the strands on the sides of the DNA, and we added DNA bases. The geometric parameters of the blunt ended junction with terminal phosphates were analyzed using x3DNA[69] (Table 2.1). Explicit modeling of possible DNA inserts (Protocol S1 and published to Zenodo[75]) suggested that a 10 base pair insertion would approximate a contiguous dsDNA geometry at the DNA:DNA junction ($CC1^{+10}$). The target crystal lattice would be quite porous with a solvent fraction of 80% (Figure 2.2, Protocol S2.2) and solvent channels large enough to permit transport of guests with diameter of 5 nm (*per* MAPCHANNELS[31]).

Table 2.1. DNA-DNA junction geometry parameters in different versions of the CC1 native crystal. The original CC1 contains 3’ terminal dangling thymines. For expansion, these thymines were truncated for two blunt ended versions: CC1-5’p and CC1-3’p.

Junction Parameters from x3dna	**CC1 (original)**	**B-DNA***	**CC1-5′P (blunt-ended)**	**CC1-3′P (blunt-ended)**
PDB code	7rva		7sgc	7sdp
Base pair step parameters	GC/CG		GC/CG	GC/CG
Shift (Å)	−0.03	0.0 ± 0.51	0.10	0.05
Slide (Å)	−0.81	0.35 ± 0.78	−1.00	−2.03 [†]
Rise (Å)	3.49	3.32 ± 0.19	3.78 [†]	3.56
Tilt (°)	1.87	0.0 ± 3.4	2.51	0.72
Roll (°)	1.25	1.4 ± 5.1	1.94	4.03
Twist (°)	36.64	35.4 ± 6.3	39.10	28.18

*Base pair step parameters from Olson *et al.* 1998.[76] C5′ to O3′ distance from x3DNA idealized B-DNA.

[†] Values that differ from canonical B-DNA by more than 2 standard deviations.

2.3.3. Porous Co-crystal 1 Crystallization

Remarkably, we obtained crystals which featured the designed DNA-DNA junctions while maintaining the two original protein-protein interactions. This was in keeping with our goal to obtain “isoreticular” crystals with expanded DNA but maintaining the same lattice topology. Here, we show the preliminary porous crystal results.

We screened for crystallization conditions for the expanded CC1^{+10bp} with a T-A rich insert sequence (Figure 2.3) on the Gryphon crystallization robot using a 96-well crystallization screen (Hampton NatrixHT). To our excitement, we achieved a porous lattice using this random crystallization condition screening (Figure 2.3A). The porous crystals grew in a *I*121 space group and a parallelepiped habit. The initial XRD resolution was 4.2Å. To improve the diffraction, repeat crystals were grown. The crystal repeatably grew in a similar growth condition (Figure 2.3B), even after replacing the cacodylate compound stoichiometrically 1:1 with MES pH 6.5 for more environmentally and personnel friendly crystal handling. Further X-ray diffraction analysis is needed to confirm the average diffraction quality of the porous lattice CC1^{+10bp} with a T-A rich insert sequence.

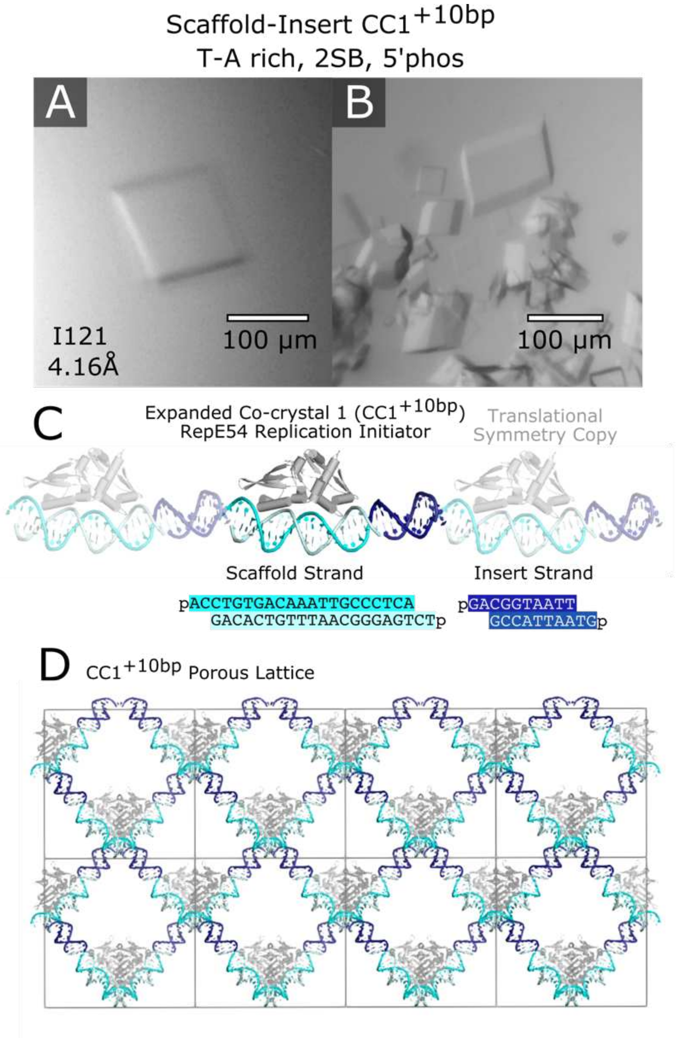


Figure 2.3. Crystallization of the target porous lattice was grown with a NatrixHT crystallization screen (Hampton). **(A)** The original crystal hit was grown in 0.01 M magnesium acetate tetrahydrate, 0.05 M sodium cacodylate trihydrate pH 6.5, and 1.3 M lithium sulfate monohydrate. **(B)** Repeat crystals grew, even when replacing sodium cacodylate trihydrate pH 6.5 with a stoichiometric equivalent of MES hydrate. **(C)** The asymmetric unit contained the scaffold protein (RepE54 initiator protein) bound to a 21-base pair duplex (scaffold strand) and a 10-base pair T-A rich duplex (insert strand) to complete the co-crystal expansion. **(D)** A crystallographic view of the experimentally validated porous lattice.

As intended, the expanded lattice had increased solvent channel diameter. To assess solvent channel geometry more rigorously, we used MAP_CHANNELS.[31] For the preliminary porous crystal, a sphere with diameter of 5.7 nm would be expected to be able to diffuse along the *x*-axis, whereas only guest spheres with diameter less than

3.9 nm would be expected to be able to diffuse laterally along the *y*- or *z*-axes. Accordingly small proteins may be able to diffuse into these crystals in all directions, whereas medium or large proteins would likely be restricted to single-file diffusion along the *x*-axis. The shape of the guest molecule will also be important, and we can expect varying transport behavior depending on whether the bound guest molecule obstructs the diffusion of another guest molecule. For example, green fluorescent protein (27 kDa; PDB code 1gfl) would hypothetically be able to diffuse through the *x*-axis pores. Next experiments include hosting small DNA-binding proteins in the porous crystals. Once published, our porous crystal would be the first example of a crystal host for XRD of DNA-binding proteins.

2.3.3. Porous Co-crystal 1 Modularity

We demonstrated that assembly of the target expanded lattice (31-mer rather than 21-mer) was modular by growing crystals using three different DNA expansion strategies (Figure 2.4). Symmetric expansion (Figure 2.4B) used a single DNA block with 5 additional base pairs flanking the scaffold protein binding segment, thereby converting a 21-mer to a 31-mer. Asymmetric expansion (Figure 2.4C) involved a single DNA block with 10 base pairs added to one side of the scaffold protein binding site. Lastly, the 2-part scaffold-insert expansion strategy (Figure 2.4D) used two separate DNA blocks, one scaffold duplex that consisted of the scaffold protein binding site (21-mer) and one insert duplex (10-mer). All three strategies led to growth of crystals (Figure 2.5 and Table 2.2) and X-ray diffraction analysis is needed to confirm all schemes are growing in the same symmetry group.

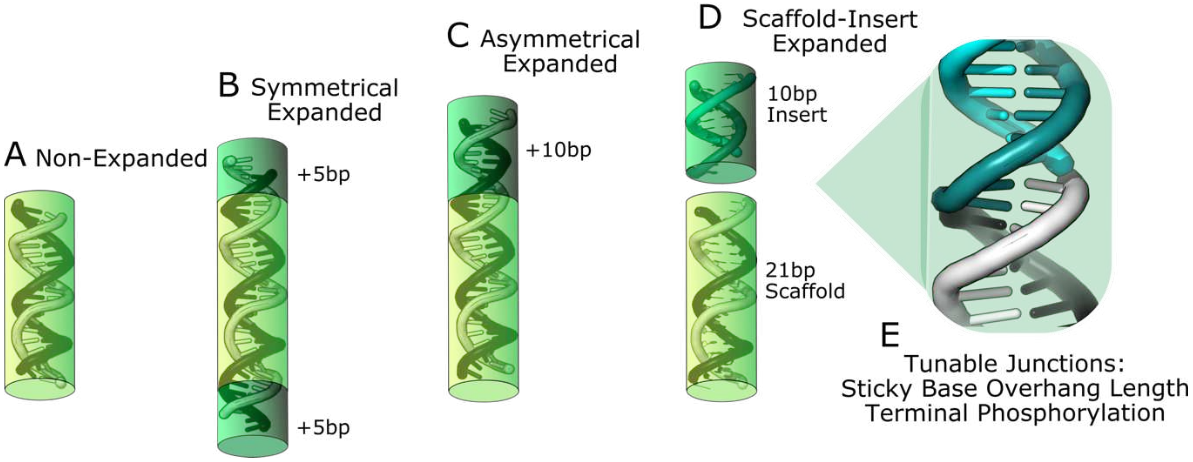
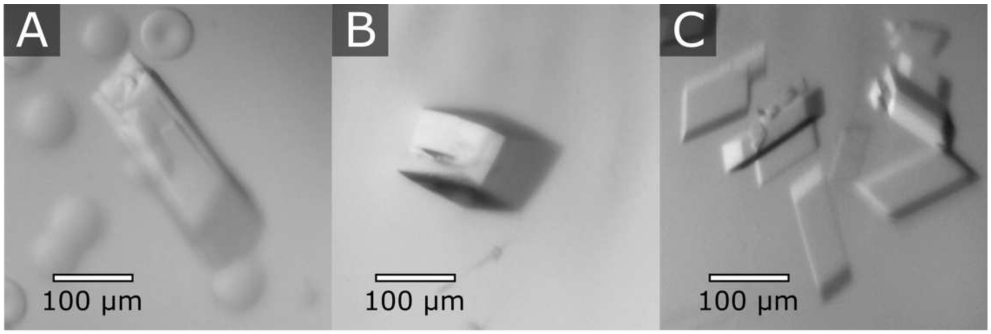


Figure 2.4. The Isoreticular Co-crystal design schemes were based on variation from the original, non-expanded duplex (A). The symmetrical expansion (B) added 5 base pairs to each side of the original duplex. The asymmetrical expansion (C) added 10 base pairs to a single side of the original duplex. The scaffold-insert expansion (D) maintained a 21 base pair scaffold strand and added a 10 base pair insert strand with matching sticky overhangs to the scaffold strand. In each expansion scheme, the DNA-DNA junctions were tunable with varied sticky base overhang lengths and terminal phosphorylation (no phosphate, 3'p or 5'p).

For isoreticular co-crystals to realize their potential as a general-purpose scaffold crystal, it is crucial to reliably assemble into the intended lattice while making changes to the *sequence* of the inserted DNA segment (Figure S2.2). To demonstrate reliable assembly regardless of the inserted DNA sequence, we grew crystals using both the symmetric and scaffold-insert strategies with two dramatically different inserted sequences: a G-C rich sequence (5'-GACGGCCCGG) and a T-A rich sequence (5'-GACGGTAATT) (Figure 2.5 and Table 2.2). Crystals consistently grew with varied sequence in a parallelogram habit. Furthermore, we improved the resolution of the scaffold crystal, with the record resolution 3.2Å. The resolution and modularity of the porous scaffold offers exciting prospects for post-crystallization guest structure determination.

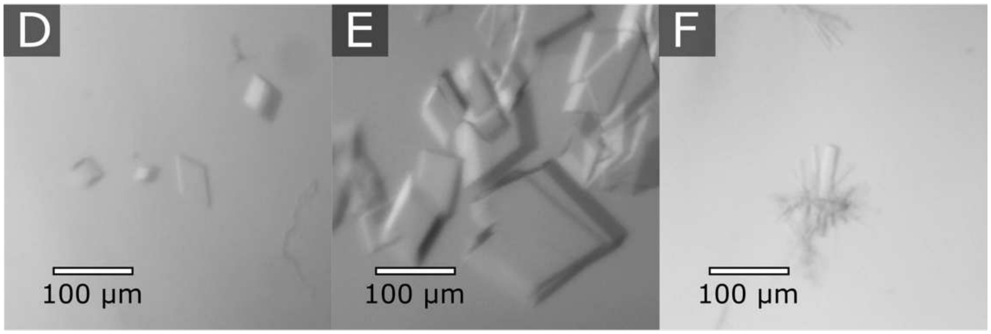
Another tunable feature at the DNA junction is the sticky overhang length (blunt ended, 1nt, 2nt, 3nt, *etc.*) and terminal phosphorylation status (no phosphate, 3’ or 5’ phosphate) (Figure 2.4E). So far, we have grown the porous crystals with blunt ends, 1-nt, 2-nt, and 3-nt sticky base overhangs and 5’ or 3’ terminal phosphates (Figure 2.5 and Table 2.2).



Symmetric CC1^{+10bp} G-C rich, 0SB, 3'phos

Symmetric CC1^{+10bp} G-C rich, 0SB, 5'phos

Symmetric CC1^{+10bp} G-C rich,1SB, 3'phos



Symmetric CC1^{+10bp} G-C rich, 3SB, 3'phos

Symmetric CC1^{+10bp} T-A rich, 2SB, 5'phos

Scaffold-Insert CC1^{+10bp} G-C rich, 2SB, 5'phos

Figure 2.5. (A-D) Porous CC1^{+10bp} variants grown to date include: Symmetric expansions with G-C rich sequences, varied sticky bases (0SB to 3SB) and 5’ or 3’ terminal phosphates, **(E)** symmetric expansion with T-A rich sequence and 2 sticky bases with 5’ terminal phosphates, and **(F)** preliminary growth of a scaffold-insert G-C rich sequence with 2 sticky bases and 5’ terminal phosphates.

Table 2.2. Porous CC1^{+10bp} building units for expanded crystal growth, including the DNA expansion scheme, sequence, sticky base overhangs, terminal phosphorylation, crystallization conditions and preliminary XRD resolutions (structure factor data were truncated using I/sigma(I) >1.5). Full duplex sequences are listed in Table S2.1.

Figure Panel	Expansion Scheme and Sequence	DNA-DNA Junction	Crystallization Condition	XRD Resolution
2.5A	CC1^{+10bp} Symmetric G-C rich (1:1.2 RepE54:DNA)	0SB 3′P	0.03 M magnesium acetate tetrahydrate 0.05 M MES pH 6.5 1.8 M lithium sulfate monohydrate	N/A
2.5B	CC1^{+10bp} Symmetric G-C rich (1:1.2 RepE54:DNA)	0SB 5′P	0.01 M magnesium acetate tetrahydrate 0.05 M MES pH 6.5 2.0 M lithium sulfate monohydrate	4.79Å
2.5C	CC1^{+10bp} Symmetric G-C rich (1:1.2 RepE54:DNA)	1SB 3′P	0.03 magnesium acetate tetrahydrate 0.05 M MES pH 6.5 1.0 M lithium sulfate monohydrate	3.24Å
2.5D	CC1^{+10bp} Symmetric G-C rich (1:1.2 RepE54:DNA)	3SB 3′P	0.02 M magnesium acetate tetrahydrate 0.05 M MES pH 6.5 1.3 M lithium sulfate monohydrate	N/A
2.5E	CC1^{+10bp} Symmetric T-A rich (1:1.2 RepE54:DNA)	2SB 5′P	0.02 M magnesium acetate tetrahydrate 0.05 M MES pH 6.5 1.3 M lithium sulfate monohydrate	4.33Å
2.5F	CC1^{+10bp} Scaffold-Insert G-C rich (1:1.2:1.2 RepE54: Scaffold:Insert)	2SB 5′P	0.03 M magnesium acetate tetrahydrate 0.05 M MES pH 6.5 1.6 M lithium sulfate monohydrate	N/A
2.3A	CC1^{+10bp} Scaffold-Insert T-A rich (1:1.2:1.2 RepE54: Scaffold:Insert)	2SB 5′P	0.01 M magnesium acetate tetrahydrate 0.05 M sodium cacodylate trihydrate pH 6.5 1.3 M Lithium sulfate monohydrate	4.16Å
2.3B	CC1^{+10bp} Scaffold-Insert T-A rich (1:1.2:1.2 RepE54: Scaffold:Insert)	2SB 5′P	0.03 M Magnesium acetate tetrahydrate 0.05 M MES pH 6.5 1.5 M Lithium sulfate monohydrate	N/A

2.4. Conclusions

In theory, the isoreticular co-crystal design can be applied to existing co-crystals in which the DNA blocks stack up in the crystal and the protein interactions are not disturbed by DNA expansion. One-dimensional expansion candidates, in which the DNA stacks up parallel throughout the crystal, may not be as highly porous but may be useful for the display of nanoarrays on the crystal surface with protruding DNA ends. Two-dimensional expansion candidates, such as the CC1 crystal expansion shown, offer highly porous solvent channels for guest macromolecule diffusion.

While the porous lattice results are preliminary, future scaffold-assisted X-ray diffraction experiments are promising. CC1 is the first example of an isoreticular co-crystal, a crystal expanded with additional DNA building units. The sequence modularity and porosity from 10 additional base pairs brings us one step closer to host DNA-binding proteins for structure determination. Furthermore, to make the porous crystals solution-independent, the crystals could be crosslinked using the chemistry in Chapters 4 and 5. Soon, co-crystals may be developed into robust, general-purpose DNA-binding molecular scaffolds.

2.5. Materials and Methods

2.5.1. Protein Cloning, Expression, and Purification

The cloning and expression of the scaffold protein, RepE54 replication initiator, was first described by Komori *et al.*[73] The CC1 RepE54 was produced at the Histone Source at Colorado State University exactly as described in Section 4.5.1. Protein purification steps were analyzed with SDS-PAGE (NuPAGE™ 4–12% Bis-Tris Gel) with MES SDS running buffer and stained with Imperial™ Protein stain. Bradford Assay using Coomassie Plus^{TM} Protein Assay Reagent was used to determine final protein concentrations.

2.5.2. DNA Duplex Annealing

The DNA oligomer sequences used for co-crystallization are listed in Table S2.1. Each oligomer contains the 19-bp iteron sequence for RepE54 protein-DNA binding, but the flanking DNA sequences vary depending on the expansion scheme and sequence. Individual oligomers were synthesized and HPLC purified by Integrated DNA Technologies and annealed in the Snow Lab. The oligomers were resuspended in CC1 oligo buffer (50 mM tris HCl, 100 mM KCl pH 7.0) and combined in equal molar ratio (1:1) with the complementary strand. The strands were annealed by heating to 94 °C for 2 min and slowly cooling to room temperature over approximately 60 minutes. The final concentration of all CC1 duplexes was 4 mM.

2.5.3. Scaffold Protein-DNA Complex Co-crystallization

The scaffold protein and DNA (1:1.2) were incubated on ice 30 minutes prior to crystallization *via* sitting drop vapor diffusion. Crystallization conditions for the porous lattice (CC1^{+10bp}) were originally found using the Colorado State University Gryphon and NatrixHT (Hampton) 96 well crystallization screen (Figure S2.3 and Table S2.2). The porous lattice was found to grow after 1 day in 0.01 M Magnesium acetate tetrahydrate, 0.05 M Sodium cacodylate trihydrate pH 6.5, and 1.3 M Lithium sulfate monohydrate. Follow up plates were crystallized by replacing sodium cacodylate trihydrate pH 6.5 with MES pH 6.5. Crystals grew to a size of 50 - 150 μm^3 in a range of 1 day to 7 days. Crystal pictures were taken with a Moticam 3.0 MP camera attached to a Motic SMZ-168 stereozoom microscope.

2.5.4. X-ray Diffraction Data Collection and Refinement

Crystals were flash frozen in liquid nitrogen using the growth conditions as the cryo-protectant. Single-crystal X-ray diffraction data was collected at the Beamline 4.2.2 using a CMOS detector from 0 to 180 degrees with an omega delta of 0.2° and an exposure time of 0.5 seconds. Data was processed with XDS.[77] The first porous co-crystal structure was solved by molecular replacement with the interpenetrating expanded CC1 (highlighted in Chapter 3) (PDB code: 7u6k; 2.38Å) and refined with PHENIX[78] and COOT.[79]

2.6. Funding

This material is based upon work supported by the National Science Foundation under Grant No. NSF DMR-2003748 and NSF DMR-1506219. The team also gratefully acknowledges support for undergraduate researchers from the Nelson Family Faculty Excellence Award.

CHAPTER 3. INTERPENETRATING, EXPANDED CO-CRYSTALS PROVIDE SEQUENCE-DEPENDENT SYNCHRONOUS GUEST INSTALLATION[3]

3.1. Overview

Biomolecular crystals have served as bio-vessels for sensing, bio-catalysis and inorganic templating. Protein-DNA co-crystals can be advantageous materials, too, with predictably designable DNA structure and robust protein-DNA and protein-protein interactions. To realize the potential of co-crystals as host crystals, we designed co-crystals with longer DNA sequences and porosity for guest molecule diffusion. Experimentally, some of the resulting designed co-crystal adopted an interpenetrating lattice, a phenomenon previously observed in metal-organic frameworks (MOFs). The interpenetrating lattice crystallized dependably in the same space group (*I*222) despite myriad modifications at the DNA-DNA junctions. Assembly was modular with respect to the DNA inserted for the expansion, providing an interchangeable DNA sequence for scaffold assisted X-ray diffraction studies. Also, the DNA-DNA junctions were tunable, accommodating varied sticky base overhang lengths and terminal phosphorylation. As a proof of concept, we used the interpenetrating scaffold crystals to separately entrap three distinct guest molecules during crystallization: guest protein, DNA and small molecule. The interpenetrating designed co-crystal could serve as a guest DNA-binding entrapment tool in applications such as drug delivery and bio-sensing.

---

[3] Parts of the work in this chapter are formatted for submission as a research article.Writing-drafting, visualization A.R.O., C.D.S; Conceptualization and formal analysis and writing-editing, A.R.O., E.T.S., C.D.S.; data curation, acquisition, investigation, and validation; A.R.O., E.T.S., S.D., A.V., C.D.S.
Orun, A. R., Shields, E. T., Dmytriw, S., Vajapayajula, A., Snow, C. D. Engineered protein-DNA co-crystals as scaffolds for X-ray diffraction of DNA-binding molecules. *Pending submission.*

3.2. Introduction

Here, we designed a scaffold co-crystal of protein and DNA that can precisely arrange guest molecules *via* programmed DNA-binding. The target crystal lattice contains two distinct protein-protein interfaces and one DNA-DNA interface. Engineered co-crystal variants consistently grew despite modifications at the DNA-DNA junction. Engineered co-crystals, albeit interpenetrating crystals with less porosity than the intended lattice, were found to host three diverse representative guest molecules (DNA, protein, small molecule), all installed *via* site-specific DNA binding during crystallization. Our crystals therefore pave the way for further crystal design with specific guest installation sites for applications like drug delivery, bio-sensing, or DNA-binding small molecule structure determination.

The original co-crystal 1 (CC1) and the preliminary porous lattice ($CC1^{+10bp}$) were both in the space group *C*121 (*i.e., I*121). Preliminary attempts to crystallize $CC1^{+10bp}$ were performed in solution conditions similar to the original CC1 (300-500 mM $MgCl_2$, 25-35% PEG 400 and 100 mM tris HCl pH 8.0).[73] Despite efforts to coax the crystals to adopt a more porous habit through DNA and protein engineering (Section 3.3.3), we consistently grew a different crystal habit, best described with the *I*222 space group. In the *I*222 crystal habit, we achieved an interpenetrating lattice in which a second copy of the target lattice was found within the body of the crystal, rotated by 180 degrees (Figure 3.1). In other words, when we expanded the DNA struts which hold together the CC1 lattice in two dimensions we obtained two interpenetrating lattices rather than a single porous lattice. Interpenetrating crystal growth is an interesting phenomenon which has been repeatedly observed in the context of Metal Organic Frameworks

(MOFs), particularly when researchers attempt to grow isoreticular MOFs with larger pores.[80-82] Despite being a less porous system, the interpenetrating CC1 shows remarkable modularity and utility for hosting DNA-binding molecules, as shown.

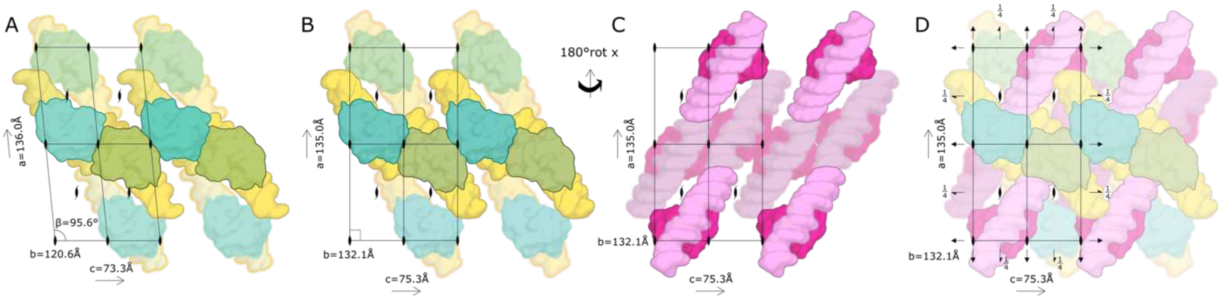


Figure 3.1. The target porous lattice (A), shown here in the *I*121 setting, was remarkably similar to (B) one half of the contents of the *I*222 experimental structure. (C) The other half of the *I*222 experimental structure is obtained *via* a 180° rotation about the *x*-axis (vertical in this diagram). (D) The full *I*222 structure requires close packing of DNA duplexes and involves additional 2-fold symmetry axes perpendicular to the *y*-axis.

3.3. Results and Discussion

In this chapter we show the expanded, interpenetrating lattice is designable, grows with remarkable consistency, and is useful to entrap guest molecules. We showed the expansion $CC1^{+10bp}$) is possible with at least 3 different DNA duplex expansion schemes: symmetric, asymmetric and scaffold-insert (Chapter 2 Figure 2.3). We obtained robust crystal growth, even when varying the sticky base overhang length from blunt ends to 4-nt. sticky base overhangs. In addition, crystals grew with variable termini: 5’ phosphate, 3’ phosphate, and lacking terminal phosphorylation. Further work is needed to quantify the effects of sticky base overhangs and terminal phosphorylation on the crystal growth rate.

Appendix II contains corresponding supplemental figures to Chapter 3.

3.3.1. The Interpenetrating Crystal Habit

The designed lattice and the experimentally obtained one were quite similar (Figure 3.1). The obtained *I*222 unit cell (75.26Å, 132.07Å, 134.99Å, 90.00°, 90.00°, 90.00° in the best resolution CC1^{+10bp} (PDB code: 7u6k)) differed from the *C*121 target model (161.43Å, 121.86Å, 73.89Å, 90.00°, 121.53°, 90.00°). To reveal the similarity (Figure 3.1), we select an alternate *I*121 setting of the design model (135.99Å, 120.57Å, 73.30Å, 90.00°, 95.63°, 90.00°) and rotate the axes for the obtained orthorhombic *I*222 crystal (so that the unit cell vectors are 134.99Å, 132.07Å, 75.26Å).

The obtained protein structure closely matches the structure in the parent structure (PDB code 7rva) and thus also the protein in the designed expanded model (Figure S3.1). Specifically, when superimposing the 210 alpha carbons found in both models, the $RMSD_{C\alpha}$ = 0.49Å. Given this alignment, the positions of the two most significant protein symmetry neighbors (both of which are rotated 180° about the *y*-axis) closely match the design model. The neighbor with the larger deviation ($RMSD_{C\alpha}$ = 1.42Å) is adjoined *via* a protein-protein interface formed between the C-terminal beta-sheets. The neighbor with the smaller deviation ($RMSD_{C\alpha}$ = 1.23Å) is connected *via* a protein-protein interface that surrounds a 2-fold symmetry axis that passes between Gly91, Pro113, and Ile116 and the symmetry copies thereof. The slightly more precise conservation of position for this neighbor may be due to the larger interface (1465 Å^2 SASA buried) compared to the C-terminal beta sheet interface (775 Å^2), *per* calculations with MSMS.[83] Considering the positioning of both protein neighbors, the small deviations are consistent with minor changes in the size of the shortest unit cell vector

(73.3 Å versus 75.3 Å), the unit cell vector that is aligned with the direction of the protein stack in the crystal.

The other two critical neighbors are connected *via* pure translational displacement and DNA:DNA stacking. If $\vec{a}$, $\vec{b}$, and $\vec{c}$ represent the unit cell edge vectors (not just the scalar edge lengths), the DNA:DNA stacking neighbors are found at $\pm\frac{\vec{a},}{2}, \pm\frac{\vec{b}}{2}, \pm\frac{\vec{c}}{2}$. To quantify the displacement of the DNA in one of these symmetry neighbors we compute $\mathrm{RMSD}_{C1'}$ = 8.60Å, comparing the 62 C1' atoms in the obtained crystal to the same atoms in the design model. We attribute the larger displacement of the neighboring DNA, compared to neighbor proteins, to the lever arm of the expanded DNA.

Whereas the RMSD numbers above refer to entire protein or DNA domains it is also interesting to quantify the deviation between the predicted crystal and obtained crystal at the key interfaces. For example, if we select 14 residues at the C-terminal beta sheet protein-protein interface, 7 residues from each symmetry partner (I180, S182, S225, I227, K229, V239, S241), we can superpose all 28 alpha and beta carbons with $\mathrm{RMSD}_{C\alpha\beta}$ = 0.75Å. Essentially, as shown in Figure S3.2, the details of the obtained interface strongly match the parent structure (PDB code 7rva) and the expanded design model built from it. We can similarly select 26 residues at the other protein-protein interface, 13 residues from each symmetry partner (A90, G91, E93, E110, S111, F112, P113, I116, K117, P118, N132, P133, Y134). We can superpose all 50 alpha and beta carbons with $\mathrm{RMSD}_{C\alpha\beta}$ = 0.56Å. In keeping with the results above, this second protein-protein interface more precisely recapitulates the interface in the parent structure (Figure S3.2).

3.3.2. Interpenetrating Co-crystal 1 Modularity

We demonstrated that assembly of the interpenetrating lattice was modular by growing crystals using the three different DNA expansion strategies mentioned in Chapter 2 Figure 2.4. In summary, the three schemes for 10 additional base pairs in each DNA strut were: a **symmetric expansion** (+5bp each side of original 21mer); a **asymmetric expansion** (+10 bp on one side of the scaffold protein binding site); and a 2-part **scaffold-insert expansion** (two separate DNA blocks, one scaffold duplex (21-mer) and one insert duplex (10-mer). All three strategies led to the growth of crystals with the same columnar habit, the same *I*222 space group, and evident interpenetrating lattice (Figure 3.2).

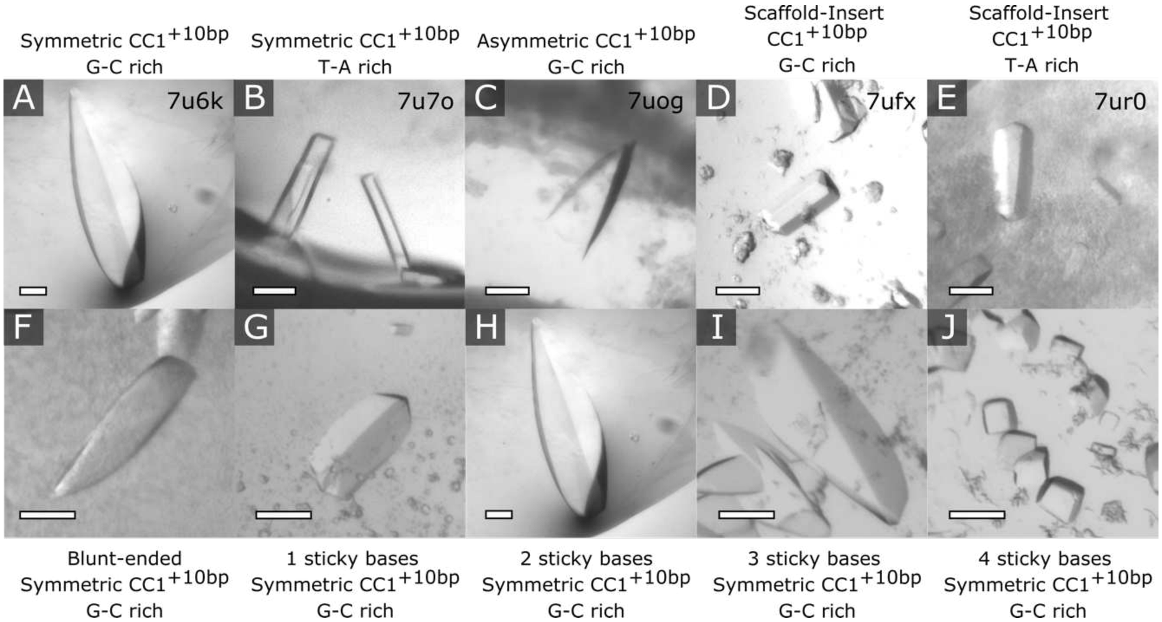


Figure 3.2. Interpenetrating CC1^{+10bp} crystal growth. The crystals are from the crystal class co-crystal 1 (CC1) of RepE54 Transcription Factor. The cognate DNA sequence for each asymmetric unit is found below the image in blue. The symmetric expansion crystals were grown with (A) G-C rich addition or (B) T-A rich addition. The asymmetric expansion crystal was grown with (C) G-C rich addition. The scaffold-insert expansion crystals were grown with (D) G-C rich insert strand or (E) T-A rich insert strand. The symmetric crystals (F-J) grew with varied sticky base overhang lengths: blunt end, 1-nt, 2-nt, 3-nt, and 4-nt. Matching crystal growth conditions and XRD resolutions are found in Table S3.1. PDB codes for A-E are in the upper right corner. Scale bars are 100 μm.

To demonstrate reliable assembly regardless of the inserted DNA sequence, we grew crystals using both the symmetric and scaffold-insert strategies with two dramatically different inserted sequences: a G-C rich sequence (5'-GACGGCCCGG) and a T-A rich sequence (5'-GACGGTAATT) (Table S3.2). Crystals consistently grew into the interpenetrating lattice. There were slight differences in favorable crystal growth conditions and crystal habit (Figure 3.2, Table S3.1), though we cannot definitively attribute these to the sequence difference (as opposed to other variables such as oligo purity and concentration). X-ray diffraction revealed that the nanostructure was essentially identical for both sequences (both in space group *I*222; unit cell dimensions Symmetric G-C rich: 75.261 132.066 134.986 90 90 90; unit cell dimensions Symmetric T-A rich: 72.141 129.124 131.29 90 90 90).

To understand how interpenetrating CC1^{+10bp} crystals were affected by changes at the DNA-DNA junction, we grew crystals with varied sticky base overhang lengths and terminal phosphorylation status. Using the symmetric expansion strategy and G-C rich sequence, we grew a variety of crystals featuring blunt ends as well as crystals with 1, 2, 3, or 4 sticky base overhangs encoding the desired assembly (Figure 3.2). While crystals with differing sticky overhang length did grow with differences in size and optimum growth conditions, we nonetheless observed crystal growth for all such systems within one family of growth conditions (300-500 mM $MgCl_2$, 25-35% PEG 400, and 100 mM tris HCl pH 8.0). We have similarly verified crystal growth (with blunt ends or 1, 2, 3 nt-overhangs) when using DNA oligos that were modified to include either 3' or 5' phosphorylation. Crystal growth was reliable despite these changes. The nucleation and growth phenomenon that underlie crystal formation are very sensitive to small differences in the initial conditions. For this reason, it is challenging to unambiguously attribute changes in crystal growth to one variable (*e.g.,* sticky overhangs or phosphorylation). Our results to date, consistent with the work of others,[84] do suggest that crystal growth might be tuned by changing the DNA-DNA junctions, but future work will be needed to carefully quantify such effects using identical growth conditions.

To further probe the modularity of the isoreticular co-crystal assembly strategy we also combined the RepE54 scaffold protein building block with DNA blocks intended to expand the parent crystal lattice by 2 DNA turns (20 bp) *per* scaffold protein binding site (Figure S3.3). Of the strategies described above, we tested the symmetric and scaffold-insert DNA expansion strategy for these large expansion attempts. Notably, if these building blocks were to assemble into the target lattice (Figure S3.3) the solvent channels would be predicted to accommodate the transport of spherical guests with diameter of 8.3 nm (*per* MAP_CHANNELS[31]). The predicted solvent content would be 87%. Given the statistical tendency for crystals with high solvent content to diffract poorly,[85] we prioritized our efforts with the 10-bp expanded crystals. Despite conducting a limited number of crystal growth trials for this system, we observed crystal growth for the CC1^{+20bp} expansion case for both symmetric and scaffold-insert expansion schemes (Figure S3.3). Initial crystals were small (10-20 µm in length), which lowered our expectations to observe diffraction (Figure S3.3C,D). However, recent crystallization optimization of CC1^{+20bp} yielded crystals large enough for data collection (Figure S3.3E) and follow-up XRD studies are underway.

3.3.3. Designs towards the Target Porous Lattice

Before crystallizing the intended porous lattice with optimized crystallization conditions, we attempted a myriad of engineering tactics to coax the interpenetrating lattice to adopt the intended porous lattice. For the interpenetrating lattice to form, DNA blocks must pack together closely (Figure 3.3). In the most closely packed region, the phosphate backbone of one DNA block is docked into the major and minor grooves of a symmetry neighbor (Figure 3.3A). While the parent crystal structure does feature one resolved magnesium that is coordinated by Asp-81 and Glu-77 from RepE54 (Figure 3.3B), our new lattice appears to be stabilized by the presence of four additional interfacial magnesium atoms (Figure 3.3C). These magnesium atoms are most tightly coordinated by the phosphate backbones of neighboring strands. The following are attempts to disturb each of these interactions to achieve the porous lattice. Ultimately, the porous lattice was obtained as described in Section 2.3.3.

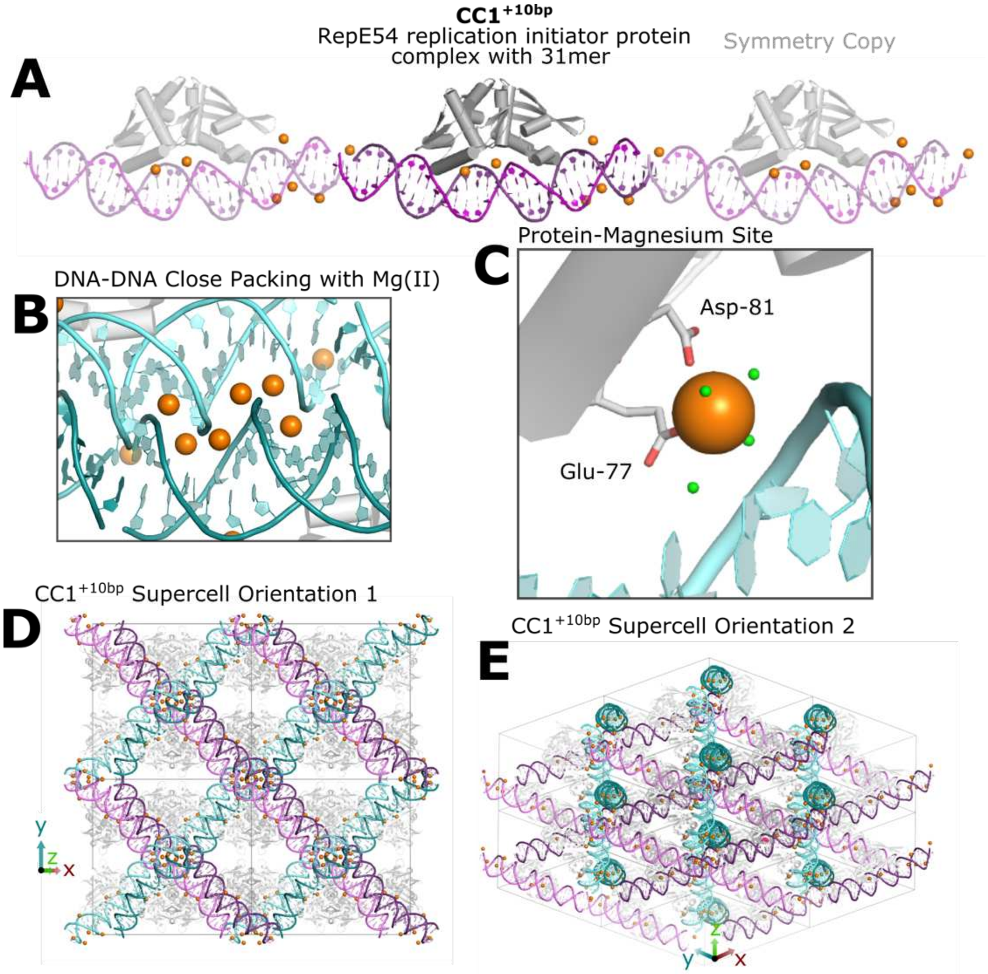
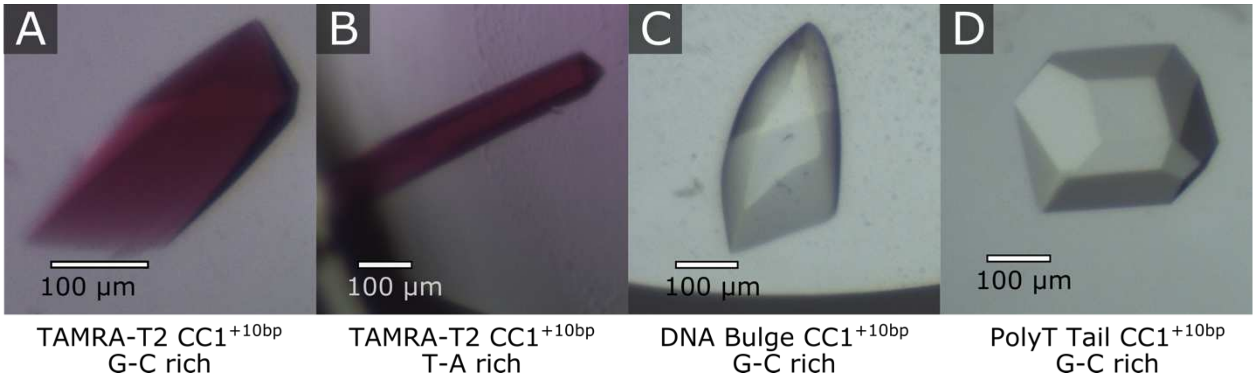


Figure 3.3. The interpenetrating CC1^{+10bp} features RepE54 Protein complexed to a 31mer DNA. (A) The DNA stacks end to end in the crystal and (B) helix-groove packing with ordered magnesium (orange spheres). The protein-Mg(II) site (C) found in the original CC1 structure is conserved in every interpenetrating crystal. The DNA pillar orientations (D,E) in the interpenetrating crystal are highlighted in four different colors.

We hypothesized that the high magnesium concentration of the crystallization condition (~400 mM $MgCl_2$ in mother liquor) was necessary to stabilize growth of the interpenetrating lattice. To test this hypothesis, we sought crystallization conditions with lower magnesium. To date, we have found one such low-magnesium crystallization condition: 160 mM $MgCl_2$, 22% PEG 400 and 80 mM tris HCl pH 8.0 (Figure S3.4). The LowMG lattice grows in a different crystal habit and preliminary X-ray diffraction analysis showed the crystal was growing in a different space group (*C*222). Further analysis of the LowMG crystal may give ideas for targeted designs of the lattice.

Following the same hypothesis that magnesium was inducing the interpenetrating form, we attempted crystal growth with calcium rather than magnesium, with hopes the larger cation may prevent tight DNA interactions. The crystals grew in the same interpenetrating lattice (*I*222) with very little differences in the DNA structure (PDB entry 8d8m, Figure S3.5). One difference was the divalent cation binding sites. There were two resolved calcium atoms near the protein-DNA interface, as in the Mg(II) structure. However, there were no cations present at the DNA-DNA stacking interface, as we saw in the high Mg(II) growth solution (Figure 3.3C). There is the possibility that the removed divalent cation sites may have been due to the resolution of the structure (3.1Å).

In another approach to block the double helix interaction, we added larger features in the DNA design to block the unintended DNA-DNA interactions. We designed three different DNA features to block the close DNA-DNA interactions by steric hinderance. DNA duplexes for the designs are found in Table S3.3. First, we added a TAMRA-labelled thymine strategically positioned such that upon self-assembly, the bulky TAMRA group may hinder the DNA stacks (Figure 3.4A,B). Instead, the crystals grew in *I*222 lattice, with the TAMRA unresolved but the DNA had distinct electron density where the TAMRA was conjugated to the thymine (more details in Section 3.3.5). Next, we made a DNA bulge with an additional adenine at the DNA stacks (Figure 3.4C). Preliminary X-ray diffraction analysis was inconclusive as the crystal diffraction was ~5.34Å and we were unable to define the space group or packing arrangement. Finally, we added a penta-thymine tail (Figure 3.4D) and the crystals grew in a different habit. Upon preliminary XRD analysis, the crystal was best explained with the *I*222 space group but the molecular replacement with the CC1^{+10bp} (PDB code 7u6k) and successive refinement were not conclusive for the crystal packing arrangement. Further analysis is necessary for the Bulge and PolyT Tail crystals to determine their crystal structures.



TAMRA-T2 CC1^{+10bp} G-C rich

TAMRA-T2 CC1^{+10bp} T-A rich

DNA Bulge CC1^{+10bp} G-C rich

PolyT Tail CC1^{+10bp} G-C rich

Figure 3.4. Crystal growth of DNA designs. Crystal growth conditions are found in Table S3.3.

3.3.4. Guest Protein Entrapment

To demonstrate the potential utility of the designed expanded crystals as molecular scaffolds, we embedded within the DNA insert a binding site (5’-TAATTA) for the engrailed homeodomain protein (EnH)[86] (Figure 3.5 and Figure S3.6). Modeling suggests that EnH binding is incompatible with the presence of a symmetry neighbor from the interpenetrating lattice (Figure S3.7). Therefore, we hypothesized that EnH binding in the crystallization experiment might compete with the interpenetrating lattice. To test this hypothesis, we first fused EnH at the genetic level to an enhanced green fluorescent protein (eGFP) (Figure S3.6 and Protocol S3.1). In principle, EnH-eGFP fluorescence might allow visualization and quantification of guest capture within the co-crystals *via* fluorescence or confocal microscopy.

Our first EnH installation experiment used the symmetric expansion strategy. In this case the EnH binding sequence was divided at the flanking DNA end region (Figure 3.5). Our hypothesis was that EnH-eGFP would bind during crystal growth as the DNA-DNA sticky base overhangs hybridize and thus reconstitute the full EnH binding site. We further hypothesized that EnH-eGFP binding would stabilize labile DNA-DNA junctions during crystal growth. We co-crystallized scaffold symmetric expanded CC1^{+10bp} with a moderate stoichiometric excess of EnH-eGFP (1:1.2:1.2 RepE54 scaffold protein, DNA, and guest EnH-eGFP). Consistent with the presence of EnH-eGFP in solution, the initial crystallization solution appeared green by eye. After crystal growth, the solution was transparent, but the crystal had not become bright green by eye. However, when illuminated with a blue laser (420 nm) using a confocal microscope, the crystals grown in the presence of EnH-eGFP were bright green (Figure 3.5). A control crystal, grown with the same scaffold protein and scaffold DNA in the absence of EnH-eGFP was not fluorescent (though the adjacent EnH-eGFP doped crystal illuminated part of the control crystal, and reflections occur at crystal facets due to refractive index changes).

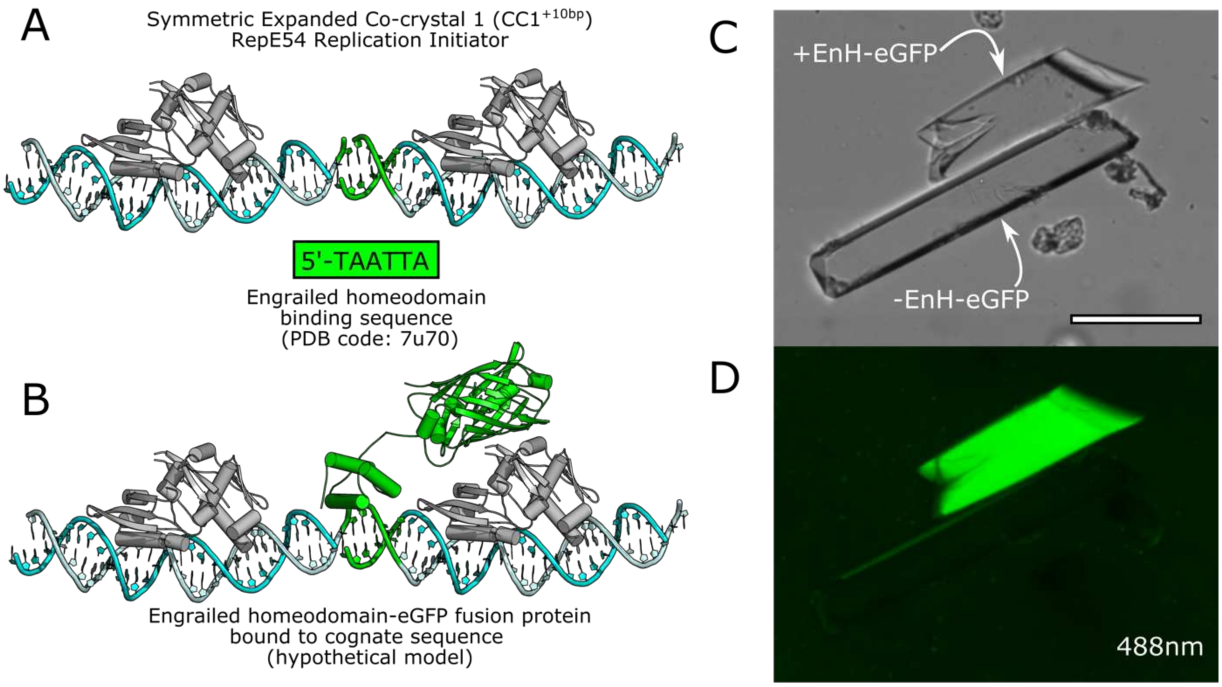


Figure 3.5. (A) The symmetric expanded CC1^{+10bp} was designed with the Engrailed homeodomain (EnH) binding site. (B) After co-crystallization with EnH-eGFP fusion protein, (C-D) the crystal fluoresced when exposed to 488 nm with confocal microscopy, indicating guest EnH-eGFP entrapment. The control crystal, grown without EnH-eGFP, did not fluoresce. Scale bar is 100 µm.

The presence of a full interpenetrating lattice would be expected to block the EnH binding site (and would leave little room for the full EnH-eGFP fusion protein) (Figure S3.7). Therefore, we were unsurprised to find that we could not observe bound EnH in crystal structures where EnH-eGFP was present during crystallization. Instead, we were surprised to find, using confocal microscopy, that EnH-eGFP was nonetheless uniformly incorporated within the co-crystals, and that incorporation required the cognate EnH DNA binding site. Control crystals lacking the EnH DNA binding site had much lower evident doping of the EnH-eGFP (Figure S3.8). Significant doping of the EnH-eGFP might correspond to isolated defects in an otherwise ideal interpenetrating double lattice. Alternately, the crystal may have larger domains that are composed of only one of the two interpenetrating lattices.

3.3.5. Guest DNA Co-crystallization

Using ICC modularity to generate scaffold crystals with specific protein binding sites is only one application. Capturing functional cargo nucleic acids represents an intriguing alternative. For example, guest DNA might encode information, or a guest RNA might have a functional therapeutic role. In both cases, the crystal might serve as a tough protective matrix. The scaffold-insert strategy described above provides a starting point if the insert is viewed as a guest molecule. To track incorporation of guest DNA we grew crystals with a stoichiometric mixture of scaffold DNA (21-mer), scaffold protein, and insert DNA (10-mer). One strand of the insert duplex was modified with the red fluorescent dye TAMRA (conjugated to the C7 atom in thymine 2) (Figure 3.6). The insert DNA was varied to be G-C rich or T-A rich, both with a strand containing a TAMRA-T2. When control crystals were set up without the insert strand, crystals did not grow (Figure S3.9). We therefore expect that the quantity of guest DNA incorporated within the crystal is stoichiometrically matched to the quantity of scaffold DNA and scaffold protein.

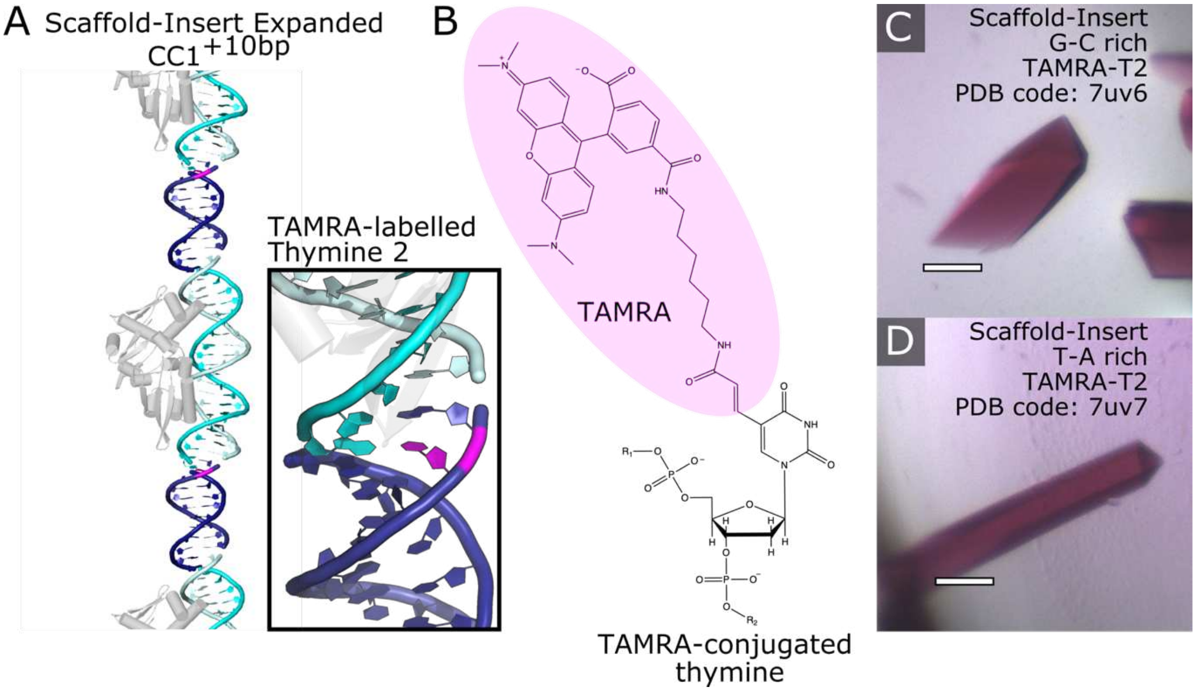


Figure 3.6. (A) The scaffold-insert expanded complex showing the scaffold DNA in cyan and the insert strand in dark blue. The TAMRA-labelled thymine 2 (magenta) is conjugated on the C7 atom (B). Co-crystals grew with the TAMRA-T2 in the interpenetrating, *I*222 space group. Crystals consistently grew for both a G-C rich insert sequence (C) and a T-A rich insert sequence (D). Substituents R_1 and R_2 are the 5' and 3' neighboring nucleobases, respectively.

Post-crystallization, the guest-loaded crystals were magenta by eye (Figure 3.6 C-D). Given the crystal structures, we assume that the crystal contains an equal stoichiometric number of scaffold DNA blocks, scaffold proteins, and insert DNA blocks. The X-ray structures matched the intended expanded (and interpenetrating) lattice. The thymine containing the TAMRA-T2 had a visible linkage to the TAMRA in the electron density (Figure 3.6) in both crystals, G-C rich and T-A rich inserts. The lack of electron density for the TAMRA ring system suggested that the fluorescent dye was not rigid in the crystal.

3.3.6. Guest Small Molecule Co-crystallization

Yet another use of a scaffold crystal is entrapment of a specific DNA-binding small molecule. Netropsin, a robustly studied minor groove DNA-binding drug, has been co-crystallized with DNA previously and shown to prefer T-A rich binding sequences, especially 5'-AATT.[87] Here, we grew CC1^{+10bp} co-crystals in the presence of a slight stoichiometric excess of netropsin (1:1.2:1.2 Protein to DNA to Netropsin). As with the previous guest protein and DNA, the presence of a DNA-binding guest did not prevent assembly of the interpenetrating lattice. Post co-crystallization, the netropsin was fully visible in the X-ray diffraction pattern (Figure 3.7). As expected from previous research on netropsin DNA binding specificity, the small molecule bound in the minor groove at the 5'-AATT sequence which was binding pinned between the RepE54 protein and the recognition sequence thereof (Figure 3.7A). We did not find any direct protein to netropsin interaction. Each netropsin amide was hydrogen bonded to DNA (Figure 3.8) resembling the quintessential Class I netropsin-DNA binding which includes bifurcated hydrogen bonds between the surrounding A-T bases.[87] The ends of netropsin were ordered, with the amidinium atom hydrogen bonded to complimentary A and T and the guanidinium end hydrogen bonded to neighboring adenines (Figure 3.8B). Future co-crystals that are sequence specific for netropsin at the DNA-DNA junction may induce small molecule binding at the helix-groove packing region of the interpenetrating crystal and produce a porous lattice instead.

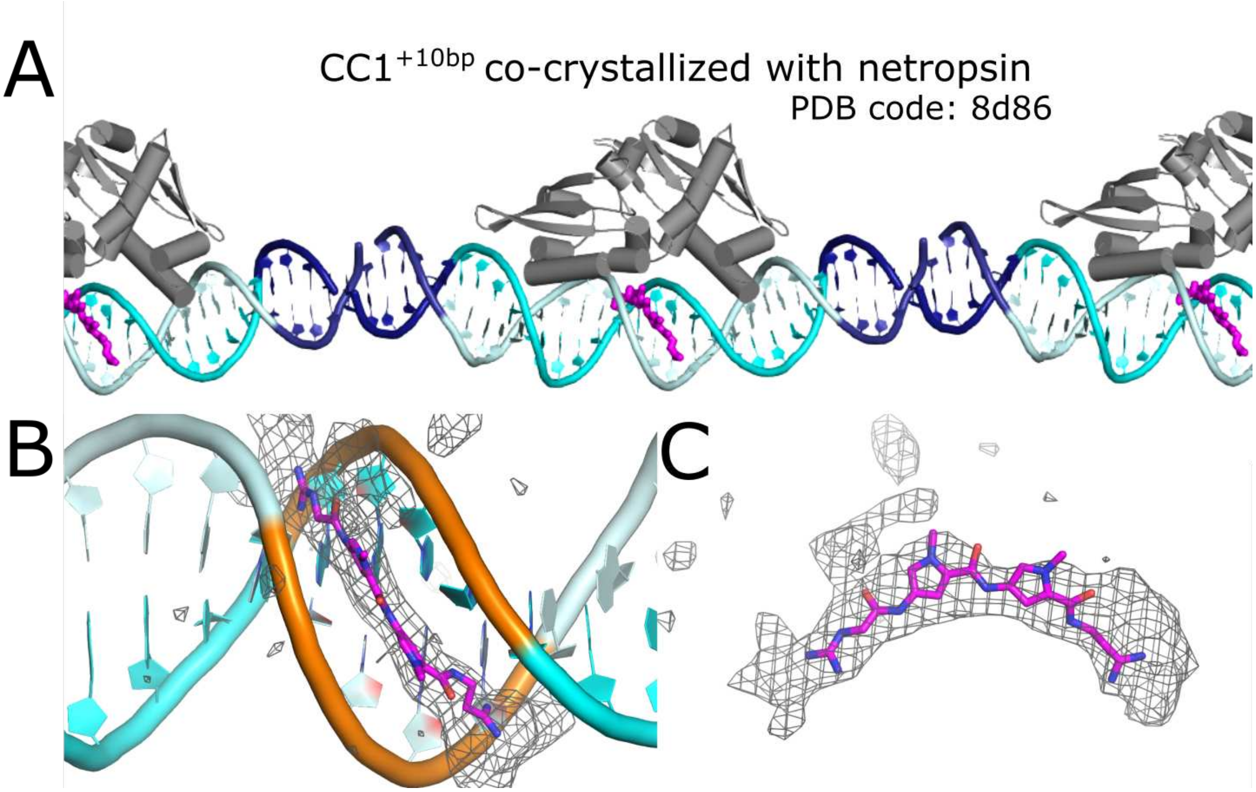


Figure 3.7. X-ray diffraction highlight of netropsin bound to CC1^{+10bp}. (A) The electron density indicated netropsin, shown in magenta, bound at the binding site 5’-AATT. The co-crystal resolution was 3.12Å. (B) A Polder omit map in grey mesh shows the electron density around netropsin in the binding pocket. The DNA backbone of the binding site is highlighted with orange. (C) The Polder map of netropsin (contour level 3 rmsd).

