Supplemental Information

Stabilizing DNA-protein co-crystals via in crystallo chemical ligation of the DNA

Abigail R. Ward, Sara Dmytriw, Ananya Vajapayajula, Christopher D. Snow

Table of Contents

Supplemental Information – Figures S1 – S13

Supplemental Information – Tables S1 – S4

Supplemental Information – Protocols S1 – S4

SUPPLEMENTAL INFORMATION - FIGURES

Figure S1. TapeStation analysis and matching gel electrophoresis. (A) The TapeStation analysis for crosslinked crystals of CC1 3' phosphate shows the ligation end product distribution. The crosslinked crystals are represented with 'L', 'M', and 'H'. The crosslink values correspond with Figure 4 in the text. (B) The corresponding TBE-urea gel shown is from the same samples as the TapeStation samples.

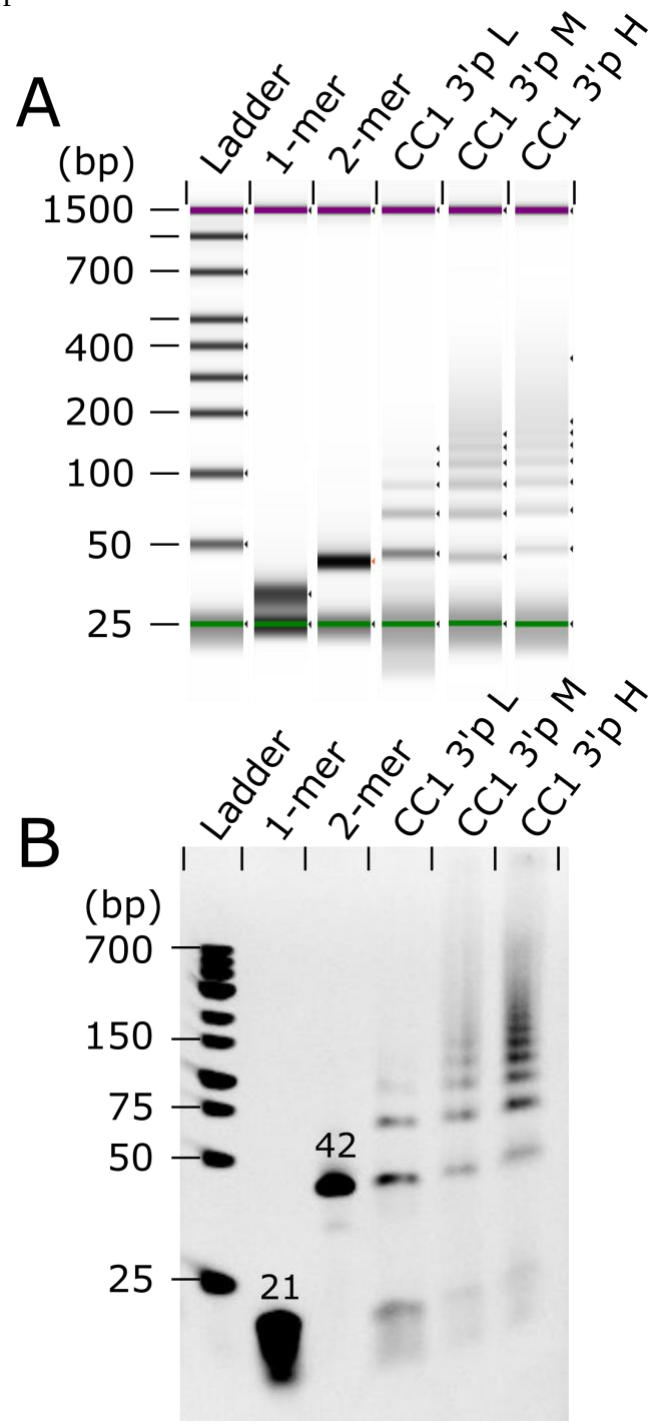
