DISSERTATION

MODULAR PROTEIN-DNA CO-CRYSTALS FOR PROGRAMMABLE SCAFFOLD ASSISTED STRUCTURE DETERMINATION

Submitted by

Ethan Shields

School of Biomedical and Chemical Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, CO

Spring 2026

Doctoral Committee:

Advisor: Christopher Snow

Christie Peebles

Margarita Herrera-Alonso

Grant Schauer

ABSTRACT

MODULAR PROTEIN-DNA CO-CRYSTALS FOR PROGRAMMABLE SCAFFOLD ASSISTED STRUCTURE DETERMINATION

The structure of biomolecules can provide key insight into understanding their underlying function. X-ray crystallography is a principal technique in structural biology used to determine atomic-level details of macromolecular structures. Despite major advances in X-ray diffraction instrumentation and computational modelling, the practical bottleneck in crystallography continues to be crystallization itself. Scaffold assisted crystallography offers an alternative to approach to traditional crystallization. Instead of crystallizing each target independently, one could grow a robust crystalline framework in advance and subsequently introduce guest molecules for structure determination. To be practically useful, such scaffold crystals must be highly porous, modular, and mechanically stable under diverse solvent conditions.

Here, a family of co-crystals (CC1) was systematically expanded through lattice engineering of its DNA and protein constituents. The programmable nature of DNA enabled rational modulation of lattice dimensions, pore architecture, and the incorporation of tailored sequence motifs that serve as anchoring sites for guest DNA-binding proteins. At the same time, the protein component was engineered through point mutations and sequence fusions—changes that could disrupt crystal formation—yet the lattice retained its overall topology and packing geometry. This combined DNA- and protein-level modularity provides an unusually broad design landscape for a crystalline system, establishing CC1 as a robust and adaptable scaffold for subsequent functional applications.

Finally, CC1 scaffold crystal variants were applied to the long-standing challenge of post-crystallization guest structure determination. Stabilized crystals accommodated substantial changes in solvent environment and enabled a range of DNA-binding proteins to diffuse into the lattice and bind their cognate sites in an ordered fashion. These bound guests produced clear electron density suitable for structural analysis. This achievement represents the first successful use of a porous protein–DNA co-crystal to determine the structure of guest macromolecular added via simple soaking, providing a definitive demonstration of scaffold-assisted crystallography and establishing CC1 as a foundational platform for future structural biology and biomaterials applications.

ACKNOWLEDGEMENTS

Thank you to all the Snow Lab members for your support, advice, and friendship

Dr. Christopher Snow

Dr. Abigail Orun

Dr. Julius Stuart

Dr. Jacob Deroo

Alec Jones

Ashlyn Chen

Caroline Slaughter

Rachel Cohen

Emma Magna

John League

Kiberly Robbins

Pegah Eizadkhah

Thank you to my collaborators for your advice and comradery

Dr. Donald Spratt

Fadwa Mekkaoui

Thank you to my committee members for your guidance and support

Dr. Christie Peebles

Dr. Grant Schauer

Dr. Margarita Herrera-Alonso

Funding

National Science Foundation Department of Materials Research Grant Number 2003748

National Science Foundation Department of Materials Research Grant Number 2310574

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1. POROUS CRYSTALINE SCAFFOLDS FOR STRUCTURE DETERMINATION

1.1. Abstract

Nanomaterial scaffolds provide an organized framework for otherwise disordered states of matter. When these scaffolds adopt crystalline order, they can be resolved in atomic detail by X-ray or electron diffraction, and guest molecules incorporated systematically within the lattice may likewise be determined. Although the use of crystalline scaffolds to assist in guest structure determination is not yet routine, porous scaffolds capable of immobilizing biomolecules have been demonstrated with varying degrees of success. Protein-based, DNA-based, and hybrid inorganic–organic frameworks have each demonstrated unique advantages, ranging from tunable porosity and chemical stability to biocompatibility and ease of functionalization. In this chapter, we review recent milestones in scaffold crystal development and compare the advantages and limitations of different classes of crystalline scaffolds.

1.2. Introduction

X-ray crystallography and cryo-electron microscopy (cryo-EM) are cornerstone techniques for resolving biomolecular structures, yet both face persistent bottlenecks in sample preparation. Crystallography depends on growing single, well-ordered crystals which may require years of exhaustive screening and still fail to yield diffraction-quality samples. Cryo-EM, while circumventing crystallization, is limited by the difficulty of forming thin, vitreous ice layers that not only preserve molecular integrity but also produce well-distributed particle orientations. In both methods, preparative challenges are often the most time-consuming and

failure-prone step. To overcome such barriers, structural biology would benefit from robust strategies to impose order on biomolecules. In this chapter we explore the use of crystalline scaffolds, variations on the idea of pre-ordered frameworks that can facilitate guest structure determination, via diffraction.

Scaffold-assisted crystallography was first proposed by Nadrian Seeman in the form of a porous three-dimensional self-assembling DNA lattice with sequence variable DNA blocks for cognate DNA-binding guests to join the crystal[1]. In contrast to the hit-and-miss screening required in traditional crystallography, the DNA scaffold would assemble independently via high-affinity nucleic acid interactions (i.e. Watson-Crick base pairing and base-stacking) and guest binding would be driven by affinity. Circumvention of sample preparation variability is a core advantage of scaffold-assisted crystallography. Other theoretical advantages include the possibility of lower guest purity requirements, lower guest quantity requirements, straightforward model preparation, suitability for high-throughput data collection, and independent scaffold optimization could permit stable diffraction in extended solvent conditions.

In recent years, porous crystalline scaffolds have been explored for a wide range of applications beyond structural biology, including molecular storage[2–4], catalysis[5–7], molecular sieves and nanoscale devices[8–10]. Importantly, several studies[11–13] have demonstrated that such scaffolds can host guest macromolecules and even enable structure determination by diffraction, though these successes remain limited in scope. Together, these advances suggest that the longstanding vision of high-throughput structure determination using a generalized porous scaffold is now within reach. This review will explore the recent progress in scaffold crystal design, with a particular focus on applications in structure determination, and the challenges that remain for a broadly applicable scaffold-assisted structure determining platform.

1.3. Nucleic Acid Scaffolds

Nucleic acids have long been recognized as versatile components in engineered nanomaterials due to their predictable, programmable base-pairing[14,15], and remarkable capacity for highly-specific self-assembly[16]. Beyond these intrinsic features, DNA is also relatively inexpensive to synthesize and benefits from a wide array of design tools and established protocols, making it an accessible basis for scaffold construction. These advantages make DNA frameworks particularly attractive for development of porous nanomaterials. Strategies developed to date include two- and three-dimensional crystals[17,18] as well as DNA origami scaffolds[19]. Each motif provides a unique approach for organizing molecular assemblies, and together they illustrate the range of possibilities offered by nucleic acid scaffolds.

Of the various nucleic acid scaffolds developed, three-dimensional crystals most closely embody Seeman's original vision[20] of employing DNA as a programmable scaffold nanomaterial for molecular architectures. These crystals are built from rationally designed DNA segments that self-assemble into larger crystallographic frameworks. A particularly noteworthy family of fully designed three-dimensional DNA crystals were based on the tensegrity triangle structural motif[21]. These crystals are formed by assembling seven DNA segments into tiles with pseudo-threefold symmetry. Building on this design motif, subsequent tensegrity tiles have incorporated three- and four- helical turns yielding crystals with increased pore size[20]. However, tensegrity triangle crystals with large tiles and large pores have diffracted to lower resolution (>5Å) limiting their ability to resolve precise molecular interactions. Low resolution concomitant with large solvent channels is consistent with general trends[22] and presumably results from the increased flexibility introduced by longer DNA segments.

Beyond the classic tensegrity DNA tile designs, several other DNA crystal scaffolds have been developed. One such example are the crystals generated by *Paukstelis et al* which incorporate both canonical B-form DNA segments and noncanonical homopurine base pairs[23]. These crystals, containing six-fold symmetry axes and large solvent channels (~9nm diameter), diffracted to 2.1Å have been used to host guest proteins[24] and demonstrated that guest proteins can remain catalytically active[5]. Another example comes from the high resolution porous DNA crystals (2.2Å) engineered by Chengde Mao and colleagues[11]. In the same study, this self-assembling lattice with moderately sized solvent channels (~22Å diameter) was modified to accommodate Hoechst 33342 as a model guest molecule, enabling the structure of Hoechst 33342 to be determined directly within the crystalline framework[11]. Together, these studies highlight the versatility of DNA crystals as porous host materials, capable of achieving atomic-level resolution while simultaneously supporting functional guests. While DNA crystals showcase the potential of nucleic acid frameworks for diffraction-based approaches, DNA scaffolds have also been adapted for cryo-EM, where their programmability and geometric precision can be harnessed to organize challenging targets for imaging.

Moving away from three-dimensional DNA crystals, DNA has also been used to construct DNA origami scaffolds for cryo-EM studies. One pivotal example comes from the Douglas lab, which demonstrated that DNA origami frameworks could be engineered to position proteins at defined orientations within a rigid lattice[25]. This “molecular goniometer” strategy not only enabled the visualization of the small DNA-binding protein BurrH by effectively increasing its apparent size but also allowed the same scaffold design to present the guest protein in multiple orientations, thereby facilitating high-resolution reconstruction[25]. Another notable demonstration was provided by Dietz and colleagues, who used a multilayer DNA origami

platform to resolve an 8 kDa thrombin-binding aptamer. By embedding the aptamer into a rigid framework, they were able to overcome the size limitations of cryo-EM and achieve structural determination of a molecule well below the conventional resolution threshold[26].

1.4. Protein Scaffolds

In contrast to the modest diffraction resolution often achieved with porous DNA scaffolds, protein crystals frequently exhibit higher-quality diffraction even in the presence of substantial solvent content[27]. This advantage stems in part from the chemical and structural diversity of proteins, which provides a far broader palette of building blocks than DNA. Protein crystal scaffolds can be rationally engineered to optimize lattice contacts, enhance stability, and improve diffraction quality[28]. Moreover, advances in computational protein design have enabled the *de novo* construction of proteins that crystallize into predetermined symmetries and architectures, dramatically expanding the design space available for scaffold development. A key example of the latter comes from David Baker and colleagues, who designed crystals with 90% solvent content yet still diffracted to 3.3Å. In the same study, they also engineered crystals with remarkable stability, capable of withstanding autoclaving at 121 °C for 40 minutes without loss of structural integrity[29].

Many protein scaffolds involve the use of fusion tags, antibodies, or small molecules to directly attach guest molecules to the crystal lattice. One such example comes from Maita, who successfully resolved the structure of guest ubiquitin within a porous protein crystal scaffold using single crystal X-ray diffraction[30]. This work highlights the challenges of directly fusing a guest to scaffold building blocks: of the five ubiquitin-fused constructs with different tether lengths tested, only two structures resulted in guest electron density for the ubiquitin fusion.

Nevertheless, a key advantage of this approach is that guest occupancy within the crystal could be precisely controlled, providing a predictable and well-ordered environment for structural analysis. Similarly, *Kojima et al* fused an 11-residue fragment of the intrinsically disordered protein (IDP) c-Myc to a scaffold protein, enabling the determination of its structure via X-ray crystallography at 1.92Å resolution[31]. The use of a scaffold crystal to observe an otherwise inaccessible induced structure is an intriguing capability that is distinct from traditional crystallography.

Protein assemblies have also been leveraged for structure determination via cryo-EM, particularly for targets that are too small or conformationally dynamic to resolve on their own. A prominent example is a self-assembling protein cage decorated with Designed Ankyrin Repeat Proteins (DARPins) from the Yeates lab. By presenting multiple copies of a target protein bound to DARPins in a rigid, symmetric arrangement, this scaffold increases both molecular weight and symmetry to facilitate high-resolution cryo-EM analysis[32]. Using this strategy, Yeates and colleagues resolved the structure of green fluorescent protein (GFP) at 3.8 Å resolution[33], as well as several variants of the oncogenic KRAS protein at ~3.0 Å[34]. However, a practical limitation of this system is that a new DARPin must be engineered for each guest protein of interest, reducing throughput and making it less immediately generalizable than other scaffold strategies.

Beyond their role in structural biology, porous protein crystals have been harnessed as functional materials across bioengineering and nanotechnology. For instance, Michael Chan and colleagues have engineered porous Cry3A protein crystals that naturally form within the bacterium *Bacillus thuringiensis*[35]. Modified Cry3A crystals have been adapted for diverse applications, including targeted cargo delivery to mammalian cells[2–4], intracellular coenzyme immobilization[36], and enhancement of biocatalytic activity *in vivo*[7]. A complementary example

comes from our group's porous protein crystals derived from the *Campylobacter jejuni* periplasmic protein (CJ)[37]. CJ crystals have been used to host nucleic acid barcodes to tag mosquitos[10] and immobilize enzymes[6]. Together, these studies highlight the versatility of porous protein crystals as modular platforms for biotechnology, where their robustness, large solvent channels, and amenability to genetic engineering can be leveraged for functions extending well beyond structural determination.

1.5. Metal Organic Framework Scaffolds

Metal-organic frameworks (MOFs) are three-dimensional crystalline materials generated by coordinating metal ions or clusters with multivalent organic ligands of defined symmetry. The "crystal sponge" method developed by *Fujita et al* adapted the cavities within MOFs to act as adventitious pockets for the structure determination of small molecules[12,38]. Unlike porous protein scaffolds, which often require the careful design of attachment sites for each guest, MOF sponges can accommodate small molecules through spontaneous incorporation under appropriate conditions. Despite the success in using MOFs to determine the structures of small molecules, the existence of large pore MOFs[39], and MOFs that offer sites for guest attachment[40,41], no clear use of MOFs to solve the structures of macromolecules has emerged. While Fujita et al, used MOF cages to encapsulate small proteins, there was no clear solvable guest protein signal[42]. Nevertheless, the development of suitably engineered MOF crystals with large pores and site-specific guest attachment strategies may offer an attractive alternative to the biomolecular scaffolds outlined above.

1.6. Specific Aims:

This dissertation describes the development and application of a novel porous protein–DNA co-crystal (CC1), composed of the replication initiator protein RepE54 and its cognate dsDNA. The work is organized around three major research aims: (1) characterization of CC1 modularity and assembly (Chapter II), (2) optimization of CC1 via protein engineering (Chapter III), and (3) guest protein absorption and structure determination within the CC1 lattice (Chapter IV).

Chapter II demonstrates the modularity of CC1. The crystal pore size was tuned by adding DNA helical turns, and the inserted DNA sequences were highly variable, with fourteen different sequences successfully incorporated into the lattice. Furthermore, CC1 assembly was modulated by varying the position of DNA nick sites within the helix.

Chapter III probes the remarkable tolerance of CC1 to protein mutations and modifications. Mutation of RepE54 residue Ile 116 to cysteine promoted the formation of an intermolecular disulfide bond across a crystallographic protein–protein interface. In addition, both a short peptide tag and separately an entire DNA-binding domain were genetically fused to the C-terminus of RepE54 to probe structural tolerance to terminal modifications. In all three cases, the CC1 crystalline lattice was preserved.

Chapter IV presents the first successful example of a modular scaffold crystal to assist in the post-crystallization soaking and structure determination of biomolecules. Several DNA-binding proteins were soaked into CC1 crystals containing cognate DNA sequences tailored for each guest protein. Their structures were resolved by X-ray diffraction (XRD) without significant loss of resolution.

Chapter V concludes by summarizing the findings and discussing future directions for the CC1 platform.

CHAPTER 2. DESIGN AND CHARACTERIZATION OF POROUS PROTEIN-DNA CO-CRYSTALS FOR BIOMOLECULAR ENGINEERING

Parts of the work in this chapter are formatted for submission as a research article. Writing-drafting and visualization ETS and CDS; conceptualization, analysis, and formal-writing ETS, CKS, FM, ENM, and CDS; data curation, acquisition, investigation, and validation ETS, CKS, CS, CDS.

Shields, E.T., Slaughter, C.K., Mekkaoui, F., Magna, E.N., Shepherd, C., Lukeman, P.S., Spratt, D.E., Snow, C.D. Modular Scaffold Crystals for Programmable Installation and Structural Observation of DNA-Binding Proteins

Pre-print available on BioRxiv: https://doi.org/10.64898/2026.03.04.709581

2.1. Abstract

Crystalline biomaterials composed of protein and DNA offer a unique combination of programmability from nucleic acids and structural precision from protein components. In this chapter, we describe the design and characterization of a novel class of protein–DNA co-crystals composed of the replication initiator protein RepE54 and its cognate DNA (21-mer), termed CC1. CC1 crystals were expanded in an isoreticular manner by inserting additional base pairs into the DNA struts, producing a family of topologically equivalent lattices—CC1**,** $CC1^{+10}$ (31-mer), and $CC1^{+21}$ (42-mer)—with progressively larger solvent channels. Systematic variation of DNA sequence, expansion symmetry, and sticky-end overhang length revealed that the CC1 framework is highly modular, and its assembly can be predictably tuned through rational design of DNA junctions. These results demonstrate that the CC1 family establishes a robust and programmable platform for constructing porous crystalline materials with controllable geometry and composition.

2.2. Introduction

Engineered biomolecular crystals composed of protein and DNA can be highly precise, tunable, and modular. Engineered crystals have been shown to host guest molecules for structure determination[11–13,43], catalyze enzymatic reactions[5–7], deliver cargo to cells[2–4], as molecular sieves[8,44], and as nanoscale devices[9]. We can broadly separate engineered crystals into three classes based on their biomolecular building blocks; DNA crystals, protein crystals, and protein-DNA co-crystals.

DNA is relatively inexpensive and quick to synthesize enabling rapid testing of new material designs. The predictable nature of B-form DNA assembly, driven by canonical base pairing, enables not only the creation of new designs but also allows for tunable growth of materials, offering precise control over their size and structure[14,45,46]. However, it has proven difficult to design and validate DNA crystals that simultaneously diffract to high resolution yet retain the porosity needed for many functional applications. DNA crystals utilize these principles to provide considerable sequence flexibility in their DNA tile components, along with notable porosity and capacity to host DNA-binding guest molecules when the crystal pores are suitable large[20,23,47,48]. In addition, DNA has been used to assemble into origami crystals that can be designed to dynamically reconfigure based on environmental stimuli[49–51], or as scaffolds for structure determination with cryo-electron microscopy(cryo-EM)[25,52]. However, the nanostructure precision of porous DNA crystals, as measured by X-ray diffraction resolution, remains limited.

Compared to DNA crystals, protein crystals have limited modularity and changes to building block proteins can easily disrupt crystallization[53–55]. Recent advances in computational techniques have significantly improved the feasibility of designing protein crystals from the

ground up[56,57]. Despite the challenges of crystallizing modified protein monomers, protein crystals provide useful options for hosting guest proteins compared to DNA crystals(via fusion at the N- or C-termini, accessible thiols, or insertion into loops)[36,58]. Furthermore, using proteins as the fundamental building blocks of crystals unlocks greater control over the physicochemical composition of the material, thanks to the variety of canonical amino acids. The diverse range of amino acid interactions facilitates precise protein packing within crystals as demonstrated by high resolution X-ray diffraction. Despite these advances, purely protein-based crystals lack the inherent modularity of nucleic acids, motivating hybridization strategies.

Herein we present a validated novel class of porous protein-DNA co-crystal. We hypothesized that a porous crystal composed of both protein and DNA could combine the benefits of modular DNA assembly and high precision from protein packing as seen in other co-crystals[59,60]. Our crystal design utilizes continuous double helical DNA struts coordinated by DNA-binding proteins that stack along the crystal x-axis creating a three-dimensional lattice. DNA junctions control crystal growth in the plane perpendicular to the protein strains. The DNA junctions are both literally and figuratively orthogonal to the protein dimension. This allows significant modularity with respect to the DNA strut lengths and the DNA sequences outside the protein-binding region.

We have previously shown that our novel Co-crystal 1(CC1), in its interpenetrating form (I222 space group), can incorporate a variety of DNA sequences, expansion motifs, and modified nucleotides. The interpenetrating crystals can be stabilized with small molecule based DNA ligation and used to host small molecules, nucleic acids, or peptide guests[43,61,62]. However, this interpenetrating CC1(ipCC1) form had limited solvent porosity compared to our hypothetical design limiting the library of possible guests. In contrast, the new porous $CC1^{+10}$ lattice (I121

space group) has larger solvent pores while conserving the same crystallographic interfaces. Like the previous interpenetrating examples, this platform has successfully incorporated a variety of DNA sequences, expansion motifs, and nucleotide modifications. In addition to retaining the modular properties previously expounded upon, this porous platform has successfully been expanded further via the insertion of longer DNA struts (21 nucleotides) increasing the solvent pore size even further than the 10 nucleotide insertions albeit at the cost of molecular precision.

2.3. Results

2.3.1. Engineering and Characterization of a Porous Co-Crystal Lattice

To develop a versatile and guest-accessible crystalline scaffold, we first focused on engineering a co-crystal lattice with large solvent channels. The base scaffold, co-crystal 1 (CC1), is composed of the replication initiator protein RepE54 and its cognate DNA (21-mer). The CC1 lattice is stabilized by stacked DNA-DNA junctions and two unique protein-protein interfaces. Notably, these protein-protein contacts are oriented perpendicular to the DNA-DNA junctions, enabling two-dimensional lattice expansion by inserting additional DNA base pairs at the junction site.

When 10 base pairs were inserted at this junction, the resulting expanded lattice crystallized in space group I222 (unit cell parameters: a = 75.26 Å, b = 132.07 Å, c = 134.99 Å; $\alpha = \beta = \gamma = 90.00°$), as observed in the highest resolution crystal structure (2.38 Å, PDB: 7U6K). The unexpected space group arose from the formation of two interpenetrating copies of the intended lattice, each rotated 180° relative to the other[43] (Figure 2.1A). This interpenetrating

architecture, referred to as ipCC1, resembles previously reported behavior in porous metal-organic frameworks (MOFs)[63].

To isolate a porous version of the $CC1^{+10}$ lattice without interpenetration, extensive crystallization screening using a Gryphon 96-well robot and the Hampton NatrixHT screen was performed. Whereas ipCC1 grew in high magnesium conditions (120–450 mM magnesium chloride, 15–30% PEG 400, and 80–100 mM Tris-HCl pH 8.0), porous $CC1^{+10}$ crystals formed under significantly milder conditions: 10 mM magnesium acetate tetrahydrate, 50 mM sodium cacodylate trihydrate (pH 6.5), and 1.3 M lithium sulfate monohydrate. Equivalent crystals grew upon substituting sodium cacodylate with MES buffer (pH 6.5).

These new porous crystals belonged to space group I121 (unit cell parameters: a = 73.782 Å, b = 132.646 Å, c = 126.768 Å; $\alpha = 90.00°$, $\beta = 90.26°$, $\gamma = 90.00°$), as confirmed by the highest resolution dataset (2.96 Å, PDB: 9YZJ, Figure 2.1). The solvent fraction of this lattice was calculated to be 80%, and the channels were large enough to permit the diffusion of globular guest proteins up to 5.04 nm in diameter, as determined by MAP_CHANNELS[64]. The difference in magnesium concentration between ipCC1 and porous $CC1^{+10}$ crystallization conditions likely reflects the divalent counterion demand for stabilizing close-packed DNA duplexes in the interpenetrated form (Figure S2.1).

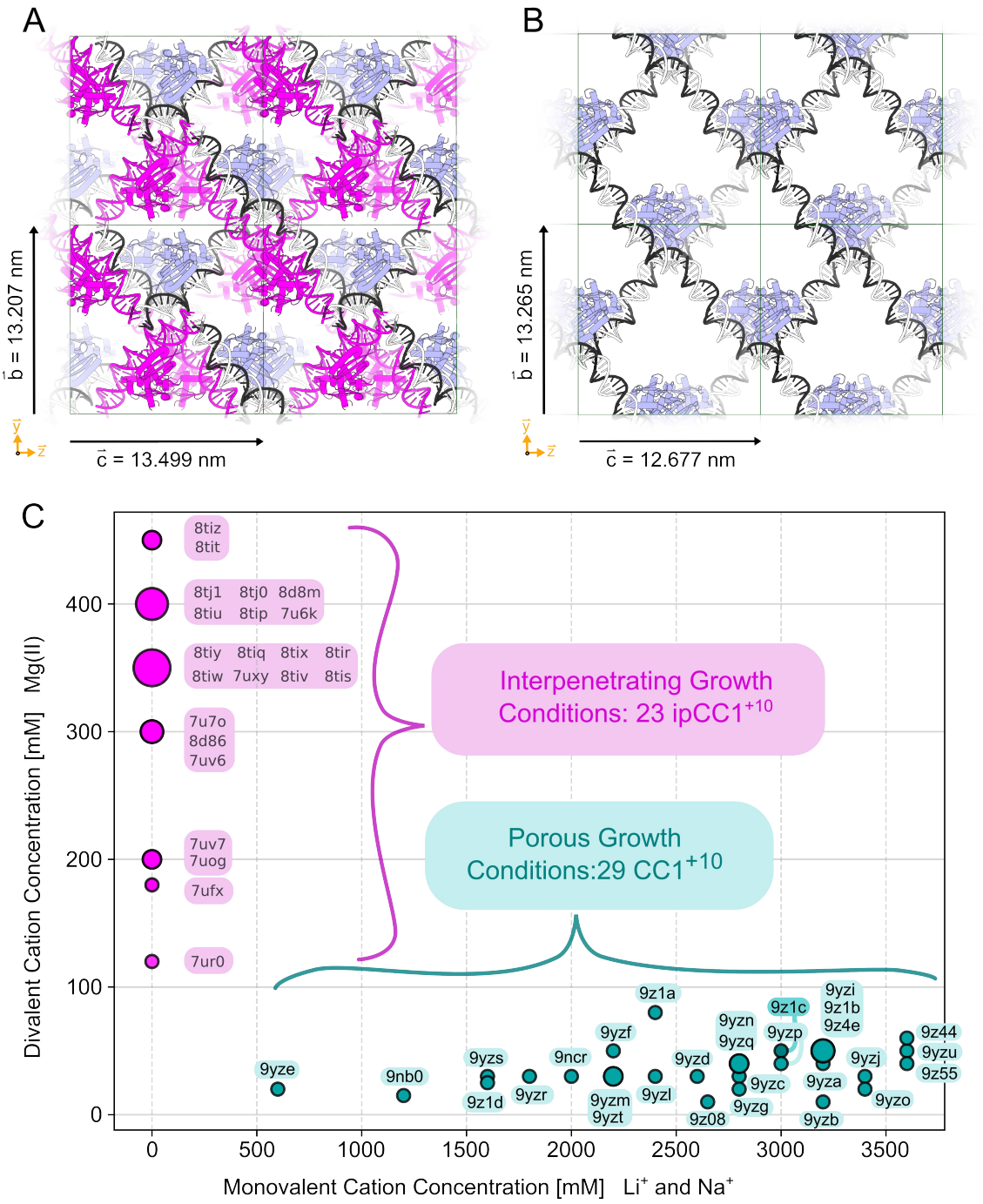


Figure 2.1. Interpenetrating vs Porous CC1 Lattices

(A) Four unit cells from the interpenetrating CC1 lattice (ipCC1^{+10}, I222 space group, PDB 7U6K), where one copy of the lattice (blue, white, black) is interwoven with another copy

rotated by 180° (magenta). **(B)** Four unit cells from the porous $CC1^{+10}$ lattice (I121 space group) shown with the major solvent channels perpendicular to the image (PDB 9YZJ). Close packing of the DNA in the ipCC1 lattice occurs at a higher magnesium concentration. For example, the 7U6K sitting drop growth condition was 400 mM $MgCl_2$, 24% PEG 400, and 80 mM Tris HCl pH 8.0 while the 9YZJ growth condition for sitting drops was 30 mM magnesium acetate, 1.7 M Li_2SO_4, and 50 mM MES pH 6.5. **(C)** Comparison of the divalent and monovalent cation concentrations in the growth conditions for interpenetrating (magenta) and porous crystals (teal). Among porous $CC1^{+10}$ crystals (Table S2) the growth condition magnesium concentration spanned 10-80 mM. The magnesium concentration was much higher for our earlier 23 $ipCC1^{+10}$ PDB entries (spanning 120 mM to 450 mM Mg(II), but most often 300-400 mM). See Figure S3 for a molecular explanation of the role of Mg(II) in the interpenetrating crystal form.

2.3.2. Modularity of the Co-Crystal Scaffold

Having established a porous and stable $CC1^{+10}$ lattice, we next investigated whether the system retained modularity and could accommodate further variations in DNA length and sequence. To this end, we employed two distinct DNA expansion strategies: a symmetric expansion, in which five nucleotides were added to each end of the RepE54 binding sequence, and an asymmetric expansion, where all ten additional nucleotides were placed on one end. Both strategies supported rapid crystal growth (1–7 days), yielding crystals with the same parallelepiped habit and I121 space group as the original porous $CC1^{+10}$.

To evaluate modularity with respect to DNA sequence, we designed a panel of 14 unique DNA inserts with varying GC content and expansion symmetry (Table S2.1). All constructs formed crystals that adopted the $CC1^{+10}$ lattice, although slight differences in optimal

crystallization conditions were observed (e.g., 300 mM lithium sulfate vs. 1.8 M lithium sulfate, Table S2.2). These condition differences could not be attributed solely to sequence variation, as the experiments involved different protein batches and oligonucleotide orders; variation in reagent purity or concentration likely also contributed. Nonetheless, X-ray diffraction revealed nearly identical unit cell dimensions and crystal contacts across all constructs, confirming that the lattice was robust to sequence changes.

To further probe the limits of modularity, we extended the DNA insert to 21 nucleotides, resulting in a 42-bp building block ($CC1^{+21}$). Despite the significant increase in DNA length, $CC1^{+21}$ crystals retained the characteristic parallelepiped habit, preserved all intended protein-protein and protein-DNA interfaces, and crystallized in the expected lattice (Fig 2.2). The final model was refined in space group I121 (unit cell parameters: a = 75.84 Å, b = 163.02 Å, c = 192.28 Å; $\alpha = 90.00°$, $\beta = 98.61°$, $\gamma = 90.00°$) using the highest-resolution dataset available (5.1 Å, PDB: 9YZK). Despite maintaining the correct architecture, $CC1^{+21}$ constructs grew less consistently and required longer periods for growth (7–28 days). Of ten designs tested only two produced crystals, and only one diffracted to sufficient resolution for structure determination (Table S2.3). Nonetheless, the successful growth of CC1, $CC1^{+10}$, and $CC1^{+21}$ crystals demonstrates the robust modularity of this system through controlled isoreticular expansion.

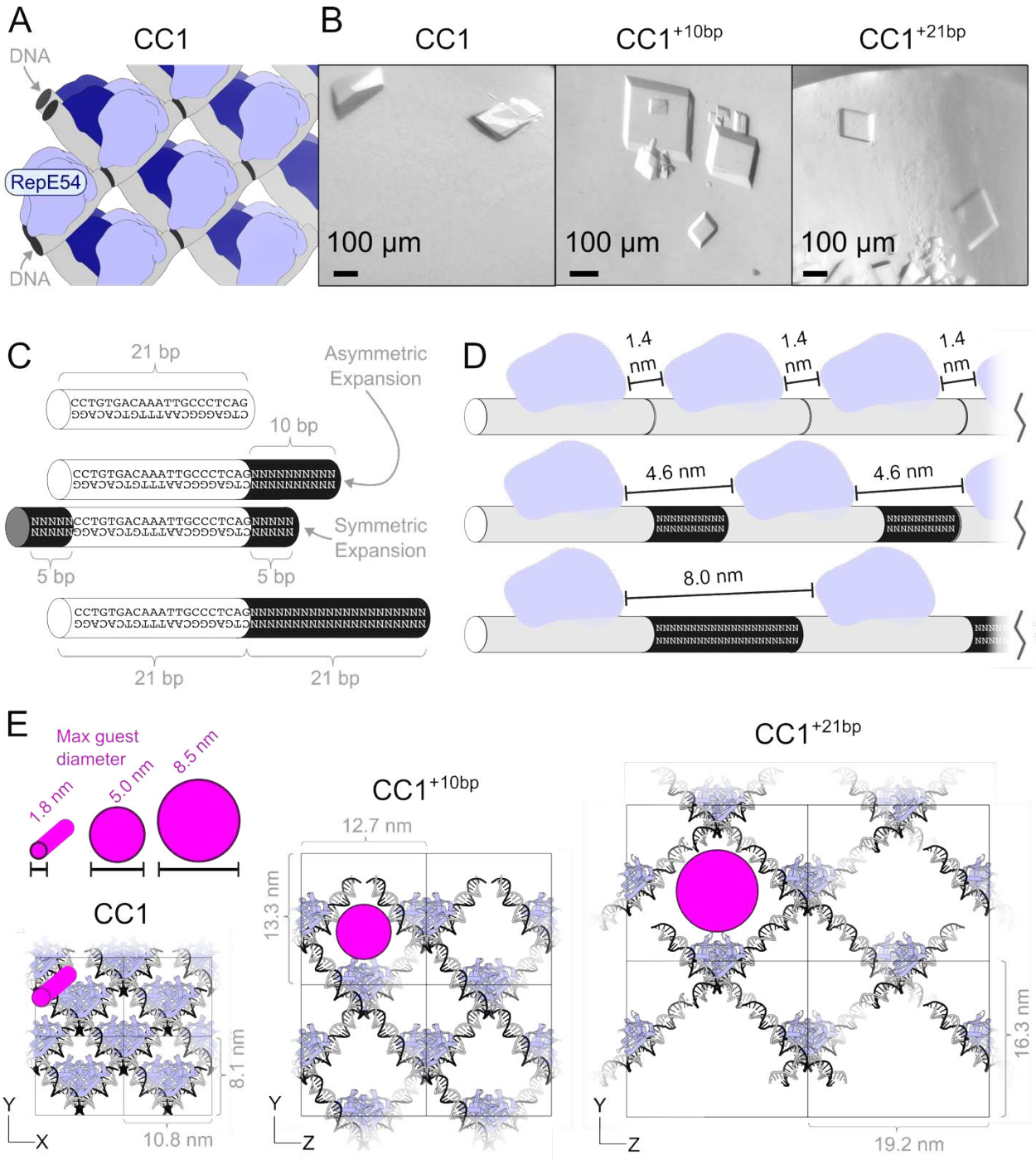


Figure 2.2. Isoreticular Expansion of CC1

(A) CC1 crystals grow via coaxial DNA:DNA interfaces (gray) in two directions and a protein:protein stack (blue shades) in the last direction. **(B)** Light microscope images of

representative CC1 family crystals with characteristic parallelepiped habit. 100 micron scale bars. **(C)** CC1 crystals are expanded by adding new dsDNA (black cylinders) to the parent sequence (white cylinders) that encodes RepE54 binding. **(D)** Expansion along the DNA increases the minimum distance between RepE54 protein (light blue) symmetry mates. **(E)** The CC1 lattice family is shown with magenta discs depicting the maximum diffusible object diameter.

2.3.3. Tunable Assembly and Junction Geometry:

Beyond varying DNA sequence, length, and expansion symmetry, the CC1 scaffold also permits precise tuning of crystal packing through modifications at the DNA–DNA junctions. To examine how base-pairing continuity and stacking geometry influence lattice formation, we varied the number of sticky-end overhangs from 0 to 3 nucleotides (nt). Crystals formed readily for constructs with 0- or 1-nt overhangs, each adopting the intended $CC1^{+10}$ lattice in the *I121* space group with nearly identical unit-cell parameters. Diffraction resolution and overall lattice quality were comparable between these two variants, indicating that short overhangs do not perturb global packing or symmetry.

By contrast, crystals containing 2-nt overhangs grew less frequently and required longer incubation times to reach full size (2-3 weeks), though the resulting crystals maintained the same lattice architecture and diffracted to similar resolution when obtained. The reduced crystallization success and slower growth suggest that extending the overhang to two complementary bases begins to introduce measurable geometric strain, narrowing the energetic window in which ordered assembly is favorable. At the extreme, the 3-nt sticky-end designs failed to crystallize entirely, despite extensive screening across multiple DNA sequences and precipitant concentrations. The reproducible absence of ordered crystals implies that longer

complementary overhangs over-stabilize the DNA junction, introducing torsional strain or misalignment of the RepE54 interfaces necessary for proper lattice propagation. In effect, while short (1–2 nt) overhangs confer additional cohesion compatible with the lattice geometry, excessive base pairing disrupts the delicate balance between the DNA and protein contact networks that maintain crystalline order.

DNA geometry comparisons between the crystal structures were performed using x3DNA[65], revealing subtle but systematic differences in junction conformation across constructs. The helical twist across blunt-ended junctions averaged 31.7° which was much higher than the ideal 20° reported for coaxially stacked but unpaired helices by Grupa *et al* [66] indicating that the blunt configuration accommodates additional rotational strain while maintaining effective base stacking. Constructs with single-nucleotide overhangs exhibited an average twist of 35.0°, closely matching the canonical B-form DNA value of 34.7°, while the single two-nucleotide overhang structure showed an average twist of 38°, consistent with a slightly overwound geometry (Figure S2.3). Taken together, these measurements demonstrate that the $CC1^{+10}$ lattice tolerates a narrow range of junction geometries while retaining crystallographic order.

2.4. Discussion

2.4.1. CC1 vs Other Systems

The CC1 family of crystals achieve a combination of high XRD resolution (~3 Å) coupled with large pores and very high solvent content (~80%). This combination of properties has proven elusive in pure DNA crystals, where the introduction of large pores has typically resulted in limited diffraction quality[22]. We attribute the improved diffraction quality of $CC1^{+10}$

to the elimination of inherently flexible holiday junctions commonly found in DNA crystals, and the "insulating" effect of protein components that both rigidify and separate adjacent DNA duplexes. These design features allow the CC1 framework to retain the geometric programmability of DNA-based lattices while benefitting from the packing precision and chemical stability characteristic of protein crystals. In this way, the CC1 family bridges the divide between DNA and protein crystallization strategies, combining the modularity and design flexibility of DNA scaffolds with the rigidity and diffraction performance of protein lattices to yield a material that is simultaneously programmable, porous, and structurally precise.

Further improvements in the diffraction quality of the CC1 family could likely be achieved through enhanced scaffold rigidity, reduction of lattice defects, and advanced multi-dataset data-analysis strategies. Increasing rigidity either by strengthening protein-protein interfaces, improving cross-links at the DNA junctions, or incorporating more torsionally constrained duplexes could increase local lattice precision and therefore resolution. Similarly, approaches that reduce crystal imperfections, such as optimizing crystal growth conditions or increasing building block purities, would decrease crystal mosaicity[67]. Finally, data analysis improvements could provide further improvements to CC1 resolution. For example, Pearce et al. collected numerous crystal datasets and were able to combine these data sets with advanced data analysis methods to obtain clear electron density maps for small molecule fragments soaked into protein crystals[68]. Another interesting comparison point is to the work of Zhang and coworkers who have grown crystals where the building blocks were entire DNA origami assemblies[69]. While such crystals represent feats of self-assembly and offer intriguing opportunities for programmed installation of third-party molecules (e.g. gold nanoparticles), the resulting crystals have not been of sufficient precision to yield high-resolution diffraction.

2.4.2. Tunable Assembly

A defining strength of the CC1 platform is the ease with which lattice assembly can be tuned through oligonucleotide design. Because DNA components are inexpensive and straightforward to synthesize, it is relatively convenient to produce and test building-block variants that systematically alter junction properties. As described previously[43], the total insert length (e.g. 10 bp) was selected to enable each building block to form Watson-Crick overhangs with flanking building blocks with minimal geometric strain. Since crystal packing in two dimensions is mediated exclusively by DNA–DNA junctions, assembly parameters can be adjusted simply by modifying the sequence, length, or complementarity of the DNA building blocks. This modularity enables controlled variation in lattice spacing, pore diameter, and inter-helical twist without requiring any change to the protein component.

Systematic exploration of DNA variants demonstrated that the CC1 lattice tolerates a broad but clearly defined range of DNA junction configurations. Crystals containing blunt-end junctions formed readily and consistently, indicating that base-stacking interactions alone are sufficient to sustain crystal order. One- or two-nucleotide overhangs also produced the intended lattice, but less reliably in the case of two-nucleotide overhangs. The reduced frequency of crystallization in the two-nucleotide overhang variants, suggests that the incremental increase of complimentary base-pairs between neighboring DNA junctions narrows the energetically favorable window for crystal assembly. Beyond this range, three-nucleotide overhangs failed to crystallize altogether despite extensive screening, implying that over-stabilization of the junction prevents proper alignment of the RepE54 interfaces resulting in the loss of crystal formation.

Although the energetic window for junction stability in the current CC1 design is relatively narrow, several strategies could potentially broaden this range. One approach would be to strengthen the CC1 protein interfaces, thereby offsetting the increased junction strength introduced by longer or more complementary overhangs. Strengthening both CC1 protein-protein interfaces through targeting mutations, could provide additional rigidity, allowing the lattice to tolerate greater torsional strain along the DNA struts. Alternatively, varying the DNA expansion length could reveal junction geometries that are more favorable than the current $CC1^{+10}$ design. Such tests would be most informative for blunt-ended expansions, as alternative expansion lengths (e.g. 9 or 11 bp) would alter rotational phasing without necessarily disrupting base stacking. In contrast, modifying the expansion length in sticky-ended constructs would likely misalign complementary overhangs and interfere with pairing. Together, these approaches would extend the design space of the CC1 lattice and lead to more reliable crystallization of new CC1 variants.