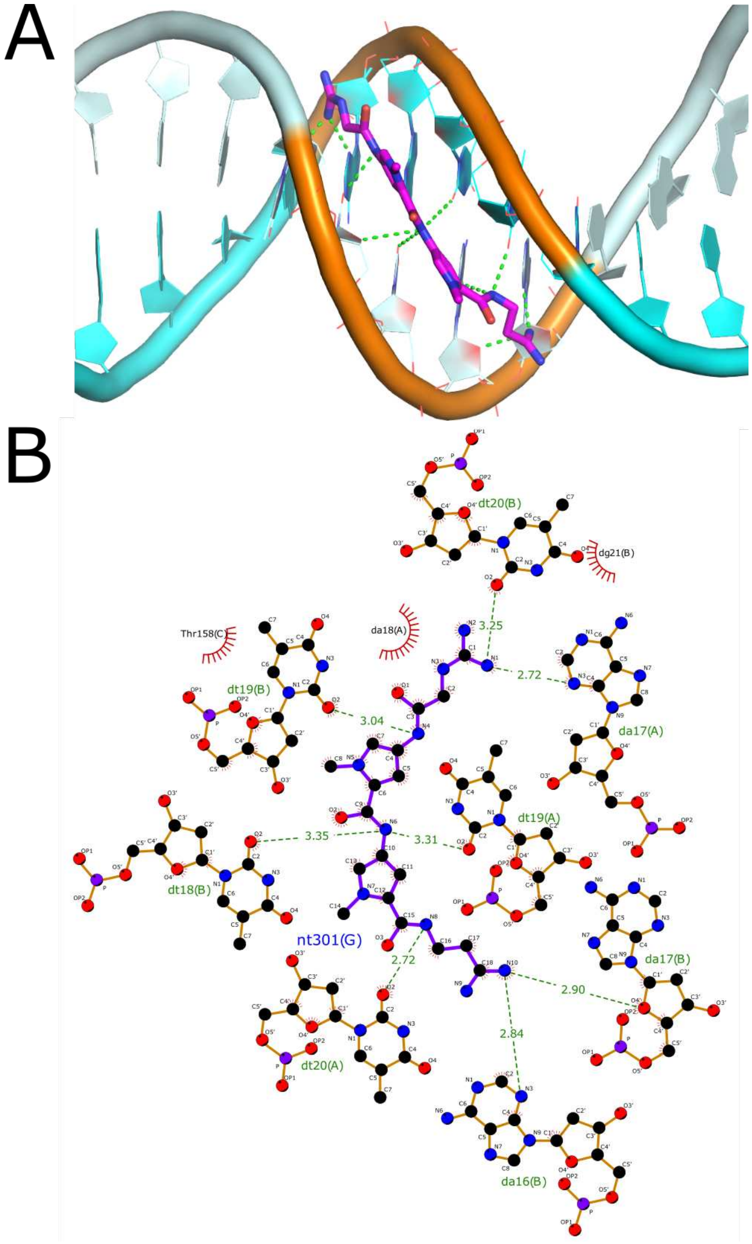


Figure 3.8. Hydrogen bonding network of netropsin bound to CC1^{+10bp}. (A) The PyMOL view of hydrogen bonds corresponds to distances and bases in the hydrogen bonding diagram. (B) Schematic of hydrogen bonding interactions between netropsin and the bases and phosphodeoxyribose backbone of the cognate DNA. The hydrogen bond cutoff distance is 3.5Å.

3.4. Conclusions

The designed host co-crystals we present here are modular and consistent scaffolds for entrapping DNA-binding molecules. In the interpenetrating crystal habit, the DNA designs repeatably obtained the same crystal symmetry (*I*222) despite changes in expansion scheme, sequence, sticky base overhang lengths, and terminal phosphorylation. Within the expanded, interpenetrating co-crystals, we installed three diverse guest molecules using the DNA-binding event: a protein, a discrete DNA insert, and a small molecule. Therefore, we conclude the crystal structure is robust and may be tunable for DNA-binding guest of interest.

The limiting feature of the interpenetrating habit co-crystals is the porosity. Ideally, one would host guest proteins *post*-crystallization in a co-crystal with large solvent pores. The current interpenetrating form allowed for entrapment of guests during crystallization. We hypothesize the interpenetrating crystal form is preferred due to the high magnesium in the growth solution and therefore, the neutral DNA strands can pack tightly. Optimization of the growth solutions may produce a porous lattice and isoreticular expanded crystals for guest installation *via* DNA-binding. Alternatively, co-crystallization with a fully intact guest binding site and in excess guest protein may force the crystal to adopt a different habit.

With few exceptions, biomolecular crystals adopt adventitious lattices. Here we show that isoreticular expansion, a concept adopted from metal organic framework engineering, can be applied to re-engineer protein-DNA co-crystals for the modular insertion of designed DNA blocks. The resulting scaffolds may have favorable biomaterial properties. For example, subjecting such an interpenetrating network to chemical ligation may result in remarkably robust biomaterial. Hosting guest molecules in these crystals is the first step in using these crystals as molecular delivery devices.

3.5. Materials and Methods

3.5.1. Protein Cloning, Expression, and Purification

The cloning and expression of the scaffold protein, RepE54 Replication Initiator, was first described by Komori *et al.*[73] The CC1 RepE54 was produced at the Histone Source at Colorado State University exactly as described in Section 4.5.1.

The CC1 protein variant RepE54(L53G,Q54G,E55G) was cloned into the PetDuet plasmid (Novagen) using Gibson cloning (Hi-Fi assembly New England Biolabs) (Protocol S3.1). The protein was expressed with a T7 promoter in *E. coli* BL21(DE3) cells. Upon addition of 0.5 mM IPTG, the cells were outgrown at 25 °C for 20 h. The cell pellets were sonicated in lysis buffer (1X PBS, 300mM NaCl, 25mM imidazole, pH 7.4) and applied to HisTrap (HisPur™ Ni-NTA resin) equilibrated with HisTrap buffer (1X PBS, 300mM NaCl, 25mM imidazole, pH 7.4). The protein was eluted with 100 mM imidazole in HisTrap buffer. Following HisTrap the samples were purified at the Histone Source with HiLoad Superdex 200 PG column (Cytiva). Eluants were stored at 15 mg/mL in CC1 storage buffer and flash frozen for storage at −80 °C.

The guest protein, Engrailed Homeodomain-eGFP Fusion (EnH-eGFP), was cloned into the PetDuet plasmid using Gibson cloning (Protocol S3.1). EnH-eGFP protein was expressed with a T7 promoter in *E. coli* BL21(DE3) cells. Upon addition of 0.5 mM IPTG, the cells were outgrown at 25 °C for 20 h. The cell pellets were sonicated in lysis buffer ((100mM tris HCl, 200mM NaCl, 10% glycerol, 10mM imidazole, pH 8.0) and applied to HisTrap (HisPur™ Ni-NTA Resin) equilibrated with HisTrap buffer (100mM tris HCl, 200mM NaCl, 10% glycerol, 10mM imidazole, pH 8.0). The protein was eluted with 100 mM imidazole in HisTrap buffer. The EnH-eGFP protein was purified further with Nuvia™ cPrime™ Hydrophobic Cation Exchange Media, equilibrated with cation exchange buffer (100mM NaCl, 100mM tris HCl, 10% glycerol, pH 6.8), and eluted with 400 mM NaCl in cation exchange buffer. Following cation exchange, the samples were purified *via* HiLoad Superdex 200 PG column (Cytiva) at the Histone Source at Colorado State University. The fractions containing EnH-eGFP protein were pooled, concentrated using Amicon Ultra-15 10 kDa MWCO centrifugal filter unit (EMD Millipore), and dialyzed with EnH-eGFP storage buffer (200mM NaCl, 100mM tris HCl, 10% glycerol, pH 8.0). EnH-eGFP protein was collected, concentrated to 42 mg/mL, and stored at −80 °C after flash freezing with liquid nitrogen.

Protein purification steps were analyzed with SDS-PAGE (NuPAGE™ 4–12% Bis-Tris Gel) with MES SDS running buffer and stained with Imperial™ Protein stain. Bradford Assay using Coomassie Plus™ Protein Assay Reagent was used to determine final protein concentrations.

3.5.2. DNA Duplex Annealing

The DNA oligomer sequences used for co-crystallization are listed in Tables S3.2 and S3.3. Each oligomer contains the 19-bp iteron sequence for RepE54 Protein-DNA binding, but the flanking DNA sequences vary depending on the expansion scheme and sequence. Individual oligomers were synthesized and HPLC purified by Integrated DNA Technologies and annealed in the Snow Lab. The oligomers were resuspended in CC1 oligo buffer (50 mM tris HCl, 100 mM KCl pH 7.0) and combined in equal molar ratio

(1:1) with the complimentary strand. The strands were annealed by heating to 94 °C for 2 min and slowly cooling to room temperature over approximately 60 minutes. The final concentration of all CC1 duplexes was 4 mM.

3.5.3. Scaffold Protein-DNA Complex Co-crystallization

The scaffold protein and DNA (1:1.2) were incubated on ice 30 minutes prior to crystallization *via* sitting drop vapor diffusion. Crystallization conditions for both the interpenetrating lattice (CC1^{+10bp}) and for the 20 bp expansion (CC1^{+20bp}) were 300-500 mM $MgCl_2$, 20-35% PEG 400 and 80-100 mM tris HCl pH 8.0. A detailed list of individual crystals, PDB codes and corresponding growth conditions are in Table S3.1. Crystals grew to a size of 50 - 300 μm^3 in a range of 24 hours to 30 days. Crystal pictures were taken with a Moticam 3.0 MP camera attached to a Motic SMZ-168 stereozoom microscope.

3.5.4. Co-crystallization with Guest Protein, DNA and Small Molecule

Crystallization conditions for all guest entrapped co-crystals are in Table S3.1. Scaffold-guest co-crystallization were grown via sitting-drop vapor diffusion. The guest fusion protein (EnH-eGFP) was incubated with scaffold protein and symmetrical expanded DNA (1:1.2:1.2 scaffold protein to DNA to EnH-eGFP) on ice 30 minutes prior to crystallization set-up. The fluorescently labelled DNA guest was co-crystallized in CC1^{+10bp} using the scaffold-insert scheme. The insert duplex was a 10-mer with 2 sticky base overhangs and a G-C rich sequence. The complimentary strand contained the thymine-2 with a TAMRA. The oligomers making up the insert was synthesized and HPLC purified by Integrated DNA Technologies and the insert duplex was annealed in the Snow Lab. Co-crystallization with the excess of each of the scaffold and insert duplexes (1:1.2:1.2 scaffold protein to scaffold DNA to insert DNA). Netropsin dihydrochloride (VWR) was resuspended in molecular biology grade water and added to the G-C rich symmetric CC1^{+10bp} complex prior to crystallization (1:1.2:1.2 scaffold protein to DNA to netropsin).

3.5.5. Confocal Microscopy

Fluorescence microscopy was performed on a Nikon Eclipse Ti spinning-disk confocal microscope with an AndoriXon Ultra 897U EMCCD camera. EnH-eGFP protein entrapment was analyzed with under visible light (DIC-N1) and laser exposure (488 nm excitation).

3.5.6. X-ray Diffraction Data Collection and Refinement

Crystals were flash frozen in liquid nitrogen using the growth conditions as the cryo-protectant. Single-crystal X-ray diffraction data was collected at the Beamline 4.2.2 using a CMOS detector from 0 to 180 degrees with an omega delta of 0.2° and an exposure time of 0.5 seconds. Data was processed with XDS.[77] The best resolution expanded structure (CC1^{+10bp} Symmetric G-C rich; PDB code: 7u6k; 2.38Å) was solved by molecular replacement with the updated original RepE54 structure (PDB code: 7rva; 1.89Å) and refined (PHENIX[78] and COOT[79]). The later expanded structures were solved with molecular replacement using the best resolution expanded structure and the same Rfree flags were applied. A flow chart of X-ray diffraction refinement for all expanded structures is given in Figure S3.10. Tables S3.4-S3.9 list X-ray diffraction and refinement statistics for all structures.

3.5.7. Guest small molecule refinement

X-ray diffraction data was obtained from the co-crystal grown with netropsin. Molecular replacement was performed with the starting model as the best resolution expanded structure (CC1^{+10bp} Symmetric G-C rich; PDB code: 7u6k; 2.38Å). We first refined the structure without the netropsin ligand and produced a discovery map. To add netropsin to the structure, we used the previous structure (PDB code: 1ztt) with the same binding site (5'-AATT) and aligned the DNA binding sequences in PyMOL.[88] The resulting orientation of netropsin was added to the discovery map structure and this CC1^{+10bp}-netropsin model was used for further refinement in PHENIX (and unselected the automatic setting "link ligand to protein"). We produced polder maps from the structure with the ligand. The final structure was analyzed for hydrogen bonds in PyMOL and LigPlot^{+} v.2.2,[89] a ligand-protein interaction diagram package.

3.6. Funding

This material is based upon work supported by the National Science Foundation under Grant No. NSF DMR 2003748 and NSF DMR 1506219. The team also gratefully acknowledges support for undergraduate researchers from the Nelson Family Faculty Excellence Award.

CHAPTER 4. STABILIZING DNA–PROTEIN CO-CRYSTALS *VIA IN CRYSTALLO* CHEMICAL LIGATION OF THE DNA[4]

4.1. Overview

Protein and DNA co-crystals are most commonly prepared to reveal structural and functional details of DNA-binding proteins when subjected to X-ray diffraction. However, biomolecular crystals are notoriously unstable in solution conditions other than their native growth solution. To achieve greater application utility beyond structural biology, biomolecular crystals should be made robust against harsh conditions. To overcome this challenge, we optimized chemical DNA ligation within a co-crystal. Co-crystals from two distinct DNA-binding proteins underwent DNA ligation with the carbodiimide crosslinking agent 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) under various optimization conditions: 5′ *vs.* 3′ terminal phosphate, EDC concentration, EDC incubation time, and repeated EDC dose. This crosslinking and DNA ligation route did not destroy crystal diffraction. In fact, the ligation of DNA across the DNA–DNA junctions was clearly revealed *via* X-ray diffraction structure determination. Furthermore, crystal macrostructure was fortified. Neither the loss of counterions in pure water, nor incubation in blood serum, nor incubation at low pH (2.0 or 4.5) led to apparent crystal

---

[4] The work in this chapter is published in *Crystals*. Conceptualization, A.R.W., S.D., A.V. and C.D.S.; data curation, A.R.W., S.D., A.V. and C.D.S.; formal analysis, A.R.W., S.D., A.V. and C.D.S.; funding acquisition, A.R.W. and C.D.S.; investigation, A.R.W., S.D., A.V. and C.D.S.; methodology, A.R.W., S.D., A.V. and C.D.S.; project administration, A.R.W. and C.D.S.; resources, C.D.S.; software, A.R.W. and C.D.S.; supervision, A.R.W. and C.D.S.; validation, A.R.W., S.D., A.V. and C.D.S.; visualization, A.R.W. and C.D.S.; writing—original draft, A.R.W. and C.D.S.; writing—review and editing, A.R.W., S.D., A.V. and C.D.S.
Ward, A.R.; Dmytriw, S.; Vajapayajula, A.; Snow, C.D. Stabilizing DNA–Protein Co-Crystals *via in crystallo* Chemical Ligation of the DNA. *Crystals* **2021**, *11*, 49.

degradation. These findings motivate the use of crosslinked biomolecular co-crystals for purposes beyond structural biology, including biomedical applications.

4.2. Introduction

Beyond serving as the fundamental components of life, proteins and DNA are also key building blocks for nanoscale self-assemblies. Biomolecular assemblies, ranging from 2D arrays to 3D crystals, are useful tools for structural biology, bio-catalysis, and biomedical applications.[4,56,90] Porous biomolecular crystals can even act as macromolecular scaffolds,[1] providing structural details to guest macromolecules.[8] However, downstream applications of interest, including X-ray diffraction, are hindered by crystal fragility and intolerance to solvent conditions other than the crystal growth solution. In this study, we establish a protocol for the chemical ligation of DNA inside of crystals and we demonstrate structural resilience of crosslinked co-crystals which may further their application utility.

DNA assembly stability is a limiting factor for DNA nanotechnology and DNA crystals. While coding DNA sticky base overhangs can drive self-assembly, the non-covalent DNA base stacking interactions and Watson–Crick hydrogen bonds that stabilize the junctions are only stable under specific conditions. For example, crystallization conditions for DNA crystals typically feature high concentrations of divalent cations such as Mg(II) to balance the negative phosphate backbone of DNA.[91] DNA–protein co-crystals may be similarly reliant on counterions, particularly if counterions stabilize the DNA–protein binding event.[53] Crystal forms that bring DNA building blocks into close proximity are very sensitive to the counterion environment, and often dissolve or convert into a disordered aggregate when placed in water. To maximize application versatility, DNA structures should ideally be robust to solution variations, not just ionic strength but also temperature and pH.[90] Introducing covalent bonds across DNA–DNA interfaces has the potential to dramatically improve crystal macro-structure stability and could also improve X-ray diffraction.

Bioconjugation, or crosslinking, is a well-established strategy to improve the structural integrity of protein and DNA crystals.[92] The protein–protein interfaces found within protein crystals tend to be rich in primary amines and carboxylic acids. If all neighboring building blocks can be covalently linked, the resulting covalent organic framework can be a robust material. In traditional protein X-ray crystallography, glutaraldehyde, a highly reactive crosslinker, can increase crystal stability in varying solution conditions, and can even improve diffraction resolution.[93-94] In our previous work on protein crystals, we have found that glyoxal offers an effective alternative to glutaraldehyde.[15,45] Chemical crosslinking and photo-crosslinking methods for DNA crystals are also established in the literature;[95-97] however, we wanted to focus on a protocol in which the crosslinking does not require a specific sequence of DNA and does not add atoms to the structure (a zero-length crosslink).

Arguably the most natural form of sequence-independent DNA crosslinking is *ligation*, where the nicks dividing stacked dsDNA blocks are removed to generate longer contiguous DNA strands. For example, Li *et al.* used T4 DNA Ligase to ligate the DNA junctions within highly porous DNA crystals.[98] This elegant approach is limited to crystals that have large enough solvent channels for enzyme ingress. Here, we sought to optimize a chemical ligation alternative to the use of ligase that would be applicable to crystals with both large and small pores.

Our chemical ligation chemistry relies on 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC), a water-soluble carbodiimide.[92] EDC is widely used, especially in protein conjugation, to crosslink primary amines to carboxylic acids. A less common chemistry for EDC is the activation of a terminal phosphate such that a suitably placed nucleophile can displace the leaving group.[99] When that nucleophile is the hydroxyl of a neighboring DNA strand, this chemistry results in a zero-length crosslink: a scar-less chemical ligation of DNA (Figure 4.1). EDC has been used to ligate dsDNA hairpins in solution,[99] to link the phosphate backbone of stacked DNA in liquid crystals[100] and to stabilize a 600 nucleotide DNA origami structure.[101] Our work represents the first ligation *via* EDC of co-crystals containing protein and DNA. We show that EDC crosslinking dramatically increases crystal stability at the macroscale and does not destroy the crystal nanostructure (*i.e.,* treated crystals are still suitable for study *via* X-ray diffraction).

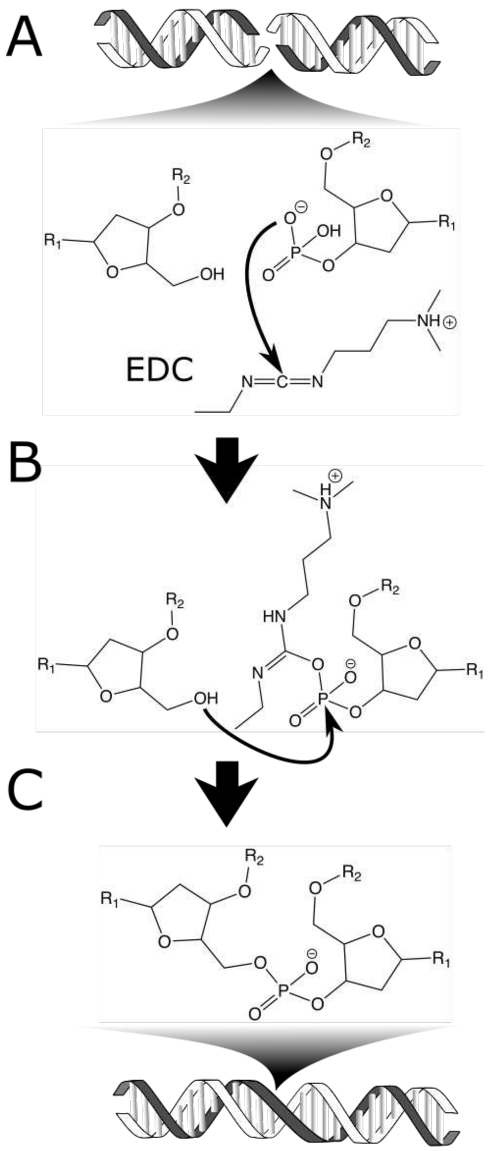


Figure 4.1. The mechanism of chemical DNA ligation with EDC. (**A**) A terminal 5′ hydroxyl and a terminal 3′ phosphate on neighboring DNA chains. The phosphate interacts with EDC to form an intermediate (**B**) and the hydroxyl displaces the reactive intermediate to form a zero-length crosslink (**C**) between the two DNA chains. R_1 is the nucleobase and R_2 is the phosphate backbone.

To demonstrate generality, we chemically ligate two different co-crystals of DNA-binding proteins containing stacked DNA–DNA interfaces (Figure 4.2). For convenience, we will refer to crystals of the RepE54 transcription factor bound to cognate 21-mer dsDNA as Co-Crystal One (CC1) (Figure 4.2A–D) and we will refer to crystals of the E2F8 transcription factor bound to cognate 15-mer dsDNA as Co-Crystal Two (CC2) (Figure 4.2E–G). The asymmetric unit for each co-crystal consists of a DNA-binding protein and short, cognate DNA duplex. Both co-crystals have existing models in the Protein Data Bank (PDB). CC1 is closely related to existing PDB entry 1rep, though the 1rep model corresponds to a crystal with differing DNA at the junction (Table S4.1). CC2 is identical to existing PDB entry 4yo2. The CC1 and CC2 crystals used in this study consist of dsDNAs that are either blunt-ended or carry terminal 5′ or 3′ phosphates (Figure 4.3). In each co-crystal system, the crosslinking variables tested were terminal 5′ *vs.* 3′ phosphates, crosslinking time, EDC concentration, and repeated EDC dose. After EDC crosslinking, co-crystals had dramatically increased structural integrity with respect to changes in the solution condition.

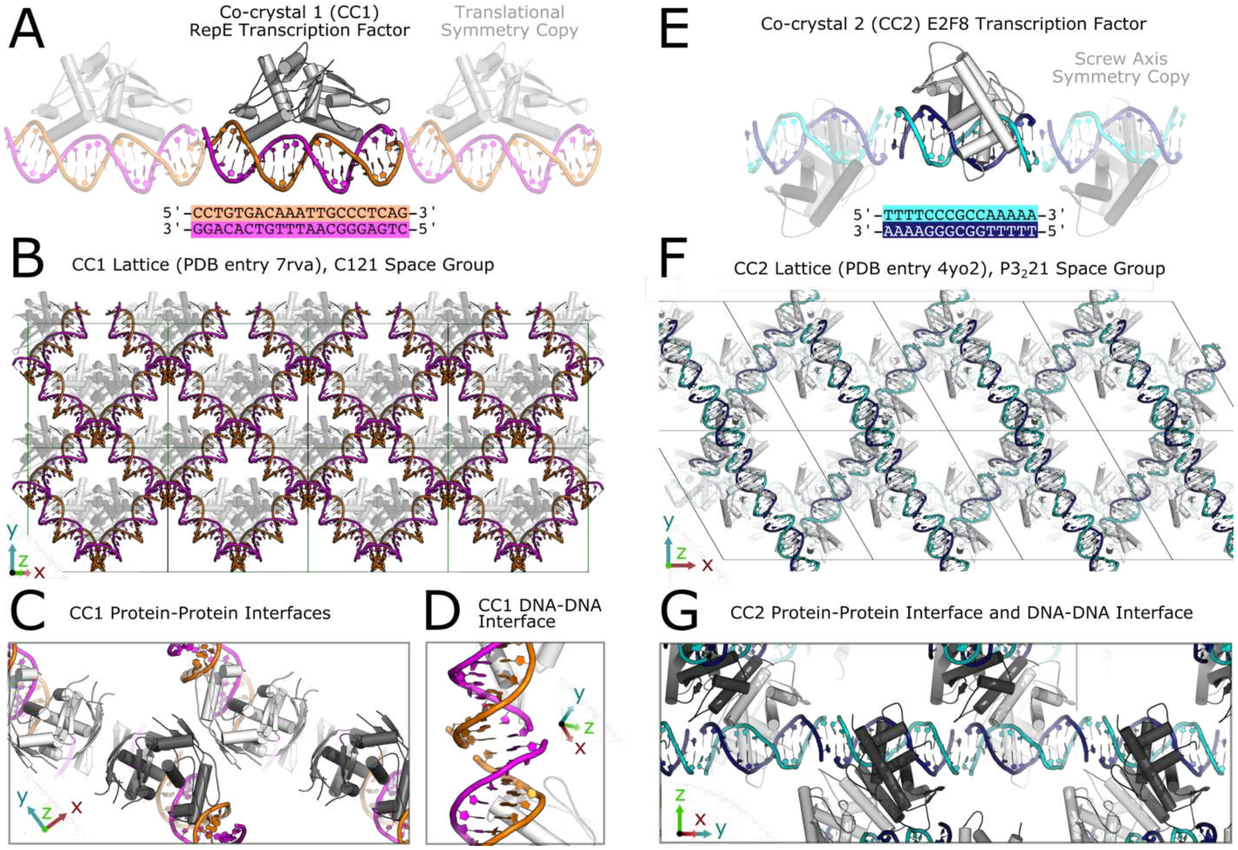


Figure 4.2. (**A**) The building block for co-crystal 1 (CC1) consists of the RepE54 transcription factor bound to 21-mer cognate DNA (represented here by PDB entry 1rep). **(B**) A collection of neighboring CC1 unit cells oriented to show the DNA stacks in 2 dimensions, with protein at 50% transparency. (**C**) The CC1 lattice has *C*121 symmetry, and all DNA–DNA junctions are symmetry equivalent to (**D**) the single DNA–DNA junction shown here. (**E**) The building block for co-crystal 2 (CC2) consists of the E2F8 transcription factor bound to 15-mer cognate DNA (represented here by PDB entry 4yo2). (**F**) A collection of neighboring CC2 unit cells oriented to show the DNA stacks in two dimensions, with protein at 50% transparency. (**G**) The CC2 lattice has *P*3$_2$21 symmetry, and all DNA–DNA junctions are symmetry equivalent to the single DNA–DNA junction shown here. Images were generated in PyMOL.

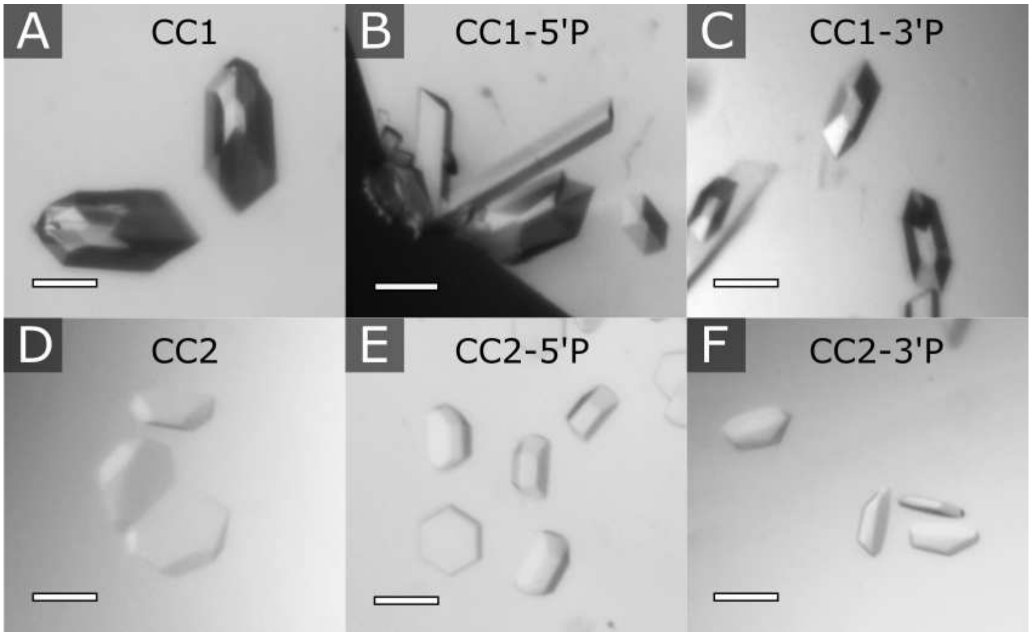


Figure 4.3. Examples of six co-crystal variants relative to a 100 micron scale bar. The CC1 crystals have *C*121 symmetry and tend to grow as monoclinic prisms: (**A**) CC1 without terminal phosphates, (**B**) CC1 with terminal 5′ phosphate, and (**C**) CC1 with terminal 3′ phosphate. In contrast, CC2 crystals have $P3_221$ symmetry and tend to grow as truncated hexagonal prisms: (**D**) CC2 without terminal phosphates, (**E**) CC2 with terminal 5′ phosphate, and (**F**) CC2 with terminal 3′ phosphate.

To show foundational feasibility for biomedical applications, we demonstrated that crosslinked co-crystals remain robust in aqueous environments, blood serum, and at pH values found in the stomach (pH 2.0) or lysosomes (pH 4.5). Therefore, the EDC crosslinking results provided here may justify further investigation of chemically ligated co-crystals or pure DNA crystals as biomaterials. For scaffold-assisted crystallography[56] it is also important to note that the crosslinked co-crystals still diffracted X-rays. The crosslinked crystals tested here diffracted nearly as well as non-crosslinked crystals (anecdotally, a typical ~0.3 Å resolution difference). Additionally, we showed that this chemical ligation method is independent of the DNA sequence at the DNA–DNA junction. For example, despite differing DNA sequences at the junctions of CC1 and CC2, chemical ligation was effective in both cases. In summary, EDC ligation is a practical approach for crosslinking DNA inside of crystals and the optimized chemical crosslinking shown can provide the stability needed for diverse downstream applications.

4.3. Results

Appendix III contains corresponding supplemental figures to Chapter 4.

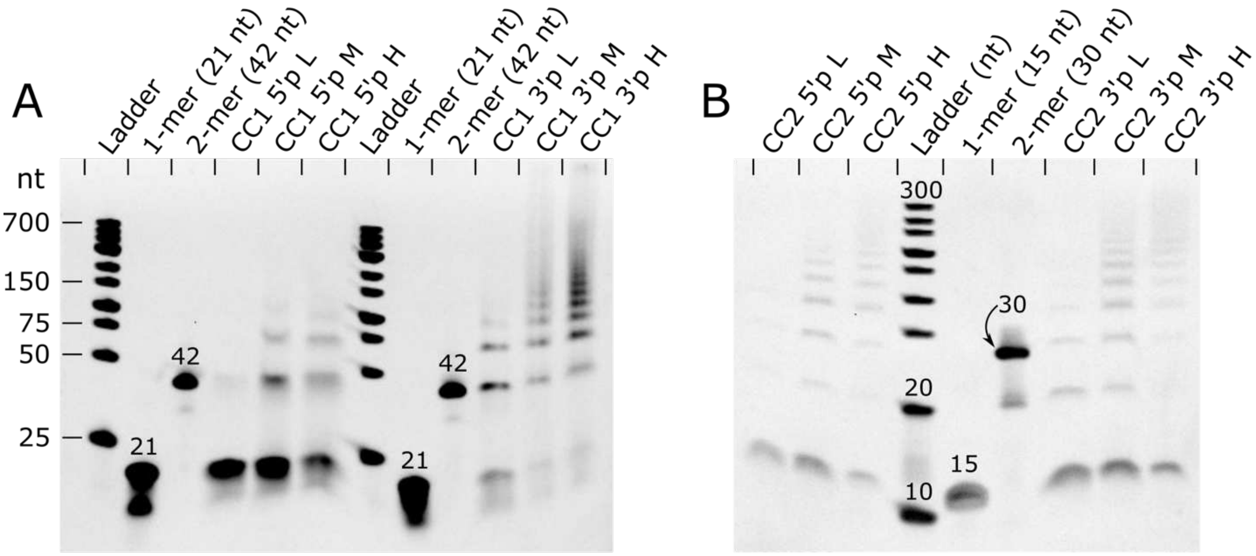
4.3.1. Chemical Ligation in Co-Crystals

Within our two co-crystal families (CC1 and CC2), we observed clear evidence of chemical ligation of stacked DNA duplexes. Both co-crystals demonstrated broadly similar ligation results, emphasizing the generality of this ligation method to co-crystals in which blunt-ended DNA blocks are suitably positioned to resemble contiguous DNA. As shown in Table 4.1, the PDB entries for the parent structures of both CC1 (7rva) and CC2 (4yo2) have junction step geometry that is reasonably comparable to contiguous B-DNA as calculated using x3DNA.[69] Except for the twist and roll across the CC2 junction (as seen in PDB entry 4yo2), all step geometry parameters are within 2 standard deviations of the B-DNA mean. It is possible that other co-crystals in which the DNA–DNA junctions have a geometry less like contiguous DNA would resist ligation. Additionally, in the preliminary crosslinking tests shown here, the crosslinking was successfully independent of the sequence at the DNA ends. CC1 has GC/CG flanking ends while CC2 has AT/TA flanking ends. The sequence independence of this ligation strategy is advantageous for DNA structure design projects where the junction sequence may be constrained for functional reasons. Table 4.3 also reports an interesting asymmetry between the two nick sites at the DNA–DNA junctions within the CC1 family of structures. We report the distance between C5′ and O3′ to avoid relying on the less certain O5′ position. For calibration, an idealized B-DNA model from x3dna had C5′ to O3′ distances of 2.73 Å for contiguous bases, but this span is variable (2.99 ± 0.17 Å) elsewhere within the dsDNA of PDB entry 7rva. One of the two CC1 nick sites, chain B, was invariably closer than chain A (*e.g.,* 3.75 Å rather than 4.22 Å in CC1-3′P), and electron density suggested that this shorter gap (chain B) was more readily ligated.

Table 4.1. DNA–DNA junction geometry parameters before and after phosphorylation and ligation. The likelihood of successful chemical ligation for stacked DNA may depend on geometry details across the junction. Here, we compare the geometry of the junction in the parent PDB models for CC1 and CC2, as well as the blunt-ended 5′ or 3′ phosphorylated CC1 crystals, to the geometry of contiguous bases in idealized B-DNA from Olson *et al.*, 1998.[76] The junctions are not symmetric, and differing distances for the two nicks across the junctions are also shown.

Junction Parameters from x3dna	**CC1**	**CC2**	**B-DNA***	**CC1-5′P**	**CC1-3′P**	**CC1-3′P EDC Heavy**
PDB code	7rva	4yo2		7sgc	7sdp	7spm
Base pair step parameters	GC/CG	AT/TA		GC/CG	GC/CG	GC/CG
Shift (Å)	−0.03	0.36	0.0 ± 0.51	0.10	0.05	0.50
Slide (Å)	−0.81	−1.27	0.35 ± 0.78	−1.00	−2.03 †	−0.52
Rise (Å)	3.49	3.61	3.32 ± 0.19	3.78 †	3.56	4.02 †
Tilt (°)	1.87	0.91	0.0 ± 3.4	2.51	0.72	1.72
Roll (°)	1.25	−15.84 †	1.4 ± 5.1	1.94	4.03	2.10
Twist (°)	36.64	20.95 †	35.4 ± 6.3	39.10	28.18	36.67
Nick distances C5′ to O3′						
Chain A: (C for CC2)	3.64	3.71	2.73	4.01	4.22	3.46 ‡
Chain B: (D for CC2)	3.45	3.86	2.73	3.39	3.75	3.30 ‡

*Base pair step parameters from Olson *et al.* 1998.[76] C5′ to O3′ distance from x3dna idealized B-DNA. †Values differ from B-DNA by more than 2 standard deviations. ‡Values are the distances in the refined “discovery” models prior to addition of the 3′ phosphate (not PDB 7spm).



L = Low Crosslink (1 dose 5 mg/mL EDC 12h)

M = Moderate Crosslink (1 dose 30 mg/mL EDC 12h)

H = Heavy Crosslink (2 doses 30 mg/mL EDC 12h)

Figure 4.4. TBE-urea gels of (**A**) CC1 and (**B**) CC2 chemical ligation. In both co-crystals, additional ligation was achieved with increased EDC concentration and a second EDC dose. (**A**) A 10% TBE-urea gel of CC1 illustrating a much-improved ligation product distribution for 3′ *vs.* 5′ phosphates. (**B**) A 15% TBE-urea gel of CC2 illustrating a modestly improved ligation product distribution for 3′ *vs.* 5′ phosphates. Assigned band sizes are given in nucleotides.

EDC crosslinking was tested for both 5′ and 3′ phosphate laden crystals. For CC1, the 3′ phosphate resulted in superior ligation yield over the 5′ phosphate in each trial (Figures 4.4 and S4.7–S4.11). On the other hand, CC2 ligation yields had a modest difference in the ligation yield for 3′ and 5′ phosphates. Given the limited dataset, it is premature to conclude that 3′ phosphates will typically give a higher ligation yield within co-crystals.

In both systems, DNA ligation was dependent on the presence of the terminal phosphates as well as on the crystal template; control crystals lacking terminal phosphates yielded no observable ligation products (Figure S4.10). Additionally, freely diffusing DNA blocks carrying terminal phosphates (but lacking the co-crystal scaffold) also yielded no observable ligation products when exposed to EDC (Figure S4.10). This second control demonstrated that the scaffold was necessary for ensuring efficient ligation of blunt-ended DNA blocks. The absence of observable ligation for building blocks in the absence of the crystal “scaffold”, precludes a systematic study of the effects of precursor ligation on crystal growth. Future work will determine, as a function of sticky overhang length, the extent to which blocks with sticky overhangs can be ligated within crystals and in solution.

Crosslinking reaction time was clearly and directly related to ligation reaction yield during the first 12 h (Figure S4.7). It was less clear if reaction yield was further improved by incubation beyond 12 h. Therefore, 12 h crosslinking incubations were used for the subsequent ligation optimization trials.

In the next series of experiments, we optimized EDC concentration for maximum ligation yield. We assayed the ligation product distribution as a function of concentration from 5 mg/mL EDC to 80 mg/mL EDC. As hypothesized, increasing the concentration of EDC increases the ligation of DNA in the co-crystals (Figures 4.4 and S4.8). In CC1 trials, we did not see a noticeable increase in ligation beyond 30 mg/mL. However, in CC2 trials, there was improved ligation at 60 mg/mL. We also subjected the co-crystals to multiple fresh doses of EDC (30 mg/mL) to determine if we could achieve near 100% ligation. For both co-crystal systems, multiple doses of EDC did increase ligation yields but qualitatively did not approach complete ligation (Figure S4.9).

Reaction buffer components were critical for successful ligation. We observed, at the outset of this project, that the presence of magnesium chloride in the crosslinking buffer appeared to interfere with the crosslinking reaction. This was problematic because the CC1 crystal growth conditions contain a significant amount of magnesium chloride. In our crystallization trials, 30–120 mM magnesium chloride was required for growth. Additionally, there is a structural Mg(II) at the DNA–protein interface coordinated by Glu77 and Asp81. To circumvent the apparent deleterious role of Mg(II) on CC1 crosslinking, we replaced magnesium chloride with sodium chloride in the wash solution for all CC1 crosslinking trials. At the conclusion of the project, we again confirmed that Mg(II) was deleterious to ligation by adding Mg(II) to the optimized ligation protocol. Specifically, we verified that supplementing the crosslinking incubation buffer with 90 mM or 110 mM $MgCl_2$ noticeably reduced the ligation yield (Figure S4.11). The exact role of Mg(II) in inhibiting the ligation reaction is not clear, but might involve reduced availability of the nucleophilic phosphate groups.

4.3.2. Ligation Model Compared to Experimental Co-Crystal Ligation

The ligation product distributions we experimentally obtained should shed light on the stochastic process of ligation. Using a destructive assay, densitometric analysis of electrophoresis results on ssDNA recovered from dissolved crystals, we quantified the population ratio of bands assigned to non-modified DNA strands as well as fused 2-mer, 3-mer, *etc.* For selected gels, we also obtained TapeStation results (Figure S4.1). The relative population of the TapeStation end-product distribution was fairly consistent with gel band populations measured with TBE-urea gels in ImageJ (Figure S4.1).

Next, we sought to calculate a global performance metric for the ligation yield, P_{LIG}, as the fraction of all possible DNA–DNA nick sites throughout a crystal that were ligated. To quantify the ligation yield throughout an entire crystal we analyzed the implications of the final DNA product distribution recovered after the crystal is dissolved and the protein components are removed. If we count the number of DNA oligos of each length (n_i) that were present in the crystal, and ignore edge effects we can estimate the total number of single strand breaks (SSB) as $N_{SSB} = \sum_i n_i$. For the same crystal, the estimated total number of original single strand breaks (regardless of final ligation status) would be $N_{JXN} = \sum_i i \cdot n_i$. For example, adding a single fused 3-mer to the crystal increases the SSB tally by one, but increases the tally of all possible junctions by three. Then, to compute the total probability of encountering SSB, we calculate:

$$P_{SSB} = \frac{N_{SSB}}{N_{JXN}} = \frac{\sum_i n_i}{\sum_i i \cdot n_i} = \frac{\sum_i n_i \big/ \sum_i n_i}{\sum_i i \cdot n_i \big/ \sum_i n_i} = \frac{1}{\sum_i i \cdot x_i} \qquad (1)$$

In the final equation, x_i is the mole fraction for i-mer oligos. Therefore, to estimate the P_{SSB}, we can use estimated mole fractions from electrophoresis and densitometry (Figures S4.2 and 4.4 and Table 4.2). Accurately calculating P_{SSB} does require including the small mole fractions for higher-order products (Table S4.3) since longer products contribute proportionally more to $\sum_i i \cdot x_i$. To estimate the uncertainty in each P_{SSB} calculation, we used 500 numerical trials in which random noise was added to $i \cdot x_i$ to mimic densitometry measurement error. We used noise comparable to $i \cdot x_i$ for the highest-order ligation products (normal variate with standard deviation 0.03), such that the smallest $i \cdot x_i$ values would regularly fall to 0 after the addition of random noise. Given the probability of encountering a single-strand break in the crystal (P_{SSB}) it is trivial to calculate the probability of each terminal phosphate having undergone ligation, since $P_{LIG} = 1 - P_{SSB}$. In the context of the random ligation model (RLM), ligation events throughout the crystal are independent and occur with equal probability at all nick sites. This is a physically plausible model if the intra-crystal transport rate for EDC exceeds the rate of reaction. Therefore, the incidence of double-strand breaks within the crystal should occur with the joint probability of independent events, $P_{DSB} = (P_{SSB})^2$.

Table 4.2. Distribution of DNA block sizes as a function of crosslinking protocol and 3′ vs. 5′ terminal phosphates. The data shown correspond with the gel lanes in Figure 4. The crosslinking protocols low, medium, and high were 1 dose of 5 mg/mL EDC for 12 h, 1 dose of 30 mg/mL EDC for 12 h, and 2 doses of 30 mg/mL EDC for 12 h each, respectively. The values in this table are weighted so that the DNA length and dye intensity contributes to the final value. Unweighted values are found in Table S2. The full table including estimated mole fractions for higher-order products is found in Table S3. P_{SSB}, $P_{LIG,}$ and P_{DSB}), were calculated for each crosslinked crystal sample. P_{SSB} uncertainties are the standard deviation in the calculated P_{SSB} after 500 trials in which simulated noise (standard deviation 0.03) was introduced into relative band intensities.

Parent Crystal	**CC1-3′P**	**CC1-3′P**	**CC1-3′P**	**CC1-5′P**	**CC1-5′P**	**CC1-5′P**
Crosslinking Protocol	low	medium	high	low	medium	high
DNA block size	[%]	[%]	[%]	[%]	[%]	[%]
1	58.7	30.0	24.9	98.6	91.6	82.1
2	18.7	16.8	14.9	1.4	7.3	9.9
3	15.2	15.3	15.6		1.0	6.3
4	5.0	11.0	11.0		0.2	1.5
5	2.4	6.3	8.4			0.2
6		6.5	6.9			
7		4.4	5.8			
8 and above		9.7	12.5			

Parent Crystal	**CC1-3′P**	**CC1-3′P**	**CC1-3′P**	**CC1-5′P**	**CC1-5′P**	**CC1-5′P**
Crosslinking Protocol	low	medium	high	low	medium	high
$P_{SSB}{}^{*}$	0.58±0.01	0.28±0.01	0.25±0.01	0.99±0.01	0.91±0.02	0.78±0.02
$P_{LIG} = 1 - P_{SSB}$	0.42±0.01	0.72±0.01	0.75±0.01	0.01±0.01	0.09±0.02	0.22±0.02
$P_{DSB} = (P_{SSB})^2$	0.33±0.01	0.08±0.01	0.06±0.005	0.97±0.02	0.83±0.04	0.61±0.03

Parent Crystal	**CC2-3′P**	**CC2-3′P**	**CC2-3′P**	**CC2-5′P**	**CC2-5′P**	**CC2-5′P**
Crosslinking Protocol	low	medium	high	low	medium	high
DNA block size	[%]	[%]	[%]	[%]	[%]	[%]
1	94.4	80.3	74.4	96.9	84.8	72.2
2	2.6	4.8	3.3	1.2	5.6	3.1
3	1.5	4.5	4.4	1.9	4.9	7.7
4	0.8	3.6	3.7		2.6	5.2
5	0.7	2.5	2.9		1.1	2.9
6		1.3	2.9		1.0	2.7
7		1.3	2.3			1.9
8 and above		1.7	6.1			4.3

Table 4.2 continued.

Parent Crystal	**CC2-3′P**	**CC2-3′P**	**CC2-3′P**	**CC2-5′P**	**CC2-5′P**	**CC2-5′P**
Crosslinking Protocol	low	medium	high	low	medium	high
P_{SSB}*	0.90±0.03	0.61±0.03	0.45±0.02	0.95±0.02	0.75±0.02	0.48±0.01
$P_{LIG} = 1 - P_{SSB}$	0.10±0.03	0.39±0.03	0.55±0.02	0.05±0.02	0.25±0.02	0.52±0.01
$P_{DSB} = (P_{SSB})^2$	0.82±0.05	0.37±0.05	0.20±0.02	0.91±0.04	0.57±0.03	0.23±0.01

*Calculated from experimental mole fractions per Equation 1. Other probabilities are calculated using the formulas shown. The double strand break probability estimate makes the assumption that ligation probability of both nicks at the same DNA:DNA junction are the same and independent. Uncertainty (Δ) propagation:

$$\Delta P_{DSB} = \sqrt{(2 \cdot P_{SSB} \cdot \Delta P_{SSB})^2}$$

This analysis of the electrophoresis experiments suggests that ~76% of the terminal phosphates within the most thoroughly crosslinked CC1-3′P crystal have undergone ligation. Furthermore, ~94% of the DNA–DNA junctions in this crystal had at least one ligated chain. The similarity in the ligation yield for the medium and high dose cases leads to an important question. What factors are limiting the yield? Incomplete ligation could result if a random population of terminal phosphates are missing, or otherwise incapable of on-target ligation. We used simulations to verify that the predicted RLM product ratio did not change when we postulated that a random subset of nick sites is incapable of ligation. This makes sense because junctions that are randomly selected to be incapable of ligation are functionally equivalent to sites that are randomly selected to be ligated last.

It may also be possible that ligating one phosphate at a DNA–DNA junction would negatively affect neighboring ligation probabilities. However, evidence for such allostery is lacking. Instead, the observed product distributions for CC1 ligation outcomes (Table 4.2), were close to the distributions predicted by the RLM (Figure S4.12). One small but consistent deviation from the RLM was a lower 2-mer, and higher 3-mer population than predicted. This observation seems to preclude the simplest negative allostery scenario (where one ligation event would reduce the probability at flanking sites). We cannot rule out the possibility that this discrepancy is an artifact associated with the gel electrophoresis densitometry.

The CC2 ligation outcomes (Table 4.2) were significantly less consistent with distributions predicted by the RLM. Once more, the 3-mer population was often higher than expected, sometimes exceeding the 2-mer population (which never happens in the RLM). This effect also seemed to extend to anomalously common 4-mers. A more striking divergence from the RLM prediction was the high population of non-ligated 1-mer blocks. Regardless of the RLM fit, the significant difference between the 1-mer mole fractions and the P_{DSB} values obtained from all the mole fractions strongly implicates that the RLM is lacking.

To investigate, we tested biased ligation model simulations. One possible explanation is that the ligation outcomes were driven partially by kinetics and molecular transport phenomena. Hypothetically, ligation sites near the crystal exterior might be more likely to be ligated than possible sites near the crystal center since reactive molecules must traverse the outer layers to react the interior. To determine the likely implications of this scenario, we conducted biased random ligation simulations (Protocol S4.3) that increased the probability of ligation events near the surface, decreased the probability at the center, and terminated the random ligation process at a set P_{SSB} threshold. Perhaps counterintuitively, this spatial bias increased the predicted 1-mer mole fraction. A high 1-mer fraction is partially consistent with the observed product distribution for CC2. The overall lower ligation yield achieved for CC2 crystals compared to CC1 is also consistent with the hypothesis that the CC2 crystal interior is systematically under-ligated. Alternately, it could be the case that one of the two symmetry-distinct nick sites in the CC2 lattice has a significantly lower ligation yield, and therefore one of the two DNA oligos will be over-represented in the 1-mer population.

4.3.3. Ligation Structural Details

Co-crystal structural details were revealed with X-ray diffraction at the Advanced Light Source beamline 4.2.2. Electrophoresis data (Figure 4.4) suggest that the CC2 DNA is stacked as intended. However, while high-resolution diffraction for CC2 crystals should be possible (3.07 Å reported by Morgunova *et al.*[102]), our CC2 crystals have, to date, yielded poor diffraction (>10 Å). Therefore, we chose to focus on the CC1 crystals as the model crystals to observe ligation *via* X-ray diffraction.

Here, we report five new crystal structures for CC1. We obtained a 1.89 Å dataset for the original co-crystal, which revealed additional details beyond the original model (PDB code: 1rep, 2.60 Å). Komori *et al.* varied the DNA building block to optimize resolution,[73] finding that dangling Ts resulted in the best data. Our updated structure provides a rationale for this empirical observation. Specifically, one of the dangling T bases is resolved, and participates in a crystallographic contact. Removing the dangling Ts decreased the resolution of our native structures from 1.9 Å to 2.7 Å (CC1-5′p) or 3.01 Å (CC1-3′p). Once crystals were crosslinked with low (15 mg/mL 12 h) and heavy (2 doses 30 mg/mL 12 h) EDC, the crystals maintained diffraction, albeit with a moderate loss in diffraction (3.14 Å and 3.28 Å, respectively).

Models were refined with PHENIX[78] and COOT.[79] The electron density for the heavily ligated DNA junction was consistent with contiguous DNA, despite omitting the terminal phosphate throughout prior refinement calculations. Figure 4.5 shows omit maps where any terminal phosphates are omitted, along with the bases flanking the junctions. The potential for overlapping electron density contributions from non-ligated and ligated phosphates makes it difficult to quantify occupancy. Nonetheless, we observed clear trends. Prior to ligation, the positions of 3′ phosphates (Figure 4.5C) or 5′ phosphates (Figure 4.5D) were reasonably clear.

Consistent with the lower distance between C5′ and O3′ for chain B (Table 4.1), the electron density was invariably higher for the right-hand nick (chain B:chain B). When contoured at 3.0 rmsd, the omit map electron density was even contiguous for the non-ligated CC1 3′P case (Figure 4.5C). Notably, the maps for crystals subjected to EDC (Figure 4.5E,F) are discovery maps in the sense that the models were refined in the absence of terminal 3′ phosphates. After light ligation (Figure 4.5E), the omit map was not clearly changed. However, after heavy ligation (Figure 4.5F), there was very strong electron density on the right and solid electron density in the left. Phosphates were added prior to submission to the PDB (entry 7spm) and our final refinement calculation for CC1 3′P High included bond length restraints between the model and its symmetry neighbor to ensure a reasonable phosphate geometry.

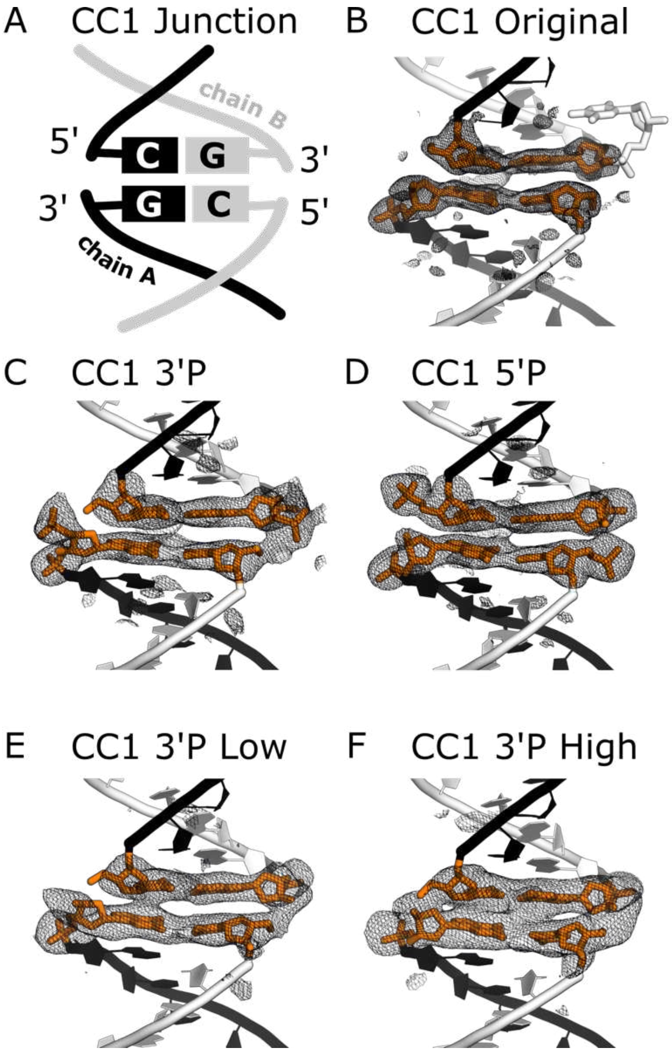


Figure 4.5. Omit maps for (**A**) the CC1 DNA–DNA junction. (**B**) Our updated model for the original structure resolves one of the dangling 5′ terminal bases (white sticks). Prior to ligation, CC1 crystals grow with either (**C**) terminal 3′ phosphates or (**D**) terminal 5′ phosphates. Whereas (**E**) low dose EDC ligation results in minor changes to the electron density for CC1 with 3′ phosphates, (**F**) high dose EDC ligation results in electron density consistent with ligated DNA. Neighboring protein is hidden for clarity. All meshes are omit maps (mF_o-DF_c) contoured at 3.0 rmsd. All four bases flanking the junction (orange sticks) were omitted. To faithfully represent COOT contours in PyMOL, we turned off automatic map normalization and instead set the contour level to 3.0 rmsd. Table S4.4 has the corresponding e/Å^3 values.

It is notable that ligation was visible in the electron density trend (Figure 4.5C–F), despite the incomplete ligation yield suggested by the electrophoresis data (Table 4.2). In principle, the clarity of the ligation sites in the electron density maps may vary depending on whether the X-ray beam is diffracting from a highly ligated region of the crystal.

4.3.4. Co-Crystal Stabilization Effects from Ligation and Crosslinking

To determine if crosslinked co-crystals may be suitable for various applications, including biomedical applications at physiologically relevant conditions, the co-crystals were crosslinked (20 h, 15 mg/mL EDC) and subjected to a panel of harsh conditions: a stomach acid mimic, a lysosomal fluid mimic, blood serum (bovine calf), and deionized water (Figure 4.6). The conditions chosen, especially the stomach acid mimic and deionized water, were challenging for native crystals (no crosslink) since DNA-containing crystals typically require stabilizing counterions.

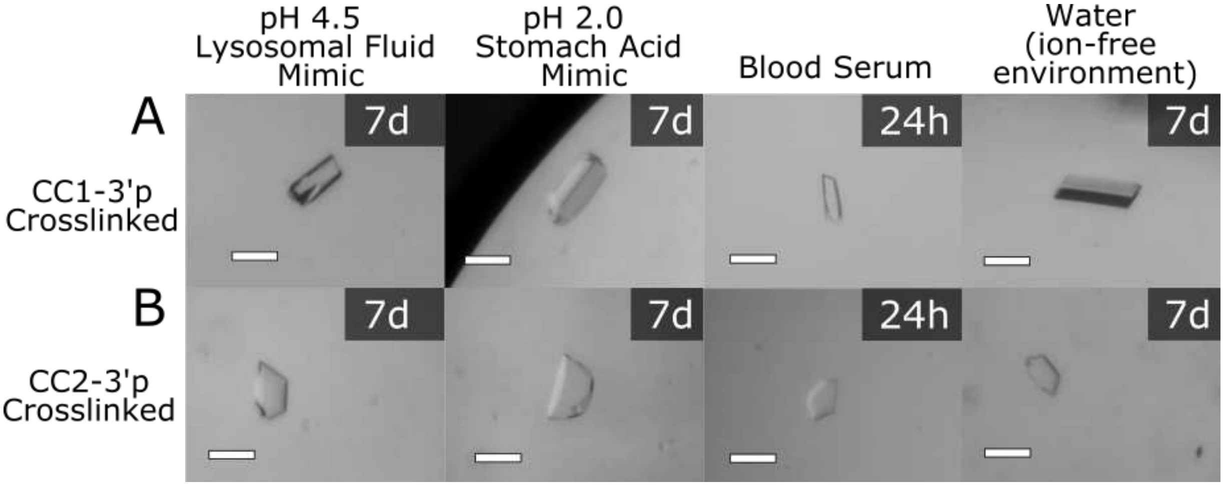


Figure 4.6. A survey of crosslinked crystals (15 mg/mL EDC 20 h) with terminal 3′ phosphates in four stringent solutions. (**A**) CC1-3′p crosslinked crystals incubated in pH 4.5, pH 2.0 and water for seven days and blood serum for twenty-four hours. (**B**) CC2-3′p crosslinked crystals incubated in pH 4.5, pH 2.0 and water for seven days and blood serum for twenty-four hours.

In the stomach acid mimic (0.01 M hydrochloric acid pH 2.0), the non-crosslinked co-crystals were observed to convert to an aggregate (Figure S4.4). Remarkably, in the stomach acid solution, the entire set of crosslinked crystals demonstrated enhanced stability, not dissolving even after 7 days. The 3′ phosphate crosslinked crystals did not change macro-structure for at least 5 days in the harshly acidic environment (Figure S4.4). Co-crystals without phosphates were also crosslinked and these crystals expanded dramatically in the acidic environment after 24 h (~430 ± 70% volume change), demonstrating the importance of the DNA ligation for crystal stability. Crosslinked co-crystals also maintained integrity in a lysosomal mimic buffer (pH 4.5) and blood serum with no measurable changes to the crystal dimensions after 24 and 72 h, respectively (Figures S4.5 and S4.6).

In deionized water, the co-crystal stability resulting from crosslinking was exceptional (Figures 4.7 and S4.3). Within one minute of transferring co-crystals to deionized water, non-crosslinked crystals (except for interesting exception CC2-3′P) completely dissolved or were converted to an aggregate. When the co-crystals were crosslinked (20 h, 15 mg/mL EDC), the crystals remained intact and lacked observable changes to their surface quality or dimensions for at least 7 days (Figures 4.7 and S4.3). Interestingly, crosslinked co-crystals without terminal phosphates remained unperturbed, just like the 3′ and 5′ phosphorylated crystals. These results indicate that the protein–protein crosslinks created within the co-crystals were sufficient to maintain macroscopic crystal integrity in water. The distinct stability of crosslinked crystals in water confirmed our hypothesis that crystals can be stabilized with new covalent crosslinks. Specifically, the non-covalent interactions that make up crystals can be stabilized with chemical crosslinking and prevent crystals from degrading rapidly in an ion-environment (deionized water).

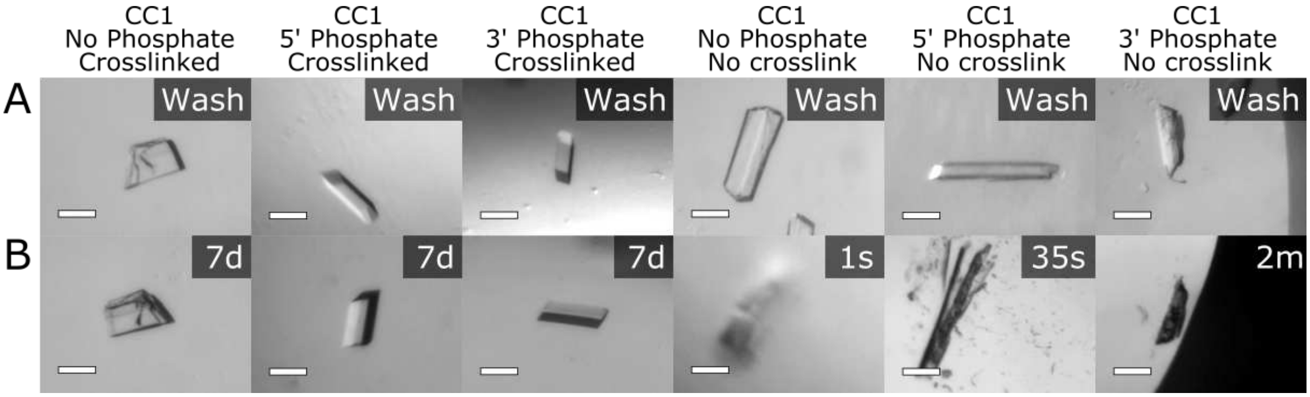


Figure 4.7. The crystals were crosslinked with 15 mg/mL EDC for 20 h and quenched with tris base pH 8.2 for 30 min prior to transfer to the wash solution. All scale bars are 100 µm. (**A**) CC1 crystals in wash solution containing 50 mM NaCl, 14% PEG 400, and 200 mM MES buffer pH 6.0. The concentrations of the wash solution matched the initial crystal growth solutions, but we replaced $MgCl_2$ with NaCl and tris HCl pH 8.0 with MES buffer pH 6.0. (**B**) CC1 crystals after transitioning to an ion-free environment (deionized water). The crosslinked crystals (left three panels) remained intact for 7 days. Non-crosslinked control crystals (right three columns) dissolved or converted to an aggregate at various immediate time points.