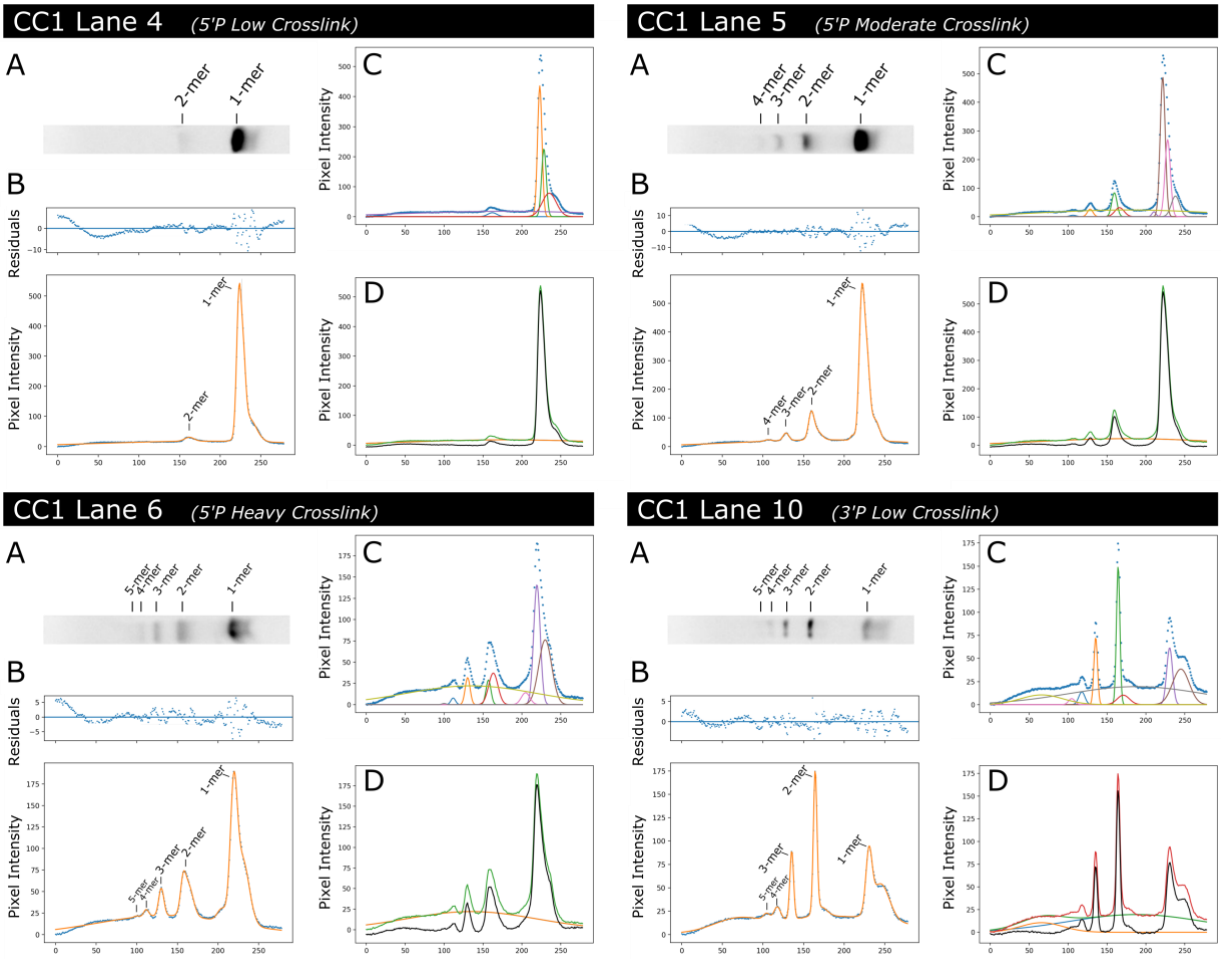
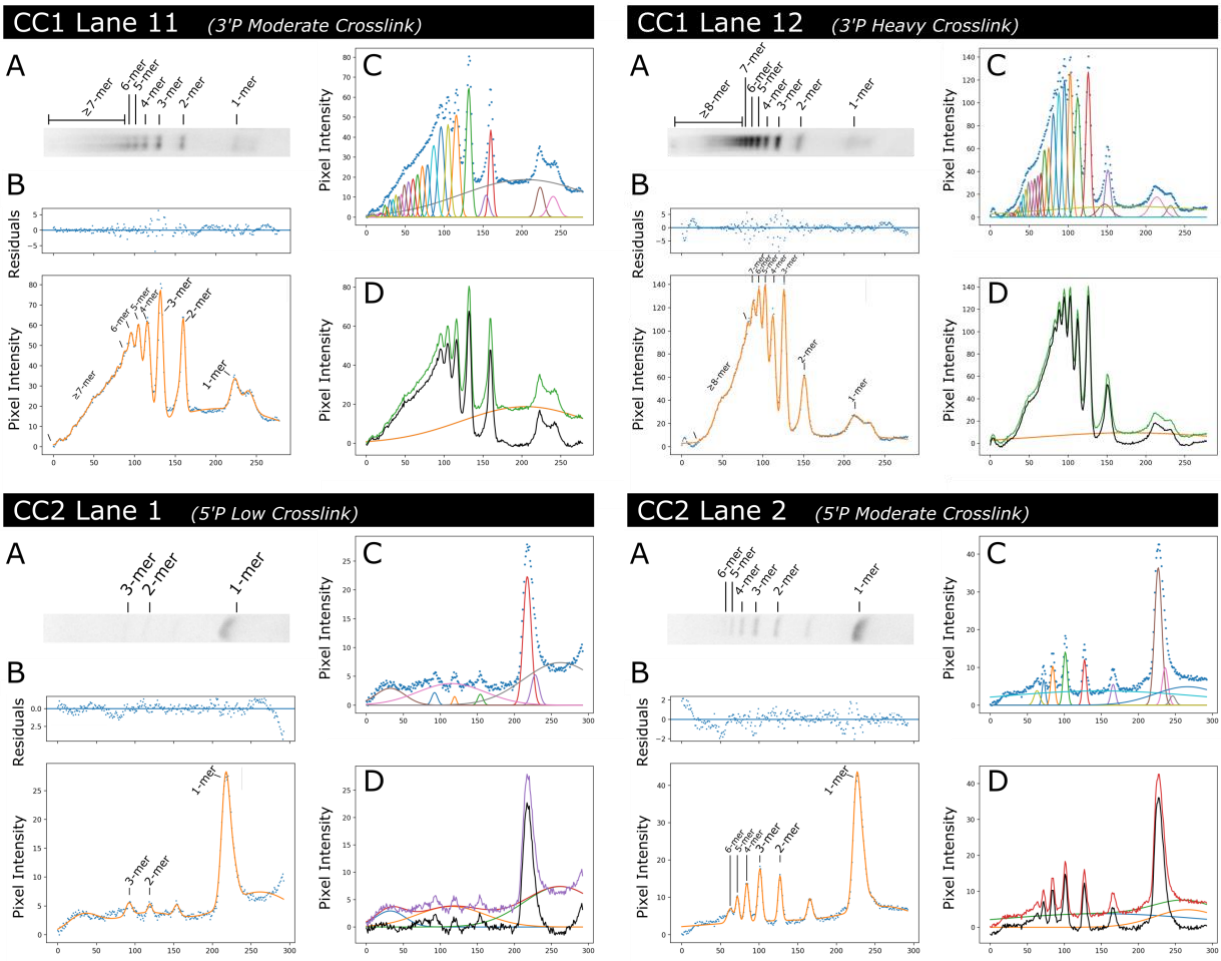
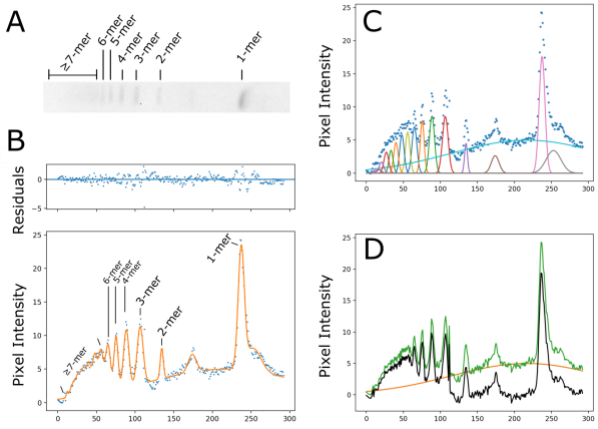


Figure S2. Densitometry results and annotation (corresponds to main text Figure 4).
For each lane, we conducted a multi-step densitometry analysis. (A) The lane was manually excised from the gel using ImageJ software. (B) The final fit (orange line) closely follows the raw intensity data (blue dots) which is the average intensity value across the lane (perpendicular to $\vec{E}$ direction). The quality of the fit is also evident in the small residuals (above). (C) The compenent gaussian functions inside the fit are shown with different colors.
(D) If we subtract the diffuse background (orange) from the full fit (green) the background curve is shown in black.