2.5. Conclusion:

The studies presented in this chapter establish the CC1 family of crystals as a versatile and structurally precise platform for modular lattice design. Through systematic variation of DNA insert length, sequence, and junction architecture, we demonstrated that this crystal class maintains order across a broad range of oligonucleotide designs. The combination of programmable DNA components and rigid protein interfaces enables controlled tuning of pore geometry and lattice spacing without altering the overall topology, making these crystals the first example of an isoreticular biomolecular crystal class.

The structural precision and high porosity of the CC1 lattice make it ideally suited for downstream applications such as hosting guest macromolecules. Because the DNA struts are

fully programmable, their sequences can be tailored to incorporate sequence-specific motifs for guest DNA-binding proteins or other nucleic acid–recognizing molecules. The combination of structural order, compositional flexibility, and programmable specificity positions CC1 as a promising host framework for investigating protein–DNA recognition and related biomolecular complexes within an ordered crystalline environment—an opportunity explored in detail in Chapter 4.

2.6. Materials and Methods

2.6.1. Protein Expression and Purification

The protein cloning and expression of the crystal protein Replication initiator RepE54 was described by Komori et al[70]. The CC1 RepE54 was expressed and purified by the Histone source at Colorado State University exactly as described in our previous ligation study[61]. In short, the protein from PDB code 7RVA was overexpressed with an 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 15mg/mL in storage buffer (100 mM sodium citrate pH 6.2, 100 mM KCl, 10 mM $MgCa_2$, and 10 % glycerol) and stored at -80C after flash freezing with liquid nitrogen.

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, 300 mM NaCl, 25 mM imidazole, pH 7.4) and applied to HisTrap (HisPurTM Ni-NTA resin) equilibrated with HisTrap buffer (1X PBS, 300 mM NaCl, 25 mM

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). RepE54(L53G, Q54G, E55G) was stored at 15 mg/mL in CC1 storage buffer and flash frozen for storage at −80 °C

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.

2.6.2. DNA Duplex Annealing

The DNA oligomer sequences used for co-crystallization are listed in Tables S2.1 and S2.3. 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. The oligomers were resuspended in CC1 oligo buffer (50 mM Tris HCl, 100 mM KCl pH 6.7) 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.

2.6.3. Scaffold Protein: DNA Complex Crystallization

The scaffold protein and DNA (1:1.2) were incubated on ice 30 minutes prior to crystallization via sitting drop vapor diffusion. Crystals were grown using vapor diffusion sitting drop on Cryschem plates (Hampton). Crystallization conditions for both the porous lattice ($CC1^{+10}$) and for the large-porous lattice ($CC1^{+21}$) were 20-50 mM Magnesium Acetate, 0.3-2.0M Lithium Sulfate, and 50mM MES pH 6.5. A detailed list of individual crystals, PDB codes

and corresponding growth conditions are in Table S2.2. Crystals grew to a size of 50 - 800 μm^3 in a range of 24 hours to 30 days. Crystal images were taken with a Moticam X5 Plus camera attached to a Motic SMZ-168 stereozoom microscope.

2.6.4. X-Ray Diffraction Data Collection and Refinement

Crystals were flash frozen in liquid nitrogen after briefly transferring the crystal to a cryo-protectant solution (matching the reservoir solution supplemented with 25% glycerol). Single crystal x-ray diffraction data was collected at either ALS beamline 4.2.2 (CMOS detector), beamline 8.2.2 (Pilatus3 detector), or beamline 8.2.1 (Dectris EIGER2 detector) from 0 to 180° with an omega delta of 0.2°and an exposure of 0.5s or 0.25s. Data was processed with XDS and refined using PHENIX and COOT. The first porous $CC1^{+10}$ dataset (pdb code: 9Z08) was refined using new RFree flags, and all subsequent $CC1^{+10}$ structures used the same flags applied. The first $CC1^{+21}$ dataset (pdb code: 9YZK) was refined using new Rfree flags. A detailed list of XRD statistics for each structure can be found in Tables S2.5-S2.13)

2.7. Acknowledgements

We thank Hataichanok (Mam) Scherman, Director of the Histone Source at Colorado State University for the expression and purification of RepE54 transcription factor; Jay Nix at ALS beamlines 4.2.2, 8.2.1, and 8.2.2 for extensive support of XRD data collection; Kay Perry at APS beamline 24-IDE for extensive support of XRD data collection.

2.8. Funding

This material is based upon work supported by the National Science Foundation under Grant No. NSF DMR 2003748 and NSF DMR 2310574.

CHAPTER 3. PROTEIN ENGINEERING WITHIN THE $CC1^{+10}$ LATTICE: STRUCTURAL TOLERANCE AND FUNCTIONAL EXPANSION

3.1. Abstract

The $CC1^{+10}$ lattice provides a modular framework for hybrid material design, yet its capacity to tolerate protein modification has remained largely unexplored. To investigate these limits, we modified the RepE54 protein component through three classes of genetic changes: point mutation, termini modification, and fusion to an additional protein domain. All variants crystallized under standard conditions and preserved the characteristic lattice architecture. As intended, an isoleucine-to-cysteine point mutant (I116C) resulted in a disulfide bond across a crystallographic interface without perturbing packing geometry. The small peptide SpyTag was incorporated successfully, demonstrating that N-terminal modification can be accommodated without disrupting lattice organization. Finally, N-terminal fusion of the 61-residue DNA-binding domain Engrailed homeodomain produced crystals identical to the parent lattice, though the pendant homeodomain domain remained disordered within solvent channels. Collectively, these results reveal that the $CC1^{+10}$ lattice retains structural integrity despite extensive alterations to its protein component, highlighting targeted protein modification of RepE54 as a powerful and previously untapped route to expand $CC1^{+10}$ modularity and functionality for future applications.

3.2. Introduction

Crystals containing proteinaceous components represent a unique ordered environment in which molecular interactions can be precisely defined, making them a privileged target for rational engineering. Protein crystal engineering has been used to stabilize existing lattices and to introduce new functions. Stabilization strategies have included introducing disulfides[71–77], and

other surface or interface mutations to improve diffraction quality[28,78,79]. Other studies have sought to assist with the crystallization of other proteins or to functionalize crystals with peptide modifications via genetic fusion[30,80,81]. Recent advances in computational protein design and structure prediction have made these engineering goals increasingly attainable[29,82].

Here, we evaluated the tolerance of the $CC1^{+10}$ lattice to genetic modifications of the protein component, RepE54, within the $CC1^{+10}$ lattice. This modular lattice, previously shown to form robustly through engineered DNA-mediated contacts and protein–protein interfaces, offers an attractive platform for modular design. To date, most modifications to the $CC1^{+10}$ lattice have focused on the nucleic acid component, including variations in sequence, insert length, and junction geometry, while protein engineering has remained largely unexplored. To probe the possibility of protein engineering within this system, we introduced a series of targeted genetic modifications to RepE54, encompassing point mutations intended to stabilize crystallographic interfaces through disulfide bond formation, short N-terminal peptide tags to test the tolerance of terminal modification and potential functionalization sites, and larger domain fusions that append an entire folded DNA-binding domain to evaluate the upper size limit of tolerated insertions.

Engineered protein variants consistently assembled into co-crystals under standard conditions, despite substantial sequence changes to RepE54. The resulting crystals preserved the defining $CC1^{+10}$ lattice architecture, maintaining the I121 space group and the network of DNA- and protein-mediated contacts that underpin lattice organization. Even the most ambitious modification—the full DNA-binding domain fusion—retained the same crystallographic packing observed in the parent structure. These results demonstrate that the CC1 scaffold is structurally robust, maintaining overall lattice integrity despite extensive alterations to its protein component. This resilience broadens the utility of $CC1^{+10}$ as a modular framework for the design of

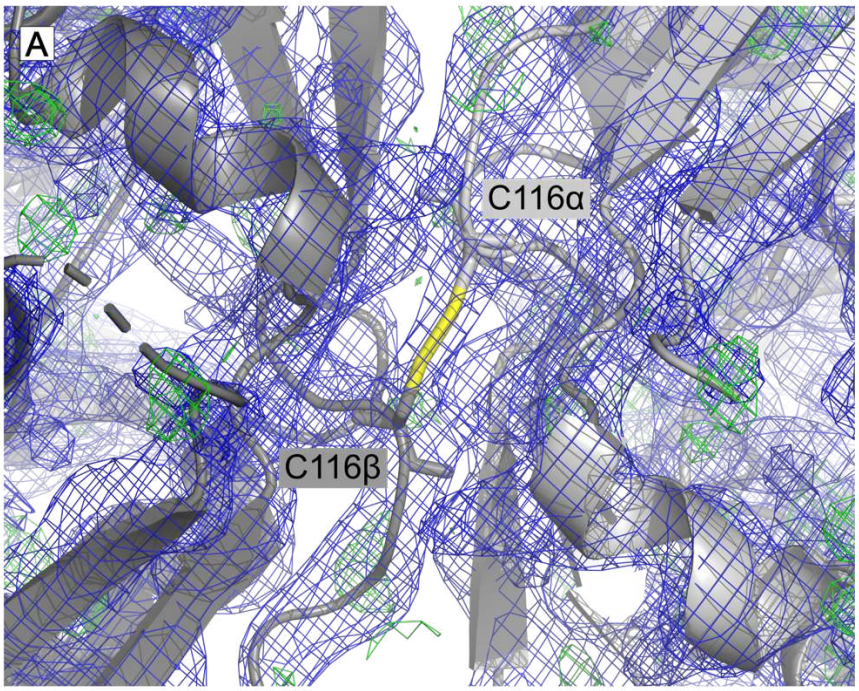
functional and responsive crystal architectures. The $CC1^{+10}_{SpyTag}$ scaffold offers the intriguing possibility of the post-crystallization covalent installation of any guest protein that is fused to the SpyCatcher domain[83].

3.3. Results and Discussion

3.3.1 Disulfide Mutation

To test whether the $CC1^{+10}$ lattice could accommodate targeted point mutations at crystallographic interfaces, we introduced a series of cysteine substitutions (5 protein variants in total) intended to promote disulfide bond formation across symmetry-related contacts. Beyond simply demonstrating the lattice's tolerance to point mutations, this strategy also aimed to explore whether such modifications could stabilize the lattice, ultimately enabling broader applications for this platform in structural biology and molecular scaffolding. Candidate sites were manually selected based on several criteria: favorable geometry, interaction energy, appropriate Cβ–Cβ distances for disulfide bond formation (as predicted by Rosetta[84] energy minimization), and interactions between identical residues on symmetry-related copies of RepE54 in the canonical $CC1^{+10}$ crystal structure. To increase confidence in the designs, AlphaFold-3[82] structural predictions of the protein mutants were used to validate the feasibility of each proposed mutation (Figure S3.1). Ultimately five residues (G91, I116, I180, I227, and V239) were selected as cysteine mutation. Because the crystal lattice contains two distinct protein-protein interfaces, mutations were introduced both individually and as matched pairs to assess crosslinking potential across all relevant contacts. (Figure 3.1)

All designed protein variants except those containing the G91C mutation were successfully expressed in *E. coli* and purified (Protocol 3.5.2). Crystallization trials were then performed on the remaining mutants under standard $CC1^{+10}$ conditions (20–50 mM magnesium acetate, 0.3–2.0 M lithium sulfate, and 50 mM MES pH 6.5). Among these, only the I116C single mutant formed crystals despite extensive crystallization efforts (>10 crystallization experiments each). Proving a negative is difficult; the other variants might still crystallize with continued effort. For the purposes of the current report, we will focus entirely on the I116C variant. To verify that the engineered Cys116 residue was solvent-accessible prior to disulfide formation, purified RepE54-I116C protein was assayed with Ellman's reagent (DTNB), which produces a chromophore upon reaction with free thiols (Figure S3.2). The observed increase in absorbance relative to wild type confirmed the presence of an additional accessible cysteine compared to the three cysteines present in wild-type RepE54 (Figure 3.1). XRD analysis of the resulting crystals showed that the $CC1^{+10}$ lattice was preserved. Moreover, there was clear electron density ($2F_C$-F_0) connecting the I116C residues across the crystallographic interface and negligible difference density (F_C-F_0) when the site was modeled as a disulfide bond.. Detailed examination of the structure revealed an Sγ–Sγ distance of 2.06 Å and a Cα–Cβ–Sγ angle of 117.1°, both consistent with canonical disulfide bond geometry[85].


A
C116α
C116β

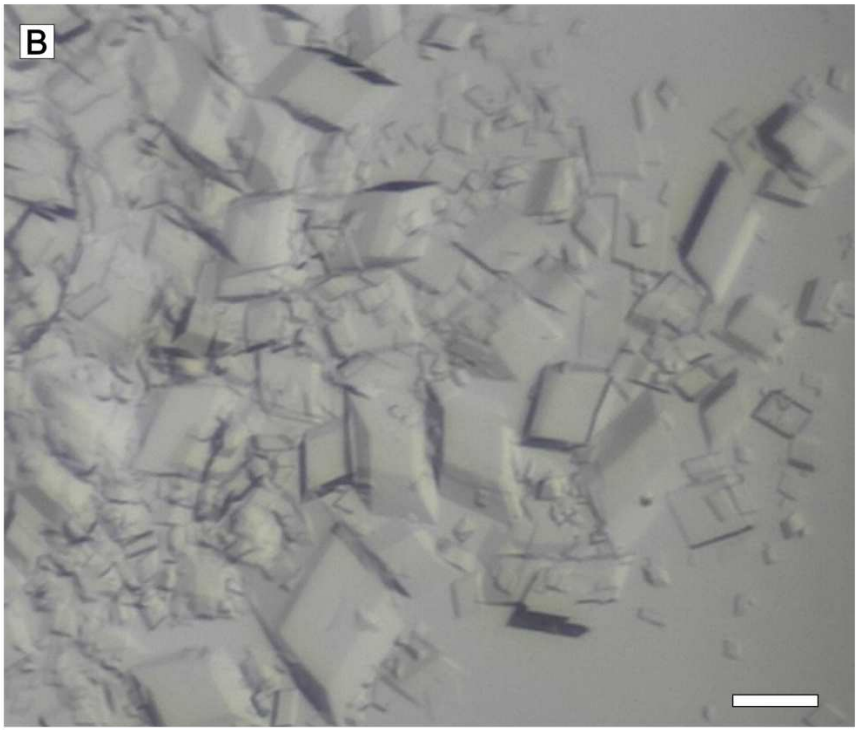

B

Figure 3.1. RepE54-I116C CC1^{+10} structure and crystal habit

(A) Crystal structure of $CC1^{+10}_{I116C}$ inter-protein disulfide bond (PDB:9Z41), $2F_C$-F_0 map (σ=1, blue), and F_C-F_0 map (σ=4, green). (B) Examples of RepE54-I116C CC1^{+10} crystals scalebar = 100 microns.

Stability trials of the resulting $CC1^{+10}_{I116C}$ crystals did not reveal a measurable increase in overall crystal stability with respect to CC1^{+10}, which is unsurprising given that the disulfide bond was introduced across only one of the two key protein-protein interfaces. Future efforts could explore additional stabilizing mutations at the protein-protein interfaces beyond cysteine crosslinks, such as more optimized hydrophobic contacts, more extensive hydrogen-bond networks or engineered salt bridges (e.g., introducing complementary acidic and basic residues like glutamate and lysine). Designed salt-bridge sites are particularly appealing because they could be converted to covalent amide bonds using carbodiimide conjugation with EDC (1-ethyl-3-(3-dimethylaminopropyl)carbodiimide). We already apply EDC to these crystals for chemical ligation of the DNA junctions[61,62]. While enhanced lattice stability is a project for the future, our results demonstrate one instance of the CC1^{+10} lattice tolerating a rationally designed point mutation at intermolecular interfaces without disrupting crystallization, a notable feat given the extreme sensitivity of crystal growth to small modifications[28,54].

3.3.2. N-terminal Peptide Fusion

To further evaluate the tolerance of the CC1^{+10} lattice to genetic modification, we designed a RepE54 variant bearing an N-terminal SpyTag peptide: $CC1^{+10}_{SpyTag}$. The short 13-amino acid SpyTag sequence can form a covalent isopeptide bond with its binding partner, SpyCatcher, and is commonly used in modular protein assembly and labeling[83]. To minimize

potential structural interference, the SpyTag was fused to the RepE54 N-terminus via a short flexible linker consisting of glycine-glycine-serine repeats $(GGS)_n$. Alphafold-3[82] was used to optimize linker length and ensure that the SpyTag did not perturb RepE54 folding and remained solvent-exposed (Figure S3.3). In addition to the N-terminal SpyTag insertion (AHIVMVDAYKPTK), the existing N-terminal solubility and histidine purification tag composed of 21 amino acids was removed and a hexahistidine purification tag (HHHHHH) was inserted to the C-terminus of RepE54. Thus, successful crystallization of the $CC1^{+10}_{SpyTag}$ variant demonstrated that both the N- and C-termini of RepE54 are permissive to significant modification.

The SpyTag–RepE54 fusion protein was successfully expressed, purified, and co-crystallized with its cognate DNA under standard $CC1^{+10}$ conditions (20–50 mM magnesium acetate, 0.3–2.0 M lithium sulfate, 50 mM MES pH 6.5). Crystals formed readily and adopted the same I121 space group and lattice packing as unmodified $CC1^{+10}$, indicating that the N-terminal peptide extension was well-tolerated. No electron density corresponding to SpyTag or the flexible linker was observed, consistent with a disordered, solvent-exposed tag (Figure 3.2). This was consistent with our design intent that SpyTag should remain flexible and not participate in crystallographic contacts or interfere with lattice packing.

Attempts to confirm *in crystallo* covalent isopeptide bond formation between SpyTag and SpyCatcher were inconclusive. Our standard EDC crosslinking protocol is likely incompatible with the SpyTag sequence, as the first step in typical EDC conjugation is a reaction with carboxylic acids, and the SpyTag–SpyCatcher chemistry utilizes the carboxylic acid present on Asp7 of the SpyTag (Figure S3.4). Testing the reaction on a non-crosslinked crystal was not possible, as $CC1^{+10}$ crystals dissolved rapidly outside of growth conditions in the absence of

EDC chemical ligation. However, we verified functional SpyTag activity in solution: when incubated with a SpyCatcher–Green Fluorescent Protein fusion (SC-GFP), formation of a covalent complex was observed by SDS–PAGE, confirming that the SpyTag remained accessible and reactive in the fusion protein (Figure S3.5). Although *in crystallo* bond formation could not be confirmed, these results demonstrate that the SpyTag–RepE54 construct retains functionality and that the $CC1^{+10}$ lattice can accommodate peptide extensions without loss of crystal formation. Future work should explore alternative crosslinking agents that are compatible with the SpyTag–SpyCatcher system, such as aldehyde-based crosslinkers. Additionally, fusing the SpyCatcher portion of the covalent peptide system directly to RepE54 could circumvent the issues with EDC; however, initial attempts using the same flexible linker design workflow to synthesize $CC1^{+10}_{SpyCatcher}$ did not yield an expressible protein.

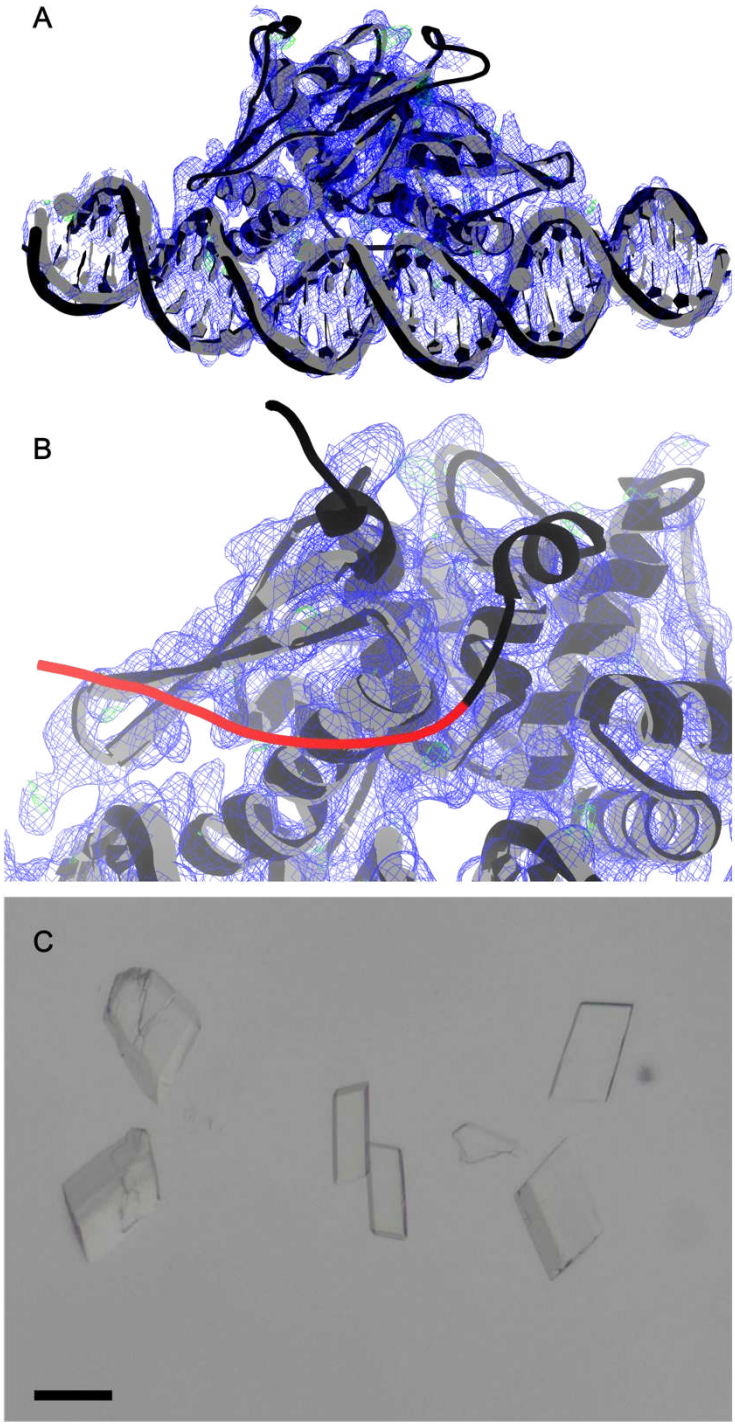

A
B
C

Figure 3.2. SpyRepE $CC1^{+10}$ structure and crystal habit

(A) Crystal structure of SpyRepE fusion protein $CC1^{+10}$ (grey, PDB:9Z42) aligned with AlphaFold-3 prediction of the same structure (black), $2F_C$-f_0 map (σ=1, blue), and F_C-F_0 map (σ=4, green). (B) Close-up view of SpyRepE $CC1^{+10}$ (grey, PDB:9Z42) aligned with FlphaFold-3 prediction of the same structure (black) with one candidate position for SyTag illustrated in red (without observable electron density), $2F_C$-F_0 map (σ=1, blue), and F_C-F_0 map (σ=4, green). (C) Examples of SpyRepE $CC1^{+10}$ crystals. Scalebar = 200 microns.

3.3.3. DNA Binding Domain Fusion

To test the tolerance of the CC1+10 lattice to larger genetic insertions, we fused an entire DNA-binding domain to the N-terminus of RepE54. Specifically, we selected the Engrailed homeodomain (EnH) from *Drosophila melanogaster*, a well-characterized transcription factor consisting of 61 amino acids. EnH was selected for its ability to bind dsDNA with a clear preference for TA rich sequences[86]. The EnH-RepE54 N-terminal fusion was designed using the same flexible linker strategy ($(GGS)_{n=4}$) to preserve the structural independence of both domains and minimize potential interference with crystallographic contacts (Figure S3.6). AlphaFold3 modeling (Fig. S3.6.) predicted that the homeodomain could bind the major groove of the expanded $CC1^{+10}$ DNA strut in several possible DNA binding registers, without disrupting protein-protein or DNA-DNA contacts within the lattice.

The EnH-RepE54 fusion protein was successfully expressed, purified, and co-crystallized with two of the four DNA designs under standard $CC1^{+10}$ crystallization conditions (20–50 mM magnesium acetate, 0.3–2.0 M lithium sulfate, and 50 mM MES pH 6.5). Crystals formed readily in both cases and maintained the same I121 space group and lattice packing as the original scaffold, indicating that the addition of a folded domain at the protein N-terminus is compatible

with $CC1^{+10}$ lattice formation. The intended dsDNA binding sites were resolved as well as expected given the 3.1 Å resolution and the elevated B-factor in the solvent exposed dsDNA struts. However, we were surprised to see no obvious electron density corresponding to the Engrailed homeodomain in either structure (Figure 3.3). The absence of discernible density suggests that the appended domain remains flexible or adopts multiple conformations that are not constrained by the lattice environment under the solvent conditions present during XRD data collection. Notably, when a very similar guest protein (even-skipped homeodomain) was soaked into $CC1^{+10}$ crystals rather than genetically fused to the scaffold, we were able to observe electron density for the untethered homeodomain (Chapter 4). Future efforts will determine if an alternative linker length ($(GGS)_{5-7}$) permits the tethered homeodomains to bind. In the meantime, the ability of the $CC1^{+10}$ scaffold to accommodate large, independently folded protein domains at the RepE54 N-terminus without perturbing lattice organization is an impressive feat of modularity.

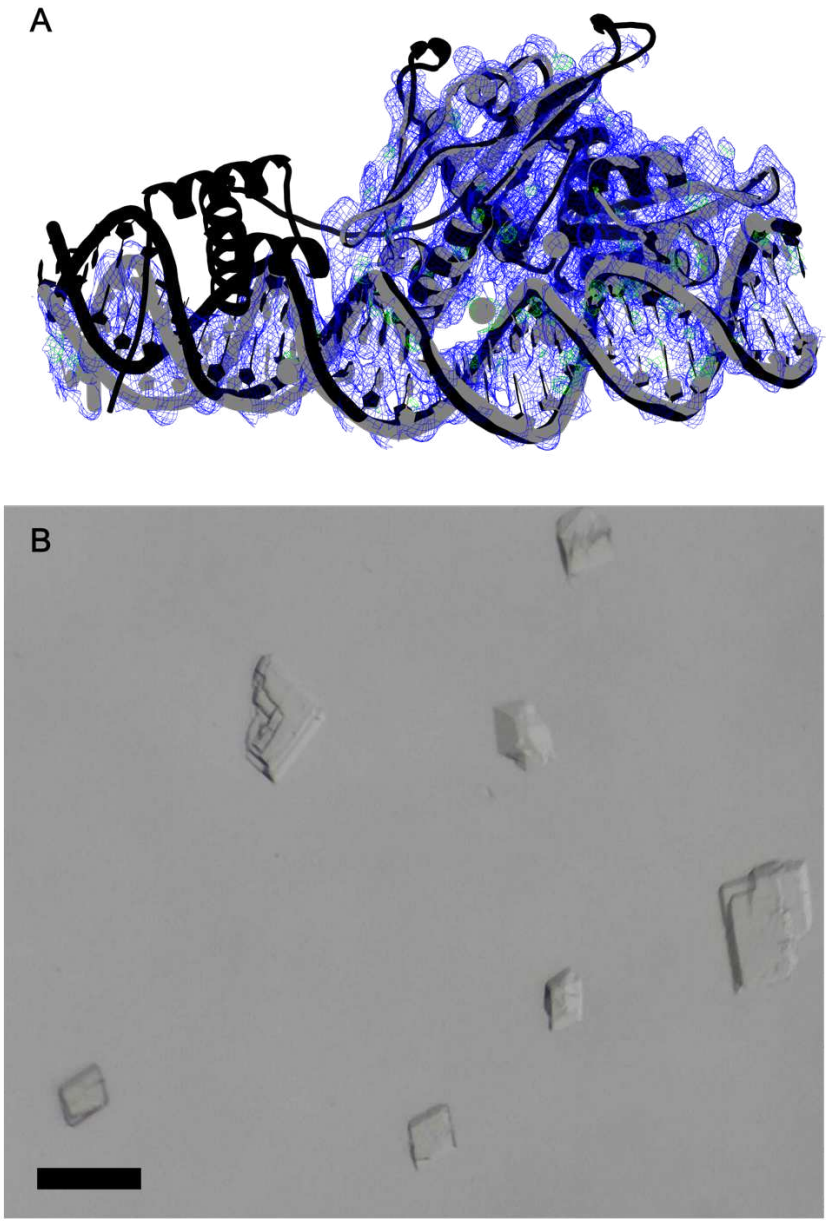


Figure 3.3. EnH-RepE CC1^{+10} structure and crystal habit

(A) Crystal structure of EnH-RepE fusion protein CC1^{+10} (grey, PDB:9Z43) aligned with AlphaFold-3 prediction of the same structure (black), 2F_C-F_0 map (σ=1, blue), and F_C-F_0 map (σ=4, green). (B) Examples of EnH-RepE CC1^{+10} crystals. Scalebar 200 microns.

3.4. Conclusions

The studies presented in this chapter demonstrate that the $CC1^{+10}$ lattice is tolerant to protein engineering, maintaining its overall crystallographic architecture despite substantial sequence and structural modifications to the RepE54 protein component. Point mutations designed to form disulfide crosslinks, insert peptide tags, and insert an entire domain all yielded co-crystals that preserved the $CC1^{+10}$ lattice and interface geometries. This tolerance has important implications for both the stability and functionality of the CC1 family of crystals. The ability to modify RepE54 through targeted mutations or terminal extensions provides new routes to enhance lattice rigidity, introduce chemical handles, or append functional domains while retaining crystalline order.

Enhancing lattice rigidity through targeted protein engineering complements the design strategies described in chapter 2, where variations in DNA geometry revealed the lattice's inherent mechanical limits. Introducing a stabilizing disulfide crosslink within the RepE54–RepE54 interface offers a route to reinforce the scaffold from the protein side, potentially reducing the subtle conformational fluctuations that limit crystal resolution. Future efforts combining protein-based stabilization with DNA-level tuning may further improve lattice order and extend the resolution range attainable in CC1 crystals.

At the same time, the successful incorporation of peptide tags and domain fusions establishes a framework for introducing functional diversity into the CC1 lattice. The ability to append short peptides or folded domains without perturbing crystal packing provides a direct means to recruit or orient guest proteins within the porous framework. This modular compatibility sets the stage for sequence-specific guest capture and in-crystal complex

formation, as explored in the following chapter. Together, these findings position CC1 as both a stabilizable and functional scaffold, bridging the transition from a structural material to an active host for protein–DNA interactions.

3.5. Materials and Methods

3.5.1. Protein Expression and Purification

The CC1 protein disulfide variant RepE54 (I116C) was cloned into the PetDuet plasmid (Novagen) using Quickchange mutagenesis (Quikchange II site-directed mutagenesis kit, Agilent). The protein was expressed with a T7 promoter in E. coli BL21(DE3) cells. Upon addition of 1.0 mM IPTG, the cells were outgrown at 28 °C for 20 h. The cell pellets were sonicated in lysis buffer (1X PBS, 300 mM NaCl, 25 mM imidazole, pH 7.4) and applied to a HisTrap HP purification column (Cytiva) equilibrated with lysis buffer. Protein was washed and eluted using a linear gradient 25-500mM imidazole in lysis buffer. Following HisTrap purification, the fraction containing RepE54-I116C were combined and dialyzed into heparin buffer (1X PBS pH 7.4, 50 mM NaCl) using Snakeskin dialysis tubing (Thermo Scientific) with a 10kDa MWCO. After dialysis RepE54-I116C was applied to a HiTrap Heparin HP column (Cytiva) equilibrated with heparin buffer. Protein was washed and eluted using a linear gradient of 50 mM-500 mM NaCl in heparin buffer. Aliquots containing RepE54-I116C were pooled and concentrated to 8 mg/mL using Amicon Ultra-15 10 kDa MWCO centrifugal filter unit (EMD Millipore) and stored at -80 °C.

The CC1 protein variant SpyTag-RepE54 (SpyRepE) was cloned into the PetDuet plasmid (Novagen) using Gibson cloning (Hi-Fi assembly New England Biolabs). The protein was expressed with a T7 promoter in E. coli BL21(DE3) cells. Upon addition of 1.0 mM IPTG,

the cells were outgrown at 28 °C for 20 h. The cell pellets were sonicated in lysis buffer (1X PBS, 300 mM NaCl, 25 mM imidazole, pH 7.4) and applied to a HisTrap HP purification column (Cytiva) equilibrated with lysis buffer. Protein was washed and eluted using a linear gradient 25-500mM imidazole in lysis buffer. Following HisTrap purification, the fraction containing SpyRepE were combined and dialyzed into heparin buffer (1X PBS pH 7.4, 50 mM NaCl) using Snakeskin dialysis tubing (Thermo Scientific) with a 10kDa MWCO. After dialysis SpyRepE was applied to a HiTrap Heparin HP column (Cytiva) equilibrated with heparin buffer. Protein was washed and eluted using a linear gradient of 50mM-500mM NaCl in heparin buffer. Aliquots containing SpyRepE were pooled and concentrated to 20 mg/mL using Amicon Ultra-15 10 kDa MWCO centrifugal filter unit (EMD Millipore) and stored at -80 °C.

The CC1 fusion protein EnH-RepE was synthesized into the pET28 plasmid (Twist Bioscience). The protein was expressed with a T7 promoter in E. coli BL21(DE3) cells. Upon addition of 1.0 mM IPTG, the cells were outgrown at 28 °C for 20 h. The cell pellets were sonicated in lysis buffer (1X PBS, 300 mM NaCl, 25mM imidazole, pH 7.4) and applied to a HisTrap HP purification column (Cytiva) equilibrated with lysis buffer. Protein was washed and eluted using a linear gradient 25-500mM imidazole in lysis buffer. Following HisTrap purification, the fraction containing EnH-RepE were combined and dialyzed into heparin buffer (1X PBS pH 7.4, 50 mM NaCl) using Snakeskin dialysis tubing (Thermo Scientific) with a 10kDa MWCO. After dialysis EnH-RepE was applied to a HiTrap Heparin HP column (Cytiva) equilibrated with heparin buffer. Protein was washed and eluted using a linear gradient of 50mM-500mM NaCl in heparin buffer. Aliquots containing EnH-RepE were pooled and concentrated to 9 mg/mL using Amicon Ultra-15 10 kDa MWCO centrifugal filter unit (EMD Millipore) and stored at -80 °C.

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**. 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. The oligomers were resuspended in CC1 oligo buffer (50 mM Tris HCl, 100 mM KCl pH 6.7) 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.

3.5.3. Scaffold Protein: DNA Complex Crystallization

The scaffold protein and DNA were mixed in a 1.2:1 ratio (440:364 μM, 55 μL total) and were incubated on ice for 30 minutes prior to crystallization via sitting drop vapor diffusion. Crystallization trials used CrysChem sitting-drop plates (Hampton Research HR3-159), with 400 μL reservoir solution per well and 4 μL crystallization drops prepared by mixing 2 μL each of the protein–DNA complex and reservoir solution directly on the platform; drops were mixed by gentle pipetting. Plates were sealed immediately after setup and incubated at 25 °C under static conditions. Crystallization conditions for both the porous lattice ($CC1^{+10}$) and for the large-porous lattice ($CC1^{+21}$) were 10-80 mM Magnesium Acetate, 0.3-1.8 M Lithium Sulfate, and 50 mM MES pH 6.5 (or occasionally 50 mM sodium cacodylate pH 6.5). A detailed list of

individual crystals, PDB codes, and corresponding growth conditions are reported in Table S2. Crystals grew to full size in a range of 24 hours to 30 days (most often in 1-2 days). Crystal images were taken with a Moticam X5 Plus camera attached to a Motic SMZ-168 stereozoom microscope.

3.5.4. X-Ray Diffraction Data Collection and Refinement

Crystals were flash frozen in liquid nitrogen after briefly transferring the crystal to a cryo-protectant solution (matching the reservoir solution supplemented with 25% glycerol). Single crystal x-ray diffraction data was collected at either ALS beamline 4.2.2(CMOS detector), beamline 8.2.2(Pilatus3 detector), or beamline 8.2.1(Dectris EIGER2 detector) from 0 to 180° with an omega delta of 0.2°and an exposure of 0.5s or 0.25s. Data was processed with XDS and refined using PHENIX and COOT. The first porous $CC1^{+10}$ dataset (pdb code: 9Z08) was refined using new R_{Free} flags, and subsequent structures used the same flags applied.

3.5.5. EDC Ligation

Co-crystals were washed in conditions comparable to their growth conditions (1.2-1.6M Li_2SO_4, 50mM $MgAc_2$, and 50mM MES pH 6.5) but lacking free protein and DNA blocks. The co-crystals were washed in 9-well glass plates (Hampton) to remove excess protein and DNA monomers. 1-Ethyl-3(3-dimethylaminopropyl)carbidiimide (EDC) (Advanced Chemtech CAS#:25952-53-8) was resuspended in the crosslinking solution (1.2-1.6M Li_2SO_4, 50mM $MgAc_2$, and 50mM MES pH 6.0) to final concentration values of 30-60 mg/mL and used immediately. The co-crystals were crosslinked in a 200µL EDC solution volume overnight (16-24 hrs). The crosslinking reaction was ended by moving the co-crystals back into the wash solution.

3.5.6. Ellman's Assay

Ellman's reagent (DTNB) was suspended in Reaction buffer consisting of 100mM sodium phosphate pH 8.0 and 1mM EDTA to a final concentration of 10 mM DTNB (Ellman's buffer). A standard curve was created by serially diluting L-cysteine from 1.6 mM to 0.05 mM (with an additional blank) in Reaction buffer. Each protein tested had its concentration normalized to 0.4 mM by diluting in their storage buffer. All samples (including L-cysteine samples) were mixed with 50uL of Ellman's buffer (0.18 mM final DTNB concentration) and incubated at room temperature for 15 minutes. Following incubation, the absorbance of the samples at 412nm was measured using a Synergy H1 Hybrid plate reader (BioTek). All samples were created in triplicate, and the resulting measurements were plotted in RStudio (Figure 3.2).

3.6. Acknowledgements

We thank Hataichanok (Mam) Scherman, PhD., Director of the Histone Source at Colorado State University for the expression and purification of RepE54 transcription factor; Jay Nix, PhD., at ALS beamlines 4.2.2, 8.2.1, and 8.2.2 for extensive support of the XRD data collection; Jacob Deroo, PhD., for assistance with Rosetta design calculations of RepE54 cysteine mutation sites.

3.7. Funding

This material is based upon work supported by the National Science Foundation under Grant No. NSF DMR 2003748 and NSF DMR 2310574.

CHAPTER 4. POST-CRYSTALLIZATION GUEST BINDING AND STRUCTURE DETERMINATION WITHIN A PROTEIN-DNA CO-CRYSTAL

Parts of the work in this chapter are formatted for submission as a research article. Writing-drafting and visualization ETS and CDS; conceptualization, analysis, and formal-writing ETS, CKS, FM, ENM, and CDS; data curation, acquisition, investigation, and validation ETS, CKS, CS, CDS.

Shields, E.T., Slaughter, C.K., Mekkaoui, F., Magna, E.N., Shepherd, C., Lukeman, P.S., Spratt, D.E., Snow, C.D. Modular Scaffold Crystals for Programmable Installation and Structural Observation of DNA-Binding Proteins

Pre-print available on BioRxiv: https://doi.org/10.64898/2026.03.04.709581

4.1. Abstract

Inducing trillions of biomolecules to self-organize into a regular crystal lattice is a notorious technical challenge that is seldom amenable to rational approaches. Instead, most practitioners find biomolecular crystals via brute force screening, an approach that often fails[53]. Over four decades ago, Seeman proposed that a porous scaffold crystal could serve to capture guest macromolecules, adding these domains to the lattice and therefore facilitating guest structure determination via X-ray diffraction[1]. Seeman proposed using designed DNA assemblies as such a scaffold. Numerous crystals composed of designed DNA tiles now exist in the Protein Data Bank, but crystals of this type with sufficiently large solvent channels to permit post-crystallization addition of guest proteins (e.g. >5 nm solvent channels) tend to diffract only to low-resolution (>5Å). Meanwhile, despite the general correlation between solvent content and poor diffraction, select porous protein crystals diffract relatively well[27]. In one remarkable case, Maita[30] genetically fused ubiquitin to the scaffold protein R1EN and was able to observe

ubiquitin in the X-ray structure, albeit in multiple conformations and only when the flexible tether between the scaffold protein and ubiquitin had a particular length. In general, genetic fusion risks disrupting crystal growth of the lattice. We have been pursuing a hybrid approach that (1) preserves the separation of scaffold crystal growth and engineering from the process of guest installation, (2) avoids the low-resolution diffraction of porous DNA crystals, and (3) retains the modular programmability benefits of incorporating DNA building blocks. Here we show that the resulting engineered family of protein-DNA co-crystals have sufficient modularity to permit growth of many distinct porous scaffold crystals, that the resulting crystals can be stabilized, that the stabilized crystals can diffract under a variety of conditions including soaking in diverse DNA-binding domains, and that these guest domains bind with sufficient affinity and specificity to be observed via X-ray diffraction. The resulting system may facilitate the rapid and high-throughput structure determination of numerous DNA-binding domains.