4.4. Discussion and Conclusions

Our strategy in this work was to identify an EDC ligation protocol (EDC concentration, incubation time, and repeated dosage regimen) that optimized reaction yield, without chasing diminishing returns. Accordingly, our final protocol uses 30 mg/mL EDC, an incubation time of 12 h, and two repeated doses within reaction sizes of approximately 200 microliters to ligate the DNA present within approximately 500 ng of co-crystals. Under these conditions, stacked DNA within co-crystals was reliably ligated to a significant extent. We used gel densitometry and detailed Gaussian peak fitting to estimate the fraction of the population for each ligated species (Figure S4.2).

Global analysis of the ligation product distribution suggested that the most thoroughly crosslinked CC1 crystals feature ligation of approximately 76% of all possible ligation sites, covalently linking about 94% of the DNA–DNA junctions through one or more covalent bond. Ligation was corroborated by single-crystal XRD where we could directly observe ligation in electron density omit maps (Figure 4.5).

Apart from small systematic deviations, the random ligation model (RLM, Protocol S4.2) was able to fit the ligation product distribution for CC1 (Figure S4.12). In contrast, the CC2 ligation results could not be fit to the RLM as accurately (Figure S4.12). In particular, the CC2 crystals appeared to have a 1-mer mole fraction that was significantly larger than the P_{DSB} estimated from all the mole fractions, which is inconsistent with the RLM. This could be explained by missing terminal phosphates, a disfavored ligation geometry for half of the nick sites, or by invoking transport limitations. Specifically, one way to boost the 1-mer mole fraction is if the exterior of the crystal has a higher ligation probability than the interior (Protocol S4.3).

Previously mentioned in the introduction, EDC ligation of DNA has been reported in the literature in the context of DNA hairpins in solution, liquid DNA crystals, and DNA origami. Notably, there has not been a consensus for whether 5′ or 3′ phosphate placement results in a superior yield. Fraccia *et al.* used 3′ phosphates for the EDC ligation of liquid DNA crystals,[100] whereas Kramer and Richert used 5′ phosphates for the EDC ligation of a DNA origami structure.[101] Giving a comparison of 5′ versus 3′ phosphates, Obianyor *et al.* showed EDC ligation of a hairpin DNA structure and reported 95% ligation yield for DNA with 3′ phosphates whereas the 5′ phosphates yielded 40% ligation.[99] They hypothesized the 3′ phosphate ligation reaction could benefit from a primary alcohol nucleophile (Figure 4.1B) and the geometry difference of the two phosphate positions could contribute to reaction yields. Our data suggest that 3′

phosphates may be superior in the context of a crystal, though comparison between CC1 and CC2 suggests that the results may be system dependent. Our XRD data (Figure 4.5) furthermore suggest that the results may vary for different nick sites within the same crystal.

It is not clear why 3′ phosphates were more readily ligated than 5′ phosphates in CC1. Conceivably, the rate limiting step for the ligation reaction may be the attack of the hydroxyl on the activated EDC intermediate. Perhaps the short-arm 3′-EDC intermediate is more accessible to the long-arm 5′ hydroxyl than a long-arm 5′-EDC intermediate is to a short-arm 3′ hydroxyl. Notably, one of the 3′ phosphates in the CC1 lattice (chain B) is close (5.65 Å) to a symmetry copy of itself (Figure S4.13), whereas the 5′ phosphate is farther (9.29 Å). Therefore, 3′ phosphate ligation might be favored due to the greater reduction in electrostatic repulsion upon ligation. However, the CC2 ligation results were more balanced (albeit still favoring 3′ phosphates), suggesting that the relative efficacy of 3′ or 5′ phosphates will be system dependent.

Analyses *via* gel electrophoresis showed that the ligation yield increased concomitant with the EDC incubation time, but also that the reaction yield appeared to plateau short of full ligation. The cause is unclear. Transport considerations and EDC conjugation to protein sites complicate reaction modeling. One consideration is that the predicted active half-life for EDC in water at 298 K is sixteen hours.[99] However, the ligation yields also appeared to plateau for repeated EDC dosing. Perhaps incomplete ligation is due to a small fraction of DNA strands lacking the necessary terminal phosphate. Alternately, perhaps some EDC-activated phosphates have been ligated to third-party molecules. Perhaps a small DNA strand population is missing a base.

Further investigation may be worthwhile prior to future work that depends on near 100% ligation.

In addition to optimizing conditions for our two co-crystals, we have established a set of generalizable guidelines for DNA ligation within co-crystals regarding optimal reaction conditions, phosphate composition, and concentration of EDC. First, it is imperative to optimize the wash solution for each respective system, eliminating components that could interfere with crosslinking. Reactive amines and carboxylic acids are obvious components to eliminate, to avoid forming off-target species. Additionally, we empirically found that it was important to minimize the concentration of the standard divalent cation Mg(II). Re-introducing 90–110 mM Mg(II) into our optimized protocol, we observed a dramatic reduction in the yield (Figure S4.11). Second, since we found that the best phosphate for ligation may depend on subtle geometry differences, we recommend testing both 5′ and 3′ phosphates for new co-crystal systems. Finally, the EDC concentration used for ligation of a new co-crystal may need to be optimized. Our co-crystals did not dissolve when introduced to crosslinking agents, with the highest concentration at 80 mg/mL. However, in past experiments, we found that the concentration of EDC in the crosslinking reaction drop can affect the integrity of co-crystals. Biomolecular crystals are typically fragile, and a drastic change in solution conditions can cause crystals to fall apart. Therefore, when working with a new system, we recommend testing a range of EDC concentrations. Dosing experiments may be necessary for systems that need a “gentle”, multistep transition to harsher conditions. These guidelines may apply to crystals composed of only DNA, as well.

With data for two example co-crystals, generalization is difficult. CC1 and CC2 differ in numerous ways (*e.g.,* DNA length of 21 bp *vs.* 15 bp, crystal space group, different base pairs spanning the DNA–DNA junction, different DNA sequences in general including flanking base pairs) which makes it difficult to determine which variables may be predictive of ligation yield. Given our observation that ligation may be very sensitive to the nick geometry (Table 4.1 and Figure 4.5), we hypothesize that several factors will be particularly important due to their influence on the nick geometry. The DNA sequence at the junction, and to a lesser extent the flanking bases, will affect the base pair stacking energy, which would be expected to change the nick geometry probability distribution. Other nick-site ligation yield differences may be driven by the crystallographic symmetry, particularly the presence or absence of neighboring groups in addition to intrinsic geometry differences between the nick site (*e.g.,* a slightly higher nick distance for chain A nick sites in CC1 crystals).

The crystal stability produced after the chemical ligation of stacked DNA within crystals opens the door for downstream applications, especially for DNA nanotechnology efforts. As shown here, even incomplete ligation can result in dramatic stabilization effects with tangible benefits to suitable application targets. No obvious EDC-induced crosslinks were visible at the two distinct protein–protein interfaces in the CC1 system. Further experiments will be needed to specifically seek and identify any EDC-induced protein–protein or DNA–protein conjugation.

It is possible that DNA ligation provided strong stabilizing effects because both CC1 and CC2 are held together by DNA–DNA junctions in two dimensions (Figures 4.6 and 4.7). Essentially, by ligating the stacked DNA in these cases we are forming longer “threads” that are woven together. If double strand breaks are distributed randomly with P_{DSB} = 0.06, then these threads might average over 100 nm in length (21 bp * 0.332 nm/bp * P_{DSB}^{-1}). Stabilization of devices or materials is intriguing if this stabilization allows them to provide or preserve functionality in various biomedical contexts (*e.g.,* in the digestive system, the blood stream, or within lysosomes). It may also be useful if crosslinking allows crystals to remain stable and diffract to high resolution under buffer conditions that mimic physiological conditions (*e.g.,* inside the nucleus), thereby allowing XRD structure determination under conditions besides the idiosyncratic conditions that allow for co-crystal growth.

Along the same lines, one traditional concern crystallographers have regarding crosslinking chemistry is that subjecting a crystal to handling, buffer changes, and reactive chemicals, can degrade the diffraction resolution. For example, subjecting crystals to the common crosslinking agent, glutaraldehyde, can rapidly degrade diffraction resolution. However, supplying aldehydes *via* gentle vapor diffusion[93] can improve outcomes. We have observed that using glyoxal and EDC can likewise result in negligible diffraction loss, particularly if the reactive chemistry is quenched.[15,45] In the case of CC1, we have once again found that carefully optimized crosslinking protocols can maintain diffraction. Another notable benefit of the EDC crosslinking method is that crystals were not “damaged” during the reaction chemistry. For comparison, when crosslinking HEWL crystals with glutaraldehyde, careful optimization was required to avoid forming cracks in the crystals.[103]

Future work may determine if the ligation yield differs for sticky overhang junctions compared to the blunt end junctions used in this work. Similarly, yield may also depend on the DNA bases that span the junction. That said, the current work suggests that the method may be sequence independent because the CC1 junction has a GC/CG and the CC2 junction has an AT/TA. In summary, the reported protocol is a reliable crosslinking strategy using the zero-length crosslinking agent EDC to affect DNA ligation at blunt-end DNA–DNA junctions held together by the co-crystal lattice. Post-ligation stability paves the way for biomedical applications.

4.5. Methods and materials

4.5.1. Protein Cloning, Expression, and Purification

The protein sequence (Protocol S4.1) of RepE54 initiator protein (CC1 protein) from PDB code 1rep was cloned into a PSB3 vector with a N-terminal 6-Histag.[73] The Histone Source at Colorado State University expressed and purified CC1 protein as follows. *E. coli* CodonPlus RIPL competent cells were transformed with the CC1 protein expression plasmid and grown at 37 °C to a density of OD_{600} 0.6 in 2xYT broth containing Ampicillin (100 mg/L) and Chloramphenicol (25 mg/L). Isopropyl-β-D-thiogalactoside (IPTG) was added at 0.4 mM and the culture was continually shaken at 37 °C for 3 h. Cells were harvested by centrifugation and resuspended in PBS buffer supplemented with 300 mM NaCl, 0.2 mM AEBSF, and 5 mM B-mercaptoethanol and were homogenized by sonication at 50% output (10 cycles of 45 s on, 120 s off). Lysate was recovered by centrifugation at 27,000× *g* for 25 min. The supernatant was loaded onto Ni Excel Sepharose resins (CV = 15 mL, Cytiva), washed and eluted by a linear gradient of 0–500 mM imidazole in resuspension buffer. The fractions containing CC1 protein were pooled, concentrated using Amicon Ultra-15 10 kDa MWCO centrifugal filter unit (EMD Millipore) and loaded onto a size-exclusion HiLoad Superdex 200 PG column (Cytiva) equilibrated with sodium citrate buffer (100 mM Sodium citrate pH 6.2, 100 mM KCl, 10 mM $MgCl_2$ and 10% glycerol). Fractions containing CC1 protein were collected, concentrated to 15 mg/mL, and stored at −80 °C after freezing with liquid nitrogen.

The E2F8 transcription factor (CC2 protein) plasmid was graciously donated by the Taipale Lab (Protocol S4.1). The protein was expressed and purified based on previous guidelines.[102] CC2 protein with a TEV protease-cleavable N-terminal thioredoxin tag was expressed with a T7 promoter in *E. coli* BL21(DE3) cells. Upon addition of 0.5 mM IPTG, the cells were outgrown at 25 °C for 20 h. The cell pellets were sonicated in lysis buffer and applied to HisTrap (HisPur^{TM} Ni-NTA Resin) equilibrated with HisTrap buffer (500 mM NaCl, 100 mM HEPES, 10 mM imidazole, 10% glycerol, 0.5 mM TCEP, pH 7.5). The protein was eluted with 200 mM imidazole in HisTrap buffer. CC2 protein was TEV cleaved from thioredoxin during dialysis using Snakeskin MWCO 10 kDa into HisTrap buffer. The cleaved product was separated from thioredoxin and TEV Protease by HisTrap, eluting with addition of HisTrap buffer. The CC2 protein was purified further with Nuvia^{TM} cPrime^{TM} Hydrophobic Cation Exchange Media, equilibrated with cation exchange buffer (50 mM NaCl, 100 mM HEPES, 10% glycerol, 0.5 mM TCEP, pH 7.5), and eluted with 100 mM NaCl in cation exchange buffer. The fractions containing CC2 protein were pooled, concentrated using Amicon Ultra-15 10 kDa MWCO centrifugal filter unit (EMD Millipore) and loaded onto a size-exclusion HiLoad Superdex 200 PG column (Cytiva) equilibrated with CC2 storage buffer (150 mM NaCl, 20 mM HEPES, 5% glycerol, 0.5 mM TCEP, pH 7.5). Size exclusion was completed at CSU's Histone

Source. Fractions containing CC2 protein were collected, concentrated to 10 mg/mL, and stored at −80 °C after flash freezing with liquid nitrogen.

All protein sample purification was analyzed with SDS-PAGE (NuPAGE^{TM} 4–12% Bis-Tris Gel) with MES SDS running buffer. Gels were stained with Imperial^{TM} Protein stain. Protein concentrations were determined with Bradford Assay using Coomassie Plus^{TM} Protein Assay Reagent.

4.5.2. DNA Duplex Annealing

DNA duplex sequences are given in Figure 4.2 and Table S4.1. The RepE54 co-crystal oligomers were designed from the original 22-mer in PDB code 1rep.[73,104] All sequences contained the 19 bp iteron sequence for DNA–protein binding, but the original duplex was truncated from a 22-mer to a 21-mer to eliminate an unresolved dangling base and to give a blunt ended DNA interaction for crosslinking. The E2F8 transcription factor co-crystal oligomers were the original duplex found in PDB code 4yo2.[102] The CC1 and CC2 oligomers were synthesized and HPLC purified by Integrated DNA Technologies with termini containing no phosphates, 5′ phosphates, or 3′ phosphates. The oligomers were resuspended: CC1 oligomers in 50 mM tris HCl, 100 mM KCl pH 7.0 and CC2 oligomers in 10 mM tris base, 150 mM NaCl, 1 mM EDTA pH 7.5. The DNA duplexes were annealed by combining cognate ssDNA oligomers in a 1:1 molar ratio, heating to 94 °C for 2 min then slowly cooling to room temperature over approximately 60 min. The final concentration of CC1 and CC2 duplexes were 4 mM and 1 mM, respectively. DNA stocks were quantified with a Qubit4 (Qubit^{TM} 1× dsDNA HS Assay Kit).

4.5.3. DNA–Protein Complex Co-Crystallization

All co-crystals were grown *via* sitting drop vapor diffusion. At 30 min prior to crystal plate setup, the protein and DNA were incubated at a 1:1.2 molar ratio. The DNA–protein complexes were kept on ice for 30 min prior to use. CC1, RepE54 transcription factor co-crystal, crystallization conditions were 30–120 mM $MgCl_2$, 2–16% PEG 400 and 100–220 mM tris HCl pH 8.0. CC2, E2F8 transcription factor co-crystal, crystallization conditions were 40–300 mM ammonium sulfate, 5% PEG 400, 5–20% PEG 3350, and 80 mM HEPES pH 7.1. Crystals grew to a span of 50–150 µm in a range of 24 h to 7 days.

4.5.4. EDC Crosslinking Co-Crystals

Co-crystals were washed in conditions similar to crystal growth conditions where growth buffer components that interfere with crosslinking were substituted (*i.e.,* primary amines, carboxylic acids, and divalent cations). The CC1 wash solution consisted of 30–120 mM NaCl (substituting for $MgCl_2$), 2–16% PEG 400 and 100–220 mM MES pH 6.0 (substituting for tris HCl pH 8.0). The CC2 wash solution consisted of 20–300 mM lithium sulfate (substituting for ammonium sulfate), 5% PEG 400, 10–30% PEG 3350 (an increase of 10% PEG 3350 compared to the growth solution), and 80 mM MES pH 6.0 (substituting for HEPES pH 7.1). The 10% additional PEG 3350 for CC2 appeared to prevent the crystals from degrading upon addition of the wash. The co-crystals were washed in 9-well glass plates (Hampton) to remove additional protein and DNA monomers and unwanted buffer components. 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) (Advanced Chemtech CAS#:25952-53-8) was resuspended in the wash solution to final concentration values ranging from 5 to 80 mg/mL and used immediately. The co-crystals were crosslinked in a 200 µL EDC solution volume for varying time points. The co-crystal crosslinking reaction was quenched by moving crystals to 1× Tris-Borate-EDTA (TBE) buffer pH 8.3 containing 3.5 M urea.

4.5.5. DNA Gel Electrophoresis and Densitometry

Crosslinked co-crystals were dissolved in 1 × TBE supplemented with 3.5 M urea and 20 μg proteinase K and incubated at 50 ºC for 12 hours. When crystals were too robust to dissolve under these harsh conditions, the crystals were heated to 94 °C for 1 h and glass crystal crushers (Hampton) were used to crush the crystals prior to chemical and enzymatic attack. The crystals were analyzed with denaturing gels (10% or 15% Novex^{TM} TBE-Urea Polyacrylamide Gel Electrophoresis) formulated with 7M urea and run with 1× TBE running buffer. DNA ladders were GeneRuler Low Range DNA Ladder (Thermo Scientific, Houston, USA) for CC1 gels and Ultra Low Range DNA Ladder (Invitrogen) for CC2 gels. The control lanes included 1-mer DNA and 2-mer DNA. While the gel conditions are expected to denature the DNA and yield ssDNA, the control lanes were first annealed as described in Section 2.3 (Table S4.1 Duplex IDs 1.1, 1.4, 2.1, and 2.4). Gels were incubated with 3 × GelRed^{TM} Nucleic Acid Gel Stain and imaged with a UVP Bioimaging System on the Ethidium Bromide setting. For further validation, selected ligation products for CC1 were also analyzed with a TapeStation D1000 ScreenTape assay (Agilent) (Figure S4.1) at CSU's Next Generation Sequencing Core. The gels and TapeStation were analyzed *via* densitometry.

4.5.6. DNA Gels and Densitometry

For densitometry analyses, we used ImageJ (1.52 k) to obtain raw x,y,intensity values for the gels shown in Figure 4.4. We averaged these data over x values and used custom Python scripts (within “cocrystal_ligation_scripts.zip” hosted on Zenodo[105]) as well as the lmfit module[105] to obtain non-linear best fits of the gel intensity. Specifically, we modeled peaks using Gaussian functions. We also modeled the background using diffuse Gaussian functions. Crystals with more crosslinking produced overlapping gel bands for higher-order ligation products. One of the benefits of using a mathematical curve fitting framework is our ability to fit (albeit approximately) these populations. Specifically, we fit the peak position trend using the well-separated gel bands corresponding to smaller ligation products. Then, we fit the highly overlapping region using extrapolated peak positions with fitting parameter restrictions implemented *via* lmfit. Inspection of the fitting results (Figure S4.2) gave us confidence that higher-order band intensity fit was reasonable.

In principle, longer DNA ligation products can adsorb a greater number of GelRed fluorophores, proportionally with the DNA length. Ignoring this effect might cause us to overestimate the ligation yield. Accordingly, we proceeded to normalize the estimated molar ratio of the ligation products (Table 4.2) by dividing each band intensity by the assigned DNA block size (divide by N for N-mer DNA blocks). The raw band intensity fits are provided in Table S4.2.

4.5.7. Random Ligation Model

As discussed in Section 4.3.2, the densitometry data could be interpreted in terms of the relative population of unfused DNA, ligated 2-mer, ligated 3-mer, *etc.* We sought to interpret these data in terms of the ligation yield. First, we used the estimated molar ratio of products from gel densitometry to estimate the fraction of potential ligation sites that were ligated. Second, to compute the expected distribution of fused DNA blocks of varying length, we implemented a simple 1D simulation in Python (Protocol S4.2, also within "cocrystal_ligation_scripts.zip" [105]) in which all 42856 nicks between DNA blocks in a 1D stack of 42,857 strands (a 300 micron stack) were equally likely to be randomly removed in each unit of time. This "random ligation model" is arguably the least complex theoretical model for the crosslinking process, ignoring transport phenomena and assuming that all possible ligation sites throughout the crystal undergo ligation randomly with equal probability *per* unit of time. We also developed a biased ligation model in which sites near the crystal interior are less likely to be ligated than sites near the crystal surface (Protocol S4.3, also within "cocrystal_ligation_scripts.zip" [105]).

4.5.8. X-ray Diffraction Data Collection, Refinement and Omit Maps

Single-crystal X-ray diffraction (XRD) data were collected for CC1 crystals containing 5′ and 3′ terminal phosphates. Crosslinked crystals with 3′ terminal phosphates were also analyzed *via* XRD. Crystals were briefly swished through cryo-protectant solution (300 mM $MgCl_2$, 30% PEG 400, and 100 mM tris HCl pH 8.0) and flash-frozen in liquid nitrogen. Frozen crystals were stored in Rigaku ACTOR Magazines (Mitegen) and shipped to the Advance Light Source Beamline 4.2.2 for data collection. Full datasets were collected on a CMOS detector from 0 to 180 degrees with an omega delta of 0.2° and an exposure time of 0.3 s. Data were processed with XDS[77] and molecular replacement and refinement within PHENIX[78] and COOT.[79] As a result, the original co-crystal for RepE54 transcription factor (2.60 Å PDB code 1rep) was updated with a higher-resolution structure (1.89 Å PDB code 7rva). The updated structure was solved with molecular replacement using the PDB code 1rep. CC1 crystal structures containing 5′ or 3′ terminal phosphates were solved *via* molecular replacement in PHENIX using the updated original CC1 as a starter model. For all structures, the same R-free flags were used during refinement in PHENIX and COOT. Structure factor data were truncated using I/sigma(I) >1.5 as a cutoff. The resulting structures were of: CC1 with terminal 5′ phosphates (PDB code: 7sgc), CC1 with terminal 3′ phosphates (PDB code: 7sdp), low EDC crosslinked (5 mg/mL EDC, 12 h) CC1 with terminal 3′ phosphates (PDB code: 7soz), and heavy EDC crosslinked (30 mg/mL EDC, 12 h, two doses) CC1 with terminal 3′ phosphates (PDB code: 7spm). Standard X-ray diffraction data quality statistics are provided in Tables 4.3 and 4.4.

Omit maps were generated for each structure, shown in Section 4.3.3. To prevent bias of the electron density at the junctions in crosslinked structures, discovery and omit maps were generated with structures containing no terminal phosphates. After generating discovery and omit maps, the terminal phosphates were added to the structures and refined for submission to the PDB. In the final PHENIX refine of heavy crosslinked CC1 terminal 3′ phosphates, a custom geometry bond restraint was added because the electron density indicated ligation at both junctions. The terminal 3′P and flanking 5′OH were given a bond length restraint of 1.59 Å, the ideal length of the phosphate-oxygen bond in the DNA backbone.[107]

Table 4.3. X-ray diffraction statistics for the updated original CC1 crystal, the CC1 crystal with terminal 5′ phosphates, and the CC1 crystal with terminal 3′ phosphates.

	Updated CC1 Original **PDB Code 7rva**	**CC1 5′p** **PDB Code 7sdp**	**CC1 3′p** **PDB Code 7sgc**
Data collection			
Light source	Synchrotron	Synchrotron	Synchrotron
Wavelength (Å)	1.0	1.0	1.0
Resolution range (Å)	33.78–1.89 (1.958–1.89) *	34.45–2.7 (2.796–2.7) *	37.68–3.01 (3.118–3.01) *
Space group	C 1 2 1	C 1 2 1	C 1 2 1
Unit cell dimensions			
a, b, c (Å)	107.578 80.715 73.299	109.161 83.051 74.349	109.142 83.089 74.593
α, β, γ (°)	90 122.63 90	90 123.948 90	90 122.691 90
Total reflections	152,923 (15,187)	51,877 (5051)	40,979 (3834)
Unique reflections	41,778 (2501)	14,436 (1425)	11,132 (1074)
Multiplicity	3.7 (3.6)	3.6 (3.5)	3.7 (3.5)
Completeness (%)	90.98 (59.52)	93.97 (92.88)	98.37 (94.79)
Mean I/sigma(I)	10.77 (1.71)	13.56 (1.69)	11.45 (1.67)
Wilson B-factor	32.28	61.1	57
R-merge	0.05735 (0.7813)	0.06375 (0.7563)	0.1406 (1.383)
R-meas	0.06726 (0.9191)	0.07496 (0.8954)	0.1652 (1.638)
R-pim	0.03484 (0.4793)	0.03905 (0.4741)	0.08579 (0.8663)
CC1/2	0.998 (0.749)	0.997 (0.669)	0.993 (0.574)
CC *	1 (0.926)	0.999 (0.895)	0.998 (0.854)
Refinement			
Reflections used in refinement	38499 (2501)	14342 (1422)	11060 (1074)
Reflections used for R-free	2007 (132)	752 (78)	586 (55)
R-work	0.2074 (0.5745)	0.1866 (0.3206)	0.1820 (0.3258)
R-free	0.2490 (0.5833)	0.2457 (0.4038)	0.2351 (0.3987)
CC (work)	0.972 (0.374)	0.968 (0.792)	0.973 (0.730)
CC (free)	0.937 (0.202)	0.964 (0.578)	0.929 (0.546)
Number of non-hydrogen atoms	2970	2800	2825
Macromolecules	2766	2770	2776
Ligands	3	3	2
Solvent	201	27	47
Protein residues	228	230	232
RMS (bonds) (Å)	0.009	0.011	0.012
RMS (angles) (°)	1.11	1.3	1.35
Ramachandran favored (%)	98.14	95.85	95.43
Ramachandran allowed (%)	1.86	4.15	4.57
Ramachandran outliers (%)	0	0	0
Rotamer outliers (%)	0	0	0
Clashscore	3.14	8.45	8.41
Average B-factor	55.73	75.32	62.28
Macromolecules	56.07	75.49	62.69
Ligands	55.14	78.12	52.05
Solvent	51.08	57.85	38.7
Number of TLS groups	10	10	10

* Values in parentheses are for high-resolution shell.

Table 4.4. X-ray diffraction statistics for the CC1 crystal with terminal 3′ phosphates and low crosslink (5 mg/mL EDC for 12 h) and the CC1 crystal with terminal 3′ phosphates and heavy crosslink (two doses of 30 mg/mL EDC for 12 h).

	CC1 3′p Low EDC Crosslink 5 mg/mL EDC for 12 h PDB Code 7soz	CC1 3′p Heavy EDC Crosslink 2 Doses of 30 mg/mL EDC for 12 h PDB Code 7spm
Data collection		
Light source	Synchrotron	Synchrotron
Wavelength (Å)	1.0	1.0
Resolution range (Å)	37.05–3.14 (3.252–3.14) *	33.78–3.28 (3.397–3.28) *
Space group	C 1 2 1	C 1 2 1
Unit cell dimensions		
a, b, c (Å)	111.993 79.181 74.903	110.817 80.174 74.723
α, β, γ (°)	90 123.023 90	90 122.899 90
Total reflections	34,361 (3280)	31,305 (3106)
Unique reflections	9433 (884)	8467 (830)
Multiplicity	3.6 (3.6)	3.7 (3.7)
Completeness (%)	95.82 (90.02)	97.95 (93.58)
Mean I/sigma(I)	10.37 (2.22)	8.42 (2.09)
Wilson B-factor	83.36	94.48
R-merge	0.08469 (0.5434)	0.1187 (0.6541)
R-meas	0.09971 (0.6391)	0.1394 (0.7672)
R-pim	0.05208 (0.3334)	0.0724 (0.3973)
CC1/2	0.997 (0.946)	0.998 (0.94)
CC *	0.999 (0.986)	0.999 (0.984)
Refinement		
Reflections used in refinement	9290 (857)	8356 (802)
Reflections used for R-free	498 (49)	448 (40)
R-work	0.1936 (0.3019)	0.2010 (0.3183)
R-free	0.2637 (0.3575)	0.2528 (0.3556)
CC (work)	0.982 (0.891)	0.985 (0.869)
CC (free)	0.915 (0.899)	0.987 (0.932)
Number of non-hydrogen atoms	2691	2649
Macromolecules	2687	2641
Ligands	1	1
Solvent	3	7
Protein residues	224	221
RMS (bonds) (Å)	0.012	0.015
RMS (angles) (°)	1.49	1.63
Ramachandran favored (%)	91.28	91.16
Ramachandran allowed (%)	8.72	8.37
Ramachandran outliers (%)	0	0.47
Rotamer outliers (%)	0	0
Clashscore	14.36	14.92
Average B-factor	109.8	130.22
Macromolecules	109.91	130.43
Ligands	30	38.22
Solvent	38.28	63.68
Number of TLS groups	10	10

* Values in parentheses are for high-resolution shell.

4.5.9. Stability Assays

Crystals were crosslinked using the Section 4.5.4 protocol, with 15 mg/mL EDC for 20 h. The EDC reaction was quenched in 50 mM tris base pH 8.0 for 30 min. The crystals were equilibrated in crosslinking wash solution for 30 min prior to looping to stringent conditions. The stability test buffers used were as follows: molecular biology grade water (CORNING), very low pH 2.0 0.01 M HCl buffer (to mimic stomach acid), a moderately low pH 4.5 citrate buffer (46 mM sodium citrate, 54.1 mM citric acid to mimic lysosomal fluid pH), and blood serum (HyClone, bovine calf serum). Pictures for each trial are in Figures S4.3–S4.6. Crystal pictures were obtained with a Moticam 3.0 MP camera attached to a Motic SMZ-168 stereozoom microscope and crystal measurements were performed in Motic Images Plus 2.0 (Figure S4.4 and Protocol S4.4).

4.6. Funding

This material is based upon work supported by the National Science Foundation under Grant No. NSF DMR 2003748 and NSF DMR 1506219. The team also gratefully acknowledges support for undergraduate researchers from the Nelson Family Faculty Excellence Award.

CHAPTER 5. TUNING CHEMICAL DNA LIGATION WITHIN DNA CRYSTALS AND CO-CRYSTALS[5]

5.1. Overview

Biomolecular crystals can serve as materials for a plethora of applications including precise guest entrapment. However, as grown, biomolecular crystals are fragile and destroyed easily in solutions other than their growth conditions. For crystals to achieve their full potential as hosts for other molecules, crystals can be made stronger with bioconjugation. Here, we optimized chemical ligation using the carbodiimide 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) within two classes of biomolecular crystals: co-crystals of DNA-binding proteins and pure DNA crystals. Both crystal classes contain DNA junctions where DNA strands stack up end-to-end. Ligation yields were studied as a function of sticky base overhang length and terminal phosphorylation status. The best ligation performance for both crystal classes was achieved with longer sticky overhangs and terminal 3'phosphates. Notably, EDC chemical ligation was achieved in crystals with pore sizes too small for intra-crystal transport ligase enzyme. X-ray diffraction for co-crystals post-ligation showed relatively similar diffraction quality to native co-crystals (~0.5 Å diffraction loss on average). Furthermore, post-assembly crosslinking produced dramatic stability improvements for

---

[5] The work in this chapter is formatted for submission as a research article. Conceptualization, writing-drafting, and visualization, A.R.O., C.D.S; writing-editing, A.R.O., A.V., E.T.S., C.D.S.; formal analysis, data curation, acquisition, investigation, and validation; A.R.O., A.V., S.D., E.T.S., R.S., C.D.S.
Orun, A.R.; Vajapayajula, A.; Dmytriw, S.; Shields, E.T.; Shrestha, R.; Snow, C.D. Tuning chemical DNA ligation within DNA crystals and co-crystals. *Pending submission.*

both DNA crystals and co-crystals in water and blood serum. The results presented may help crystals containing DNA to achieve broader application utility, including as scaffolds for structural biology.

5.2. Introduction

Crystals made from protein and DNA are versatile materials that precisely order molecules, self-assemble, and have tunable growth.[52,71,108-109] Porous crystals were shown to host molecules for structure determination,[8,45,56,61] enhance enzymatic activity[5,21] and serve as molecular sieves.[2,13,22,110] Engineered crystals with DNA building blocks, from pure DNA crystals[24] to hybrid protein-DNA crystals (Chapters 2,3) have been designed to serve as a scaffold for DNA-binding molecules.[1,90] However, crystals made from biological units are held together by weak non-covalent bonds that are only stable in very specific solutions. As grown, crystals are likely to dissolve when introduced to anything other than their growth solution. Even more, crystals with DNA building units are particularly sensitive to changes in divalent cations. Crystal stability must be enhanced to fully realize the potential of crystals as biomaterials[25] and molecular scaffolds.[1,24,90] Herein, we optimize a chemical DNA ligation strategy to provide stability for crystals containing DNA stacks (both pure DNA crystals and protein-DNA co-crystals). Enhanced stability makes crystals robust in water and blood serum for biomedical applications. The chemical ligation yields shown in DNA crystals set a new record, exceeding previous biological ligation yields and providing a method for crystalline DNA junction ligation unhindered by the crystal porosity or steric accessibility to the DNA-DNA junctions.

Bioconjugation, also called crosslinking in this context, is a technique to make crystals strong for application utility[92] by introducing covalent bonds between the neighboring units in the crystal. Crystal crosslinking techniques for crystals containing DNA are established in the literature.[95-97] However, most of the established protocols involve highly reactive DNA crosslinking agents that introduce permanent links quickly but with unnatural bonds and severe handling disadvantages.[95]

An alternative DNA crosslinking strategy is ligation, a zero-length bond between neighboring DNA-DNA phosphate backbones. In crystals where DNA lines up end-to-end, either with a blunt or sticky base overlap, ligation can heal the nick in the DNA. In other words, upon ligation, two neighboring duplexes become one contiguous piece of DNA. Under the right pH and solution conditions, chemical ligation can be achieved with the reactive carbodiimide crosslinker, 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC).[100] Recently, we reported EDC ligation in co-crystals containing stacking DNA blocks.[74] The questions remained: Can we achieve ligation in purely DNA crystals? What role does sticky base overhang length and phosphorylation status play in ligation of co-crystal and DNA crystals? The crosslinking of both classes of crystals containing DNA may lead to applications in therapeutic deliveries and scaffold-assisted structure determination.

Advantages of EDC ligation include reactivity regardless of crystal porosity, maintained crystal nano-scale assembly and X-ray diffraction quality, and reaction tunability for the crystal target and crystal application. Notably, EDC ligation forms a zero-length crosslink, chemically identical to the biological ligation achieved with T4 DNA ligase. EDC reacts with a phosphate to form an intermediate, and the leaving group is removed by the hydroxyl on a neighboring strand to form a healed phosphate bond and a contiguous strand of DNA (Chapter 4 Figure 4.1).

Li *et al.* previously ligated *porous* DNA crystals with a biological approach, T4 DNA ligase.[98] Now, we show EDC chemical ligation in DNA crystals, a method feasible for crystals with smaller pore size. We also report successful ligation of each symmetry distinct junction within tensegrity triangle crystals. The optimized protocol for co-crystal and DNA crystal ligation includes a careful buffer exchange. For engineering purposes, EDC ligation opens the flexibility of terminal phosphorylation status (vary 5' or 3' phosphate) and the nick site can hypothetically be in a less accessible site (compared to the case for T4 DNA ligase).

5.3. Results and Discussion

Appendix IV contains corresponding supplemental figures to Chapter 5.

5.3.1. Chemical ligation in Co-crystals

In our first study of EDC ligation templated by co-crystals,[74] two example co-crystals in which the DNA stacks up end-to-end were used as preliminary examples for *in crystallo* EDC ligation. One of these ligation targets, Co-crystal 1 (CC1), was the co-crystal of Replication Initiator Protein RepE54 and cognate DNA sequence. The original instance of CC1 (RepE54 and 21mer with 3' dangling T's) contained DNA duplexes stacking end-to-end. The structure was improved from 2.6 Å (PDB code: 1rep)[73] to 1.89Å (PDB code: 7rva).[74] For the purposes of studying ligation in co-crystals, the original dangling T's were truncated to make a blunt ended 21mer (PDB codes: 7sgc and 7sdp). The blunt ended CC1 was a canvas to study chemical EDC ligation in co-crystals and the role of terminal phosphates. The blunt ended junction was decorated with 5’ or 3’ terminal phosphates. In the case of CC1, the 3’ terminal phosphates ligated more readily than 5’ phosphates. Our next question became what is the role of sticky base overhangs in EDC ligation?

Recently, we designed an expanded, interpenetrating CC1 crystal lattice to serve as a molecular scaffold (Chapter 3). To vary the sticky end lengths at the DNA junctions, we used the designed CC1 with ten additional base pairs (CC1^{+10bp}) (Figure 1). Specifically, expanded duplexes (31mer) were crystallized with the CC1 RepE54 protein to form the template co-crystals used in this study (Figure 1A). For the purposes of studying ligation yields with respect to variations in the junctions, we varied sticky overhang length (0, 1, 2, and 3 nt overhang) and terminal phosphorylation status (3’ or 5’ phosphate) (Figure 1B). Despite the plethora of variations at the junctions, the co-crystals grew in the same *I*222 space group and contained the same contacts with symmetry copies (Table S5.1). Therefore, CC1^{+10bp} variants could be compared side-by-side for ligation yields.

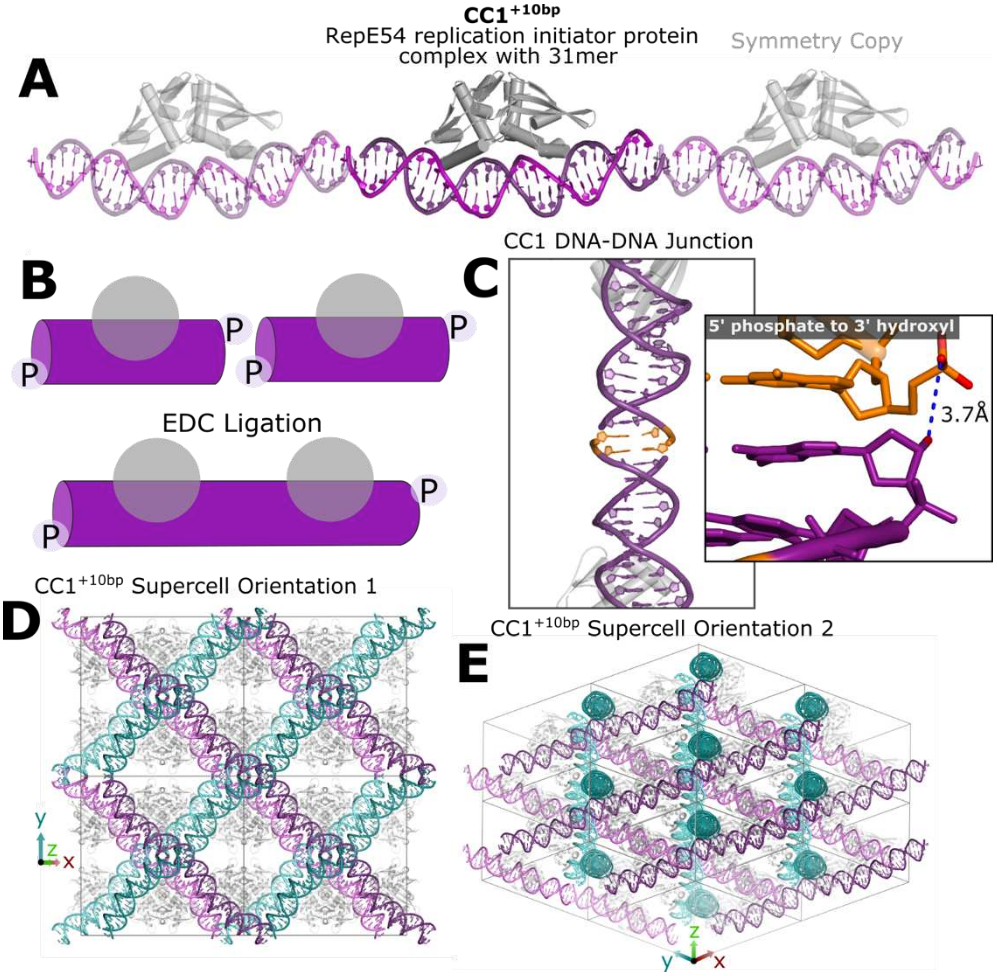


Figure 5.1. The protein-DNA co-crystal used to study the effects of sticky base overhang length on EDC ligation. (A) Replication initiator protein complexed to a 31mer cognate DNA sequence with 10 additional flanking base pairs, named CC1^{+10bp}. (B) A schematic view of the DNA blocks in the crystal with terminal phosphates. Upon ligation, the blocks become a single block of DNA. (C) The DNA-DNA junction of CC1^{+10bp} with 2 sticky base overhangs and 5' terminal phosphates (PDB code: 7rva). The distance between the 5' phosphoryl oxygen and the neighboring 3' hydroxyl oxygen is measured in PyMOL as 3.7Å. We varied the sticky base length (0-3 SB) and terminal phosphorylation status (5' or 3' phosphates) to study ligation yields as a function of both variables. (D) Supercell orientation 1 shows four distinct versions of the DNA stack directions in the crystal, highlighted in four different colors. (E) Supercell orientation 2 shows another view of the four different DNA stack directions in the crystal.

First, a gentle buffer exchange from the crystallization conditions to a crosslinking wash was necessary to remove compounds that interfere with EDC ligation and supernatant protein-DNA monomers. Three components of the CC1^{+10bp} crystallization conditions (300-500 mM $MgCl_2$, 25-35% PEG 400, and 100 mM tris HCl pH 8.0) were modified. First, divalent cations were previously found to interfere with co-crystal ligation.[74] Therefore, Mg(II) was replaced 1:1 stoichiometrically with Na(I). Second, to adjust the pH from 8.0 to 6.0 to increase the phosphate reactivity with EDC,[92] the main crystallization buffer, tris HCl, was replaced with MES hydrate. After a buffer exchange, the crystals maintained their visible macroscopic structure and were transferred to crosslink wash containing EDC for ligation.

Parallel crosslinking trials were performed for crystals with varied sticky base overhang lengths and terminal phosphorylation status. Crystals were crosslinked with the optimal EDC ligation conditions found previously (30 EDC mg/mL for 12h) and results were compared after one and two doses of EDC (Figure 5.2). We hypothesized that longer sticky base overhangs would improve EDC ligation yields because the sticky end cohesion would orient the phosphate and flanking hydroxyl for the EDC reaction. The ligation yields (Figure 5.2 and Table 5.2) were quantified with TBE-urea gel electrophoresis and gel densitometry (our custom Python scripts published on Zenodo[105]) and ligation product calculations described previously (Chapter 4 and Table 5.1).[74] Figure 5.2A shows the first dose of EDC ligation with varied sticky base overhangs (0, 1, 2, and 3 nt.) and phosphorylation (5'P *vs.* 3'P). Independent of sticky base overhang length, the 3'phosphates had higher ligation yields than 5' phosphates. Better yields with 3' terminal phosphates over 5' terminal phosphates in the blunt ended CC1^{+10bp} (zero sticky bases) concurred with our findings in the native, blunt-ended CC1 crystals.[74] In EDC dose 1, the 0SB and 1SB with 3' terminal phosphates had slightly better yields than the 2SB and 3SB. After two EDC doses, the effects of sticky overhangs became more prominent, especially for the 3'phosphates. After the second dose of EDC, the 2SB overhang crystal yielded the highest ligation product for the $\text{CC1}^{+10\text{bp}}$ crystal variants, with ~54% ligation (Table 5.2).

Figure 5.2. TBE-urea gels of (A) CC1 after one EDC dose and (B) CC1 after two EDC doses. Additional ligation was achieved with longer sticky overhangs and 3' phosphates after two doses. (A) A 10% TBE-urea gel of CC1 illustrating a slightly-improved ligation product distribution for 3' *vs.* 5' phosphates. (B) A 10% TBE-urea gel of CC1 illustrating a dramatically improved ligation product distribution for 3' *vs.* 5' phosphates. Assigned band sizes are given in bp. Gel densitometry for each gel is in Figures S5.1 and S5.2.

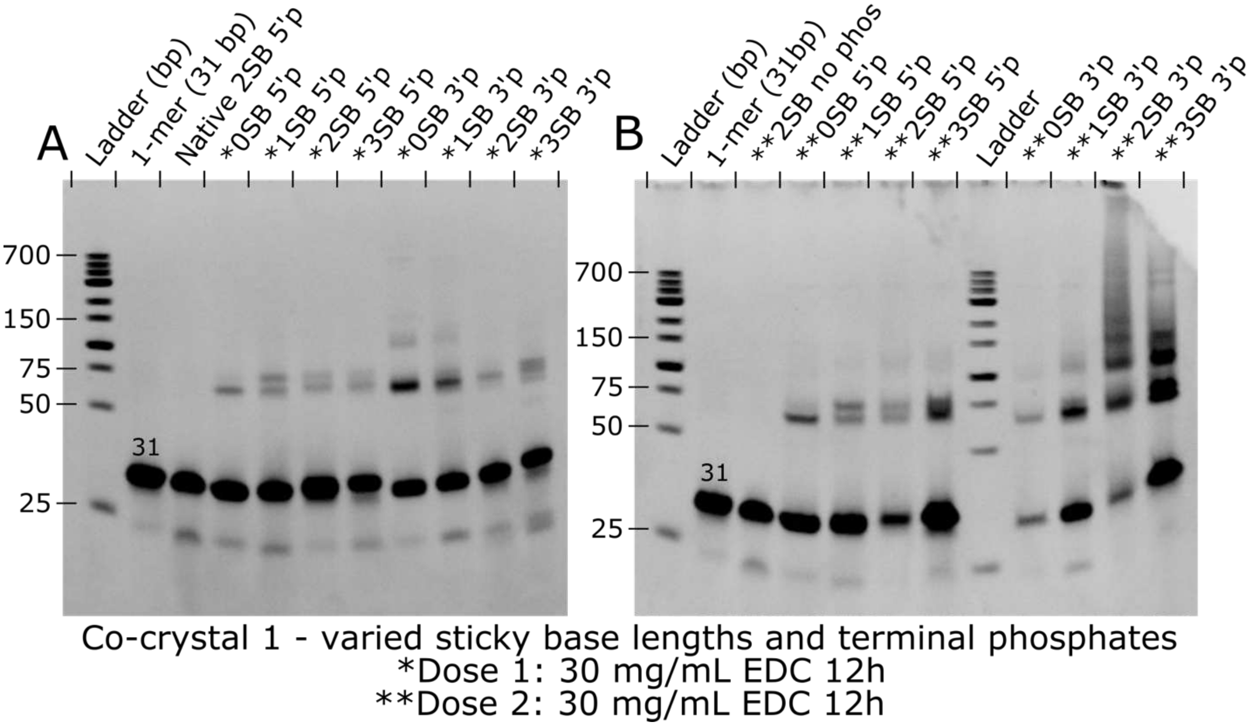


Table 5.1. Equations used for ligation product calculations. These are originally explained in our first co-crystal ligation[74] and Chapter 4.

Equation 1	$P_{SSB} = 1/\sum_i i \cdot x_i$
Equation 2	$P_{LIG} = 1 - P_{SSB}$
Equation 3	$P_{DSB} = (P_{SSB})^2$
Equation 4	$P_{DLIG} = (P_{LIG})^2$

Table 5.2. Distribution of DNA block sizes as a function of crystal sticky base overhang length and 3’ vs. 5’ terminal phosphates. The data shown corresponds with the gel lanes in Fig. 5.2. The crosslinking protocol was either *1 dose or **2 doses of 30 mg/mL EDC for 12 hours. The values in this table are weighted, dividing by DNA length to account for the increased dye intensity with DNA length. Unweighted values are found in Table S5.2. The full table including estimated mole fractions for higher-order products is found in Table S5.3. P_{SSB}, P_{LIG}, P_{DSB}, and P_{DLIG} were calculated for each crosslinked crystal sample. Probabilities of double strand breaks or double strand ligations assume ligation events are independent. P_{SSB} uncertainties are the standard deviation in the calculated P_{SSB} after 500 trials in which simulated noise (standard deviation 0.03) was introduced into relative band intensities.

Co-crystals EDC Dose One (Figure 5.2A)								
Parent Co-crystal	**0SB 5′P**	**1SB 5′P**	**2SB 5′P**	**3SB 5′p**	**0SB 3′P**	**1SB 3′P**	**2SB 3′P**	**3SB 3′p**
DNA block size	[%]	[%]	[%]	[%]	[%]	[%]	[%]	[%]
1	92.9	91.8	94.6	93.4	77.4	85.1	94.1	93.4
2	7.1	8.2	5.4	6.6	19.8	13.6	5.7	6.5
3					2.1	0.6	0.2	0.1
4					0.3	0.1		
5					7×10^{-9}	0.1		
6					0.1	0.2		
7					0.2	0.2		
8					0.2	0.1		

Parent Co-crystal	**0SB 5′P**	**1SB 5′P**	**2SB 5′P**	**3SB 5′P**	**0SB 3′P**	**1SB 3′P**	**2SB 3′P**	**3SB 3′P**
${P_{SSB}}^{*}$	0.93 ±0.01	0.92 ±0.01	0.95 ±0.01	0.94 ±0.01	0.77 ±0.02	0.82 ±0.03	0.94 ±0.02	0.93 ±0.02
$P_{LIG} = 1 - P_{SSB}$	0.07 ±0.01	0.08 ±0.01	0.05 ±0.01	0.06 ±0.01	0.23 ±0.02	0.18 ±0.03	0.06 ±0.02	0.07 ±0.02
$P_{DSB} = (P_{SSB})^2$	0.87 ±0.02	0.85 ±0.02	0.90 ±0.02	0.88 ±0.02	0.59 ±0.04	0.68 ±0.04	0.88 ±0.03	0.87 ±0.03
$P_{DLIG} = (P_{LIG})^2$	4 ± 2 $\cdot10^{-3}$	6 ± 2 $\cdot10^{-3}$	3 ± 1 $\cdot10^{-3}$	4 ± 2 $\cdot10^{-3}$	0.05 ±0.01	0.03 ±0.01	4 ± 2 $\cdot10^{-3}$	5 ± 2 $\cdot10^{-3}$

Table 5.2 continued.

Co-crystals EDC Dose Two (Figure 5.2B)

Parent Co-crystal	0SB 5′P	1SB 5′P	2SB 5′P	3SB 5′p	0SB 3′P	1SB 3′P	2SB 3′P	3SB 3′p
DNA block size	[%]	[%]	[%]	[%]	[%]	[%]	[%]	[%]
1	90.9	92.6	83.1	92.2	78.0	82.3	42.4	61.3
2	7.1	6.5	14.4	7.0	18.1	13.9	23.6	23.0
3	0.8	0.3	1.1	0.3	2.6	2.8	16.2	8.0
4	0.6	0.2	$1\text{x}10^{-8}$	0.2	0.5	0.4	5.9	3.8
5	0.1	0.2	0.2	0.03	$2\text{x}10^{-8}$	0.1	4.3	1.44
6	0.2	0.2	0.5	0.1	0.2	0.1	2.6	0.8639
7	0.1	0.1	0.3	0.03	0.1	0.2	1.5	0.6
8 and above	0.1		0.3	0.12	0.44	0.19	3.5	1.1

Parent Co-crystal	0SB 5′P	1SB 5′P	2SB 5′P	3SB 5′P	0SB 3′P	1SB 3′P	2SB 3′P	3SB 3′P
$P_{SSB}{}^{*}$	0.86 ±0.03	0.89 ±0.03	0.79 ±0.03	0.84 ±0.04	0.73 ±0.03	0.77 ±0.03	0.39 ±0.01	0.57 ±0.02
$P_{LIG} = 1 - P_{SSB}$	0.14 ±0.03	0.11 ±0.03	0.21 ±0.03	0.16 ±0.04	0.27 ±0.03	0.23 ±0.03	0.61 ±0.01	0.43 ±0.02
$P_{DSB} = (P_{SSB})^2$	0.74 ±0.05	0.78 ±0.05	0.62 ±0.05	0.70 ±0.06	0.53 ±0.04	0.59 ±0.04	0.15 ±0.01	0.33 ±0.02
$P_{DLIG} = (P_{LIG})^2$	0.02 ±0.01	0.01 ±0.01	0.04 ±0.01	0.03 ±0.01	0.08 ±0.02	0.05 ±0.01	0.37 ±0.02	0.18 ±0.02

To further assess how the EDC ligation effects the nanoscale stability of the co-crystals, we obtained the structures of crystals before and after EDC ligation (Table 5.3). On average, co-crystals after 2 doses of ligation lost ~0.5Å resolution. Also, the 3SB crystals were not diffracting; these crystals were small for XRD (~30 μm span) and the native crystals were hit-or-miss in diffraction quality. Although no protein-protein crosslinks stood out in the electron density, further studies could assess the EDC crosslinking in the protein of the co-crystals.

Table 5.3. Diffraction resolution of each CC1^{+10bp} variant before and after EDC ligation with two doses of EDC (30mg/mL EDC for 12h). Diffraction resolution estimates for I/Sigma(I) > 1.5 cutoff.

Sticky Base Overhang Length	Terminal Phosphorylation	Diffraction Resolution No Crosslink	Diffraction Resolution 2 Doses EDC (30 mg/mL EDC 12h)
0 SB	5'phos	2.79 Å	3.35 Å
0 SB	3'phos	2.46 Å	3.11 Å
1 SB	5'phos	2.54 Å	2.91 Å
1 SB	3'phos	3.11 Å	3.15 Å
2 SB	No phos	2.84 Å	3.11 Å
2 SB	5'phos	2.38 Å	3.24 Å
2 SB	3'phos	2.54 Å	3.15 Å
3 SB	5'phos	3.07 Å	N/A
3 SB	3'phos	3.14 Å	N/A

5.3.2. Chemical Ligation in DNA Crystals

It is not clear the extent to which protein-protein EDC conjugation is stabilizing CC1. Therefore, we sought to test the chemical ligation in the absence of protein. 3D DNA crystals self-assemble from three-armed tensegrity triangles *via* sticky end cohesion.[24,26,71] As the DNA crystals only have DNA-DNA contacts, the only reaction with EDC should be DNA-DNA junction ligation. Each DNA tile has seven single strands: one central strand, three edge strands and three crossover strands (Figure 5.3). Formerly, variations in the sticky end cohesion and overhang sequences were studied to improve the X-ray diffraction quality.[26] Here, we grew crystals from this optimization with either 1 or 2 nt overhangs with sequences in Table S5.4 and Figure S5.3. At the DNA-DNA junctions, the strands stack such that upon ligation, crossover strands (S1) ligate to a flanking crossover strand (S1), resulting in ligation products S1-S1 (28mer), S1-S1-S1 (42mer), *etc.* The same ligation pattern occurs for the edge strands (S2), where a 21mer single edge strand becomes a 42mer double edge strand.

The central strand (S3) simply ligates with itself to form a ring. To assess the role of sticky base overhang length and terminal phosphorylation status in DNA crystals, DNA crystals were grown with varied terminal phosphates (5' or 3'). The first case was with terminal phosphates only on the crossover strands (S1(5'P or 3'P)-S2-S3). Then we explored the ligation yields when all three strands were phosphorylated (S1(5'P or 3'P)-S2(5'P or 3'P)-S3(5'P or 3'P)).

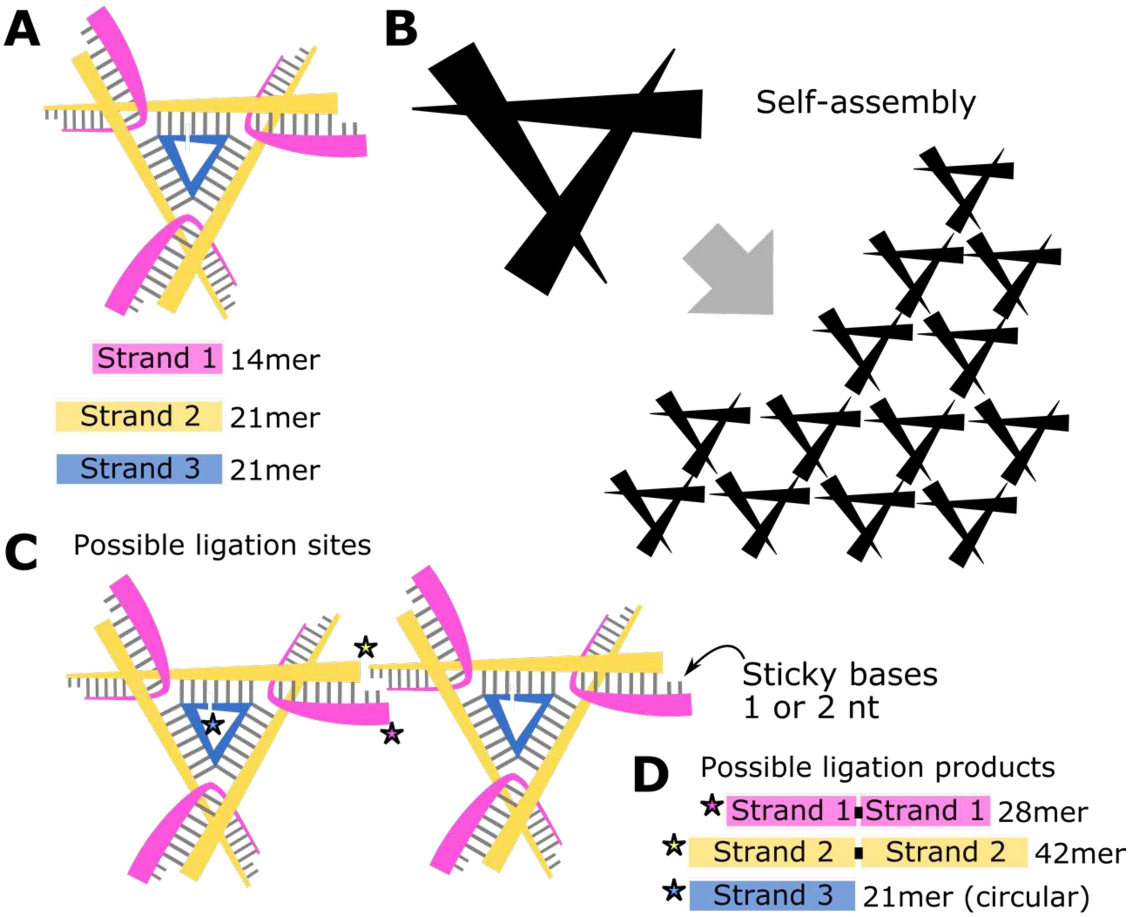


Figure 5.3. (A) Tensegrity triangle crystals are grown by mixing a stoichiometric ratio of three unique strands: Strand 1 (magenta), Strand 2 (yellow) and Strand 3 (blue). (B) Upon annealing, the DNA tiles self-assemble with sticky end cohesion at the DNA-DNA junctions. (C) At the sticky overhang region, the possible ligations are labeled with blue stars. We varied the sticky base overhangs (1 or 2 nt) to study the effects on ligation yields. (D) The possible ligation products are a polymerization of strand 1 to neighboring strand 1, represented with a single ligation (28mer product). Also, strand 2 can ligate to a neighboring strand 2, represented with a single ligation (42mer product). Strand 3 can ligate to itself, forming a circular product.

As with the co-crystals, the first step for DNA crystal ligation was a gentle buffer exchange. The DNA tile crystallization conditions (10-60 mM tris base pH 8.5, 5-30 mM acetic acid, 0.5-3 mM EDTA, and 125-750 mM magnesium acetate) were assessed for EDC ligation favorability. The same concepts applied: remove reactive species such as amines and carboxylic acids that hamper EDC ligation and make the solution conditions conductive to EDC ligation. In the resulting wash solution, we replaced tris base pH 8.5 with stoichiometric equivalent MES hydrate pH 6.0 for optimal phosphate reactivity. We also removed EDTA and acetic acid. Finally, we replaced magnesium acetate with a 1:1 stoichiometric equivalent magnesium chloride. In the end, the DNA crystal ligation wash buffer only contained 10-60 mM MES pH 6.0 and 125-750 mM magnesium chloride with concentrations matching the original growth condition concentrations of tris base and magnesium acetate, respectively. The DNA crystals maintained their visible macroscopic structure and were subsequently crosslinked in a wash solution supplemented with 30 mg/mL EDC.

Before comparing ligation yields in crystal variants, we determined if magnesium hindered DNA crystal ligation yields. Buffer exchange to fully remove Mg(II) could not be done without destroying the as-grown crystals. Therefore, the first dose of EDC ligation was performed with magnesium present and a subsequent dose was completed after a stoichiometric 1:1 replacement of Mg (II) with Na (I). When we replaced Mg (II) with Na (I), the ligation yields increased significantly for crystals containing 5’ phosphates (Figures S5.4 and S5.5). However, after four doses of EDC, there was no quantitative difference in ligation yields (Table 5.4) for repeated ligation in magnesium solution versus the protocol that involves initial ligation in magnesium and repeated ligation in sodium. Therefore, we cannot conclude that magnesium hinders EDC ligation in the DNA crystals.

With the optimized EDC reaction solution conditions in hand, the roles of sticky base overhang and terminal phosphorylation position were assessed. Crystals containing terminal phosphates only on the crossover strands (S1(5'P or 3'P)-S2-S3) and 1 or 2 nt overhangs were compared (Figure 5A). Again, the optimized EDC conditions were used: with one dose corresponding to 30 mg/mL EDC for 12h. Crystals were subjected to four doses of EDC, with the first dose in magnesium and subsequent doses with sodium. In Figure 5A, we compare the ligation products after four EDC doses. The Strand 1 ligation product distribution was readily interpretable; due to the lack of Strand 2 and 3 ligation these strands remained constant at 21 bases where they did not interfere with densitometry on the remaining bands. Like co-crystals, DNA crystal junctions with 3' terminal phosphates ligated more readily than 5' terminal phosphates, for both sticky base variants (1 nt and 2 nt). In the case of the sticky overhang length, the two sticky base overhangs promoted EDC ligation in the DNA crystals. The increased Strand 1 ligation yields were more prominent between 1nt and 2nt overhangs with 5' phosphates, where the 1nt 5'phos had 40% ligation versus the 2nt 5'phos with 65% ligation. For further comparison of all four doses of ligation, see Supplemental Figures S5.4-S5.5. As discussed in the Supporting Information for our previous report,[74] if ligation is random then the remaining 1-mer mole fraction after ligation provides another estimate (P_{SSB}') for the total strand break probability. Here, the correlation was reasonably strong ($R^2 = 0.8663$) between P_{SSB}' and the P_{SSB} calculated from the mole fractions via Equation 1 (Figure S5.6). The random ligation model is consistent with the idea that catalysis is slow relative to the intra-crystal transport of EDC, so that ligation is equally likely to occur throughout the body of the crystal.