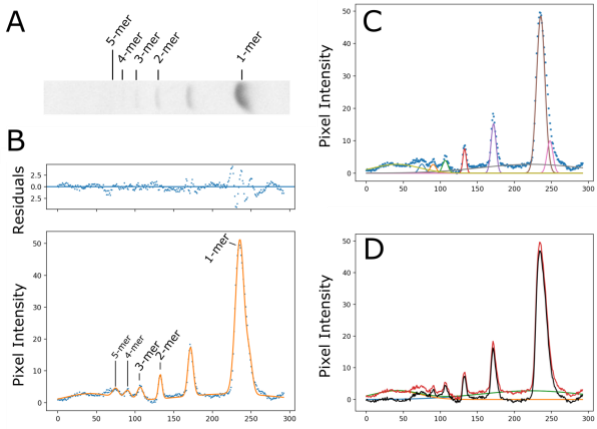



CC1 Lane 11 (3'P Moderate Crosslink)
CC1 Lane 12 (3'P Heavy Crosslink)
CC2 Lane 1 (5'P Low Crosslink)
CC2 Lane 2 (5'P Moderate Crosslink)
A
B
C
D
Residuals
Pixel Intensity
≥7-mer
6-mer
5-mer
4-mer
3-mer
2-mer
1-mer
≥8-mer
7-mer

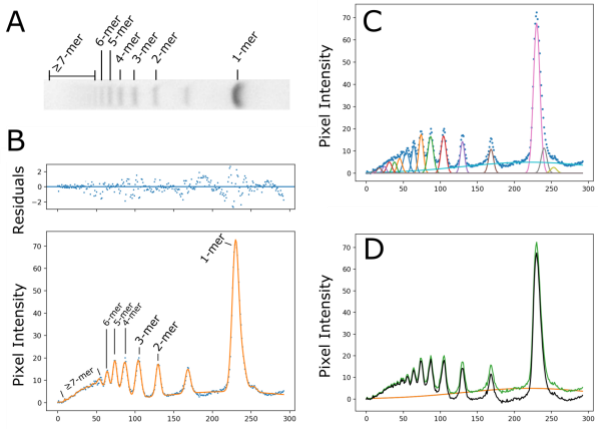
CC2 Lane 3 *(5'P Heavy Crosslink)*



CC2 Lane 7 *(3'P Low Crosslink)*



CC2 Lane 8 *(3'P Moderate Crosslink)*



CC2 Lane 9 *(3'P Heavy Crosslink)*

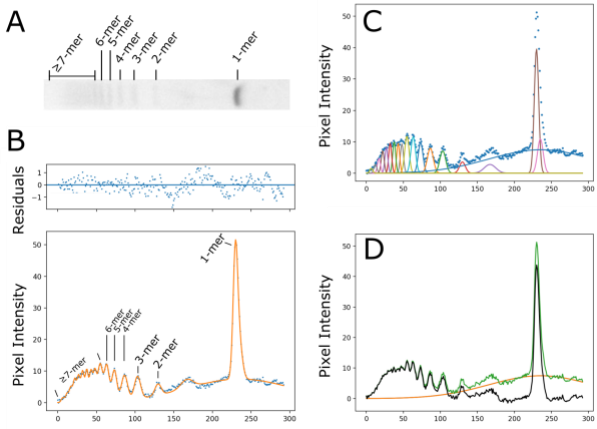


Figure S3. Co-crystal stability test - water. The crystals were crosslinked at 15 mg/mL EDC for 20 hours and quenched with Tris base pH 8.2 for 30 minutes prior to transfer to the wash solution. All scale bars are 100 µm. (A) CC1 crystals in wash solution containing 50 mM NaCl, 14% PEG 400, and 200 mM MES buffer pH 6.0. The concentrations of the wash solution matched the initial crystal growth solutions but we replaced $MgCl_2$ with NaCl and Tris HCl pH 8.0 with MES buffer pH 6.0. (B) CC1 crystals after transfer to deionized water. The crosslinked crystals (left three panels) remained intact for at least 7 days. The non-crosslinked crystals (right three panels) dissolved or converted to an aggregate at various immediate timepoints. (C) CC2 crystals in wash solution containing 160 mM lithium sulfate, 5% PEG 400, 13.3% PEG 3350 and 80 mM MES buffer pH 6.0. The concentrations of the wash solution matched the initial crystal growth solutions but we replaced ammonium sulfate with lithium sulfate and HEPES buffer pH 7.1 with MES buffer pH 6.0. (D) CC2 crystals after transfer to deionized water. The crosslinked crystals (left three panels) remained intact for at least 7 days. The non-crosslinked crystals (right three panels) dissolved or converted to an aggregate at various immediate timepoints. Surprisingly, non-crosslinked CC2 3' phosphate crystals did not dissolve after 20 hours in water.