4.2. Introduction

X-ray crystallography remains the most widely used technique for determining macromolecular structures at atomic resolution. Despite continuous advances in automation, screening, and computational prediction[29], most crystallization campaigns fail to yield diffracting crystals[54]. One strategy to overcome these limitations is to use pre-formed crystalline scaffolds that provide an ordered, solvent-accessible framework capable of hosting guest molecules for structure determination. This concept has been applied in both X-ray crystallography and cryo-electron microscopy to create ordered scaffolds that assist in the structure determination of proteins.

DNA-based scaffolds have been used to host both small molecules and proteins[5,11,24]. Yeates and colleagues extended this concept by employing a DNA origami scaffold to resolve

the structure of the small DNA-binding protein BurrH[32–34], illustrating how programmed nucleic acid architectures can serve as precision alignment tools for crystallographic or cryo-EM analysis. In addition, protein-based scaffolds have also been used as scaffolds for structure determination. Maita solved the structure of ubiquitin by genetically fusing it to their R1EN crystal framework[30], and Kojima et al. used a similar strategy to determine the structure of a fragment of the intrinsically disordered protein c-Myc[31]. Despite these advances, such successes remain rare, and guest incorporation frequently disrupts lattice order or exceeds the geometric constraints of the host framework.

The $CC1^{+10}$ lattice provides an ideal framework for testing scaffold-assisted structure determination. Its large solvent channels and mechanically stable architecture make it well suited for post-crystallization loading of guest proteins. Previous work using EDC crosslinking showed that CC1 crystals remain stable outside their growth conditions, allowing them to withstand buffer exchanges required for guest binding[61,62]. Building on this stability, we demonstrate that the $CC1^{+10}$ lattice can incorporate a diverse set of model systems—including homeodomains, coiled-coils, and zinc-finger motifs—under solvent conditions optimized for guest association. Because binding occurs within a pre-formed, diffracting lattice, local concentrations of guest molecules reach levels sufficient to stabilize even weak or transient DNA–protein interactions, enabling structural determination of complexes that are otherwise inaccessible to conventional crystallization.

4.3. Results

4.3.1. Asynchronous Loading of Guest proteins

To evaluate the versatility of $CC1^{+10}$ as a general molecular scaffold, we engineered the DNA insert to present targeted binding sites for members of three distinct DNA-binding domain (DBD) families: homeodomains, coiled-coils, and zinc-fingers, families selected for their structural diversity, canonical DNA-binding modes, and known crystal structures. These families are foundational across diverse biological systems, underscoring the scaffold's potential to accommodate a wide range of functional guests. As specific model guest proteins, we tested several homeodomain variants from *Drosophila melanogaster*, including Even-skipped (Eve-HD)[87], Engrailed homeodomain fused to eGFP (EnH-eGFP)[86,88], Ultrabithorax (UBX-HD)[89], Antennapedia (Antp-HD)[90], an engineered variant of the DNA-binding domain of the *Saccharomyces cerevisiae* GCN4 transcription factor (bZip)[91,92] and the C-clamp domain of human HDBP1 (C-clamp)[93]. Prior to experimental testing, each guest domain was individually modeled within the CC1 lattice using AlphaFold-3 to confirm compatibility with pore dimensions and to rule out steric clashes with the scaffold or symmetry mates (Figure S4.1).

An initial round of experiments revealed good X-ray diffraction for a variety of porous $CC1^{+10}$ crystals that had been stabilized with EDC, before and after soaking in high concentrations of UBX-HD guest protein. As detailed in Table S4.1., the incubation buffers were a mixture of $CC1^{+10}$ growth conditions and conditions that yielded strong UBX-HD binding as measured by ITC[94]. We selected these buffers to closely resemble conditions that had previously yielded diffraction since protein–DNA crystals are typically unstable outside of their growth conditions; even small changes in salt concentration or pH can disrupt lattice integrity, often leading to crystal degradation. Fluorophore-conjugated guest variants were observed to

concentrate within the crystals via light microscopy and confocal microscopy (Figure S4.2-S4.4). However, despite the apparent strong guest uptake within the crystals and the clear electron density maps for the target binding sites (Figure S4.5), there was no obvious electron density corresponding to the guest proteins.

Therefore, we next sought to determine if the crystals would diffract well under conditions optimized for guest DNA binding using the homeodomains as a model system. To identify solvent conditions favoring guest protein binding we turned to *in vitro* fluorescence polarization (FP) to evaluate binding to the same DNA oligonucleotides used for $CC1^{+10}$ crystallization, augmented by a 3′-terminal FAM fluorophore on one strand (Figure S4.5). Eve-HD binding was highly sensitive to ionic composition (Figure S4.6), with maximal fluorescence polarization change observed in the buffer containing 100 mM sodium chloride and 5mM magnesium chloride as the principal salts.

To confirm that the FP-optimized conditions were compatible with structural analysis, we crystallized CC1 under standard growth conditions and then subjected the crystals to chemical ligation using EDC, following our previously published protocol[61,62]. Crosslinked crystals were subsequently mechanically transferred using cryo loops into the low salt guest-binding buffers identified by FP. X-ray diffraction data collected from these transferred crystals revealed the complete loss of diffraction in the case of the condition where maximal affinity was observed (50 mM potassium chloride, no calcium chloride). However, another FP condition containing divalent cations (50 mM potassium chloride, 4 mM calcium chloride) retained diffraction comparable to $CC1^{+10}$ crystals as grown (Table S4.1.). Compared to the small guest proteins tested here, RepE54 binding of its cognate DNA appears to be remarkably insensitive to salt with crystals forming in conditions with much higher salt concentrations (e.g. up to 2.0 M Li^{+}).

Using the FP-optimized buffer conditions, we successfully loaded six guest proteins into pre-formed, crosslinked CC1 crystals and confirmed ordered binding by X-ray crystallography. Electron density corresponding to the DNA-binding domains was observed for each guest, including Eve-HD, UBX-HD, AntP-HD, EnH-eGFP, bZip, and C-clamp *were all visualized in expected binding modes*. For the Engrailed homeodomain GFP fusion (EnH-GFP), we observed clear density for the EnH domain but not for the flexibly tethered eGFP domain. This result highlights that even a relatively large protein fusion can diffuse through the crystal lattice and bind its DNA target, although flexibility or disorder may obscure density for non-DNA-binding domains. All guests were introduced post-crystallization and EDC crosslinking, demonstrating for the first time that asynchronous loading (soaking) is sufficient for guest incorporation into a protein–DNA co-crystal lattice. The resolution of the resulting structures varied by guest: Eve-HD-loaded crystals diffracted to 3.1 Å, UBX-HD to 3.2 Å, AntP-HD to 3.1 Å, EnH–GFP to 3.9 Å, bZip to 3.1 Å, and the C-clamp to 7.2 Å. Structures for the model guest proteins were consistent with known conventional crystal structures (Table S4.5), despite featuring DBD:DNA pairings new to the Protein Data Bank. Surprisingly, while Eve-HD did bind the expected target sequence register, a "bonus copy" of EVE-HD bound even more clearly to an adventitious site opposite RepE54 (see Discussion). This site consists of base pairs 10-15 (henceforth Register 1. R1:10–15). Together, these results confirm that the CC1 scaffold can accommodate a wide range of structurally and phylogenetically diverse DNA-binding proteins, supporting its future utility as a platform technology for obtaining rapid, high-throughput structural information on DNA–protein interactions under multiple conditions.

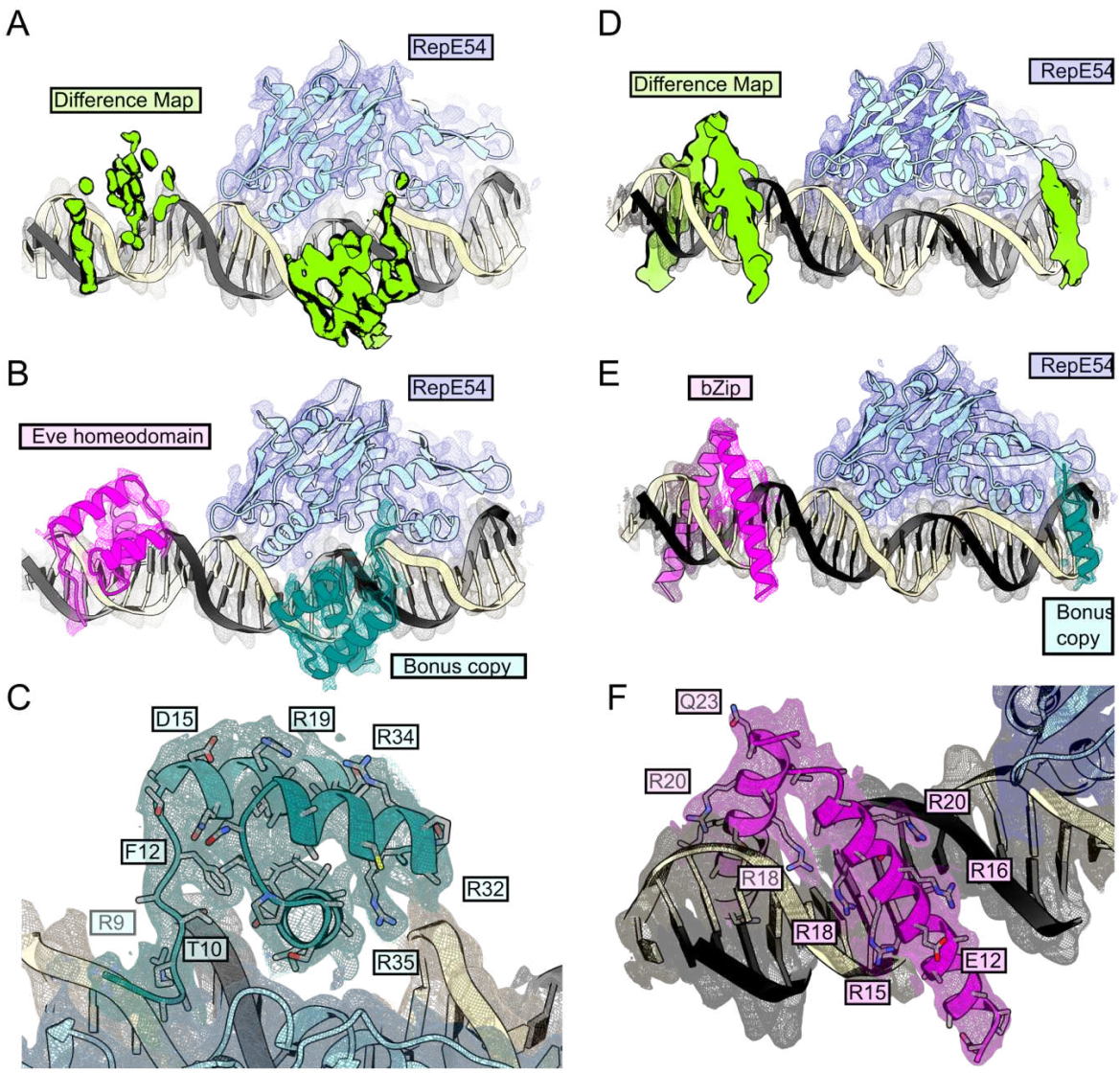


Figure 4.1. Guest Structure Determination of Eve-HD and bZip

(A) Discovery map of Eve-HD from 9YZM.pdb with F_0-F_C difference map contoured to σ = 4.0 (green). (B) Crystal structure of 9YZM.pdb with $2F_C$-F_0 map contoured to σ = 1.0, Eve-HD bound to canonical sequence 5' TAATTA 3' (magenta) and bonus copy of Eve-HD bound to R1 site (teal). (C) Close up view of Eve-HD bound to R1 sequence with key resolved residues shown as sticks and $2F_C$-F_0 map contoured to σ = 1.0 (teal). (D) Discovery map of bZip from 9YZU.pdb with F_0-F_C difference map contoured to σ = 4.0 (green). (B) Crystal structure of

9YZU.pdb with $2F_C$-F_0 map contoured to σ = 1.0, bZip bound to canonical sequence 5' ATGAGTCAT 3' (magenta) and bonus partial structure of bZip (teal). (C) Close up view of bZip bound to canonical sequence with key resolved residues shown as sticks and $2F_C$-F_0 map contoured to σ = 1.0 (magenta).

4.3.2. Molecular Goniometer

To further probe the versatility of $CC1^{+10}$ for asynchronous guest loading, we systematically shifted the binding register/sequence of EVE-HD and bZip along the DNA strut. We designed five DNA variants in which the homeodomain or bZip binding register/sequence was systematically shifted along the DNA strut (Figure S4.7). Eve-HD or bZip were successfully loaded into crosslinked CC1 crystals for all five DNA variants, and clear electron density was observed in each case (Figure 4.2). Surprisingly, in all five structures, bonus density was observed in unintended adventitious sites.

To determine whether the ionic environment influences the preference for R1, we soaked EDC ligated $CC1^{+10}$ crystals with Eve-HD under conditions with different ionic strengths. In all cases we observed clear electron density in the same R1 site within the core 21-mer scaffold DNA footprint (Table S4.5). While the overall binding position of Eve-HD within R1 remained conserved, comparison of the resulting structures revealed slight differences between the hydrogen bonding network and electrostatic interactions between Eve-HD and the local DNA duplex found in R1 (Figure 4.2G). The ability to obtain multiple structures with varying solution conditions points to the future capacity of scaffold-assisted crystallography as a high-throughput structure determination platform technology. Persistent alternate site binding across all ionic environments suggests that the driving forces for binding register selectivity are not driven by adventitious salt-dependent interactions.

Binding to alternate DNA sites was not unique to Eve-HD. All homeodomains tested displayed similar behavior and in the case of UBX-HD, a second noncanonical site (R2:18-23, 5' TCATAA 3') was also observed in addition to the first noncanonical site (R1:10-15, 5' AATTGC 3') (Figure S4.8). Isothermal titration calorimetry (ITC) measurements show that Eve-HD and UBX-HD bind these off-target sequences with significantly lower affinity in solution than the canonical motifs reported in the literature[95], yet they were consistently preferred within the crystal lattice (Figure S4.9). This reversal of specificity suggests that additional structural factors govern register selection in the scaffold. DNA-binding is well known to be driven by DNA helical structure as well as sequence[96]. The scaffold protein RepE54 compresses the minor groove opposite its site (Figure S4.10), and a compressed minor groove is a favorable binding location for the homeodomain N-terminal residues[97]. Detailed analysis of the DNA backbone conformation revealed that the DNA strut adopts different curvature when a guest is bound compared to when it is not (Figure S4.11). This further supports the idea that local DNA shape, rather than sequence alone, contributes to register selection. Lattice contacts may also play a role, as positioning Eve-HD at this site enables favorable electrostatic interactions between symmetry neighbors (Figure S4.12). In sum, our findings indicate that binding register selection within the CC1 scaffold is driven by a combination of sequence recognition, shape recognition, and the lattice environment.

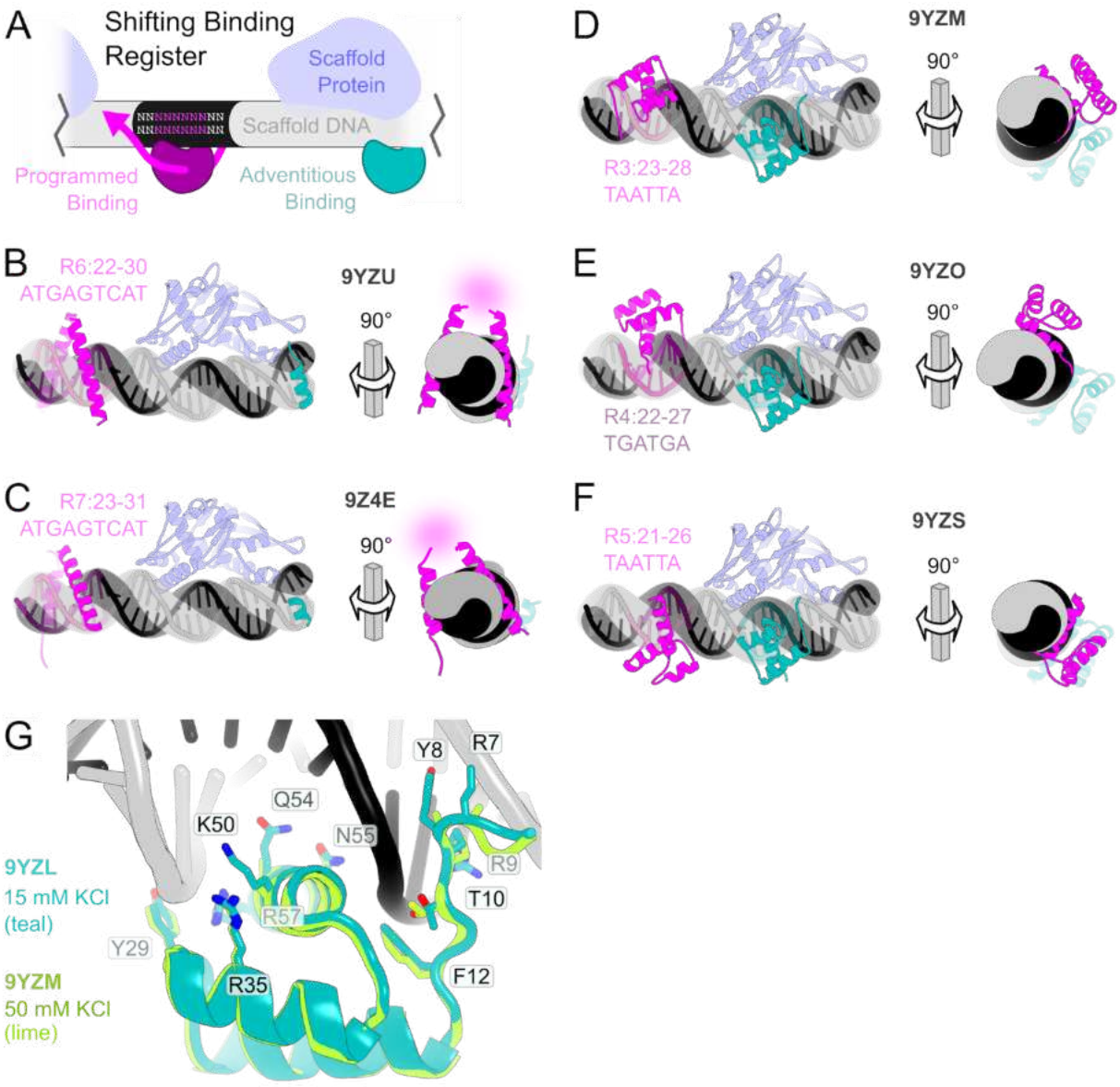


Figure 4.2. Guest structures with shifted and varied binding sites

(A) Schematic representation of $CC1^{+10}$ with shifting programmed binding sites (magenta) and adventitious binding (teal). Each crystal structure below is likewise shown with guest proteins bound to target locations (magenta) and bonus sites (teal). Each model is also shown with a 90° rotated view. **(B)** $CC1^{+10}$ with asymmetric insert 5'ATGAGTCATA 3' loaded with bZip (PDB

9YZU). **(C)** $CC1^{+10}$ with asymmetric insert 5'GATGAGTCAT 3' loaded with bZip (PDB 9Z4E), with unresolved putative C-terminal disulfide (magenta blur). **(D)** $CC1^{+10}$ with asymmetric insert 5'-GCTAATTAGGC-3' loaded with EVE-HD (PDB 9YZM). **(E)** $CC1^{+10}$ with asymmetric insert 5'-TTGATGAGCAG-3' loaded with EVE-HD (PDB 9YZO). **(F)** $CC1^{+10}$ with asymmetric insert 5'-TAATTAGGCCG-3' loaded with EVE-HD (PDB 9YZS). **(G)** Comparison of Eve-HD bound to R1 site under two different monovalent cation concentrations (PDBs 9YZL and 9YZM) with a subset of the resolved residues shown with the stick representation.

4.4. Discussion

4.4.1. Guest Protein Selection

The experiments described here demonstrate, for the first time, that guest protein structures can be determined following post-crystallization incorporation into a pre-formed crystalline scaffold. Using the $CC1^{+10}$ lattice, we successfully introduced multiple distinct DNA-binding domains—including homeodomains, coiled-coils, and zinc-fingers—into chemically stabilized crystals and resolved each at atomic detail. Despite differences in charge, size, and binding mechanism, all guests preserved their characteristic folds and canonical DNA-binding topologies, confirming that the CC1 framework provides an ordered yet permissive environment for guest capture. The ability of the lattice to accommodate such diverse molecular geometries without loss of diffraction quality underscores its exceptional mechanical stability and structural tolerance, establishing $CC1^{+10}$ as a robust and generalizable scaffold for protein incorporation.

This result represents a fundamental departure from traditional crystallographic approaches. Conventional determination of DNA-binding domain structures depends on brute-force screening of crystallization conditions, a process that is often slow, low-yielding, and

unpredictable. Even deliberate fusion strategies achieved only limited success because flexible linkers or appended domains frequently disrupted crystal packing or introduced disorder[30]. In contrast, the CC1 system decouples the processes of crystallization and guest incorporation entirely: the scaffold is first grown and chemically stabilized under conditions optimized for lattice order, and only then exposed to the guest protein under solution conditions favorable for DNA binding. This asynchronous loading strategy eliminates the need for genetic fusion or co-expression and, because the scaffold can be reprogrammed through minor alterations in its DNA sequence, enables reuse of the same crystalline host architecture across many distinct guest systems.

The successful visualization of guest domains within the CC1 lattice establishes a practical and conceptual foundation for scaffold-assisted crystallography. It confirms that macromolecular structure determination can proceed not from a stochastic search for crystal contacts, but from the rational design of a pre-ordered host that positions target molecules by molecular recognition. In doing so, this system realizes Seeman's original vision of using DNA-based order to mediate the crystallization problem[1], while extending it into a hybrid protein–DNA framework capable of atomic-level precision. More broadly, these results show that crystallography can be reimagined as a modular, design-driven process where lattice architecture and guest capture are independently programmable yet structurally compatible.

4.4.2. Size Constraints:

While the proof-of-concept experiments establish that $CC1^{+10}$ can incorporate diverse DNA-binding domains, the physical dimensions of the lattice ultimately govern which macromolecules can be accommodated. Highly porous crystals like $CC1^{+10}$ are unusual among

biomolecular lattices in that the solvent channels exceed the size of the protein building blocks themselves, creating a continuous network for guest diffusion. Analysis with *MAP_CHANNELS* revealed that the largest sphere capable of diffusing through the lattice interior has a diameter of approximately 5.0 nm, defining the upper limit for unhindered guest transport. Many globular proteins fall below this threshold; for example, a typical fluorescent protein approximates to a 3 × 5 nm cylinder and should be capable of navigating the pore network without steric hindrance. Indeed, crystals soaked with the EnH–eGFP fusion protein revealed clear electron density for the Engrailed homeodomain component, while the flexibly tethered eGFP moiety remained unresolved.

These observations emphasize that while the $CC1^{+10}$ lattice readily admits larger fusion constructs, the conformational freedom of attached domains can limit their structural visibility. Flexible linkers, while beneficial for preserving independent domain folding, permit a wide range of orientations that make resolution of the fusion domain elusive. Future iterations of this system could mitigate such disorder by introducing architectural elements that physically or sterically constrain the guest domain. One strategy would be to replace long, flexible linkers with shorter, semi-rigid connectors that limit rotational freedom without imposing excessive strain. Another strategy could involve direct fusion between the RepE54 scaffold protein and another DNA-binding protein effectively “pinning” the guest domain in a consistent orientation. Such approaches would extend the CC1 framework beyond DNA-binding proteins, enabling incorporation of diverse functional domains and further building on the intrinsic modularity of the lattice.

4.4.3. Scaffold Binding Driving Forces

To serve as a platform technology for the study of protein–DNA interactions, the host crystals must be able to resolve both expected and unexpected complexes. Despite the use of model proteins, this capacity is already in ample evidence. A striking feature of our system is the consistent preference of homeodomains for noncanonical binding registers within the CC1 scaffold, even when canonical sequence motifs were present. This behavior was observed across all homeodomains tested and was not limited to a single site. Homeodomain recognition relies on both sequence and shape-based readout[97,98], and in this context the RepE54 component of the CC1 scaffold provides a rich environment: it bends the scaffold DNA and compresses the minor groove opposite its binding site, creating a thermodynamically favorable pocket for the positively charged N-terminal arm of the homeodomain. Moreover, the R1 register partially matches canonical recognition motifs. ITC measurements supported the critical role of lattice deformations and contacts, confirming that canonical sites are preferred in solution. Notably, the unexpected binding of UBX-HD to the additional register R2 was confirmed by ITC as a moderately strong recognition sequence in solution.

Importantly, the ability of CC1 scaffolds to withstand high local guest concentrations means that even relatively weak interactions can become structurally detectable. By taking advantage of mass action, otherwise invisible protein–DNA complexes can be stabilized and visualized, providing access to a broader landscape of binding modes than is typically observable in solution. “Bonus density” was not unique to the homeodomain family: for the bZip guest, which interacts exclusively with the DNA major groove, we likewise observed additional electron density at an unintended register, although the canonical site remained the predominant binding location. The absence of minor groove contacts in this system suggests that other

factors—such as lattice packing, local electrostatics, or DNA deformability—also contribute to alternate site recognition.

Collectively, these findings reveal that binding register selection within the CC1 lattice arises from an interplay of sequence, DNA geometry, and lattice context. The scaffold's intrinsic curvature and minor-groove compression favors particular registers, while the ordered environment enhances weak or transient associations that would otherwise escape detection in solution. In this sense, the CC1 lattice acts not as a passive container but as an active determinant of molecular recognition that can bias binding toward energetically favorable configurations shaped by the geometry of the host framework. This emergent behavior highlights the capacity of scaffold-assisted crystallography to uncover recognition modes inaccessible to traditional approaches and underscores the importance of local lattice environment as a tunable parameter in future scaffold design.

4.5. Conclusions

The studies presented in this chapter establish that the $CC1^{+10}$ lattice provides a robust and modular framework capable of supporting scaffold-assisted structure determination. By demonstrating post-crystallization loading and structural resolution of multiple DNA-binding domains, this work transforms the concept of a programmable crystalline host from theory into practice. Beyond decoupling lattice formation from guest incorporation, the system enables systematic exploration of molecular recognition as a function of DNA sequence, local shape, and ionic environment. Analysis of guest binding behavior revealed that the CC1 scaffold itself can modulate site selectivity, uncovering previously unrecognized binding modes for homeodomains within a pre-ordered environment. Together, these findings illustrate how scaffold-assisted

crystallography can both capture known interactions and reveal new ones, broadening our understanding of molecular recognition within crystalline assemblies and underscoring the potential of hybrid scaffolds as experimental platforms for structural discovery.

4.6. Materials and Methods

4.6.1. Protein Expression and Purification

The protein cloning and expression of the crystal protein Replication initiator RepE54 was described by Komori et al[70]. The CC1 RepE54 was expressed and purified by the Histone source at Colorado State University exactly as described in Chapter 2 and in our previous ligation study[61].

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 (100 mM tris HCl, 200 mM NaCl, 10% glycerol, 10 mM imidazole, pH 8.0) and applied to HisTrap (HisPur™ Ni-NTA Resin) equilibrated with HisTrap buffer (100 mM tris HCl, 200 mM NaCl, 10% glycerol, 10 mM 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 (100 mM NaCl, 100 mM 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 (200 mM NaCl, 100 mM 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.

The other homeodomain proteins used in this work (Eve-HD, UBX-HD, and AntP-HD) were all expressed and purified by our collaborators in the Spratt lab (Clark University). A detailed protocol was described by *Orlomoski et al*[95].

bZip protein was commercially synthesized (GenScript) and resuspended in 100 mM sodium chloride,100 mM sodium phosphate (pH 7.4). Following resuspension, samples were incubated on ice for 12 hours to promote disulfide bond formation prior to downstream applications.

C-clamp protein was commercially synthesized (LifeTein) in two variants: an unlabeled construct and a C-terminal TAMRA-conjugation construct. Lyophilized material was resuspended in 50 mM Tris-HCl (pH 7.4) supplemented with 1 mM $ZnCl_2$. To remove excess zinc, samples were concentrated and buffer-exchanged using Amicon Ultra centrifugal filter units (3 kDa MWCO; Millipore) prior to downstream applications

4.6.2. 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. The oligomers were

resuspended in CC1 oligo buffer (50 mM Tris HCl, 100 mM KCl pH 6.7) 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.

4.6.3. Scaffold Protein: DNA Complex Crystallization

The scaffold protein and DNA were mixed in a 1.2:1 ratio (440:364 μM, 55 μL total) and were incubated on ice for 30 minutes prior to crystallization via sitting drop vapor diffusion. Crystallization trials used CrysChem sitting-drop plates (Hampton Research HR3-159), with 400 μL reservoir solution per well and 4 μL crystallization drops prepared by mixing 2 μL each of the protein–DNA complex and reservoir solution directly on the platform; drops were mixed by gentle pipetting. Plates were sealed immediately after setup and incubated at 25 °C under static conditions. Crystallization conditions for both the porous lattice ($CC1^{+10}$) and for the large-porous lattice ($CC1^{+21}$) were 10-80 mM Magnesium Acetate, 0.3-1.8 M Lithium Sulfate, and 50 mM MES pH 6.5 (or occasionally 50 mM sodium cacodylate pH 6.5). A detailed list of individual crystals, PDB codes, and corresponding growth conditions are reported in Table S2. Crystals grew to full size in a range of 24 hours to 30 days (most often in 1-2 days). Crystal images were taken with a Moticam X5 Plus camera attached to a Motic SMZ-168 stereozoom microscope.

4.6.4. X-Ray Diffraction Data Collection and Refinement

Crystals were flash frozen in liquid nitrogen after briefly transferring the crystal to a cryo-protectant solution (matching the reservoir solution supplemented with 25% glycerol). Single crystal x-ray diffraction data was collected at either ALS beamline 4.2.2(CMOS detector),

beamline 8.2.2(Pilatus3 detector), or beamline 8.2.1(Dectris EIGER2 detector) from 0 to 180° with an omega delta of 0.2°and an exposure of 0.5s or 0.25s. Data was processed with XDS and refined using PHENIX and COOT. The first porous $CC1^{+10}$ dataset (pdb code: 9Z08) was refined using new RFree flags, and subsequent structures used the same flags applied.

4.6.5. EDC Ligation

Co-crystals were washed in conditions similar to their growth conditions (1.2-1.6M Li_2SO_4, 50mM $MgAc_2$, and 50mM MES pH 6.5). The co-crystals were washed with 200uL in 9-well glass plates (Hampton) to remove additional protein and DNA monomers. 1-Ethyl-3(3-dimethylaminopropyl)carbidiimide (EDC) (Advanced Chemtech CAS#:25952-53-8) was resuspended in the crosslinking solution (1.2-1.6M Li_2SO_4, 50mM $MgAc_2$, and 50mM MES pH 6.0) to final concentration values of 30-60 mg/mL and used immediately. The co-crystals were crosslinked in a 200µL EDC solution volume overnight (20-24 hrs). The crosslinking reaction was ended by moving the co-crystals back into the wash solution.

4.6.6. Fluorescence Polarization Probe Duplexing

The DNA oligomer sequences used for fluorescence polarization measurements match a sequence from a $CC1^{+10}$ structure or contains the canonical binding site for a guest protein but lacks terminal phosphate modification and one strand is modified with a 3′-terminal 6-carboxyfluorescein (FAM) to enable fluorescence-based detection (Figure S4.3.). Individual oligomers were synthesized by Integrated DNA Technologies. The oligomers were resuspended in CC1 oligo buffer (50 mM Tris HCl, 100 mM KCl pH 6.7) 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 concentration of all

annealed FP duplexes was 1mM. FP duplexes were diluted to a final concentration of 20nM in the solvent conditions being tested for each guest.

4.6.7. Fluorescence Polarization Data Collection

The first DNA oligomer sequences used for fluorescence polarization measurements (Fig. S9) matched a full $CC1^{+10}$ DNA block (31 bp) , contained the canonical binding site (5'-TAATTA-3') for a guest homeodomain, lacked terminal phosphate modification, and had one strand modified with a 3′-terminal 6-carboxyfluorescein (FAM) for fluorescence detection. A second truncated DNA duplex (11 bp) was prepared to probe binding in the absence of the RepE54 binding site. Individual oligomers were again synthesized by Integrated DNA Technologies, and purified via standard desalting. The oligomers were resuspended in CC1 oligo buffer (50 mM Tris HCl, 100 mM KCl, pH 6.7) 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 then allowed to cool to room temperature for one hour. The stock solution concentration of all annealed FP duplexes was 1 mM. Prior to 1:1 mixing with the protein (see below), FP duplexes were diluted to a concentration of 20 nM in the solvent conditions being tested.

Experiments were carried out in black, flat-bottom 384-well plates (Greiner Bio-One, Ref#:781900) in a final volume of 30 µL per well (15 µL FP duplex and 15 µL of protein). Each well contained 10 nM final concentration of FP duplex in binding buffer. Guest proteins were serially diluted in their storage buffer starting from 50 µM, generating a 15-point dilution series (final concentration range from 50 µM to ~1.5 nM), with one well containing no protein as a fluorescence polarization baseline.

After adding guest protein, the plates were incubated at room temperature for 30-45 minutes to allow binding equilibrium to be reached. FP was measured on a CLARIOStar Plus

(BMG Laboratories) plate reader using a 485 nm Ex: 535 nm Em filter. Initial readings were performed in duplicate (Fig. S10AB). Later readings were measured with eightfold replication (Fig. S10C). Baseline polarization values were measured using protein-free blanks. Fluorescence polarization data were analyzed using a custom Python script (https://github.com/cdsnow/FPfitting) that averages replicate measurements, calculates standard errors, and fits each titration curve to a 1:1 binding model (Fig. S10). The script fits the protein–DNA complex concentration at each titration point, incorporating the fixed final 10 nM concentration of fluorescent duplex via the quadratic Morrison equation to account for ligand depletion. Nonlinear regression, weighted by the inverse variance (with a minimum uncertainty threshold of 5% of the dynamic range to prevent overweighting), was used to simultaneously optimize K_d, the baseline polarization value, and the bound-state polarization value. One outlier data point at 200 mM NaCl and 6.25 µM was excluded from the analysis. The resulting fits (Fig. S10) were plotted on a log-scaled protein concentration axis and used to estimate K_d values for the salt scouting conditions. The provided Python fitting scripts also allow simple Langmuir binding equation fitting. We omit the simple Langmuir fits here since the results were redundant with the more sophisticated model (Fig. S10).

4.6.8. Isothermal Titration Calorimetry (ITC)

The binding affinity of EVE-HD and UBX-HD to dsDNA sequences R1 and R2 (Fig. S16) were measured using an Affinity ITC Auto Low Volume System (190 µL) equipped with a gold cell (TA Instruments; New Castle, DE). ITC runs were performed and controlled using the built-in ITCRun software. The sample cell (set to 25**°C)** containing 20 µM HD in 20 mM Tris, 50 mM KCl, 4 mM $CaCl_2$ (pH 7.4) was titrated with 2 µL x 20 injections of 200 µM dsDNA. All

runs were performed in duplicate to ensure reproducibility. Isotherms and titration curves were plotted using Microsoft Excel.

Python scripts (https://github.com/cdsnow/FPfitting) were used to fit the data and obtain thermodynamic parameters for the HD-dsDNA interaction (i.e, K_d, binding stoichiometry, ΔH, and ΔS). Specifically, integrated heat data (kJ mol^{-1} of injectant vs. molar ratio) were fit to the Wiseman isotherm, a closed-form model for 1:1 binding (80). The four adjustable parameters were the apparent stoichiometry (N), a dimensionless affinity parameter $c = K_a [M]$ (the isotherm is algebraically equivalent to the traditional parameterization with $c = n K_a [M]$), the molar enthalpy of binding (ΔH), and a constant offset accounting for the heat of dilution. Nonlinear least-squares regression was performed with the trust-region reflective algorithm as implemented in scipy.optimize.curve_fit (SciPy v1.16.3), using unweighted residuals. The macromolecule concentration was fixed at 20 μM for all experiments and the temperature was fixed at 298.15 K. Thermodynamic quantities were derived from the fitted parameters using standard relationships: $K_d = [M]/c$, $\Delta G = RT \ln K_d$, and $-T\Delta S = \Delta G - \Delta H$.

Uncertainties in the fitted parameters were obtained from the diagonal of the covariance matrix returned by the least-squares optimizer. For derived quantities that depend on multiple fitted parameters (notably $-T\Delta S$, which depends on both c and ΔH), uncertainties were propagated analytically using the full covariance matrix, including off-diagonal terms, to account for parameter correlations. As an independent check, residual-resampling bootstrap analysis (81) was performed with 1,000 trials per dataset: residuals from the best-fit model were resampled with replacement, added back to the model prediction to generate synthetic datasets, and each synthetic dataset was re-fit to yield a bootstrap distribution of all parameters. Bootstrap standard deviations (with Bessel's correction) served as the reported uncertainties.

4.6.9. Crystal guest Loading

Crosslinked co-crystals containing DNA sequences for guest proteins were washed at room temperature for 1 hour to remove components that could interfere with guest protein binding and to equilibrate the crystals to the impending guest loading soak. We used wash solution that matched the subsequent guest loading buffers, and these guest loading buffers (Table S7) were varied to 1) demonstrate the capacity of the crystals to diffract in different conditions and 2) determine how the salt concentration changed the resulting electron density map. Accordingly, the wash solutions consisted of 15-50 mM KCl, 2-4 mM $CaCl_2$, 10 mM Tris HCl, pH 7.4, and 10% glycerol. A fresh 180 µL of guest loading / wash solution was mixed with 20-30 µL of guest protein, and the washed co-crystals were moved into the guest loading solution using cryo loops (the concentrations in Table S4.4 are the final concentrations after mixing). The co-crystals sitting in guest loading solutions were sealed for 24-72 hours to allow guest proteins to diffuse and reach binding equilibrium within the crystal lattice.

4.7. Acknowledgements

We thank Hataichanok (Mam) Scherman, PhD., Director of the Histone Source at Colorado State University for the expression and purification of RepE54 transcription factor; Jay Nix, PhD. at ALS beamlines 4.2.2, 8.2.1, and 8.2.2 for extensive support of the XRD data collection; Donald Spratt, PhD. and Fadwa Mekkaoui for providing homeodomain guest proteins and performing ITC experiments; Grant Schauer, PhD. for fluorescence polarization expertise.

4.8 Funding

This material is based upon work supported by the National Science Foundation under Grant No. NSF DMR 2003748 and NSF DMR 2310574.

CHAPTER 5. SUMMARY AND FUTURE DIRECTIONS

From design to experimental validation, the novel isoreticular co-crystals developed through this research have enabled the possibility of high-throughput crystallography for DNA-binding proteins representing the successful realization of scaffold assisted structure determination after four decades[1]. With continued development and optimization, expanded co-crystal variants may facilitate the structural determination of previously uncharacterized DNA-binding molecules by X-ray diffraction.

Notably, the nanoscale structural control enabled by CC1 scaffolds should have applications beyond structural biology. In a biomedical context, CC1-based scaffolds could accelerate the identification of therapeutic targets among human DNA-binding proteins. Even weak DNA interactions may be observed among these future targets if driven by mass action, a phenomenon we have already observed in the case of the homeodomain and bZip binding to adventitious sites within the $CC1^{+10}$ lattice. CC1 may serve as a general-purpose molecular sieve for applications in sensing and catalysis. A variety of stable CC1 variants could ultimately be deployed for advanced separation operations on complex mixtures of DNA-binding proteins. In addition, the combination of programmable DNA geometry and protein-mediated rigidity, established in earlier chapters, provides a generalizable strategy for creating stable yet tunable crystalline environments. Together these elements define a modular design space in which both pore dimensions and chemical composition can be systematically controlled, positioning CC1 as a versatile hybrid material platform whose potential extends beyond the directions explored here.

In Chapter 2, the engineered porous co-crystal CC1 was characterized and experimentally validated. The expanded variant, $CC1^{+10}$, demonstrated a high degree of modularity with respect to both lattice topology and DNA sequence. Despite significant porosity (50Å diameter), the XRD resolution was good (up to ~3.0Å). Assembly of the scaffold was shown to be tunable by

adjusting the number of sticky-end base pairs and the length of the inserted DNA duplex. Looking ahead, further efforts will focus on expanding DNA-insert length beyond current limits to probe the upper bounds of lattice accommodation and stability. Parallel work, led by Callie Slaughter, is exploring incorporation of non-canonical nucleic acids into the CC1 framework, potentially enabling new modes of interaction, chemical functionality, or environmental responsiveness within the lattice.

In Chapter 3, the CC1 scaffold was further engineered through targeted modifications of the RepE54 protein to enhance stability, functional versatility, and guest compatibility. A disulfide bond was introduced across one of the key protein interfaces, showing that these contacts can be altered without disrupting crystal formation. In addition, a short peptide tag and an auxiliary DNA-binding domain were fused to RepE54, demonstrating that even substantial modifications to the protein component are tolerated and still yield the characteristic CC1 lattice. Building on these results, future work will explore the introduction of additional disulfide bonds or other interface-stabilizing mutations to further rigidify the scaffold and improve diffraction quality. Parallel efforts will expand the range of peptide tag modifications, such as other covalent linking peptides[99], which will be systematically evaluated by X-ray diffraction for structural compatibility. Finally, fusion-based constructs incorporating additional DNA-binding domains will be refined with the long-term goal of achieving structure determination of the fused domains within the lattice.