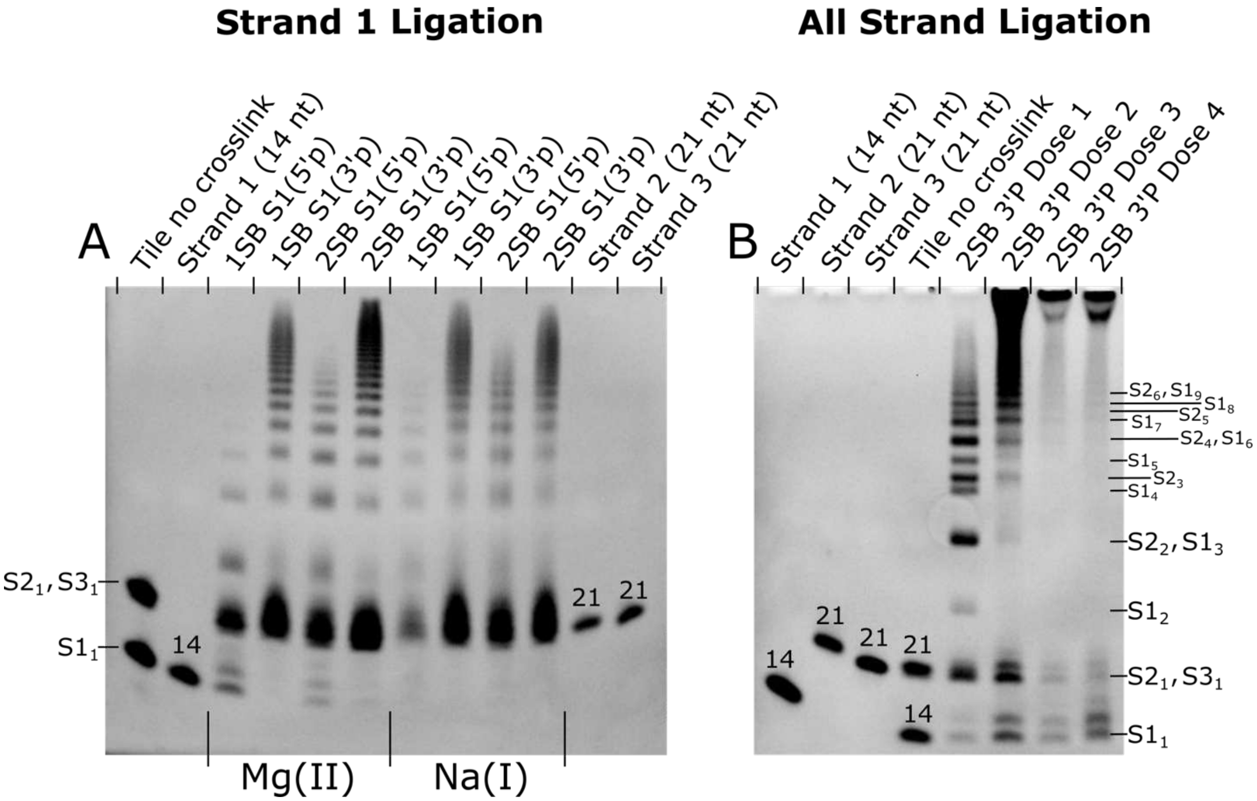


Figure 5.4. TBE-urea gels of EDC ligation for DNA crystals. Crystals with modifications only on strand 1 are shown (A) with varied sticky base overhangs (1SB and 2SB) and varied terminal phosphates (5'phos or 3'phos) after four doses of EDC. Furthermore, the same crystal classes are compared side by side in buffers with Mg(II) or with later rounds of ligation occurring after a buffer exchange into Na(I). Up to four EDC doses were applied to (B) crystals with 2 sticky bases and 3' phosphates on all strands revealing an apparent high ligation yield. After three EDC doses, lower size products disappear except for bands attributed to monomer strands. Assigned band sizes are given in bases. Gel densitometry for gel A is in Figure S5.7 and Table 5.4. Gel densitometry for gel B doses one and two are in Figure S5.8 and Table 5.5.

Table 5.4. Distribution of DNA block sizes as a function of crystal sticky base overhang length and 3’ vs. 5’ terminal phosphates. The data shown corresponds with the gel lanes in Fig. 5.4A. The crosslinking protocol was four doses of 30 mg/mL EDC for 12 hours and crystal samples are compared after four doses. The EDC buffer component indicates whether the crystal was crosslinked in the presence of magnesium or transferred to a sodium buffer (replacing magnesium). The values in this table are weighted so that the DNA length and dye intensity contributes to the final value. Unweighted values are found in Table S5.5. The full table including estimated mole fractions for higher-order products is found in Table S5.6. P_{SSB}, P_{LIG} were calculated for each crosslinked crystal sample. Double strand ligation is not possible here because only Strand 1 was phosphorylated.

DNA Crystal Strand 1 Ligation Products (Figure 5.4A)

Parent DNA	1SB 5′P	1SB 3′P	2SB 5′P	2SB 3′P	1SB 5′P	1SB 3′P	2SB 5′P	2SB 3′P
EDC buffer component	Mg(II)	Mg(II)	Mg(II)	Mg(II)	Na(I)	Na(I)	Na(I)	Na(I)
DNA block size	[%]	[%]	[%]	[%]	[%]	[%]	[%]	[%]
1	51.2	8.4	12.4	8.9	58.6	3.9	6.0	3.5
2	35.0	12.3	41.9	8.1	17.4	8.5	13.9	1.2
3	10.0	10.2	17.6	7.1	11.3	5.8	15.2	8.6
4	3.2	8.7	10.2	7.1	5.0	5.8	13.6	8.0
5	0.7	9.0	6.3	7.2	2.7	8.1	13.0	9.7
6		7.9	4.3	7.2	2.0	8.9	11.9	10.5
7		9.1	3.0	5.3	1.4	8.7	10.5	10.1
8 and above		34.3	4.3	49.3	1.7	50.2	16.0	48.4

Parent DNA	1SB 5′P	1SB 3′P	2SB 5′P	2SB 3′P	1SB 5′P	1SB 3′P	2SB 5′P	2SB 3′P
EDC buffer component	Mg(II)	Mg(II)	Mg(II)	Mg(II)	Na(I)	Na(I)	Na(I)	Na(I)
$P_{SSB}{}^{*}$	0.598 ±0.012	0.154 ±0.004	0.325 ±0.009	0.140 ±0.004	0.511 ±0.014	0.136 ±0.004	0.205 ±0.006	0.138 ±0.004
$P_{LIG} = 1 - P_{SSB}$	0.402 ±0.012	0.846 ±0.004	0.675 ±0.009	0.860 ±0.004	0.489 ±0.014	0.864 ±0.004	0.795 ±0.06	0.862 ±0.004

*Calculated from experimental mole fractions per Equation 1 (Table 5.1). P_{LIG} is derived from P_{SSB} as shown.

Following the comparison of strand one (S1) phosphorylation, we sought to achieve a fully ligated crystal with each junction phosphorylated. Crystals containing terminal phosphates on each strand (S1, S2 and S3) were grown and we compared gel densitometry results for four EDC doses. In Figure 5B, crystals with 2 sticky base overhangs and 3’ terminal phosphates had efficient ligation in the crystal. A significant quantity of the DNA loaded into these gels appeared to remain in the loading wells. For crystals ligated with 3 or 4 doses there was additionally a visible band for a long product. Future work may accurately size this long product band (i.e. with a gel that is optimized for longer DNA strands) and identify the constituent strands (e.g. using nanopore sequencing). We hypothesize that the apparent strong ligation yield at higher doses may reflect entangling of ligated DNA threads which prevent full denaturation and separation of the crystal into constituent ssDNA strands. The diminishing relative intensity of the $S3_1$ band with repeated ligation doses (relative to $S1_1$) may reflect circularized S3 becoming entangled with longer strands. Entangling complicates attempts to quantify ligation performance. Additionally, as ligation approaches completion, remaining smaller ligation product bands may increasingly reflect finite crystal size effects rather than incomplete ligation chemistry yield. Even in a hypothetical fully ligated crystal some strands would be short due to intersections with the crystal surfaces. We attribute the notable remaining population of small species as originating from these finite crystal effects, or perhaps from incomplete phosphorylation. While we did not attempt to further quantify the ligation yield for crystals subjected to 3 or 4 doses, we were able to estimate the ligation product mole fractions for one or two EDC doses (Table S5.7). Future work to accurately assess the yield for the in-crystal ligation yield of circular S3 will require solution synthesis of control circular S3 to establish how the circular strand runs on these gels. We also performed the same experiment with 2 sticky base overhangs and 5’ terminal phosphates. As with the Strand 1 ligations, after four EDC doses, the 3’ phosphates resulted in more ligation than the 5’ phosphates (Figure S5.9).

Further studies are needed to determine the effects of EDC ligation on X-ray diffraction quality for pure DNA crystals. As previously noted by Ohayon et al., tensegrity triangle crystals have marginal X-ray diffraction resolution.[26] In limited testing of the non-crosslinked crystals, we have not obtained diffraction suitable for structure determination. For future investigation of EDC chemical ligation of DNA crystals, it may be advantageous to select an alternative DNA crystal with more favorable initial diffraction.

5.3.3. Post-Ligation Stability in Harsh Conditions

For biomedical applications, engineered biomolecular crystals should be robust in solution conditions other than their growth solutions. After EDC ligation, the co-crystals and DNA crystals were subjected to both an ion-free environment (D.I. water) and to blood serum (bovine calf). Crosslinked DNA crystals with 2 sticky base overhangs and 3’ phosphates on each strand with two doses of EDC ligation (sister crystals of the ligation in Figure 5.4B Lane 6) showed amazing stability in these harsh conditions (in contrast to the immediate destruction of non-ligated crystals) (Figure 5.5). The stability in these crosslinked crystals can only be attributed to the addition of ligation (compared to the co-crystals which may also benefit from protein-protein crosslinks).

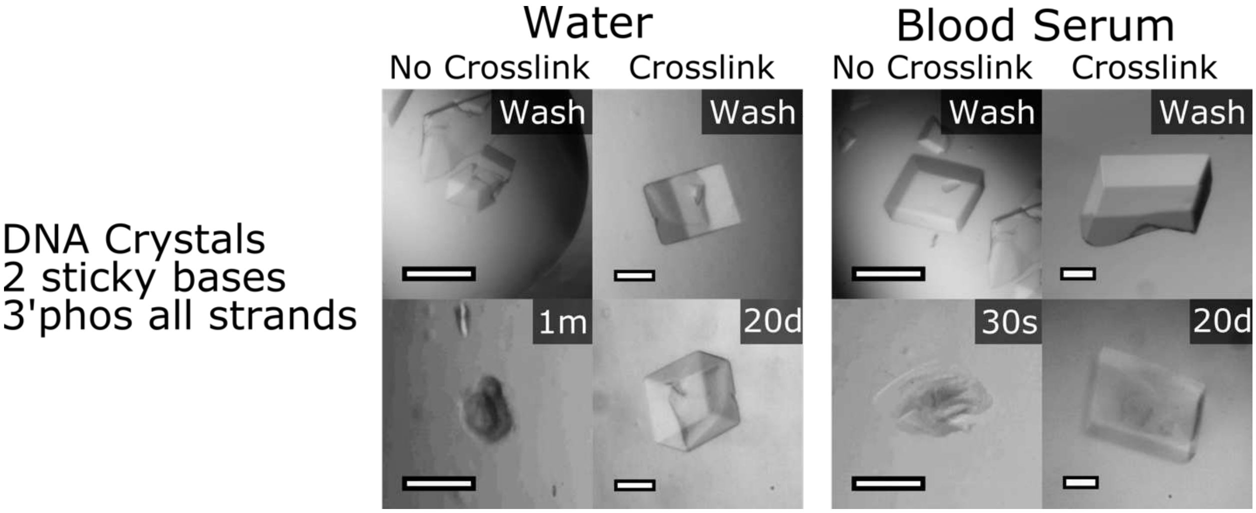


Figure 5.5. DNA Crystals in water and blood serum. Comparison of native and crosslinked crystal stability in harsh conditions. The DNA crystals had 2 sticky base overhangs and 3’ terminal phosphates on each strand (2SB S1(3’P)-S2(3’P)-S3(3’P)). Crystals were crosslinked with 2 doses of 30 mg/mL EDC for 12 hours. After quenching the EDC reaction in tris base, crystals were transferred to the ligation wash solution (without EDC). All scale bars are 100 µm.

DNA crystals with a terminal phosphate on only strand 1 (2SB S1(3’P)-S2-S3) were also subjected to water and blood serum (Figure S5.10). When quenching the EDC reaction in 50 mM tris base pH 8.3, the crosslinked DNA crystals (2SB S1(3’P)-S2-S3) visibly cracked. After adding these crystals to water, their macrostructure was expanded and remained as such for 20 days. In blood serum, the crosslinked crystal (2SB S1(3’P)-S2-S3) maintained the parallelepiped macrostructure for 20 days. Tunable degradation may be feasible by tuning the ligation yield, which would enable certain applications (*e.g.,* extended release of therapeutic cargo molecules).

Co-crystals with varied sticky base overhangs and terminal phosphorylation status were subjected to four harsh conditions: ion-free environment (water), blood serum, stomach acid mimic (pH 2.0) and lysosomal fluid mimic (pH 4.5). Non-crosslinked co-crystals degraded rapidly in all four conditions (Figure S5.11). After the co-crystals were crosslinked with two doses of EDC (30 mg/mL EDC for 12 h), quenched in 50 mM tris base pH 8.0, and washed in a growth solution mimic, the crystals were transferred to the harsh conditions. In Figure 5.6, we highlight the incredible stability of crosslinked co-crystals with 2 sticky base overhangs and 3' terminal phosphates. In all four harsh conditions, the crystals maintain their macrostructure after 21 days (albeit with arguable expanded structure in pH 2.0) (Figures S5.12-S5.15). Directly assessing crystal size and shape is challenging due to the crystal habit which presents a different cross-section shape depending on the viewing angle. In a more detailed analysis, (Figures S5.11-S5.14) the only crystal which appeared to (arguably) change size was the crystal incubated in pH 2.0 where unfolding of the protein might be responsible for the size change. Even after 5 months in these challenging conditions, the crystal in water and pH 4.5 appeared to maintain the same macrostructure.

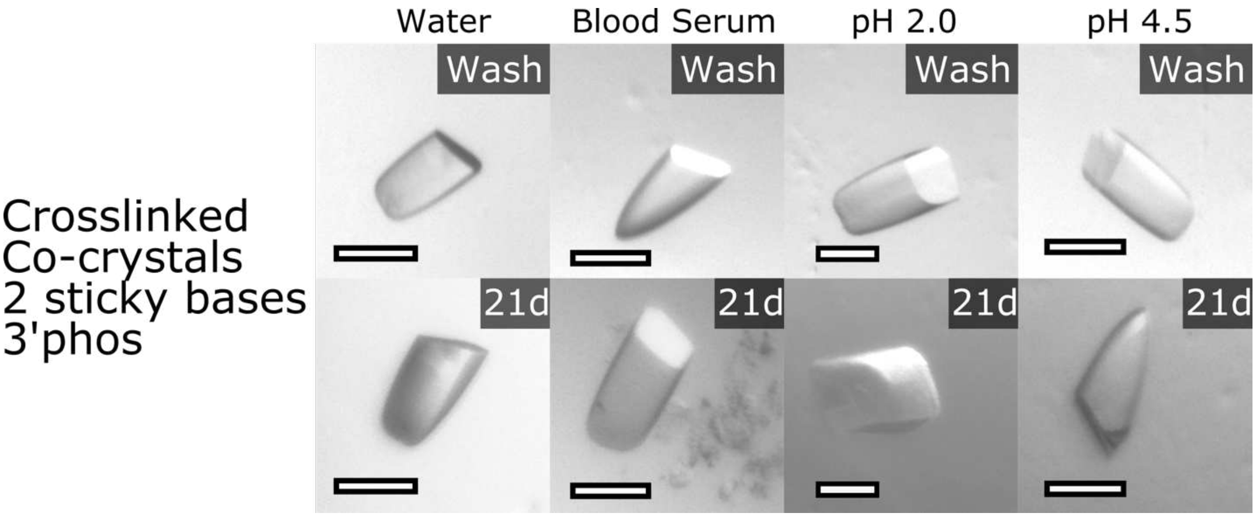


Figure 5.6. Crosslinked Co-crystals in four different harsh conditions. The co-crystals shown have 2 sticky base overhangs and 3'phosphates. Crystals were crosslinked with 2 doses of 30 mg/mL EDC for 12 hours. After quenching the EDC reaction in tris base, crystals were transferred to ligation wash solution (without EDC). All scale bars are 100 μm.

It is important to note that co-crystals without terminal phosphates were also stabilized by the two doses of EDC ligation. Therefore, we cannot attribute the full stability of the co-crystals to DNA ligation. While therefore attribute some co-crystal stability to the protein components, but it is not obvious from the crystal structure which intermolecular protein-protein crosslinks might be responsible. Regardless, the stability provided by EDC crosslinking in co-crystals may make them useful for biomedical applications.

5.4. Conclusions

In conclusion, chemical EDC ligation was shown to be effective in two crystal classes that contain DNA blocks stacking end-to-end. The roles of sticky base overhangs and terminal phosphate placement were assessed. In DNA crystals, two-base sticky overhang junctions with 3'terminal phosphates outperformed the alternatives (1 sticky base or 5' phosphates). With 3' terminal phosphates at each DNA junction in the DNA crystal, the apparent ligation yield was very high after 3 EDC doses (though bands attributed to monomer strands were still visible, potentially due to finite crystal effects). (Figure 5.4). In co-crystals, ligation yield trends as a function of sticky overhang length were not readily apparent in gel electrophoresis analysis after one EDC dose. However, after 2 EDC doses a marked increase in ligation yield was observed for crystals with 2 sticky base overhangs and 3' terminal phosphates. As in our first study of EDC ligation in co-crystals,[74] 3' terminal phosphates exceeded ligation yields of 5' terminal phosphates.

In addition to optimizing the crystal building units for ligation, we found a gentle buffer exchange was necessary for both crystal types to achieve chemical ligation. *Li et* *al.* ligated porous DNA crystals with enzymatic ligation but they specifically needed to avoid salt concentrations higher than 100 mM Mg(II) because it would interfere with T4 DNA ligase. Although Mg(II) also decreased EDC ligation yields for crystals with 5' phosphates, we were able to remove Mg(II) for successive crosslinking rounds in pursuit of higher ligation yields. Also, crystals may be stabilized with EDC ligation *post-assembly* regardless of crystal porosity, whereas T4 ligase function requires large solvent channels and a lack of steric hindrance around nick sites.

For biomedical applications, crystals were shown to be robust in water and blood serum (additionally, co-crystals in pH 2.0 and pH 4.5). Further studies may assess the cytocompatibility of crosslinked DNA crystals and co-crystals *in vivo* for drug delivery. In theory, ligation yields should be tunable for the application needed, including controlled crystal degradation.

5.5. Methods and Materials

5.5.1. Protein Cloning, Expression, and Purification

RepE54 initiator protein (CC1 protein) was expressed and purified by the Colorado State University Histone Source as described previously.[74] In short, the protein from PDB code 1rep was overexpressed with a N-terminal 6-Histag in *E. coli* CodonPlus RIPL competent cells. Sonicated cell lysate was purified with Ni Excel Sepharose (Cytiva) and HiLoad Superdex 200 PG column (Cytiva). The resulting CC1 protein was concentrated to 15 mg/mL in storage buffer (100 mM sodium citrate pH 6.2, 100 mM KCl, 10 mM $MgCl_2$, and 10% glycerol) and stored at –80 °C after flash freezing with liquid nitrogen.

5.5.2. Oligomers and Co-crystal Duplex Annealing

Oligomers were synthesized and HPLC purified by Integrated. DNA tensegrity triangle strands (Table S5.4) were resuspended in molecular biology grade water (VWR) and combined 3:3:1 Strand 1 to Strand 2 to Strand 3 to form the tile at 60 μM. Co-crystal DNA duplexes are listed in Table S5.1. All 31-bp duplex sequences contained a conserved 19-bp iteron sequence for RepE54 protein-DNA binding but had varied sticky base overhang lengths and terminal phosphorylation status (no phosphate, 5' or 3' phosphate). The oligomers were resuspended in 50 mM tris HCl, 100 mM KCl pH 7.0. Duplexes were annealed by combining cognate oligomers 1:1, heating to 94 °C for 2 min then slowly cooling to room temperature over approximately 60 min. The final duplex concentrations were 4 mM. DNA stocks were quantified using a Quibit4 (Qubit^{TM} 1 × dsDNA HS Assay Kit).

5.5.3. DNA-Protein Complex Co-crystallization

Co-crystals were grown *via* sitting drop vapor diffusion in 24-well Intelli plates (Hampton) with 0.5 μL protein-DNA complex and 0.5 μL crystallization liquor. Protein and DNA were combined (1:1.2) and the complex was incubated on ice for 30 min prior to crystal plate setup. CC1 crystallization conditions were 300-500 mM $MgCl_2$, 25-35% PEG 400, and 100 mM tris HCl pH 8.0. Crystals grew to a span of 50-300 μm in a range of 24 h to 7 days.

5.5.4. DNA Tile Crystallization

The DNA crystals were grown using sitting drop vapor diffusion in Hampton 24-well Cryschem M plates. Crystals were grown by combining in the well 14 μM DNA tile and 10-60 mM tris base pH 8.5, 5-30 mM acetic acid, 0.5-3 mM EDTA, and 125-750 mM magnesium acetate (exact crystallization conditions for each DNA tile variant are listed in Table S5.8). The reservoir contained 1.75 M ammonium sulfate. Self-assembly was induced by incubating crystal plates at 60ºC for 10 mins, and then allowing crystals to cool to 25 ºC over 1h in a Heratherm Oven (Thermo Scientific). Crystals grew to a size of 25-200 μm^3 in a range of 24 h to 7 days.

5.5.5. EDC Crosslinking

Co-crystals and DNA crystals were washed prior to chemical EDC ligation to remove reactive buffer components (*i.e.,* primary amines, carboxylic acids), remove supernatant protein and/or DNA, and to achieve optimal pH 6.0 for phosphate reactivity with EDC. The CC1 wash condition consisted of 300-500 mM NaCl (substituting for $MgCl_2$), 25-35% PEG 400, and 100 mM MES pH 6.0 (substituting for tris HCl pH 8.0). The co-crystal wash completely removed Mg(II), whereas in the DNA crystal wash, Mg(II) was not fully removed until after the first EDC crosslink dose. For EDC dose one, DNA crystal wash condition consisted of 10-60 mM MES pH 6.0 (substituting for tris base pH 8.5) and 125-750 mM $MgCl_2$ (substituting for magnesium acetate). In subsequent EDC doses, the magnesium was replaced fully with wash consisting of 10-60 mM MES pH 6.0 and 125-750 mM NaCl.

Both co-crystals and DNA crystals were washed in 9-well glass plates (Hampton) to remove additional protein and/or DNA monomers and unwanted buffer components. 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) (Advanced Chemtech CAS#255952-53-8) was resuspended in the wash solution to 30 mg/mL and used immediately. The crystals were crosslinked in a 200 μL EDC solution volume for 12-

hour doses. The crosslinking reaction was quenched by moving crystals to 1 × Tris-Borate-EDTA (TBE) buffer pH 8.3 containing 3.5 M urea.

5.5.6. DNA Gel Electrophoresis and Gel Densitometry

Crosslinked co-crystals were dissolved in 1 × TBE supplemented with 3.5 M urea and 20 μg proteinase K and incubated at 50 ºC for 12 hours. The DNA crystals were dissolved in 3.5M Urea in 1 × TBE and heated to 90 ºC for 10 minutes. The crystals were analyzed with denaturing gels (10% Novex^{TM} TBE-Urea Polyacrylamide Gel Electrophoresis) formulated with 7M urea and run with 1× TBE running buffer. The co-crystal DNA ladder was GeneRuler Low Range DNA Ladder (Thermo Scientific). Gels were incubated with 3 x GelRed^{TM} Nucleic Acid Gel Stain and imaged with a UVP Bioimaging System on the Ethidium Bromide setting. For gel densitometry, we used the technique established in our previous ligation study (Python scripts for fitting densitometry peaks with Gaussians).[74] The raw band intensity fits are provided in Tables S5.2, S5.5, and S5.7 and the densitometry fits using Gaussians are provided in Figures S5.1-S5.2 and S5.7-S5.8.

5.5.7. Stability Assays

Crystals were crosslinked with 2 doses of 30 mg/mL EDC and the EDC reaction was quenched in 50 mM tris base pH 8.0 for 30 min. The crystals were equilibrated in crosslinking wash solution for 30 min prior to looping to stringent conditions. The stability test buffers used were as follows: molecular biology grade water (CORNING), very low pH 2.0 0.01 M HCl buffer (to mimic stomach acid), a moderately low pH 4.5 citrate buffer (46 mM sodium citrate, 54.1 mM citric acid to mimic lysosomal fluid pH), and blood serum (HyClone, bovine calf blood serum). Microscope images for each trial are in Figures S5.9-S5.14.

5.5.8. X-ray Diffraction Data Collection, Refinement, and Omit Maps

Single-crystal X-ray diffraction (XRD) data were collected for CC1^{+10bp} co-crystals containing blunt ends, 1-nt, 2-nt and 3-nt sticky base overhangs with 5’ or 3’ phosphates. In addition, XRD was obtained for a co-crystal with 2-nt sticky base overhangs and no terminal phosphates. Crystals were crosslinked with 2 doses of 30 mg/mL EDC for 12 h and flash frozen directly from EDC crosslink (no quench) for XRD. Crosslinked co-crystals with 3SB overhangs were not diffracting well enough for structure determination, possibly due to small crystal size (~25 μm span) but also could be the crosslinking disrupted the nanoscale order.

5.6. Funding

This material is based upon work supported by the National Science Foundation under Grant No. NSF DMR 2003748 and NSF DMR 1506219. The team also gratefully acknowledges support for undergraduate researchers from the Nelson Family Faculty Excellence Award.

CHAPTER 6. SUMMARY AND FUTURE DIRECTIONS

From design to experimental validation, the novel isoreticular co-crystals developed through this research bring us significantly closer to the longstanding goal of a molecular scaffold for X-ray diffraction studies.[1] Upon further development, expanded co-crystals may serve as a molecular pegboard to host DNA-binding molecules, revealing never-before seen DNA-binding molecules in X-ray diffraction. In medicine, future structures revealed with scaffold-assisted crystallography can advance our knowledge of human proteins for therapeutic targets. Beyond structural biology, the porous co-crystal could act as a general-purpose molecular sieve for sensing, catalysis, and inorganic templating. The prospects for our designed co-crystals are not limited to the following proposed directions.

In **Chapter 2**, the first engineered porous co-crystal was designed and experimentally validated. Further development to create a versatile vessel include confirming sequence and DNA scheme modularity and making expanded crystals robust with EDC crosslinking. The expanded co-crystals have many exciting possible applications. One avenue already in exploration is using the co-crystals as a host for scaffold-assisted structural biology. Preliminary work is being explored by Ethan Shields and Callie Slaughter in the Snow Lab. Briefly, towards the goal of scaffold-assisted X-ray diffraction, crystals were designed with binding sites for two different mini-protein guests: the Engrailed Homeodomain (PDB code: 3hdd)[86] and the C-Clamp (PDB code: 7dta)[111] (Figure 6.1). Both selected guests have existing structures, and they will serve as proof-of-principle targets for X-ray diffraction of DNA-binding protein guests. These guests were chosen strategically to highlight the sequence modularity of the scaffold crystal. For instance, the engrailed homeodomain binding sequence is 5’-TAATTA while the C-Clamp binding sequence is 5’-GCCGGG. Once the post-crystallization host diffraction is achieved, the scaffold may be used to host a variety of guest Protein, DNA, RNA, and small molecule DNA-binding guests. Furthermore, as I pursue Post-Doctoral training in the Blind Lab at Vanderbilt University Medical Center, I am enthusiastic to host guest nuclear receptor protein targets for direct medical impact in diabetes and cancer.

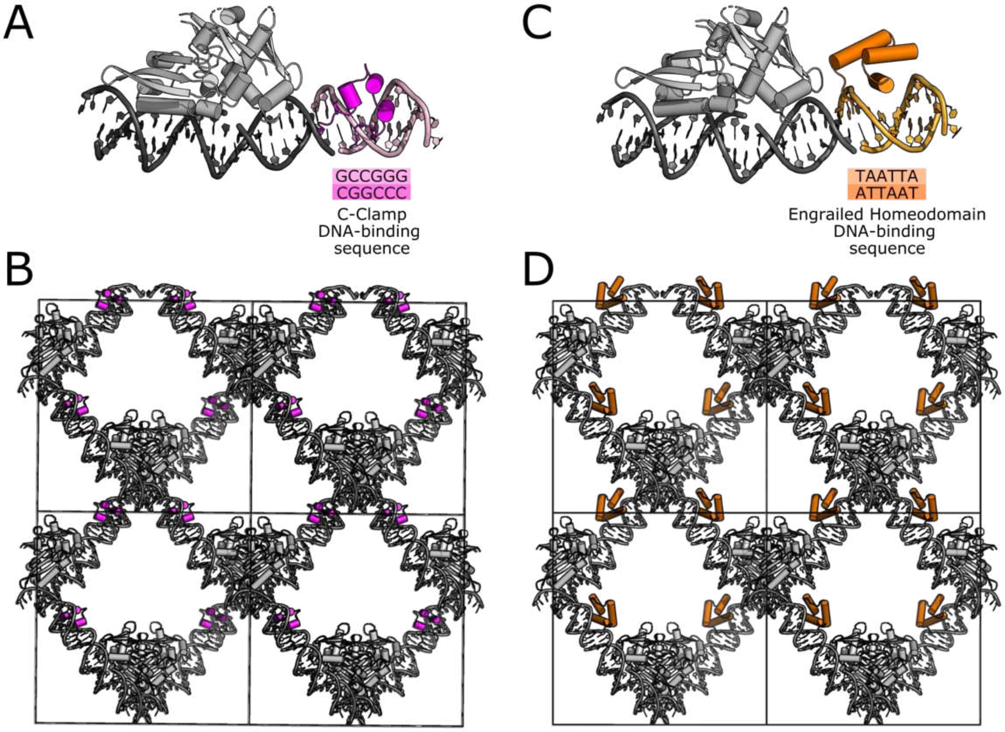


Figure 6.1. Models of the two DNA-binding guest proteins within CC1^{+10bp} expanded, porous crystal. The C-Clamp binding site (A) is a G-C rich sequence and the hypothetical model (B) of the porous lattice with the C-Clamp bound shows ample room for the guest and guest diffusion. The Engrailed Homeodomain site (C) is a T-A rich sequence and the hypothetical model (D) of Engrailed in the porous lattice also shows ample room for guest entrapment.

An additional follow up is needed for the crystallization screens as additional crystallization conditions may yield the porous lattice (or other interesting crystal forms).

CC1^{+10bp} variants were screened with the 96-well conditions (NatrixHT and IndexHT Hampton). There are many crystal hits remaining on these plates that should undergo X-ray diffraction analysis. One crystal hit that was grown in the presence of EnH-eGFP (co-crystallized with the guest target binding sequence) is visibly green. Other crystal forms other than the porous lattice presented may be elucidated from these crystal screen plates.

In **Chapter 3**, the interpenetrating, expanded co-crystal were shown to self-assemble in the presence of guest molecules for site-specific entrapment of protein, DNA and small molecules. The crystals were modular and grew consistently despite myriad DNA variations, including varied DNA expansion scheme, insert sequence, sticky overhang lengths, and terminal phosphorylation. While the interpenetrating lattice is unfavorable for structural biology applications, the lower porosity of the interpenetrating co-crystal could be helpful for drug delivery applications if entrapped cargo macromolecules are protected against extraneous nuclease attack in the body. To realize isoreticular co-crystal directions in biomedical applications, it will be critical to assess possible cytotoxic effects of introducing porous co-crystals *in vivo*. In the exploration of biological chemistry, the interpenetrating expanded crystal that templates the DNA-DNA junction can serve to study various chemistries, that then can be revealed at atomic detail in X-ray diffraction. For example, thymine dimerization via UV-crosslinking can be studied in our designed crystals.

In **Chapter 4**, chemical EDC ligation was shown to produce tremendous stability within two different classes of co-crystals. The bioconjugation technique we presented in the journal *Crystals* was the first example of DNA ligation templated by a 3D biomolecular crystal. Biomolecular nanotechnologists can use our optimized crosslinking techniques and guidelines to stabilize crystals and other DNA assemblies for applications of interest. Future classes of isoreticular co-crystals, and hypothetically any crystals with stacking DNA-DNA junctions, may be crosslinked with EDC and the guidelines presented may serve to advance the progress more rapidly. Also, the EDC ligation chemistry, although it did not enhance diffraction quality, may ultimately prove to be a tool for crystallographers with inherently flexible crystals. Further work is needed to study the exact protein-protein EDC crosslinked sites in co-crystals. Ultimately, it would also be advantageous to engineer the protein-protein interfaces to facilitate controlled site-specific polymerization of the protein components. Hypothetically, a future crosslinked expanded co-crystal might be covalently linked at each interface: protein-protein, protein-DNA and DNA-DNA. In the future, a combination of EDC and formaldehyde bioconjugation may achieve all three interfacial crosslinks.

In **Chapter 5**, EDC crosslinking chemistry was studied in DNA crystals and interpenetrating co-crystals, and the role of sticky end DNA junctions and terminal phosphates was assessed. We concluded that ligations are likely to benefit from using terminal 3’ phosphates. Also, in DNA crystals (and sometimes in the co-crystals), longer sticky base overhangs were advantageous for ligation. We showed chemical ligation can be applied to pure 3D DNA crystals and the stability achieved after our optimized crosslinking chemistry may help DNA crystal engineers towards the long-standing goal of using DNA crystals as molecular scaffolds. The crystal stability produced at the macroscale and nanoscale may help the host-crystal strategy realize its full potential as a *post*-crystallization host. Lastly, the crystal robustness produced by ligation at the molecular level provides solution independence, enabling a plethora of applications for these high-precision biomaterials.

BIBLIOGRAPHY

(1) Seeman, N. C. Nucleic Acid Junctions and Lattices. *J. Theor. Biol.* **1982**, *99* (2), 237–247.

(2) Abe, S.; Ueno, T. Design of Protein Crystals in the Development of Solid Biomaterials. *RSC Adv.* **2015**, *5* (27), 21366–21375.

(3) Abe, S.; Tabe, H.; Ijiri, H.; Yamashita, K.; Hirata, K.; Atsumi, K.; Shimoi, T.; Akai, M.; Mori, H.; Kitagawa, S.; Ueno, T. Crystal Engineering of Self-Assembled Porous Protein Materials in Living Cells. *ACS Nano* **2017**, *11* (3), 2410–2419.

(4) Hartje, L. F.; Snow, C. D. Protein Crystal Based Materials for Nanoscale Applications in Medicine and Biotechnology. *Wiley Interdiscip Rev Nanomed Nanobiotechnol* **2019**, *11* (4), e1547.

(5) Kowalski, A. E.; Johnson, L. B.; Dierl, H. K.; Park, S.; Huber, T. R.; Snow, C. D. Porous Protein Crystals as Scaffolds for Enzyme Immobilization. *Biomater. Sci.* **2019**, *7* (5), 1898–1904.

(6) Tabe, H.; Abe, S.; Hikage, T.; Kitagawa, S.; Ueno, T. Porous Protein Crystals as Catalytic Vessels for Organometallic Complexes. *Chemistry – An Asian Journal* **2014**, *9* (5), 1373–1378.

(7) Ueno, T. Porous Protein Crystals as Reaction Vessels. *Chemistry – A European Journal* **2013**, *19* (28), 9096–9102.

(8) Maita, N. Crystal Structure Determination of Ubiquitin by Fusion to a Protein That Forms a Highly Porous Crystal Lattice. *J. Am. Chem. Soc.* **2018**, *140* (42), 13546–13549.

(9) Bukowska, M. A.; Grütter, M. G. New Concepts and Aids to Facilitate Crystallization. *Curr. Opin. Struct. Biol.* **2013**, *23* (3), 409–416.

(10) Griffin, L.; Lawson, A. Antibody Fragments as Tools in Crystallography. *Clin Exp Immunol* **2011**, *165* (3), 285–291.

(11) Ni, T. W.; Tezcan, F. A. Structural Characterization of a Microperoxidase Inside a Metal-Directed Protein Cage. *Angewandte Chemie International Edition* **2010**, *49* (39), 7014–7018.

(12) Hoshino, M.; Khutia, A.; Xing, H.; Inokuma, Y.; Fujita, M. The Crystalline Sponge Method Updated. *IUCrJ* **2016**, *3* (2), 139–151.

(13) Huber, T. R.; Hartje, L. F.; McPherson, E. C.; Kowalski, A. E.; Snow, C. D. Programmed Assembly of Host–Guest Protein Crystals. *Small* **2017**, *13* (7), 1602703.

(14) Hashimoto, T.; Ye, Y.; Matsuno, A.; Ohnishi, Y.; Kitamura, A.; Kinjo, M.; Abe, S.; Ueno, T.; Yao, M.; Ogawa, T.; Matsui, T.; Tanaka, Y. Encapsulation of Biomacromolecules by Soaking and Co-Crystallization into Porous Protein Crystals of Hemocyanin. *Biochemical and Biophysical Research Communications* **2019**, *509* (2), 577–584.

(15) Hartje, L. F.; Bui, H. T.; Andales, D. A.; James, S. P.; Huber, T. R.; Snow, C. D. Characterizing the Cytocompatibility of Various Cross-Linking Chemistries for the Production of Biostable Large-Pore Protein Crystal Materials. *ACS Biomater. Sci. Eng.* **2018**.

(16) Inokuma, Y.; Yoshioka, S.; Ariyoshi, J.; Arai, T.; Hitora, Y.; Takada, K.; Matsunaga, S.; Rissanen, K.; Fujita, M. X-Ray Analysis on the Nanogram to Microgram Scale Using Porous Complexes. *Nature* **2013**, *495* (7442), 461–466.

(17) Deng, H.; Grunder, S.; Cordova, K. E.; Valente, C.; Furukawa, H.; Hmadeh, M.; Gándara, F.; Whalley, A. C.; Liu, Z.; Asahina, S.; Kazumori, H.; O'Keeffe, M.; Terasaki, O.; Stoddart, J. F.; Yaghi, O. M. Large-Pore Apertures in a Series of Metal-Organic Frameworks. *Science* **2012**, *336* (6084), 1018–1023.

(18) Lee, S.; Kapustin, E. A.; Yaghi, O. M. Coordinative Alignment of Molecules in Chiral Metal-Organic Frameworks. *Science* **2016**, *353* (6301), 808–811.

(19) Fujita, D.; Suzuki, K.; Sato, S.; Yagi-Utsumi, M.; Yamaguchi, Y.; Mizuno, N.; Kumasaka, T.; Takata, M.; Noda, M.; Uchiyama, S.; Kato, K.; Fujita, M. Protein Encapsulation within Synthetic Molecular Hosts. *Nat Commun* **2012**, *3*, 1093.

(20) Paukstelis, P. J.; Nowakowski, J.; Birktoft, J. J.; Seeman, N. C. Crystal Structure of a Continuous Three-Dimensional DNA Lattice. *Chem. Biol.* **2004**, *11* (8), 1119–1126.

(21) Geng, C.; Paukstelis, P. J. DNA Crystals as Vehicles for Biocatalysis. *J. Am. Chem. Soc.* **2014**, *136* (22), 7817–7820.

(22) Paukstelis, P. J. Three-Dimensional DNA Crystals as Molecular Sieves. *J. Am. Chem. Soc.* **2006**, *128* (21), 6794–6795.

(23) McNeil, R.; Paukstelis, P. J. Core-Shell and Layer-by-Layer Assembly of 3D DNA Crystals. *Adv. Mater. Weinheim* **2017**, *29* (28).

(24) Zheng, J.; Birktoft, J. J.; Chen, Y.; Wang, T.; Sha, R.; Constantinou, P. E.; Ginell, S. L.; Mao, C.; Seeman, N. C. From Molecular to Macroscopic via the Rational Design of a Self-Assembled 3D DNA Crystal. *Nature* **2009**, *461* (7260), 74–77.

(25) Liu, D.; Wang, M.; Deng, Z.; Walulu, R.; Mao, C. Tensegrity: Construction of Rigid DNA Triangles with Flexible Four-Arm DNA Junctions. *J. Am. Chem. Soc.* **2004**, *126* (8), 2324–2325.

(26) Ohayon, Y.; Hernandez, C.; Chandrasekaran, A. R.; Wang, X.; Abdallah, H.; Jong, M. A.; Mohsen, M.; Sha, R.; Birktoft, J. J.; Lukeman, P. S.; Chaikin, P. M.; Ginell, S. L.; Mao, C.; Seeman, N. C. Designing Higher Resolution Self-Assembled 3D DNA Crystals via Strand Terminus Modifications. *ACS Nano* **2019**, *13* (7), 7957–7965.

(27) Stahl, E.; Praetorius, F.; de Oliveira Mann, C. C.; Hopfner, K.-P.; Dietz, H. Impact of Heterogeneity and Lattice Bond Strength on DNA Triangle Crystal Growth. *ACS Nano* **2016**, *10* (10), 9156–9164.

(28) Wang, T.; Sha, R.; Birktoft, J.; Zheng, J.; Mao, C.; Seeman, N. C. A DNA Crystal Designed to Contain Two Molecules per Asymmetric Unit. *J. Am. Chem. Soc.* **2010**, *132* (44), 15471–15473.

(29) Simmons, C. R.; Zhang, F.; Birktoft, J. J.; Qi, X.; Han, D.; Liu, Y.; Sha, R.; Abdallah, H. O.; Hernandez, C.; Ohayon, Y. P.; Seeman, N. C.; Yan, H. Construction and Structure Determination of a Three-Dimensional DNA Crystal. *J. Am. Chem. Soc.* **2016**, *138* (31), 10047–10054.

(30) Zhang, F.; Simmons, C. R.; Gates, J.; Liu, Y.; Yan, H. Self-Assembly of a 3D DNA Crystal Structure with Rationally Designed Six-Fold Symmetry. *Angewandte Chemie International Edition* **2018**, *57* (38), 12504–12507.

(31) Juers, D. H.; Ruffin, J. MAP_CHANNELS: A Computation Tool to Aid in the Visualization and Characterization of Solvent Channels in Macromolecular Crystals. *J Appl Crystallogr* **2014**, *47* (Pt 6), 2105–2108.

(32) Brady, R. A.; Brooks, N. J.; Foderà, V.; Cicuta, P.; Di Michele, L. Amphiphilic-DNA Platform for the Design of Crystalline Frameworks with Programmable Structure and Functionality. *J. Am. Chem. Soc.* **2018**, *140* (45), 15384–15392.

(33) Zhang, C.; Tian, C.; Guo, F.; Liu, Z.; Jiang, W.; Mao, C. DNA-Directed Three-Dimensional Protein Organization. *Angewandte Chemie International Edition* **2012**, *51* (14), 3382–3385.

(34) Zhang, T.; Hartl, C.; Frank, K.; Heuer-Jungemann, A.; Fischer, S.; Nickels, P. C.; Nickel, B.; Liedl, T. 3D DNA Origami Crystals. *Advanced Materials* **2018**, *30* (28), 1800273.

(35) Jeliazkov, J. R.; Robinson, A. C.; E, B. G.-M.; Berger, J. M.; Gray, J. J. Toward Computational Design of Protein Crystals with Improved Resolution. *bioRxiv* **2019**, 657262.

(36) Leibly, D. J.; Arbing, M. A.; Pashkov, I.; DeVore, N.; Waldo, G. S.; Terwilliger, T. C.; Yeates, T. O. A Suite of Engineered GFP Molecules for Oligomeric Scaffolding. *Structure* **2015**, *23* (9), 1754–1768.

(37) King, N. P.; Lai, Y.-T. Practical Approaches to Designing Novel Protein Assemblies. *Curr Opin Struct Biol* **2013**, *23* (4), 632–638.

(38) Cannon, K. A.; Ochoa, J. M.; Yeates, T. O. High-Symmetry Protein Assemblies: Patterns and Emerging Applications. *Curr. Opin. Struct. Biol.* **2019**, *55*, 77–84.

(39) Mou, Y.; Yu, J.-Y.; Wannier, T. M.; Guo, C.-L.; Mayo, S. L. Computational Design of Co-Assembling Protein-DNA Nanowires. *Nature* **2015**, *525* (7568), 230–233.

(40) Subramanian, R. H.; Smith, S. J.; Alberstein, R. G.; Bailey, J. B.; Zhang, L.; Cardone, G.; Suominen, L.; Chami, M.; Stahlberg, H.; Baker, T. S.; Tezcan, F. A. Self-Assembly of a Designed Nucleoprotein Architecture through Multimodal Interactions. *ACS Cent. Sci.* **2018**, *4* (11), 1578–1586.

(41) Sontz, P. A.; Bailey, J. B.; Ahn, S.; Tezcan, F. A. A Metal Organic Framework with Spherical Protein Nodes: Rational Chemical Design of 3D Protein Crystals. *J. Am. Chem. Soc.* **2015**, *137* (36), 11598–11601.

(42) Bailey, J. B.; Zhang, L.; Chiong, J. A.; Ahn, S.; Tezcan, F. A. Synthetic Modularity of Protein-Metal-Organic Frameworks. *J. Am. Chem. Soc.* **2017**, *139* (24), 8160–8166.

(43) Liu, Y.; Huynh, D. T.; Yeates, T. O. A 3.8 Å Resolution Cryo-EM Structure of a Small Protein Bound to an Imaging Scaffold. *Nat Commun* **2019**, *10* (1), 1864.

(44) Haworth, N. L.; Gready, J. E.; George, R. A.; Wouters, M. A. Evaluating the Stability of Disulfide Bridges in Proteins: A Torsional Potential Energy Surface for Diethyl Disulfide. *Mol Simul* **2007**, *33* (6–8), 475–485.

(45) Huber, T. R.; McPherson, E. C.; Keating, C. E.; Snow, C. D. Installing Guest Molecules at Specific Sites within Scaffold Protein Crystals. *Bioconjugate Chem.* **2018**, *29* (1), 17–22.

(46) Koshiyama, T.; Kawaba, N.; Hikage, T.; Shirai, M.; Miura, Y.; Huang, C.-Y.; Tanaka, K.; Watanabe, Y.; Ueno, T. Modification of Porous Protein Crystals in Development of Biohybrid Materials. *Bioconjug. Chem.* **2010**, *21* (2), 264–269.

(47) Derewenda, Z. S. Application of Protein Engineering to Enhance Crystallizability and Improve Crystal Properties. *Acta Crystallographica Section D Biological Crystallography* **2010**, *66* (5), 604–615.

(48) 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* **2013**, *69* (Pt 12), 2513–2523.

(49) Salgado, E. N.; Radford, R. J.; Tezcan, F. A. Metal-Directed Protein Self-Assembly. *Acc. Chem. Res.* **2010**, *43* (5), 661–672.

(50) Abe, S.; Maity, B.; Ueno, T. Functionalization of Protein Crystals with Metal Ions, Complexes and Nanoparticles. *Current Opinion in Chemical Biology* **2018**, *43*, 68–76.

(51) Kowalski, A. E.; Huber, T. R.; Ni, T. W.; Hartje, L. F.; Appel, K. L.; Yost, J. W.; Ackerson, C. J.; Snow, C. D. Gold Nanoparticle Capture within Protein Crystal Scaffolds. *Nanoscale* **2016**, *8* (25), 12693–12696.

(52) Rupp, B. *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*; Garland Science, 2009.

(53) Hollis, T. Crystallization of Protein-DNA Complexes. *Methods Mol Biol* **2007**, *363*, 225–237.

(54) Aggarwal, A. K. Crystallization of DNA Binding Proteins with Oligodeoxynucleotides. *Methods* **1990**, *1* (1), 83–90.

(55) Saida, F. Statistical Analysis of 15 Dimensions in the Crystallization Space for Protein-DNA Complexes. *Protein Pept. Lett.* **2006**, *13* (9), 929–939.

(56) Ward, A. R.; Snow, C. D. Porous Crystals as Scaffolds for Structural Biology. *Current Opinion in Structural Biology* **2020**, *60*, 85–92.

(57) Nguyen, T. K.; Negishi, H.; Abe, S.; Ueno, T. Construction of Supramolecular Nanotubes from Protein Crystals. *Chem. Sci.* **2019**, *10* (4), 1046–1051.

(58) Kojima, M.; Abe, S.; Ueno, T. Engineering of Protein Crystals for Use as Solid Biomaterials. *Biomater. Sci.* **2022**, *10* (2), 354–367.

(59) Hong, F.; Jiang, S.; Lan, X.; Narayanan, R. P.; Šulc, P.; Zhang, F.; Liu, Y.; Yan, H. Layered-Crossover Tiles with Precisely Tunable Angles for 2D and 3D DNA Crystal Engineering. *J. Am. Chem. Soc.* **2018**, *140* (44), 14670–14676.

(60) Simmons, C. R.; MacCulloch, T.; Zhang, F.; Liu, Y.; Stephanopoulos, N.; Yan, H. A Self-Assembled Rhombohedral DNA Crystal Scaffold with Tunable Cavity Sizes and High-Resolution Structural Detail. *Angewandte Chemie International Edition* **2020**, *59* (42), 18619–18626.

(61) Ernst, P.; Plückthun, A.; Mittl, P. R. E. Structural Analysis of Biological Targets by Host:Guest Crystal Lattice Engineering. *Sci Rep* **2019**, *9* (1), 1–11.

(62) Martín-Benito, J.; Bertrand, S.; Hu, T.; Ludtke, P. J.; McLaughlin, J. N.; Willardson, B. M.; Carrascosa, J. L.; Valpuesta, J. M. Structure of the Complex between the Cytosolic Chaperonin CCT and Phosducin-like Protein. *Proceedings of the National Academy of Sciences* **2004**, *101* (50), 17410–17415.

(63) Tereshko, V.; Uysal, S.; Koide, A.; Margalef, K.; Koide, S.; Kossiakoff, A. A. 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* **2008**, *17* (7), 1175–1187.

(64) Nawarathnage, S.; Soleimani, S.; Mathis, M. H.; Bezzant, B. D.; Ramírez, D. T.; Gajjar, P.; Bunn, D. R.; Stewart, C.; Smith, T.; Pedroza Romo, M. J.; Brown, S.; Doukov, T.; Moody, J. D. Crystals of TELSAM-Target Protein Fusions That Exhibit Minimal Crystal Contacts and Lack Direct Inter-TELSAM Contacts. *Open Biol* **2022**, *12* (3), 210271.

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

(66) Partridge, B. E.; Winegar, P. H.; Han, Z.; Mirkin, C. A. Redefining Protein Interfaces within Protein Single Crystals with DNA. *J. Am. Chem. Soc.* **2021**, *143* (23), 8925–8934.

(67) Coté, M. L.; Yohannan, S. J.; Georgiadis, M. M. Use of an N-Terminal Fragment from Moloney Murine Leukemia Virus Reverse Transcriptase to Facilitate Crystallization and Analysis of a Pseudo-16-Mer DNA Molecule Containing G-A Mispairs. *Acta Crystallogr D Biol Crystallogr* **2000**, *56* (Pt 9), 1120–1131.

(68) Hays, F. A.; Teegarden, A.; Jones, Z. J. R.; Harms, M.; Raup, D.; Watson, J.; Cavaliere, E.; Ho, P. S. How Sequence Defines Structure: A Crystallographic Map of DNA Structure and Conformation. *Proceedings of the National Academy of Sciences* **2005**, *102* (20), 7157–7162.

(69) Lu, X.-J.; Olson, W. K. 3DNA: A Versatile, Integrated Software System for the Analysis, Rebuilding and Visualization of Three-Dimensional Nucleic-Acid Structures. *Nat Protoc* **2008**, *3* (7), 1213–1227.

(70) Zadeh, J. N.; Steenberg, C. D.; Bois, J. S.; Wolfe, B. R.; Pierce, M. B.; Khan, A. R.; Dirks, R. M.; Pierce, N. A. NUPACK: Analysis and Design of Nucleic Acid Systems. *Journal of Computational Chemistry* **2011**, *32* (1), 170–173.

(71) Li, Z.; Zheng, M.; Liu, L.; Seeman, N. C.; Mao, C. 5′-Phosphorylation Strengthens Sticky-End Cohesions. *J. Am. Chem. Soc.* **2021**, *143* (37), 14987–14991.

(72) Holder, T. Crystallographic Symmetry Related Commands, version 2012; PymolWiki, **2012**.

(73) Komori, H.; Matsunaga, F.; Higuchi, Y.; Ishiai, M.; Wada, C.; Miki, K. Crystal Structure of a Prokaryotic Replication Initiator Protein Bound to DNA at 2.6 A Resolution. *EMBO J* **1999**, *18* (17), 4597–4607.

(74) Ward, A. R.; Dmytriw, S.; Vajapayajula, A.; Snow, C. D. Stabilizing DNA–Protein Co-Crystals via Intra-Crystal Chemical Ligation of the DNA. *Crystals* **2022**, *12* (1), 49.

(75) Orun, A. R.; Snow, C. D. Scripts for Designing Expanded, Isoreticular Co-Crystals and Crystal Datasets with Guest Protein.; *Zenodo*, **2022**. https://doi.org/10.5281/zenodo.6646957.

(76) Olson, W. K.; Gorin, A. A.; Lu, X.-J.; Hock, L. M.; Zhurkin, V. B. DNA Sequence-Dependent Deformability Deduced from Protein–DNA Crystal Complexes. *PNAS* **1998**, *95* (19), 11163–11168.

(77) Kabsch, W. XDS. *Acta Crystallogr D Biol Crystallogr* **2010**, *66* (Pt 2), 125–132.

(78) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. *PHENIX* : A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallographica Section D Biological Crystallography* **2010**, *66* (2), 213–221.

(79) Cowtan, K.; Emsley, Paul. Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallographica Section D Biological Crystallography* **2004**, *60* (12), 2126–2132.

(80) Hasell, T.; Culshaw, J. L.; Chong, S. Y.; Schmidtmann, M.; Little, M. A.; Jelfs, K. E.; Pyzer-Knapp, E. O.; Shepherd, H.; Adams, D. J.; Day, G. M.; Cooper, A. I. Controlling the Crystallization of Porous Organic Cages: Molecular Analogs of Isoreticular Frameworks Using Shape-Specific Directing Solvents. *J. Am. Chem. Soc.* **2014**, *136* (4), 1438–1448.

(81) Safarifard, V.; Rodríguez-Hermida, S.; Guillerm, V.; Imaz, I.; Bigdeli, M.; Azhdari Tehrani, A.; Juanhuix, J.; Morsali, A.; Casco, M. E.; Silvestre-Albero, J.; Ramos-Fernandez, E. V.; Maspoch, D. Influence of the Amide Groups in the CO2/N2 Selectivity of a Series of Isoreticular, Interpenetrated Metal–Organic Frameworks. *Crystal Growth & Design* **2016**, *16* (10), 6016–6023.

(82) Tranchemontagne, D. J.; Hunt, J. R.; Yaghi, O. M. Room Temperature Synthesis of Metal-Organic Frameworks: MOF-5, MOF-74, MOF-177, MOF-199, and IRMOF-0. *Tetrahedron* **2008**, *64* (36), 8553–8557.

(83) Sanner, M. F.; Olson, A. J.; Spehner, J. C. Reduced Surface: An Efficient Way to Compute Molecular Surfaces. *Biopolymers* **1996**, *38* (3), 305–320.

(84) Li, Z.; Zheng, M.; Liu, L.; Seeman, N. C.; Mao, C. 5′-Phosphorylation Strengthens Sticky-End Cohesions. *J. Am. Chem. Soc.* **2021**, *143* (37), 14987–14991.

(85) Weichenberger, C. X.; Afonine, P. V.; Kantardjieff, K.; Rupp, B. The Solvent Component of Macromolecular Crystals. *Acta Crystallogr D Biol Crystallogr* **2015**, *71* (Pt 5), 1023–1038.

(86) Fraenkel, E.; Rould, M. A.; Chambers, K. A.; Pabo, C. O. Engrailed Homeodomain-DNA Complex at 2.2 å Resolution: A Detailed View of the Interface and Comparison with Other Engrailed Structures11Edited by T. Richmond. *Journal of Molecular Biology* **1998**, *284* (2), 351–361.

(87) Goodwin, K. D.; Long, E. C.; Georgiadis, M. M. A Host–Guest Approach for Determining Drug–DNA Interactions: An Example Using Netropsin. *Nucleic Acids Research* **2005**, *33* (13), 4106–4116.

(88) Delano, W. *The PyMOL Molecular Graphics System.*; **2002**.

(89) Laskowski, R. A.; Swindells, M. B. LigPlot+: Multiple Ligand–Protein Interaction Diagrams for Drug Discovery. *J. Chem. Inf. Model.* **2011**, *51* (10), 2778–2786.

(90) Paukstelis, P. J.; Seeman, N. C. 3D DNA Crystals and Nanotechnology. *Crystals* **2016**, *6* (8), 97.

(91) Russo Krauss, I.; Merlino, A.; Vergara, A.; Sica, F. An Overview of Biological Macromolecule Crystallization. *Int J Mol Sci* **2013**, *14* (6), 11643–11691.

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

(93) Lusty, C. J. A Gentle Vapor-Diffusion Technique for Cross-Linking of Protein Crystals for Cryocrystallography. *Journal of Applied Crystallography* **1999**, *32* (1), 106–112.

(94) Yan, E.-K.; Cao, H.-L.; Zhang, C.-Y.; Lu, Q.-Q.; Ye, Y.-J.; He, J.; Huang, L.-J.; Yin, D.-C. Cross-Linked Protein Crystals by Glutaraldehyde and Their Applications. *RSC Adv.* **2015**, *5* (33), 26163–26174.

(95) Zhang, D.; Paukstelis, P. J. Enhancing DNA Crystal Durability through Chemical Crosslinking. *Chembiochem* **2016**, *17* (12), 1163–1170.

(96) Abdallah, H. O.; Ohayon, Y. P.; Chandrasekaran, A. R.; Sha, R.; Fox, K. R.; Brown, T.; Rusling, D. A.; Mao, C.; Seeman, N. C. Stabilisation of Self-Assembled DNA Crystals by Triplex-Directed Photo-Cross-Linking. *Chem. Commun.* **2016**, *52* (51), 8014–8017.

(97) Gerling, T.; Kube, M.; Kick, B.; Dietz, H. Sequence-Programmable Covalent Bonding of Designed DNA Assemblies. *Science Advances* **2018**, *4* (8), eaau1157.

(98) Li, Z.; Liu, L.; Zheng, M.; Zhao, J.; Seeman, N. C.; Mao, C. Making Engineered 3D DNA Crystals Robust. *J. Am. Chem. Soc.* **2019**, *141* (40), 15850–15855.

(99) Obianyor, C.; Newnam, G.; Clifton, B.; Grover, M. A.; Hud, N. V. Impact of Substrate-Template Stability, Temperature, Phosphate Location, and Nick-Site Base Pairs on Non-Enzymatic DNA Ligation: Defining Parameters for Optimization of Ligation Rates and Yields with Carbodiimide Activation. *bioRxiv* **2019**, 821017.

(100) Fraccia, T. P.; Smith, G. P.; Zanchetta, G.; Paraboschi, E.; Yi, Y.; Walba, D. M.; Dieci, G.; Clark, N. A.; Bellini, T. Abiotic Ligation of DNA Oligomers Templated by Their Liquid Crystal Ordering. *Nature Communications* **2015**, *6*, 6424.

(101) Kramer, M.; Richert, C. Enzyme-Free Ligation of 5'-Phosphorylated Oligodeoxynucleotides in a DNA Nanostructure. *Chem Biodivers* **2017**, *14* (9).

(102) Morgunova, E.; Yin, Y.; Jolma, A.; Dave, K.; Schmierer, B.; Popov, A.; Eremina, N.; Nilsson, L.; Taipale, J. Structural Insights into the DNA-Binding Specificity of E2F Family Transcription Factors. *Nat Commun* **2015**, *6*, 10050.

(103) Yan, E.-K.; Lu, Q.-Q.; Zhang, C.-Y.; Liu, Y.-L.; He, J.; Chen, D.; Wang, B.; Zhou, R.-B.; Wu, P.; Yin, D.-C. Preparation of Cross-Linked Hen-Egg White Lysozyme Crystals Free of Cracks. *Scientific Reports* **2016**, *6* (1), 34770.

(104) Bi, S.; Pollard, A. M.; Yang, Y.; Jin, F.; Sourjik, V. Engineering Hybrid Chemotaxis Receptors in Bacteria. *ACS Synth. Biol.* **2016**, *5* (9), 989–1001.

(105) Ward, A. R.; Snow, C. D. *Scripts for Modeling Chemical Ligation of DNA Junctions within Biomolecular Crystals*; Zenodo, **2021**.

(106) Newville, M.; Stensitzki, T.; Allen, D. B.; Ingargiola, A. *LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python*; Zenodo, **2014**.

(107) Gelbin, A.; Schneider, B.; Clowney, L.; Hsieh, S.-H.; Olson, W. K.; Berman, H. M. Geometric Parameters in Nucleic Acids: Sugar and Phosphate Constituents. *J. Am. Chem. Soc.* **1996**, *118* (3), 519–529.

(108) Falkner, J. C.; Al-Somali, A. M.; Jamison, J. A.; Zhang, J.; Adrianse, S. L.; Simpson, R. L.; Calabretta, M. K.; Radding, W.; Phillips; Colvin, V. L. Generation of Size-Controlled, Submicrometer Protein Crystals. *Chem. Mater.* **2005**, *17* (10), 2679–2686.