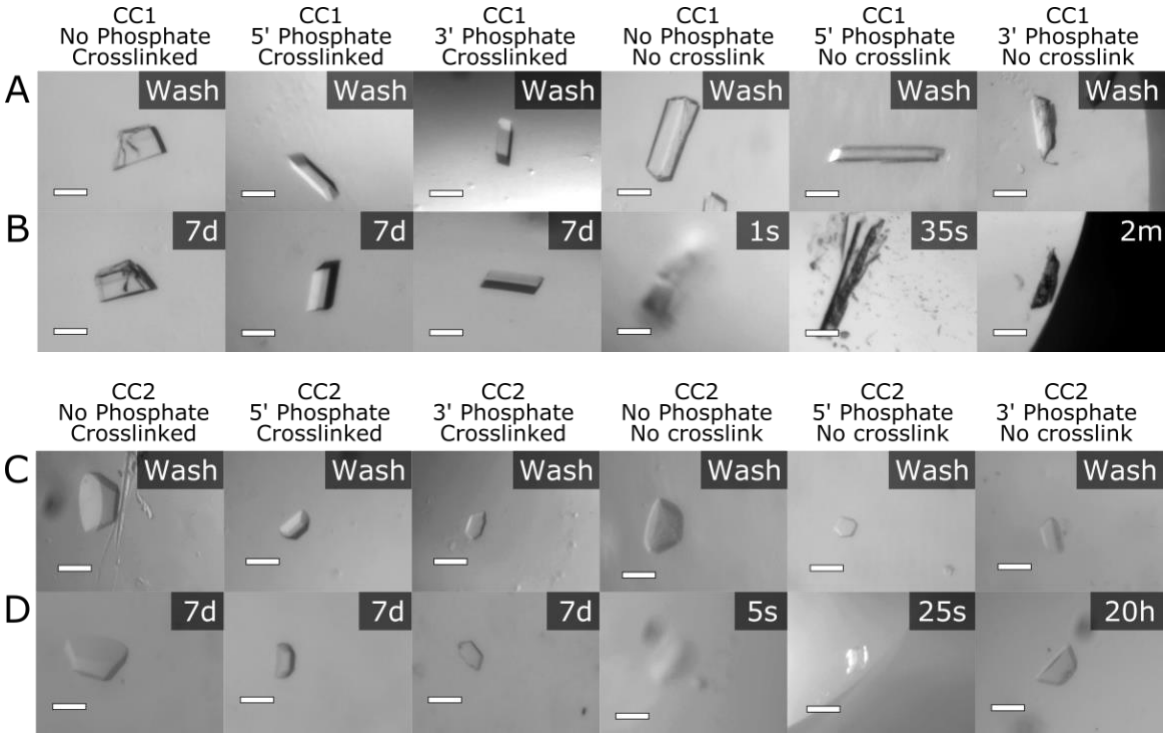


Figure S4. Co-crystal stability test – very low pH 2.0 to mimic stomach acid. The crystals were crosslinked at 15 mg/mL EDC for 20 hours and quenched with Tris base pH 8.2 for 30 minutes prior to transfer to the wash solution. All scale bars are 100 µm. Percent volume increases were measured using Protocol S4. (A) CC1 crystals in wash solution containing 50 mM NaCl, 14% PEG 400, and 200 mM MES buffer pH 6.0. The concentrations of the wash solution matched the initial crystal growth solutions, but we replaced $MgCl_2$ with NaCl and Tris HCl pH 8.0 with MES buffer pH 6.0. (B) CC1 crystals after transitioning to a pH 2.0 stomach acid mimic (0.01M HCl). The crosslinked crystals (left three panels) showed varied results after 24 hours. The crosslinked CC1 no phosphates expanded ~430% ± 70% in volume after 24 hours from the original washed crystal. The crystals with terminal phosphates did not noticeably change in macro-structure. The non-crosslinked crystals (right three panels) dissolved or converted to an aggregate at various immediate timepoints. (C) 5 days in pH 2.0 buffer the crosslinked crystals showed varied results. The CC1 no phosphate and 5' phosphate expanded by ~440% ± 70% and ~140% ± 40% in volume after 5 days from the original washed crystal. The CC1 3' phosphate did not change in macro-structure. (D) After 7 days in pH 2.0 buffer, each of the crosslinked crystals expanded in volume from the original washed crystal: CC1 no phosphate by ~460% ± 70%, CC1 5' phosphate by ~240% ± 60%, and CC1 3' phosphate by ~62% ± 0%. (E) CC2 crystals in wash solution containing 160 mM lithium sulfate, 5% PEG 400, 13.3% PEG 3350 and 80 mM MES buffer pH 6.0. The concentrations of the wash solution matched the initial crystal growth solutions, but we replaced ammonium sulfate with lithium sulfate and HEPES buffer pH 7.1 with MES buffer pH 6.0. (F) CC2 crystals after transitioning to a pH 2.0 stomach acid mimic (0.01M HCl). The crosslinked crystals (left three panels) expanded in volume after 24 hours from the original washed crystal: CC2 no phosphate by ~640% ± 20%, CC2 5' phosphate by ~1390% ± 60%, and CC2 3' phosphate by ~440% ± 40%. The non-crosslinked crystals (right three panels) appeared to degrade into a protein aggregate at various immediate timepoints. (G) In contrast, after 7 days, CC2 crosslinked crystals still had an expanded macro-structure but they did not dissolve.