In Chapter 4, the expanded scaffold $CC1^{+10}$ was used to host a variety of DNA-binding proteins for structure determination. Guest proteins—including representatives from homeodomain, zinc-finger, and coiled-coil families—were successfully visualized in the electron density maps of $CC1^{+10}$ co-crystals, demonstrating clear incorporation across multiple DNA-

binding registers and under varying solvent conditions. These results establish $CC1^{+10}$ as a versatile platform for high-throughput X-ray crystallography of DNA-binding proteins, offering structural access to otherwise challenging targets and enabling comparative analyses of protein behavior across different binding geometries and environmental contexts. Rather than conventional crystals, where adventitious crystal contacts can occasionally induce local changes to the molecular structure, the modularity of the CC1 platform allows researchers to observe a DNA-binding-protein of interest with interactions that vary by shifting the binding register.

Looking ahead, future work will leverage this platform to solve the structures of uncharacterized DNA-binding proteins, particularly those difficult to crystallize by traditional methods. In parallel, efforts are underway to re-engineer the CC1 lattice to accommodate non-DNA-binding proteins through covalent or affinity-based installation, broadening the scope of this system beyond nucleic acid recognition and toward general-purpose structural analysis. It has also not escaped our notice that guest proteins fused to multiple binding motifs could be precisely positioned within the lattice. In the future, dynamic control over the guest position would represent an interesting type of structure-resolved, massively high-throughput nanotechnology.

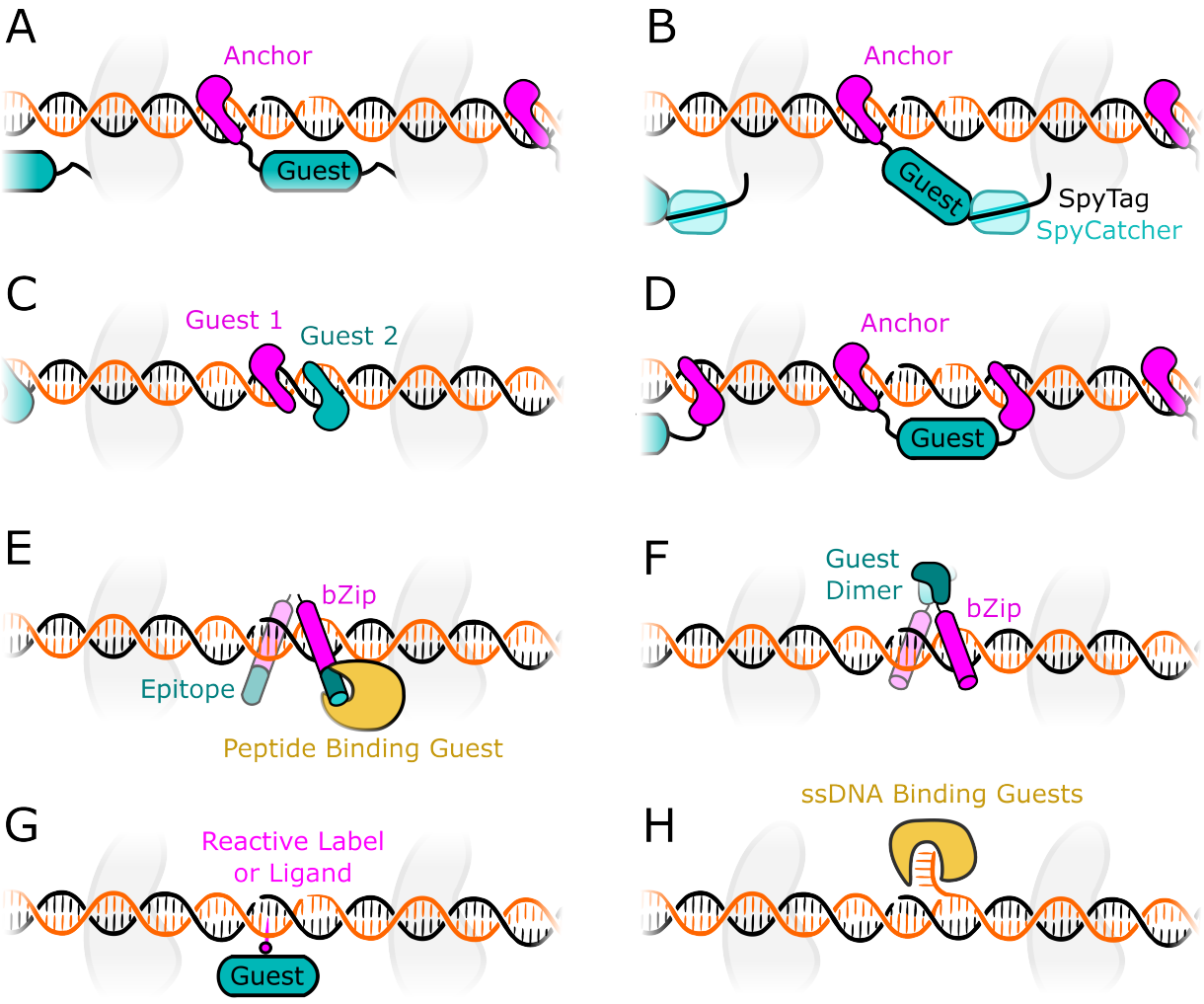


Figure 5.1. Bionanotechnology Extensions to the Scaffold Crystal Platform Technology.

While the current work demonstrates how either programmed or adventitious binding can be used to install DNA-binding domains, this scaffold family should permit additional control over guest orientation via a variety of approaches, motivating the analogies to a 3D molecular pegboard or additive manufacturing on the molecular scale. Ultimately, these may be sufficient to induce order in guest domains that lack intrinsic DNA binding affinity or specificity. Notably, while adventitious binding sites do not prevent programmed guest installation for structural biology, adventitious sites could interfere with the pursuit of ultimate control. Therefore, future

work will focus on eliminating off-target installation. One key concept is to move beyond single attachment sites to dual anchoring strategies. **(A)** Genetic fusion of the guest protein to the scaffold protein RepE54 and a DBD anchor. While sizable changes to the scaffold protein risk degrading expression, stability, and crystallization, we have successfully grown $CC1^{+10}$ crystals in preliminary work where an entire homeodomain was genetically fused to RepE54. Rapidly growing crystals where the anchor binding site is moved could be a systematic way to reduce variability in the guest binding position. **(B)** To limit the perturbation to the scaffold protein expression, stability, and crystallization propensity, we can engineer a RepE54 variant that offers a covalent installation motif like SpyTag (115). We have grown $CC1^{+10}$ crystals containing RepE54-ST in preliminary results. In principle, guest proteins of interest could be expressed as a genetic fusion to SpyCatcher and covalently captured by the scaffold protein SpyTag in solution or *in crystallo*. **(C)** For small guests or larger struts (e.g. $CC1^{+21}$), it should be possible to study the interactions between the guests, interactions that could provide additional stabilizing contacts to reduce disorder. The feasibility of this strategy is boosted by the proof of concept observed multiple times in this report of the same DBD bound to two registers. **(D)** Another route to capitalize on the existing guest DNA binding domains is to fuse guests of interest to N- and C-terminal DBD. We envision a modular system that can vary both the protein linkers as well as the DBD binding locations on the scaffold crystal, providing multiple methods to "tighten the straps". **(E)** Existing DBD (e.g. homeodomains or bZip as shown) can be augmented to display bait motifs for additional guest molecules, thereby converting into structure determination "adaptors". For example, adaptor domains could display helical peptide motifs (e.g. Sun Tag, ALFA tag) suitable to recruit additional guest proteins (e.g. peptide-binding scFv constructs). **(F)** Successful installation of the bZip parallel homodimer provides an interesting avenue to

potentially pursue scaffold-assisted structure determination of homodimer guest domains. A simple expedient would be to add the bZip domains as a N- or C-terminal fusion, optionally with computational protein design of the linker regions to maximize construct rigidity. **(G)** The ability to rapidly synthesize a variety of porous scaffold crystals with varying DNA inserts should be yet another fruitful avenue. A large variety of reactive chemical moieties (e.g. click chemistry components) or bait ligands can be displayed to enhance guest capture. In prior work, we grew ipCC1^{+10} crystals with DNA mismatches in the form of a 1 nt bulge (35). In preliminary work, we have grown CC1^{+10} crystals with a variety of modifications including phosphorothioates, a G:G mismatch, pendant amines, a DNA:RNA hybrid strut, pendant carboxylic acids, and pendant fluorophores. The ability to grow CC1+10 crystals with significant variations in the DNA bode well for future attempts to use such variations to capture cognate proteins (e.g. for mismatch recognition). **(H)** In preliminary work, we also have grown CC1^{+10} crystals with pendant ssDNA segments of 4, 6, or 15 nt, with multiple display sites and diffraction to 3.2Å. In addition to highlighting the remarkable modularity of the system, these crystals provide a motif suitable to capture proteins that bind ssDNA rather than dsDNA. The capture of nucleic acid guests is another exciting research direction that is underway. Executing multiple rounds of strand installation as pioneered by Hao *et al*. is an intriguing form of nanotechnology (17), and we ultimately aim to observe such transitions via ordered changes in the electron density map. Figure prepared in Inkscape.

6. BIBLIOGRAPHY

(1) Seeman, N. C. Nucleic Acid Junctions and Lattices. *J. Theor. Biol.* **1982**, *99* (2), 237–247. https://doi.org/10.1016/0022-5193(82)90002-9.

(2) Yang, Z.; Lee, M. M. M.; Chan, M. K. Efficient Intracellular Delivery of P53 Protein by Engineered Protein Crystals Restores Tumor Suppressing Function in Vivo. *Biomaterials* **2021**, *271*, 120759. https://doi.org/10.1016/j.biomaterials.2021.120759.

(3) Yang, Z.; Sun, J. K.-L.; Lee, M. M.; Chan, M. K. Restoration of P53 Activity via Intracellular Protein Delivery Sensitizes Triple Negative Breast Cancer to Anti-PD-1 Immunotherapy. *J. Immunother. Cancer* **2022**, *10* (9), e005068. https://doi.org/10.1136/jitc-2022-005068.

(4) Yang, Z.; Yang, M.; Chow, H.-M.; Tsang, S. Y.; Lee, M. M.; Chan, M. K. Cytosolic Delivery of CDK4/6 Inhibitor P16 Protein Using Engineered Protein Crystals for Cancer Therapy. *Acta Biomater.* **2021**, *135*, 582–592. https://doi.org/10.1016/j.actbio.2021.08.054.

(5) Geng, C.; Paukstelis, P. J. DNA Crystals as Vehicles for Biocatalysis. *J. Am. Chem. Soc.* **2014**, *136* (22), 7817–7820. https://doi.org/10.1021/ja502356m.

(6) 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. https://doi.org/10.1039/C8BM01378K.

(7) Yekta, R.; Xiong, X.; Li, J.; Heater, B. S.; Lee, M. M.; Chan, M. K. Mechanoresponsive Protein Crystals for NADH Recycling in Multicycle Enzyme Reactions. *J. Am. Chem. Soc.* **2024**, *146* (28), 18817–18822. https://doi.org/10.1021/jacs.4c04725.

(8) Woloszyn, K.; Vecchioni, S.; Ohayon, Y. P.; Lu, B.; Ma, Y.; Huang, Q.; Zhu, E.; Chernovolenko, D.; Markus, T.; Jonoska, N.; Mao, C.; Seeman, N. C.; Sha, R. Augmented DNA Nanoarchitectures: A Structural Library of 3D Self-Assembling Tensegrity Triangle Variants. *Adv. Mater.* **2022**, *34* (49), 2206876. https://doi.org/10.1002/adma.202206876.

(9) Hao, Y.; Kristiansen, M.; Sha, R.; Birktoft, J. J.; Hernandez, C.; Mao, C.; Seeman, N. C. A Device That Operates within a Self-Assembled 3D DNA Crystal. *Nat. Chem.* **2017**, *9* (8), 824–827. https://doi.org/10.1038/nchem.2745.

(10) Stuart, J. D.; Hartman, D. A.; Gray, L. I.; Jones, A. A.; Wickenkamp, N. R.; Hirt, C.; Safira, A.; Regas, A. R.; Kondash, T. M.; Yates, M. L.; Driga, S.; Snow, C. D.; Kading, R. C. Mosquito Tagging Using DNA-Barcoded Nanoporous Protein Microcrystals. *PNAS Nexus* **2022**, *1* (4), pgac190. https://doi.org/10.1093/pnasnexus/pgac190.

(11) Zhang, C.; Zhao, J.; Lu, B.; Seeman, N. C.; Sha, R.; Noinaj, N.; Mao, C. Engineering DNA Crystals toward Studying DNA–Guest Molecule Interactions. *J. Am. Chem. Soc.* **2023**, *145* (8), 4853–4859. https://doi.org/10.1021/jacs.3c00081.

(12) 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. https://doi.org/10.1038/nature11990.

(13) Simmons, C. R.; Buchberger, A.; Henry, S. J. W.; Novacek, A.; Fahmi, N. E.; MacCulloch, T.; Stephanopoulos, N.; Yan, H. Site-Specific Arrangement and Structure Determination of Minor Groove Binding Molecules in Self-Assembled Three-Dimensional DNA Crystals. *J. Am. Chem. Soc.* **2023**, *145* (48), 26075–26085. https://doi.org/10.1021/jacs.3c07802.

(14) Seeman, N. C.; Sleiman, H. F. DNA Nanotechnology. *Nat. Rev. Mater.* **2017**, *3* (1), 17068. https://doi.org/10.1038/natrevmats.2017.68.

(15) Dong, Y.; Yao, C.; Zhu, Y.; Yang, L.; Luo, D.; Yang, D. DNA Functional Materials Assembled from Branched DNA: Design, Synthesis, and Applications. *Chem. Rev.* **2020**, *120* (17), 9420–9481. https://doi.org/10.1021/acs.chemrev.0c00294.
(16) Yin, P.; Choi, H. M. T.; Calvert, C. R.; Pierce, N. A. Programming Biomolecular Self-Assembly Pathways. *Nature* **2008**, *451* (7176), 318–322. https://doi.org/10.1038/nature06451.
(17) 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. https://doi.org/10.1021/jacs.8b07180.
(18) Stephanopoulos, N. Hybrid Nanostructures from the Self-Assembly of Proteins and DNA. *Chem* **2020**, *6* (2), 364–405. https://doi.org/10.1016/j.chempr.2020.01.012.
(19) Castro, C. E.; Kilchherr, F.; Kim, D.-N.; Shiao, E. L.; Wauer, T.; Wortmann, P.; Bathe, M.; Dietz, H. A Primer to Scaffolded DNA Origami. *Nat. Methods* **2011**, *8* (3), 221–229. https://doi.org/10.1038/nmeth.1570.
(20) 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. https://doi.org/10.1038/nature08274.
(21) 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. https://doi.org/10.1021/ja031754r.
(22) 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. https://doi.org/10.1021/jacs.6b06508.
(23) Paukstelis, P. J.; Nowakowski, J.; Birktoft, J. J.; Seeman, N. C. Crystal Structure of a Continuous Three-Dimensional DNA Lattice.
(24) Paukstelis, P. J. Three-Dimensional DNA Crystals as Molecular Sieves. *J. Am. Chem. Soc.* **2006**, *128* (21), 6794–6795. https://doi.org/10.1021/ja061322r.
(25) Aksel, T.; Yu, Z.; Cheng, Y.; Douglas, S. M. Molecular Goniometers for Single-Particle Cryo-Electron Microscopy of DNA-Binding Proteins. *Nat. Biotechnol.* **2021**, *39* (3), 378–386. https://doi.org/10.1038/s41587-020-0716-8.
(26) Khoshouei, A.; Kempf, G.; Mykhailiuk, V.; Griessing, J. M.; Honemann, M. N.; Kater, L.; Cavadini, S.; Dietz, H. Designing Rigid DNA Origami Templates for Molecular Visualization Using Cryo-EM. *Nano Lett.* **2024**, acs.nanolett.4c00915. https://doi.org/10.1021/acs.nanolett.4c00915.
(27) Jones, A. A.; Snow, C. D. Porous Protein Crystals: Synthesis and Applications. *Chem. Commun.* **2024**, *60* (45), 5790–5803. https://doi.org/10.1039/D4CC00183D.
(28) Derewenda, Z. S. Application of Protein Engineering to Enhance Crystallizability and Improve Crystal Properties. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (5), 604–615. https://doi.org/10.1107/S090744491000644X.
(29) Li, Z.; Wang, S.; Nattermann, U.; Bera, A. K.; Borst, A. J.; Yaman, M. Y.; Bick, M. J.; Yang, E. C.; Sheffler, W.; Lee, B.; Seifert, S.; Hura, G. L.; Nguyen, H.; Kang, A.; Dalal, R.; Lubner, J. M.; Hsia, Y.; Haddox, H.; Courbet, A.; Dowling, Q.; Miranda, M.; Favor, A.; Etemadi, A.; Edman, N. I.; Yang, W.; Weidle, C.; Sankaran, B.; Negahdari, B.; Ross, M. B.;

Ginger, D. S.; Baker, D. Accurate Computational Design of Three-Dimensional Protein Crystals. *Nat. Mater.* **2023**, *22* (12), 1556–1563. https://doi.org/10.1038/s41563-023-01683-1.
(30) 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. https://doi.org/10.1021/jacs.8b07512.
(31) Kojima, M.; Abe, S.; Furuta, T.; Hirata, K.; Yao, X.; Kobayashi, A.; Kobayashi, R.; Ueno, T. High-Throughput Structure Determination of an Intrinsically Disordered Protein Using Cell-Free Protein Crystallization. *Proc. Natl. Acad. Sci.* **2024**, *121* (25), e2322452121. https://doi.org/10.1073/pnas.2322452121.
(32) Liu, Y.; Gonen, S.; Gonen, T.; Yeates, T. O. Near-Atomic Cryo-EM Imaging of a Small Protein Displayed on a Designed Scaffolding System. *Proc. Natl. Acad. Sci.* **2018**, *115* (13), 3362–3367. https://doi.org/10.1073/pnas.1718825115.
(33) 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. https://doi.org/10.1038/s41467-019-09836-0.
(34) Castells-Graells, R.; Meador, K.; Arbing, M. A.; Sawaya, M. R.; Gee, M.; Cascio, D.; Gleave, E.; Debreczeni, J. É.; Breed, J.; Leopold, K.; Patel, A.; Jahagirdar, D.; Lyons, B.; Subramaniam, S.; Phillips, C.; Yeates, T. O. Cryo-EM Structure Determination of Small Therapeutic Protein Targets at 3 Å-Resolution Using a Rigid Imaging Scaffold. *Proc. Natl. Acad. Sci.* **2023**, *120* (37), e2305494120. https://doi.org/10.1073/pnas.2305494120.
(35) Nair, M. S.; Lee, M. M.; Bonnegarde-Bernard, A.; Wallace, J. A.; Dean, D. H.; Ostrowski, M. C.; Burry, R. W.; Boyaka, P. N.; Chan, M. K. Cry Protein Crystals: A Novel Platform for Protein Delivery. *PLOS ONE* **2015**, *10* (6), e0127669. https://doi.org/10.1371/journal.pone.0127669.
(36) Sun, Q.; Heater, B. S.; Li, T. L.; Ye, W.; Guo, Z.; Chan, M. K. Cry3Aa*SpyCatcher Fusion Crystals Produced in Bacteria as Scaffolds for Multienzyme Coimmobilization. *Bioconjug. Chem.* **2022**, *33* (2), 386–396. https://doi.org/10.1021/acs.bioconjchem.2c00003.
(37) 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. https://doi.org/10.1002/smll.201602703.
(38) Inokuma, Y.; Arai, T.; Fujita, M. Networked Molecular Cages as Crystalline Sponges for Fullerenes and Other Guests. *Nat. Chem.* **2010**, *2* (9), 780–783. https://doi.org/10.1038/nchem.742.
(39) Zhang, X.; Chen, Z.; Liu, X.; Hanna, S. L.; Wang, X.; Taheri-Ledari, R.; Maleki, A.; Li, P.; Farha, O. K. A Historical Overview of the Activation and Porosity of Metal–Organic Frameworks. *Chem. Soc. Rev.* **2020**, *49* (20), 7406–7427. https://doi.org/10.1039/D0CS00997K.
(40) Chen, Y.; Lu, W.; Schröder, M.; Yang, S. Analysis and Refinement of Host–Guest Interactions in Metal–Organic Frameworks. *Acc. Chem. Res.* **2023**, *56* (19), 2569–2581. https://doi.org/10.1021/acs.accounts.3c00243.
(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. https://doi.org/10.1021/jacs.5b07463.
(42) 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* (1), 1093. https://doi.org/10.1038/ncomms2093.

(43) Orun, A. R.; Shields, E. T.; Dmytriw, S.; Vajapayajula, A.; Slaughter, C. K.; Snow, C. D. Modular Protein–DNA Cocrystals as Precise, Programmable Assembly Scaffolds. *ACS Nano* **2023**, *17* (14), 13110–13120. https://doi.org/10.1021/acsnano.2c07282.
(44) Stuart, J. D.; Hartman, D. A.; Gray, L. I.; Jones, A. A.; Wickenkamp, N. R.; Hirt, C.; Safira, A.; Regas, A. R.; Kondash, T. M.; Yates, M. L.; Driga, S.; Snow, C. D.; Kading, R. C. Mosquito Tagging Using DNA-Barcoded Nanoporous Protein Microcrystals. *PNAS Nexus* **2022**, *1* (4), pgac190. https://doi.org/10.1093/pnasnexus/pgac190.
(45) 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. *Proc. Natl. Acad. Sci.* **2005**, *102* (20), 7157–7162. https://doi.org/10.1073/pnas.0409455102.
(46) Da Rosa, G.; Grille, L.; Calzada, V.; Ahmad, K.; Arcon, J. P.; Battistini, F.; Bayarri, G.; Bishop, T.; Carloni, P.; Cheatham Iii, T.; Collepardo-Guevara, R.; Czub, J.; Espinosa, J. R.; Galindo-Murillo, R.; Harris, S. A.; Hospital, A.; Laughton, C.; Maddocks, J. H.; Noy, A.; Orozco, M.; Pasi, M.; Pérez, A.; Petkevičiūtė-Gerlach, D.; Sharma, R.; Sun, R.; Dans, P. D. Sequence-Dependent Structural Properties of B-DNA: What Have We Learned in 40 Years? *Biophys. Rev.* **2021**, *13* (6), 995–1005. https://doi.org/10.1007/s12551-021-00893-8.
(47) Lu, B.; Vecchioni, S.; Ohayon, Y. P.; Sha, R.; Woloszyn, K.; Yang, B.; Mao, C.; Seeman, N. C. 3D Hexagonal Arrangement of DNA Tensegrity Triangles. *ACS Nano* **2021**, *15* (10), 16788–16793. https://doi.org/10.1021/acsnano.1c06963.
(48) 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. *Angew. Chem.* **2018**, *130* (38), 12684–12687. https://doi.org/10.1002/ange.201807223.
(49) Yan, X.; Wang, Y.; Ma, N.; Yu, Y.; Dai, L.; Tian, Y. Dynamically Reconfigurable DNA Origami Crystals Driven by a Designated Path Diagram. *J. Am. Chem. Soc.* **2023**, *145* (7), 3978–3986. https://doi.org/10.1021/jacs.2c10755.
(50) Wang, Y.; Yan, X.; Zhou, Z.; Ma, N.; Tian, Y. pH-Induced Symmetry Conversion of DNA Origami Lattices. *Angew. Chem. Int. Ed.* **2022**, *61* (40), e202208290. https://doi.org/10.1002/anie.202208290.
(51) Wang, Y.; Dai, L.; Ding, Z.; Ji, M.; Liu, J.; Xing, H.; Liu, X.; Ke, Y.; Fan, C.; Wang, P.; Tian, Y. DNA Origami Single Crystals with Wulff Shapes. *Nat. Commun.* **2021**, *12* (1), 3011. https://doi.org/10.1038/s41467-021-23332-4.
(52) Aksel, T.; Yu, Z.; Cheng, Y.; Douglas, S. M. Molecular Goniometers for Single-Particle Cryo-EM of DNA-Binding Proteins. February 28, 2020. https://doi.org/10.1101/2020.02.27.968883.
(53) D'Arcy, A.; Stihle, M.; Kostrewa, D.; Dale, G. Crystal Engineering: A Case Study Using the 24 kDa Fragment of the DNA Gyrase B Subunit from *Escherichia Coli*. *Acta Crystallogr. D Biol. Crystallogr.* **1999**, *55* (9), 1623–1625. https://doi.org/10.1107/S0907444999008136.
(54) Dale, G. E.; Oefner, C.; D'Arcy, A. The Protein as a Variable in Protein Crystallization. *J. Struct. Biol.* **2003**, *142* (1), 88–97. https://doi.org/10.1016/S1047-8477(03)00041-8.
(55) Holcomb, J.; Spellmon, N.; Zhang, Y.; Doughan, M.; Li, C.; Yang, Z.; 1 Department of Microbiology, Immunology, and Biochemistry, Wayne State University School of Medicine, Detroit, MI, USA; 2 Center for Molecular and Translational Medicine, Georgia State University, Atlanta, GA, USA. Protein Crystallization: Eluding the Bottleneck of X-Ray Crystallography. *AIMS Biophys.* **2017**, *4* (4), 557–575. https://doi.org/10.3934/biophy.2017.4.557.

(56) Li, Z.; Wang, S.; Nattermann, U.; Bera, A. K.; Borst, A. J.; Bick, M. J.; Yang, E. C.; Sheffler, W.; Lee, B.; Seifert, S.; Nguyen, H.; Kang, A.; Dalal, R.; Lubner, J. M.; Hsia, Y.; Haddox, H.; Courbet, A.; Dowling, Q.; Miranda, M.; Favor, A.; Etemadi, A.; Edman, N. I.; Yang, W.; Sankaran, B.; Negahdari, B.; Baker, D. Accurate Computational Design of 3D Protein Crystals. November 19, 2022. https://doi.org/10.1101/2022.11.18.517014.
(57) Lanci, C. J.; MacDermaid, C. M.; Kang, S.; Acharya, R.; North, B.; Yang, X.; Qiu, X. J.; DeGrado, W. F.; Saven, J. G. Computational Design of a Protein Crystal. *Proc. Natl. Acad. Sci.* **2012**, *109* (19), 7304–7309. https://doi.org/10.1073/pnas.1112595109.
(58) 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. *Biochem. Biophys. Res. Commun.* **2019**, *509* (2), 577–584. https://doi.org/10.1016/j.bbrc.2018.12.096.
(59) 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. https://doi.org/10.1038/nature14874.
(60) 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. https://doi.org/10.1021/jacs.1c04191.
(61) Orun, A.; Dmytriw, S.; Vajapayajula, A.; Snow, C. Stabilizing DNA–Protein Co-Crystals via Intra-Crystal Chemical Ligation of the DNA. *Crystals* **2021**, *12* (1), 49. https://doi.org/10.3390/cryst12010049.
(62) Orun, A. R.; Slaughter, C. K.; Shields, E. T.; Vajapayajula, A.; Jones, S.; Shrestha, R.; Snow, C. D. Tuning Chemical DNA Ligation within DNA Crystals and Protein–DNA Cocrystals. *ACS Nanosci. Au* **2024**, acsnanoscienceau.4c00013. https://doi.org/10.1021/acsnanoscienceau.4c00013.
(63) Gupta, M.; Vittal, J. J. Control of Interpenetration and Structural Transformations in the Interpenetrated MOFs. *Coord. Chem. Rev.* **2021**, *435*, 213789. https://doi.org/10.1016/j.ccr.2021.213789.
(64) 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* (6), 2105–2108. https://doi.org/10.1107/S160057671402281X.
(65) Li, S.; Olson, W. K.; Lu, X.-J. Web 3DNA 2.0 for the Analysis, Visualization, and Modeling of 3D Nucleic Acid Structures. *Nucleic Acids Res.* **2019**, *47* (W1), W26–W34. https://doi.org/10.1093/nar/gkz394.
(66) Grupa, U.; Liebl, K.; Zacharias, M. Orientation Dependence of DNA Blunt-End Stacking Studied by Free-Energy Simulations. *J. Phys. Chem. B* **2021**, *125* (51), 13850–13857. https://doi.org/10.1021/acs.jpcb.1c07829.
(67) Ayyer, K.; Yefanov, O. M.; Oberthür, D.; Roy-Chowdhury, S.; Galli, L.; Mariani, V.; Basu, S.; Coe, J.; Conrad, C. E.; Fromme, R.; Schaffer, A.; Dörner, K.; James, D.; Kupitz, C.; Metz, M.; Nelson, G.; Xavier, P. L.; Beyerlein, K. R.; Schmidt, M.; Sarrou, I.; Spence, J. C. H.; Weierstall, U.; White, T. A.; Yang, J.-H.; Zhao, Y.; Liang, M.; Aquila, A.; Hunter, M. S.; Robinson, J. S.; Koglin, J. E.; Boutet, S.; Fromme, P.; Barty, A.; Chapman, H. N. Macromolecular Diffractive Imaging Using Imperfect Crystals. *Nature* **2016**, *530* (7589), 202–206. https://doi.org/10.1038/nature16949.
(68) Pearce, N. M.; Krojer, T.; Bradley, A. R.; Collins, P.; Nowak, R. P.; Talon, R.; Marsden, B. D.; Kelm, S.; Shi, J.; Deane, C. M.; Von Delft, F. A Multi-Crystal Method for Extracting

Obscured Crystallographic States from Conventionally Uninterpretable Electron Density. *Nat. Commun.* **2017**, *8* (1), 15123. https://doi.org/10.1038/ncomms15123.
(69) Zhang, T.; Hartl, C.; Frank, K.; Heuer-Jungemann, A.; Fischer, S.; Nickels, P. C.; Nickel, B.; Liedl, T. 3D DNA Origami Crystals. *Adv. Mater.* **2018**, *30* (28), 1800273. https://doi.org/10.1002/adma.201800273.
(70) Komori, H. Crystal Structure of a Prokaryotic Replication Initiator Protein Bound to DNA at 2.6 A Resolution. *EMBO J.* **1999**, *18* (17), 4597–4607. https://doi.org/10.1093/emboj/18.17.4597.
(71) Hiromoto, T.; Ikura, T.; Honjo, E.; Blaber, M.; Kuroki, R.; Tamada, T. Creation of Cross-Linked Crystals With Intermolecular Disulfide Bonds Connecting Symmetry-Related Molecules Allows Retention of Tertiary Structure in Different Solvent Conditions. *Front. Mol. Biosci.* **2022**, *9*, 908394. https://doi.org/10.3389/fmolb.2022.908394.
(72) Quistgaard, E. M. A Disulfide Polymerized Protein Crystal. *Chem Commun* **2014**, *50* (95), 14995–14997. https://doi.org/10.1039/C4CC07326F.
(73) Negron Teron, K. I.; Das, C. Cocrystallization of Ubiquitin–Deubiquitinase Complexes through Disulfide Linkage. *Acta Crystallogr. Sect. Struct. Biol.* **2023**, *79* (11), 1044–1055. https://doi.org/10.1107/S2059798323008501.
(74) Xiang, Y.; Huang, R.-H.; Liu, X.-Z.; Zhang, Y.; Wang, D.-C. Crystal Structure of a Novel Antifungal Protein Distinct with FIve Disulfide Bridges from Eucommia Ulmoides Oliver at an Atomic Resolution. *J. Struct. Biol.* **2004**.
(75) Liu, T.; Wang, Y.; Luo, X.; Li, J.; Reed, S. A.; Xiao, H.; Young, T. S.; Schultz, P. G. Enhancing Protein Stability with Extended Disulfide Bonds.
(76) Hirakawa, H.; Kakitani, A.; Nagamune, T. Introduction of Selective Intersubunit Disulfide Bonds into Self-assembly Protein Scaffold to Enhance an Artificial Multienzyme Complex's Activity. *Biotechnol. Bioeng.* **2013**, *110* (7), 1858–1864. https://doi.org/10.1002/bit.24861.
(77) Dombkowski, A. A. Protein Disulfide Engineering. *FEBS Lett.* **2014**.
(78) Derewenda, Z. S. Rational Protein Crystallization by Mutational Surface Engineering. *Structure* **2004**, *12* (4), 529–535. https://doi.org/10.1016/j.str.2004.03.008.
(79) Derewenda, Z. S.; Vekilov, P. G. Entropy and Surface Engineering in Protein Crystallization. *Acta Crystallogr. D Biol. Crystallogr.* **2006**, *62* (1), 116–124. https://doi.org/10.1107/S0907444905035237.
(80) Thorsen, T. S.; Matt, R.; Weis, W. I.; Kobilka, B. K. Modified T4 Lysozyme Fusion Proteins Facilitate G Protein-Coupled Receptor Crystallogenesis. *Structure* **2014**, *22* (11), 1657–1664. https://doi.org/10.1016/j.str.2014.08.022.
(81) Liu, S.; Li, W. Protein Fusion Strategies for Membrane Protein Stabilization and Crystal Structure Determination. *Crystals* **2022**, *12* (8), 1041. https://doi.org/10.3390/cryst12081041.
(82) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; Bodenstein, S. W.; Evans, D. A.; Hung, C.-C.; O'Neill, M.; Reiman, D.; Tunyasuvunakool, K.; Wu, Z.; Žemgulytė, A.; Arvaniti, E.; Beattie, C.; Bertolli, O.; Bridgland, A.; Cherepanov, A.; Congreve, M.; Cowen-Rivers, A. I.; Cowie, A.; Figurnov, M.; Fuchs, F. B.; Gladman, H.; Jain, R.; Khan, Y. A.; Low, C. M. R.; Perlin, K.; Potapenko, A.; Savy, P.; Singh, S.; Stecula, A.; Thillaisundaram, A.; Tong, C.; Yakneen, S.; Zhong, E. D.; Zielinski, M.; Žídek, A.; Bapst, V.; Kohli, P.; Jaderberg, M.; Hassabis, D.; Jumper, J. M. Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3. *Nature* **2024**, *630* (8016), 493–500. https://doi.org/10.1038/s41586-024-07487-w.

(83) Zakeri, B.; Fierer, J. O.; Celik, E.; Chittock, E. C.; Schwarz-Linek, U.; Moy, V. T.; Howarth, M. Peptide Tag Forming a Rapid Covalent Bond to a Protein, through Engineering a Bacterial Adhesin. *Proc. Natl. Acad. Sci.* **2012**, *109* (12). https://doi.org/10.1073/pnas.1115485109.
(84) Das, R.; Baker, D. Macromolecular Modeling with Rosetta. *Annu. Rev. Biochem.* **2008**, *77* (1), 363–382. https://doi.org/10.1146/annurev.biochem.77.062906.171838.
(85) Schmidt, B.; Ho, L.; Hogg, P. J. Allosteric Disulfide Bonds. *Biochemistry* **2006**, *45* (24), 7429–7433. https://doi.org/10.1021/bi0603064.
(86) Fraenkel, E.; Rould, M. A.; Chambers, K. A.; Pabo, C. O. Engrailed Homeodomain-DNA Complex at 2.2 AÊ Resolution: A Detailed View of the Interface and Comparison with Other Engrailed Structures.
(87) Hirsch, J. A.; Aggarwal, A. K. Structure of the Even-Skipped Homeodomain Complexed to AT-Rich DNA: New Perspectives on Homeodomain Specificity. *EMBO J.* **1995**, *14* (24), 6280–6291. https://doi.org/10.1002/j.1460-2075.1995.tb00318.x.
(88) Royant, A.; Noirclerc-Savoye, M. Stabilizing Role of Glutamic Acid 222 in the Structure of Enhanced Green Fluorescent Protein. *J. Struct. Biol.* **2011**, *174* (2), 385–390. https://doi.org/10.1016/j.jsb.2011.02.004.
(89) Passner, J. M.; Ryoo, H. D.; Shen, L.; Mann, R. S.; Aggarwal, A. K. Structure of a DNA-Bound Ultrabithorax–Extradenticle Homeodomain Complex. *Nature* **1999**, *397* (6721), 714–719. https://doi.org/10.1038/17833.
(90) Fraenkel, E.; Pabo, C. O. Comparison of X-Ray and NMR Structures for the Antennapedia Homeodomain–DNA Complex. *Nat. Struct. Mol. Biol.* **1998**, *5* (8), 692–697. https://doi.org/10.1038/1382.
(91) Keller, W.; König, P.; Richmond, T. J. Crystal Structure of a bZIP/DNA Complex at 2.2 Å: Determinants of DNA Specific Recognition. *J. Mol. Biol.* **1995**, *254* (4), 657–667. https://doi.org/10.1006/jmbi.1995.0645.
(92) Cao, W.; Liu, L.; Lai, L.; Tang, Y. Molecular Recognition: Monomer of the Yeast Transcriptional Activator GCN4 Recognizes Its Dimer DNA Binding Target Sites Specifically. *Sci. China Ser. B Chem.* **2000**, *43* (5), 466–476. https://doi.org/10.1007/BF02969492.
(93) 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. https://doi.org/10.1021/jacs.1c00599.
(94) Mekkaoui, F.; Drewell, R. A.; Dresch, J. M.; Spratt, D. E. Experimental Approaches to Investigate Biophysical Interactions between Homeodomain Transcription Factors and DNA. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **2025**, *1868* (1), 195074. https://doi.org/10.1016/j.bbagrm.2024.195074.
(95) Orlomoski, R.; Bogle, A.; Loss, J.; Simons, R.; Dresch, J. M.; Drewell, R. A.; Spratt, D. E. Rapid and Efficient Purification of Drosophila Homeodomain Transcription Factors for Biophysical Characterization. *Protein Expr. Purif.* **2019**, *158*, 9–14. https://doi.org/10.1016/j.pep.2019.02.001.
(96) Rohs, R.; West, S. M.; Sosinsky, A.; Liu, P.; Mann, R. S.; Honig, B. The Role of DNA Shape in Protein–DNA Recognition. *Nature* **2009**, *461* (7268), 1248–1253. https://doi.org/10.1038/nature08473.
(97) Joshi, R.; Passner, J. M.; Rohs, R.; Jain, R.; Sosinsky, A.; Crickmore, M. A.; Jacob, V.; Aggarwal, A. K.; Honig, B.; Mann, R. S. Functional Specificity of a Hox Protein Mediated by

the Recognition of Minor Groove Structure. *Cell* **2007**, *131* (3), 530–543. https://doi.org/10.1016/j.cell.2007.09.024.
(98) Noyes, M. B.; Christensen, R. G.; Wakabayashi, A.; Stormo, G. D.; Brodsky, M. H.; Wolfe, S. A. Analysis of Homeodomain Specificities Allows the Family-Wide Prediction of Preferred Recognition Sites. *Cell* **2008**, *133* (7), 1277–1289. https://doi.org/10.1016/j.cell.2008.05.023.
(99) Keeble, A. H.; Yadav, V. K.; Ferla, M. P.; Bauer, C. C.; Chuntharpursat-Bon, E.; Huang, J.; Bon, R. S.; Howarth, M. DogCatcher Allows Loop-Friendly Protein-Protein Ligation. *Cell Chem. Biol.* **2022**, *29* (2), 339-350.e10. https://doi.org/10.1016/j.chembiol.2021.07.005.