(109) Saoji, M.; Zhang, D.; Paukstelis, P. J. Probing the Role of Sequence in the Assembly of Three-Dimensional DNA Crystals. *Biopolymers* **2015**, *103* (11), 618–626.

(110) Hashimoto, T.; Ye, Y.; Ui, M.; Ogawa, T.; Matsui, T.; Tanaka, Y. Protein Encapsulation in the Hollow Space of Hemocyanin Crystals Containing a Covalently Conjugated Ligand. *Biochemical and Biophysical Research Communications* **2019**, *514* (1), 31–36.

(111) Duan, B.; Fu, D.; Zhang, C.; Ding, P.; Dong, X.; Xia, B. Selective Nonmethylated CpG DNA Recognition Mechanism of Cysteine Clamp Domains. *J. Am. Chem. Soc.* **2021**, *143* (20), 7688–7697.

APPENDIX I. SUPPLEMENTAL INFORMATION – CHAPTER 2

Engineered protein-DNA co-crystals as scaffolds for DNA-binding molecules

Table of Contents

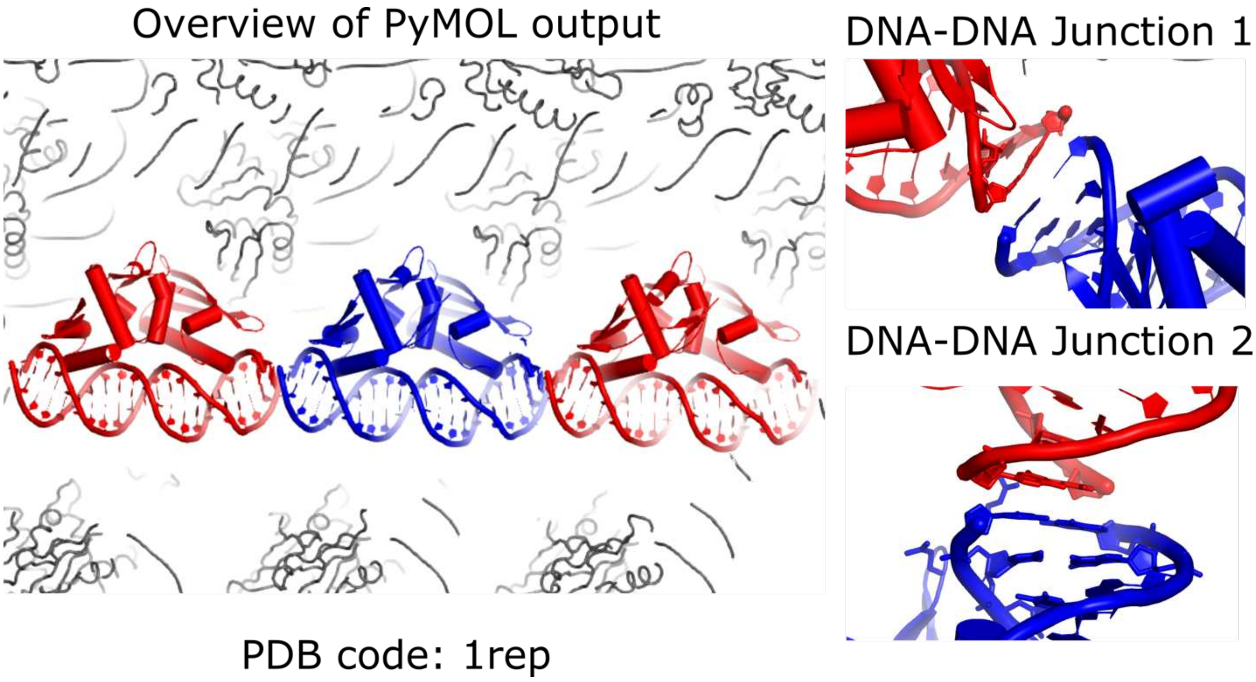


Figure S2.1. Check for co-crystal stacks and base pairs at the DNA-DNA junctions. The Snow Lab custom Python script generates a pdb file for inspection in PyMOL. The asymmetric copy is highlighted in blue and flanking symmetric copies with C1’ that are potentially stacked are highlighted in red. Both are represented with cartoon protein and DNA. The remainder of the protein and DNA in the crystal are minimal as to guide the eye for inspection of the junctions. The pdb output has two scenes saved, one for each DNA-DNA junction that may be stacked. Junction 1 and 2 shows the scene zoomed in, highlighting the stacked DNA junctions in PDB entry 1rep.

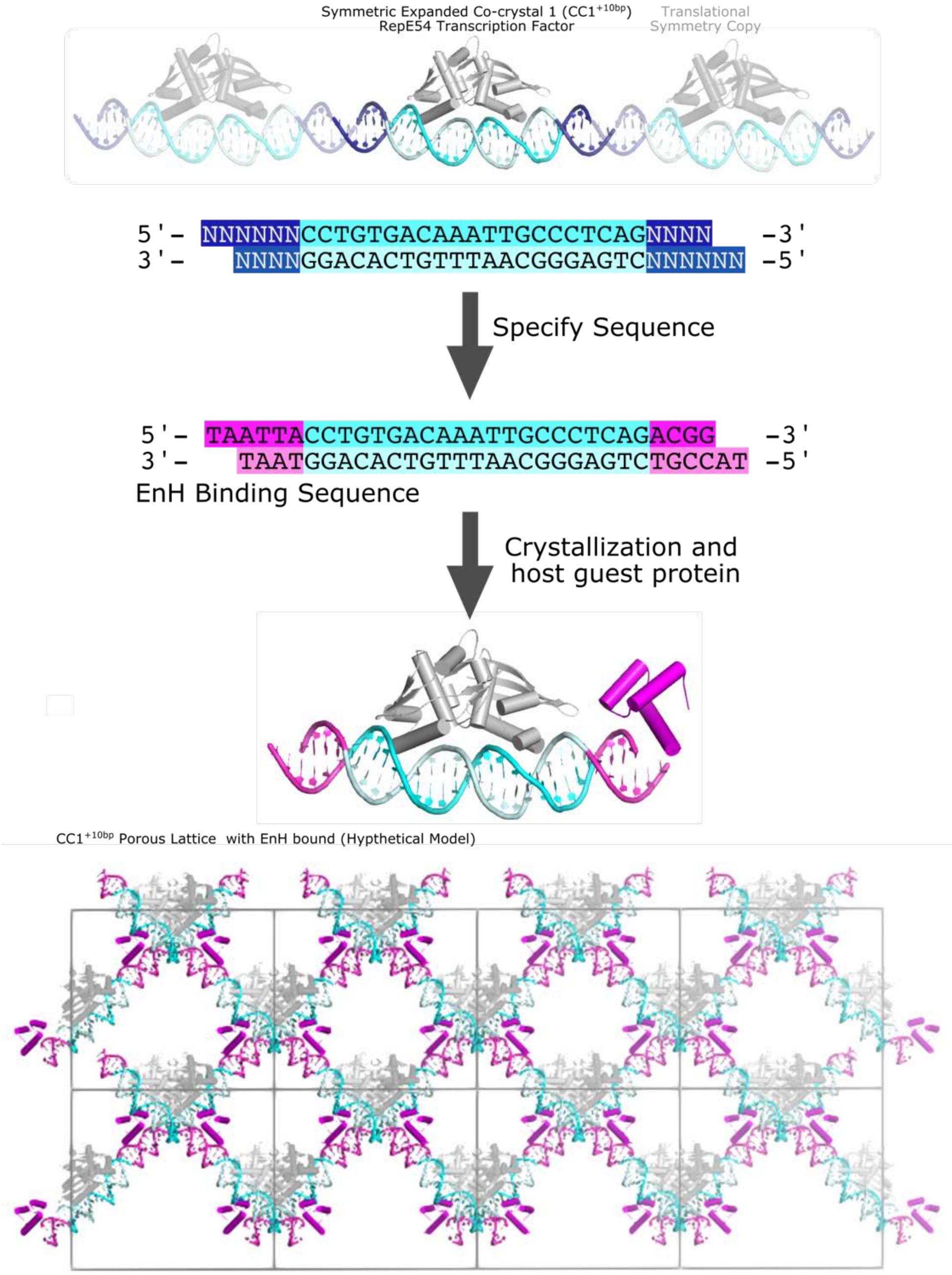


Figure S2.2. Expanded, modular DNA to host DNA-binding guests. An example shown where we can specify the expanded sequence for a DNA-binding protein, Engrailed Homeodomain (EnH). Post-crystallization, one may hypothetically host EnH at the binding site in each asymmetric unit at the EnH binding site for full guest incorporation in the co-crystal X-ray diffraction pattern.

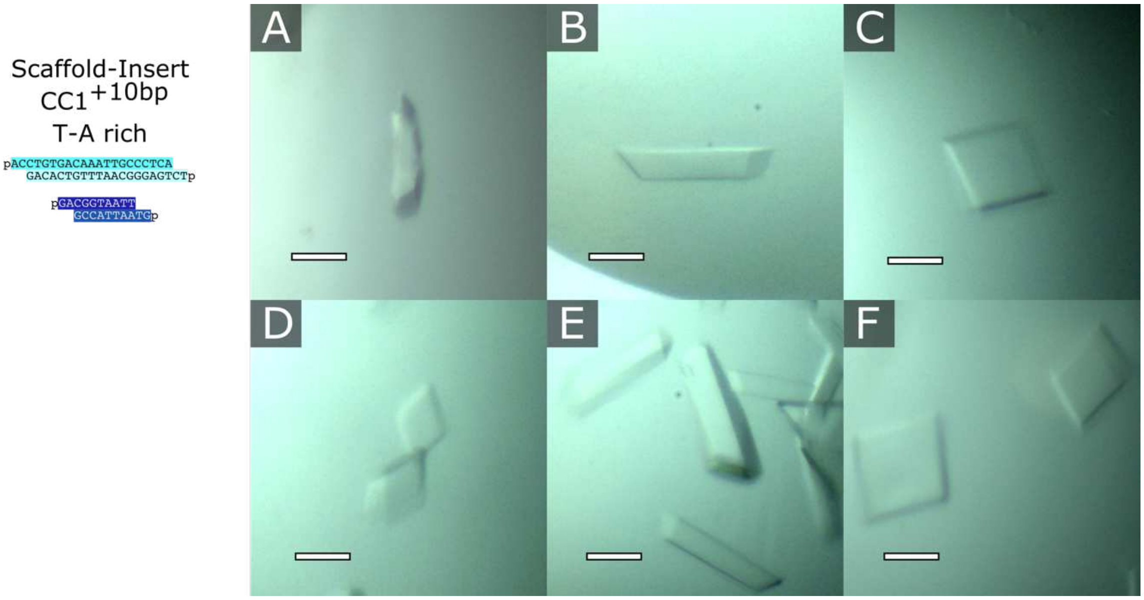


Figure S2.3. Crystallization screen results. Gryphon crystallization screening resulted in crystal growth after 2 days. Corresponding crystallization conditions are in Table S2.2. The IndexHT screen produced crystals A and B. The NatrixHT screen produced crystals C-F. After data collection and refinement, crystal C was the intended porous lattice in space group *I*121.

Table S2.1. DNA duplexes for porous crystal growth.

DNA-DNA Junction	Duplex Description	Duplex Sequence
0SB 3'phos	CC1^{+10bp} Symmetric G-C rich (1:1.2 RepE54:DNA)	5′ CCCGGACCTGTGACAAATTGCCCTCAGACGGp 3′ 3′ pGGGCCTGGACACTGTTTAACGGGAGTCTGCC 5′
0SB 5'phos	CC1^{+10bp} Symmetric G-C rich (1:1.2 RepE54:DNA)	5′ pCCCGGACCTGTGACAAATTGCCCTCAGACGG 3′ 3′ GGGCCTGGACACTGTTTAACGGGAGTCTGCCp 5′
1SB 3'phos	CC1^{+10bp} Symmetric G-C rich (1:1.2 RepE54:DNA)	5′ CCCGGACCTGTGACAAATTGCCCTCAGACGGp 3′ 3′ pGGCCTGGACACTGTTTAACGGGAGTCTGCCG 5′
3SB 3'phos	CC1^{+10bp} Symmetric G-C rich (1:1.2 RepE54:DNA)	5′ CCCGGACCTGTGACAAATTGCCCTCAGACGGp 3′ 3′ pCCTGGACACTGTTTAACGGGAGTCTGCCGGG 5′
2SB 5'phos	CC1^{+10bp} Symmetric T-A rich (1:1.2 RepE54:DNA)	5′ pTAATTACCTGTGACAAATTGCCCTCAGACGG 3′ 3′ TAATGGACACTGTTTAACGGGAGTCTGCCATp 5′
2SB 5'phos	CC1^{+10bp} Scaffold-Insert G-C rich (1:1.2:1.2 RepE54:Scaffold DNA:Insert DNA)	Scaffold 5′ pACCTGTGACAAATTGCCCTCA 3′ 3′ GACACTGTTTAACGGGAGTCTp 5′ Insert 5′ pGACGGCCCGG 3′ 3′ GCCGGGCCTGp 5′
2SB 5'phos	CC1^{+10bp} Scaffold-Insert T-A rich (1:1.2:1.2 RepE54:Scaffold DNA:Insert DNA)	Scaffold 5′ pACCTGTGACAAATTGCCCTCA 3′ 3′ GACACTGTTTAACGGGAGTCTp 5′ Insert 5′ pGACGGTAATT 3′ 3′ GCCATTAATGp 5′

Table S2.2. Gryphon crystallization conditions and X-ray diffraction quality.

Screen	Crystal	Crystallization condition	X-ray Diffraction Resolution (space group)
IndexHT	A	0.1 M tris pH 8.5, 0.3 M magnesium formate dihydrate	
IndexHT	B	5% v/v tacsimate pH 7.0, 0.1 M HEPES pH 7.0, 10% w/v polyethylene glycol monomethyl ether 5,000	3.15 Å (*I*121)
NatrixHT	C	0.01 M magnesium acetate tetrahydrate, 0.05 M sodium cacodylate trihydrate pH 6.5, 1.3 M lithium sulfate monohydrate	4.16 Å (*I*121)
NatrixHT	D	0.01 M magnesium acetate tetrahydrate, 0.05 M sodium cacodylate trihydrate pH 6.5, 2 M ammonium sulfate	
NatrixHT	E	0.2 M potassium chloride, 0.025 M magnesium sulfate hydrate, 0.05 M HEPES sodium pH 7.0, 20% v/v polyethylene glycol 200	
NatrixHT	F	0.05 M sodium cacodylate trihydrate pH 6.0, 35% v/v tacsimate^{TM} pH 6.0, 0.001 M spermine	

Protocol S2.2. Estimate crystal solvent fraction.

To estimate the solvent fraction we use CCTBX* and the following script, adapted from a script by Keitaro Yamashita (https://gist.github.com/keitaroyam/7a06528b81a2f7eb86853adc27bec8fd:) When applied to our original target model for CC1 expanded by 10 bp, this script estimates a mask-based solvent content of 85% and solvent content of 80.6% on the basis of having within each asymmetric unit 62 DNA bases, 214 structured amino acids, and 49 disordered amino acids.

```
#!/usr/bin/env python
import iotbx.pdb
import mmtbx.utils
from mmtbx.scaling.matthews import p_vm_calculator

def run(pdbin, missingN=0):
  pdb_inp = iotbx.pdb.input(pdbin)
  xray_structure = pdb_inp.xray_structure_simple()
  print("Read from %s" % pdb_in)
  xray_structure.show_summary(prefix="  ")
  print()
  print("Mask-based calculation")
  print(" Solvent Content:",
mmtbx.utils.f_000(xray_structure=xray_structure).solvent_fraction)
  print()

  hierarchy = pdb_inp.construct_hierarchy()
  counts = hierarchy.overall_counts()
  num_aa = counts.resname_classes.get("common_amino_acid", 0)
  num_na = counts.resname_classes.get("common_rna_dna", 0)
  print("Sequence-based calculation")
  print(" Number of amino acids: %d + %d" % (num_aa,missingN))
  print(" Number of       bases: %d" % num_na)

  vm_calc = p_vm_calculator(xray_structure.crystal_symmetry(),
n_residues=(num_aa+missingN), n_bases=num_na)
  print(" Solvent Content: ", vm_calc.solc(vm=vm_calc.vm(copies=1)))
# run()

if __name__ == "__main__":
  import sys
  try:
    pdb_in, missingN = sys.argv[1], int(sys.argv[2])
  except:
    print('Provide input pdb and a number of non-structured amino acids to consider')
    sys.exit()
  run(pdb_in, missingN)
```

The symmetry definition for our original model is:

```
CRYST1  161.431  121.858   73.891  90.00 121.53  90.00 C 1 2 1       4
```

*Grosse-Kunstleve, R. W.; Sauter, N. K.; Moriarty, N. W.; Adams, P. D. The Computational Crystallography Toolbox: Crystallographic Algorithms in a Reusable Software Framework. *Journal of Applied Crystallography* **2002**, *35* (1), 126–136.

APPENDIX II. SUPPLEMENTAL INFORMATION – CHAPTER 3

Interpenetrating, expanded co-crystals provide sequence-dependent synchronous guest installation

Table of Contents

Table S3.6. X-ray diffraction statistics for CC1^{+10bp} Scaffold-Insert G-C rich PDB entry 7ufx and CC1^{+10bp} Scaffold-Insert T-A rich PDB entry 7ur0.

Table S3.7. X-ray diffraction statistics for TAMRA-labelled guest DNA co-crystallized with CC1^{+10bp} Scaffold-Insert G-C rich PDB entry 7uv6 and CC1^{+10bp} Scaffold-Insert T-A rich PDB entry 7uv7.

Table S3.8. X-ray diffraction statistics for guest netropsin co-crystallized with CC1^{+10bp} Symmetric G-C rich PDB entry 8d86.

Table S3.9. X-ray diffraction statistics for RepE54 scaffold protein variant CC1^{+10bp} Symmetric G-C rich PDB entry 7uxy.

Protocol S3.1. Protein sequences for cloning and overexpression in *E. coli*.

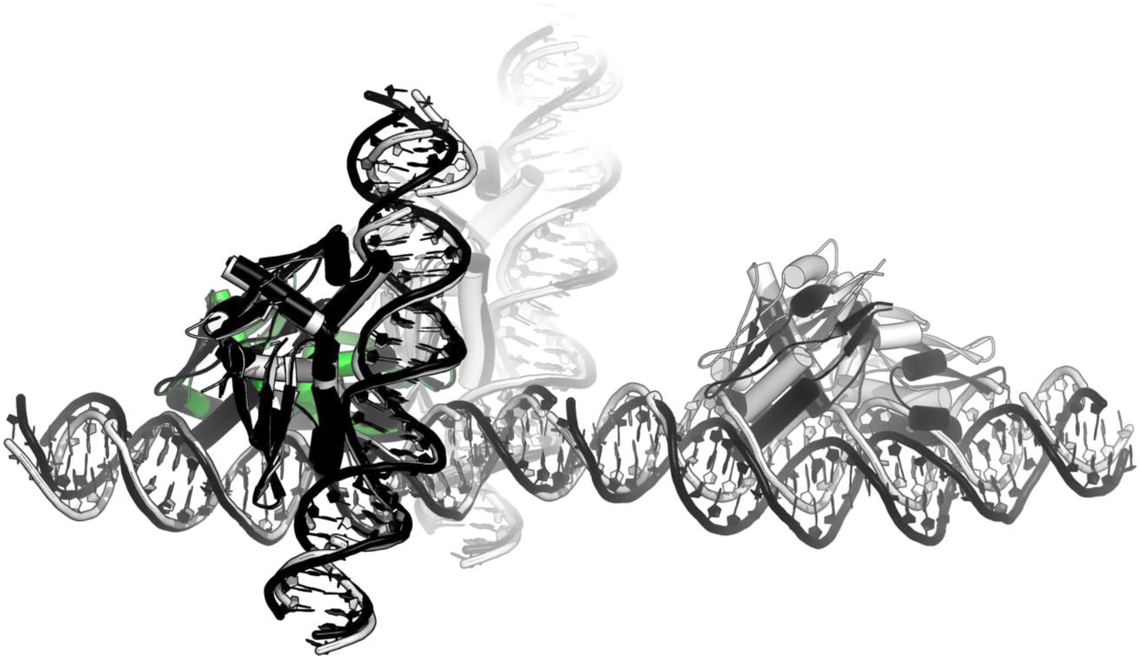


Figure S3.1. Superposition of design and experimental proteins. The central building block (protein-DNA complex) is connected to neighbors via 3 distinct symmetry operations. Thus a useful minimal construct to illustrate these symmetry operations is a local portion of the lattice consisting of a central protein-DNA complex, one symmetry neighbor (translation only) connected via a DNA:DNA stacking interface and two symmetry neighbors (translation and 180 degree rotation) connected via protein:protein interfaces. This figure shows the differences between the design model (white and green) and the neighbors as found in the I222 interpenetrating lattice (black). The I222 lattice was superimposed by aligning the central protein (black) to the central protein in the design model (green). It is possible to see that the neighbors connected via the protein-protein interfaces closely match the design model whereas the neighbor connected via DNA-DNA stacking has a larger difference.

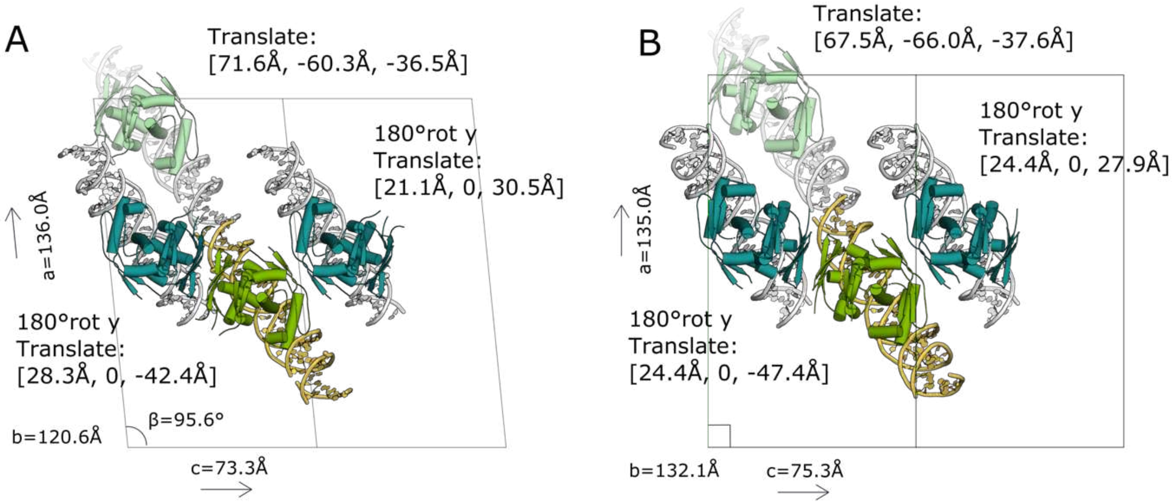


Figure S3.2. Comparison of designed and interpenetrating lattice. Four protein-DNA complexes in the CC1 lattice represent the three distinct interfaces in the crystal: the two protein-protein contacts and the one DNA-DNA junction. (A) The four building units of the hypothetical porous model crystal with unit cell dimensions. (B) The same four building units in the experimental interpenetrating model. Each are labeled with the "symmetry operation" needed to get the neighbor (where rotations are applied about the center of geometry). This view highlights the remarkable similarity of the obtained protein-protein interfaces to the interfaces in the design model.

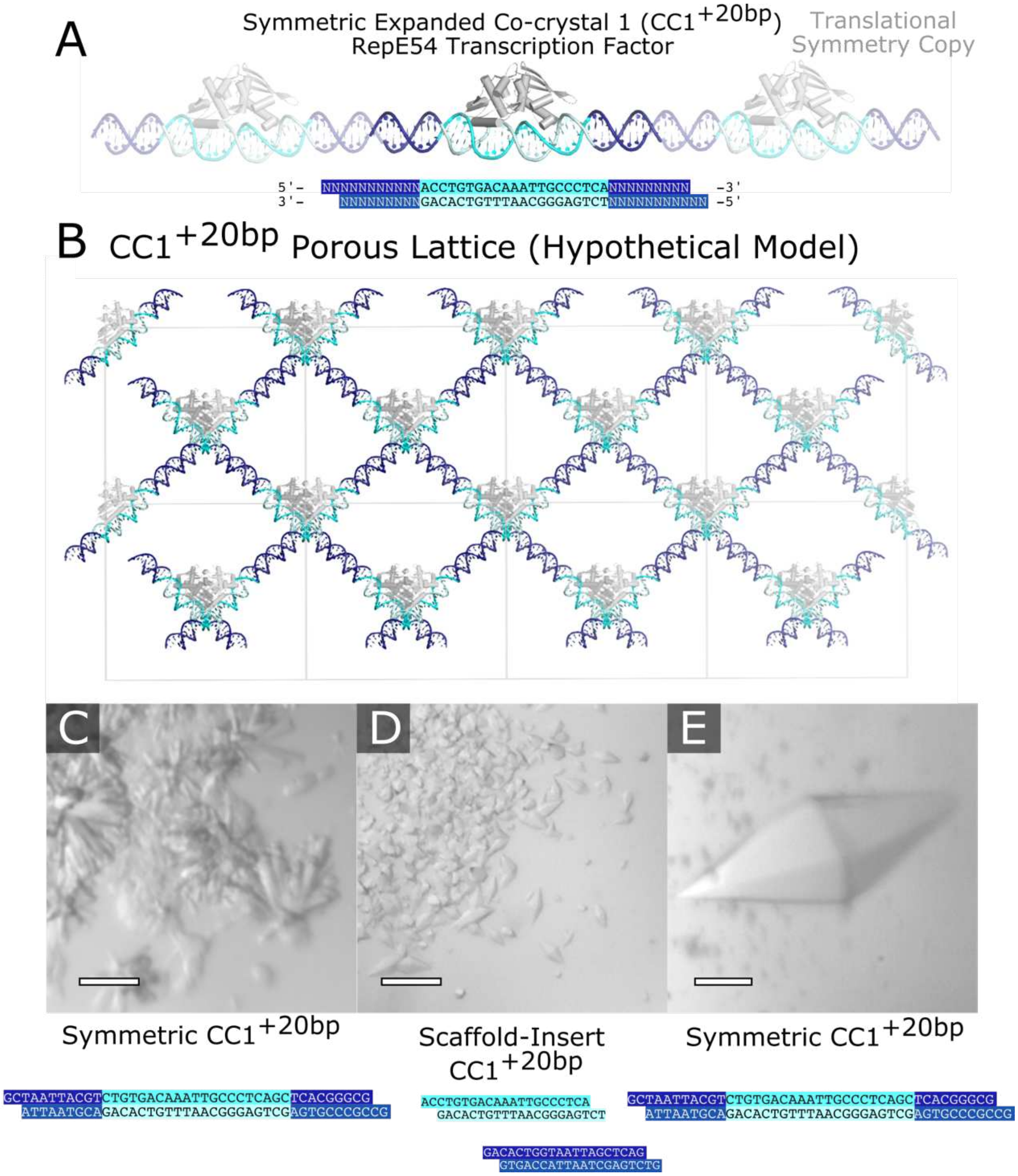
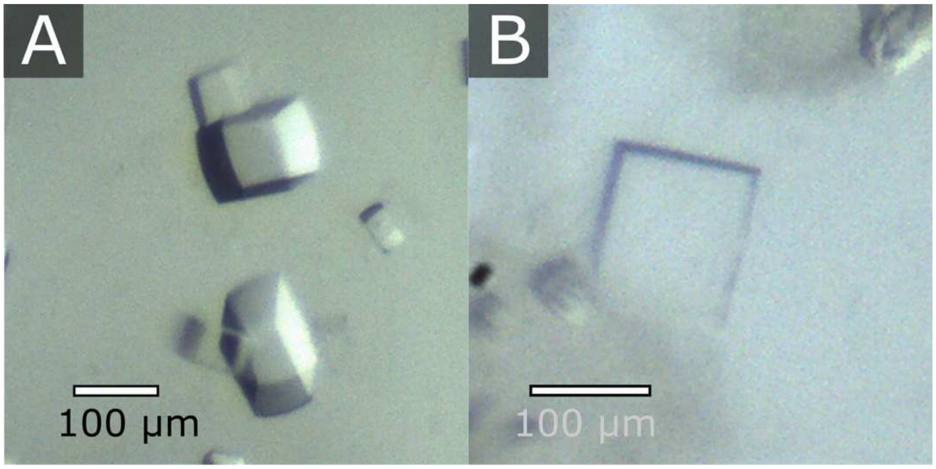


Figure S3.3. CC1 20 bp expansion and crystal growth. The asymmetric unit of a hypothetical 20 bp expanded crystal (A) where the N's represent bases in the DNA that one may choose the complimentary sequence. The CC1^{+20bp} crystal lattice (B) is highly porous. Preliminary crystallization attempts using the symmetric scheme (C) and the scaffold-insert scheme (D) have yielded crystals too small for X-ray diffraction validation. A recent symmetric expansion crystal (E) yielded mediocre diffraction quality in P$6_5$22, and molecular replacement yielded a model in which the long DNA curved into large spirals rather than adopting the intended parallel stacks. Crystallization conditions for (E) symmetric CC1^{+20bp} were 250 mM $MgCl_2$, 30% PEG 400, and 100 mM tris HCl pH 8.0.



Symmetric T-A rich Symmetric G-C rich

Figure S3.4. LowMG crystal habit. (A) CC1^{+10bp} symmetric T-A rich crystal grown in 160 mM $MgCl_2$, 18% PEG 400, and 80 mM tris HCl pH 8.0. The same DNA duplex was used for PDB code 7u7o in Table S3.2. (B) CC1^{+10bp} symmetric G-C rich crystal grown in150 $MgCl_2$, 15% PEG 400, and 80 mM tris HCl pH 8.0. The same DNA duplex was used for PDB code 7u6k in Table S3.2.

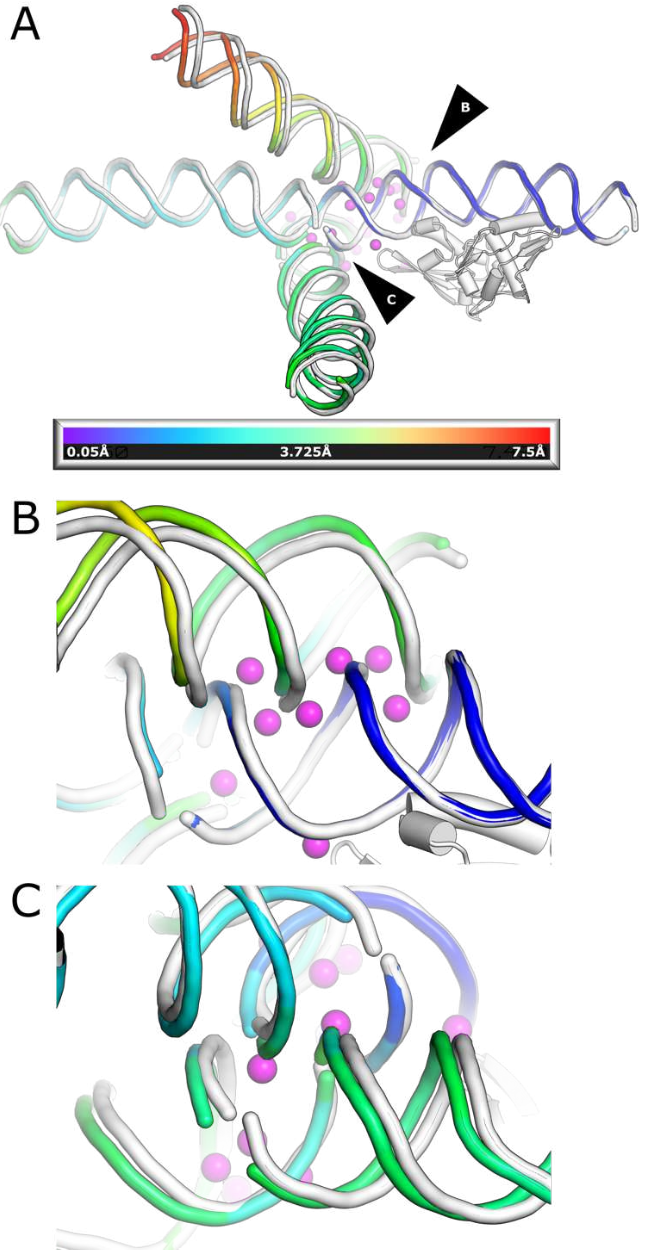


Figure S3.5. Crystal growth in Calcium *vs.* Magnesium. The packing between dsDNA was noticeably tighter in the structure (7u6k, white) where magnesium (Magenta spheres) was present at high concentration. When the structure obtained in calcium (8d8m, rainbow) is aligned to best superimpose one copy of the protein, we obtain a very close match (0.15Å rmsd over 210 alpha carbons). The DNA block bound to this protein moves very little when going from Mg to Ca. We computed the movement of each corresponding C1' atom, and dark blue corresponds to small changes. The neighboring DNA blocks that are connected *via* DNA:DNA stacking (one is shown on the left) also had modest shifts in the position with most C1' atoms moving less than 2Å. In contrast, the tightly packed DNA-DNA interfaces where multiple magnesium atoms were found in in 7u6k were noticeably less tightly packed in the calcium structure, where no obvious counterions were found at the same site.

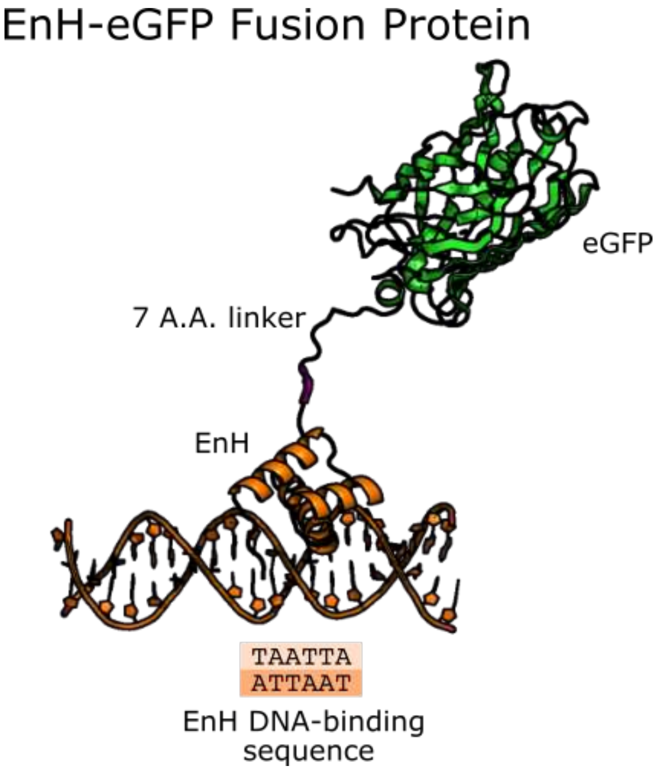


Figure S3.6. Protein guest, Engrailed Homeodomain, fused to eGFP (EnH-eGFP). Hypothetical PyMOL visual model of Engrailed Homeodomain (EnH; PDB code: 3hdd) fused to eGFP (PDB code: 4eul) by a 7 amino acid linker. The EnH-eGFP fusion protein (Protocol S3.1) is bound to the EnH cognate DNA-binding sequence, highlighted in orange.

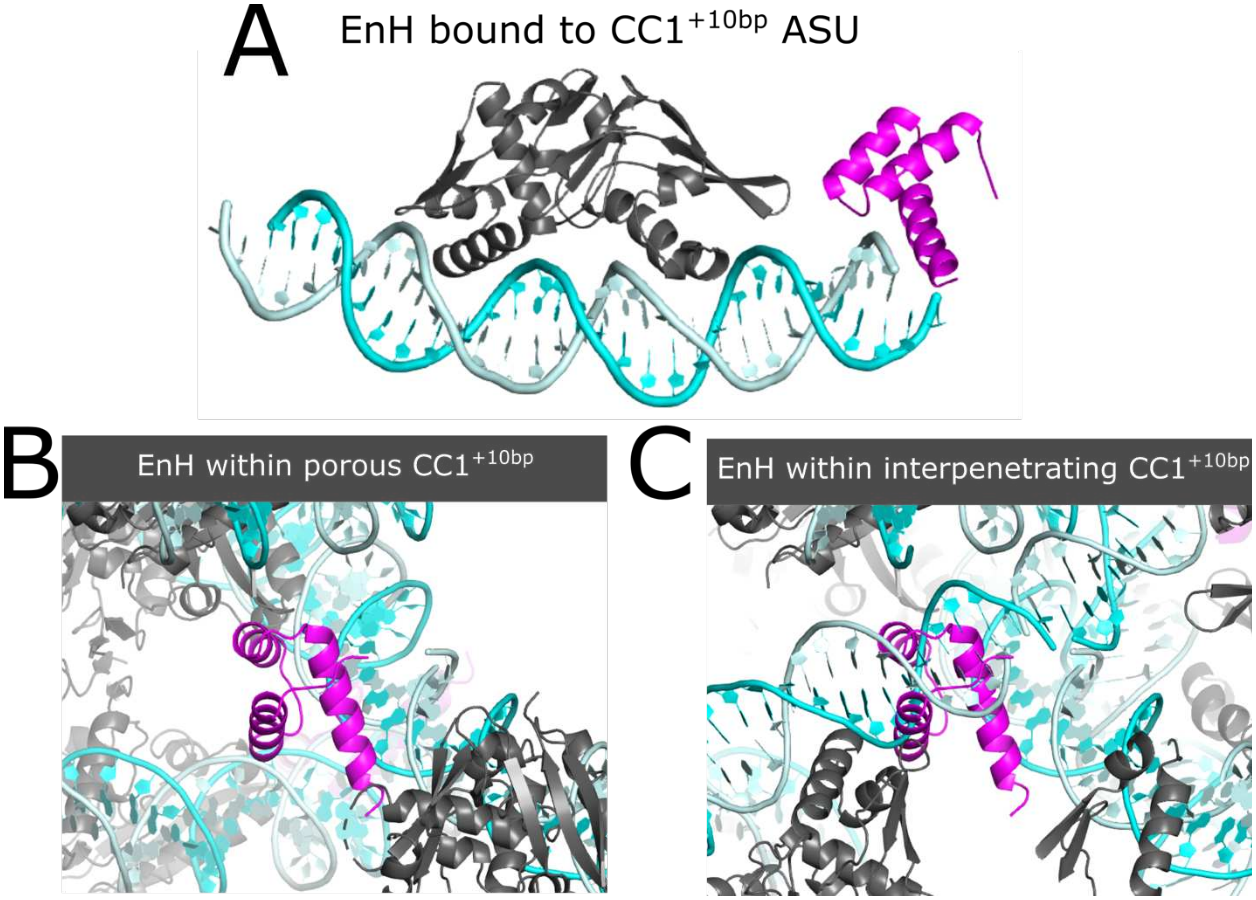


Figure S3.7. Interpenetrating CC1 blocks full guest protein entrapment. (A) The hypothetical model of the engrailed homeodomain protein (magenta) bound at the junction of the symmetric expanded CC1. Upon crystallization, in the hypothetical porous lattice (B), EnH would have ample room to bind to the DNA. However, in the interpenetrating lattice (C), EnH would not be able to bind as the closely packed DNA from the interpenetrating lattice (orange) would block the protein binding site.

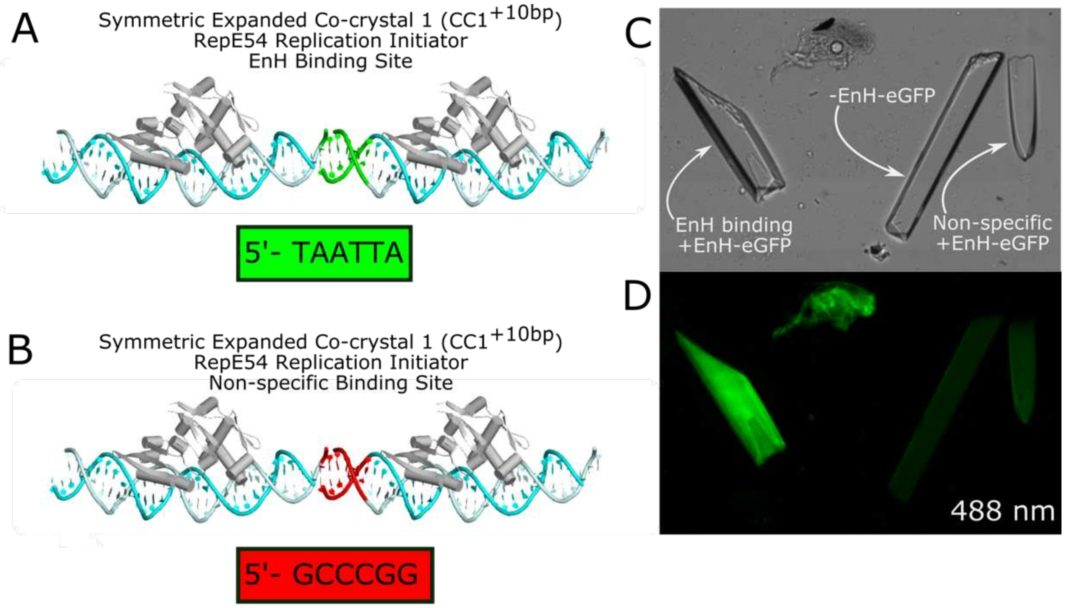


Figure S3.8. Sequence-specific fluorescent protein entrapment. Crystals with either the (A) EnH binding site or (B) an unrelated G-C rich sequence were grown in the presence of EnH-eGFP (a molar ratio of 1:1.2:1.2 RepE54 protein to DNA to EnH-eGFP). Under confocal microscopy (C-D) the EnH binding crystal co-crystallized with EnH-eGFP (EnH binding +EnH-eGFP) fluoresced brightly when exposed to 488 nm light. In contrast, the crystals lacking the EnH DNA binding site had much lower evident doping of the EnH-eGFP (C-D). The control crystal grown in the absence EnH-eGFP (-EnH-eGFP) had even lower background fluorescence.

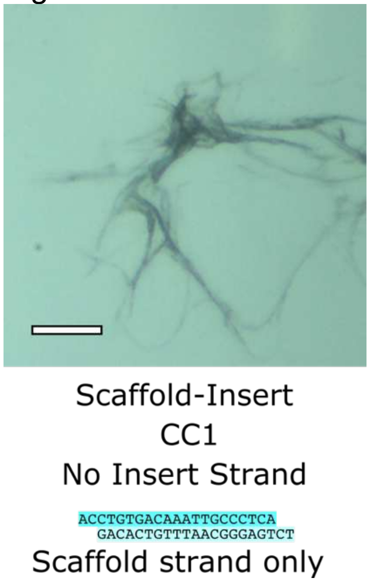


Figure S3.9. Scaffold-Insert expanded control crystals (no insert strand). The RepE54 scaffold protein was mixed in a stoichiometric ratio with the scaffold strand only (1:1.2 protein to DNA). Crystallization screening was set up in the interpenetrating crystal growth solutions (25 mM to 400 mM $MgCl_2$, 15% to 30% PEG 400, and 100 mM tris HCl pH 8.0). No crystals grew and the well above shows only precipitant.

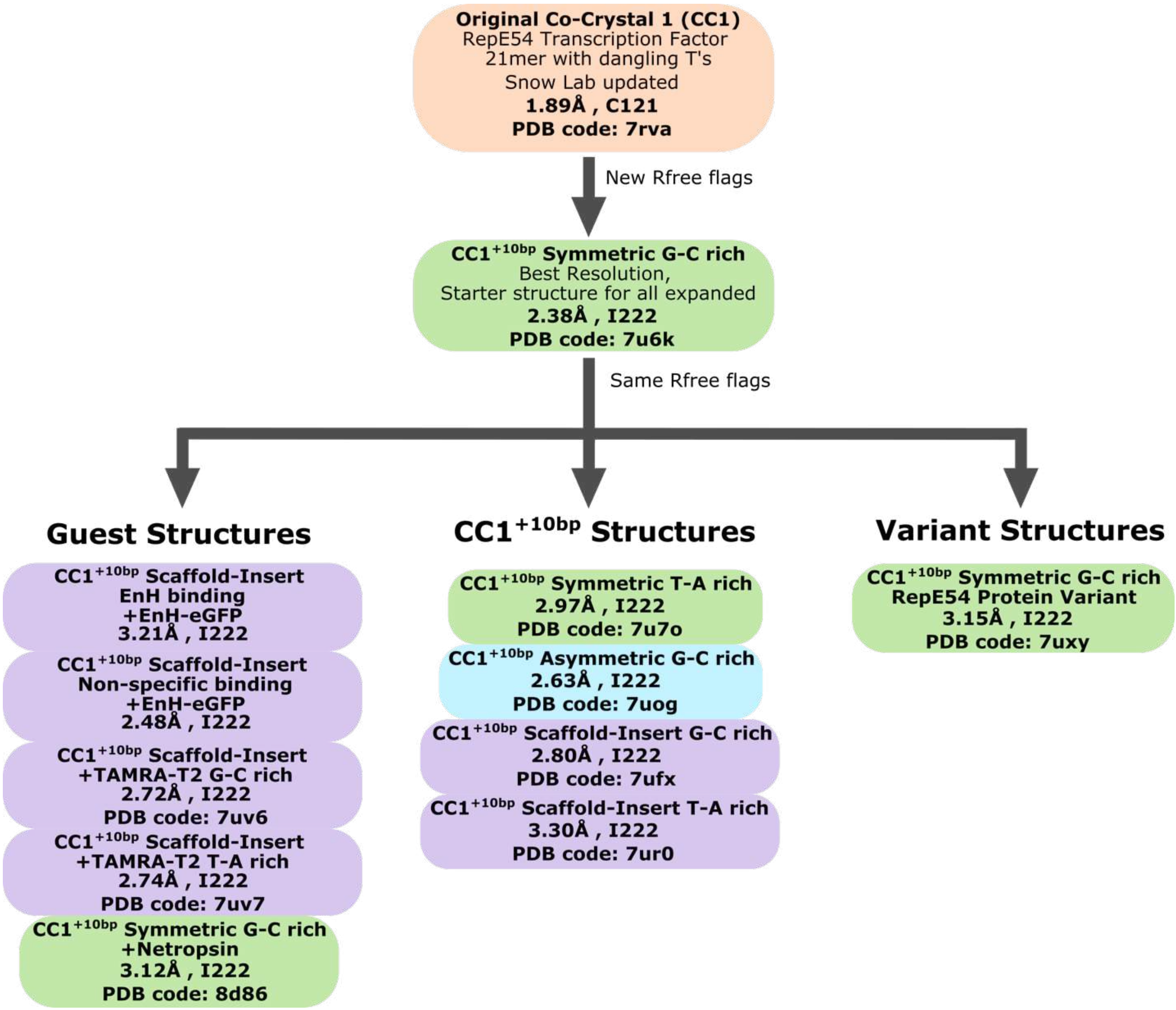


Figure S3.10. Flow chart of CC1^{+10bp} X-ray diffraction data refinement.

Table S3.1. Crystallization conditions for interpenetrating CC1^{+10bp} (corresponds to Figure 3.2).

PDB code	Crystal Description	Crystallization Condition
7u6k	CC1^{+10bp} Symmetric G-C rich (1:1.2 RepE54:DNA)	400 mM $MgCl_2$, 24% PEG 400, 80 mM tris HCl pH 8.0
7u7o	CC1^{+10bp} Symmetric T-A rich (1:1.2 RepE54:DNA)	300 mM $MgCl_2$, 28% PEG 400, 80 mM tris HCl pH 8.0
7uog	CC1^{+10bp} Asymmetric G-C rich (1:1.2 RepE54:DNA)	200 mM $MgCl_2$, 30% PEG 400, 100 mM tris HCl pH 8.0
7ufx	CC1^{+10bp} Scaffold-Insert G-C rich (1:1.2:1.2 RepE54:Scaffold DNA:Insert DNA)	180 mM $MgCl_2$, 20% PEG 400, 100 mM tris HCl pH 8.0
7ur0	CC1^{+10bp} Scaffold-Insert T-A rich (1:1.2:1.2 RepE54:Scaffold DNA:Insert DNA)	220 mM $MgCl_2$, 28% PEG 400, 100 mM tris HCl pH 8.0
Zenodo[74]	CC1^{+10bp} Scaffold-Insert G-C rich + EnH-eGFP (1:1.2:1.2:1.2 RepE54:Scaffold DNA: Insert DNA: EnH-eGFP)	400 mM $MgCl_2$, 25% PEG 400, 100 mM tris HCl pH 8.0
Zenodo[74]	CC1^{+10bp} Scaffold-Insert T-A rich + EnH-eGFP (1:1.2:1.2:1.2 RepE54:Scaffold DNA: Insert DNA: EnH-eGFP)	300 mM $MgCl_2$, 25% PEG 400, 100 mM tris HCl pH 8.0
7uv6	CC1^{+10bp} Scaffold-Insert G-C rich + TAMRA-T2 (1:1.2:1.2 RepE54:Scaffold DNA:Insert DNA)	300 mM $MgCl_2$, 15% PEG 400, 100 mM tris HCl pH 8.0
7uv7	CC1^{+10bp} Scaffold-Insert T-A rich + TAMRA-T2 (1:1.2:1.2 RepE54:Scaffold DNA:Insert DNA)	200 mM $MgCl_2$, 30% PEG 400, 100 mM tris HCl pH 8.0
8d86	CC1^{+10bp} Symmetric G-C rich +Netropsin (1:1.2:1.2 RepE54:DNA:Netropsin)	300 mM $MgCl_2$, 25% PEG 400, 100 mM tris HCl pH 8.0
7uxy	CC1^{+10bp} Symmetric G-C rich RepE54 Variant (1:1.2 RepE54:DNA)	350 mM $MgCl_2$, 30% PEG 400, 100 mM tris HCl pH 8.0

Table S3.2. DNA oligonucleotide sequences.

PDB code	Duplex Description	Duplex Sequence
7u6k and 8d86	CC1$^{+10\text{bp}}$ Symmetric G-C rich Two sticky bases 5' terminal phosphates	5′ – pCCCGGACCTGTGACAAATTGCCCTCAGACGG – 3′ 3′ – GCCTGGACACTGTTTAACGGGAGTCTGCCGGp – 5′
7u7o	CC1$^{+10\text{bp}}$ Symmetric T-A rich Two sticky bases 5' terminal phosphates	5′ – pTAATTACCTGTGACAAATTGCCCTCAGACGG – 3′ 3′ – TAATGGACACTGTTTAACGGGAGTCTGCCATp – 5′
7uog	CC1$^{+10\text{bp}}$ Asymmetric G-C rich Two sticky bases no terminal phosphates	5′ – ACCTGTGACAAATTGCCCTCAGACCCGCCTT – 3′ 3′ – GACACTGTTTAACGGGAGTCTGGGCGGAATG – 5′
7ufx and +EnH-eGFP	CC1$^{+10\text{bp}}$ Scaffold-Insert G-C rich Two sticky bases 5' terminal phosphates	Scaffold 5′ – pACCTGTGACAAATTGCCCTCA – 3′ 3′ – GACACTGTTTAACGGGAGTCTp – 5′ Insert 5′ – pGACGGCCCGG – 3′ 3′ – GCCGGGCCTGp – 5′
7ur0 and +EnH-eGFP	CC1$^{+10\text{bp}}$ Scaffold-Insert T-A rich Two sticky bases 5' terminal phosphates	Scaffold 5′ – pACCTGTGACAAATTGCCCTCA – 3′ 3′ – GACACTGTTTAACGGGAGTCTp – 5′ Insert 5′ – pGACGGTAATT – 3′ 3′ – GCCATTAATGp – 5′
7uv6	CC1$^{+10\text{bp}}$ Scaffold-Insert G-C rich + TAMRA-T2 Two sticky bases no terminal phosphates	Scaffold 5′ – ACCTGTGACAAATTGCCCTCA – 3′ 3′ – GACACTGTTTAACGGGAGTCT – 5′ Insert 5′ – GACGGCCCGG – 3′ 3′ – GCCGGGCC(T-TAMRA)G – 5′
7uv7	CC1$^{+10\text{bp}}$ Scaffold-Insert T-A rich + TAMRA-T2 Two sticky bases no terminal phosphates	Scaffold 5′ – ACCTGTGACAAATTGCCCTCA – 3′ 3′ – GACACTGTTTAACGGGAGTCT – 5′ Insert 5′ – GACGGTAATT – 3′ 3′ – GCCATTAA(T-TAMRA)G – 5′
7uxy	CC1$^{+10\text{bp}}$ Scaffold-Insert G-C rich One sticky base 5' terminal phosphates	5′ – pCCCGGACCTGTGACAAATTGCCCTCAGACGG – 3′ 3′ – GGCCTGGACACTGTTTAACGGGAGTCTGCCGp – 5′

Table S3.3. DNA oligonucleotide sequences with large features for steric hinderance.

Duplex Description	Duplex Sequence
CC1^{+10bp} Scaffold-Insert G-C rich + TAMRA-T2 Two sticky bases no terminal phosphates	Scaffold 5′ — ACCTGTGACAAATTGCCCTCA — 3′ 3′ — GACACTGTTTAACGGGAGTCT — 5′ Insert 5′ — GACGGCCCGG — 3′ 3′ — GCCGGGCC(T-TAMRA)G — 5′
CC1^{+10bp} Scaffold-Insert T-A rich + TAMRA-T2 Two sticky bases no terminal phosphates	Scaffold 5′ — ACCTGTGACAAATTGCCCTCA — 3′ 3′ — GACACTGTTTAACGGGAGTCT — 5′ Insert 5′ — GACGGTAATT — 3′ 3′ — GCCATTAA(T-TAMRA)G — 5′
CC1^{+10bp} Bulge G-C rich Two sticky bases no terminal phosphates	5′ — CCCGGACCCGTGACAAATTGCCCTCAGACGG — 3′ 3′ — GCCTGGGACACTGTTTAACGGGAGTCTGCCGG — 5′
CC1^{+10bp} PolyT Tail G-C rich Scaffold-Insert	Scaffold 5′ — ACCTGTGACAAATTGCCCTCA — 3′ 3′ — GACACTGTTTAACGGGAGTCT — 5′ Insert 5′ — AGACGGCCCG — 3′ 3′ — TGCCGGGCCTTTTTT — 5′
CC1^{+10bp} Scaffold-Insert T-A rich Two sticky bases 5’ terminal phosphates	Scaffold 5′ — pACCTGTGACAAATTGCCCTCA — 3′ 3′ — GACACTGTTTAACGGGAGTCTp — 5′ Insert 5′ — pGACGGTAATT — 3′ 3′ — GCCATTAATGp — 5′

Table 3.4. X-ray diffraction statistics for CC1^{+10bp} Symmetric G-C rich PDB entry 7u6k and CC1^{+10bp} Symmetric T-A rich PDB entry 7u7o.

	CC1^{+10bp} Symmetric G-C rich PDB code 7u6k	**CC1^{+10bp} Symmetric T-A rich PDB code 7u7o**
Data collection		
Light source	Synchrotron	Synchrotron
Wavelength (Å)	1.0	1.0
Resolution range (Å)	38.62 - 2.38 (2.465 - 2.38)*	37.42 - 2.97 (3.076 - 2.97)*
Space group	I 2 2 2	I 2 2 2
Unit cell dimensions		
a, b, c (Å)	75.261 132.066 134.986	72.141 129.124 131.29
α, β, γ (°)	90 90 90	90 90 90
Total reflections	200956 (19783)	91831 (9545)
Unique reflections	27318 (2666)	13009 (1292)
Multiplicity	7.4 (7.4)	7.1 (7.4)
Completeness (%)	99.54 (99.33)	96.91 (99.61)
Mean I/sigma(I)	27.30 (1.83)	17.03 (1.27)
Wilson B-factor	60.54	60.92
R-merge	0.04734 (1.164)	0.104 (1.529)
R-meas	0.05096 (1.253)	0.1126 (1.645)
R-pim	0.0187 (0.4589)	0.04266 (0.6028)
CC1/2	1 (0.723)	0.998 (0.666)
CC *	1 (0.916)	1 (0.894)
Refinement		
Reflections used in refinement	27267 (2665)	12639 (1292)
Reflections used for R-free	2719 (262)	1264 (130)
R-work	0.1826 (0.2995)	0.2460 (0.4603)
R-free	0.2239 (0.3633)	0.2590 (0.4636)
CC (work)	0.964 (0.792)	0.937 (0.774)
CC (free)	0.949 (0.639)	0.941 (0.740)
Number of non-hydrogen atoms	3167	3149
Macromolecules	3060	3121
Ligands	7	1
Solvent	100	27
Protein residues	210	216
RMS (bonds) (Å)	0.011	0.005
RMS (angles) (°)	1.21	0.90
Ramachandran favored (%)	99.01	98.08
Ramachandran allowed (%)	0.99	1.92
Ramachandran outliers (%)	0	0
Rotamer outliers (%)	0	0
Clashscore	3.97	10.07
Average B-factor	82.60	74.57
Macromolecules	83.08	64.14
Ligands	86.32	55.47
Solvent	67.79	51.96
Number of TLS groups	8	11

* Values in parentheses are for high-resolution shell.

Table S3.5. X-ray diffraction statistics for CC1^{+10bp} Asymmetric G-C rich PDB entry 7uog.

	CC1^{+10bp} Asymmetric G-C rich PDB code 7uog
Data collection	
Light source	Synchrotron
Wavelength (Å)	1.0
Resolution range (Å)	38.77 - 2.63 (2.724 - 2.63)*
Space group	I 2 2 2
Unit cell dimensions	
a, b, c (Å)	74.901 131.576 135.926
α, β, γ (°)	90 90 90
Total reflections	148790 (14998)
Unique reflections	20312 (1999)
Multiplicity	7.3 (7.5)
Completeness (%)	99.61 (99.80)
Mean I/sigma(I)	22.23 (1.81)
Wilson B-factor	71.19
R-merge	0.06093 (1.054)
R-meas	0.06562 (1.133)
R-pim	0.02414 (0.4117)
CC1/2	0.999 (0.779)
CC *	1 (0.936)
Refinement	
Reflections used in refinement	20287 (1996)
Reflections used for R-free	2021 (201)
R-work	0.1951 (0.3127)
R-free	0.2221 (0.3533)
CC (work)	0.959 (0.786)
CC (free)	0.930 (0.769)
Number of non-hydrogen atoms	3122
Macromolecules	3081
Ligands	5
Solvent	36
Protein residues	214
RMS (bonds) (Å)	0.011
RMS (angles) (°)	1.35
Ramachandran favored (%)	98.54
Ramachandran allowed (%)	1.46
Ramachandran outliers (%)	0
Rotamer outliers (%)	0
Clashscore	5.20
Average B-factor	88.03
Macromolecules	78.38
Ligands	88.56
Solvent	72.29
Number of TLS groups	8

* Values in parentheses are for high-resolution shell.

Table S3.6. X-ray diffraction statistics for CC1^{+10bp} Scaffold-Insert G-C rich PDB entry 7ufx and CC1^{+10bp} Scaffold-Insert T-A rich PDB entry 7ur0.

	CC1^{+10bp} Symmetric G-C rich PDB code 7ufx	CC1^{+10bp} Symmetric T-A rich PDB code 7ur0
Data collection		
Light source	Synchrotron	Synchrotron
Wavelength (Å)	1.0	1.0
Resolution range (Å)	37.4 - 2.8 (2.9 - 2.8)*	38.61 - 3.3 (3.418 - 3.3)*
Space group	I 2 2 2	I 2 2 2
Unit cell dimensions		
a, b, c (Å)	74.809 129.493 137.118	74.407 127.493 135.478
α, β, γ (°)	90 90 90	90 90 90
Total reflections	123126 (12453)	68924 (6970)
Unique reflections	16759 (1662)	10028 (967)
Multiplicity	7.3 (7.5)	6.9 (7.1)
Completeness (%)	99.58 (99.70)	97.79 (96.94)
Mean I/sigma(I)	28.22 (1.86)	5.91 (2.58)
Wilson B-factor	74.50	71.06
R-merge	0.05759 (1.117)	0.3851 (1.218)
R-meas	0.06201 (1.201)	0.4175 (1.314)
R-pim	0.02282 (0.4378)	0.1595 (0.4899)
CC1/2	0.999 (0.696)	0.99 (0.904)
CC *	1 (0.906)	0.998 (0.975)
Refinement		
Reflections used in refinement	16728 (1662)	9824 (951)
Reflections used for R-free	1665 (159)	973 (98)
R-work	0.1946 (0.3964)	0.2278 (0.2576)
R-free	0.2257 (0.4532)	0.2730 (0.2935)
CC (work)	0.951 (0.743)	0.945 (0.912)
CC (free)	0.946 (0.717)	0.903 (0.935)
Number of non-hydrogen atoms	3075	3076
Macromolecules	3057	3066
Ligands	6	2
Solvent	12	8
Protein residues	210	212
RMS (bonds) (Å)	0.005	0.004
RMS (angles) (°)	0.91	0.77
Ramachandran favored (%)	98.51	99.02
Ramachandran allowed (%)	1.49	0.98
Ramachandran outliers (%)	0	0
Rotamer outliers (%)	0	0
Clashscore	7.05	7.39
Average B-factor	89.83	114.41
Macromolecules	82.94	107.41
Ligands	90.59	120.14
Solvent	81.40	63.89
Number of TLS groups	10	10

* Values in parentheses are for high-resolution shell.

Table S3.7. X-ray diffraction statistics for TAMRA-labelled guest DNA co-crystallized with CC1^{+10bp} Scaffold-Insert G-C rich PDB entry 7uv6 and CC1^{+10bp} Scaffold-Insert T-A rich PDB entry 7uv7.