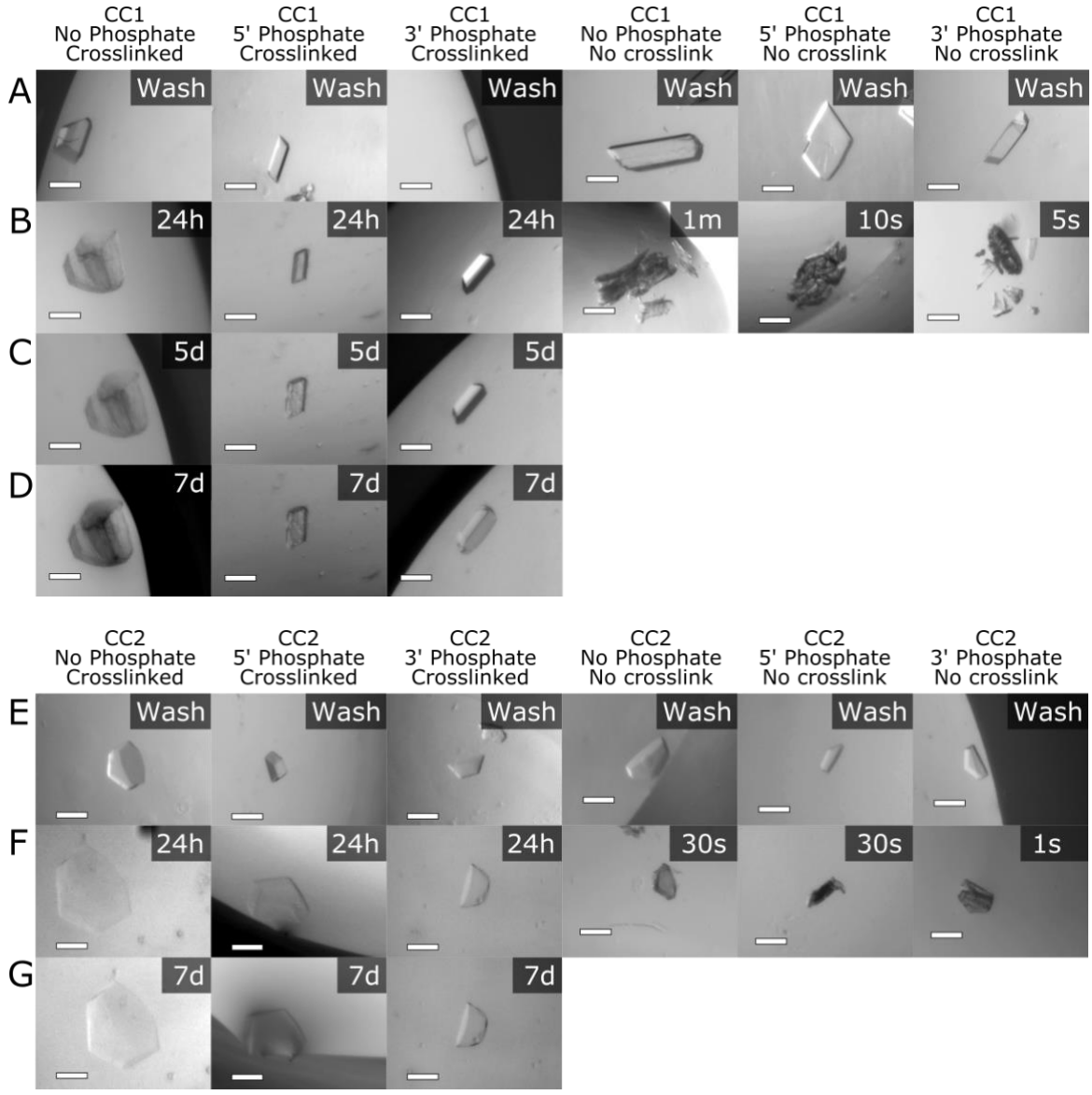
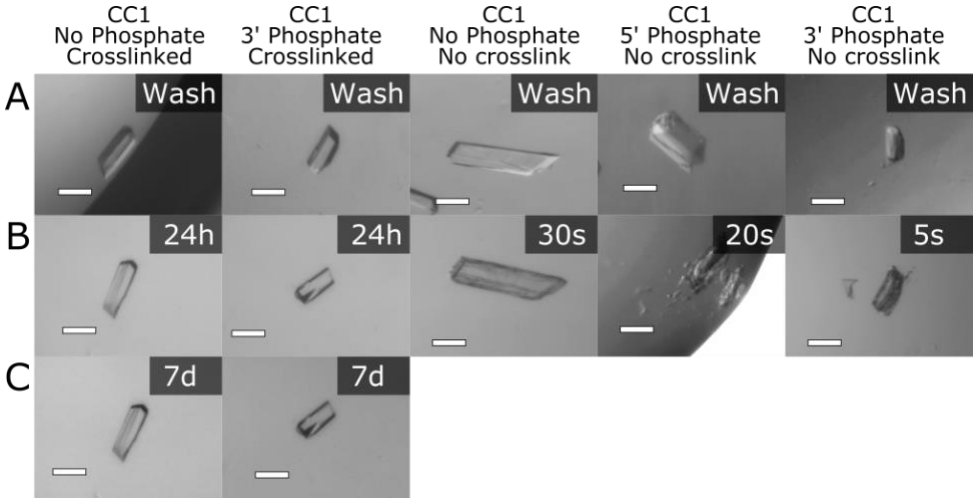

CC1
No Phosphate
Crosslinked
CC1
5' Phosphate
Crosslinked
CC1
3' Phosphate
Crosslinked
CC1
No Phosphate
No crosslink
CC1
5' Phosphate
No crosslink
CC1
3' Phosphate
No crosslink
A
Wash
Wash
Wash
Wash
Wash
Wash
B
24h
24h
24h
1m
10s
5s
C
5d
5d
5d
D
7d
7d
7d
CC2
No Phosphate
Crosslinked
CC2
5' Phosphate
Crosslinked
CC2
3' Phosphate
Crosslinked
CC2
No Phosphate
No crosslink
CC2
5' Phosphate
No crosslink
CC2
3' Phosphate
No crosslink
E
Wash
Wash
Wash
Wash
Wash
Wash
F
24h
24h
24h
30s
30s
1s
G
7d
7d
7d

Figure S5. Co-crystal stability test – moderately low pH 4.5 to mimic lysosomal fluid. The crystals were crosslinked at 15 mg/mL EDC for 20 hours and quenched with Tris base pH 8.2 for 30 minutes prior to transfer to the wash solution. All scale bars are 100 µm. (A) CC1 crystals in wash solution containing 50 mM NaCl, 14% PEG 400, and 200 mM MES buffer pH 6.0. The concentrations of the wash solution matched the initial crystal growth solutions but we replaced $MgCl_2$ with NaCl and Tris HCl pH 8.0 with MES buffer pH 6.0. (B) CC1 crystals after transitioning to a pH 4.5 lysosomal fluid mimic (46 mM sodium citrate, 54.1 mM citric acid). The crosslinked crystals (left two panels) were void of macrostructural changes after 24 hours whereas the uncrosslinked crystals (right three panels) either visibly degraded or dissolved rapidly. (C) After 7 days in pH 2.0 buffer, each of the crosslinked crystals lacked obvious changes. (D) CC2 crystals in wash solution containing 160 mM lithium sulfate, 5% PEG 400, 13.3% PEG 3350 and 80 mM MES buffer pH 6.0. The concentrations of the wash solution matched the initial crystal growth solutions but we replaced ammonium sulfate with lithium sulfate and HEPES buffer pH 7.1 with MES buffer pH 6.0. (E) CC2 crystals after transitioning to a pH 4.5 lysosomal fluid mimic (46 mM sodium citrate, 54.1 mM citric acid). The crosslinked crystals lacked macrostructural changes after 24 hours (left three panels). Surprisingly, the non-crosslinked crystals remained relatively unperturbed after a short time in the harsh condition (right three panels). They dissolved after ~24 hours (not shown). (G) After 7 days, CC2 crosslinked crystals remained without visible damage.