APPENDIX I. SUPPLEMENTAL INFORMATION FOR CHAPTER 2
DESIGN AND CHARACTERIZATION OF POROUS PROTEIN-DNA CO-CRYSTALS FOR BIOMOLECULAR ENGINEERING

Table of Contents

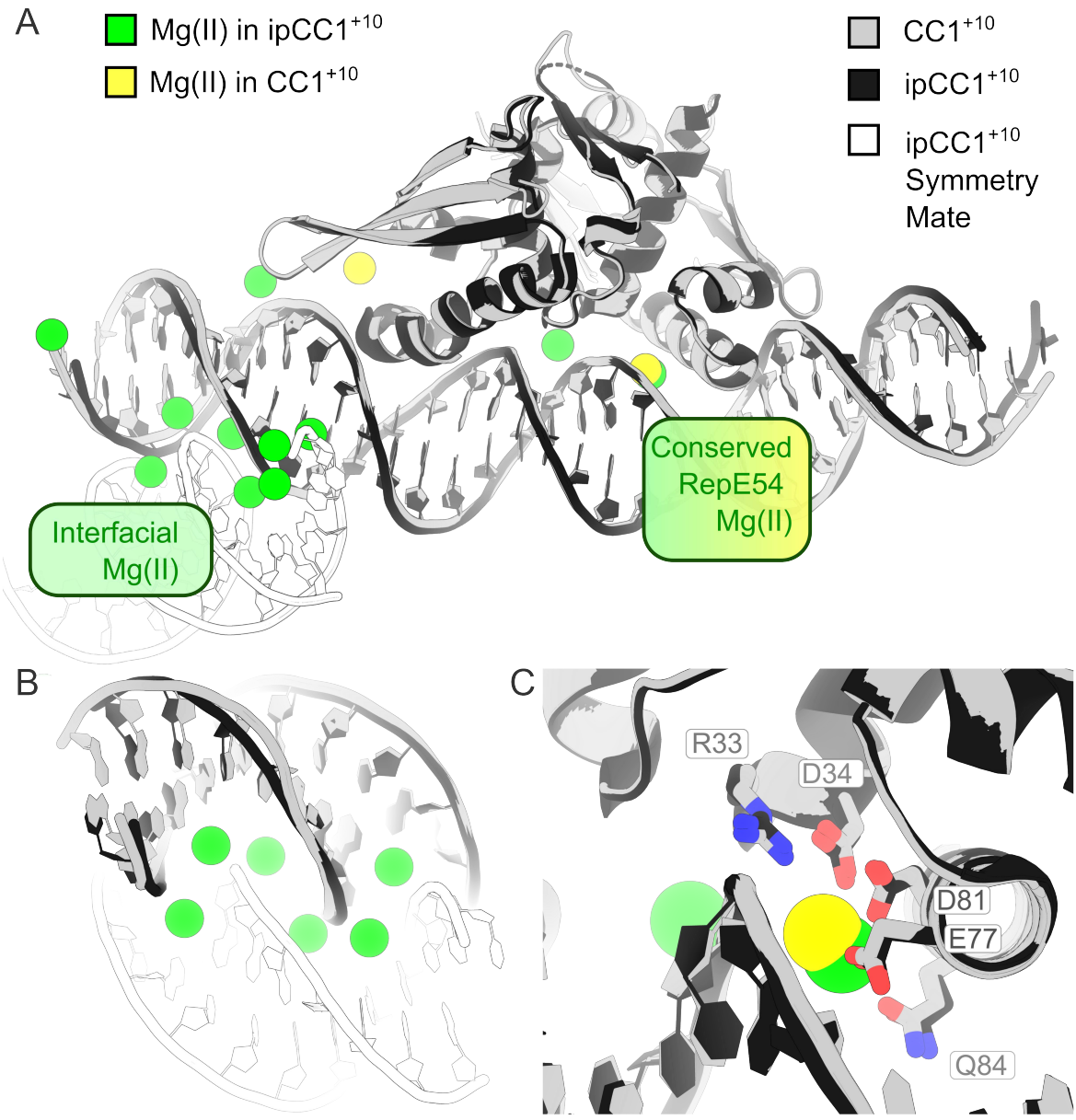


Fig. S2.1. CC1^{+10} vs ipCC1 ASU resolved ion coordinates

A key difference between ipCC1 growth and CC1 growth was the reduction in magnesium concentration. **(A)** Overview, showing an entire asymmetric unit for CC1^{+10} (gray, PDB:9NCR) superimposed onto ipCC1^{+10} (black, PDB:7U6K), where we also show the ipCC1^{+10} interpenetrating symmetry mate (white). The ipCC1^{+10} magnesium ions (green) dramatically outnumber the CC1^{+10} magnesium ions (yellow) and sodium ion (purple). **(B)** Close-up view of

the interdigitating DNA in ipCC1^{+10} , and six interfacial Mg(II). Thus, interfacial magnesium contacts appear to drive the close packing of DNA in the interpenetrating lattice. **(C)** We note that complete removal of divalent cations could hinder RepE54 binding, since one Mg(II) exists at the interface of RepE54 and the DNA, coordinated by R33, D34, E77, D81, and Q84. Figure prepared in PyMOL and Inkscape.

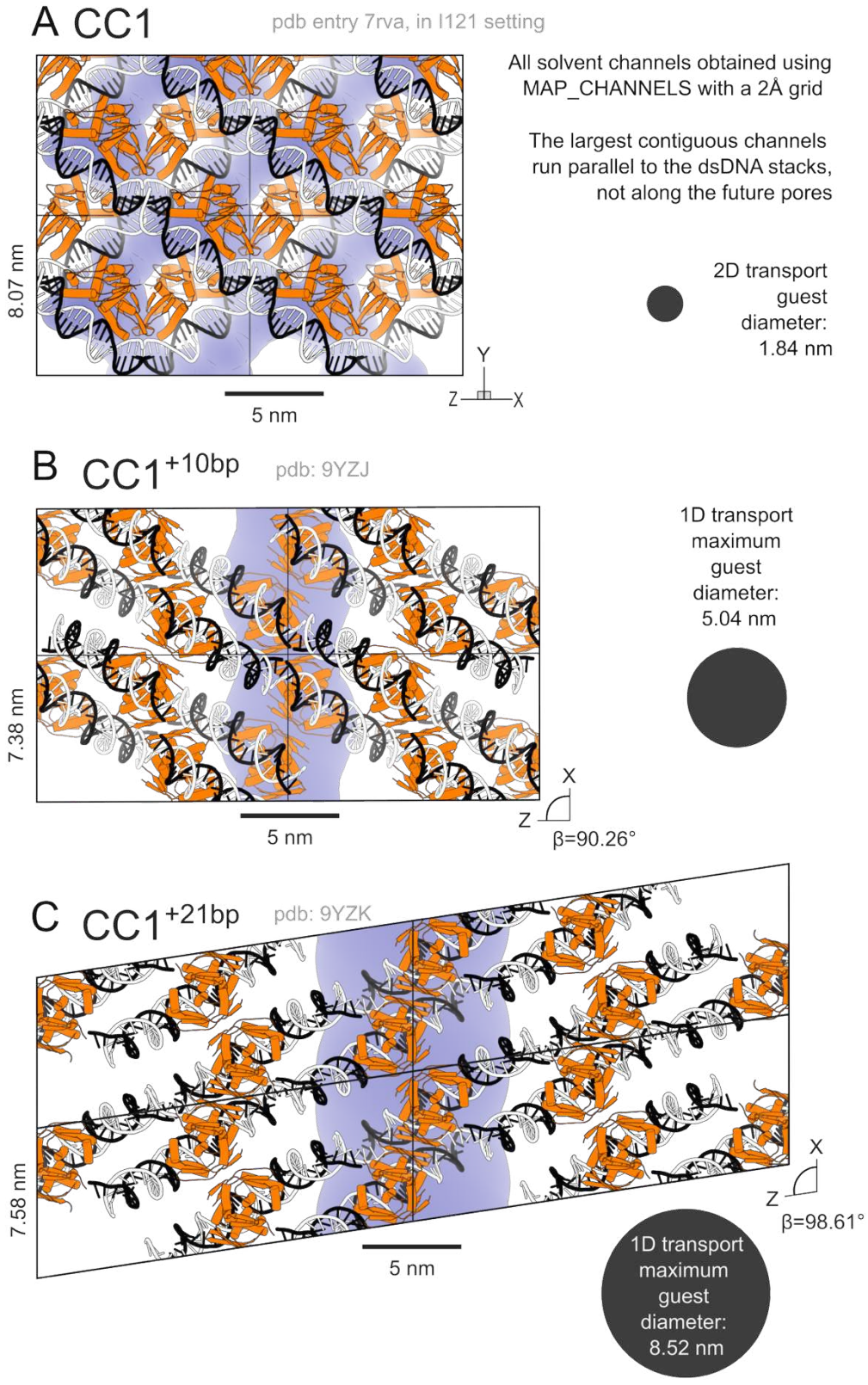

A CC1
pdb entry 7rva, in I121 setting
All solvent channels obtained using MAP_CHANNELS with a 2Å grid
The largest contiguous channels run parallel to the dsDNA stacks, not along the future pores
2D transport guest diameter: 1.84 nm
8.07 nm
5 nm
Y
Z
X
B CC1+10bp
pdb: 9YZJ
1D transport maximum guest diameter: 5.04 nm
7.38 nm
5 nm
X
Z
β=90.26°
C CC1+21bp
pdb: 9YZK
7.58 nm
5 nm
X
Z
β=98.61°
1D transport maximum guest diameter: 8.52 nm

Fig. S2.2. MAP_CHANNELS analysis of CC1 family of crystals

(A) MAP_CHANNELS (67) analysis of non-expanded CC1 (PDB 7RVA) using a 2Å grid revealed that the largest contiguous solvent channels ran along the coaxial dsDNA columns and were too small for macromolecular intra-crystal diffusion. **(B)** MAP_CHANNELS analysis of $CC1^{+10}$ (specifically PDB:9YZJ for this figure) using a 2 Å grid resulted in a calculated pore diameter of 5.04 nm. As a size comparison, we show to-scale eGFP (PDB:6YLQ) inscribed in a 5.04 nm circle, with an orientation selected to maximize the visible eGFP cross-section. **(C)** MAP_CHANNELS analysis of $CC1^{+21}$ using a 2 Å grid resulted in a calculated pore diameter of 8.52 nm (PDB 9YZK). See Table S8 for MAP_CHANNELS analysis of *all* new structures.

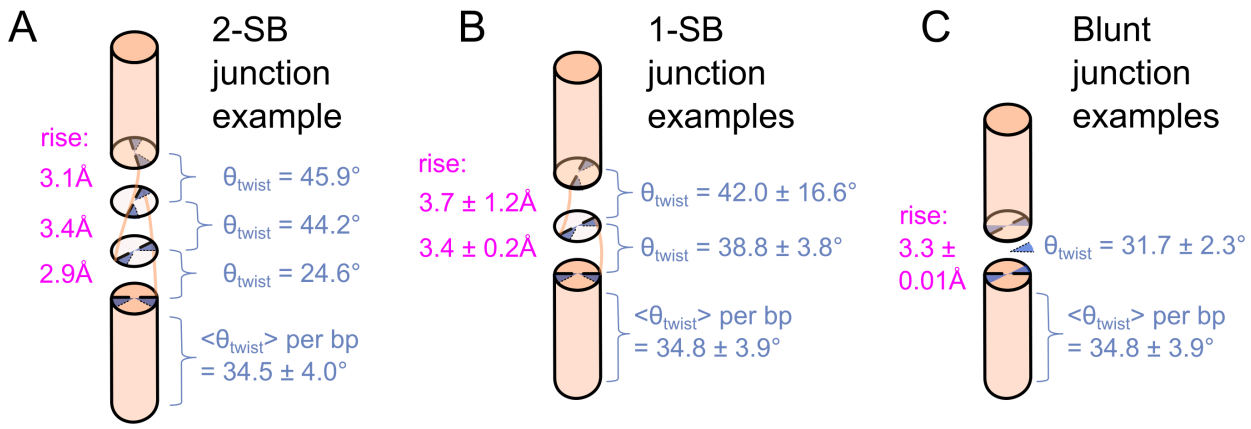


Fig. S2.3. x3DNA schematic of CC1^{+10} junction geometries

We used the program x3DNA[65] to extract DNA geometry parameters including rise (magenta) and twist (blue). **(A)** CC1^{+10} two sticky-base interface analysis of one model (PDB 9YZB). **(B)** CC1^{+10} one sticky-base interface x3DNA analysis of eight one-sticky-base CC1^{+10} structures (PDB 9YZA, 9NCR, 9NB0, 9YZC, 9YZD, 9YZE, 9YZF, and 9YZG). The data may be sufficient to note a tendency to slightly overwind DNA at the interface since the twist exceeds the average value for baseline B-DNA. **(C)** CC1^{+10} blunt end interface x3DNA analysis of two blunt-end CC1^{+10} structures (PDB 9YZI and 9YZJ). Notably, successful growth of blunt-end CC1 lattices may allow more tunable porous scaffolds via future growth of a wider variety of expansion sizes (i.e. beyond +10 and +21 bp).

The blunt variants were of particular interest because they allowed us to probe whether removing sticky overhangs would alter the helical twist across the interface, revealing otherwise hidden geometric strain introduced by overhang pairing. However, the average twist across the two blunt-ended examples was 31.7°, which was only 2.1° less than the twist values observed for sticky-end interfaces (33.8° ± 3.8). Figure prepared in Inkscape

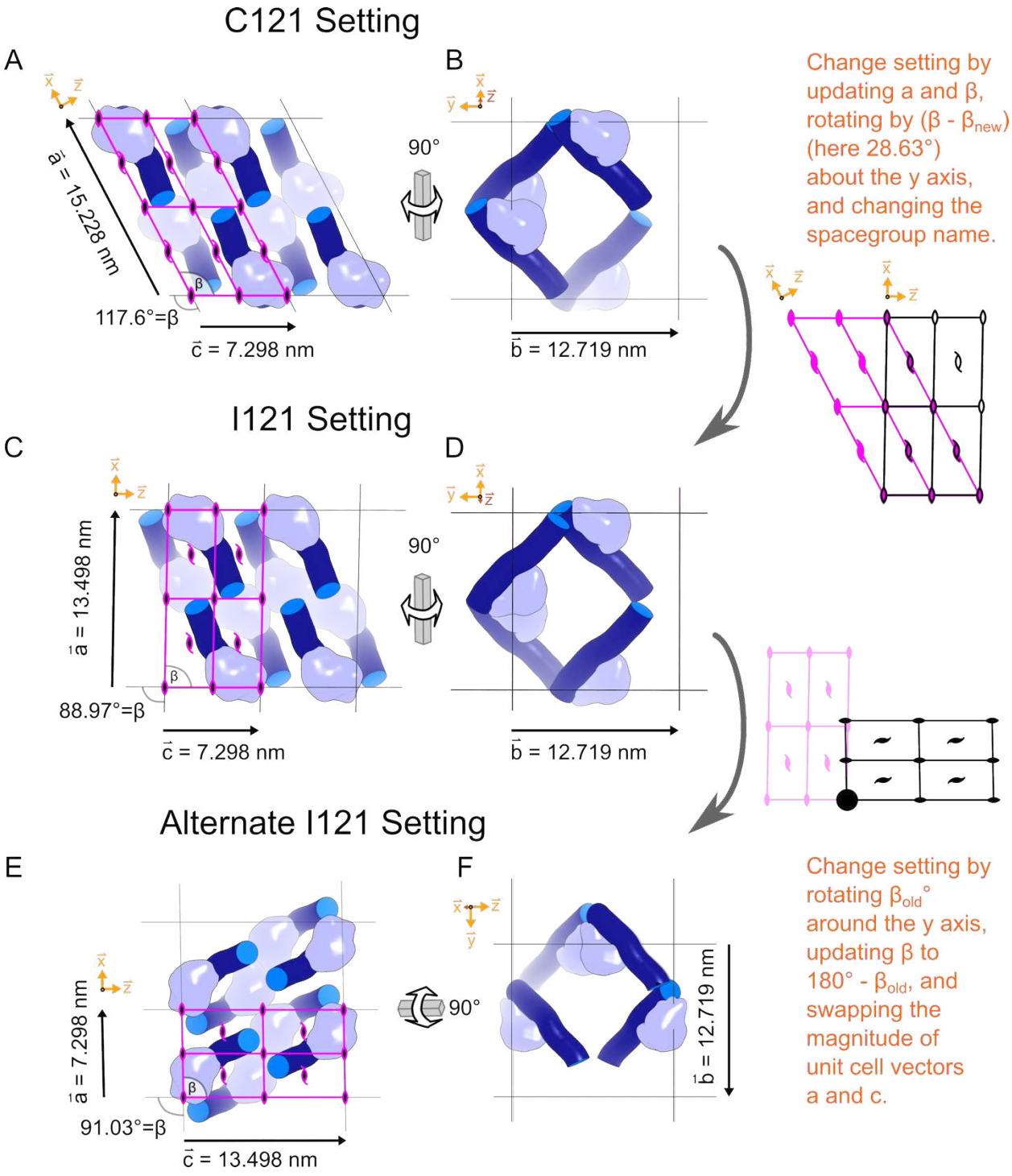


Fig. S2.4. Equivalent descriptions of the CC1 expanded lattice.

This lattice setting schematic uses a theoretical $CC1^{+10}$ model (expanded 7RVA). The monoclinic space group (#5) has several conventions. Typically, C121 is favored over I121,

obtuse β angles are preferred to acute, and the smallest reduced cell is favored. In our case, the C121 setting **(A)** is convenient for depiction of the lattice on the page because the coaxial DNA columns vectors lie in the plane perpendicular to the z-axis (which is not coincident with unit cell vector *c*). Starting from this standard view, a 90° rotation **(B)** puts the y-axis into the horizontal direction, and directs the c vector towards the reader, revealing the nanopores parallel to the unit cell vector *c*. However, the I121 setting **(C)** is also useful because the β value is close to 90°, and the I121 setting facilitates comparison to the I222 space group (adopted by the former interpenetrating lattice, e.g. PDB:7U6K) where β = 90° exactly. Also, the unit cell vector *c* that is parallel to the major nanopores **(D)** is now nearly aligned to the Cartesian z-axis. Finally, since the near-90° β values for CC1 crystals may sometimes fall below 90° (as shown here), they can violate the obtuse β convention. Therefore, one alternate setting is to rotate about the unique monoclinic y-axis (with the 2-fold axes) so that what used to be the *c* vector becomes an *a* vector aligned to the x-axis. **(E)** In this alternate setting, the nanopores are parallel to the *a* vector, which is aligned to the x-axis. The standard depiction of monoclinic cells (i.e. in space group diagrams) aligns the +*c* vector to the horizontal, rather than aligning the +*a* vector to the vertical. Thus, a horizontal 90° rotation **(F)** reveals the nanopores, since the x-axis is nearly perpendicular to the image. Figure prepared in PyMOL and Inkscape.

PDB Code	CC1^{+10} DNA Sequences
9Z08	5' – ACCTGTGACAAATTGCCCTCA – 3′ 3′ – GACACTGTTTAACGGGAGTCT – 5′ 5′ – GACGGTAATT – 3' 3′ – GCCATTAATG – 5'
9NCR	5′ – GCCGGACCTGTGACAAATTGCCCTCAGCCCG – 3′ 3′ – GGCCTGGACACTGTTTAACGGGAGTCGGGCC – 5′
9YZA	5′ – GCCGGACCTGTGACAAATTGCCCTCAGCCCG – 3′ 3′ – GGCCTGGACACTGTTTAACGGGAGTCGGGCC – 5′
9YZB	5' – TAATTACCTGTGACAAATTGCCCTCAGACGG – 3' 3' – TAATGGACACTGTTTAACGGGAGTCTGCCAT – 5'
9Z1A	5′ – CCTGTGACAAATTGCCCTCAGCCGCGCAGGC – 3′ 3′ – GACACTGTTTAACGGGAGTCGGCGCGTCCGG – 5′
9YZC	5′ – CCTGTGACAAATTGCCCTCAGCGCGCGCGCG – 3′ 3′ – GACACTGTTTAACGGGAGTCGCGCGCGCGCG – 5′
9NB0	5' – CCTGTGACAAATTGCCCTCAGGCTTGATGAG – 3' 3' – GACACTGTTTAACGGGAGTCCGAACTACTCG – 5'
9YZD	5' – CCTGTGACAAATTGCCCTCAGCTAATTAGGC – 3' 3′ – GACACTGTTTAACGGGAGTCGATTAATCCGG – 5′
9YZE	5' – CCTGTGACAAATTGCCCTGCTTGATGAGCAG – 3' 3′ – GACACTGTTTAACGGGACGAACTACTCGTCG – 5′
9YZF	5' – CCTGTGACAAATTGCCCTCATAATTAGGCCG – 3' 3′ – GACACTGTTTAACGGGAGTATTAATCCGGCG – 5′
9YZG	5' – CCTGTGACAAATTGCCCTCAGATGAGTCATA – 3' 3′ – GACACTGTTTAACGGGAGTCTACTCAGTATG – 5′
9YZI	5' – CCTGTGACAAATTGCCCTCAGCCCGGCCGGA – 3' 3′ – GGACACTGTTTAACGGGAGTCGGGCCGGCCT – 5′
9YZJ	5' – CCTGTGACAAATTGCCCTCAGCGTAATTAGG – 3' 3′ – GGACACTGTTTAACGGGAGTCGCATTAATCC – 5′
9Z55	5' – CCTGTGACAAATTGCCCTCAGCATGAGTCAT – 3' 3′ – GACACTGTTTAACGGGAGTCGTACTCAGTAG – 5′

Table S2.1 CC1^{+10} Oligonucleotide sequences

PDB Code:	Crystallization Condition:
9Z08	10 mM magnesium acetate, 1300 mM lithium sulfate, 50 mM sodium cacodylate pH 6.5
9YZA	40 mM magnesium acetate, 1600 mM lithium sulfate, 50 mM MES pH 6.5
9NCR	30 mM magnesium acetate, 1000 mM lithium sulfate, 50 mM MES pH 6.5
9YZB	10 mM magnesium acetate, 1600 mM lithium sulfate, 50 mM MES pH 6.5
9Z1A	80 mM magnesium acetate, 1200 mM lithium sulfate, 50 mM MES pH 6.5
9YZC	30 mM magnesium acetate, 1400 mM lithium sulfate, 50 mM MES pH 6.5
9NB0	15 mM magnesium acetate, 600 mM lithium sulfate, 50 mM MES pH 6.5
9YZD	30 mM magnesium acetate, 1300 mM lithium sulfate, 50 mM MES pH 6.5
9YZE	20 mM magnesium acetate, 300 mM lithium sulfate, 50 mM MES pH 6.5
9YZF	50 mM magnesium acetate, 1100 mM lithium sulfate, 50 mM MES pH 6.5
9YZG	20 mM magnesium acetate, 1400 mM lithium sulfate, 50 mM MES pH 6.5
9YZI	50 mM magnesium acetate, 1600 mM lithium sulfate, 50 mM MES pH 6.5
9YZJ	30 mM magnesium acetate, 1700 mM lithium sulfate, 50 mM MES pH 6.5
9Z55	40 mM magnesium acetate, 1800 mM lithium sulfate, 50 mM MES pH 6.5

Table S2.2. $CC1^{+10}$ crystallization conditions

Crystallization/PDB code:	dsDNA Sequence
9YZK	5′ – CGGGATGAGGTCTGAGGGCAATTTGTCACAGGTGGGAAGATA – 3′ 3′ – CCTACTCCAGACTCCCGTTAAACAGTGTCCACCCTTCTATGC – 5′
Yes, no diffraction	5′ – CTGGATGAGGTCTGAGGGCAATTTGTCACAGGTGGGAAGATA – 3′ 3′ – ACCTACTCCAGACTCCCGTTAAACAGTGTCCACCCTTCTATG – 5′
No	5′ – CTGGATGAGGTCTGAGGGCAATTTGTCACAGGTGGGAAGATA – 3′ 3′ – CCTACTCCAGACTCCCGTTAAACAGTGTCCACCCTTCTATGA – 5′
No	5′ – CTGGATGAGGTCTGAGGGCAATTTGTCACAGGTGGGAAGATAC – 3′ 3′ – ACCTACTCCAGACTCCCGTTAAACAGTGTCCACCCTTCTATGG – 5′
No	5′ – TGGGATCGGATGATAACAGGTCTGAGGGCAATTTGTCACAGG – 3′ 3′ – CACCCTAGCCTACTATTGTCCAGACTCCCGTTAAACAGTGTC – 5′
No	5′ – GCTGTGACAAATTGCCCTCAGGATGCTGTTGCGCTTGGACGC – 3′ 3′ – ACACTGTTTAACGGGAGTCCTACGACAACGCGAACCTGCGGC – 5′
No	5′ – CTGGATGAGGTCTGAGGGCAATTTGTCACAGGTGGGAAGATA – 3′ 3′ – GACCTACTCCAGACTCCCGTTAAACAGTGTCCACCCTTCTAT – 5′
No	5′ – TACATGCCGGACCTGTGACAAATTGCCCTCAGCCCGGGTAAT – 3′ 3′ – TGTACGGCCTGGACACTGTTTAACGGGAGTCGGGCCCATTAA – 5′
No	5′ – ACTGGATGAGGTCTGAGGGCAATTTGTCACAGGTGGGAAGAT – 3′ 3′ – CCTACTCCAGACTCCCGTTAAACAGTGTCCACCCTTCTATGA – 5′
No	5′ – TGGAAGCCGCCTGTGACAAATTGCCCTCAGTGCTGTTGCGCT – 3′ 3′ – CTTCGGCGGACACTGTTTAACGGGAGTCACGACAACGCGAAC – 5′

Table S2.3. CC1^{+21} oligonucleotide sequences

PDB Code:	Crystallization Condition:
9YZK	10 mM magnesium acetate, 1300 mM lithium sulfate, 50 mM sodium cacodylate pH 6.5

Table S2.4. CC1^{+21} Crystallization conditions

PDB	EDC Ligated?	Guest Status	MAP_CHANNELS Solvent Content	MAP_CHANNELS Maximum Radius 1D [Å]	Reso-lution [Å]
9NB0	No	no	0.82	25.6	3.2
9NCR	No	no	0.83	27	3.2
9YZA	No	no	0.83	27.2	3.1
9YZB	No	no	0.83	27.4	4.1
9YZC	No	no	0.83	27.8	3.2
9YZD	No	no	0.83	27.2	3.1
9YZE	No	no	0.84	26.4	4.1
9YZF	No	no	0.83	27	3.1
9YZG	No	no	0.82	26.6	3.1
9YZI	No	no	0.84	26.4	4
9YZJ	No	no	0.83	25.2	3
9YZL	Yes	EVE-HD	0.79	25.4	3.0
9YZM	Yes	EVE-HD	0.78	26.2	3.1
9YZN	Yes	EVE-HD	0.8	26.4	3.2
9YZO	Yes	EVE-HD	0.79	25.6	3.0
9YZP	Yes	UBX-HD	0.82	27.4	3.7
9YZQ	Yes	EnH-eGFP	0.82	26.4	3.8
9YZR	Yes	UBX-HD	0.81	20.6	3.4
9YZS	Yes	EVE-HD	0.78	22	3.1
9YZT	Yes	ANTP-HD	0.79	20	3.3
9YZU	Yes	bZip	0.8	25.8	3.1
9Z08	No	no	0.84	26.6	4.2
9Z1A	No	no	0.83	27.8	3.1
9Z1B	Yes	no	0.83	28	3.9
9Z1C	Yes	no	0.83	26.4	3.0
9Z1D	Yes	no	0.83	27	3.1
9Z44	Yes	C-clamp	0.84	28.4	7.2
9Z4E	Yes	bZip	0.81	22.8	3.6
9Z55	No	no	0.83	25.6	3.9
9YZK	No	no	0.9	42.6	5.2

Table S2.5. MAP_CHANNELS calculations for all $CC1^{+10}$ and $CC1^{+21}$ structures

PDB code	9Z08
Data collection	
Light source	Synchrotron
Wavelength (Å)	1
Resolution Range (Å)	37.31-4.01 (4.44 -4.22)
Space group	I121
Unit cell dimensions	
a, b, c (Å)	75.045, 133.505, 135.653
α, β, ɣ (°)	90.00, 96.08, 90.00
Unique reflections	19241 (1068)
Multiplicity	1.0 (1.0)
Completeness (%)	97.62 (86.56)
Mean I/sigma(I)	3.51 (0.76)
Wilson B-factor	146.58
R-merge	0.0891 (1.3)
R-meas	0.1260 (1.839)
R-pim	0.0891 (1.3)
CC1/2	0.996 (0.465)
Refinement	
Reflections used in refinement	9405 (1068)
Reflections used for Rfree	940 (117)
R-work	0.2398 (0.3558)
R-free	0.2537 (0.4238)
Number of non-hydrogen atoms	2980
Macromolecules	2980
Ligands	2
Solvent	0
Protein residues	213
RMS (bonds) (Å)	0.21
RMS (angles) (°)	0.41
Ramachandran favored (%)	98
Ramachandran allowed (%)	2
Ramachandran outliers (%)	0
Rotamer outliers (%)	0
Clashscore	3
Average B-factor	181

Table S2.6. X-ray diffraction statistics for $CC1^{+10}$ PDB entry 9Z08

PDB code	9YZA	9NCR
Data collection		
Light source	Synchrotron	Synchrotron
Wavelength (Å)	1	1
Resolution Range (Å)	46.67-3.05 (3.12-3.05)	34.76-3.24 (3.32-3.24)
Space group	I121	I121
Unit cell dimensions		
a, b, c (Å)	74.35, 124.35, 137.54	76.03, 124.86, 139.09
α, β, ɣ (°)	90.00, 90.30, 90.00	90.00, 91.85, 90.00
Unique reflections	39868 (218)	42728 (79)
Multiplicity	1.8 (3.5)	1 (1.1)
Completeness (%)	95.37 (99.1)	96.62 (90.90)
Mean I/sigma(I)	6.6 (1.07)	5.56 (0.93)
Wilson B-factor	102.99	104.68
R-merge	0.079 (2.277)	0.052 (1.415)
R-meas	0.094 (2.695)	0.074 (1.415)
R-pim	0.050 (1.428)	0.052 (1.415)
CC1/2	0.999 (0.603)	0.998 (0.417)
Refinement		
Reflections used in refinement	22801(1255)	20000 (1217)
Reflections used for Rfree	2279 (142)	1994 (130)
R-work	0.2959 (0.6554)	0.2203 (0.4211)
R-free	0.3090 (0.5821)	0.2310 (0.4611)
Number of non-hydrogen atoms	3044	3060
Macromolecules	3044	3060
Ligands	2	2
Solvent	0	0
Protein residues	217	216
RMS (bonds) (Å)	0.17	0.18
RMS (angles) (°)	0.41	0.41
Ramachandran favored (%)	98	98
Ramachandran allowed (%)	2	2
Ramachandran outliers (%)	0	0
Rotamer outliers (%)	1	0
Clashscore	7	7
Average B-factor	132	179

Table S2.7. X-ray diffraction statistics for $CC1^{+10}$ PDB entries 9YZA and 9NCR

PDB code	9YZB	9Z1A
Data collection		
Light source	Synchrotron	Synchrotron
Wavelength (Å)	1	1
Resolution Range (Å)	38.81-3.19 (3.27-3.19)	46.81-3.10 (3.17-3.10)
Space group	I121	I121
Unit cell dimensions		
a, b, c (Å)	75.53, 130.59, 135.53	74.75, 124.53, 137.76
α, β, ɣ (°)	90.00, 90.13, 90.00	90.00, 90.31, 90.00
Unique reflections	38587 (2738)	21447 (553)
Multiplicity	1.0 (1.0)	3.2 (3.2)
Completeness (%)	90.74 (88.62)	100 (80.7)
Mean I/sigma(I)	3.36 (0.88)	3.2 (0.9)
Wilson B-factor	79.7	100.86
R-merge	0.1158 (3.925)	0.191 (0.904)
R-meas	0.1638 (5.551)	0.224 (1.06)
R-pim	0.1158 (3.925)	0.127 (0.603)
CC1/2	0.996 (0.252)	0.999 (0.739)
Refinement		
Reflections used in refinement	19949 (966)	21445 (1107)
Reflections used for Rfree	1987 (101)	2140 (122)
R-work	0.3045 (0.4672)	0.3070 (0.6290)
R-free	0.3281 (0.4944)	0.3351 (0.6628)
Number of non-hydrogen atoms	3110	2936
Macromolecules	3110	2936
Ligands	1	
Solvent	0	0
Protein residues	221	216
RMS (bonds) (Å)	0.16	0.14
RMS (angles) (°)	0.42	0.34
Ramachandran favored (%)	98	98
Ramachandran allowed (%)	2	2
Ramachandran outliers (%)	0	0
Rotamer outliers (%)	0	0
Clashscore	4	3
Average B-factor	101	124

Table S2.8. X-ray diffraction statistics for $CC1^{+10}$ PDB entries 9ZYB and 9Z1A

PDB code	9YZC	9NB0
Data collection		
Light source	Synchrotron	Synchrotron
Wavelength (Å)	1	1
Resolution Range (Å)	46.80-3.16 (3.23-3.16)	39.55-3.20 (3.28-3.20)
Space group	I121	I121
Unit cell dimensions		
a, b, c (Å)	76.03, 124.86, 139.09	75.53, 119.24, 139.79
α, β, ɣ (°)	90.00, 91.85, 90.00	90.00, 90.63, 90.00
Unique reflections	21911 (4255)	19630 (3755)
Multiplicity	3.5 (3.6)	3.2 (2.0)
Completeness (%)	99.1 (99.7)	97.61 (96.06)
Mean I/sigma(I)	4.2 (1.1)	14.6 (3.13)
Wilson B-factor	100.86	109.47
R-merge	0.186 (1.062)	0.02144 (0.213)
R-meas	0.221 (1.251)	0.03032 (0.3012)
R-pim	0.118 (0.655)	0.02144 (0.213)
CC1/2	0.995 (0.737)	1 (0.995)
Refinement		
Reflections used in refinement	21911 (1160)	19481 (1186)
Reflections used for Rfree	2190 (138)	1963 (137)
R-work	0.2893 (0.4323)	0.1785 (0.3437)
R-free	0.2994 (0.4269)	0.1968 (0.3910)
Number of non-hydrogen atoms	3070	3070
Macromolecules	3070	3070
Ligands	5	1
Solvent	0	0
Protein residues	215	216
RMS (bonds) (Å)	0.16	0.42
RMS (angles) (°)	0.43	0.73
Ramachandran favored (%)	98	98
Ramachandran allowed (%)	2	2
Ramachandran outliers (%)	0	0
Rotamer outliers (%)	0	0
Clashscore	5	7
Average B-factor	98	165

Table S2.9. X-ray diffraction statistics for CC1^{+10} PDB entries 9YZC and 9NB0

PDB code	**9YZD**	**9YZE**

Data collection		
Light source	Synchrotron	Synchrotron
Wavelength (Å)	1	1
Resolution Range (Å)	45.29-3.05 (3.16-3.05)	45.86-4.07 (4.28-4.07)
Space group	I121	I121
Unit cell dimensions		
a, b, c (Å)	76.03, 124.86, 139.09	76.03, 124.86, 139.09
α, β, ɣ (°)	90.00, 91.85, 90.00	90.00, 91.85, 90.00
Unique reflections	23118 (1365)	26992 (3349)
Multiplicity	3.4 (3.5)	3.2 (2.9)
Completeness (%)	96.91 (99.4)	92.5 (78.7)
Mean I/sigma(I)	7.4 (1.1)	2.2 (1.0)
Wilson B-factor	92.15	95.14
R-merge	0.099 (0.913)	0.796 (0.731)
R-meas	0.118 (1.080)	0.945 (0.886)
R-pim	0.063 (0.572)	0.526 (0.493)
CC1/2	0.973 (0.779)	0.954(0.373)
Refinement		
Reflections used in refinement	23116 (1236)	9975 (1135)
Reflections used for Rfree	2278 (129)	987 (122)
R-work	0.2535 (0.4228)	0.2470 (0.3735)
R-free	0.2645 (0.4521)	0.2757 (0.4248)
Number of non-hydrogen atoms	3040	2981
Macromolecules	3040	2981
Ligands	1	1
Solvent	0	0
Protein residues	216	217
RMS (bonds) (Å)	0.23	0.18
RMS (angles) (°)	0.47	0.45
Ramachandran favored (%)	98	98
Ramachandran allowed (%)	2	2
Ramachandran outliers (%)	0	0
Rotamer outliers (%)	0	0
Clashscore	6	4
Average B-factor	163	131

Table S2.10. X-ray diffraction statistics for $CC1^{+10}$ PDB entries 9YZD and 9YZE

PDB code	9YZF	9YZG
Data collection		
Light source	Synchrotron	Synchrotron
Wavelength (Å)	1	1
Resolution Range (Å)	46.22-3.07 (3.13-3.07)	46.33-3.09 (3.16-3.09)
Space group	I121	I121
Unit cell dimensions		
a, b, c (Å)	74.55, 121.56, 142.89	74.17, 127.04, 131.72
α, β, ɣ (°)	90.00, 95.17, 90.00	90.00, 91.20, 90.00
Unique reflections	23419 (4252)	22096 (3992)
Multiplicity	3.5 (3.7)	3.5 (3.6)
Completeness (%)	98.5 (98.7)	98.7 (99.0)
Mean I/sigma(I)	6.6 (1.3)	7.7 (1.0)
Wilson B-factor	90.27	97.8
R-merge	0.101 (0.900)	0.096 (1.394)
R-meas	0.120 (1.055)	0.114 (1.639)
R-pim	0.064 (0.547)	0.061 (0.854)
CC1/2	0.997 (0.853)	0.993 (0.669)
Refinement		
Reflections used in refinement	23302 (1267)	21971 (1209)
Reflections used for Rfree	2314 (138)	2190 (134)
R-work	0.2323 (0.5056)	0.2736 (0.4703)
R-free	0.2408 (0.5519)	0.3007 (0.4898)
Number of non-hydrogen atoms	3099	3059
Macromolecules	3099	3059
Ligands	1	2
Solvent	0	0
Protein residues	219	217
RMS (bonds) (Å)	0.19	0.16
RMS (angles) (°)	0.43	0.42
Ramachandran favored (%)	97	97
Ramachandran allowed (%)	2	2
Ramachandran outliers (%)	1	1
Rotamer outliers (%)	0	0
Clashscore	5	6
Average B-factor	150	159

Table S2.11. X-ray diffraction statistics for $CC1^{+10}$ PDB entries 9YZF and 9YZG

PDB code	9YZI	9YZJ
Data collection		
Light source	Synchrotron	Synchrotron
Wavelength (Å)	1	1
Resolution Range (Å)	46.70-4.00 (4.21-4.00)	46.01-2.97 (3.16-2.97)
Space group	I121	I121
Unit cell dimensions		
a, b, c (Å)	74.90, 121.62, 138.04	73.78, 132.65, 126.77
α, β, ɣ (°)	99.00, 90.41, 90.00	90.00, 90.26, 90.00
Unique reflections	10367 (2874)	24810 (4023)
Multiplicity	4.0 (3.3)	3.5 (3.5)
Completeness (%)	98.4 (96.6)	98.4 (99)
Mean I/sigma(I)	3.2 (0.9)	11.4 (1.0)
Wilson B-factor	151.42	99.16
R-merge	0.342 (1.551)	0.065 (1.102)
R-meas	0.403 (1.982)	0.078 (1.301)
R-pim	0.211 (1.067)	0.042 (0.684)
CC1/2	0.958 (0.291)	0.998 (0.798)
Refinement		
Reflections used in refinement	10313 (1291)	21738 (1199)
Reflections used for Rfree	1022 (144)	2172 (126)
R-work	0.2961 (0.3818)	0.2721 (0.4371)
R-free	0.3216 (0.4159)	0.2917 (0.5107)
Number of non-hydrogen atoms	2853	2896
Macromolecules	2853	2896
Ligands	4	4
Solvent	0	0
Protein residues	203	221
RMS (bonds) (Å)	0.26	0.45
RMS (angles) (°)	0.51	0.71
Ramachandran favored (%)	97	96
Ramachandran allowed (%)	3	3
Ramachandran outliers (%)	0	1
Rotamer outliers (%)	0	3
Clashscore	8	17
Average B-factor	224	156

Table S2.12. X-ray diffraction statistics for CC1^{+10} PDB entries 9YZI and 9YZJ

PDB code	9Z55
Data collection	
Light source	Synchrotron
Wavelength (Å)	1

Resolution Range (Å)	64.94-3.94 (4.12-3.94)
Space group	C121
Unit cell dimensions	
a, b, c (Å)	146.48, 129.88, 73.32
α, β, ɣ (°)	90.00, 118.58, 90.00
Unique reflections	10668 (3014)
Multiplicity	7.0 (7.1)
Completeness (%)	99.5 (99.9)
Mean I/sigma(I)	8.8 (2.1)
Wilson B-factor	189.68
R-merge	0.095 (0.804)
R-meas	0.103 (0.867)
R-pim	0.039 (0.322)
CC1/2	0.998 (0.926)
Refinement	
Reflections used in refinement	10599 (1189)
Reflections used for Rfree	1058 (127)
R-work	0.2968 (0.4202)
R-free	0.3161 (0.4580)
Number of non-hydrogen atoms	2862
Macromolecules	2862
Ligands	0
Solvent	0
Protein residues	216
RMS (bonds) (Å)	0.25
RMS (angles) (°)	0.49
Ramachandran favored (%)	96
Ramachandran allowed (%)	4
Ramachandran outliers (%)	0
Rotamer outliers (%)	0
Clashscore	6
Average B-factor	299

Table S2.13. X-ray diffraction statistics for $CC1^{+10}$ PDB entry 9Z55

PDB code	**9YZK**
Data collection	
Light source	Synchrotron
Wavelength (Å)	1
Resolution Range (Å)	47.53-5.10 (5.37-5.10)
Space group	I121
Unit cell dimensions	
a, b, c (Å)	75.84, 163.02, 192.28
α, β, ɣ (°)	90.00, 98.61, 90.00
Unique reflections	14099 (3944)
Multiplicity	3.4 (3.1)
Completeness (%)	98.4 (97.7)
Mean I/sigma(I)	4.1 (0.2)
Wilson B-factor	304.94
R-merge	0.119 (5.581)
R-meas	0.142 (6.708)
R-pim	0.075 (3.673)
CC1/2	0.997 (0.196)
Refinement	
Reflections used in refinement	9121 (1050)
Reflections used for Rfree	910 (117)
R-work	0.2763 (0.5614)
R-free	0.3040 (0.6073)
Number of non-hydrogen atoms	3367
Macromolecules	3367
Ligands	1
Solvent	0
Protein residues	224
RMS (bonds) (Å)	0.22
RMS (angles) (°)	0.52
Ramachandran favored (%)	97
Ramachandran allowed (%)	3
Ramachandran outliers (%)	0
Rotamer outliers (%)	0
Clashscore	6
Average B-factor	383

PDB code	**9YZK**
Data collection	

Light source	Synchrotron
Wavelength (Å)	1
Resolution Range (Å)	47.53-5.10 (5.37-5.10)
Space group	I121
Unit cell dimensions	
a, b, c (Å)	75.84, 163.02, 192.28
α, β, ɣ (°)	90.00, 98.61, 90.00
Unique reflections	14099 (3944)
Multiplicity	3.4 (3.1)
Completeness (%)	98.4 (97.7)
Mean I/sigma(I)	4.1 (0.2)
Wilson B-factor	304.94
R-merge	0.119 (5.581)
R-meas	0.142 (6.708)
R-pim	0.075 (3.673)
CC1/2	0.997 (0.196)
Refinement	
Reflections used in refinement	9121 (1050)
Reflections used for Rfree	910 (117)
R-work	0.2763 (0.5614)
R-free	0.3040 (0.6073)
Number of non-hydrogen atoms	3367
Macromolecules	3367
Ligands	1
Solvent	0
Protein residues	224
RMS (bonds) (Å)	0.22
RMS (angles) (°)	0.52
Ramachandran favored (%)	97
Ramachandran allowed (%)	3
Ramachandran outliers (%)	0
Rotamer outliers (%)	0
Clashscore	6
Average B-factor	383

Table S2.14. X-ray diffraction statistics for $CC1^{+21}$ PDB entry 9YZK

RepE54 Transcription Factor protein sequence.