	CC1^{+10bp} Symmetric G-C rich PDB code 7uv6	CC1^{+10bp} Symmetric T-A rich PDB code 7uv7
Data collection		
Light source	Synchrotron	Synchrotron
Wavelength (Å)	1.0	1.0
Resolution range (Å)	38.93 - 2.72 (2.817 - 2.72)*	37.85 - 3.02 (3.128 - 3.02)*
Space group	I 2 2 2	I 2 2 2
Unit cell dimensions		
a, b, c (Å)	74.085 128.495 137.26	73.493 131.392 132.464
α, β, γ (°)	90 90 90	90 90 90
Total reflections	17957 (1779)	88410 (9654)
Unique reflections	7.3 (7.4)	12804 (1284)
Multiplicity	99.35 (99.72)	6.9 (7.5)
Completeness (%)	20.25 (1.77)	96.74 (99.69)
Mean I/sigma(I)	74.98	20.98 (2.97)
Wilson B-factor	0.06592 (1.13)	80.95
R-merge	0.07107 (1.216)	0.07113 (0.6448)
R-meas	0.02637 (0.4476)	0.07743 (0.6925)
R-pim	0.999 (0.739)	0.03014 (0.2514)
CC1/2	1 (0.922)	0.996 (0.91)
CC *	17957 (1779)	0.999 (0.976)
Refinement		
Reflections used in refinement	17903 (1778)	12544 (1284)
Reflections used for R-free	1783 (173)	1253 (130)
R-work	0.1956 (0.3819)	0.2312 (0.5146)
R-free	0.2241 (0.4458)	0.2681 (0.5103)
CC (work)	0.953 (0.762)	0.937 (0.731)
CC (free)	0.948 (0.538)	0.904 (0.482)
Number of non-hydrogen atoms	3115	3053
Macromolecules	3077	3044
Ligands	5	1
Solvent	33	8
Protein residues	213	212
RMS (bonds) (Å)	0.012	0.005
RMS (angles) (°)	1.48	0.83
Ramachandran favored (%)	99.02	98.53
Ramachandran allowed (%)	0.98	1.47
Ramachandran outliers (%)	0.00	0
Rotamer outliers (%)	0.00	0
Clashscore	6.45	10.14
Average B-factor	91.19	114.32
Macromolecules	79.80	99.70
Ligands	83.93	117.29
Solvent	70.69	99.53
Number of TLS groups	10	10

* Values in parentheses are for high-resolution shell.

Table S3.8. X-ray diffraction statistics for guest netropsin co-crystallized with CC1^{+10bp} Symmetric G-C rich PDB entry 8d86.

	CC1^{+10bp} Symmetric G-C rich Co-crystallized with Netropsin PDB code 8d86
Data collection	
Light source	Synchrotron
Wavelength (Å)	1.0
Resolution range (Å)	38.32 - 3.12 (3.232 - 3.12)*
Space group	I 2 2 2
Unit cell dimensions	
a, b, c (Å)	74.15 132.731 134.261
α, β, γ (°)	90 90 90
Total reflections	86552 (8880)
Unique reflections	12130 (1194)
Multiplicity	7.1 (7.4)
Completeness (%)	99.24 (99.17)
Mean I/sigma(I)	11.38 (1.84)
Wilson B-factor	74.24
R-merge	0.134 (1.18)
R-meas	0.1449 (1.269)
R-pim	0.05436 (0.4648)
CC1/2	0.998 (0.898)
CC *	0.999 (0.973)
Refinement	
Reflections used in refinement	12075 (1191)
Reflections used for R-free	1201 (121)
R-work	0.1932 (0.3835)
R-free	0.2332 (0.4910)
CC (work)	0.953 (0.894)
CC (free)	0.947 (0.749)
Number of non-hydrogen atoms	3141
Macromolecules	3088
Ligands	36
Solvent	17
Protein residues	214
RMS (bonds) (Å)	0.014
RMS (angles) (°)	1.57
Ramachandran favored (%)	97.09
Ramachandran allowed (%)	2.91
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.00
Clashscore	5.32
Average B-factor	94.34
Macromolecules	87.74
Ligands	97.07
Solvent	72.02
Number of TLS groups	8

* Values in parentheses are for high-resolution shell.

Table S3.9. X-ray diffraction statistics for RepE54 scaffold protein variant CC1^{+10bp} Symmetric G-C rich PDB entry 7uxy.

	RepE54 (L53G,Q54G,E55G) CC1^{+10bp} Symmetric G-C rich PDB code 7uxy
Data collection	
Light source	Synchrotron
Wavelength (Å)	1.0
Resolution range (Å)	34.21 - 3.15 (3.263 - 3.15)*
Space group	I 2 2 2
Unit cell dimensions	
a, b, c (Å)	74.11 129.426 136.826
α, β, γ (°)	90 90 90
Total reflections	81733 (8570)
Unique reflections	11671 (1147)
Multiplicity	7.0 (7.5)
Completeness (%)	98.05 (99.31)
Mean I/sigma(I)	15.68 (4.99)
Wilson B-factor	64.14
R-merge	0.127 (0.5253)
R-meas	0.1377 (0.566)
R-pim	0.05251 (0.2084)
CC1/2	0.996 (0.962)
CC *	0.999 (0.99)
Refinement	
Reflections used in refinement	11517 (1147)
Reflections used for R-free	1148 (116)
R-work	0.2472 (0.3172)
R-free	0.2646 (0.3192)
CC (work)	0.936 (0.910)
CC (free)	0.940 (0.899)
Number of non-hydrogen atoms	3133
Macromolecules	3092
Ligands	8
Solvent	33
Protein residues	213
RMS (bonds) (Å)	0.006
RMS (angles) (°)	0.95
Ramachandran favored (%)	99.02
Ramachandran allowed (%)	0.98
Ramachandran outliers (%)	0
Rotamer outliers (%)	0
Clashscore	3.57
Average B-factor	85.76
Macromolecules	83.32
Ligands	66.27
Solvent	59.77
Number of TLS groups	8

* Values in parentheses are for high-resolution shell.

Protocol S3.1. Protein sequences for cloning and overexpression in *E. coli*.

RepE54 Transcription Factor (CC1) protein sequence
RepE54 Transcription Factor (sequence below) was cloned into PSB3 Vector (Addgene plasmid # 82027) with an N-terminal His-tag.

MRGSHHHHHHGSMAETAVINHKKRKNSPRIVQSNDLTEAAYSLSRDQKRMLYLFVDQIRKSDGTLQEHDGI
CEIHVAKYAEIFGLTSAEASKDIRQALKSFAGKEVVFYRPEEDAGDEKGYESFPWFIKPAHSPSRGLYSVHINPY
LIPFFIGLQNRFTQFRLSETKEITNPYAMRLYESLCQYRKPDGSGIVSLKIDWIIERYQLPQSYQRMPDFRRRFL
QVCVNEINSRTPMRLSYIEKKKGRQTTHIVFSFRDITSMTTG**

RepE54^{L53G,Q54G,E55G} Transcription Factor Variant (CC1) protein sequence
RepE54 Transcription Factor (sequence below) was cloned into pETDuet-1 vector (Novagen) with an N-terminal His-tag. Mutated residues are highlighted in red.

MRGSHHHHHHGSMAETAVINHKKRKNSPRIVQSNDLTEAAYSLSRDQKRMLYLFVDQIRKSDGTGGGHDG
ICEIHVAKYAEIFGLTSAEASKDIRQALKSFAGKEVVFYRPEEDAGDEKGYESFPWFIKPAHSPSRGLYSVHINPY
LIPFFIGLQNRFTQFRLSETKEITNPYAMRLYESLCQYRKPDGSGIVSLKIDWIIERYQLPQSYQRMPDFRRRFL
QVCVNEINSRTPMRLSYIEKKKGRQTTHIVFSFRDITSMTTG**

Engrailed Homeodomain fused to eGFP (EnH-eGFP) protein sequence
The EnH-eGFP fusion protein was cloned pETDuet-1 vector (Novagen). The protein sequence contained the EnH sequence, followed by a 7 amino acid linker before the eGFP and a C-terminal His-tag.

MEKRPRTAFSSEQLARLKREFNENRYLTERRRQQLSSELGLNEAQIKIWFQNKRAKIKKSTSQFYLNEMVSKG
EELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDH
MKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHN
VYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVL
LEFVTAAGITLGMDELYKHHHHHH**

Appendix III. SUPPORTING INFORMATION FOR CHAPTER 4

Stabilizing DNA–Protein Co-Crystals via in crystallo Chemical Ligation of the DNA.

Table of Contents

Protocol S4.3. Spatial biased random ligation model.

Protocol S4.4. Crystal measurements.

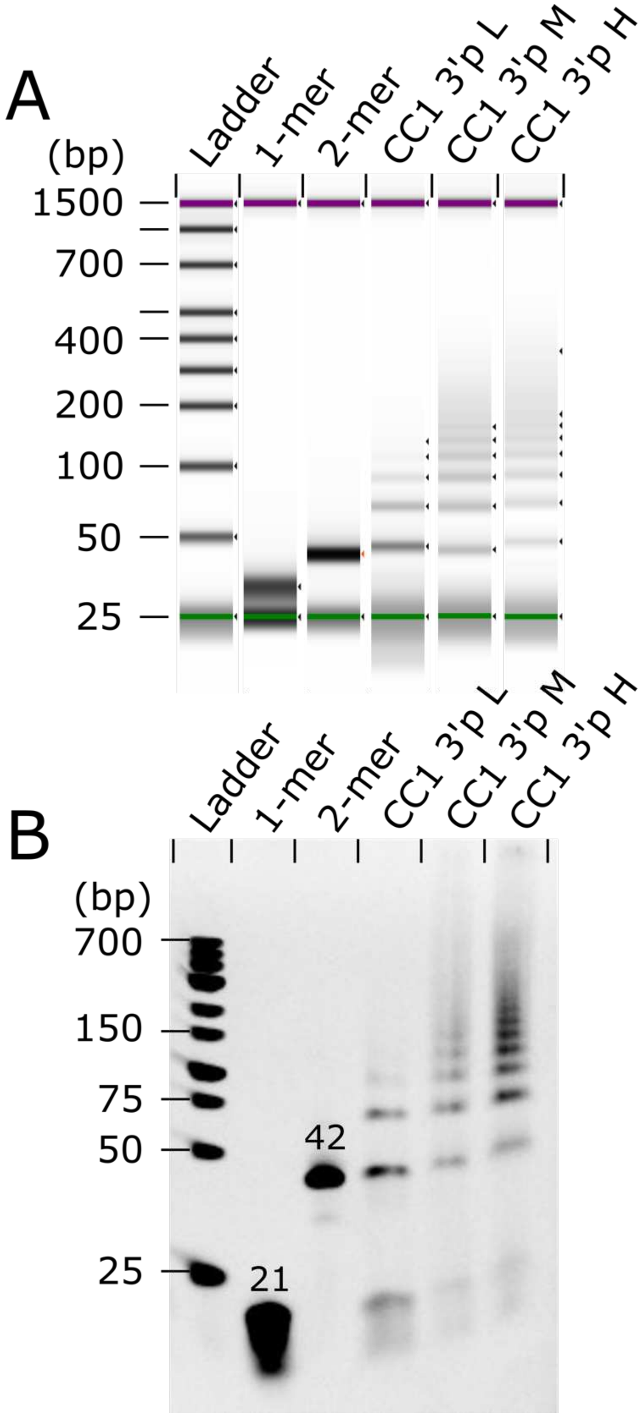


Figure S4.1. TapeStation analysis and matching gel electrophoresis. (A) The TapeStation analysis for crosslinked crystals of CC1 3' phosphate shows the ligation end product distribution. The crosslinked crystals are represented with 'L', 'M', and 'H'. The crosslink values correspond with Figure 4 in the text. (B) The corresponding TBE-urea gel shown is from the same samples as the TapeStation samples.

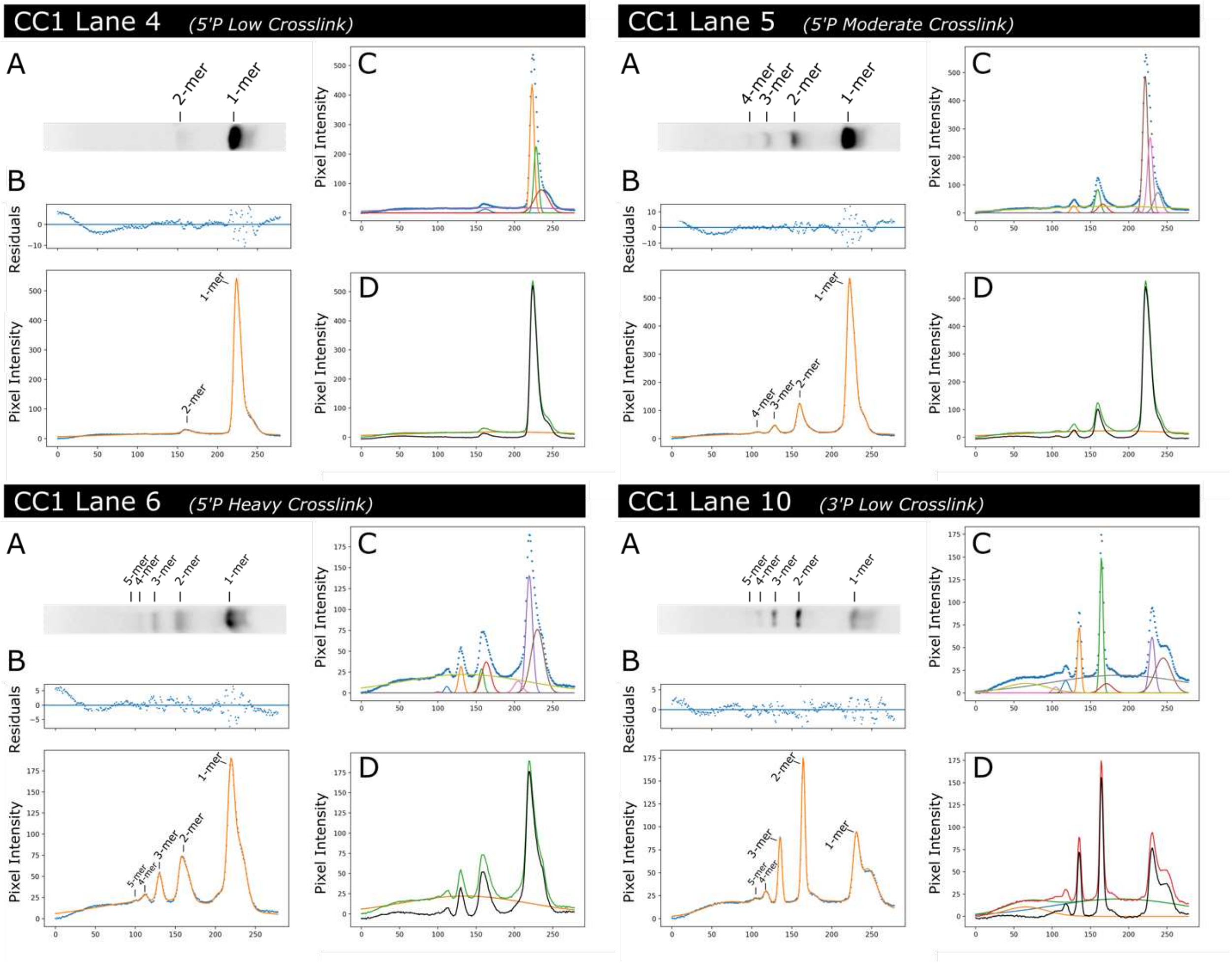


Figure S4.2. Densitometry results and annotation (corresponds to main text Figure 4.4). For each lane, we conducted a multi-step densitometry analysis. (A) The lane was manually excised from the gel using ImageJ software. (B) The final fit (orange line) closely follows the raw intensity data (blue dots) which is the average intensity value across the lane (perpendicular to $\vec{E}$ direction). The quality of the fit is also evident in the small residuals (above). (C) The component gaussian functions inside the fit are shown with different colors. (D) If we subtract the diffuse background (orange) from the full fit (green) the background curve is shown in black.

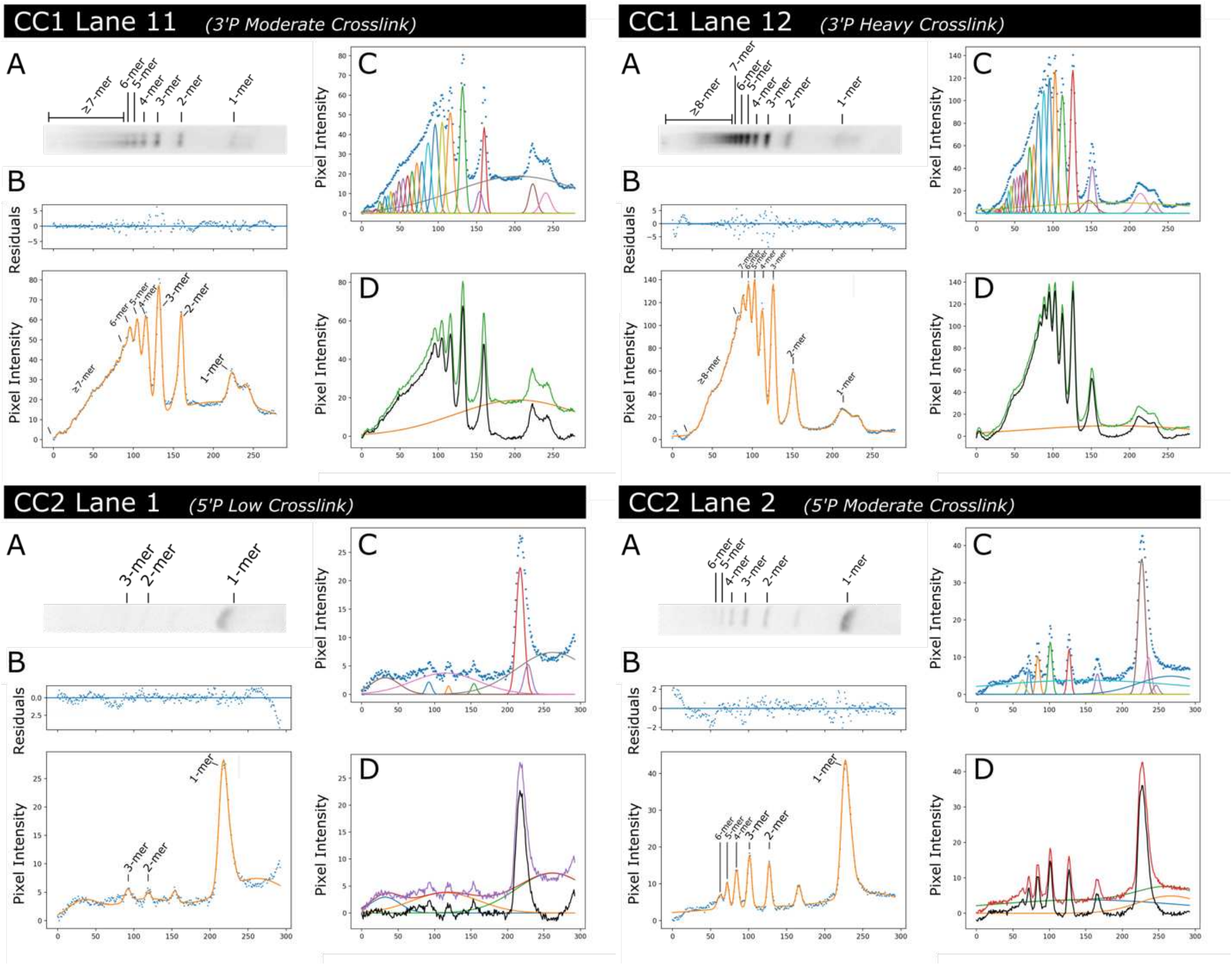


Figure S4.2. continued

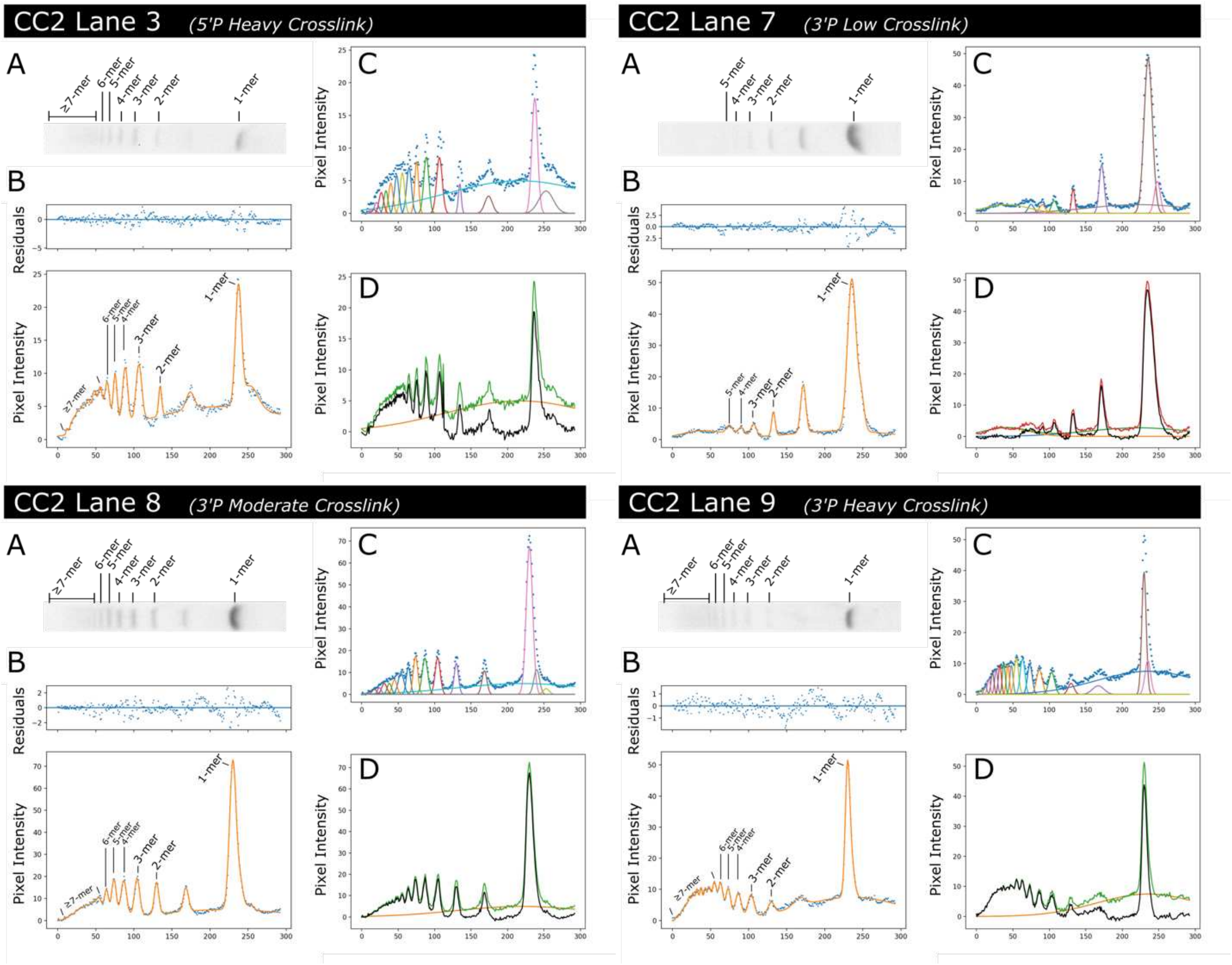


Figure S4.2. continued

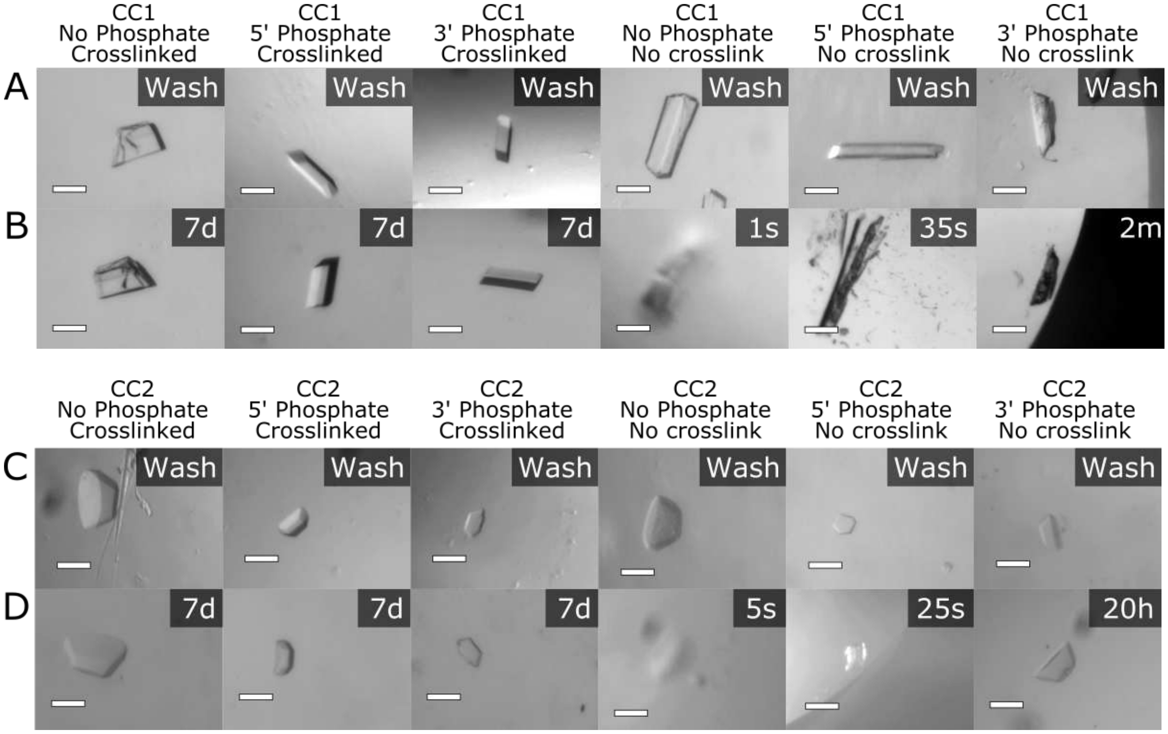


Figure S4.3. Co-crystal stability test - water. The crystals were crosslinked at 15 mg/mL EDC for 20 hours and quenched with tris base pH 8.2 for 30 minutes prior to transfer to the wash solution. All scale bars are 100 µm. (A) CC1 crystals in wash solution containing 50 mM NaCl, 14% PEG 400, and 200 mM MES buffer pH 6.0. The concentrations of the wash solution matched the initial crystal growth solutions but we replaced $MgCl_2$ with NaCl and tris HCl pH 8.0 with MES buffer pH 6.0. (B) CC1 crystals after transfer to deionized water. The crosslinked crystals (left three panels) remained intact for at least 7 days. The non-crosslinked crystals (right three panels) dissolved or converted to an aggregate at various immediate timepoints. (C) CC2 crystals in wash solution containing 160 mM lithium sulfate, 5% PEG 400, 13.3% PEG 3350 and 80 mM MES buffer pH 6.0. The concentrations of the wash solution matched the initial crystal growth solutions but we replaced ammonium sulfate with lithium sulfate and HEPES buffer pH 7.1 with MES buffer pH 6.0. (D) CC2 crystals after transfer to deionized water. The crosslinked crystals (left three panels) remained intact for at least 7 days. The non-crosslinked crystals (right three panels) dissolved or converted to an aggregate at various immediate timepoints. Surprisingly, non-crosslinked CC2 3’ phosphate crystals did not dissolve after 20 hours in water.

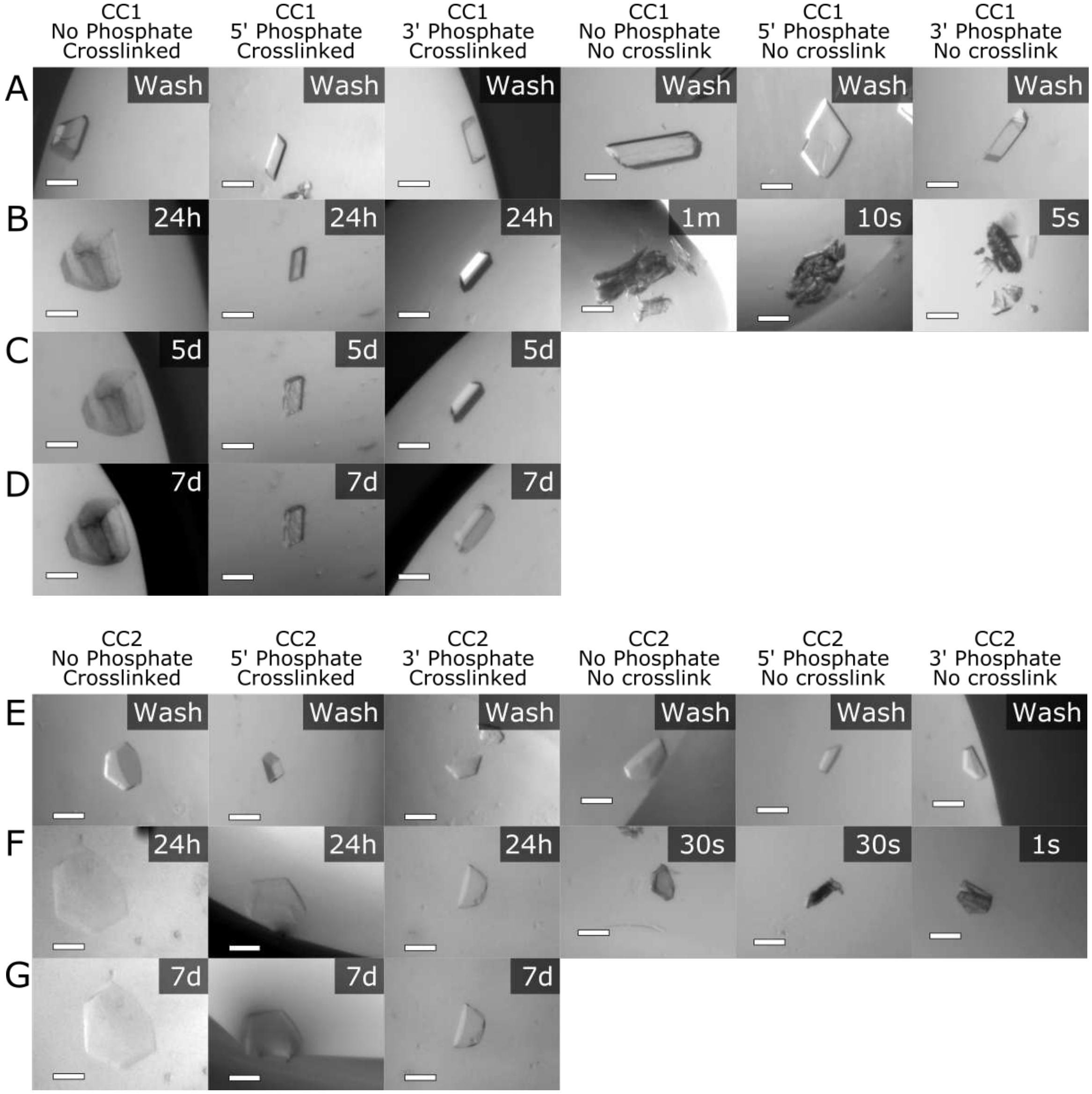


Figure S4.4. Co-crystal stability test – very low pH 2.0 to mimic stomach acid. The crystals were crosslinked at 15 mg/mL EDC for 20 hours and quenched with tris base pH 8.2 for 30 minutes prior to transfer to the wash solution. All scale bars are 100 µm. Percent volume increases were measured using Protocol S4. (A) CC1 crystals in wash solution containing 50 mM NaCl, 14% PEG 400, and 200 mM MES buffer pH 6.0. The concentrations of the wash solution matched the initial crystal growth solutions, but we replaced $MgCl_2$ with NaCl and tris HCl pH 8.0 with MES buffer pH 6.0. (B) CC1 crystals after transitioning to a pH 2.0 stomach acid mimic (0.01M HCl). The crosslinked crystals (left three panels) showed varied results after 24 hours. The crosslinked CC1 no phosphates expanded ~430% ± 70% in volume after 24 hours from the original washed crystal. The crystals with terminal phosphates did not noticeably change in macro-structure. The non-crosslinked crystals (right three panels) dissolved or converted to an aggregate at various immediate timepoints. (C) 5 days in pH 2.0 buffer the crosslinked crystals showed varied results. The CC1 no phosphate and 5' phosphate expanded by

~440% ± 70% and ~140% ± 40% in volume after 5 days from the original washed crystal. The CC1 3’ phosphate did not change in macro-structure. (D) After 7 days in pH 2.0 buffer, each of the crosslinked crystals expanded in volume from the original washed crystal: CC1 no phosphate by ~460% ± 70%, CC1 5’ phosphate by ~240% ± 60%, and CC1 3’ phosphate by ~62% ± 0%. (E) CC2 crystals in wash solution containing 160 mM lithium sulfate, 5% PEG 400, 13.3% PEG 3350 and 80 mM MES buffer pH 6.0. The concentrations of the wash solution matched the initial crystal growth solutions, but we replaced ammonium sulfate with lithium sulfate and HEPES buffer pH 7.1 with MES buffer pH 6.0. (F) CC2 crystals after transitioning to a pH 2.0 stomach acid mimic (0.01M HCl). The crosslinked crystals (left three panels) expanded in volume after 24 hours from the original washed crystal: CC2 no phosphate by ~640% ± 20%, CC2 5’ phosphate by ~1390% ± 60%, and CC2 3’ phosphate by ~440% ± 40%. The non-crosslinked crystals (right three panels) appeared to degrade into a protein aggregate at various immediate timepoints. (G) In contrast, after 7 days, CC2 crosslinked crystals still had an expanded macro-structure but they did not dissolve.

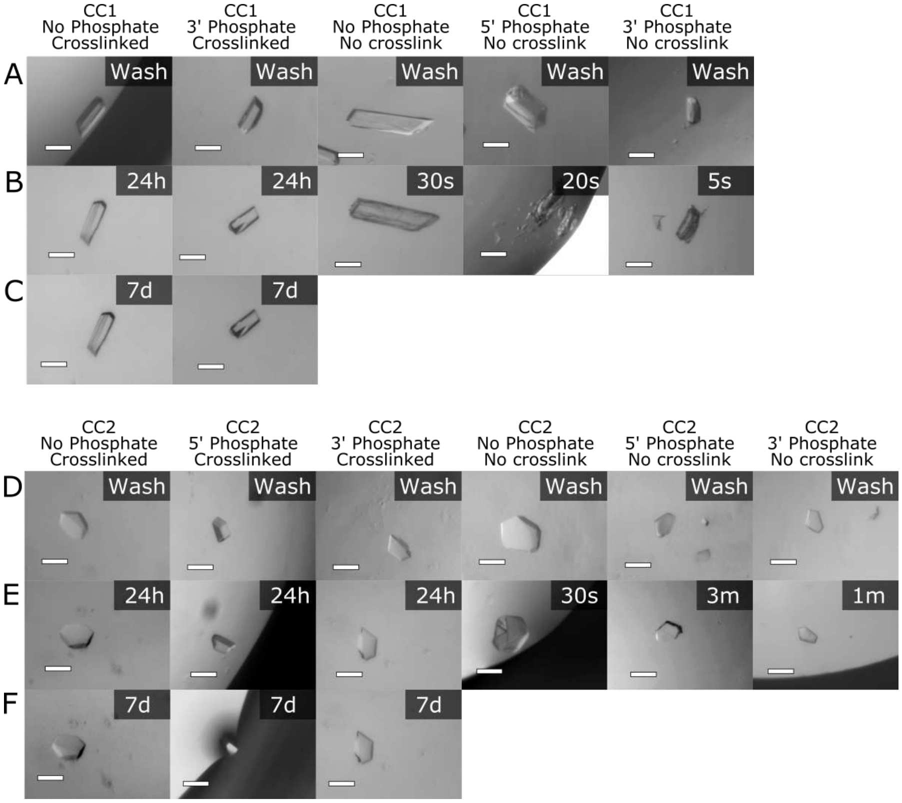


Figure S4.5. Co-crystal stability test – moderately low pH 4.5 to mimic lysosomal fluid. The crystals were crosslinked at 15 mg/mL EDC for 20 hours and quenched with tris base pH 8.2 for 30 minutes prior to transfer to the wash solution. All scale bars are 100 µm. (A) CC1 crystals in wash solution containing 50 mM NaCl, 14% PEG 400, and 200 mM MES buffer pH 6.0. The concentrations of the wash solution matched the initial crystal growth solutions but we replaced $MgCl_2$ with NaCl and tris HCl pH 8.0 with MES buffer pH 6.0. (B) CC1 crystals after transitioning to a pH 4.5 lysosomal fluid mimic (46 mM sodium citrate, 54.1 mM citric acid). The crosslinked crystals (left two panels) were void of macrostructural changes after 24 hours whereas the non-crosslinked crystals (right three panels) either visibly degraded or dissolved rapidly. (C) After 7 days in pH 2.0 buffer, each of the crosslinked crystals lacked obvious changes. (D) CC2 crystals in wash solution containing 160 mM lithium sulfate, 5% PEG 400, 13.3% PEG 3350 and 80 mM MES buffer pH 6.0. The concentrations of the wash solution matched the initial crystal growth solutions but we replaced ammonium sulfate with lithium sulfate and HEPES buffer pH 7.1 with MES buffer pH 6.0. (E) CC2 crystals after transitioning to a pH 4.5 lysosomal fluid mimic (46 mM sodium citrate, 54.1 mM citric acid). The crosslinked crystals lacked macrostructural changes after 24 hours (left three panels). Surprisingly, the non-crosslinked crystals remained relatively unperturbed after a short time in the harsh condition (right three panels). They dissolved after ~24 hours (not shown). (G) After 7 days, CC2 crosslinked crystals remained without visible damage.

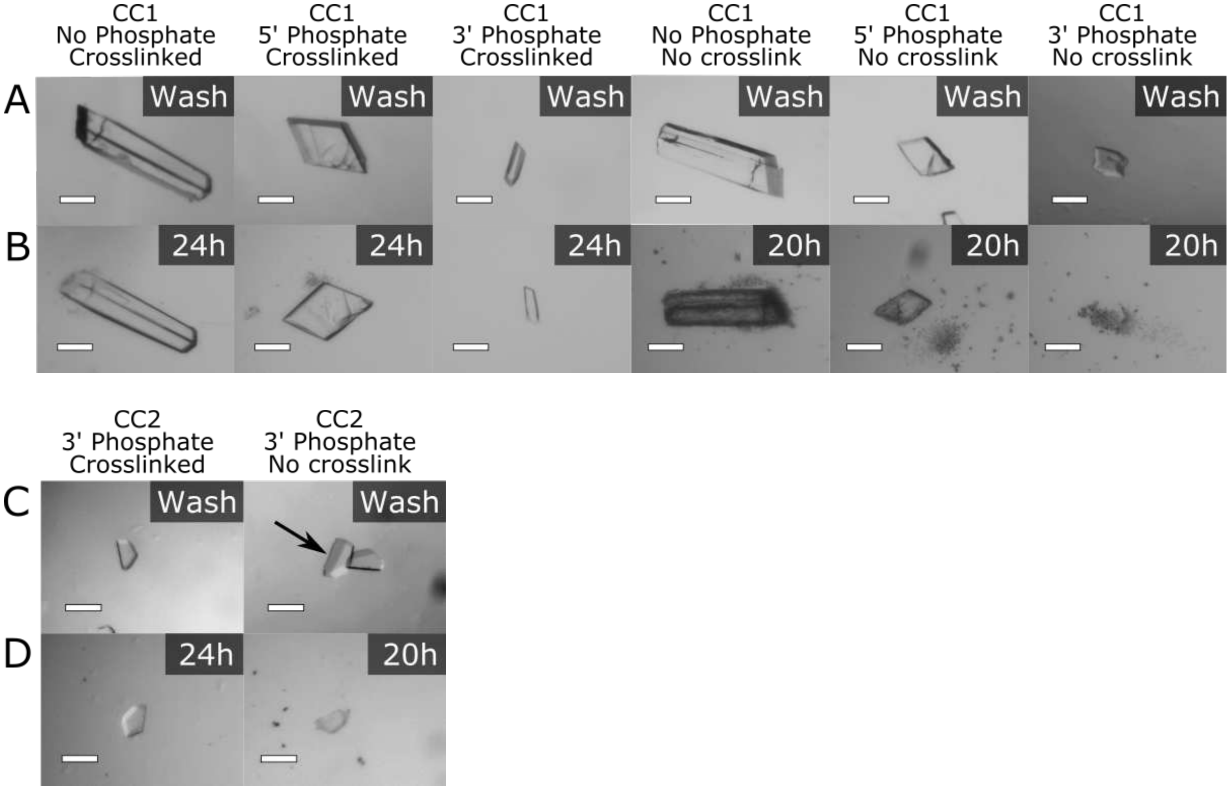


Figure S4.6. Co-crystal stability test – blood serum. Co-crystals were crosslinked at 15 mg/mL EDC for 20 hours and quenched with tris base pH 8.2 for 30 minutes prior to transfer to the wash solution. All scale bars are 100 µm. (A) CC1 crystals in wash solution containing 80 mM NaCl, 5% PEG 400, and 160 mM MES buffer pH 6.0. The concentrations of the wash solution matched the initial crystal growth solutions, but we replaced $MgCl_2$ with NaCl and tris HCl pH 8.0 with MES buffer pH 6.0. (B) CC1 crystals after transfer to blood serum (HyClone, Bovine Calf Serum Product #: SH30073.02). The crosslinked crystals (left three panels) remained intact for at least 24 hours. The non-crosslinked crystals (right three panels) dissolved or degraded after 20 hours in the serum. (C) CC2 crystals in wash solution containing 240 mM lithium sulfate, 5% PEG 400, 22% PEG 3350 and 120 mM MES buffer pH 6.0. The concentrations of the wash solution matched the initial crystal growth solutions, but we replaced ammonium sulfate with lithium sulfate and HEPES buffer pH 7.1 with MES buffer pH 6.0. (D) CC2 crystals after transitioning to blood serum. The crosslinked crystal (left panel) remained intact for at least 24 hours in blood serum. The non-crosslinked crystals (right panel) lost macrostructure after 20 hours in blood serum.

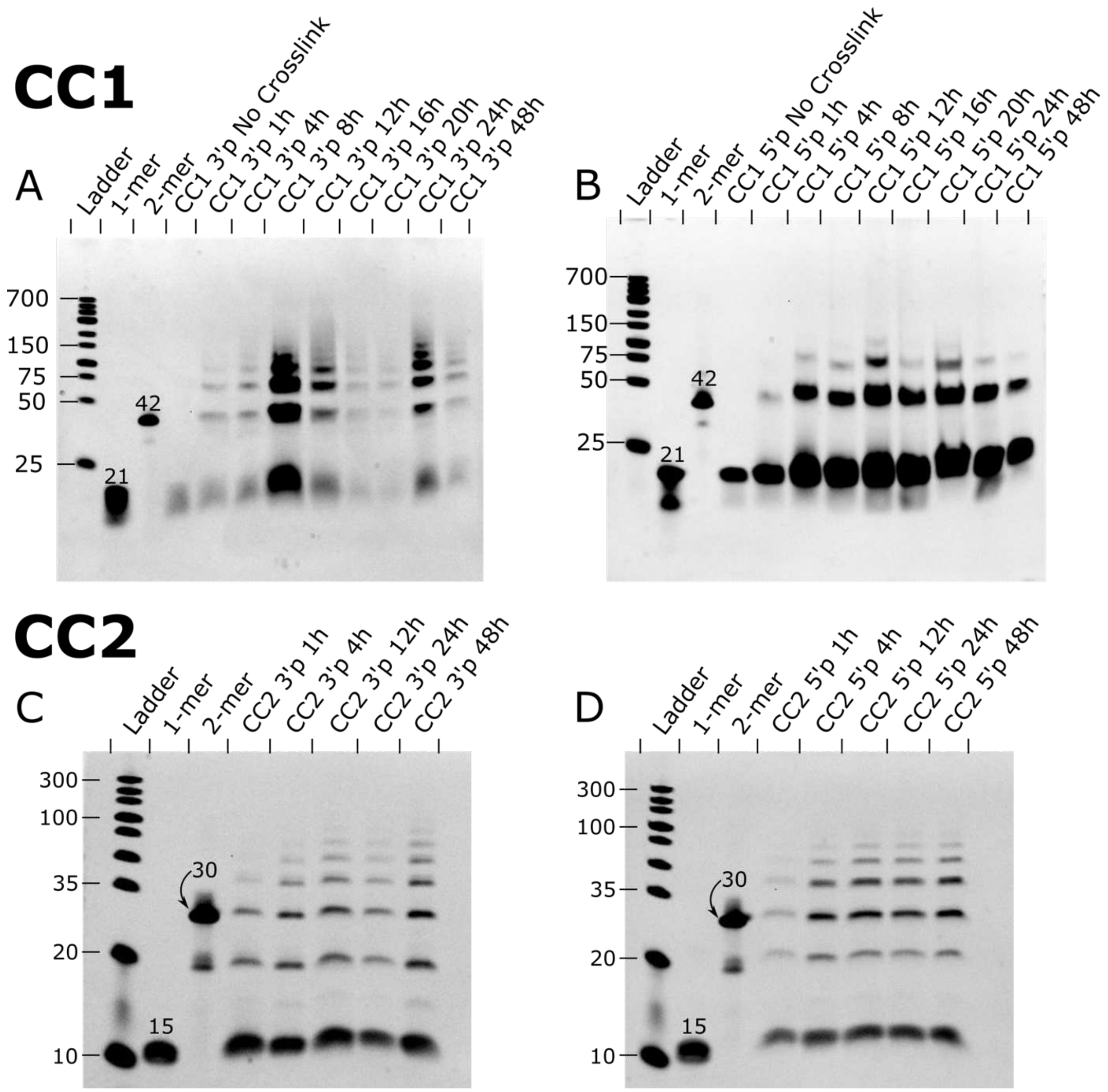


Figure S4.7. Gel electrophoresis of varied EDC crosslink time. All crosslink lanes shown in this figure were crystals dissolved after incubation in 15 mg/mL EDC solution for the time point indicated on the gel figure (1 hour to 48 hours). Each crystal was dissolved and analyzed on a TBE-urea gel. (A) CC1 co-crystal samples with 3' phosphates (B) CC1 co-crystal samples with 5' phosphates (C) CC2 co-crystal samples with 3' phosphates and (D) CC2 co-crystal samples with 5' phosphates.

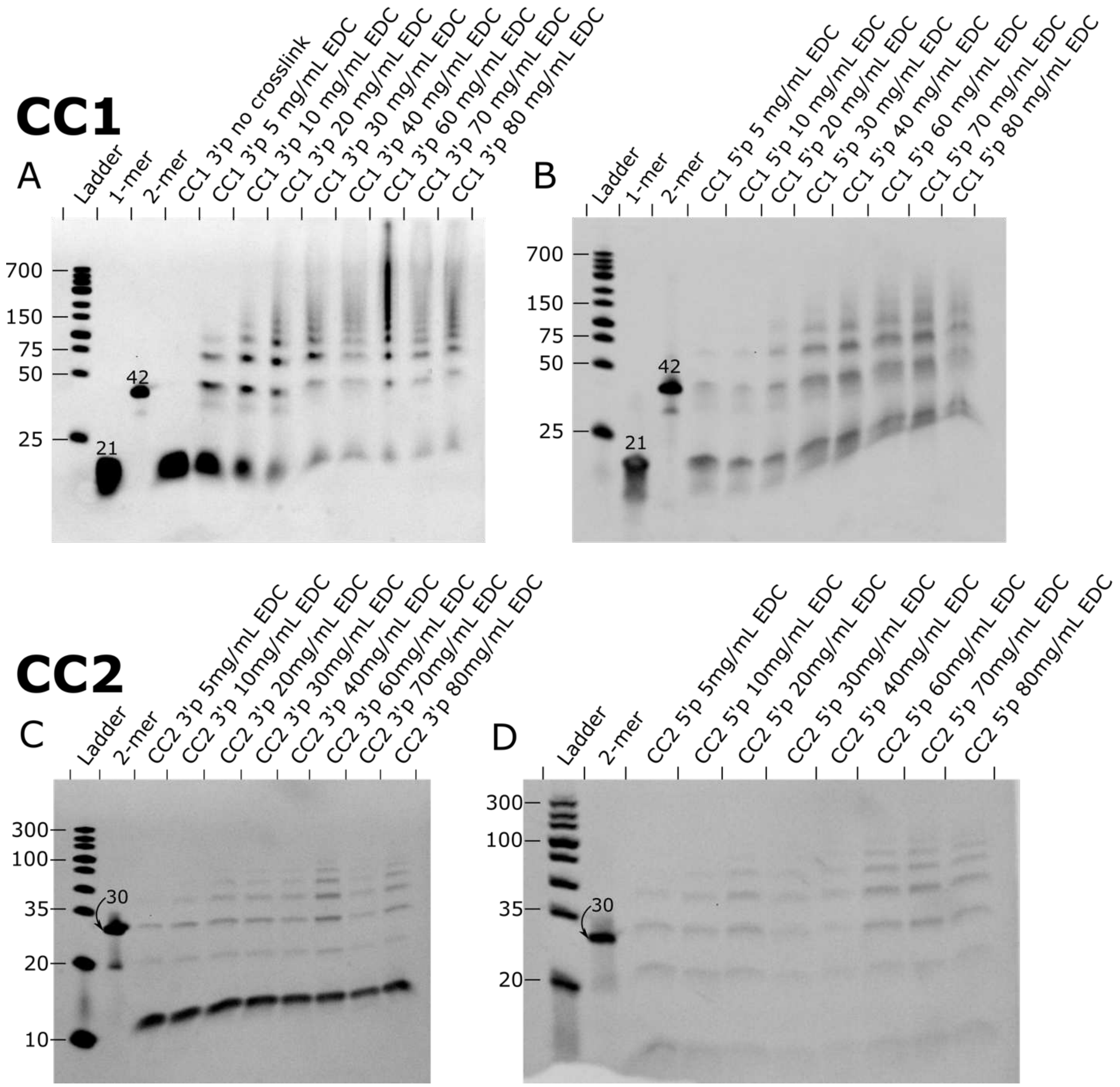


Figure S4.8. Gel electrophoresis of varied EDC crosslink concentration. All crosslink lanes shown in this figure were crystals dissolved after incubation for 12 hours in an EDC solution at the concentration indicated on the gel figure (5 mg/mL EDC to 80 mg/mL). Each crystal sample was dissolved and analyzed on a TBE-urea gel. (A) CC1 co-crystal samples with 3' phosphates (B) CC1 co-crystal samples with 5' phosphates (C) CC2 co-crystal samples with 3' phosphates and (D) CC2 co-crystal samples with 5' phosphates.

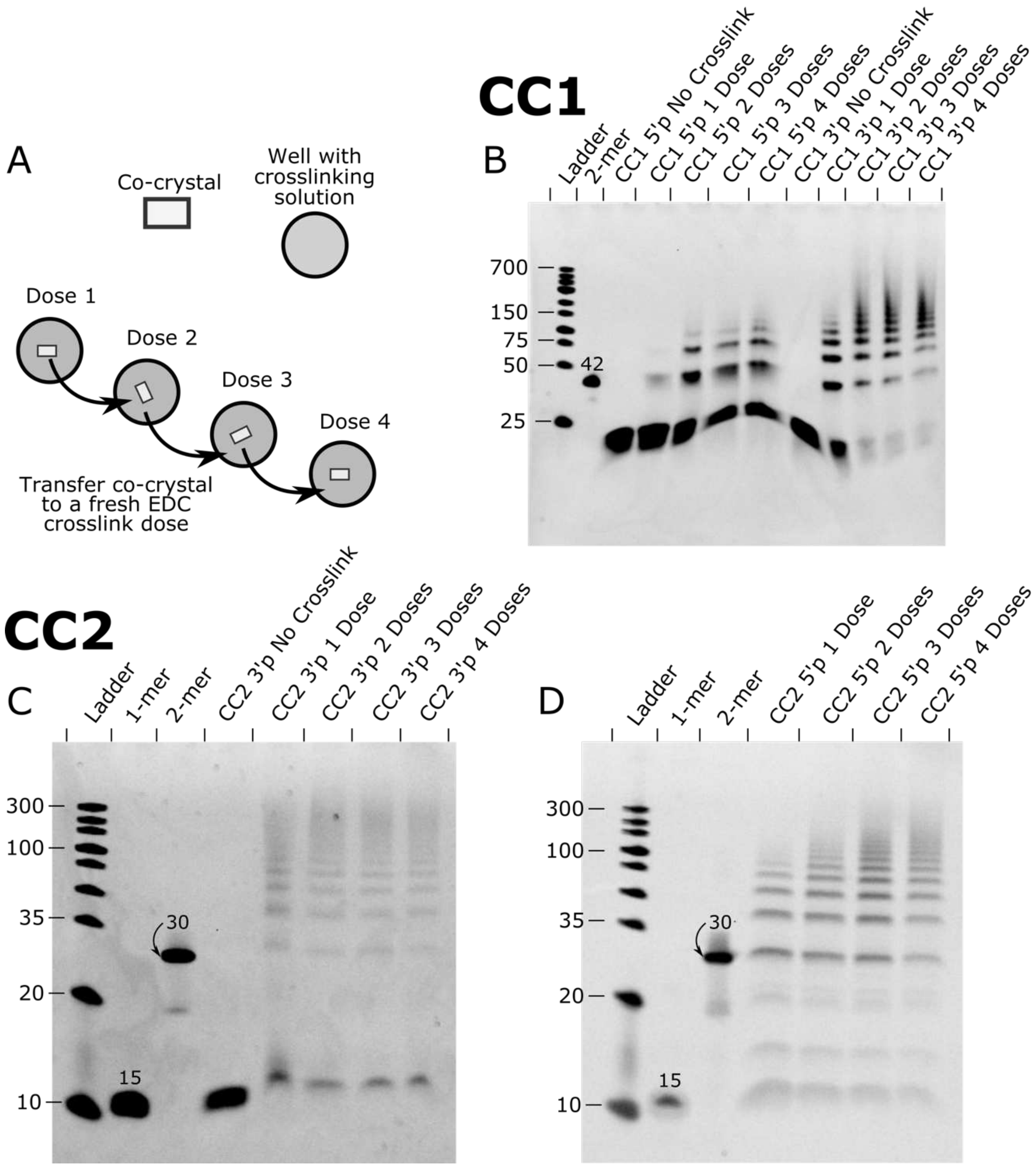


Figure S4.9. Schematic and gel electrophoresis of varied EDC crosslink dose. All crosslink lanes shown in this figure were crystals dissolved after incubation in a set number of doses (1 to 4 doses) of 30 mg/mL EDC solution for 12 hours. Each crystal was dissolved and run on a TBE-urea gel. (A) A schematic of the crosslinking dose experiment. The co-crystal is transferred to a fresh 30 mg/mL EDC dose every 12 hours. (B) CC1 co-crystal samples with 5' phosphate and with 3' phosphates (C) CC2 co-crystal samples with 3' phosphates and (D) CC2 co-crystal samples with 5' phosphates.

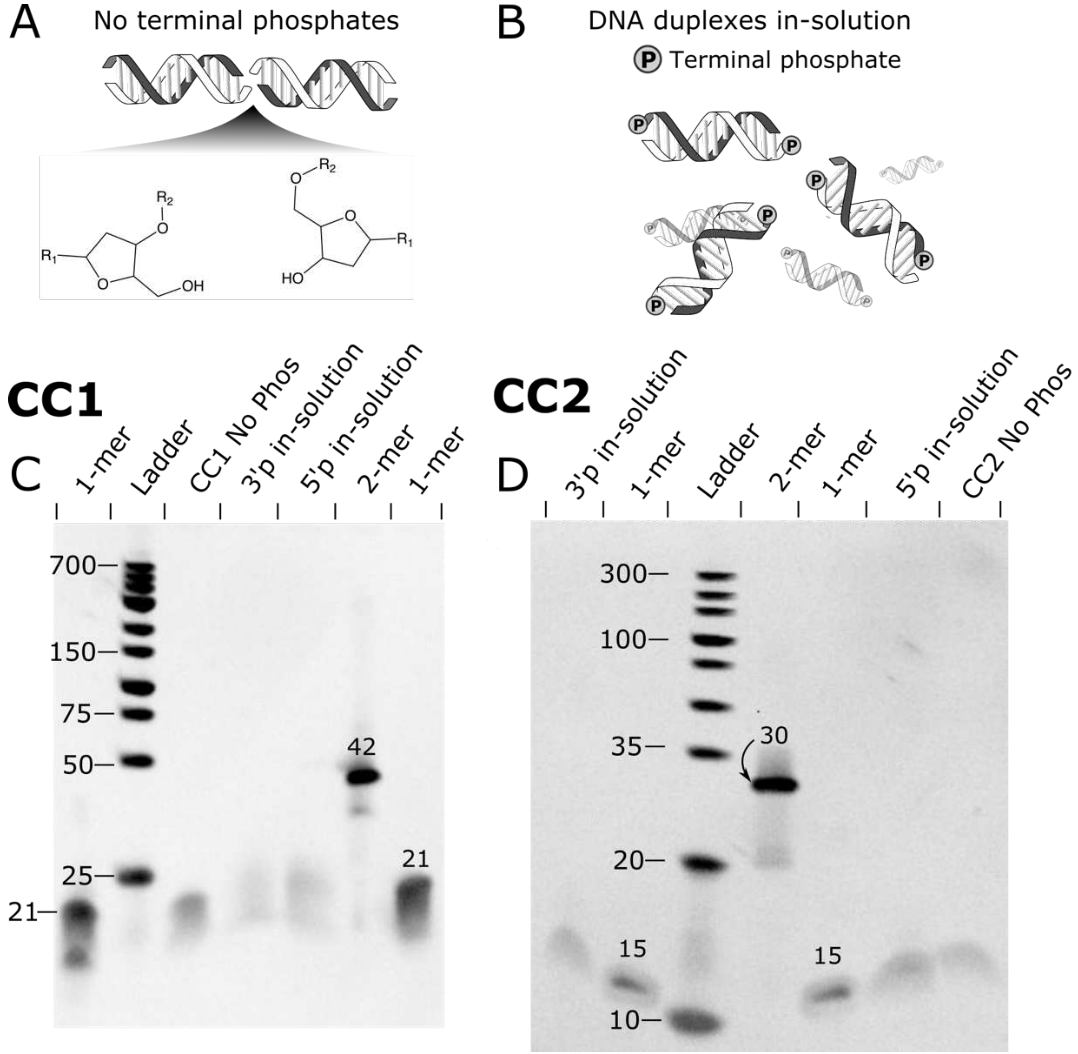


Figure S4.10. Schematic and gel electrophoresis of the controls – crystals with no terminal phosphates and duplexes with terminal phosphates in-solution. All crosslink trials shown in this figure were subjected to a 30 mg/mL EDC solution for 20 hours. Each sample was dissolved and run on a TBE-urea gel. (A) A schematic of the co-crystal DNA-DNA junction with no terminal phosphates. The terminal hydroxyls cannot react with EDC to ligate the flanking DNA duplexes. R_1 is the nucleobase and R_2 is the phosphate backbone. (B) We hypothesize DNA duplexes in-solution are randomly ordered, such that the DNA ends do not frequently stack as in the co-crystals. Therefore, the blunt ended duplexes cannot be ligated with EDC crosslinking as shown in the gels C and D. (C) CC1 co-crystal with no terminal phosphate and CC1 DNA duplexes 2.5 ng/μL (192nM) in-solution with terminal 3' phosphates or 5' phosphates. (D) CC2 co-crystal with no terminal phosphates and CC2 DNA duplexes 2.5 ng/μL (192nM) in-solution with terminal 3' phosphates or 5' phosphates.

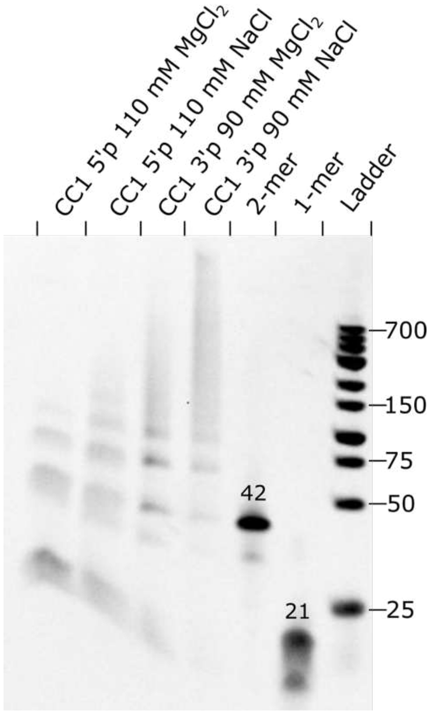


Figure S4.11. Magnesium chloride's effect on the EDC crosslinking of CC1 crystals. All crosslink trials shown in this figure were subjected to a 30 mg/mL EDC solution for 20 hours. Each CC1 sample was dissolved and run on a TBE-urea gel. The magnesium chloride was replaced with sodium chloride and the resulting crosslinks were compared.