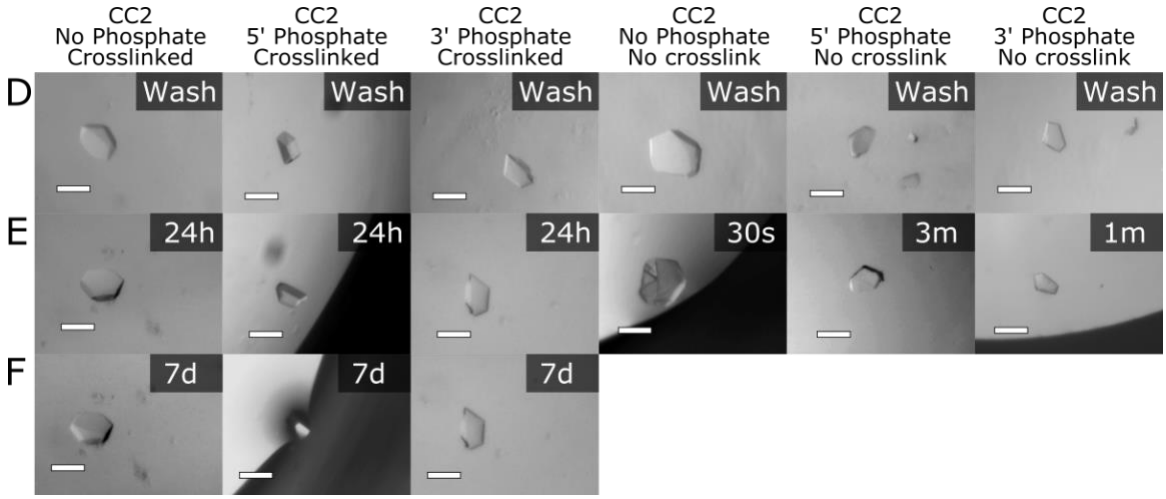
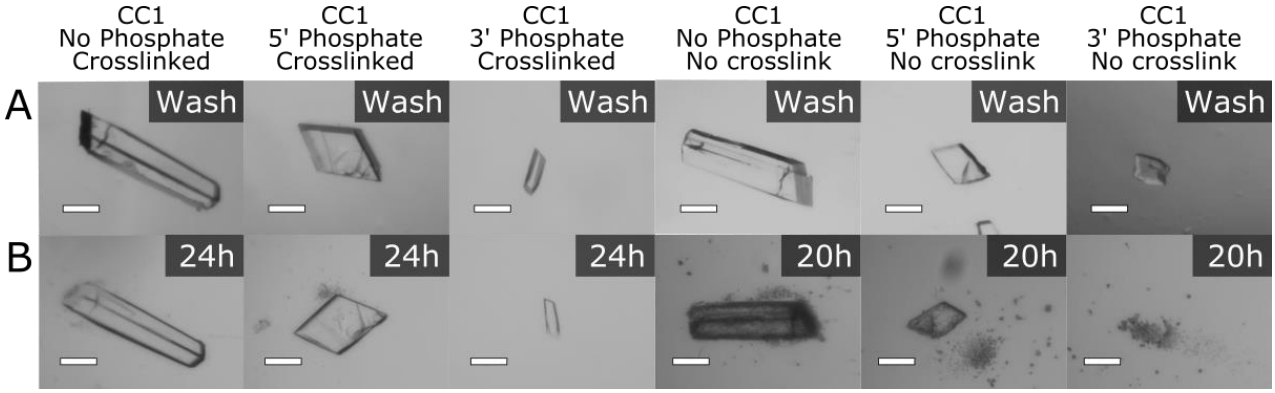

CC2
No Phosphate
Crosslinked
CC2
5' Phosphate
Crosslinked
CC2
3' Phosphate
Crosslinked
CC2
No Phosphate
No crosslink
CC2
5' Phosphate
No crosslink
CC2
3' Phosphate
No crosslink
D
Wash
Wash
Wash
Wash
Wash
Wash
E
24h
24h
24h
30s
3m
1m
F
7d
7d
7d

Figure S6. Co-crystal stability test – blood serum. The crystals were crosslinked at 15 mg/mL EDC for 20 hours and quenched with Tris base pH 8.2 for 30 minutes prior to transfer to the wash solution. All scale bars are 100 µm. (A) CC1 crystals in wash solution containing 80 mM NaCl, 5% PEG 400, and 160 mM MES buffer pH 6.0. The concentrations of the wash solution matched the initial crystal growth solutions, but we replaced $MgCl_2$ with NaCl and Tris HCl pH 8.0 with MES buffer pH 6.0. (B) CC1 crystals after transfer to blood serum (HyClone, Bovine Calf Serum Product #: SH30073.02). The crosslinked crystals (left three panels) remained intact for at least 24 hours. The non-crosslinked crystals (right three panels) dissolved or degraded after 20 hours in the serum. (C) CC2 crystals in wash solution containing 240 mM lithium sulfate, 5% PEG 400, 22% PEG 3350 and 120 mM MES buffer pH 6.0. The concentrations of the wash solution matched the initial crystal growth solutions, but we replaced ammonium sulfate with lithium sulfate and HEPES buffer pH 7.1 with MES buffer pH 6.0. (D) CC2 crystals after transitioning to blood serum. The crosslinked crystal (left panel) remained intact for at least 24 hours in blood serum. The non-crosslinked crystals (right panel) lost macrostructure after 20 hours in blood serum.



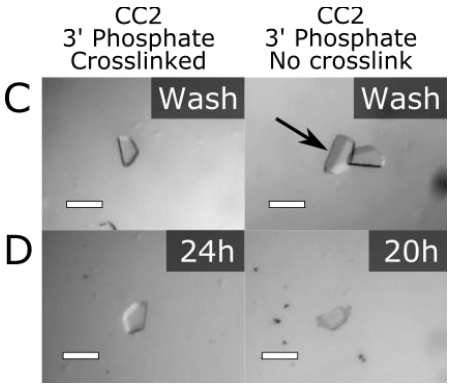


Figure S7. Gel electrophoresis of varied EDC crosslink time. All crosslink lanes shown in this figure were crystals dissolved after incubation in 15 mg/mL EDC solution for the time point indicated on the gel figure (1 hour to 48 hours). Each crystal was dissolved and analyzed on a TBE-urea gel. (A) CC1 co-crystal samples with 3' phosphates (B) CC1 co-crystal samples with 5' phosphates (C) CC2 co-crystal samples with 3' phosphates and (D) CC2 co-crystal samples with 5' phosphates.