RepE54 transcription factor (sequence below) was cloned into pSB3 vector (Addgene plasmid #82027) with an N-terminal His-tag.

MRGSHHHHHHGSMAETAVINHKKRKNSPRIVQSNDLTEAAYSLSRDQKRMLYLFVDQI
RKSDGTLQEHDGICEIHVAKYAEIFGLTSAEASKDIRQALKSFAGKEVVFYRPEEDAGDE
KGYESFPWFIKPAHSPSRGLYSVHINPYLIPFFIGLQNRFTQFRLSETKEITNPYAMRLYES
LCQYRKPDGSGIVSLKIDWIIERYQLPQSYQRMPDFRRRFLQVCVNEINSRTPMRLSYIEK
KKGRQTTHIVFSFRDITSMTTG**

RepE54 (L53G, Q54G, E55G) Transcription factor variant 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.

MRGSHHHHHHGSMAETAVINHKKRKNSPRIVQSNDLTEAAYSLSRDQKRMLYLFVDQI
RKSDGTGGGHDGICEIHVAKYAEIFGLTSAEASKDIRQALKSFAGKEVVFYRPEEDAGDE
KGYESFPWFIKPAHSPSRGLYSVHINPYLIPFFIGLQNRFTQFRLSETKEITNPYAMRLYES
LCQYRKPDGSGIVSLKIDWIIERYQLPQSYQRMPDFRRRFLQVCVNEINSRTPMRLSYIEK
KKGRQTTHIVFSFRDITSMTTG**

Protocol S2.1. Protein sequences for cloning and overexpression in *E.coli*

APPENDIX II. SUPPORTING INFORMATION FOR CHAPTER 3. PROTEIN ENGINEERING WITHIN THE $CC1^{+10}$ LATTICE: STRUCTURAL TOLERANCE AND FUNCTIONAL EXPANSION

Table of Contents

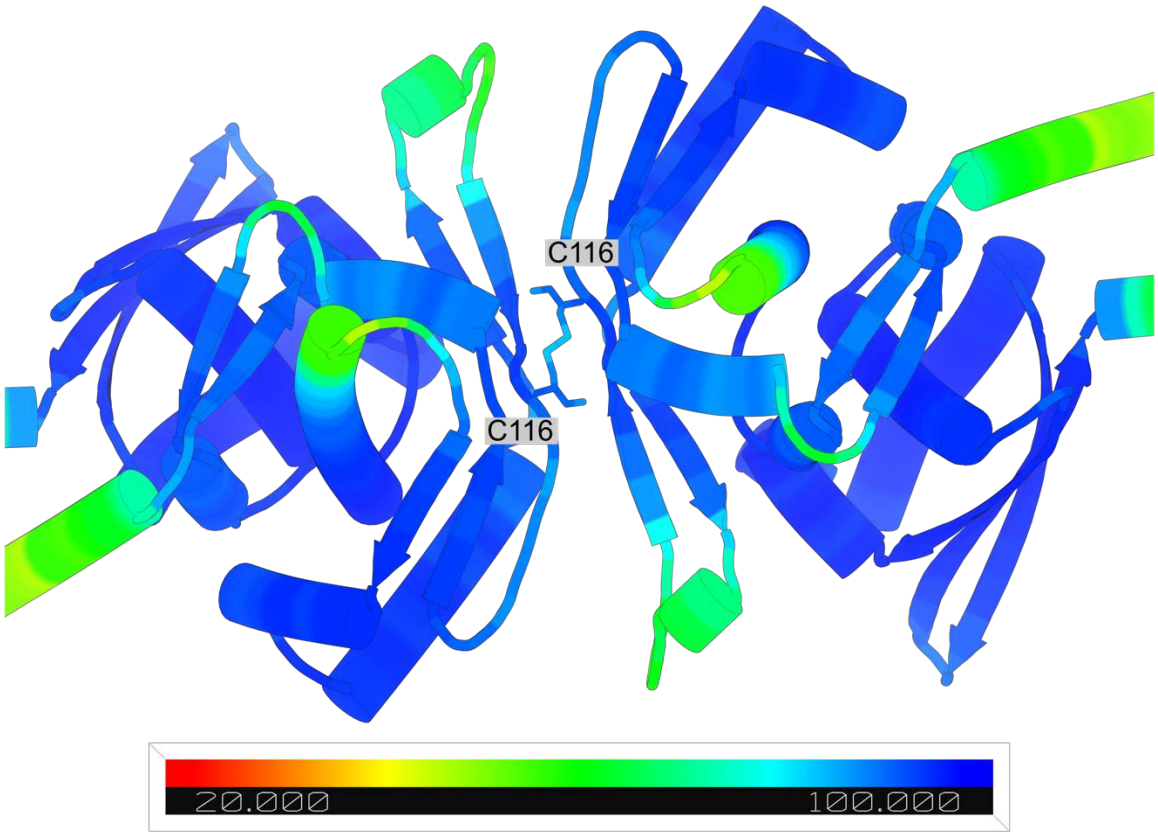


Fig. S3.1. Alphafold-3 modelling of CC1 cysteine mutation site I116C

Alphafold-3 structure prediction of RepE54-I116C mutant with residue 116 shown as sticks colored by pLDDT.

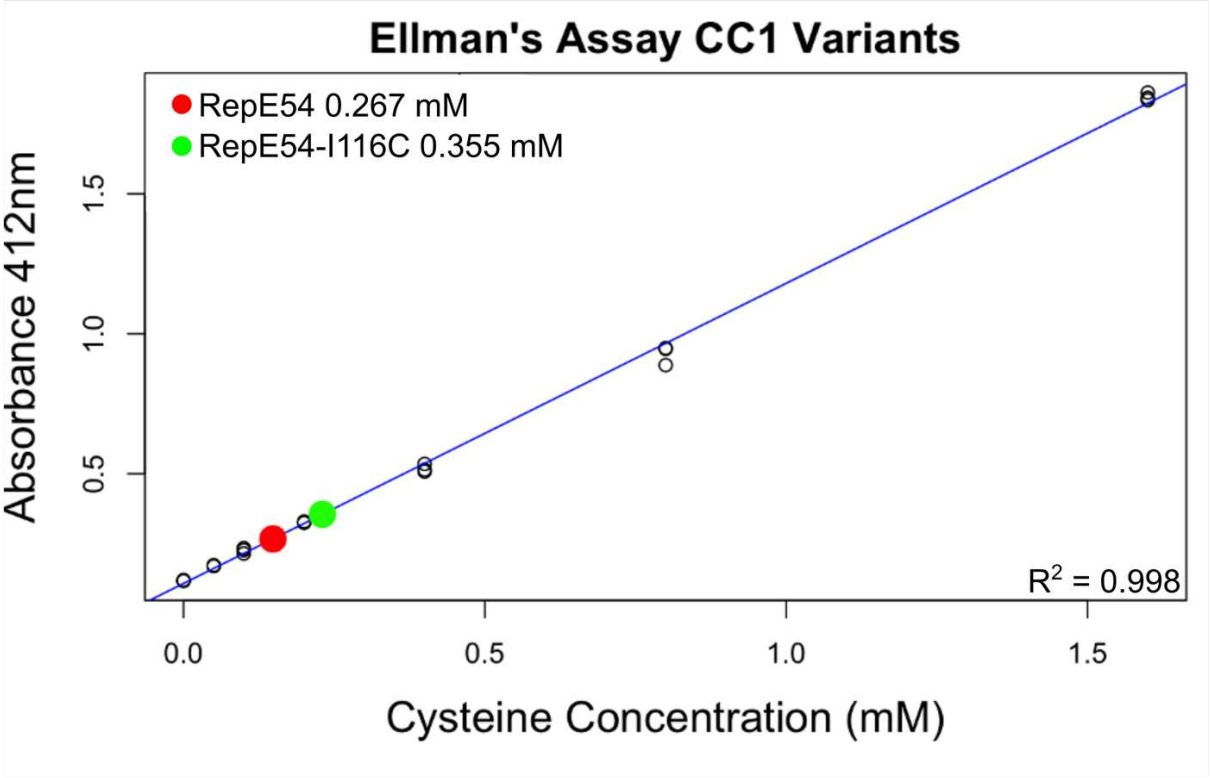


Fig. S3.2. Ellman's assay of wild-type RepE54 and cysteine mutant I116C

Ellman's assay results of wild-type RepE54 and RepE54-I116C mutant, a detailed protocol can be found in chapter 3 (Protocol 3.5.6).

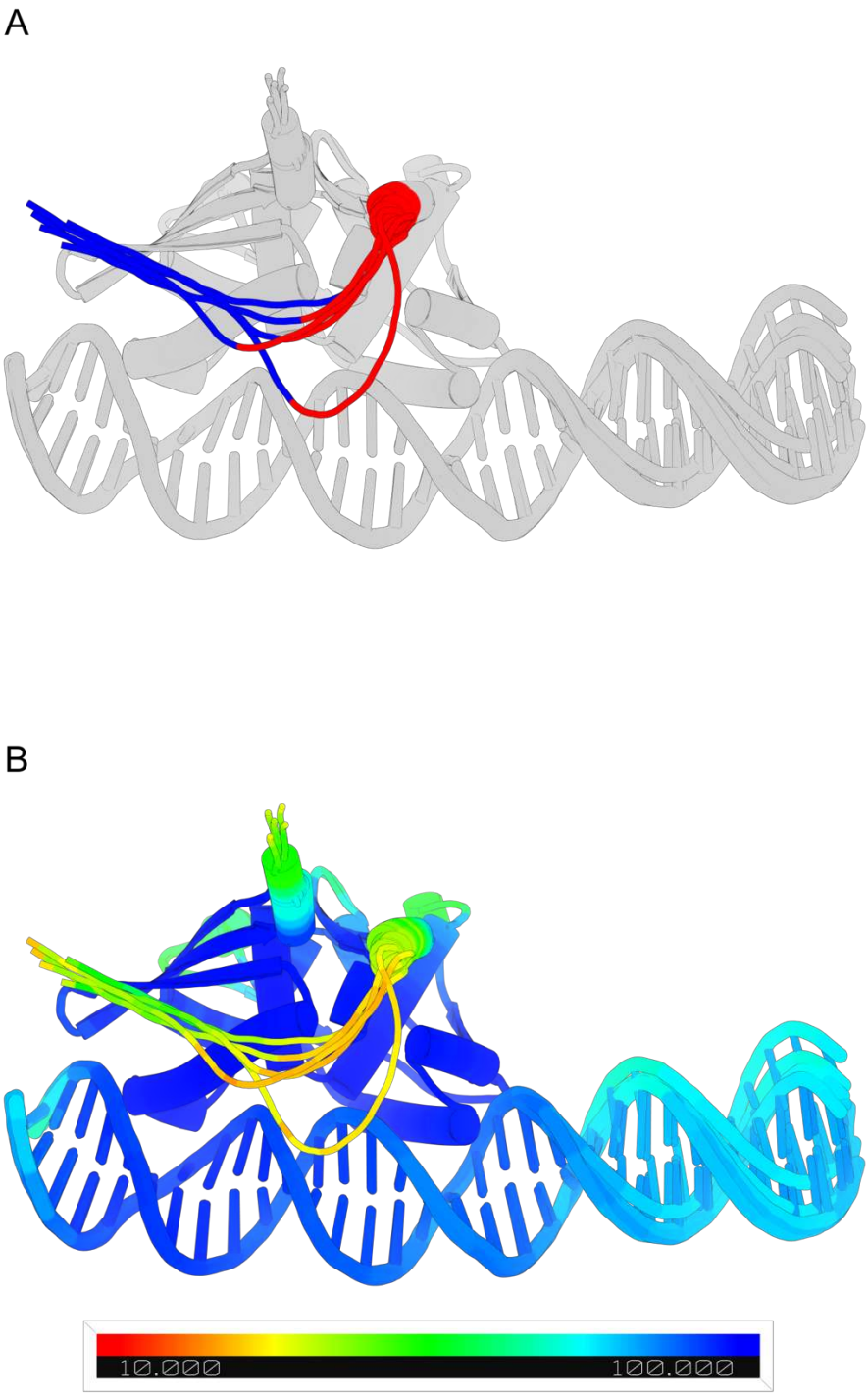


Fig. S3.3. Spytag-RepE54 fusion linker design

(**A**) Alphafold-3 $CC1^{+10}$ ASU with N-terminal Spytag peptide fusion (blue) and varied flexible linker lengths $(GGS)_{2\text{-}6}$ (red). (**B**) Alphafold-3 $CC1^{+10}$ ASU with N-terminal Spytag and varied flexible linker lengths $(GGS)_{2\text{-}6}$ colored by pLDDT.

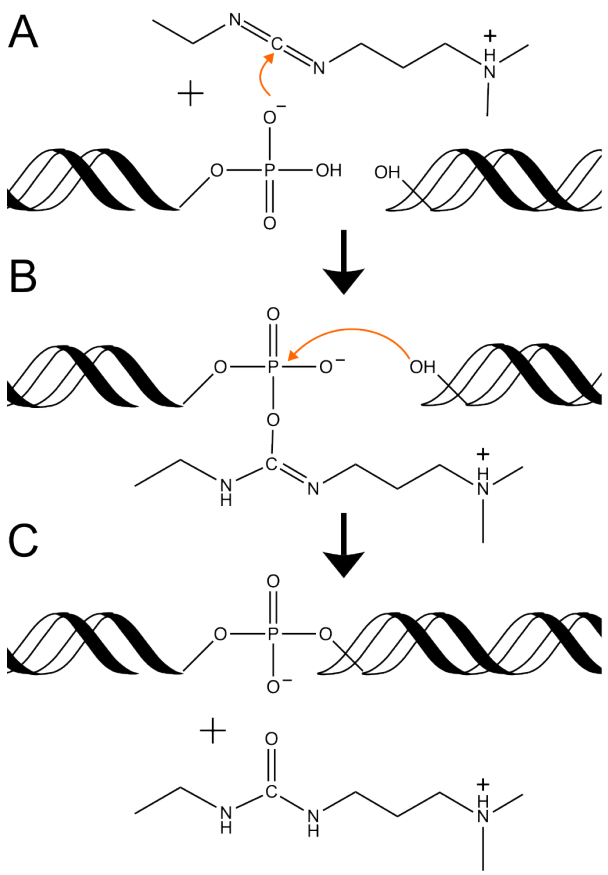


Fig. S3.4. EDC crosslinking mechanism

(**A**) A terminal 5′ hydroxyl and a terminal 3′ phosphate on neighboring DNA chains. The phosphate interacts with EDC to form an intermediate (**B**) The hydroxyl displaces the reactive intermediate (**C**) Formation of zero length crosslink between neighboring DNA chains.

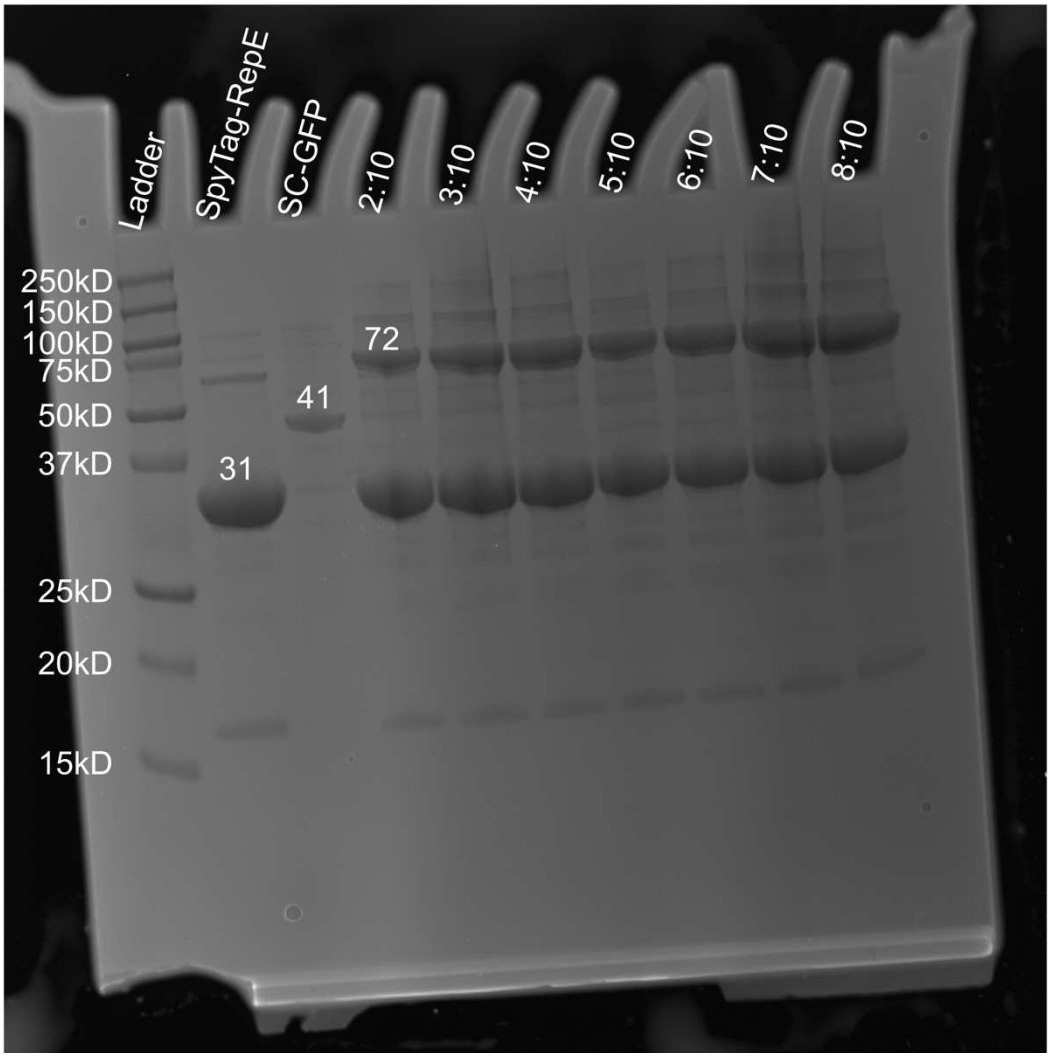


Fig. S3.5. SpyTag-RepE54 and SpyCatcher-GFP Conjugation

SDS-PAGE of SpyTag-RepE54 and SpyCatcher-GFP run independently. Other lanes contain samples of SpyTag-RepE54 mixed in excess with increasing concentrations of SpyCatcher-GFP and incubated at 30°C for 30 minutes and run on the gel.

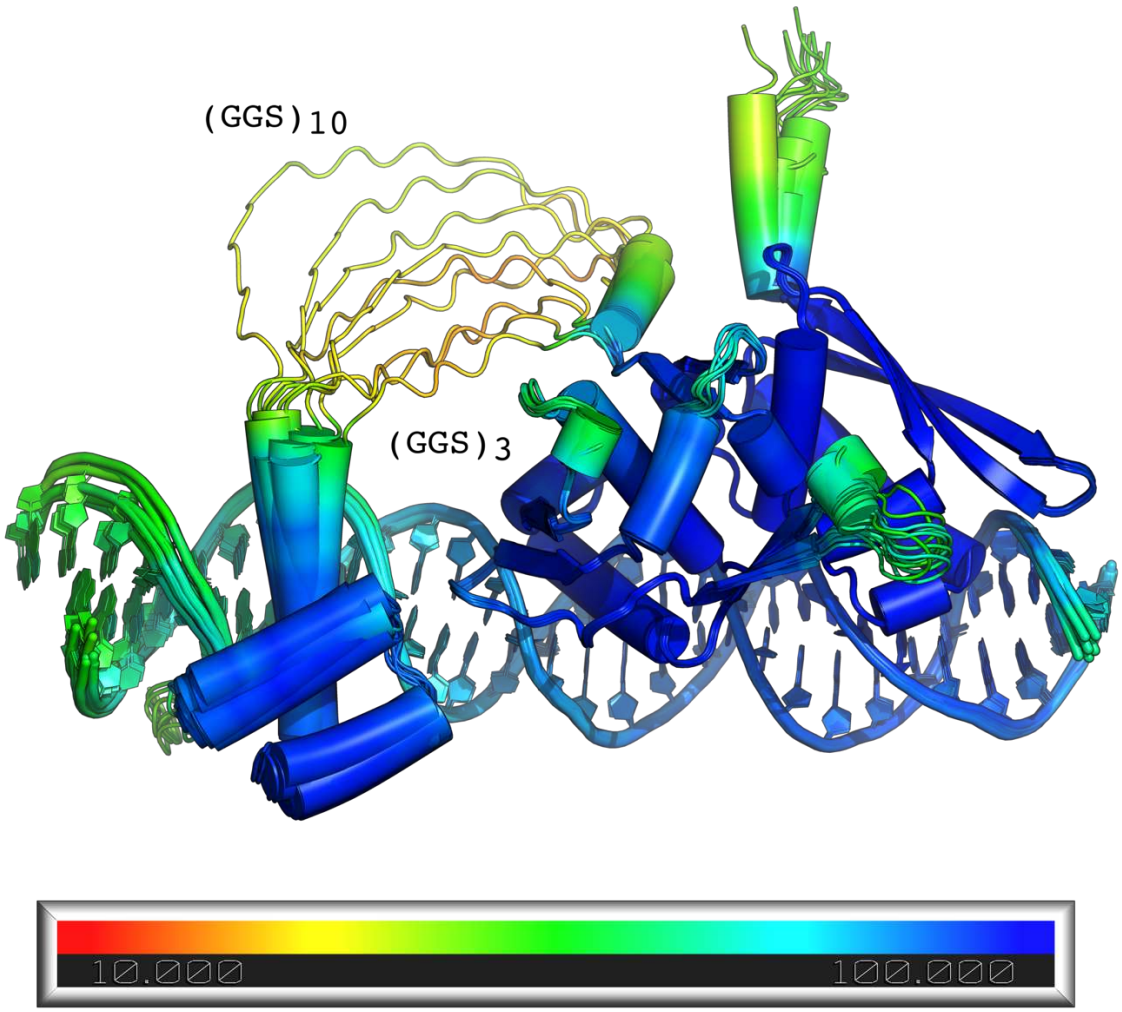


Fig. S3.6. EnH-RepE54 fusion linker design

Alphafold-3 structure predictions of EnH-RepE54 fusion proteins with varied flexible linker lengths, $(GGS)_3$-$(GGS)_{10}$ colored by pLDDT.

PDB Code	Crystallization Condition
9Z41	20mM magnesium acetate, 1.6M lithium sulfate, 50mM MES pH 6.5
9Z42	40mM magnesium acetate, 900mM lithium sulfate, 50mM MES pH 6.5
9Z43	50mM Magnesium acetate, 800mM Lithium sulfate, 50mM MES pH 6.5

Table S3.1. $CC1^{+10}$ mutant protein crystallization conditions

PDB Code	$CC1^{+10}$ DNA Sequences
9Z41	5′ - CCTGTGACAAATTGCCCTCAGCCGCGCAGGC - 3′ 3′ - GACACTGTTTAACGGGAGTCGGCGCGTCCGG - 5′
9Z42	5′ - GCCGGACCTGTGACAAATTGCCCTCAGCCCG - 3′ 3′ - GGCCTGGACACTGTTTAACGGGAGTCGGGCC - 5′
9Z43	5' - CCTGTGACAAATTGCCCTCAGCTAATTAGGC - 3' 3′ - GACACTGTTTAACGGGAGTCGATTAATCCGG - 5′

Table S3.2. $CC1^{+10}$ mutant protein oligonucleotide sequences

Dataset name	050624_1B-12_1	120624_1B-09_1
PDB code	9Z41	9Z42
Data collection		
Light source	Synchrotron	Synchrotron
Wavelength (Å)	1	1
Resolution Range (Å)	47.79-3.07 (3.07-3.14)	45.90-3.13 (3.21-3.13)
Space group	I121	I121
Unit cell dimensions		
a, b, c (Å)	75.05, 133.50, 136.65	73.66, 126.96, 133.02
α, β, ɣ (°)	90.00, 96.08, 90.00	90.00, 90.31, 90.00
Unique reflections	23263 (3724)	21512 (3928)
Multiplicity	7.1 (7.3)	3.5 (3.6)
Completeness (%)	99.4 (99.1)	99.4 (99.8)
Mean I/sigma(I)	10.2 (1.0)	4.3 (1.0)
Wilson B-factor	123.41	96.13
R-merge	0.095 (1.829)	0.145 (1.370)
R-meas	0.102 (1.968)	0.173 (1.615)
R-pim	0.039 (0.722)	0.093 (0.848)
CC1/2	0.998 (0.882)	0.995 (0.596)
Refinement		
Reflections used in refinement	22204 (203)	21443 (1411)
Reflections used for Rfree	2225 (26)	2009 (152)
R-work	0.2404 (0.4699)	0.2547 (0.4335)
R-free	0.2523 (0.4709)	0.2799 (0.4732)
Number of non-hydrogen atoms	3060	2957
Macromolecules	3060	2957
Ligands	3	3
Solvent	0	0
Protein residues	225	221
RMS (bonds) (Å)	0.47	0.18
RMS (angles) (°)	0.77	0.43
Ramachandran favored (%)	94	98
Ramachandran allowed (%)	5	2
Ramachandran outliers (%)	1	0
Rotamer outliers (%)	3	0
Clashscore	13	5
Average B-factor	168	148

Table S3.3. X-ray diffraction statistics for CC1^{+10} PDB entries 9Z41 and 9Z42

Dataset name	041325_1D-09_1
PDB code	9Z43
Data collection	
Light source	Synchrotron
Wavelength (Å)	1
Resolution Range (Å)	45.90-3.02 (3.09-3.02)
Space group	I121
Unit cell dimensions	
a, b, c (Å)	72.36, 130.83, 128.87
α, β, ɣ (°)	90.00, 90.90, 90.00
Unique reflections	23282 (3745)
Multiplicity	3.5 (3.6)
Completeness (%)	98.7 (98.6)
Mean I/sigma(I)	8.5 (1.2)
Wilson B-factor	132.39
R-merge	0.097 (1.149)
R-meas	0.115 (1.348)
R-pim	0.061 (0.699)
CC1/2	0.999 (0.768)
Refinement	
Reflections used in refinement	23150 (1325)
Reflections used for Rfree	2304 (143)
R-work	0.2983 (0.6139)
R-free	0.3191 (0.6591)
Number of non-hydrogen atoms	3016
Macromolecules	3016
Ligands	5
Solvent	0
Protein residues	219
RMS (bonds) (Å)	0.15
RMS (angles) (°)	0.39
Ramachandran favored (%)	98
Ramachandran allowed (%)	1
Ramachandran outliers (%)	1
Rotamer outliers (%)	0
Clashscore	5
Average B-factor	122

Table S3.4. X-ray diffraction statistics for $CC1^{+10}$ PDB entry 9Z43

RepE54 (L53G, Q54G, E55G, I116C) Transcription factor variant 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.

MRGSHHHHHHGSMAETAVINHKKRKNSPRIVQSNDLTEAAYSLSRDQKRMLYLFVDQI
RKSDGTGGGHDGICEIHVAKYAEIFGLTSAEASKDIRQALKSFAGKEVVFYRPEEDAGDE
KGYESFPWFCKPAHSPSRGLYSVHINPYLIPFFIGLQNRFTQFRLSETKEITNPYAMRLYE
SLCQYRKPDGSGIVSLKIDWIIERYQLPQSYQRMPDFRRRFLQVCVNEINSRTPMRLSYIE
KKKGRQTTHIVFSFRDITSMTTG**

Spytag-RepE54 fusion protein sequence.

Spytag-RepE54 fusion protein (sequence below) was cloned into the pETDuet-1 vector (Novagen) with a C-terminal His-tag.

MAHIVMVDAYKPTKGGSGGSKKRKNSPRIVQSNDLTEAAYSLSRDQKRMLYLFVDQIR
KSDGTLQEHDGICEIHVAKYAEIFGLTSAEASKDIRQALKSFAGKEVVFYRPEEDAGDEK
GYESFPWFIKPAHSPSRGLYSVHINPYLIPFFIGLQNRFTQFRLSETKEITNPYAMRLYESL
CQYRKPDGSGIVSLKIDWIIERYQLPQSYQRMPDFRRRFLQVCVNEINSRTPMRLSYIEK
KKGRQTTHIVFSFRDITSMTTHHHHHH**

Engrailed homeodomain – RepE54 (EnH-RepE54) fusion protein sequence.

EnH-RepE54 fusion protein (sequence below) was commercially synthesized into the pET28 vector (Twist Bioscience) with a C-terminal His-tag and an additional N-terminal His-tag upstream in the same reading frame.

MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGSMEKRPRTAFSSEQLARLKREFNE
NRYLTERRRQQLSSELGLNEAQIKIWFQNKRAKIKKSTGGSGGSGGSGGSKRKNSPRIVQ
SNDLTEAAYSLSRDQKRMLYLFVDQIRKSDGTLQEHDGICEIHVAKYAEIFGLTSAEASK
DIRQALKSFAGKEVVFYRPEEDAGDEKGYESFPWFIKPAHSPSRGLYSVHINPYLIPFFIG
LQNRFTQFRLSETKEITNPYAMRLYESLCQYRKPDGSGIVSLKIDWIIERYQLPQSYQRM
PDFRRRFLQVCVNEINSRTPMRLSYIEKKKGRQTTHIVFSFRDITSMTTGHHHHHH**

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

Fusion protein structure prediction was performed using the AlphaFold 3 online server. Glycine–glycine–serine (GGS) tripeptide repeats were inserted between the amino acid sequences of RepE54, each fusion target (SpyTag or Engrailed homeodomain), and the $CC1^{+10}$ oligonucleotide sequence. Each fusion target was screened with linkers containing 2–10 GGS repeats. Predicted models were evaluated based on three criteria: (1) proper folding of each protein domain, (2) minimal strain or distortion within the flexible linker, and (3) high overall model confidence (mean pLDDT > 0.75).

Protocol S3.2. Alphafold-3 modelling of flexible fusion linkers

APPENDIX III. SUPPORTING INFORMATION FOR CHAPTER 4.
POST-CRYSTALLIZATION GUEST BINDING AND STRUCTURE DETERMINATION WITHIN A PROTEIN-DNA CO-CRYSTAL

Table of Contents

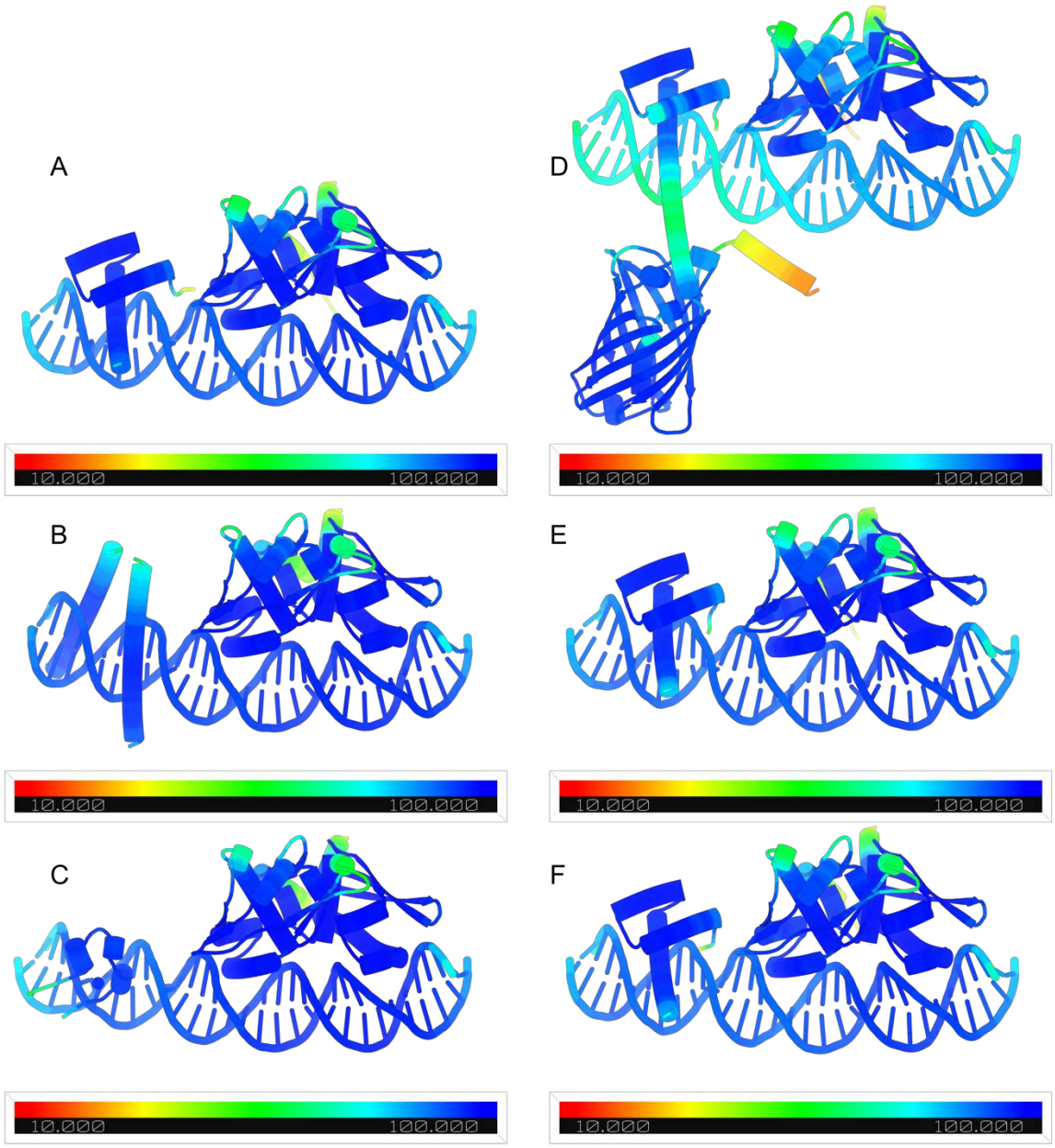


Fig. S4.1. Alphafold-3 structure prediction of Guest Proteins bound to CC1^{+10} ASU.

(A) AF3 prediction of AntP-HD bound to CC1^{+10} ASU with insert 5' TGATGAGCAG 3' colored by pLDDT. **(B)** AF3 prediction of bZip bound to CC1^{+10} ASU with insert 5' ATGAGTCATA 3' colored by pLDDT. **(C)** AF3 prediction of C-clamp bound to CC1^{+10} ASU with insert 5' CCCGGCCGGA 3' colored by pLDDT. **(D)** AF3 prediction of EnH-eGFP bound

to CC1^{+10} ASU with insert 5’ TGATGAGCAG 3’ colored by pLDDT. **(E)** AF3 prediction of Eve-HD bound to CC1^{+10} ASU with insert 5’ TGATGAGCAG 3’ colored by pLDDT. **(F)** AF3 prediction of UBX-HD bound to CC1^{+10} ASU with insert 5’ TGATGAGCAG 3’ colored by pLDDT.

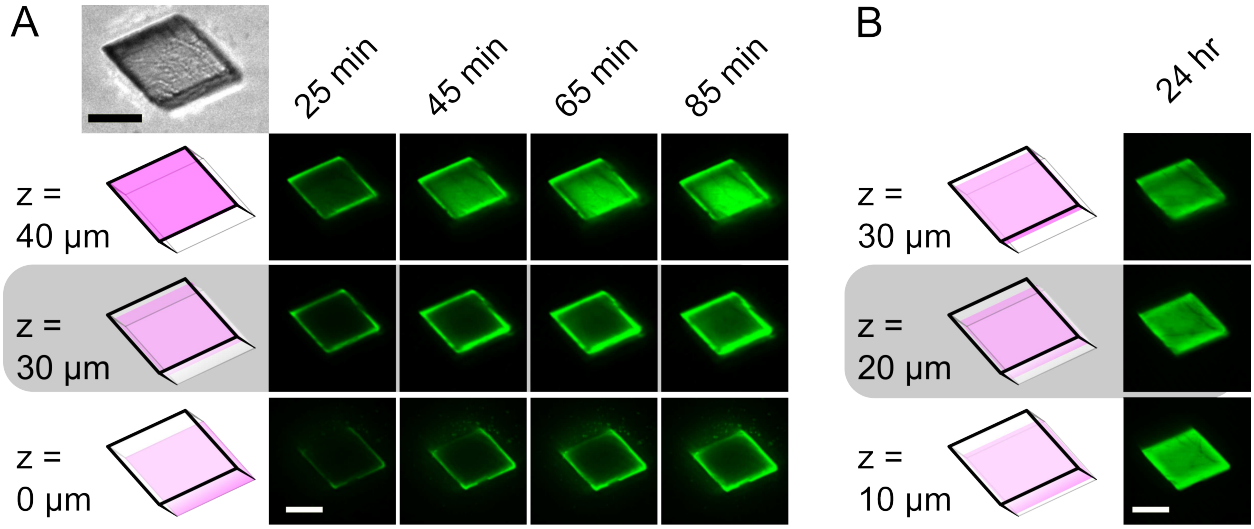


Fig. S4.2. Confocal microscopy of a CC1^{+10} crystal adsorbing UBX-HD

(A) Fluorescein-trace-labeled UBX-HD (23 µM) was diluted 1:200 with unlabeled UBX-HD and used to soak a CC1^{+10} crystal containing an asymmetric 11-nt insert 5'TAATTAGGCCG 3'. We then obtained a time-resolved, 1 µm z-stack. Planes (pink in the schematic) near the microscope slide (z = 0 µm), near the crystal top (z = 40 µm), or interior (z = 30 µm) were selected to show the anisotropic guest loading pattern. A differential interference contrast image (inset top left) is aligned to the z-plane schematics for comparison. **(B)** After 24 hr, the laser intensity was reduced from 20% to 12% to avoid over-saturation. Loading was comparable for 3 internal planes evenly spaced within the crystal interior. For confocal microscopy, dilute trace labeling was helpful to avoid extinction effect complications (Fig. S12-S14). All scale bars are 50 µm.