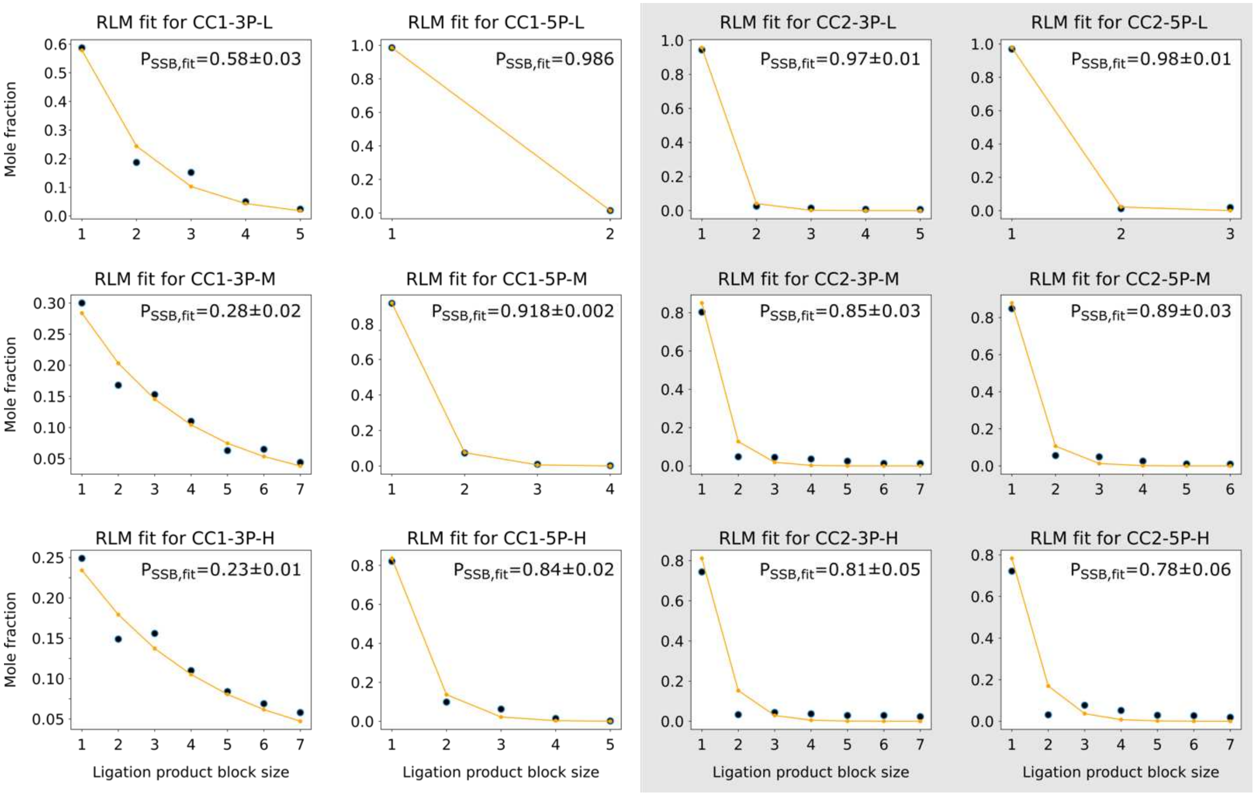


Figure S4.12. Random ligation model fit to mole fractions derived from gel densitometry

Experimental gel densitometry data (black discs), with mole fractions corrected for DNA length (Table 2), were fit to the geometric decay predicted by the random ligation model (RLM, orange line, Protocol S2) using non-linear least squares via the scipy.optimize.curve_fit function. The CC1 ligation product distribution for 3’ phosphates (column 1) and 5’ phosphates (column 2) was fit well with the RLM, albeit with a small systematic deficit of 2-mer and surfeit of 3-mer. The CC1 fit P_{SSB} values (inset, along with the standard error) also closely matched the P_{SSB} values calculated from the mole fractions. In contrast, the RLM provided an inferior fit for CC2 data (higher standard errors), overestimating the 1-mer and 2-mer fractions and underestimating higher-order product mole fractions. While CC2 fit P_{SSB} values only somewhat exceed the measured 1-mer mole fractions, underlined entries significantly deviate from P_{SSB} values derived from main text Equation 1 (reprised below). The biased ligation model (Protocol S3), or asymmetry in the ligation yield for one of the two nick sites, could address this CC2 discrepancy.

Crosslinking Protocol	low	medium	high	low	medium	high
Parent Crystal	**CC1-3′P**	**CC1-3′P**	**CC1-3′P**	**CC1-5′P**	**CC1-5′P**	**CC1-5′P**
P_{SSB}	0.58	0.28	0.25	0.99	0.91	0.78
Parent Crystal	**CC2-3′P**	**CC2-3′P**	**CC2-3′P***	**CC2-5′P**	**CC2-5′P**	**CC2-5′P**
P_{SSB}	0.90	**0.61**	**0.45**	0.95	**0.75**	**0.48**

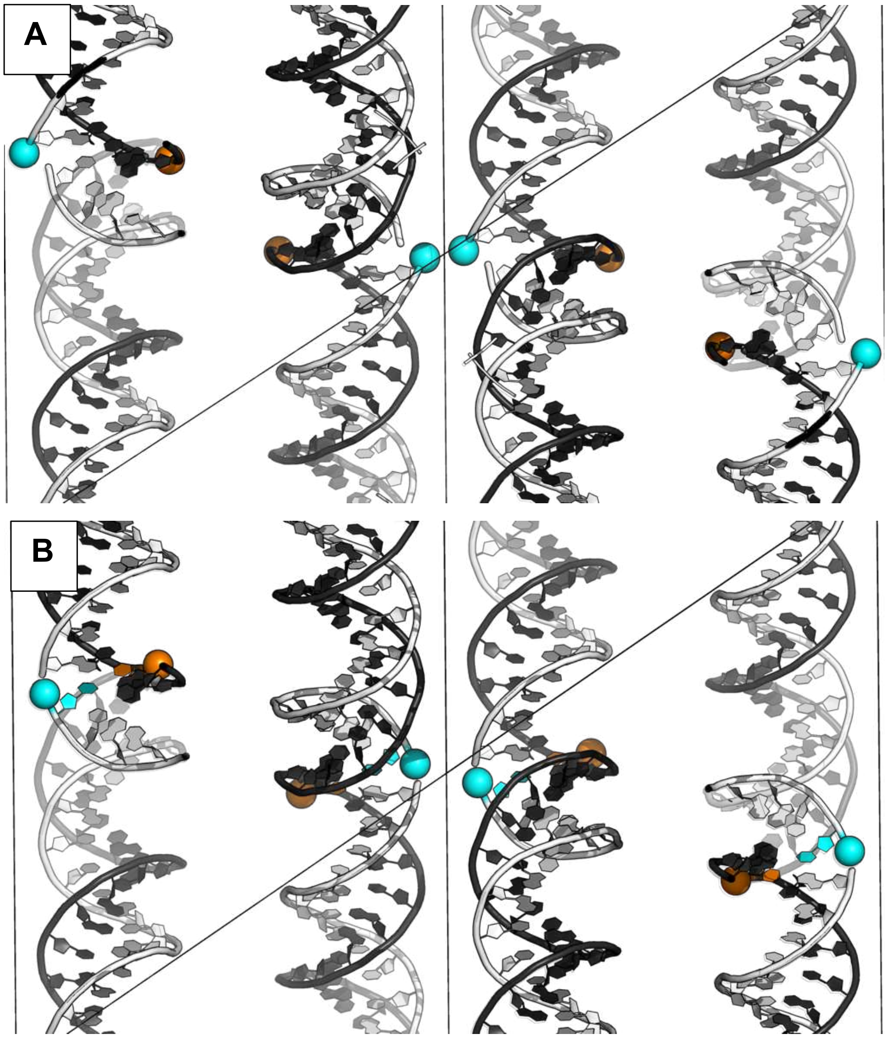


Figure S4.13. Terminal phosphate positions and symmetry
(A) CC1-3’P in our new PDB entry 7sdp has 3’ terminal phosphates on chain A (orange spheres) and chain B (cyan spheres). Due to crystallographic symmetry, the chain B 3’phosphates are quite close (5.65Å) which might contribute to increased ligation propensity for the chain B nick compared to the chain A nick. Protein is hidden for clarity. **(B)** CC1-5’P in our new PDB entry 7sgc has 5’ terminal phosphates on chain A (orange spheres) and chain B (cyan spheres). Notably, the chain B 5’phosphates are farther apart (9.29Å) than the 3’ phosphate case above, which may decrease the electrostatic repulsion driving force for ligation. Protein is hidden for clarity.

Table S4.1. DNA oligonucleotide sequences used in this study.

Duplex ID	Duplex Description	Duplex Sequence
1.0	CC1 Original Duplex (PDB codes 7rva and 1rep)	5′ — CCTGTGACAAATTGCCCTCAGT — 3′ 3′ — TGGACACTGTTTAACGGGAGTC — 5′
1.1	CC1 No Terminal Phosphates	5′ — CCTGTGACAAATTGCCCTCAG — 3′ 3′ — GGACACTGTTTAACGGGAGTC — 5′
1.2	CC1 5’ Terminal Phosphates (PDB code 7sgc)	5′ — pCCTGTGACAAATTGCCCTCAG — 3′ 3′ — GGACACTGTTTAACGGGAGTCp — 5′
1.3	CC1 3’ Terminal Phosphates (PDB code 7sdp)	5′ — CCTGTGACAAATTGCCCTCAGp — 3′ 3′ — pGGACACTGTTTAACGGGAGTC — 5
1.4	CC1 No Terminal Phosphates 2-mer	5′—CCTGTGACAAATTGCCCTCAGCCTGTGACAAATTGCCCTCAG—3′ 3′—GGACACTGTTTAACGGGAGTCGGACACTGTTTAACGGGAGTC—5′
2.1	CC2 Original Duplex and No Terminal Phosphates (PDB code 4yo2)	5′ — TTTTCCCGCCAAAAA — 3′ 3′ — AAAAGGGCGGTTTTT — 5′
2.2	CC2 5’ Terminal Phosphates	5′ — pTTTTCCCGCCAAAAA — 3′ 3′ — AAAAGGGCGGTTTTTp — 5′
2.3	CC2 3’ Terminal Phosphates	5′ — TTTTCCCGCCAAAAAp — 3′ 3′ — pAAAAGGGCGGTTTTT — 5′
2.4	CC2 No Terminal Phosphates 2-mer	5′—TTTTCCCGCCAAAAATTTTCCCGCCAAAAA-3′ 3′—AAAAGGGCGGTTTTTAAAAGGGCGGTTTTT—5′

Table S4.2. Ligation percentages from gel densitometry (unweighted)
The data below corresponds with Table 2 in the text. Shown here is the *unweighted* distribution where the length of the DNA and resulting dye binding quantity is not incorporated in the percentages.

Parent Crystal	**CC1-3'P**	**CC1-3'P**	**CC1-3'P**	**CC1-5'P**	**CC1-5'P**	**CC1-5'P**
Crosslinking Protocol	low	medium	high	low	medium	high
DNA block size	[%]	[%]	[%]	[%]	[%]	[%]
1	49.5	8.3	6.2	97.1	83.5	74.7
2	31.5	9.3	7.4	2.9	13.3	18.0
3	12.8	12.6	11.7		2.6	5.8
4	4.2	12.1	10.9		0.6	1.3
5	2.0	8.8	10.5			0.2
6		10.7	10.4			
7		8.5	10.1			
8 and above		29.7	32.7			

Parent Crystal	**CC2-3'P**	**CC2-3'P**	**CC2-3'P**	**CC2-5'P**	**CC2-5'P**	**CC2-5'P**
Crosslinking Protocol	low	medium	high	low	medium	high
DNA block size	[%]	[%]	[%]	[%]	[%]	[%]
1	85.2	49.1	34.3	92.2	64.0	34.8
2	4.7	5.9	3.1	2.3	8.4	3.0
3	4.0	8.3	6.1	5.4	11.1	11.1
4	2.9	8.7	6.8		7.8	10.1
5	3.1	7.5	6.7		4.2	7.0
6		4.9	7.9		4.4	7.8
7		5.4	7.6			6.3
8 and above		10.1	27.4			19.8

Table S4.3. Full version of densitometry output Table 4.2.
This version of Table 4.2 includes the small mole fractions for higher-order peaks

Parent Crystal	**CC1-3′P**	**CC1-3′P**	**CC1-3′P**	**CC1-5′P**	**CC1-5′P**	**CC1-5′P**
Crosslinking Protocol	low	medium	high	low	medium	high
DNA block size	[%]	[%]	[%]	[%]	[%]	[%]
1	58.7	30.0	24.9	98.6	91.6	82.1
2	18.7	16.8	14.9	1.4	7.3	9.9
3	15.2	15.3	15.6		1.0	6.3
4	5	11.0	11.0		0.2	1.5
5	2.4	6.3	8.4			0.2
6		6.5	6.9			
7		4.4	5.8			
8		2.1	4.2			
9		1.8	2.1			
10		1.4	2.0			
11		1.2	0.9			
12		0.7	0.8			
13		0.5	0.6			
14		0.4	0.6			
15		0.4	0.5			
16		0.4	0.3			
17		0.3	0.2			
18		0.1	0.2			
19		0.03	0.1			
20		0.1	0.1			
21		0.07	0.02			
22		0.07	2E-9			
23		0.009	0.02			
24		0.004	0.02			
25		0.02	0.01			
26		0.05	0.009			
27			0.001			

Parent Crystal	**CC1-3′P**	**CC1-3′P**	**CC1-3′P**	**CC1-5′P**	**CC1-5′P**	**CC1-5′P**
Crosslinking Protocol	low	medium	high	low	medium	high
$P_{SSB}{}^{*}$	0.58±0.01	0.28±0.01	0.25±0.01	0.99±0.01	0.91±0.02	0.78±0.02
$P_{LIG} = 1 - P_{SSB}$	0.42±0.01	0.72±0.01	0.75±0.01	0.01±0.01	0.09±0.02	0.22±0.02
$P_{DSB} = (P_{SSB})^2$	0.33±0.01	0.08±0.01	0.06±0.005	0.97±0.02	0.83±0.04	0.61±0.03
$P_{DLIG} = (P_{LIG})^2$	0.18±0.01	0.52±0.01	0.56±0.02	$(2\pm4)\cdot10^{-4}$	$(1\pm0.4)\cdot10^{-2}$	0.05±0.01

*Calculated from experimental mole fractions per Equation 1. Other probabilities are calculated using the formulas shown here. Double strand break/ligation probability estimates make the simplistic assumption that ligation probability of both nicks at the same DNA:DNA junction are the same and independent. Uncertainty (Δ)propagation: $\Delta P_{DSB} = \sqrt{(2 \cdot P_{SSB} \cdot \Delta P_{SSB})^2}$

Table S4.3. continued.

Parent Crystal	**CC2-3′P**	**CC2-3′P**	**CC2-3′P**	**CC2-5′P**	**CC2-5′P**	**CC2-5′P**
Crosslinking Protocol	low	medium	high	low	medium	high
DNA block size	[%]	[%]	[%]	[%]	[%]	[%]
1	94.4	80.3	74.4	96.9	84.8	72.2
2	2.6	4.8	3.3	1.2	5.6	3.1
3	1.5	4.5	4.4	1.9	4.9	7.7
4	0.8	3.6	3.7		2.6	5.2
5	0.7	2.5	2.9		1.1	2.9
6		1.3	2.9		1.0	2.7
7		1.3	2.3			1.9
8		0.6	1.1			1.5
9		0.4	1.2			1.1
10		0.4	1.0			0.7
11		0.2	0.9			0.6
12		0.1	0.7			0.3
13		0.01	0.6			0.1
14		0.01	0.3			0.04
15			0.2			2E-8
16			0.1			

Parent Crystal	**CC2-3′P**	**CC2-3′P**	**CC2-3′P**	**CC2-5′P**	**CC2-5′P**	**CC2-5′P**
Crosslinking Protocol	low	medium	high	low	medium	high
$P_{SSB}{}^{*}$	0.90±0.03	0.61±0.03	0.45±0.02	0.95±0.02	0.75±0.02	0.48±0.01
$P_{LIG} = 1 - P_{SSB}$	0.10±0.03	0.39±0.03	0.55±0.02	0.05±0.02	0.25±0.02	0.52±0.01
$P_{DSB} = (P_{SSB})^2$	0.82±0.05	0.37±0.05	0.20±0.02	0.91±0.04	0.57±0.03	0.23±0.01
$P_{DLIG} = (P_{LIG})^2$	0.01±0.01	0.15±0.02	0.30±0.02	$(2\pm2)\cdot10^{-3}$	0.06±0.01	0.27±0.01

*Calculated from experimental mole fractions per Equation 1. Other probabilities are calculated using the formulas shown here. Double strand break/ligation probability estimates make the simplistic assumption that ligation probability of both nicks at the same DNA:DNA junction are the same and independent. Uncertainty (Δ) propagation: $\Delta P_{DSB} = \sqrt{(2 \cdot P_{SSB} \cdot \Delta P_{SSB})^2}$

Table S4.4. Absolute electron density values for the Figure 4.5 electron density maps

The electron density contours in the Figure 4.5 meshes are set to 3.0 to match COOT 3.0 rmsd contours. By loading the same maps into COOT, and setting the contours to 3.0 rmsd, we can also obtain the contour value for the electron density on an absolute scale.

System	**CC1 original**	**CC1 5’P**	**CC1 3’p**	**CC1 3’P Low**	**CC1 3’P High**
Figure panel	Fig. 5B	Fig. 5D	Fig. 5C	Fig. 5E	Fig. 5F
e Å^{-3}	0.2202 e Å^{-3}	0.1856 e Å^{-3}	0.2085 e Å^{-3}	0.1556 e Å^{-3}	0.1466 e Å^{-3}
Final PDB entry	7rva	7sdp	7sgc	7soz	7spm

Protocol S4.1. Protein sequences for cloning and overexpression in *E. coli*.

RepE54 Transcription Factor (CC1) protein sequence

RepE54 Transcription Factor (sequence below) was cloned into PSB3 Vector (Addgene plasmid # 82027) with an N-terminal His-tag.

MRGSHHHHHHGSMAETAVINHKKRKNSPRIVQSNDLTEAAYSLSRDQKRMLYLFVDQIRKSDGTLQEHDGI
CEIHVAKYAEIFGLTSAEASKDIRQALKSFAGKEVVFYRPEEDAGDEKGYESFPWFIKPAHSPSRGLYSVHINPY
LIPFFIGLQNRFTQFRLSETKEITNPYAMRLYESLCQYRKPDGSGIVSLKIDWIIERYQLPQSYQRMPDFRRRFL
QVCVNEINSRTPMRLSYIEKKKGRQTTHIVFSFRDITSMTTG**

E2F8 Transcription Factor (CC2) protein sequence

The expression plasmid (pETG20A-SBP) containing E2F8 Transcription Factor was donated by the Taipale Lab. The protein sequence contained an N-terminal His-tag and Thioredoxin with a **TEV cleavage site** before the remainder of the E2F8 Transcription Factor.

SDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIR
GIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGHMHHHHHHGTMTSLYKKAGLT**ENLYFQ|G**QP
SRKEKSLGLLCHKFLARYPNYPNPAVNNDICLDEVAEELNVERRRIYDIVNVLESLHMVSRLAKNRYTWHGRH
NLNKTLGTLKSIGEENKYAEQIMMIKKKEYEQEFDFIKSYSIEDHIIKSNTGPNGHPDMCFVELPGVEFRAASV
NSRKDKSLRVMSQKFVMLFLVSTPQIVSLEVAAKILIGEDHVEDLDKSKFKTKIRRLYDIANVLSSLDLIKKVHVT
EERGRKP

Protocol S4.2. Random Ligation Model: Simulation and Mathematics.
To calculate the ligation product size distribution expected for random ligation throughout the crystal we use both simulation and mathematical arguments, which agree to high precision.

Simulation: Using a Python script [Reference 23 in main text], we repeatedly generated a non-ligated 1D stack of 42,856 ssDNA-ssDNA junctions to mimic a crystal height of 300 μm (assuming each DNA strand is 21 nt and 7 nm tall). Each junction has a nick site. We proceeded to ligate nick sites randomly without replacement until the total probability of encountering a single-strand-break, P_{SSB}) fell below the target threshold. For varying target P_{SSB} thresholds (0.1, 0.2, … 0.9) we ran 500 replica simulations. After each random ligation phase, we calculated the ligation product distribution including 1-mer DNA strands where both flanking nicks were not ligated, the number of 2-mer DNA strands where 1 ligation event was still flanked by strand-breaks, and so on. We converted the end product distribution into mole fractions for each size block $(P_1, P_2, P_3, \dots, P_n)$ and took the mean value over the 500 replicas. Error bars are not shown in Fig. Protocol S2 because they are too small to see; the standard error of the mean is below 0.0002 in every case.

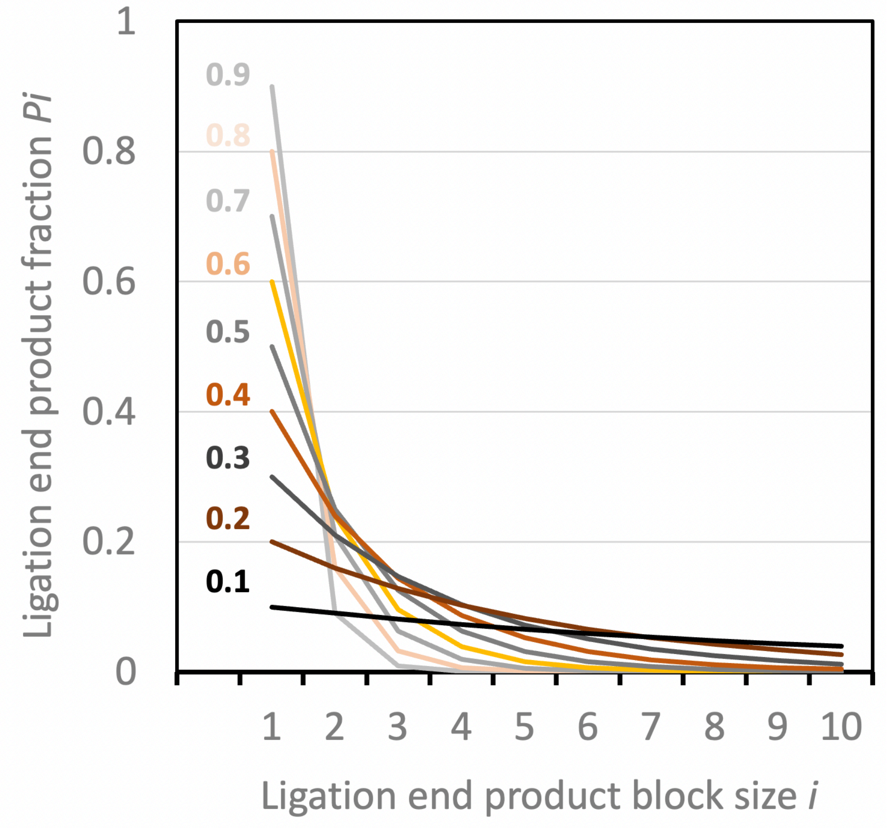


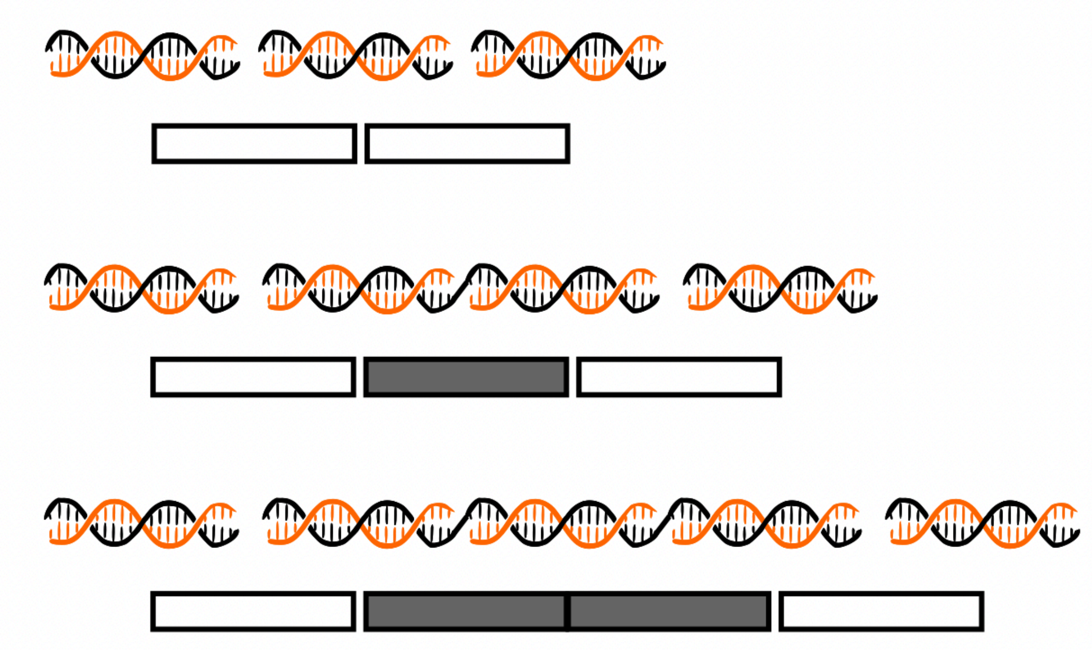
Figure Protocol S4.2. Product distributions from simulated random ligation. Each trace is labeled on the left with the P_{SSB} threshold value used to halt the ligation simulation.

Mathematics:
The simulation data indicated a clear pattern in the outcomes where the probability of finding 1-mer DNA strands after ligation (P_1) was essentially identical to the probability of strand-breaks (P_{SSB}), throughout the DNA stack: $P_1 = P_{SSB}$. This pattern is visible on the left-hand side of Fig. Protocol S2. Furthermore, the probability of finding 2-mer DNA blocks (P_2) was less than the probability of 1-mer blocks by a factor of $(1 - P_{SSB})$, such that $P_2 = (1 - P_{SSB})P_1$. The pattern persisted such that $P_{i+1} = (1 - P_{SSB})P_i$.

The reduced likelihood of finding larger fused strands can be rationalized by considering that the probability that any random strand remains a 1-mer after the ligation phase should logically be linked to the joint probability of independent events (finding a strand-break at two adjacent junctions): $P_1 \propto {P_{SSB}}^2$. Having two adjacent strand breaks is cartooned below as two empty rectangles. In contrast, 2-mers require a non-strand-break (filled rectangle) *in addition to the same flanking strand-breaks*: $P_2 \propto (1 - P_{SSB})(P_{SSB})^2$. Each additional increment in the ligated product size is associated with another factor of $(1 - P_{SSB})$. In other words, we assume that:

$$P_{i+1} = C(1 - P_{SSB})^i(P_{SSB})^2 \qquad \textit{Equation 1.}$$

where C is a proportionality constant that is not yet determined, and i is a counting integer that corresponds to the block size minus 1. When ligating randomly, finding two strand-breaks at adjacent positions is always more probable that finding two strand-breaks with a specified number of intervening non- strand-breaks.



To finish the derivation, we can use the properties of this geometric series. Starting from the generic formula for $r < 1$:

$$\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$$

We can replace r with $(1 - P_{SSB})$:

$$\sum_{i=0}^{\infty}(1 - P_{SSB})^i = \frac{1}{1-(1-P_{SSB})} = \frac{1}{P_{SSB}} \qquad \textit{Equation 2}$$

The 1-mer, 2-mer, and higher order ligation product mole fractions should sum to 1:

$$1 = \sum_{i=0}^{\infty} P_{i+1} \qquad \textit{Equation 3}$$

Substituting Equation 1 into Equation 3:

$$1 = \sum_{i=0}^{\infty} C \cdot {P_{SSB}}^2 (1 - P_{SSB})^i = C \cdot {P_{SSB}}^2 \sum_{i=0}^{\infty}(1 - P_{SSB})^i \qquad \textit{Equation 4}$$

Substituting Equation 2 into Equation 4 we can solve for the proportionality constant C:

$$1 = C \cdot {P_{SSB}}^2 \left(\frac{1}{P_{SSB}}\right) = C \cdot P_{SSB} \qquad \therefore \qquad C = \frac{1}{P_{SSB}}$$

Therefore, we can simplify Equation 1:

$$P_{i+1} = P_{SSB}(1 - P_{SSB})^i \qquad \textit{Equation 5}$$

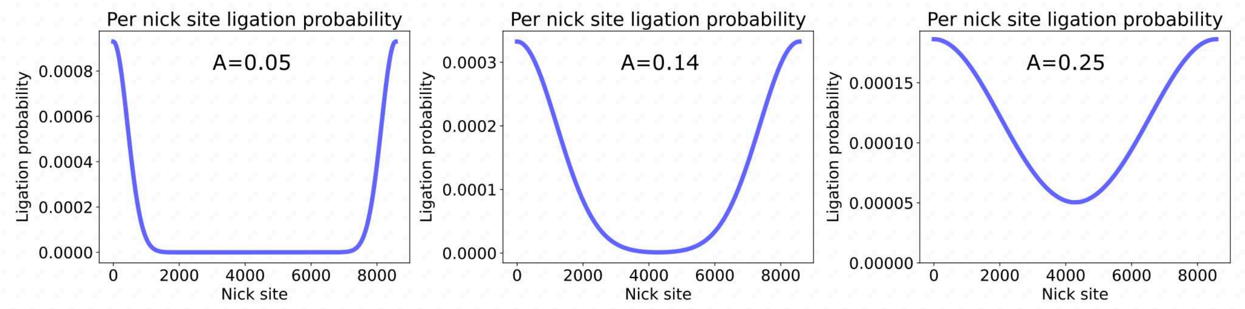
Equation 5 explains why the observed probability for observing 1-mers is equivalent to the double strand break probability:

$$P_1 = P_{0+1} = P_{SSB}(1 - P_{SSB})^0 = P_{SSB}$$

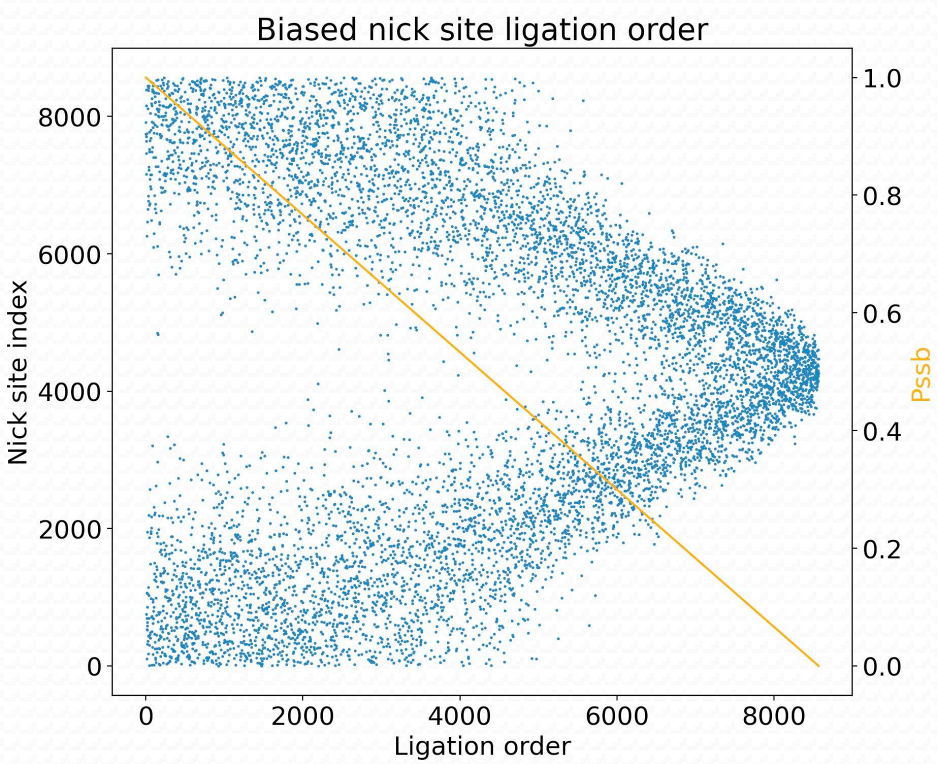
Notably, the random ligation model suggests that the 1-mer population fraction is equivalent to the double-strand-break probability, $P_1 = P_{SSB}$ and that a constant factor equivalent to $1 - P_{SSB}$ describes the population decrease with incrementing block size. Despite the approximations of the random ligation model, we can use these relationships to supply two additional estimates of the double-strand-break probability: $P'_{SSB} = P_1$ and, if κ is the best-fit value for the geometric decline in the population of higher-order ligation products with increasing block size, $P''_{SSB} = 1 - \kappa$. In the case of CC1, P'_{SSB} estimates were notably similar to the P_{SSB} estimates calculated from all the mole fractions (Table 2). In contrast, P'_{SSB} estimates for CC2 were systematically lower than P_{SSB}. P''_{SSB} estimates appear to be a less accurate estimation method because higher-order block populations do not cleanly decline with a reliable geometric factor either when normalized (Table 2) or prior to normalization (SI Table 2). Presumably these small mole fractions are sensitive to noise.

Protocol S4.3. Biased Ligation Model.

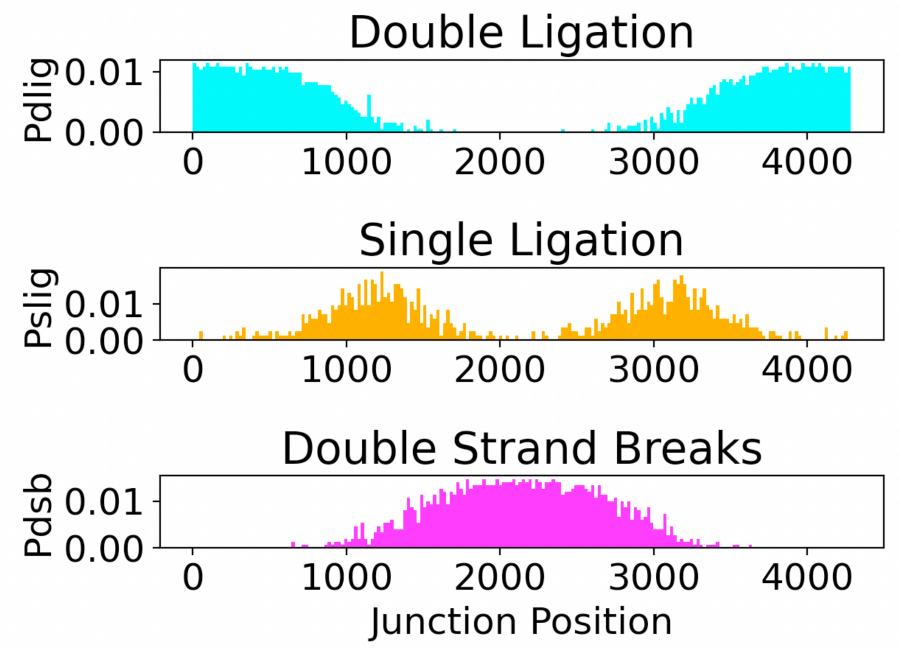
To determine whether transport limitations would be a reasonable explanation for the observed significant deviations from the RLM for CC2 crystals, we ran simulations of ligation where nick sites closer to the surface of the crystal were more likely to be ligated. We used a simple empirical model where the crystal is represented as a 1D stack of DNA strands separated by nicks. Nicks near the edges of this stack had a higher probability of ligation, driven by a single tunable parameter, ***A***. This parameter represents the width of normal distributions centered on the left and right edges of the stack. This is a simple implementation of the idea that reactive small molecules are less likely to reach DNA-DNA junctions at the center of crystal (unless the transport rate is significantly faster than the reaction rate). For example, here are the per-nick-site ligation probability distributions for A = 0.05, 0.14, and 0.25 are shown below:



If we proceed to randomly select nick sites for ligation (without replacement) according to the A=0.14 probability distribution, we are much more likely to ligate nick sites at the edges of the crystal first, as shown from left-to-right in the plot below. With this strong spatial bias, sites close to the center of the crystal are ligated late, with a strong sequential character, as the total probability of single strand breaks (P_{SSB}) drops below 0.4 (shown in orange on the right-hand axis).



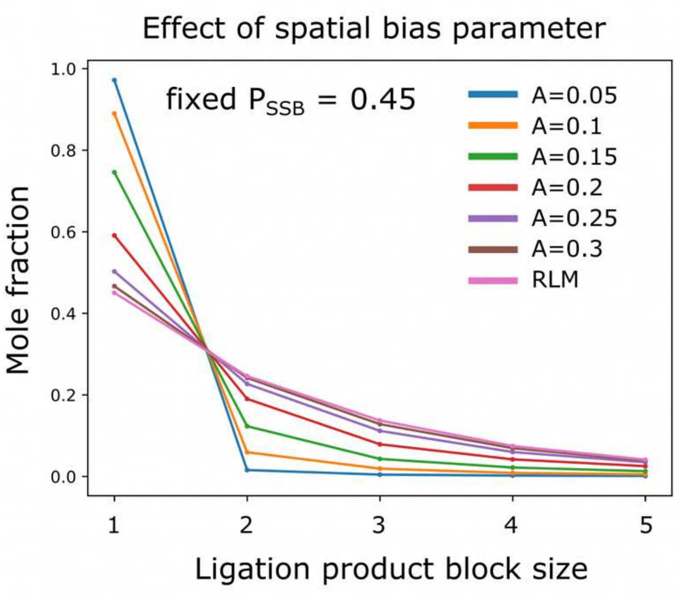
If we halt the ligation process when total P_{SSB} matches an experimentally derived target from the densitometry data (e.g. $P_{SSB} = 0.45$ for CC2-3'P High in Table 4.2), this model suggests a possible spatial distribution of doubly-ligated junctions (P_{DLIG}), singly-ligated junctions (P_{SLIG}), and double strand breaks (P_{0LIG}). For example, the biased ligation model results for A = 0.14 and total $P_{SSB} = 0.45$ are shown below:



Compared to the size distribution for the random ligation model (RLM), this uneven distribution of ligation events throughout the crystal has the net effect of increasing the mole fraction for 1-mer blocks (i.e. in the under-ligated center of the crystal), and decreasing the mole fraction for low-order ligation products, while preserving the relative mole fraction for higher-order ligation products (i.e. in the over-ligated periphery).

In this way, we can use the single additional fitting constant (*A*) to improve the fit for the experimentally observed ligation product mole fractions for CC2, while fixing the P_{SSB} at the value derived from experimental densitometry.

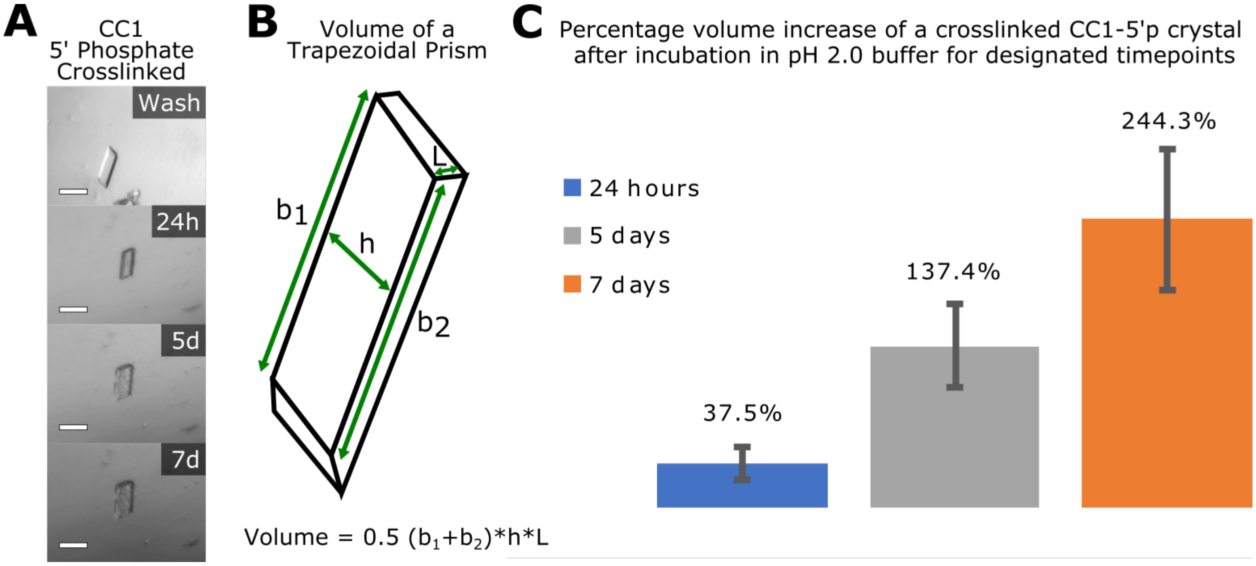
In the case, of CC2-3’P under the strongest ligation conditions, the densitometry analysis (Table 4.2) suggested that $P_{SSB} = 0.45$, but the mole fraction for the 1-mer was estimated to be 74%, which is inconsistent with the RLM. The analysis here suggests that spatial bias in the ligation may resolve this discrepancy. The best fit biased ligation model was for A=0.14 (the case shown above).



Protocol S4.4. Crystal measurements

The stability tests revealed interesting results when the crystals were subjected to very low pH 2.0 to mimic stomach acid (Fig. S4). Specifically, the crosslinked crystals expanded after incubation at pH 2.0. To measure these expansions, crystal pictures were obtained with a Moticam 3.0MP camera attached to a Motic SMZ-168 stereozoom microscope and crystal measurements were performed in Motic Images Plus 2.0.

The crystal measurement example shown is for the crosslinked CC1 5’p after incubating at pH 2.0 for 24 hours, 5 days, and 7 days (Figure Protocol S4.4 A). The crystals resembled trapezoidal prisms, and therefore the measurements were defined to calculate the volume of a trapezoidal prism (Figure Protocol S4.4 B). Measuring crystal dimensions in this way is challenging, with the largest likely source of error being subjective edge definitions. Therefore, to quantify dimensions (and inferred volumes), three people measured the designated edges. The volume was calculated for each set of measurements and the average percent change in volume and the standard error (for the 3 volume estimates) were calculated (Figure Protocol S4.4 C). Significant figures reported in the text were determined by the value of the standard errors.



Appendix IV. SUPPORTING INFORMATION FOR CHAPTER 5

Tuning chemical ligation of DNA within DNA crystals and co-crystals.

Table of Contents

Figure S5.14. Co-crystal stability test – moderately low pH 4.5 to mimic lysosomal fluid.

Figure S5.15. Co-crystal stability test – blood serum.

Table S5.1. Co-crystal variants, duplexes and crystal growth information.

Table S5.2. Co-crystal ligation percentages from gel densitometry (unweighted).

Table S5.3. Co-crystal full version of densitometry output Table 5.2.

Table S5.4. Tensegrity triangle strands for crystal growth.

Table S5.5. DNA crystal ligation percentages from gel densitometry (unweighted).

Table S5.6. DNA crystal full version of densitometry output Table 5.4.

Table S5.7. DNA Crystal 3’P all strands ligation product distribution (corresponds to the gel in Figure 5.4B.

Table S5.8. DNA tile crystallization conditions.

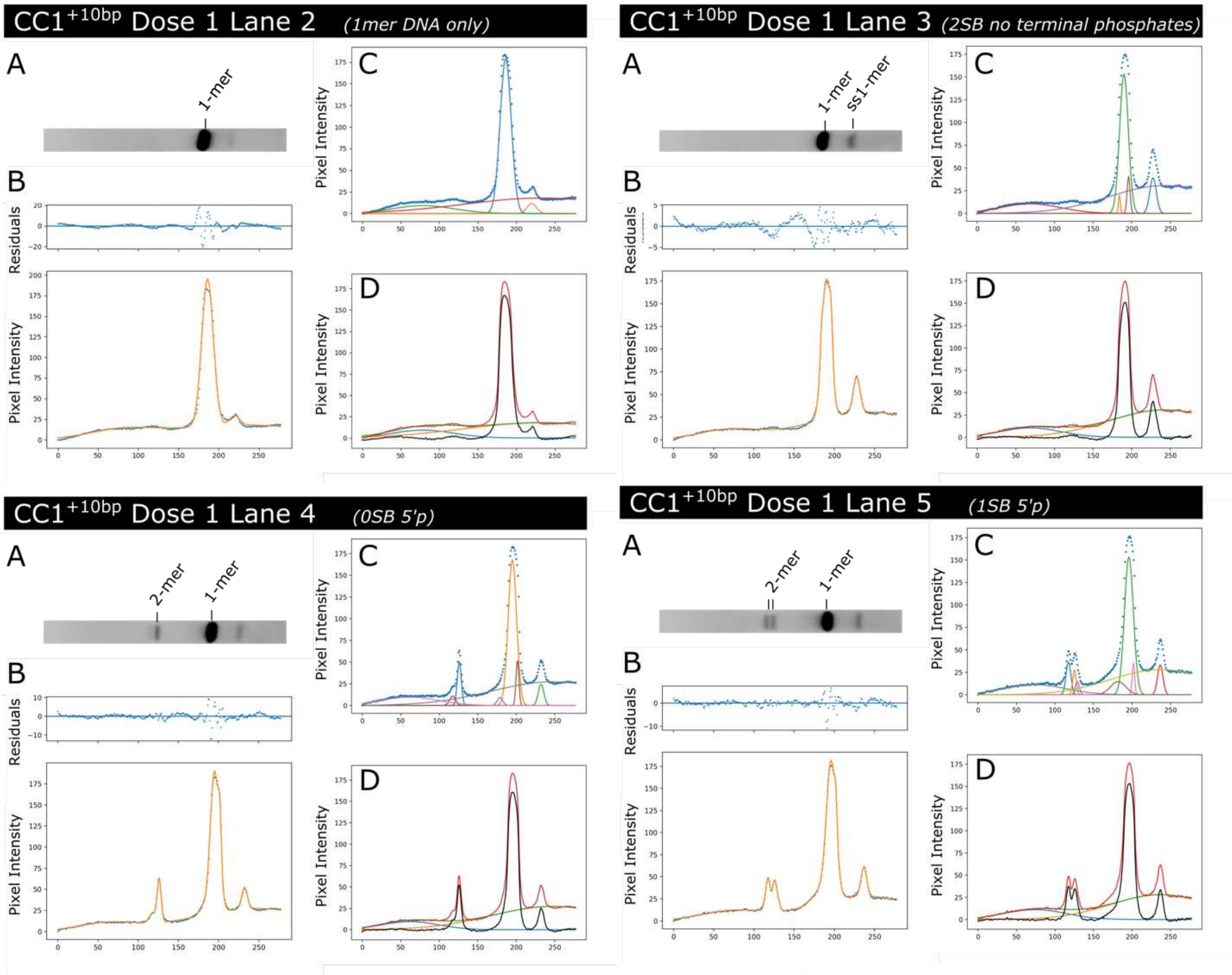


Figure S5.1. Densitometry results and annotation for co-crystals after one EDC dose (corresponds to Figure 5.2A in Chapter 5).
For each lane, we conducted a multi-step densitometry analysis. (A) The lane was manually excised from the gel using ImageJ software. (B) The final fit (orange line) closely follows the raw intensity data (blue dots) which is the average intensity value across the lane (perpendicular to $\vec{E}$ direction). The quality of the fit is also evident in the small residuals (above). (C) The component Gaussian functions inside the fit are shown with different colors. (D) If we subtract the diffuse background (orange) from the full fit (green) the background curve is shown in black. A satellite band that runs faster than the 1-mer is also attributed to 1-mer since it also appears in the control Lane 2 which consisted of only previously annealed 31mer DNA that was heat and chemically denatured on the gel.

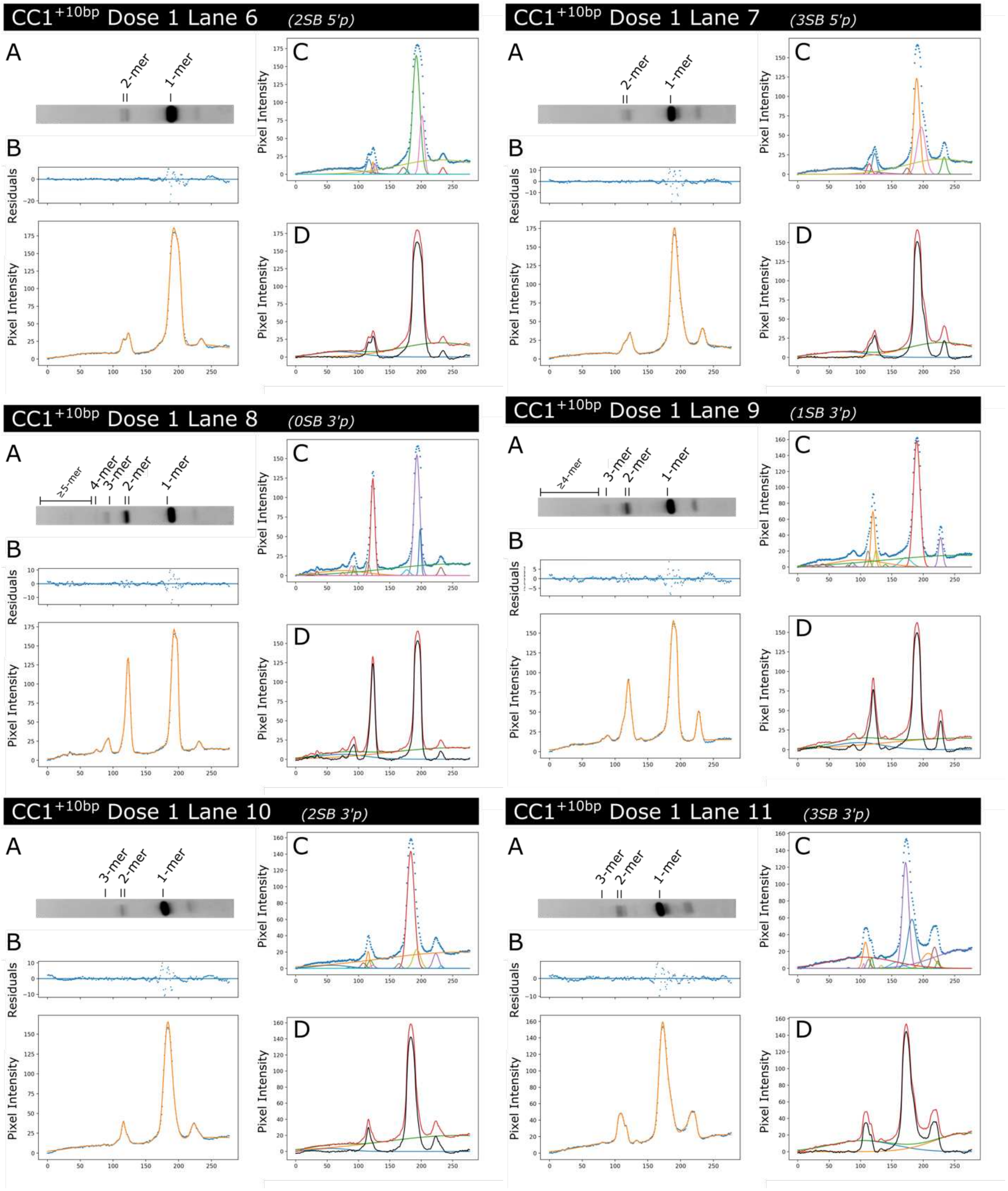


Figure S5.1. continued.

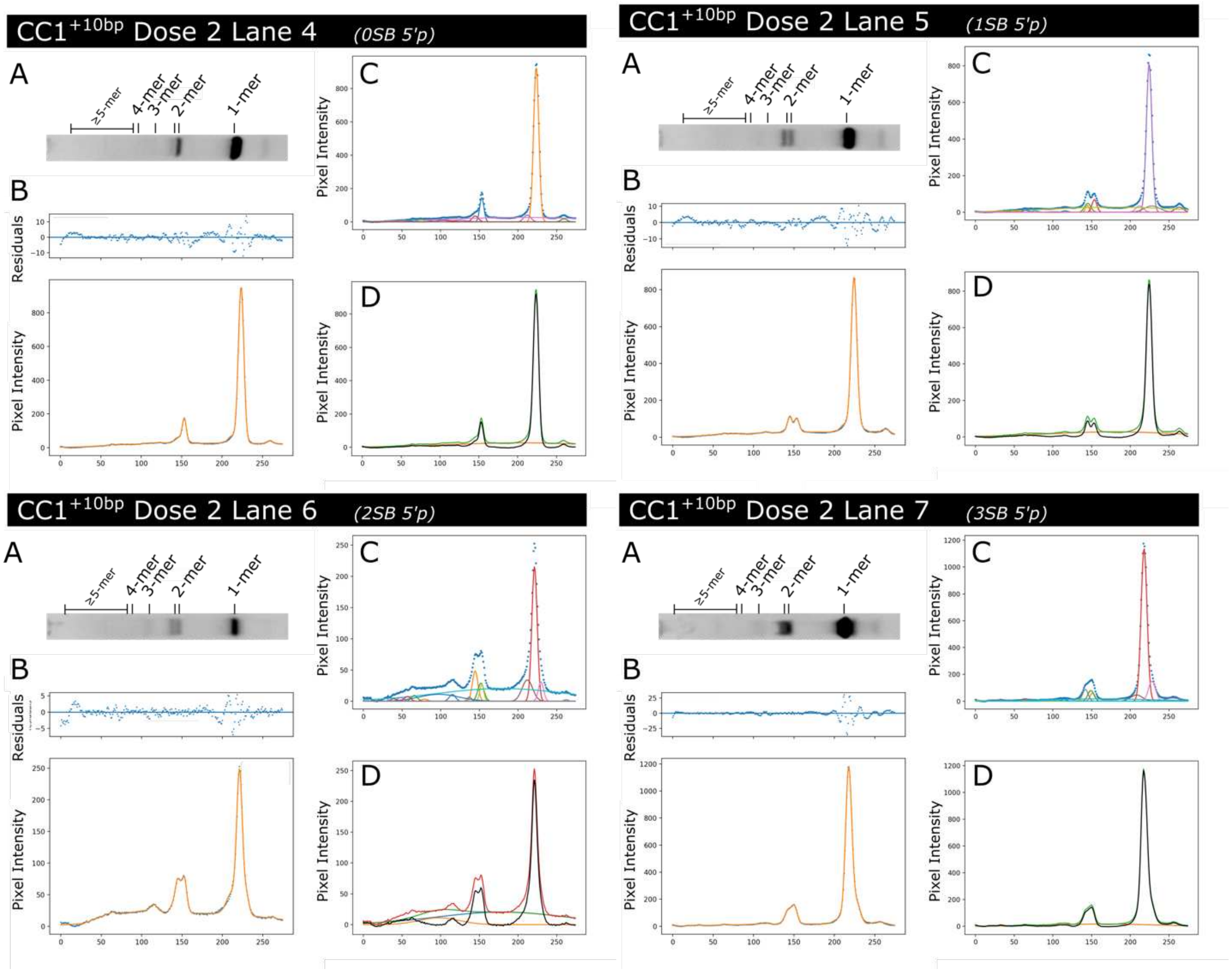


Figure S5.2. Densitometry results and annotation for co-crystals after two EDC doses (corresponds to Figure 5.2B in Chapter 5).
For each lane, we conducted a multi-step densitometry analysis. (A) The lane was manually excised from the gel using ImageJ software. (B) The final fit (orange line) closely follows the raw intensity data (blue dots) which is the average intensity value across the lane (perpendicular to $\vec{E}$ direction). The quality of the fit is also evident in the small residuals (above). (C) The component gaussian functions inside the fit are shown with different colors. (D) If we subtract the diffuse background (orange) from the full fit (green) the background curve is shown in black. A satellite band that runs faster than the 1-mer is also attributed to 1-mer since it also appears in the control Lane 2 which consisted of only previously annealed 31mer DNA that was heat and chemically denatured on the gel.

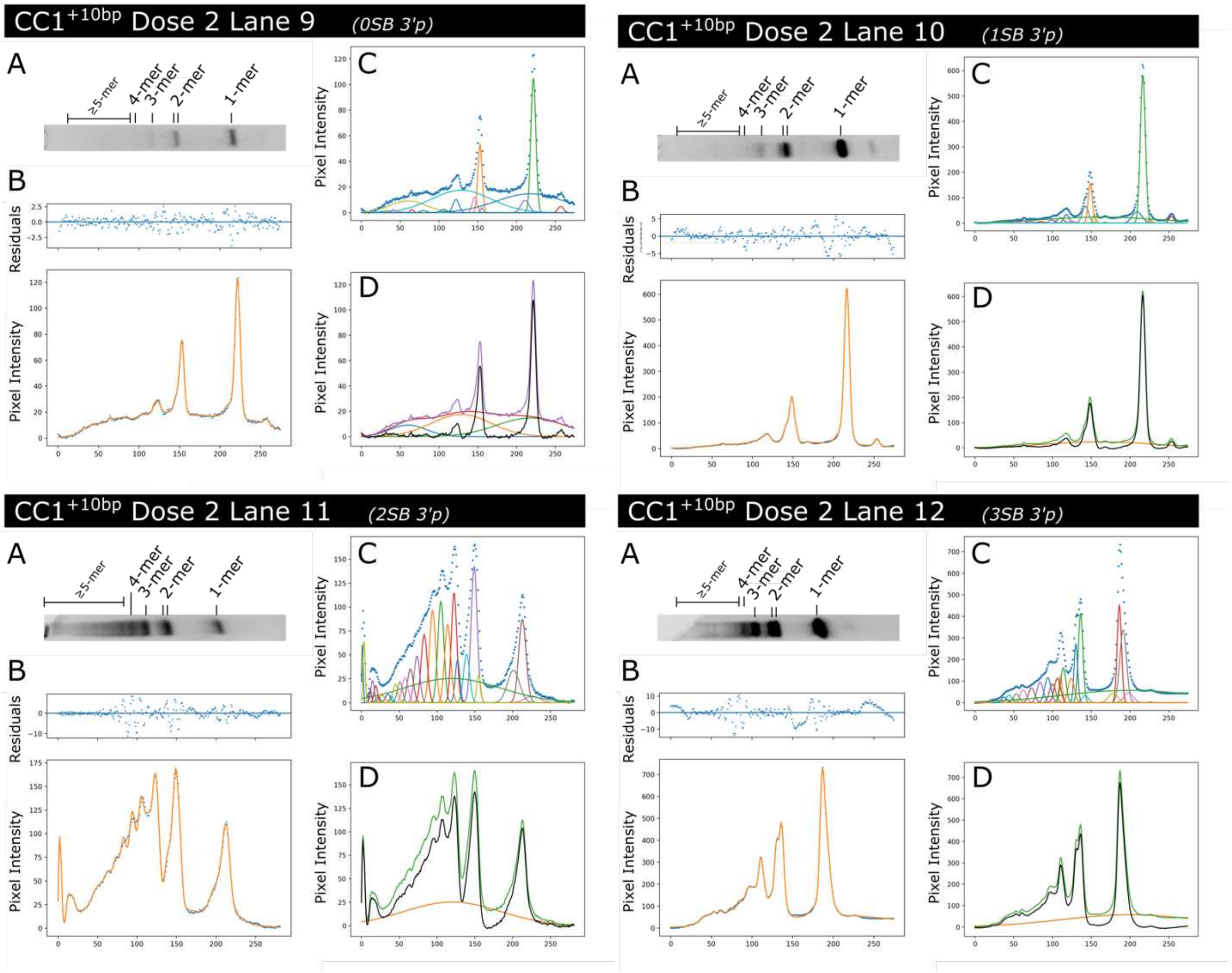


Figure S5.2. continued.

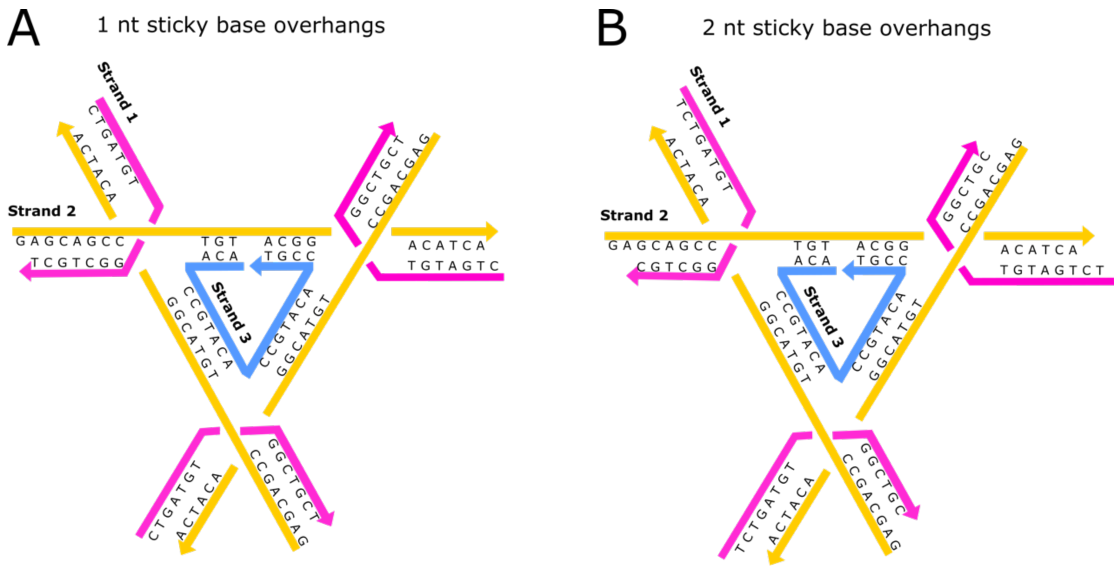


Figure S5.3. Tensegrity Tiles. (A) The DNA tiles with 1nt sticky base overhangs and (B) the DNA tiles with 2nt sticky base overhangs were crystallized with varied terminal phosphorylation at the 5' or 3' ends for chemical ligation comparisons.