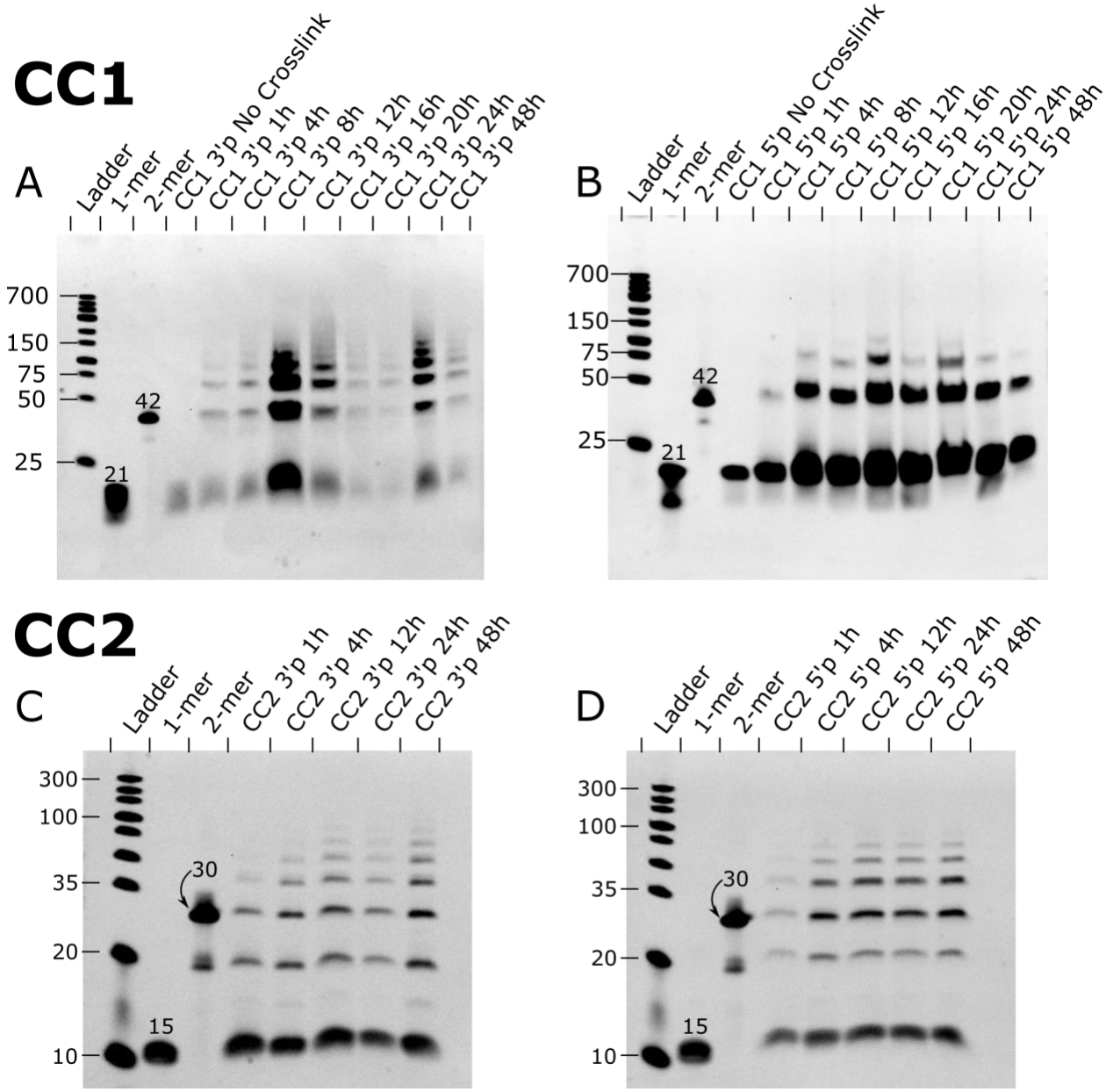


Figure S8. Gel electrophoresis of varied EDC crosslink concentration. All crosslink lanes shown in this figure were crystals dissolved after incubation for 12 hours in an EDC solution at the concentration indicated on the gel figure (5 mg/mL EDC to 80 mg/mL). Each crystal sample was dissolved and analyzed on a TBE-urea gel. (A) CC1 co-crystal samples with 3' phosphates (B) CC1 co-crystal samples with 5' phosphates (C) CC2 co-crystal samples with 3' phosphates and (D) CC2 co-crystal samples with 5' phosphates.