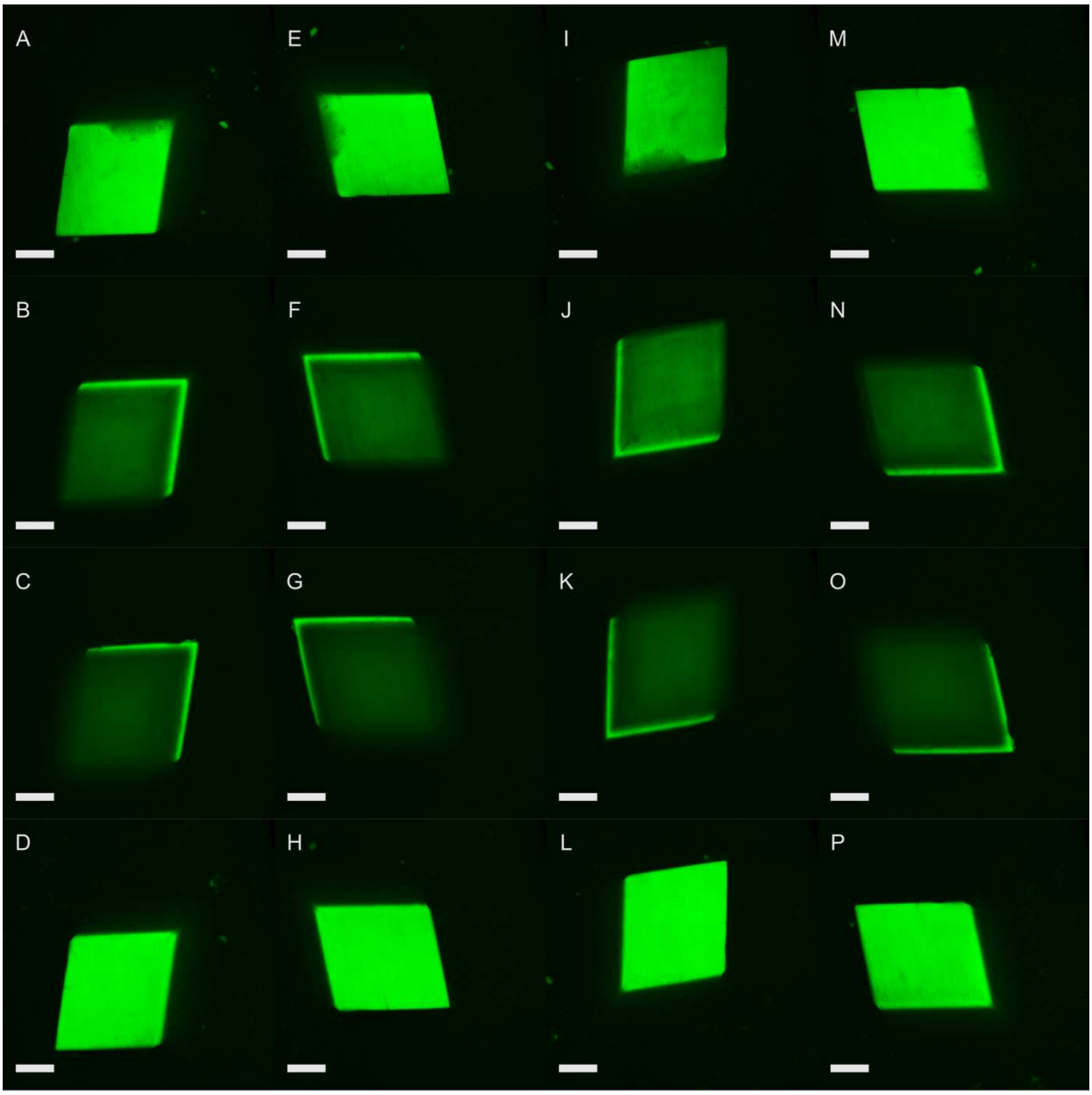
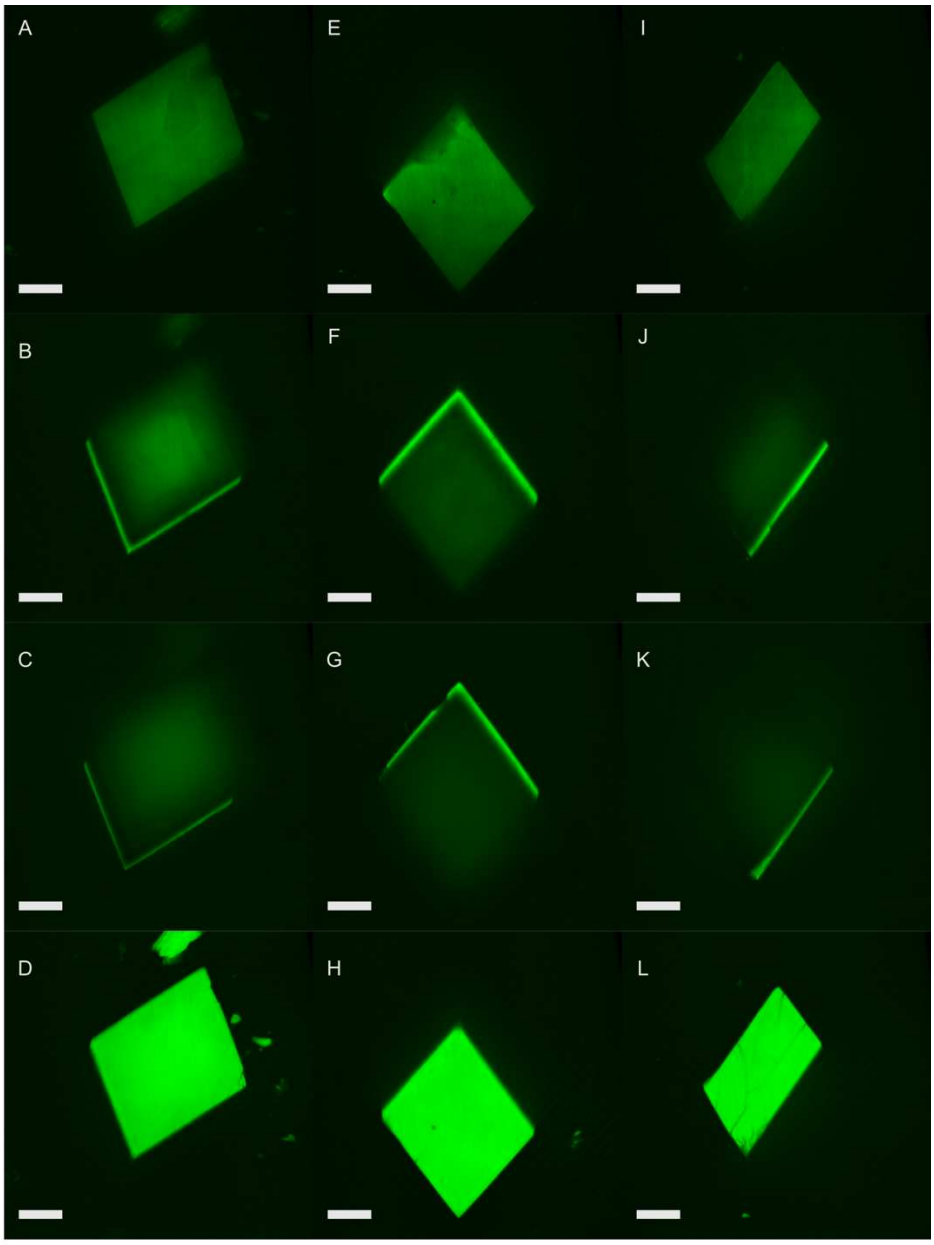


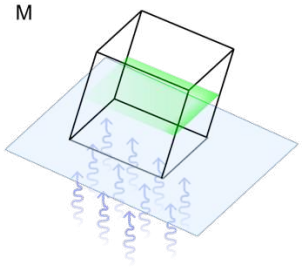
Fig. S4.3. CC1^{+10} Confocal Loading of UBX-HD Trace-labeled with NHS-Fluorescein

Our inverted confocal microscope (Nikon Eclipse Ti spinning-disk confocal microscope, Protocol S4.4) illuminates samples from below. Crystals with high internal fluorophore concentrations can exhibit extinction effects where emission intensity is reduced due to excitation absorption by intervening layers. (**A–D**) Optical sections of a CC1+10 crystal (identical DNA sequence to 9YZF.pdb) loaded with 80 µM UBX-HD trace-labeled with NHS-fluorescein. (**A**) Bottom optical section. (**B**) Mid-plane section showing reduced fluorescence

emission in the crystal interior, but strong emission from the two overhanging edges (lacking intervening crystal to absorb the excitation). (**C**) Z-slice (z=X) that is even farther from the microscope slide showing more pronounced extinction in the crystal interior. (**D**) Z-slice 10 µm above the bottom face corresponding to the brightest plane within the z-stack. (**E–H**), (**I–L**), and (**M-P**) The same four optical sections after rotating the crystal 90°, 180°, and 270° counter-clockwise, respectively. Scalebars are 50 microns.

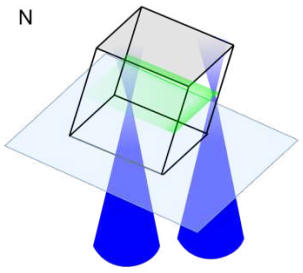


M



Out of plane confocal emission is suppressed, but not out of plane absorption

N



Consider two excitation beamlet from the rotating array at one moment in time, each sweeping out a diffraction limited spot.

Fig. S4.4. Orthogonal confocal views of the same $CC1^{+10}$ crystal loaded with UBX-HD trace-labeled with NHS-Fluorescein

To rule out heterogenous loading of guest protein we obtained Z-stack confocal images for multiple views of the same crystal. (**A–D**) Optical sections of a $CC1^{+10}$ crystal (same crystal as Fig.S17) loaded with 80 μM UBX-HD trace-labeled with NHS-fluorescein. (**A**) Bottom z-slice closest to the microscope slide. (**B**) Mid-plane (z=20μm) showing reduced fluorescence emission in the crystal interior, but strong emission from the two overhanging edges (lacking intervening crystal to absorb the excitation). (**C**) Z-slice (z=57.5μm) that is even farther from the slide showing more pronounced extinction in the crystal interior. (**D**) Z-slice 10 μm above the bottom face corresponding to the brightest plane within the z-stack. **(E–H**) and (**I–L**), the same four z-slices after flipping the crystal 180° and 90° respectively. (**M**) Schematic representation of parallelepiped crystal resting on a microscope slide, with illumination of a confocal plane (green) from below on an inverted confocal microscope. (**N**) Schematic representation of how the excitation light path length through crystalline material varies depending on the location, with overhanging edges experiencing minimal extinction.

These different views of the same crystal provide strong evidence that the striking observed fluorescence emission z-stack patterns were primarily due to variable excitation extinction. In panel A, a small defect notch in the crystal is visible just to the right of the top corner. The brightest plane in the crystal (panel D) was 10 microns above this defect. After the crystal was flipped 180°, this same small defect is now visible in panel G, just to the left of the top corner at a new height of 42.5 microns above the slide. In the new orientation, the notch is part of an overhang and is therefore illuminated despite most of the crystal being dim. In the second orientation, the bright uniform layer (panel H) was again 10 microns above the slide. This

layer was 47.5 microns above the slide in the original orientation (similar to panel C) and was therefore mostly dim other than the two overhanging edges. Thus, the bright zones in the crystal were dominated by proximity to unimpeded illumination rather than heterogeneous loading. As a final proof, the crystal was placed onto one of the smaller facets. Again, the bright z-plane was parallel to the slide and 10 microns above. In this orientation, only one facet had a slight overhang. No matter the crystal orientation, well-illuminated z-planes appeared to be uniformly loaded.

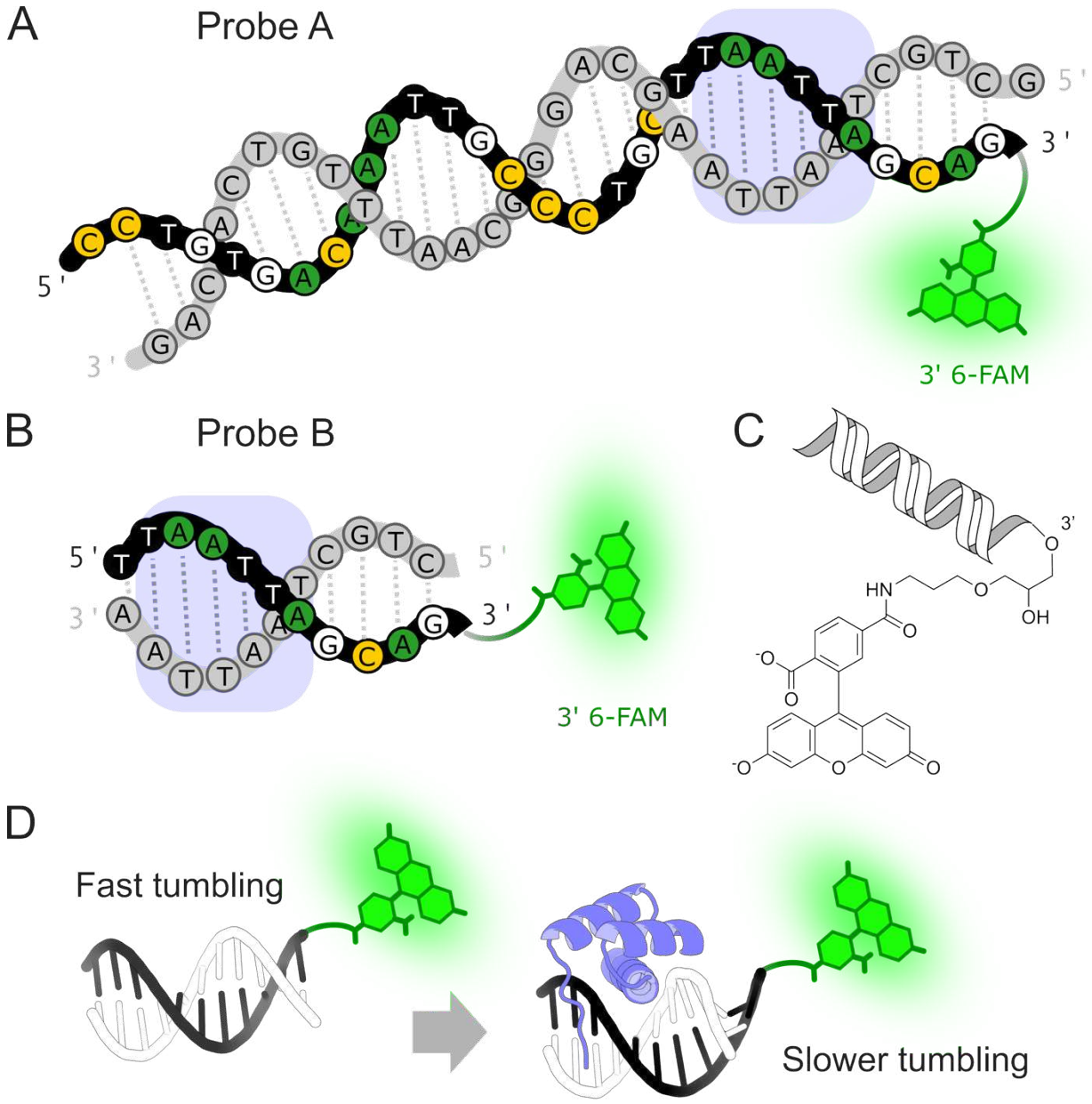


Fig. S4.5. Fluorescence Polarization Probe Schematic.

(**A**) A full-length fluorescence polarization probe A (also used for $CC1^{+10}$ crystallization trials) with a 3' 6-FAM conjugation (Integrated DNA Technologies). The location of a canonical sequence for homeodomain binding (5' TAATTA 3') is marked with a blue backdrop. (**B**) Truncated fluorescence polarization probe B. **(C)** Depiction of the fluorophore chemical structure and linker to the 3' OH of DNA **(D)** Schematic illustration of EVE-HD binding to

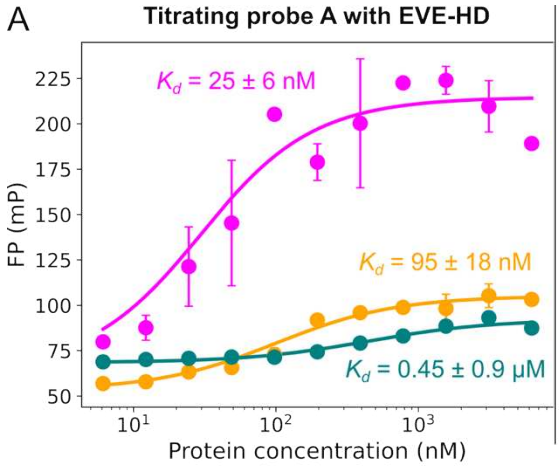
Probe B. Figure prepared in Inkscape, ChemDraw Professional (Version 25.5.0.6237, Revvity Signals Software, Inc.), and PyMOL.

The sequences below show the target homeodomain site (bold & underlined). Probe A also has the 5'-AATT-3' R1 site within the RepE54 binding footprint (bold).

Probe A:
5'-CCTGTGACA**AATT**GCCCTGCT**TAATTA**GCAG/36-FAM/-3'
 3'-GACACTGT**TTAA**CGGGACGA**ATTAAT**CGTCG-5'

Probe B:
5'-T**TAATTA**GCAG/36-FAM/-3'
3'-A**ATTAAT**CGTC-5'

A



50 mM KCl
10 mM Tris HCl pH 7.4

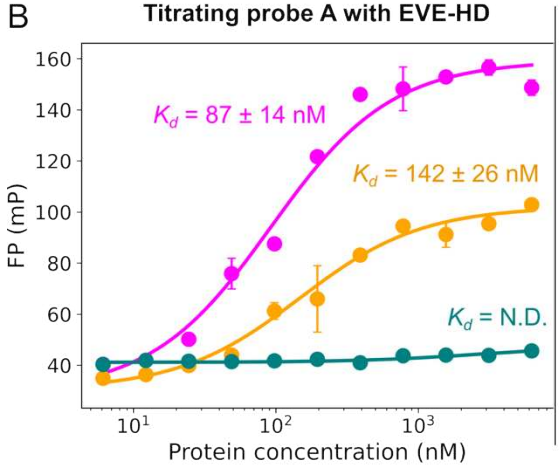
Purpose: Test Role of Divalent Salt

50 mM KCl, 4 mM $CaCl_2$
10 mM Tris HCl pH 7.4

1.6 M Li_2SO_4, 40 mM $Mg(CH_3COO)_2$
50 mM MES pH 6.5

Purpose: Verify EVE-HD binding inhibition by the high-salt CC1 mother liquor

B



100 mM NaCl, 5 mM $MgCl_2$
20 mM HEPES pH 7.4

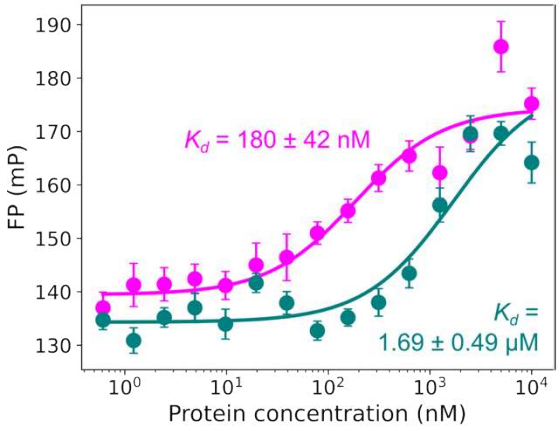
Purpose: Sodium Scanning

200 mM NaCl, 5 mM $MgCl_2$
20 mM HEPES pH 7.4

Purpose: Sodium Scanning

1 M NaCl, 5 mM $MgCl_2$
20 mM HEPES pH 7.4

C



10 mM KCl, 4 mM $CaCl_2$
10 mM Tris HCl pH 7.4
Purpose: Test Minimal Salt

100 mM NaCl, 5 mM $MgCl_2$
20 mM HEPES pH 7.4
Purpose: Compare Probes A and B with a common buffer.

D **Morrison Quadratic Binding Model:**

$$FP = B + (T - B) \times \frac{(K_d + P_t + L_t) - \sqrt{(K_d + P_t + L_t)^2 - 4P_tL_t}}{2L_t}$$

Fixed parameters:

FP: Anisotropy (mP)
P_t: Total Protein Concentration (nM)
L_t: Total Probe Concentration = 10 nM

Fit parameters:

B: Free Probe Signal (mP)
T: Bound Probe Signal (mP)
K_d: Dissociation Constant (nM)

Fig. S4.6. Fluorescence Polarization Data

We call these experiments “scouting” because the goal was to quickly find at least one salt regime consistent with DBD binding, rather than a systematic survey of guest binding as a function of solution conditions. Confidently ruling out (or building confidence in) alternate stoichiometry (e.g. 2:1 or 1:2 protein:DNA) would require less noisy experimental data, an effort counter to the scouting goal. Therefore, all fits used a 1:1 binding model. Each fit was suitable for semi-quantitative assessment of binding, with R^2 values above 0.9. (**A**) Fluorescence polarization fit of EVE-HD binding to the longer probe A (Figure S4.4.A) under three solvent conditions. Points are the mean of 2 replicates, and the error bars are the standard error of the mean (SEM) for better comparison with the data below. (**B**) Fluorescence polarization fit of EVE-HD binding to probe A (Fig. S9A) under three solvent conditions (each with 2 replicates) with varying sodium chloride concentrations. (**C**) Fluorescence polarization fit of a second homeodomain, ANTP-HD binding to probe B (Figure S4.4.B) under two solvent conditions (each with 8 replicates). (**D**) To account for depletion effects, we avoid the assumption that the total protein concentration equals the free protein concentration by using the quadratic (Morrison) equation. The Python scripts (https://github.com/cdsnow/FPfitting) fits the protein–DNA complex concentration at each titration point, incorporating the fixed final 10 nM concentration of fluorescent duplex to account for ligand depletion. The FP data can be found in the same repository. Nonlinear regression weighted by the inverse variance was used to simultaneously optimize K_d, the baseline polarization, and the bound-state polarization. For comparison, the scripts also perform simple Langmuir fitting. The values are quite close as shown in the tabulated results below and the conclusions drawn remain the same. Figure prepared with matplotlib and Inkscape.

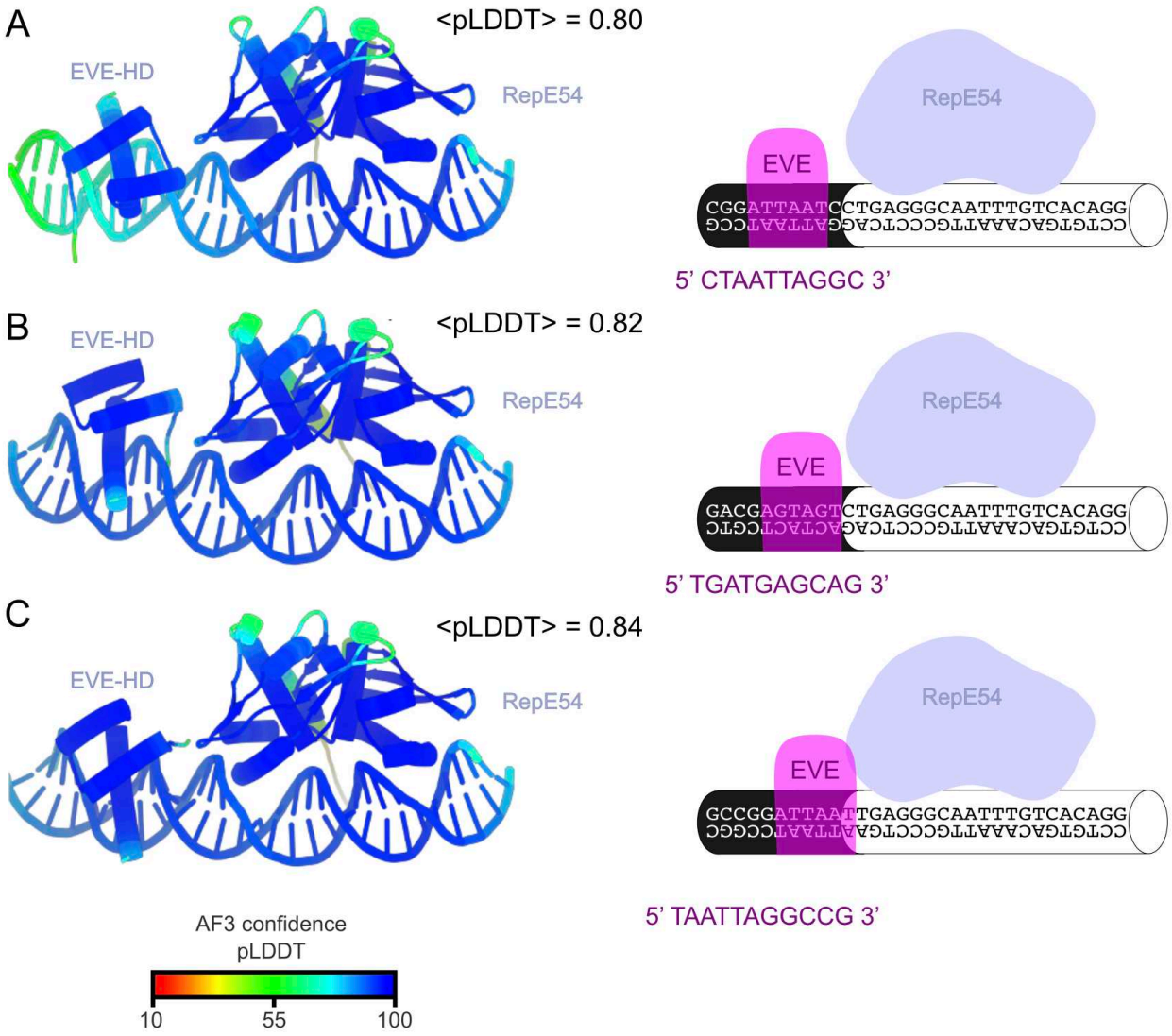
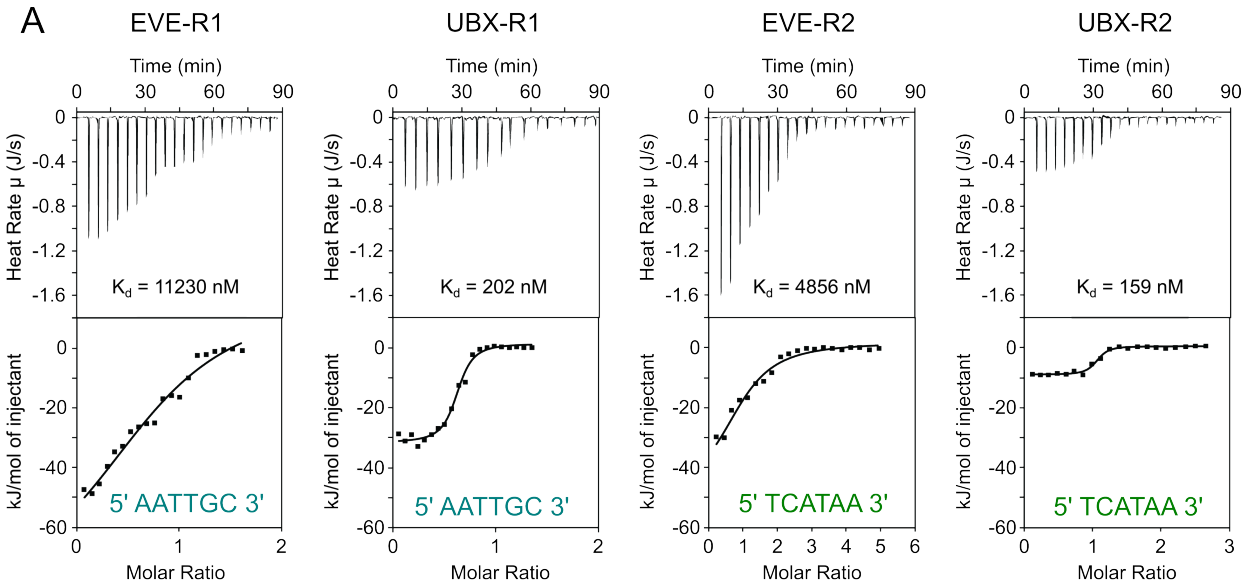
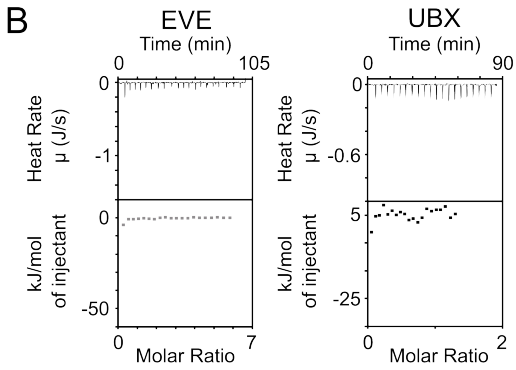


Fig. S4.7. Alphafold-3 structure prediction of Eve-HD bound to CC1^{+10} ASU with moving binding sites

All depicted models are AF3 predictions colored by pLDDT. We report the mean pLDDT for the complete model. (**A**) CC1^{+10} with insert 5' CTAATTAGGC 3' and EVE-HD. (**B**) CC1^{+10} with insert 5' TGATGAGCAG 3' and EVE-HD. (**C**) CC1^{+10} with insert 5' TAATTAGGCCG 3' and EVE-HD.



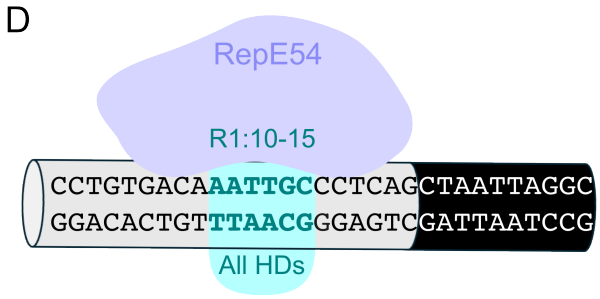


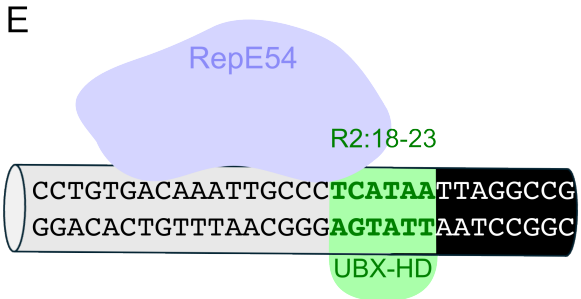
C

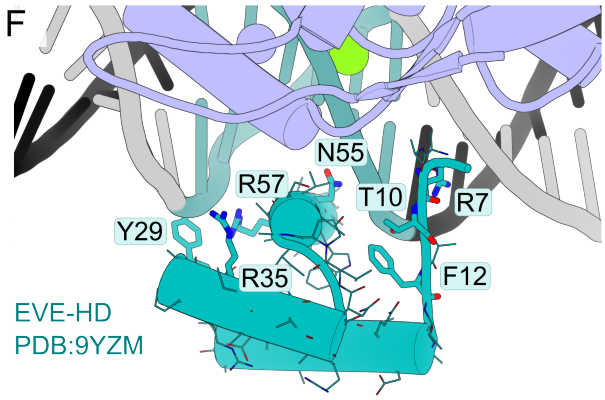
Thermodynamic Fit Results

Target Sequence	Protein	K_D	ΔH (kJ/mol)	-TΔS (kJ/mol)
R1: 5'-AATTGC-3'	EVE	8.6 ± 2.9 μM	-104 ± 21	75 ± 22
	UBX	189 ± 51 nM	-32.6 ± 1.1	-5.8 ± 1.6
R2: 5'-TCATAA-3'	EVE	6.5 ± 3.0 μM	-44 ± 10	15 ± 11
	UBX	147 ± 53 nM	-9.43 ± 0.31	-29.6 ± 1.1

5'-GCGCGCCGGCGGCCC-3'
Negative Control → No Binding







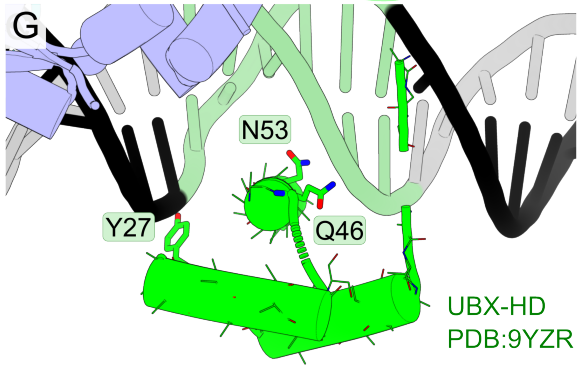


Fig. S4.8. ITC Measurements of Noncanonical Homeodomain Binding Sites and Structures Thereof

(**A)** Isothermal titration calorimetry (ITC) results of EVE-HD and UBX-HD binding to R1 (5' GCGCGCAATTGCGCCC 3') and R2 (5' GCGCGCTCATAAGCCC 3') sequences. **(B)** ITC results with a negative control duplex. **(C)** Fits for thermodynamic parameters were obtained with Python scripts (see Supplementary Methods and https://github.com/cdsnow/ITCfitting). (**D**) Schematic representation of R1 binding site along the full $CC1^{+10}$ DNA sequence from 9YZM.pdb and a PyMOL illustration of EVE-HD bound to the R1 site (teal). (**E**) Schematic representation of R2 binding site along the full $CC1^{+10}$ DNA sequence from 9YZR.pdb and a PyMOL illustration of UBX-HD bound to the R2 site (green). Figure prepared in Excel, Inkscape, and PyMOL. **(F-G)** Close views of the corresponding PDB models 9YZM and 9YZR.

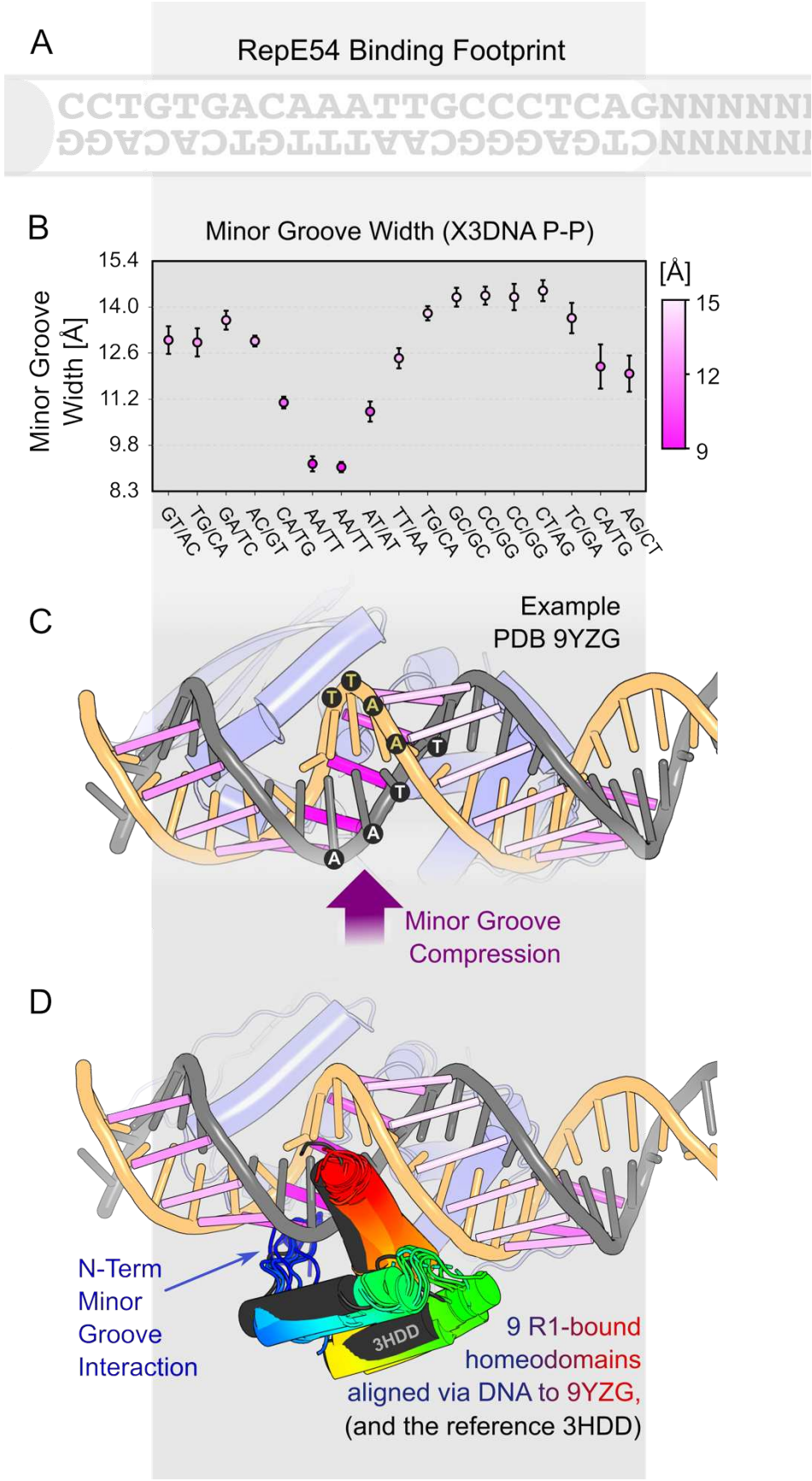

A
RepE54 Binding Footprint
CCTGTGACAAATTGCCCTCAGNNNNNN
GGACACTGTTTAACGGGAGTCNNNNNN
B
Minor Groove Width (X3DNA P-P)
Minor Groove Width [Å]
15.4
14.0
12.6
11.2
9.8
8.3
[Å]
15
12
9
GT/AC
TG/CA
GA/TC
AC/GT
CA/TG
AA/TT
AA/TT
AT/AT
TT/AA
TG/CA
GC/GC
CC/GG
CC/GG
CT/AG
TC/GA
CA/TG
AG/CT
C
Example
PDB 9YZG
Minor Groove
Compression
D
N-Term
Minor
Groove
Interaction
3HDD
9 R1-bound
homeodomains
aligned via DNA to 9YZG,
(and the reference 3HDD)

Fig. S4.9. Minor Groove Compression of $CC1^{+10}$ at Noncanonical Homeodomain Binding Site R1

(A) We typically conserved a 21 bp RepE54 binding window. **(B)** X3DNA (110) computed minor groove widths (mean and standard deviation) were tabulated for 17 $CC1^{+10}$ models that lack guest molecules: 9NB0, 9NCR, 9YZA, 9YZB, 9YZC, 9YZD, 9YDE, 9YDF, 9YDG, 9YDI, 9YZJ, 9Z08, 9Z1A, 9Z1B, 9Z1C, 9Z1D, 9Z55. **(C)** The X3DNA minor groove width is an average of two P-P distances flanking each base pair step. Using 9YZG as a representative example model, we depict each averaged vector using a variable width/color cylinder (increased magenta saturation and radius for short distances). The central 5'-AATT-3' sequence is labeled. As expected (113), our observed narrow minor groove coincides with our AT-rich region. **(D)** Nine additional $CC1^{+10}$ models have a homeodomain bound to the R1 register (9YZL, 9YZM, 9YZN, 9YZO, 9YZP, 9YZQ, 9YZR, 9YZS, 9YZT). To depict the position of those homeodomains relative to the compressed minor groove, we align each homeodomain-bound model (PyMOL align of the DNA chains) to the same 9YZG reference model and show only the R1: bound homeodomain. The C-terminal helices (red) bind the major groove opposite RepE54, and the varying homeodomain N-termini (blue) can insert an arginine into the compressed minor groove (113), a groove that is only partially obstructed by RepE54 (light blue). We also aligned the 5'-TAATTA-3' duplex in EnH-HD model 3HDD to the R1 site 5'-AATT-3' duplex and show the EnH-HD from 3HDD in black. Figure prepared in Inkscape, matplotlib, and PyMOL.

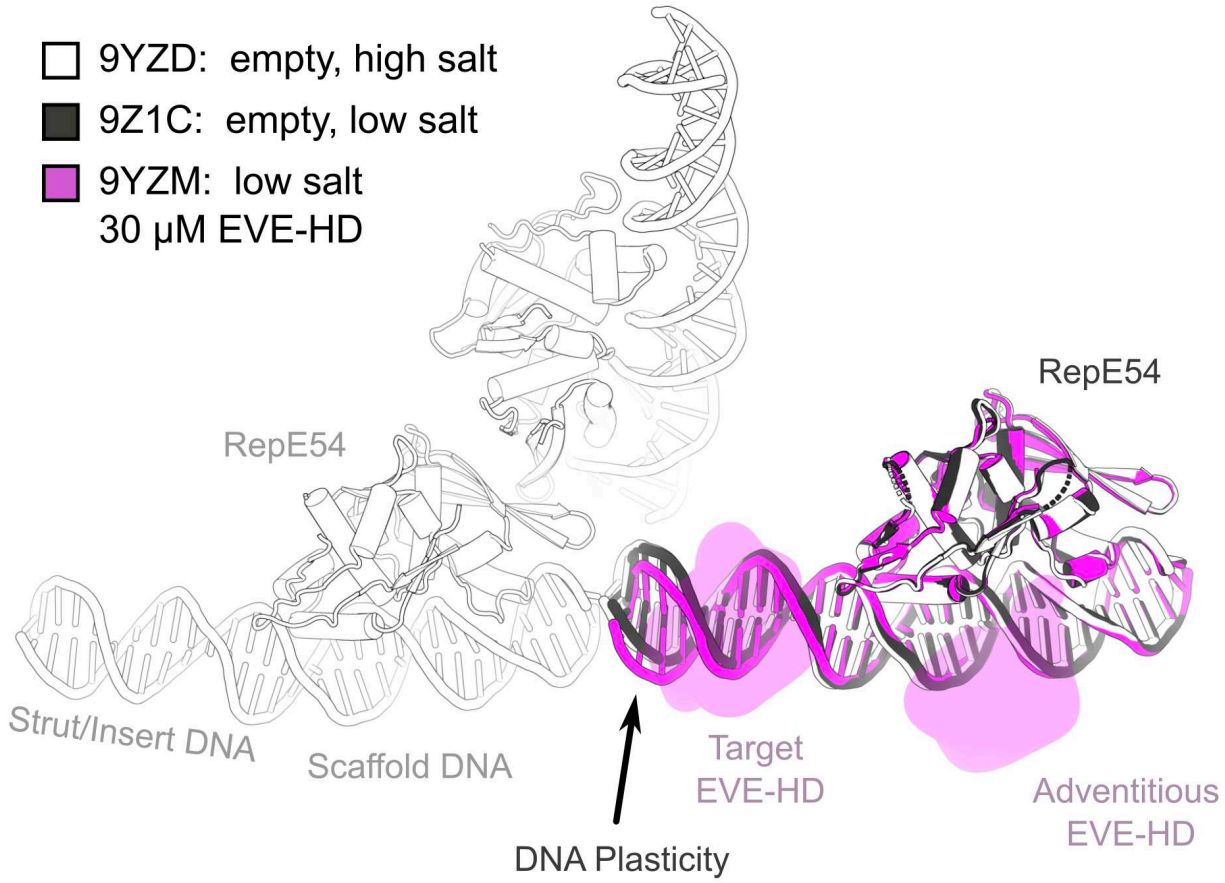


Fig. S4.10. DNA Cantilever comparison of $CC1^{+10}$ ASUs with identical DNA inserts

$CC1^{+10}$ building blocks with insert 5' CTAATTAGGC 3' and XRD collected under three different solvent conditions:

White) 30 mM Magnesium acetate, 1.3 M Lithium sulfate, 50 mM MES, pH 6.5 (PDB 9YZD).