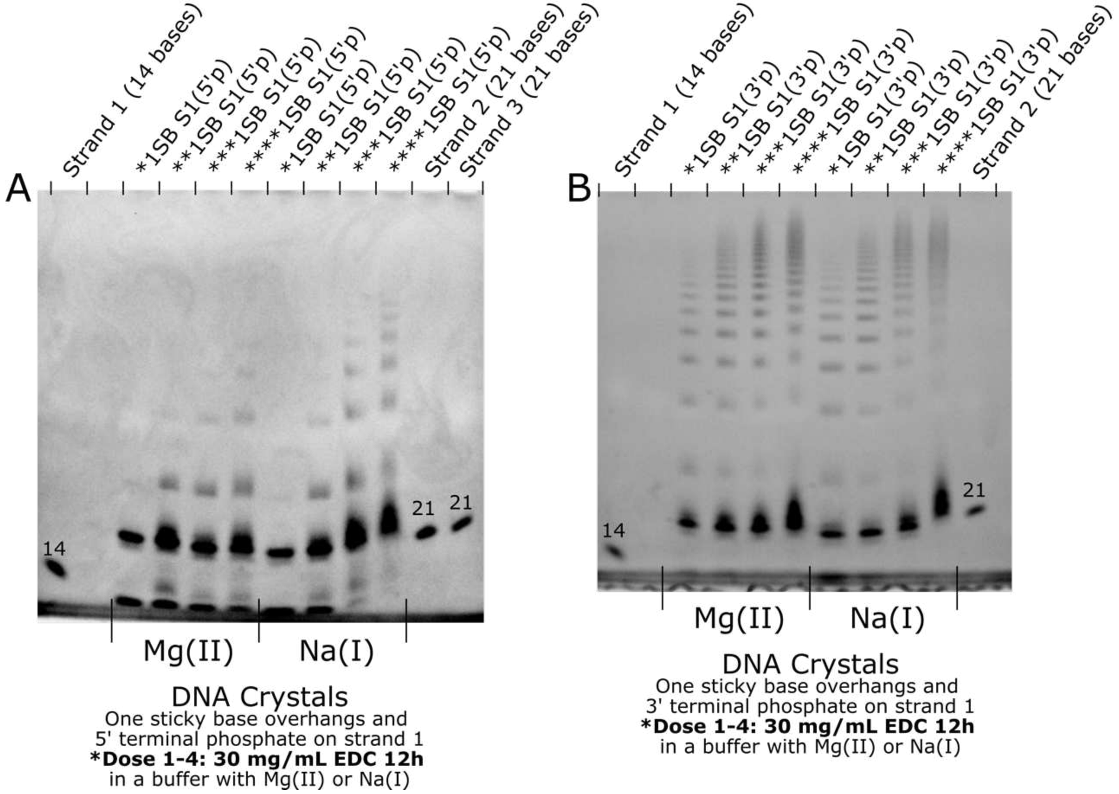


Figure S5.4. Gel electrophoresis of DNA crystals with 1 sticky base overhangs and 5' *vs.* 3' terminal phosphates on Strand 1 only. Gel electrophoresis of (A) DNA crystals with 1 sticky base overhangs and 5' terminal phosphates on Strand 1 only and (B) DNA crystals with 1 sticky base overhangs and 3' terminal phosphates on Strand 1 only.

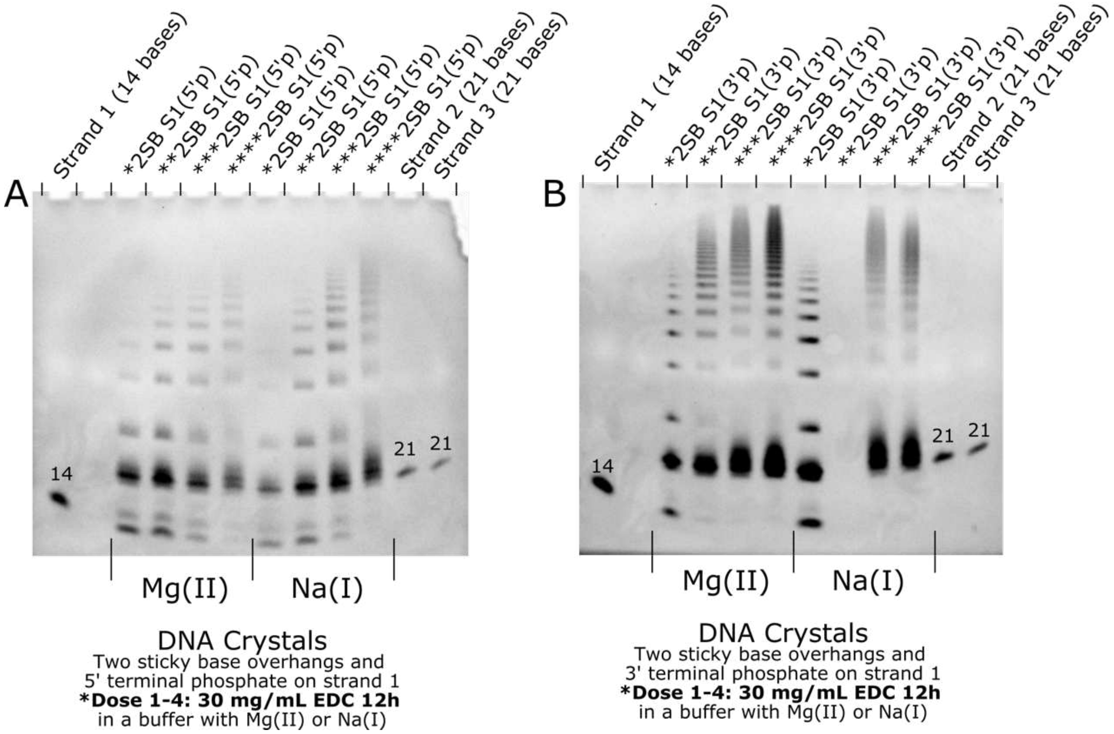
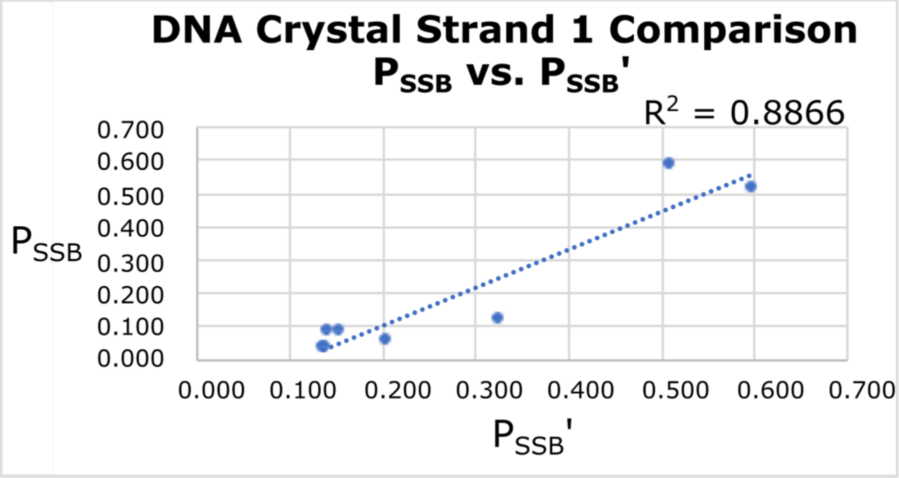


Figure S5.5. Gel electrophoresis of DNA crystals with 2 sticky base overhangs and 5' *vs.* 3' terminal phosphates on Strand 1 only. Gel electrophoresis of (A) DNA crystals with 2 sticky base overhangs and 5' terminal phosphates on Strand 1 only and (B) DNA crystals with 2 sticky base overhangs and 3' terminal phosphates on Strand 1 only.



Parent DNA	1SB 5'P	1SB 3'P	2SB 5'P	2SB 3'P	1SB 5'P	1SB 3'P	2SB 5'P	2SB 3'P
EDC buffer component	Mg(II)	Mg(II)	Mg(II)	Mg(II)	Na(I)	Na(I)	Na(I)	Na(I)
P_{SSB}	0.598	0.154	0.325	0.140	0.511	0.136	0.205	0.138
P_{SSB}'	0.512	0.084	0.124	0.089	0.586	0.039	0.060	0.035

Figure S5.6. Comparison of probability of single stranded breaks. P_{SSB} are the probability of single stranded break calculated from the mole probabilities, as shown in Table 5.5. P_{SSB}' is the value of remaining 1mer after ligation from gel densitometry.

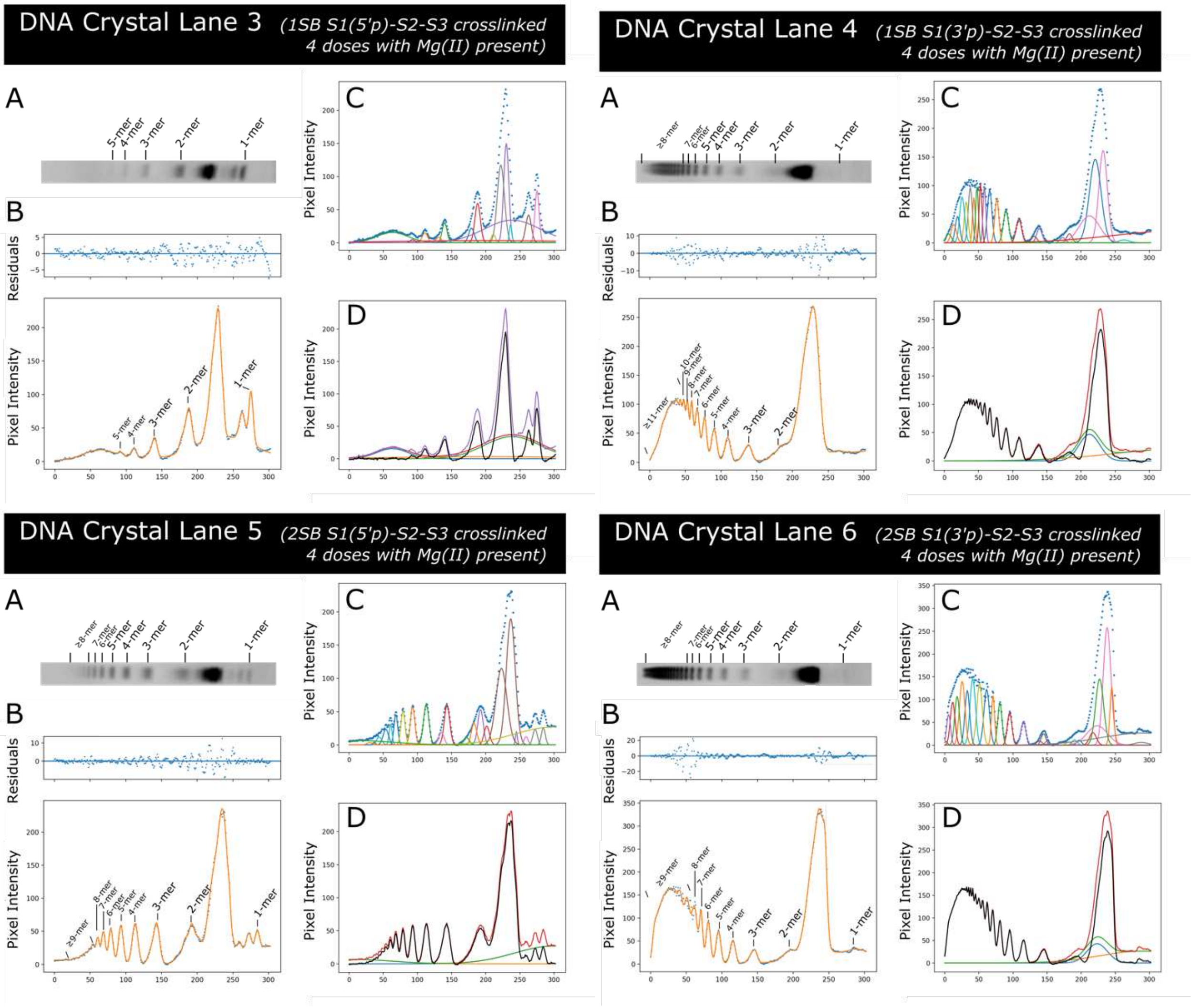
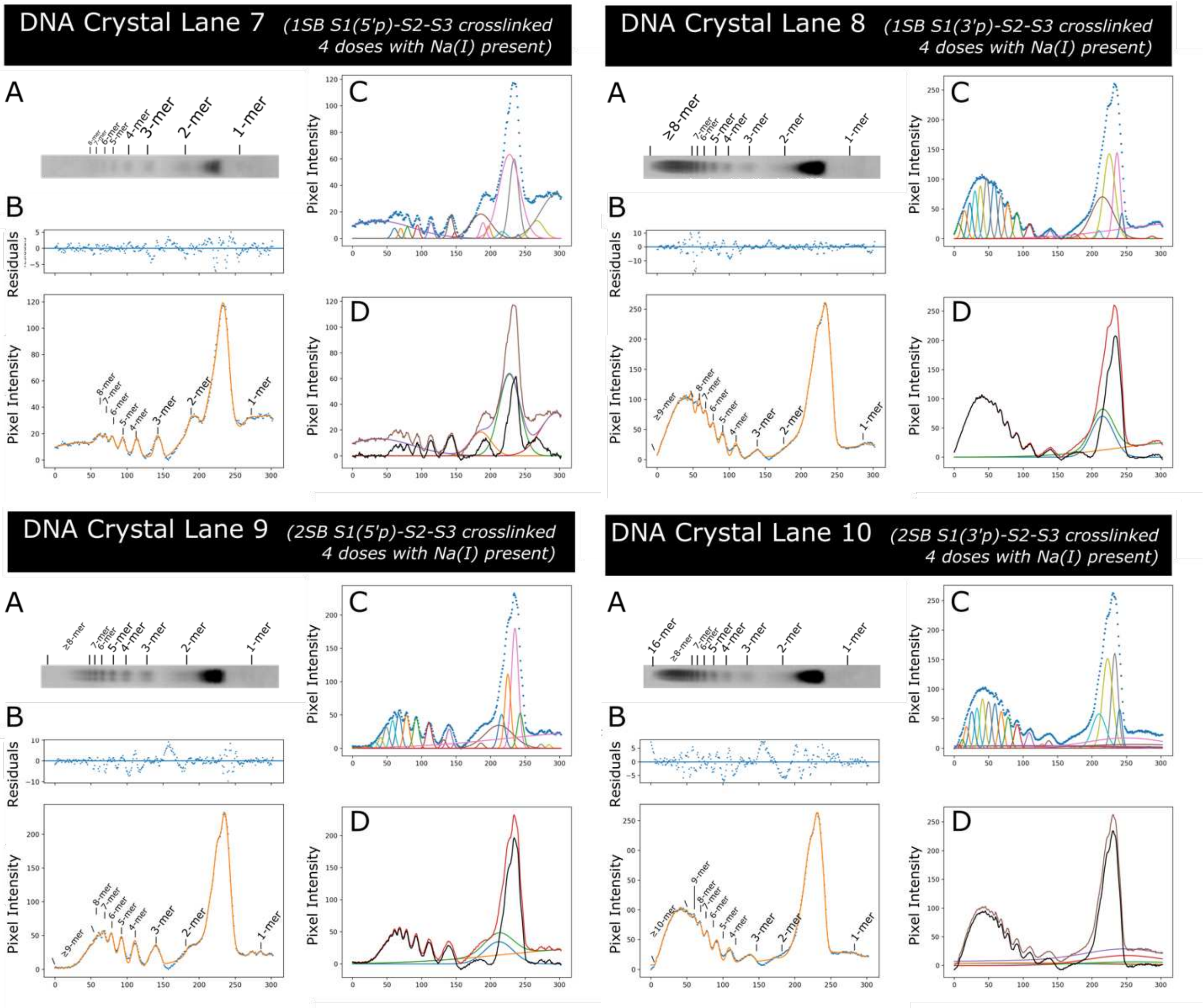


Figure S5.7. Densitometry results and annotation for DNA crystals with strand 1 modifications (corresponds to Figure 5.4A in Chapter 5).

For each lane, we conducted a multi-step densitometry analysis. (A) The lane was manually excised from the gel using ImageJ software. (B) The final fit (orange line) closely follows the raw intensity data (blue dots) which is the average intensity value across the lane (perpendicular to $\vec{E}$ direction). The quality of the fit is also evident in the small residuals (above). (C) The component gaussian functions inside the fit are shown with different colors. (D) If we subtract the diffuse background (orange) from the full fit (green) the background curve is shown in black.


DNA Crystal Lane 7 *(1SB S1(5'p)-S2-S3 crosslinked 4 doses with Na(I) present)*

DNA Crystal Lane 8 *(1SB S1(3'p)-S2-S3 crosslinked 4 doses with Na(I) present)*

DNA Crystal Lane 9 *(2SB S1(5'p)-S2-S3 crosslinked 4 doses with Na(I) present)*

DNA Crystal Lane 10 *(2SB S1(3'p)-S2-S3 crosslinked 4 doses with Na(I) present)*

Figure S5.7. continued.

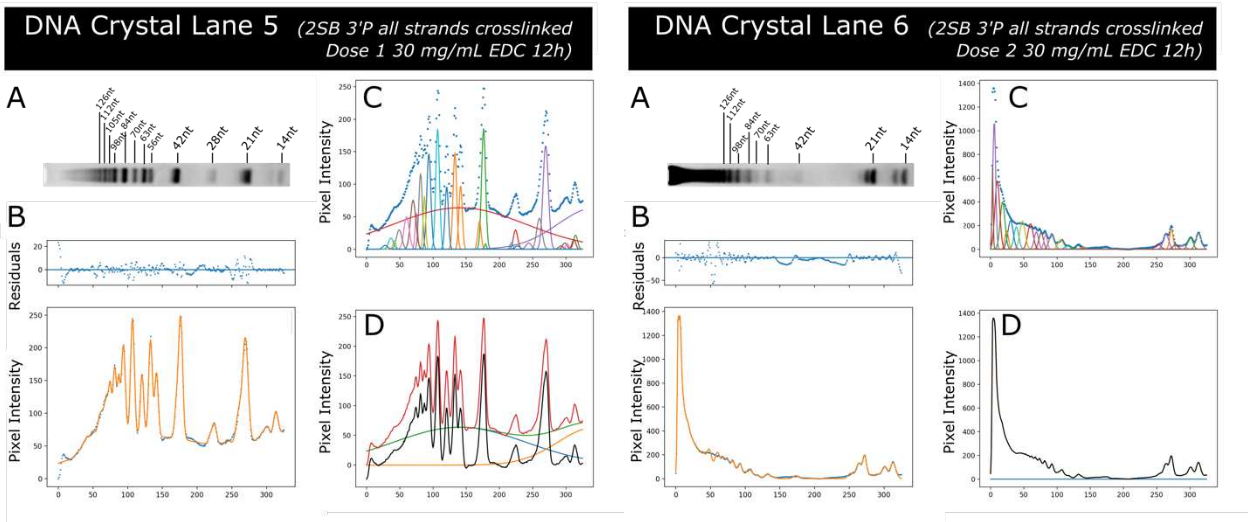


Figure S5.8. Densitometry results and annotation for DNA crystals with 3' phosphates on all strands and 2 sticky base overhangs (corresponds to Figure 5.4B in Chapter 5). For each lane, we conducted a multi-step densitometry analysis. (A) The lane was manually excised from the gel using ImageJ software. (B) The final fit (orange line) closely follows the raw intensity data (blue dots) which is the average intensity value across the lane (perpendicular to $\vec{E}$ direction). The quality of the fit is also evident in the small residuals (above). (C) The component gaussian functions inside the fit are shown with different colors. (D) If we subtract the diffuse background (orange) from the full fit (green) the background curve is shown in black.

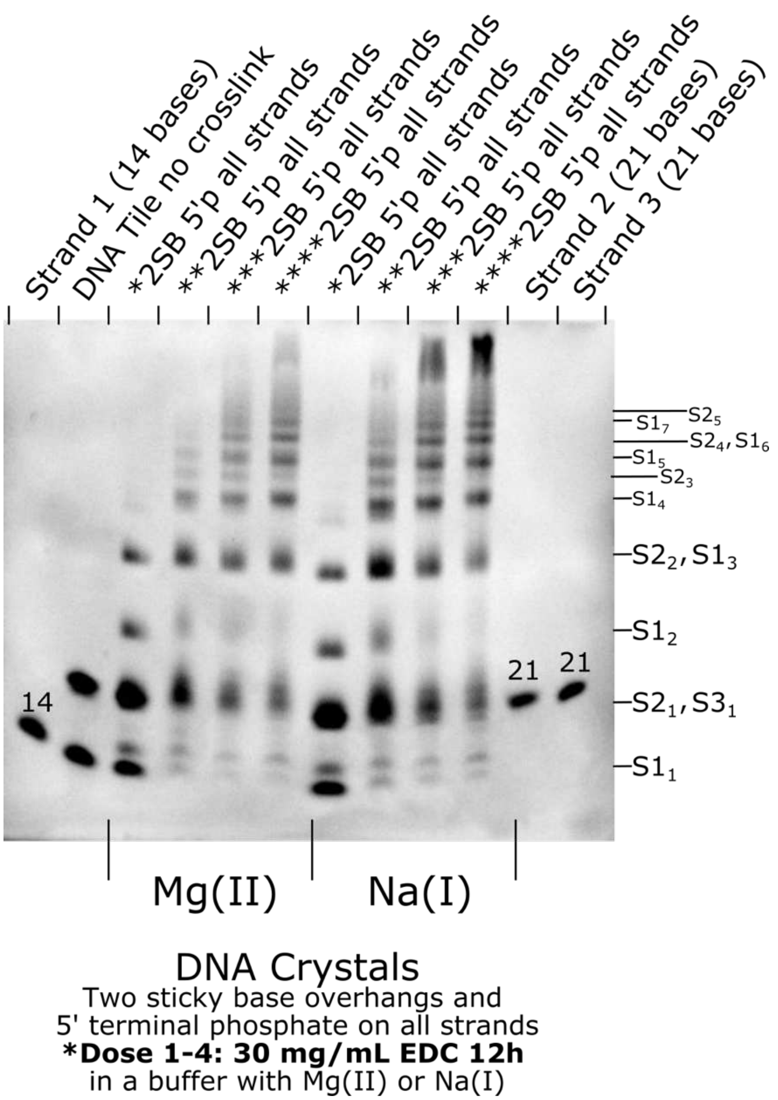
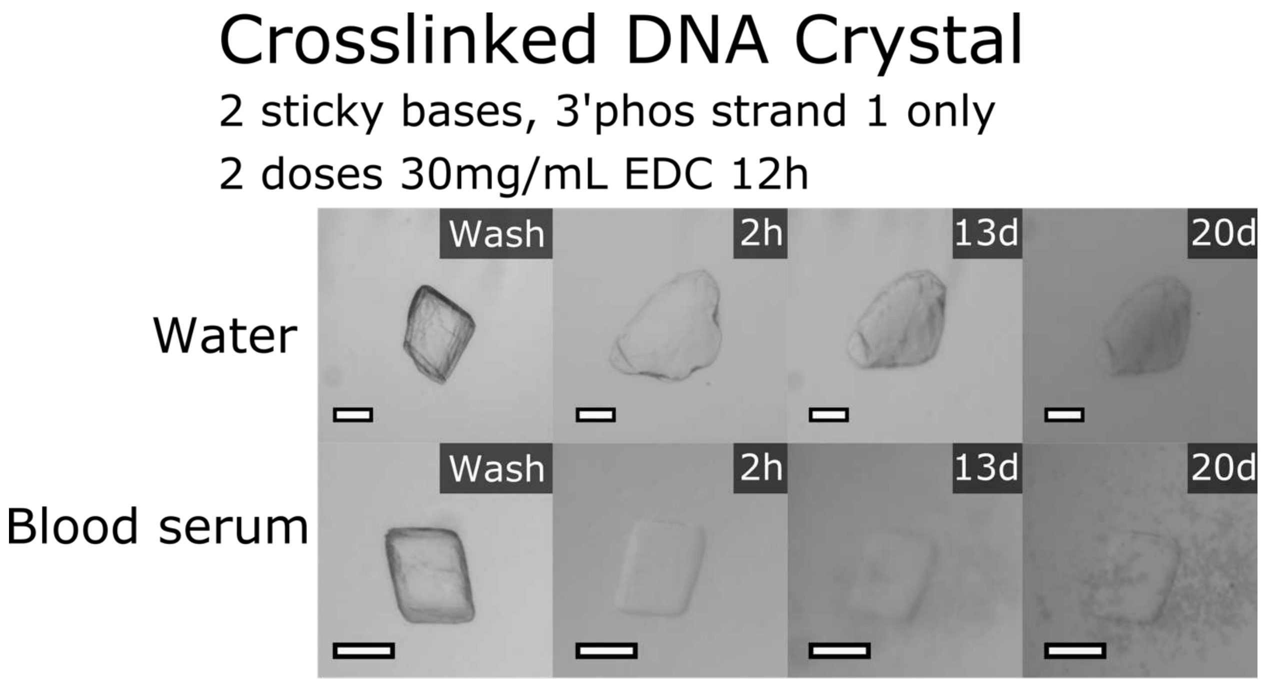


Figure S5.9. Gel electrophoresis of DNA crystals with 2 sticky base overhangs and 5’ terminal phosphates on all strands.



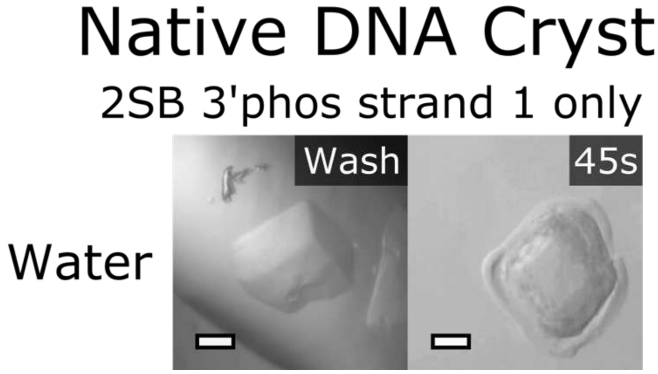


Figure S5.10. DNA Crystal stability test – water and blood serum. The tensegrity triangle DNA crystals shown had 2 sticky base overhangs and a terminal 3’phosphate on strand 1 only. Crosslinked crystals were crosslinked with two doses of 30 mg/mL EDC for 12 hours and quenched with tris base pH 8.2 for 30 minutes prior to transfer to the wash solution. All scale bars are 100 µm. Tensegrity Triangle crystals with 3’phosphates on strand one were crosslinked and transferred to wash solution: 40 mM MES pH 6.0 and 125 mM NaCl**.** The crosslinked crystals, after transfer to water or blood serum, remained intact for 20 days. The native, non-crosslinked crystal degraded in water after 45 seconds.

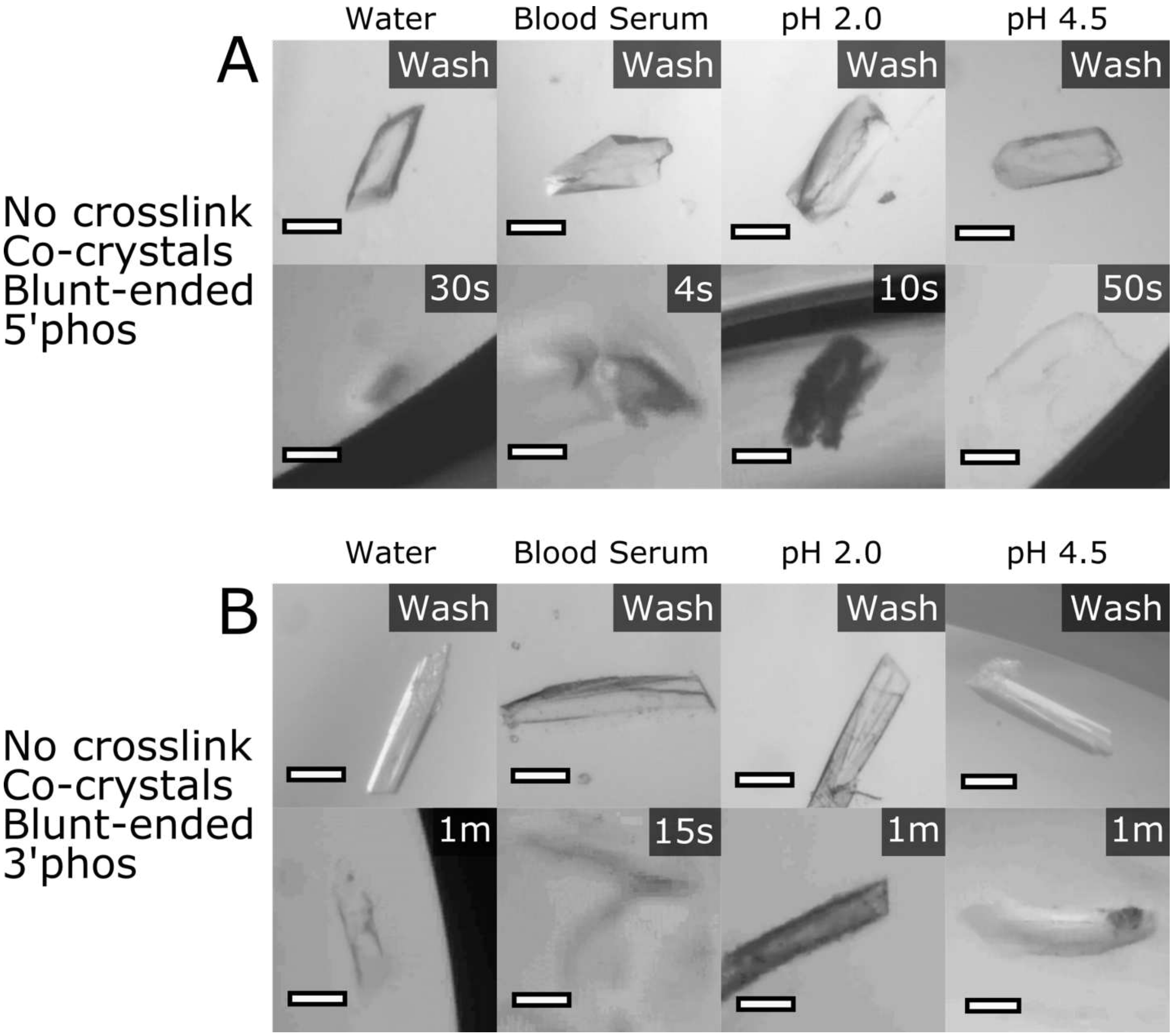


Figure S5.11. Native co-crystal stability test. The crystals grown were in 350 mM $MgCl_2$, 30% PEG 400, and 100 mM tris HCl pH 8.0. All scale bars are 100 μm. (A) CC1 blunt ended 5’terminal phosphate crystals in wash solution matching the growth solution and after transfer to the harsh condition listed above: water, blood serum, pH 2.0 and pH 4.5. (B) CC1 blunt ended 3’terminal phosphate crystals in wash solution matching the growth solution and after transfer to the harsh condition listed above: water, blood serum, pH 2.0 and pH 4.5. All crystals either dissolved or converted to an aggregate after 1 minute.

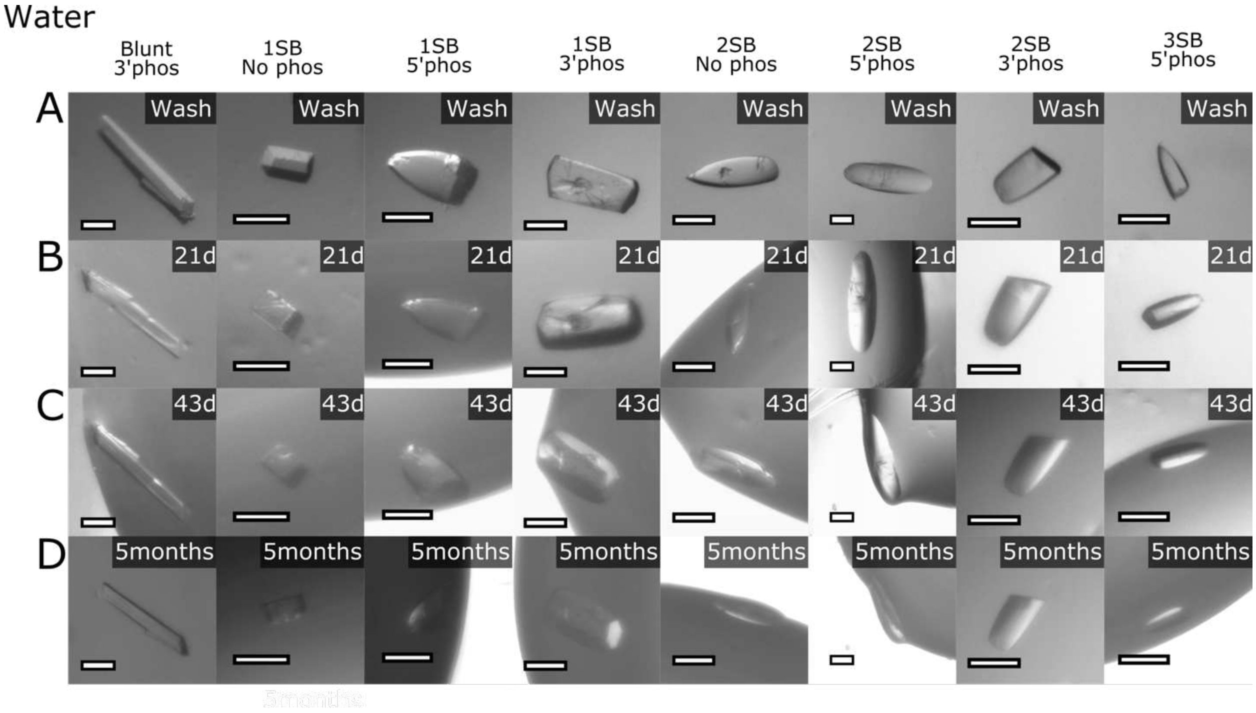


Figure S5.12. Co-crystal stability test – water. The crystals were crosslinked with two doses of 30 mg/mL EDC for 12 hours and quenched with tris base pH 8.2 for 30 minutes prior to transfer to the wash solution. All scale bars are 100 µm. (A) CC1 crystals in wash solution containing 300-500 mM NaCl, 25-35% PEG 400, and 100 mM MES buffer pH 6.0. The concentrations of the wash solution matched the initial crystal growth solutions, but we replaced MgCl$_2$ with NaCl and tris HCl pH 8.0 with MES buffer pH 6.0. (B) CC1^{+10bp} with varied sticky base overhangs and terminal phosphorylation (listed above each panel) after transfer to deionized water for (B) 21 days (C) 43 days and (D) 5 months. All crystals, regardless of phosphorylation or sticky base overhang length, remained intact.

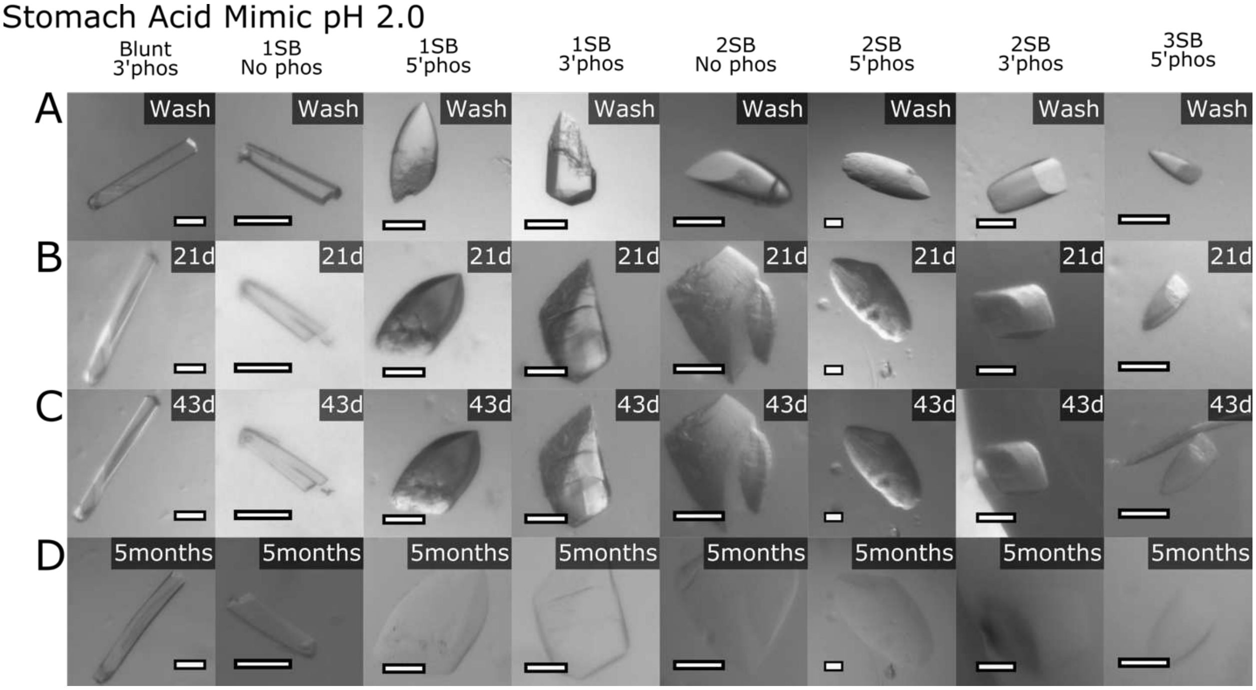


Figure S5.13. Co-crystal stability test – very low pH 2.0 to mimic stomach acid. The crystals were crosslinked with two doses of 30 mg/mL EDC for 12 hours and quenched with tris base pH 8.2 for 30 minutes prior to transfer to the wash solution. All scale bars are 100 µm. (A) CC1 crystals in wash solution containing 300-500 mM NaCl, 25-35% PEG 400, and 100 mM MES buffer pH 6.0. The concentrations of the wash solution matched the initial crystal growth solutions, but we replaced $MgCl_2$ with NaCl and tris HCl pH 8.0 with MES buffer pH 6.0. (B) CC1^{+10bp} with varied sticky base overhangs and terminal phosphorylation (listed above each panel) after transfer to a pH 2.0 stomach acid mimic (0.01M HCl) for (B) 21 days (C) 43 days and (D) 5 months.

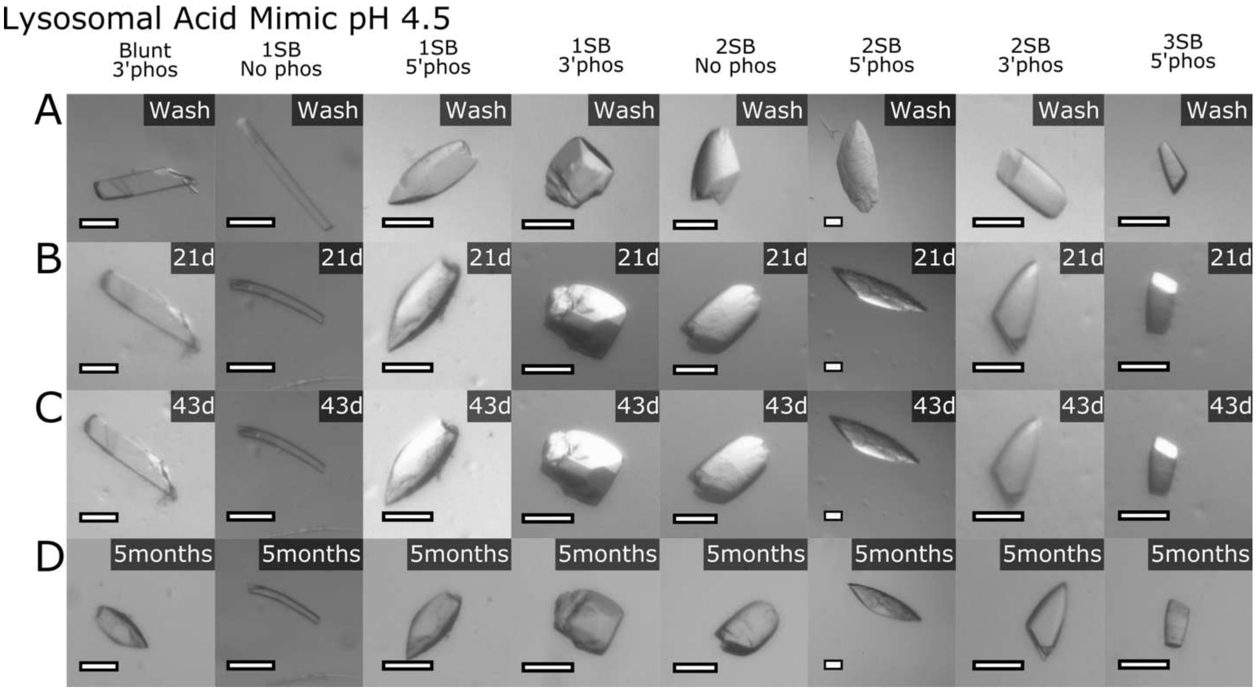


Figure S5.14. Co-crystal stability test – moderately low pH 4.5 to mimic lysosomal fluid. The crystals were crosslinked with two doses of 30 mg/mL EDC for 12 hours and quenched with tris base pH 8.2 for 30 minutes prior to transfer to the wash solution. All scale bars are 100 µm. (A) CC1 crystals in wash solution containing 300-500 mM NaCl, 25-35% PEG 400, and 100 mM MES buffer pH 6.0. The concentrations of the wash solution matched the initial crystal growth solutions, but we replaced $MgCl_2$ with NaCl and tris HCl pH 8.0 with MES buffer pH 6.0. (B) $CC1^{+10bp}$ with varied sticky base overhangs and terminal phosphorylation (listed above each panel) after transfer to a pH 4.5 lysosomal fluid mimic (46 mM sodium citrate, 54.1 mM citric acid) for (B) 21 days (C) 43 days and (D) 5 months.

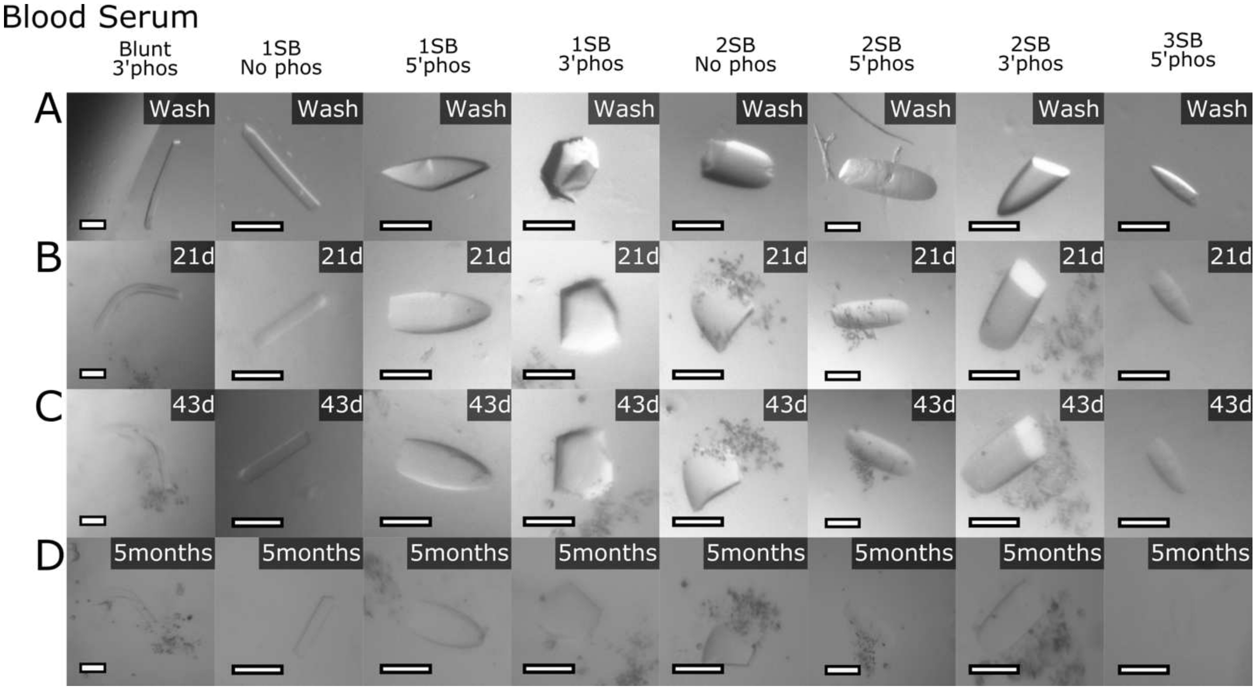


Figure S5.15. Co-crystal stability test – blood serum. The crystals were crosslinked with two doses of 30 mg/mL EDC for 12 hours and quenched with tris base pH 8.2 for 30 minutes prior to transfer to the wash solution. All scale bars are 100 µm. (A) CC1 crystals in wash solution containing 300-500 mM NaCl, 25-35% PEG 400, and 100 mM MES buffer pH 6.0. The concentrations of the wash solution matched the initial crystal growth solutions, but we replaced $MgCl_2$ with NaCl and tris HCl pH 8.0 with MES buffer pH 6.0. (B) CC1^{+10bp} with varied sticky base overhangs and terminal phosphorylation (listed above each panel) after transfer to to blood serum (HyClone, Bovine Calf Serum Product #: SH30073.02) for (B) 21 days (C) 43 days and (D) 5 months. All crystals, regardless of phosphorylation or sticky base overhang length, remained intact after 21 days. Although, the blunt ended 3'phosphate crystal was bent. After 43 days, the the crystals were still intact and the blunt 3'phosphate crystal was still bent. After 5 months, crystals remained, although possibly smaller than original size.

Table S5.1. Co-crystal variants, duplexes and crystal growth information.

Sticky base overhang and terminal phosphorylation	**DNA Duplexes** CC1^{+10bp} Symmetric G-C rich (1:1.2 RepE54:DNA)	**Crystallization Condition**
0SB 5’phos	5′ pCCCGGACCTGTGACAAATTGCCCTCAGACGG 3′ 3′ GGGCCTGGACACTGTTTAACGGGAGTCTGCCp 5′	400 mM MgCl$_2$, 25% PEG 400, 100 mM tris HCl pH 8.0
0SB 3’phos	5′ CCCGGACCTGTGACAAATTGCCCTCAGACGGp 3′ 3′ pGGGCCTGGACACTGTTTAACGGGAGTCTGCC 5′	350 mM MgCl$_2$, 25% PEG 400, 100 mM tris HCl pH 8.0
1SB 5’phos	5′ pCCCGGACCTGTGACAAATTGCCCTCAGACGG 3′ 3′ GGCCTGGACACTGTTTAACGGGAGTCTGCCGp 5′	350 mM MgCl$_2$, 30% PEG 400, 100 mM tris HCl pH 8.0
1SB 3’phos	5′ CCCGGACCTGTGACAAATTGCCCTCAGACGGp 3′ 3′ pGGCCTGGACACTGTTTAACGGGAGTCTGCCG 5′	350 mM MgCl$_2$, 25% PEG 400, 100 mM tris HCl pH 8.0
2SB 5’phos	5′ pCCCGGACCTGTGACAAATTGCCCTCAGACGG 3′ 3′ GCCTGGACACTGTTTAACGGGAGTCTGCCGGp 5′	300 mM MgCl$_2$, 25% PEG 400, 100 mM tris HCl pH 8.0
2SB 3’phos	5′ CCCGGACCTGTGACAAATTGCCCTCAGACGGp 3′ 3′ pGCCTGGACACTGTTTAACGGGAGTCTGCCGG 5′	400 mM MgCl$_2$, 30% PEG 400, 100 mM tris HCl pH 8.0
2SB no phos	5′ CCCGGACCTGTGACAAATTGCCCTCAGACGG 3′ 3′ GCCTGGACACTGTTTAACGGGAGTCTGCCGG 5′	450 mM MgCl$_2$, 25% PEG 400, 100 mM tris HCl pH 8.0
3SB 5’phos	5′ pCCCGGACCTGTGACAAATTGCCCTCAGACGG 3′ 3′ CCTGGACACTGTTTAACGGGAGTCTGCCGGGp5′	300 mM MgCl$_2$, 25% PEG 400, 100 mM tris HCl pH 8.0
3SB 3’phos	5′ CCCGGACCTGTGACAAATTGCCCTCAGACGGp 3′ 3′ pCCTGGACACTGTTTAACGGGAGTCTGCCGGG 5′	450 mM MgCl$_2$, 20% PEG 400, 100 mM tris HCl pH 8.0

Table S5.2. Co-crystal ligation percentages from gel densitometry (unweighted). The data below corresponds with Table 2 in the text. Shown here is the *unweighted* distribution where the length of the DNA and resulting dye binding quantity is not incorporated in the percentages.

Co-crystals EDC Dose One (Figure 5.2A)

Parent Co-crystal	**0SB 5′P**	**1SB 5′P**	**2SB 5′P**	**3SB 5′P**	**0SB 3′P**	**1SB 3′P**	**2SB 3′P**	**3SB 3′P**
DNA block size	[%]	[%]	[%]	[%]	[%]	[%]	[%]	[%]
1	86.8	84.8	89.7	87.7	60.7	71.9	88.7	87.6
2	13.2	15.2	10.3	12.3	31.2	23.0	10.8	12.2
3					5.0	1.6	0.6	0.1
4					0.8	0.4		
5					0.0	0.6		
6					0.4	1.1		
7					0.8	0.9		
8 and above					1.0	0.6		

Co-crystals EDC Dose Two (Figure 5.2B)

Parent Co-crystal	**0SB 5′P**	**1SB 5′P**	**2SB 5′P**	**3SB 5′P**	**0SB 3′P**	**1SB 3′P**	**2SB 3′P**	**3SB 3′P**
DNA block size	[%]	[%]	[%]	[%]	[%]	[%]	[%]	[%]
1	79.9	84.1	66.6	83.9	59.9	66.0	21.5	46.9
2	12.4	11.8	23.2	12.8	27.8	22.2	24.9	31.0
3	2.2	0.7	2.7	0.7	5.9	6.8	21.8	12.4
4	2.2	0.7	0.0	0.7	1.5	1.4	8.5	2.8
5	0.6	0.8	0.8	0.1	0.0	0.7	3.6	0.2
6	1.0	1.0	2.3	0.5	1.0	0.7	0.2	0.0
7	0.7	0.5	1.9	0.2	0.8	1.0	0.9	1.1
8 and above	1.0	0.4	2.6	1.0	3.1	1.2	18.7	5.6

Table S5.3. Co-crystal full version of densitometry output Table 5.2.

This version of Table 5.2 includes the small mole fractions for higher-order peaks (still derived from reweighted band intensities).

Co-crystals EDC Dose One (Figure 5.2A)

Parent Co-crystal	0SB 5′P	1SB 5′P	2SB 5′P	3SB 5′p	0SB 3′P	1SB 3′P	2SB 3′P	3SB 3′p
DNA block size	[%]	[%]	[%]	[%]	[%]	[%]	[%]	[%]
1	92.9	91.8	94.6	93.4	77.4	85.1	94.1	93.4
2	7.1	8.2	5.4	6.6	19.8	13.6	5.7	6.5
3					2.1	0.6	0.2	0.1
4					0.3	0.1		
5					$7\text{x}10^{-9}$	0.1		
6					0.1	0.2		
7					0.2	0.2		
8					0.2	0.1		

Parent Co-crystal	0SB 5′P	1SB 5′P	2SB 5′P	3SB 5′P	0SB 3′P	1SB 3′P	2SB 3′P	3SB 3′P
$P_{SSB}{}^{*}$	0.93 ±0.01	0.92 ±0.01	0.95 ±0.01	0.94 ±0.01	0.77 ±0.02	0.82 ±0.03	0.94 ±0.02	0.93 ±0.02
$P_{LIG} = 1 - P_{SSB}$	0.07 ±0.01	0.08 ±0.01	0.05 ±0.01	0.06 ±0.01	0.23 ±0.02	0.18 ±0.03	0.06 ±0.02	0.07 ±0.02
$P_{DSB} = (P_{SSB})^2$	0.87 ±0.02	0.85 ±0.02	0.90 ±0.02	0.88 ±0.02	0.59 ±0.04	0.68 ±0.04	0.88 ±0.03	0.87 ±0.03
$P_{DLIG} = (P_{LIG})^2$	$4 \pm 2 \cdot 10^{-3}$	$6 \pm 2 \cdot 10^{-3}$	$3 \pm 1 \cdot 10^{-3}$	$4 \pm 2 \cdot 10^{-3}$	0.05 ±0.01	0.03 ±0.01	$4 \pm 2 \cdot 10^{-3}$	$5 \pm 2 \cdot 10^{-3}$

---

*Calculated from experimental mole fractions *per* Equation 1 (Table 5.1). Other probabilities are derived from P_{DSB} as shown (Equations 2-4, Table 5.1).

Table S5.3 continued.

Co-crystals EDC Dose Two (Figure 5.2B)								
Parent Co-crystal	**0SB 5′P**	**1SB 5′P**	**2SB 5′P**	**3SB 5′p**	**0SB 3′P**	**1SB 3′P**	**2SB 3′P**	**3SB 3′p**
DNA block size	[%]	[%]	[%]	[%]	[%]	[%]	[%]	[%]
1	90.9	92.6	83.1	92.2	78.0	82.3	42.4	61.3
2	7.1	6.5	14.4	7.0	18.1	13.9	23.6	23.0
3	0.8	0.3	1.1	0.3	2.6	2.8	16.2	8.0
4	0.6	0.2	$1x10^{-8}$	0.2	0.5	0.4	5.9	3.8
5	0.1	0.2	0.2	0.03	$2x10^{-8}$	0.1	4.3	1.44
6	0.2	0.2	0.5	0.1	0.2	0.1	2.6	0.8639
7	0.1	0.1	0.3	0.03	0.1	0.2	1.5	0.6
8	0.1		0.2	0.02	0.0	0.1	1.0	0.4
9			0.1	0.05	0.1	0.04	0.6	0.2
10				0.02	0.1	0.04	0.4	0.2
11				0.01	0.04	0.01	0.3	0.2
12				$1x10^{-6}$	0.1		0.1	0.1
13				0.02	0.1		0.1	
14							0.1	
15							0.1	
16							0.1	
17							0.2	
18							0.1	
19							$1x10^{-9}$	
20							$1x10^{-9}$	
21							0.4	

Parent Co-crystal	**0SB 5′P**	**1SB 5′P**	**2SB 5′P**	**3SB 5′P**	**0SB 3′P**	**1SB 3′P**	**2SB 3′P**	**3SB 3′P**
${P_{SSB}}^{*}$	0.86 ±0.03	0.89 ±0.03	0.79 ±0.03	0.84 ±0.04	0.73 ±0.03	0.77 ±0.03	0.39 ±0.01	0.57 ±0.02
$P_{LIG} = 1 - P_{SSB}$	0.14 ±0.03	0.11 ±0.03	0.21 ±0.03	0.16 ±0.04	0.27 ±0.03	0.23 ±0.03	0.61 ±0.01	0.43 ±0.02
$P_{DSB} = (P_{SSB})^2$	0.74 ±0.05	0.78 ±0.05	0.62 ±0.05	0.70 ±0.06	0.53 ±0.04	0.59 ±0.04	0.15 ±0.01	0.33 ±0.02
$P_{DLIG} = (P_{LIG})^2$	0.02 ±0.01	0.01 ±0.01	0.04 ±0.01	0.03 ±0.01	0.08 ±0.02	0.05 ±0.01	0.37 ±0.02	0.18 ±0.02

**Calculated from experimental mole fractions *per* Equation 1 (Table 5.1). Other probabilities are derived from P_{DSB} as shown (Equations 2-4, Table 5.1).

Table S5.4. Tensegrity triangle strands for crystal growth. The strands below were synthesized by IDT with either 5’ or 3’ terminal phosphates.

Strand ID	Oligomer Sequence
Strand 1 (S1) 1 SB	5′ – CTGATGTGGCTGCT – 3′
Strand 1 (S1) 2 SB	5′ – TCTGATGTGGCTGC – 3′
Strand 2 (S2) 1SB or 2SB	5′ – GAGCAGCCTGTACGGACATCA – 3′
Strand 3 (S3) 1SB or 2SB	5′ – ACACCGTACACCGTACACCGT – 3′

Table S5.5. DNA crystal ligation percentages from gel densitometry (unweighted). The data below corresponds with Table 5.4 in the text. Shown here is the *unweighted* distribution where the length of the DNA and resulting dye binding quantity is not incorporated in the percentages.

DNA Crystal Strand 1 Ligation Products (Figure 5.4A)								
Parent DNA	**1SB 5′P**	**1SB 3′P**	**2SB 5′P**	**2SB 3′P**	**1SB 5′P**	**1SB 3′P**	**2SB 5′P**	**2SB 3′P**
EDC buffer component	**Mg(II)**	**Mg(II)**	**Mg(II)**	**Mg(II)**	**Na(I)**	**Na(I)**	**Na(I)**	**Na(I)**
DNA block size	[%]	[%]	[%]	[%]	[%]	[%]	[%]	[%]
1	30.6	1.3	4.2	1.2	15.1	0.6	1.2	0.5
2	41.8	3.8	28.4	2.3	8.9	2.6	5.7	0.3
3	17.9	4.7	17.9	3.0	8.7	2.6	9.3	3.5
4	7.6	5.4	13.8	4.0	5.1	3.5	11.2	4.4
5	2.1	6.9	10.7	5.0	3.5	6.1	13.3	6.6
6		7.3	8.7	6.1	3.2	8.1	14.6	8.6
7		9.8	7.1	5.2	2.5	9.2	15.1	9.7
8 and above		58.7	13.1	73.2	3.5	67.2	29.5	66.3

Table S5.6. DNA crystal full version of densitometry output Table 5.4.

This version of Table 5.4 includes the small mole fractions for higher-order peaks (still derived from reweighted band intensities).

DNA Crystal Strand 1 Ligation Products (Figure 5.4A)

Parent DNA	1SB 5′P	1SB 3′P	2SB 5′P	2SB 3′P	1SB 5′P	1SB 3′P	2SB 5′P	2SB 3′P
EDC buffer component	**Mg(II)**	**Mg(II)**	**Mg(II)**	**Mg(II)**	**Na(I)**	**Na(I)**	**Na(I)**	**Na(I)**
DNA block size	[%]	[%]	[%]	[%]	[%]	[%]	[%]	[%]
1	51.2	8.4	12.4	8.9	58.6	3.9	6.0	3.5
2	35.0	12.3	41.9	8.1	17.4	8.5	13.9	1.2
3	10.0	10.2	17.6	7.1	11.3	5.8	15.2	8.6
4	3.2	8.7	10.2	7.1	5.0	5.8	13.6	8.0
5	0.7	9.0	6.3	7.2	2.7	8.1	13.0	9.7
6		7.9	4.3	7.2	2.0	8.9	11.9	10.5
7		9.1	3.0	5.3	1.4	8.7	10.5	10.1
8		5.2	1.4	9.1	1.7	11.0	7.0	11.3
9		4.8	1.8	8.9		10.3	5.1	10.5
10		4.1	0.8	8.7		7.6	2.1	10.0
11		3.0	0.1	5.7		6.9	1.1	6.6
12		5.1	0.2	7.0		5.9	0.2	6.2
13		3.3		4.3		3.9	0.3	2.9
14		4.3		3.5		3.1	0.02	0.8
15		2.1		1.9		1.5	0.2	0.1
16		1.6		0.2				$5\text{x}10^{-5}$
17		0.8						

Parent DNA	1SB 5′P	1SB 3′P	2SB 5′P	2SB 3′P	1SB 5′P	1SB 3′P	2SB 5′P	2SB 3′P
EDC buffer component	**Mg(II)**	**Mg(II)**	**Mg(II)**	**Mg(II)**	**Na(I)**	**Na(I)**	**Na(I)**	**Na(I)**
$P_{SSB}{}^{*}$	0.598 ±0.012	0.154 ±0.004	0.325 ±0.009	0.140 ±0.004	0.511 ±0.014	0.136 ±0.004	0.205 ±0.006	0.138 ±0.004
$P_{LIG} = 1 - P_{SSB}$	0.402 ±0.012	0.846 ±0.004	0.675 ±0.009	0.860 ±0.004	0.489 ±0.014	0.864 ±0.004	0.795 ±0.06	0.862 ±0.004

---

*Calculated from experimental mole fractions per Equation 1 (Table 5.1). P_{LIG} is derived from P_{SSB} as shown.

Table S5.7. DNA Crystal 3’P all strands ligation product distribution (corresponds to the gel in Figure 5.4B. The percentages are weighted based on the number of nucleotides in the sequence to contribute GelRed dye binding to the oligomer lengths. The length of oligomers are estimated based on the expected ligation products. For example, strand 1 (14mer) ligates to strand 1 (14mer), forming a 28mer. Strand 2 (21mer) ligates to strand 2 (21mer) to form a 42mer. Strand 3 ligates 5’ to 3’ on the same strand, forming a circular product.

Percentage Oligomer	Dose 1 (Lane 5)	Dose 2 (Lane 6)
% 14mer	6.99	18.69
% 17mer	3.83	10.84
% 19mer		3.62
% 21mer	35.94	13.14
% 22mer	1.59	11.36
% 23mer		3.60
% 28mer	4.50	0.00
% 42mer	14.62	1.15
% 56mer	4.02	0.00
% 63mer	5.76	1.57
% 70mer	3.18	0.79
% 84mer	5.58	1.84
% 98mer	4.10	2.45
% 105mer	1.30	0.00
% 112mer	4.94	1.35
% 126mer	2.19	1.78
% 140mer	0.78	1.64
% 147mer	0.18	2.83
% 154mer	0.38	3.00
% 168mer		1.97
% 182mer		2.57
% 189mer		1.04
% 196mer		3.58
% 210mer		4.79
% 224mer		5.09
% 231mer or greater		1.32

Table S5.8. DNA tile crystallization conditions. All crystals were grown with sitting drop vapor diffusion plates (Hampton) with 1.75M Ammonium sulfate in the reservoir and the following crystallization conditions in the well. The plates were annealed for 10 min at 60 °C and cooled to room temperature in ~60 min.

Sticky base overhang and terminal phosphorylation	**Crystallization Condition**
1SB S1(5'P)-S2-S3	14 µM DNA tile, 60 mM tris base pH 8.5, 30 mM acetic acid, 3 mM EDTA, 750 mM magnesium acetate
1SB S1(3'P)-S2-S3	14 µM DNA tile, 15 mM tris base pH 8.5, 7.5 mM acetic acid, 0.75 mM EDTA, 187.5 mM magnesium acetate
2SB S1(5'P)-S2-S3	14 µM DNA tile, 15 mM tris base pH 8.5, 7.5 mM acetic acid, 0.75 mM EDTA, 187.5 mM magnesium acetate
2SB S1(5'P)-S2-S3	14 µM DNA tile, 15 mM tris base pH 8.5, 7.5 mM acetic acid, 0.75 mM EDTA, 187.5 mM magnesium acetate
2SB S1(5'P)-S2(5'P)-S3(5'P)	14 µM DNA tile, 3 mM tris base pH 8.5, 1.5 mM acetic acid, 0.15 mM EDTA, 37.5 mM magnesium acetate
2SB S1(3'P)-S2(3'P)-S3(3'P)	14 µM DNA tile, 60 mM tris base pH 8.5, 30 mM acetic acid, 3 mM EDTA, 750 mM magnesium acetate