Black) 50 mM KCl, 4 mM $CaCl_2$, 10% glycerol, 10 mM Tris HCl, pH 7.4 (PDB 9Z1C).

Magenta) 50 mM KCl, 4 mM $CaCl_2$, 10% glycerol, 10 mM Tris HCl pH 7.4 and loaded with 30 μM EVE-HD (PDB 9YZM).

Using PyMOL align, the models from 9Z1C (black) and 9YZM (magenta) were aligned via their RepE54 domain onto 9YZD (white). All biomolecules are shown as cartoons other than the guest EVE-HD domains from 9YZM which were depicted (for visual clarity) as partially transparent

smoothed gaussian surfaces (gaussian_resolution = 4). Figure prepared in PyMOL (82) and Inkscape.

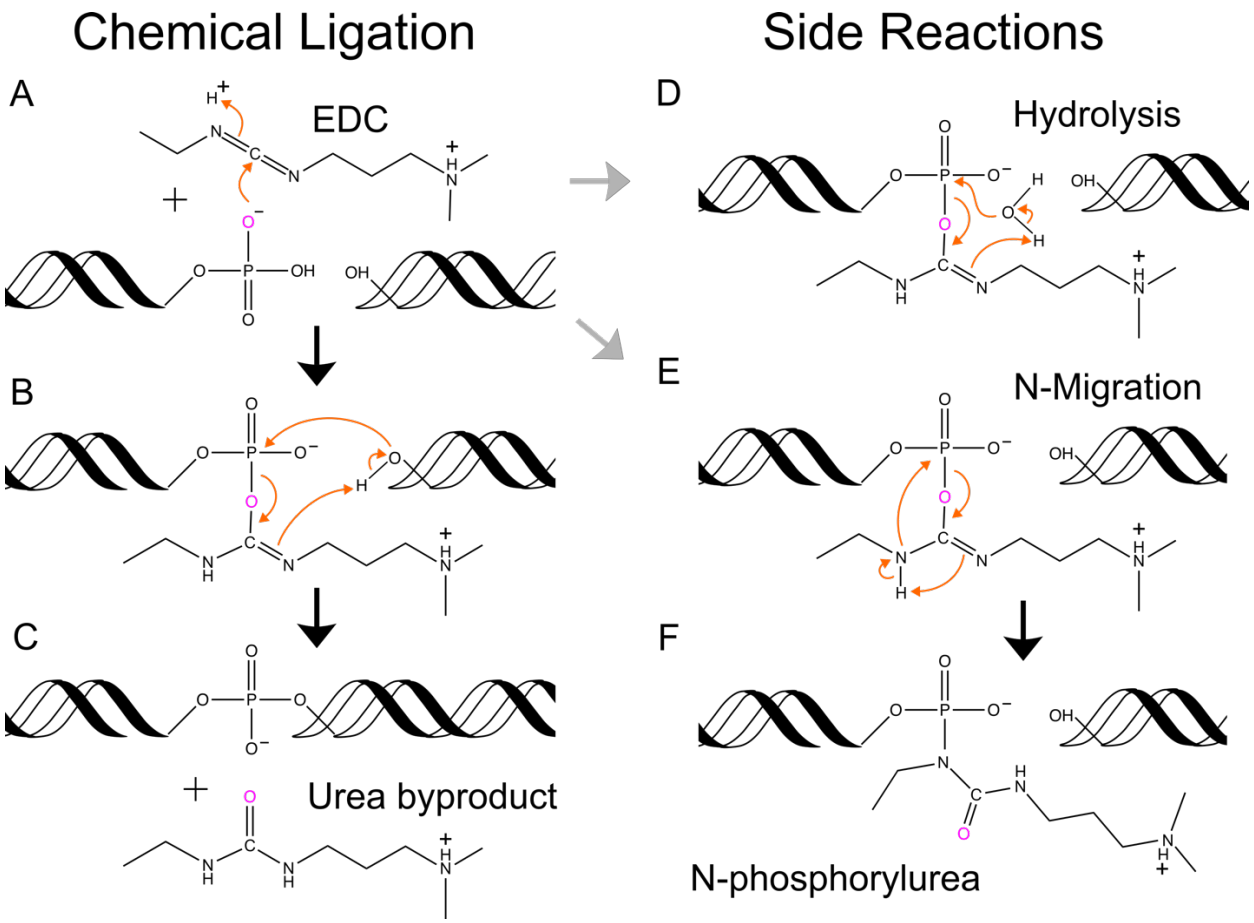


Fig. S4.11. EDC crosslinking mechanism

As demonstrated in prior publications, we use EDC for the chemical ligation of coaxial DNA interfaces within the crystals. Here we elaborate on the proposed chemical ligation mechanism and depict potential side reactions. **(A–C)** Productive Ligation. **(A)** A terminal phosphate (e.g., 3′-P) is activated by EDC to form a reactive O-phosphorylisourea intermediate. **(B)** The terminal hydroxyl (e.g., 5′-OH) of the neighboring strand attacks the phosphorus. **(C)** Formation of the native phosphodiester bond and release of the stable "urea" byproduct. Our prior work showed strong crystal stabilization despite incomplete ligation yields.

Here we provide schematics for reactions that could lead to incomplete ligation. These include: **(D)** Hydrolysis. Water competes as a nucleophile, attacking the activated phosphorus. This regenerates the starting phosphate and consumes EDC but does not cap the DNA. **(E–F)** N-

Migration (Dead End). **(E)** The activated intermediate might undergo an intramolecular rearrangement where an isourea nitrogen attacks the phosphorus (117). **(F)** This would result in a stable, inert N-phosphorylurea adduct, which would cap the DNA and prevent productive ligation. Figure prepared with ChemDraw Professional (Version 25.5.0.6237, Revvity Signals Software, Inc.).

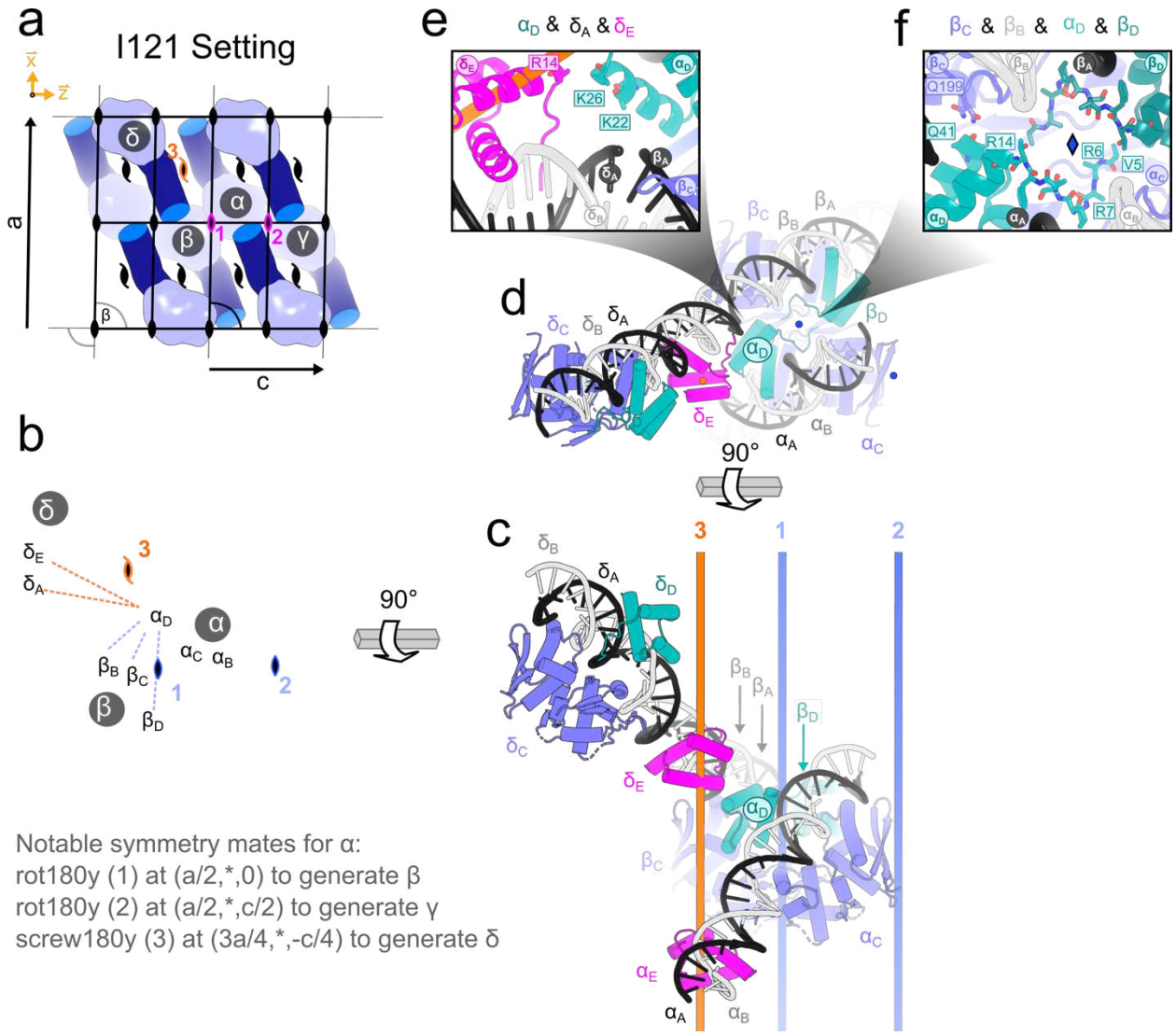


Fig. S4.12. Notable symmetry neighbors for the CC1 lattice: PDB entry 9YZM

(a) For one building block, **α**, there are three neighbors that interact with the guest homeodomains. Copies **β** and **γ** are related to **α** via 2-fold symmetry axes (#1 and #2, magenta). In contrast, copy **δ** is related to **α** via a 2-fold screw axis (#3, orange). In this orientation, the major pores run horizontally (along the unit cell vector c). **(b)** A simplified diagram shows dashed lines between the constituent chains (A,B,C,D,E) within copy **α** and its neighbors, switching the color of the 2-fold axes to blue. **(c)** After a 90-degree rotation, each symmetry copy is depicted with two DNA chains (A light gray, and B, black), one RepE54 protein (blue, chain

C), a homeodomain at the unexpected binding site opposite RepE54 (teal, chain D), and a homeodomain at the intended site (magenta, chain E). In this orientation, the 2-fold axes are depicted as blue cylinders and the 2-fold screw axis is depicted as an orange cylinder. **(d)** After another 90-degree rotation the axes are once more perpendicular to the page. **(e)** This close-up view shows interactions of the homeodomain $\boldsymbol{\alpha_D}$. Specifically, the $\boldsymbol{\alpha_D}$ and $\boldsymbol{\delta_E}$ guest-guest interaction has a minimum distance of 3.8Å. Arg14 in $\boldsymbol{\delta_E}$ is projecting towards $\boldsymbol{\alpha_D}$. Arg14 is not resolved but could potentially find a favorable interaction with the exposed carbonyls (helix dipole) of residues 25 and 26. The $\boldsymbol{\delta_E}$ helix dipole (residues 14-26) is also oriented favorably with respect to the $\boldsymbol{\alpha_D}$ helix dipole (residues 14-26). With a beta-carbon to phosphate oxygen distances of 7.6-7.7Å, the unresolved Lys22 in $\boldsymbol{\alpha_D}$ is likely interacting favorably with the DNA in $\boldsymbol{\delta_A}$ and $\boldsymbol{\beta_A}$. **(f)** This close-up view shows other interactions of the homeodomain $\boldsymbol{\alpha_D}$ which is very close (3.0Å) to its symmetry neighbor across the 2-fold axis (blue diamond). The nature of the interaction is not very clear and residues prior to Val5 are not resolved. The N-terminal tail of $\boldsymbol{\alpha_D}$ is also likely interacting with its own scaffold RepE54 $\boldsymbol{\alpha_C}$ (distance 4.8Å), via steric occlusion and interactions with disordered Arg7 and Tyr8. The $\boldsymbol{\alpha_D}$ helix dipole (residues 14-26) is also close to (5.3Å) and oriented favorably (positive N-terminal end) with respect to the negative DNA backbone ($\boldsymbol{\beta_B}$). The unresolved Arg14 could also interact favorably with the $\boldsymbol{\beta_B}$ DNA. Finally, there is very direct interaction (2.8Å) between the C-terminal end of the $\boldsymbol{\alpha_D}$ alpha helix 32-41 with multi-conformation Gln199 on RepE54 ($\boldsymbol{\beta_C}$).

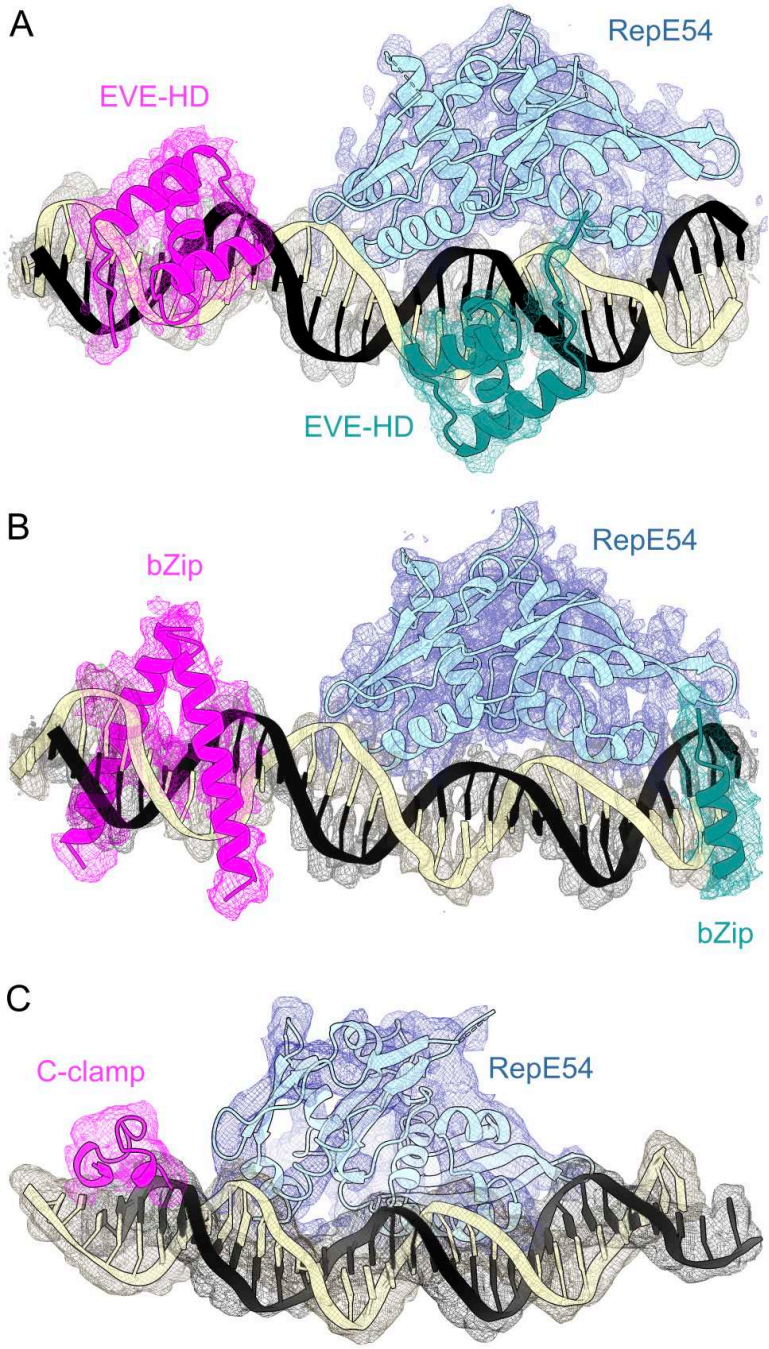


Fig. S4.13. Representative PHENIX omit maps of each guest structure Motif

Each panel is a PHENIX simulated annealing omit map (46). **(A)** 9YZM.pdb with omitted EVE-HD shown in magenta & teal. **(B)** 9YZU.pdb with omitted bZip shown in magenta & teal. **(C)** 9Z44.pdb with omitted C-clamp shown in magenta. All meshes individually contoured at 1.0 using ChimeraX (sdLevel 1.0). Omit maps of this type are the standard method to verify that

electron density map features are not the result of model bias (45–47). Omit maps confirmed that the guest molecule density was well supported by the data. Figure prepared with scripted ChimeraX and Inkscape.

Notably, “Bonus density” was not unique to the homeodomain family. For the bZip guest, which interacts exclusively with DNA major grooves, we likewise observed additional electron density at an unanticipated register (with no notable overlap with the canonical bZip recognition sequence), although in this case the canonical target site was better resolved (Fig. 3E). The partial density was consistent with a bZip helix segment fragment but lacked resolvable side chains and was not obviously stabilized by neighboring lattice contacts. We hypothesize that this site is a low-affinity complex that was only occupied due to mass action (during loading the initial bZip concentration outside the crystal was 33 μM in a 200 μL drop).

PDB Code	XRD condition
9Z1B	50mM KCl, 4mM CaCl2, 10% glycerol, 10mM Tris HCl pH7.4
9Z1C	50mM KCl, 4mM CaCl2, 10% glycerol, 10mM Tris HCl pH7.4
9Z1D	10mM HEPES, 50mM NaCl, 2.5mM MgCl2, pH 6.5, 800mM Lithium sulfate, 15mM Magnesium acetate, 25mM MES, pH 6.5, 80uM UBX

Table S4.1. $CC1^{+10}$ miscellaneous solvent XRD collection statistics

PDB code	9Z1B	9Z1C
Data collection		
Light source	Synchrotron	Synchrotron
Wavelength (Å)	1	1
Resolution Range (Å)	42.58-3.88 (4.06-3.88)	45.27-3.04 (3.11-3.04)
Space group	I121	I121
Unit cell dimensions		
a, b, c (Å)	72.98, 126.16, 135.73	72.72, 129.49, 131.28
α, β, ɣ (°)	90.00, 91.06, 90.00	90.00, 90.95, 90.00
Unique reflections	22544 (4078)	23207 (4199)
Multiplicity	3.5 (3.6)	3.5 (3.6)
Completeness (%)	98.8 (99.0)	99.0 (99.4)
Mean I/sigma(I)	7.8 (1.6)	10.3 (1.3)
Wilson B-factor	121.72	92.33
R-merge	0.118 (0.747)	0.085 (1.153)
R-meas	0.139 (0.876)	0.100 (1.355)
R-pim	0.073 (0.455)	0.053 (0.706)
CC1/2	0.998 (0.878)	1.0 (0.833)
Refinement		
Reflections used in refinement	10987 (1257)	23075 (1310)
Reflections used for Rfree	1111 (146)	2298 (142)
R-work	0.2743 (0.3454)	0.3000 (0.5731)
R-free	0.3081 (0.3631)	0.3251 (0.5991)
Number of non-hydrogen atoms	2905	2916
Macromolecules	2905	2916
Ligands	11	4
Solvent	0	0
Protein residues	216	216
RMS (bonds) (Å)	0.18	0.15
RMS (angles) (°)	0.4	0.38
Ramachandran favored (%)	97	98
Ramachandran allowed (%)	2	2
Ramachandran outliers (%)	1	0
Rotamer outliers (%)	0	0
Clashscore	6	5
Average B-factor	200	122

Table S4.2. X-ray diffraction statistics for $CC1^{+10}$ PDB entries 9Z1B and 9Z1C

PDB code	9Z1D
Data collection	
Light source	Synchrotron
Wavelength (Å)	1

Resolution Range (Å)	39.72-3.06 (3.13-3.06)
Space group	I121
Unit cell dimensions	
a, b, c (Å)	72.38, 124.54, 136.54
α, β, ɣ (°)	90.00, 90.17, 90.00
Unique reflections	22756 (4075)
Multiplicity	3.6 (3.6)
Completeness (%)	99.4 (98.9)
Mean I/sigma(I)	9.1 (0.9)
Wilson B-factor	113.12
R-merge	0.129 (1.300)
R-meas	0.153 (1.523)
R-pim	0.081 (0.789)
CC1/2	0.724 (0.869)
Refinement	
Reflections used in refinement	22199 (1001)
Reflections used for Rfree	2227 (112)
R-work	0.2929 (0.7702)
R-free	0.3184 (0.9681)
Number of non-hydrogen atoms	2848
Macromolecules	2848
Ligands	1
Solvent	0
Protein residues	216
RMS (bonds) (Å)	0.17
RMS (angles) (°)	0.43
Ramachandran favored (%)	98
Ramachandran allowed (%)	2
Ramachandran outliers (%)	0
Rotamer outliers (%)	0
Clashscore	4
Average B-factor	153

Table S4.3. X-ray diffraction statistics for CC1^{+10} PDB entry 9Z1D

PDB Code:	Guest Protein	Loading Condition
9YZL	Eve-HD	15mM KCl, 4mM CaCl2, 10% glycerol, 10mM Tris HCl pH 7.4, 30μM Eve-HD
9YZM	Eve-HD	50mM KCl, 4mM CaCl2, 10% glycerol, 10mM Tris HCl pH 7.4, 30μM Eve-HD
9YZN	Eve-HD	25mM KCl, 2mM CaCl2, 10% glycerol, 10mM Tris HCl pH 7.4, 30μM Eve-HD
9YZO	Eve-HD	50mM KCl, 4mM CaCl2, 10% glycerol, 10mM Tris HCl pH 7.4, 30μM Eve-HD
9YZP	UBX-HD	50mM KCl, 4mM CaCl2, 10% glycerol, 10mM Tris HCl pH 7.4, 40μM UBX-HD
9YZQ	EnH-eGFP	50mM KCl, 4mM CaCl2, 10% glycerol, 10mM Tris HCl pH 7.4, 22μM EnH-eGFP
9YZR	UBX-HD	35mM KCl, 4mM CaCl2, 10% glycerol, 10mM Tris HCl pH 7.4, 80μM UBX-HD
9YZS	Eve-HD	50mM KCl, 4mM CaCl2, 10% glycerol, 10mM Tris HCl pH 7.4, 30μM Eve-HD
9YZT	AntP-HD	50mM KCl, 4mM CaCl2, 10% glycerol, 10mM Tris HCl pH 7.4, 23μM AntP-HD
9Z44	C-clamp	50mM KCl, 4mM CaCl2, 10% glycerol, 10mM Tris HCl pH 7.4, 90μM C-clamp
9YZU	bZip	50mM KCl, 4mM CaCl2, 10% glycerol, 10mM Tris HCl pH 7.4, 33μM bZip
9Z4E	bZip	50mM KCl, 4mM CaCl2, 10% glycerol, 10mM Tris HCl pH 7.4, 100μM bZip

Table S4.4. $CC1^{+10}$ guest loading conditions

Structure	Binding Sites	Comparison	RMSD
9YZL	(R1:10-14, AATTGC), (R5:23-28, TAATTA)	1JGG	R1=0.775, R5=N/A
9YZM	(R1:10-14, AATTGC), (R5:23-28, TAATTA)	1JGG	R1=0.594, R5=0.697
9YZN	(R1:10-14, AATTGC), (R4:22-27, TGATGA)	1JGG	R1=0.627, R4=0.547
9YZO	(R1:10-14, AATTGC), (R4:22-27, TGATGA)	1JGG	R1=0.643, R4=0.587
9YZS	(R1:10-14, AATTGC), (R3:21-26, TAATTA)	1JGG	R1=0.545, R3=0.462
9YZP	(R1:10-14, AATTGC), (R4:22-27, TGATGA)	1B8I	R1=0.524, R4=N/A
9YZR	(R1:10-14, AATTGC), (R2:18-23, TCATAA)	1B8I	R1=0.936, R2=0.632
9YZQ	(R1:10-14, AATTGC), (R4:22-27, TGATGA)	3HDD	R1=0.515, R4=N/A
9YZT	(R1:10-14, AATTGC), (R2:21-26, TAATTA)	4XID	R1=0.439, R2=0.389
9Z44	(R1:1-5, GCCGG)	7DTA	R1 = 0.582
9YZU	(R1:22-30, ATGAGTCA)	AlphaFold-3	R1 = 0.456
9Z4E	(R1:23-31, ATGAGTCA)	AlphaFold-3	R1 = 0.504

Table S4.5. Comparison of guest structures to existing PDB structures

RMSD values were calculated in PyMOL using the *align* command by superimposing the guest-protein chain from each reference structure onto the corresponding guest-protein chain in the experimental $CC1^{+10}$ structures.

PDB code	9YZL	9Z4E
Data collection		
Light source	Synchrotron	Synchrotron
Wavelength (Å)	1	1
Resolution Range (Å)	48.31-3.02 (3.19-3.02)	65.07-3.63 (3.78-3.63)
Space group	I 1 2 1	C 1 2 1
Unit cell dimensions		
a, b, c (Å)	73.36, 120.29, 134.40	159.04, 115.33, 72.72
α, β, ɣ (°)	90.00, 91.47, 90.00	90.00, 116.50, 90.00
Unique reflections	22518 (3649)	13262 (3166)
Multiplicity	3.5 (3.6)	7.1 (7.2)
Completeness (%)	98.3 (98.5)	99.2 (99.6)
Mean I/sigma(I)	3.8 (0.9)	10 (1.9)
Wilson B-factor	86.65	149.49
R-merge	0.198 (1.428)	0.104 (0.886)
R-meas	0.234 (1.675)	0.113 (0.955)
R-pim	0.124 (0.868)	0.042 (0.352)
CC1/2	0.966 (0.622)	0.999 (0.902)
Refinement		
Reflections used in refinement	22417 (1243)	13202 (1303)
Reflections used for Rfree	2223 (137)	1319 (153)
R-work	0.2972 (0.4570)	0.2473 (0.3725)
R-free	0.3056 (0.4581)	0.2853 (0.4225)
Number of non-hydrogen atoms	3490	3172
Macromolecules	3490	3172
Ligands	2	3
Solvent	0	0
Protein residues	280	266
RMS (bonds) (Å)	0.18	0.20
RMS (angles) (°)	0.44	0.41
Ramachandran favored (%)	97	98
Ramachandran allowed (%)	3	2
Ramachandran outliers (%)	0	0
Rotamer outliers (%)	0	0
Clashscore	6	9
Average B-factor	139	214

Table S4.6. X-ray diffraction statistics for $CC1^{+10}$ PDB entries 9YZL and 9Z4E

PDB code	9ZYM	9ZYN
Data collection		
Light source	Synchrotron	Synchrotron
Wavelength (Å)	1	1
Resolution Range (Å)	44.48-3.07 (3.14-3.07)	46.80-3.16 (3.23-3.16)
Space group	I 1 2 1	I 1 2 1
Unit cell dimensions		
a, b, c (Å)	73.78, 125.19, 131.23	76.03, 124.86, 139.09
α, β, ɣ (°)	90.00, 92.26, 90.00	90.00, 91.85, 90.00
Unique reflections	22050 (4010)	23358 (4253)
Multiplicity	3.5 (3.7)	3.5 (3.6)
Completeness (%)	98.9 (99.3)	98.7 (99.4)
Mean I/sigma(I)	7.6 (1.3)	10.5 (1.9)
Wilson B-factor	87.54	93.44
R-merge	0.123 (1.054)	0.077 (0.819)
R-meas	0.145 (1.237)	0.092 (0.962)
R-pim	0.077 (0.643)	0.049 (0.501)
CC1/2	0.998 (0.741)	0.999 (0.863)
Refinement		
Reflections used in refinement	21973 (1229)	21733 (1189)
Reflections used for Rfree	2194 (131)	2166 (136)
R-work	0.2594 (0.4232)	0.2618 (0.3895)
R-free	0.2691 (0.4468)	0.2827 (0.4011)
Number of non-hydrogen atoms	3733	3699
Macromolecules	3733	3699
Ligands	2	1
Solvent	0	0
Protein residues	325	329
RMS (bonds) (Å)	0.16	0.14
RMS (angles) (°)	0.37	0.35
Ramachandran favored (%)	96	97
Ramachandran allowed (%)	3	2
Ramachandran outliers (%)	1	1
Rotamer outliers (%)	0	0
Clashscore	4	3
Average B-factor	139	117

Table S4.7. X-ray diffraction statistics for CC1^{+10} PDB entries 9YZM and 9YZN

PDB code	9YZO	9YZP
Data collection		
Light source	Synchrotron	Synchrotron
Wavelength (Å)	1	1
Resolution Range (Å)	46.68-2.96 (3.02-2.96)	47.16-3.67 (3.80-3.67)
Space group	I 1 2 1	I 1 2 1
Unit cell dimensions		
a, b, c (Å)	74.11, 118.26, 141.68	73.92, 126.61, 137.89
α, β, ɣ (°)	90.00, 90.89, 90.00	90.00, 90.92, 90.00
Unique reflections	24985 (4020)	13732 (3297)
Multiplicity	3.5 (3.6)	3.3 (3.4)
Completeness (%)	97.8 (98.2)	98.4 (99.3)
Mean I/sigma(I)	7.4 (0.9)	2.8 (0.7)
Wilson B-factor	90.27	157.05
R-merge	0.098 (1.102)	0.202 (1.619)
R-meas	0.116 (1.293)	0.243 (1.923)
R-pim	0.062 (0.671)	0.132 (1.026)
CC1/2	0.997 (0.663)	0.997 (0.368)
Refinement		
Reflections used in refinement	24816 (1216)	13695 (1228)
Reflections used for Rfree	2467 (140)	1356 (127)
R-work	0.2898 (0.4673)	0.2720 (0.4252)
R-free	0.3121 (0.5115)	0.2949 (0.4167)
Number of non-hydrogen atoms	3650	3177
Macromolecules	3650	3177
Ligands	1	1
Solvent	0	0
Protein residues	329	272
RMS (bonds) (Å)	0.22	0.45
RMS (angles) (°)	0.50	0.69
Ramachandran favored (%)	96	98
Ramachandran allowed (%)	3	2
Ramachandran outliers (%)	1	0
Rotamer outliers (%)	1	0
Clashscore	10	10
Average B-factor	104	226

Table S4.8. X-ray diffraction statistics for $CC1^{+10}$ PDB entries 9YZO and 9YZP

PDB code	9YZQ	9YZR
Data collection		
Light source	Synchrotron	Synchrotron
Wavelength (Å)	1	1
Resolution Range (Å)	46.92-3.75 (3.90-3.75)	45.29-3.39 (3.48-3.39)
Space group	I 1 2 1	I 1 2 1
Unit cell dimensions		
a, b, c (Å)	74.12, 122.40, 140.44	76.03, 124.86, 139.09
α, β, ɣ (°)	90.00, 90.38, 90.00	90.00, 91.85, 90.00
Unique reflections	15454 (3568)	21916 (3999)
Multiplicity	3.4 (3.4)	3.5 (3.6)
Completeness (%)	97.5 (95.1)	98.8 (99.3)
Mean I/sigma(I)	4.1 (0.4)	6.5 (1.1)
Wilson B-factor	135.03	92.55
R-merge	0.159 (4.541)	0.134 (1.332)
R-meas	0.190 (5.424)	0.159 (1.566)
R-pim	0.102 (2.925)	0.085 (0.816)
CC1/2	0.998 (0.436)	0.998 (0.658)
Refinement		
Reflections used in refinement	12659 (1231)	17764 (1200)
Reflections used for Rfree	1263 (144)	1767 (127)
R-work	0.2951 (0.4783)	0.2652 (0.4375)
R-free	0.3382 (0.4722)	0.2976 (0.4531)
Number of non-hydrogen atoms	3198	3409
Macromolecules	3198	3409
Ligands	2	1
Solvent	0	0
Protein residues	272	298
RMS (bonds) (Å)	0.38	0.14
RMS (angles) (°)	0.63	0.37
Ramachandran favored (%)	97	97
Ramachandran allowed (%)	2	3
Ramachandran outliers (%)	1	0
Rotamer outliers (%)	0	0
Clashscore	7	5
Average B-factor	221	146

Table S4.9. X-ray diffraction statistics for $CC1^{+10}$ PDB entries 9YZQ and 9YZR

PDB code	9YZS	9YZT
Data collection		
Light source	Synchrotron	Synchrotron
Wavelength (Å)	1	1
Resolution Range (Å)	46.92-3.08 (3.15-3.08)	46.13-3.30 (3.39-3.30)
Space group	I 1 2 1	I 1 2 1
Unit cell dimensions		
a, b, c (Å)	74.19, 122.96, 135.18	72.75, 126.04, 134.95
α, β, ɣ (°)	90.00, 91.86, 90.00	90.00, 90.24, 90.00
Unique reflections	22226 (4000)	22292 (4017)
Multiplicity	3.5 (3.6)	3.5 (3.6)
Completeness (%)	98.9 (98.9)	98.8 (98.5)
Mean I/sigma(I)	6.7 (1.3)	6.1 (1.6)
Wilson B-factor	92.55	118.28
R-merge	0.099 (1.101)	0.186 (1.488)
R-meas	0.143 (1.293)	0.220 (1.754)
R-pim	0.076 (0.673)	0.115 (0.919)
CC1/2	0.998 (0.777)	0.864 (0.730)
Refinement		
Reflections used in refinement	22106 (1250)	17991 (1183)
Reflections used for Rfree	1250 (142)	1798 (144)
R-work	0.2728 (0.4119)	0.2827 (0.4684)
R-free	0.2906 (0.4524)	0.2921 (0.4972)
Number of non-hydrogen atoms	3815	3624
Macromolecules	3815	3624
Ligands	2	1
Solvent	0	0
Protein residues	329	315
RMS (bonds) (Å)	0.21	0.14
RMS (angles) (°)	0.41	0.37
Ramachandran favored (%)	95	97
Ramachandran allowed (%)	4	3
Ramachandran outliers (%)	1	0
Rotamer outliers (%)	0	0
Clashscore	6	6
Average B-factor	161	149

Table S4.10. X-ray diffraction statistics for $CC1^{+10}$ PDB entries 9YZS and 9YZT

PDB code	9YZU	9Z44
Data collection		
Light source	Synchrotron	Synchrotron
Wavelength (Å)	1	1
Resolution Range (Å)	63.71-3.05 (3.12-3.05)	47.99-7.20
Space group	C 1 2 1	I 1 2 1
Unit cell dimensions		
a, b, c (Å)	152.28, 127.19, 72.98	73.49, 127.65, 141.18
α, β, ɣ (°)	90.00, 117.60, 90.00	90.00, 91.98, 90.00
Unique reflections	23188 (4136)	1895
Multiplicity	7.0 (6.6)	1.0
Completeness (%)	98.6 (97.6)	99.14
Mean I/sigma(I)	11.3 (2.0)	2.12
Wilson B-factor	85.24	312.80
R-merge	0.130 (1.060)	0.688
R-meas	0.140 (1.151)	0.818
R-pim	0.053 (0.443)	0.437
CC1/2	0.999 (0.891)	0.301
Refinement		
Reflections used in refinement	23071 (1353)	1895
Reflections used for Rfree	2235 (118)	192
R-work	0.2247 (0.3769)	0.2804
R-free	0.257 (0.3390)	0.3453
Number of non-hydrogen atoms	3445	2821
Macromolecules	3445	2821
Ligands	2	0
Solvent	0	0
Protein residues	278	237
RMS (bonds) (Å)	0.24	0.16
RMS (angles) (°)	0.46	0.42
Ramachandran favored (%)	98	96
Ramachandran allowed (%)	2	3
Ramachandran outliers (%)	0	1
Rotamer outliers (%)	0	0
Clashscore	5	4
Average B-factor	124	309

Table S4.11. X-ray diffraction statistics for CC1^{+10} PDB entries 9YZU and 9Z44

Protocol S4.1. Protein sequences for cloning, overexpression in *E.coli*, and $CC1^{+10}$ guest protein sequences

RepE54 Transcription Factor protein sequence.

RepE54 transcription factor (sequence below) was cloned into pSB3 vector (Addgene plasmid #82027) with an N-terminal His-tag.

MRGSHHHHHHGSMAETAVINHKKRKNSPRIVQSNDLTEAAYSLSRDQKRMLYLFVDQI
RKSDGTLQEHDGICEIHVAKYAEIFGLTSAEASKDIRQALKSFAGKEVVFYRPEEDAGDE
KGYESFPWFIKPAHSPSRGLYSVHINPYLIPFFIGLQNRFTQFRLSETKEITNPYAMRLYES
LCQYRKPDGSGIVSLKIDWIIERYQLPQSYQRMPDFRRRFLQVCVNEINSRTPMRLSYIEK
KKGRQTTHIVFSFRDITSMTTG**

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.

MEKRPRTAFSSEQLARLKREFNENRYLTERRRQQLSSELGLNEAQIKIWFQNKRAKIKKS
TSQFYLNEMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTT
GKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKT
RAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKI
RHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTA
AGITLGMDELYKHHHHHH**

Even-skipped Homeodomain (Eve-HD) protein sequence.

GTGSVRRYRTAFTRDQLGRLEKEFYKENYVSRPRRCELAAQLNLPESTIKVWFQNRRM

KDKRQR

Ultrabithorax Homeodomain (UBX-HD) protein sequence.

GTRRRGRQTYTRYQTLELEKEFHTNHYLTRRRRIEMAHALCLTERQIKIWFQNRRMKLK
KEI

Antennapedia Homeodomain (AntP-HD) protein sequence.

GTRKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWK
KEN

C-clamp domain of human HDBP1 (C-clamp) protein sequence.

GDAKKCRKVYGMERRDLWCTACRWKKACQRFLD

DNA-binding domain of the *Saccharomyces cerevisiae* GCN4 transcription factor (bZip) protein

sequence.

DPAALKRARNTEAARRSRARKLQRMKKQC

Protocol S4.2. Alphafold-3 modelling of guests into CC1 lattice

To evaluate potential guest installation sites in the porous $CC1^{+10}$ scaffold, we used AlphaFold-3 (AF3)[82] in combination with PyMOL to model guest protein binding within the context of the full crystal lattice. This protocol outlines the workflow used to assess guest compatibility with the $CC1^{+10}$ insert DNA.

Alphafold-3 structure prediction calculations were made containing: (1) the RepE54 scaffold protein sequence, (2) the full $CC1^{+10}$ insert DNA sequence containing the intended guest-binding site, and (3) the sequence of the DNA-binding guest protein. For each guest family, we tested multiple registers by inserting the canonical DNA-binding motif (e.g., 5′-TAATTA-3′ for homeodomains) into different positions along the insert DNA.

The top five predicted structures (with pLDDT > 0.70) were loaded into PyMOL and visually inspected for consistency in guest binding location. To assess lattice compatibility, we manually appended the CRYS1 record (containing space group and unit cell details) from a representative experimental $CC1^{+10}$ PDB file to the AF3 model file. The AF3-predicted DNA duplex was then aligned to the corresponding DNA segment in the $CC1^{+10}$ asymmetric unit using RMSD-based structural alignment to conserve ASU coordinates.

After alignment, we used PyMOL's supercell plugin to expand the asymmetric unit into a full lattice. The resulting lattice environment was inspected to assess steric clashes, crystal packing compatibility, and accessibility of the guest protein to solvent channels. Models were prioritized based on high model confidence, realistic guest–DNA interactions, and sterically feasible placement within the lattice.

Protocol S4.3 Trace labeling of UBX-HD with NHS-Fluorescein

UBX-HD was dialyzed into labeling buffer (50 mM Na_2PO_4, 300 mM NaCl, pH 8.1) using Slide-A-Lyzer™ dialysis pucks (1 kDa MWCO; ThermoFisher) to remove free amines and equilibrate the sample for N-hydroxysuccinimide (NHS) conjugation. The protein concentration was 200 μM prior to labeling. NHS-fluorescein (Thermo Scientific; 5 mg/mL, ~10 mM) was added at a 1:5 dye:protein molar ratio, and the reaction was incubated overnight at 4 °C with gentle agitation. With five surface lysine residues and the N-terminus, the dye conjugation location and extent is presumably heterogeneous. To minimize photobleaching, dye-handling steps were performed in opaque 1.5 mL microcentrifuge tubes and under low-light conditions. Unreacted dye was removed by dialysis against fresh labeling buffer (50 mM Na_2PO_4, 300 mM NaCl, pH 8.1, 1 kDa MWCO) with multiple buffer exchanges until no residual dye was visible in the dialysate. The labeled protein was concentrated to the working concentration using a 3 kDa MWCO Amicon centrifugal filter unit (Millipore) and stored on ice for immediate use, or aliquoted and flash-frozen for long-term storage at −80 °C.

Protocol S4.4. Loaded crystal imaging

Crystals loaded with fluorescently labeled UBX-HD were imaged using a Nikon Eclipse Ti spinning-disk confocal microscope equipped with an AndoriXon Ultra 879U EMCCD camera. Fluorescein signal was collected using 488 nm laser excitation and a 515 nm emission filter. Images were acquired using a 10X objective lens, and Z-stacks were collected using 2.5 μm step sizes for a total of 37 steps in all experiments. Laser power, exposure time, and all other optical parameters were held constant across experiments to enable direct comparison of fluorescence intensity between crystals. Image acquisition and processing were performed using Nikon NIS-Elements software.