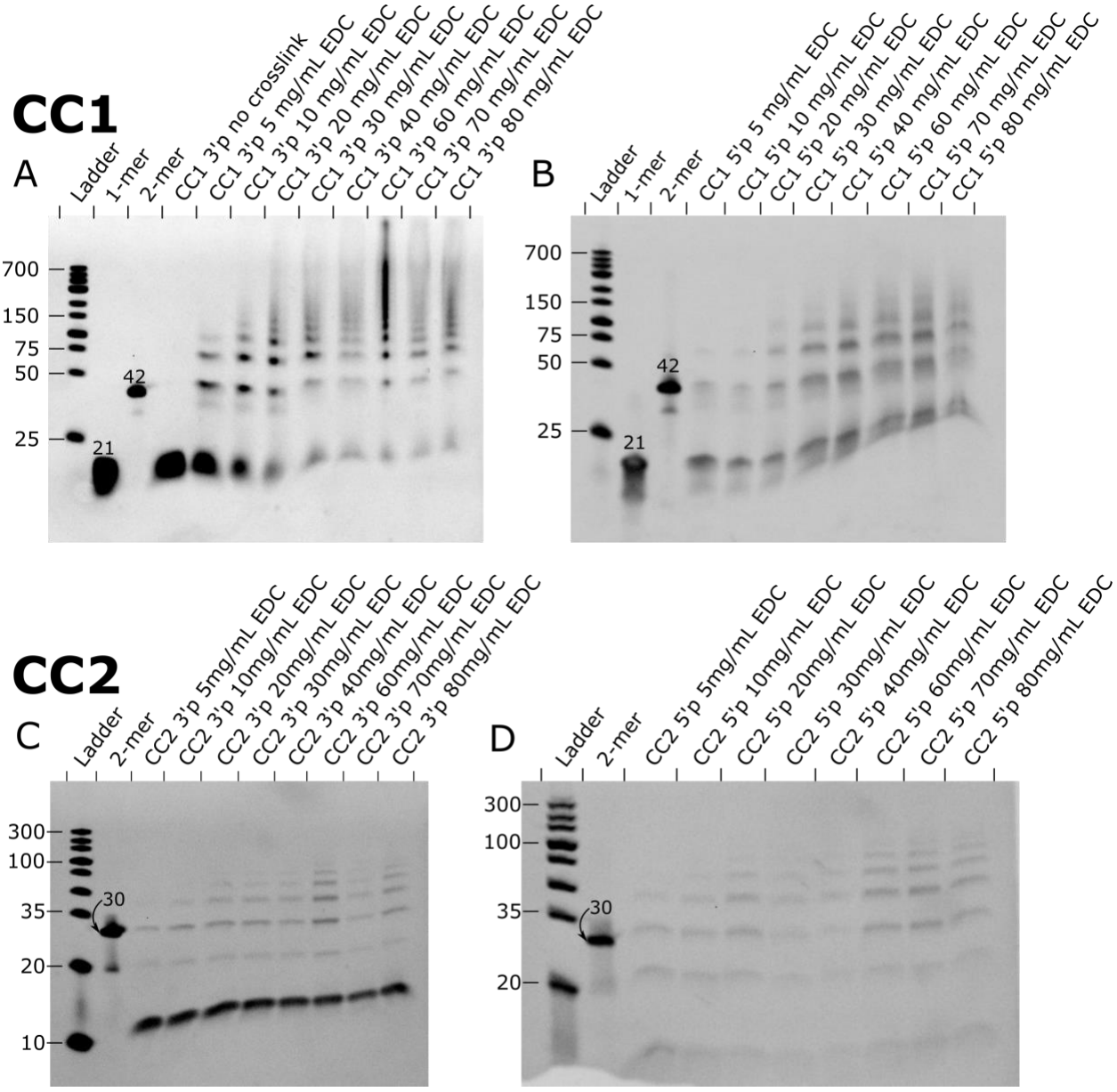


Figure S9. Schematic and gel electrophoresis of varied EDC crosslink dose. All crosslink lanes shown in this figure were crystals dissolved after incubation in a set number of doses (1 to 4 doses) of 30 mg/mL EDC solution for 12 hours. Each crystal was dissolved and run on a TBE-urea gel. (A) A schematic of the crosslinking dose experiment. The co-crystal is transferred to a fresh 30 mg/mL EDC dose every 12 hours. (B) CC1 co-crystal samples with 5' phosphate and with 3' phosphates (C) CC2 co-crystal samples with 3' phosphates and (D) CC2 co-crystal samples with 5' phosphates.

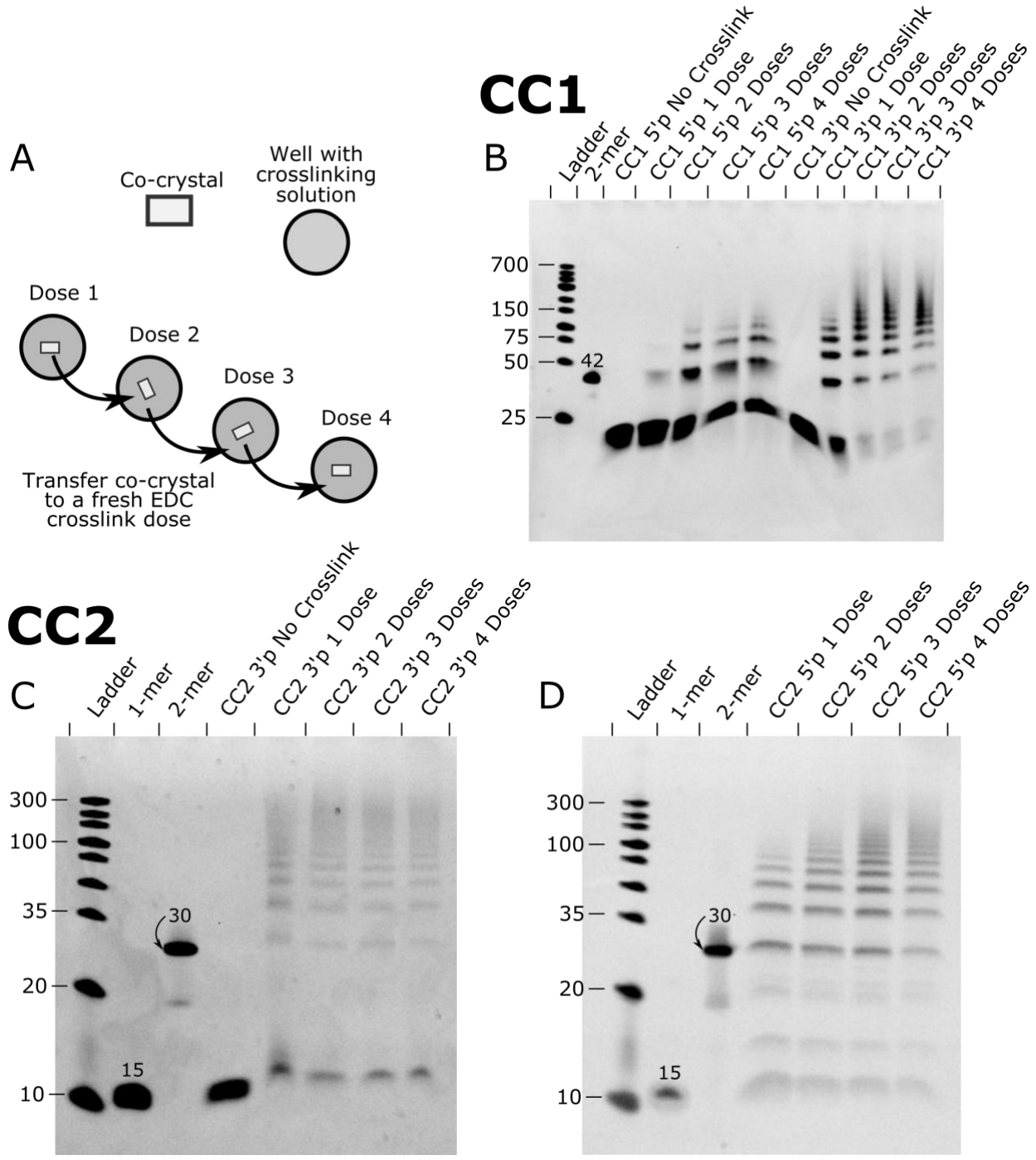


Figure S10. Schematic and gel electrophoresis of the controls – crystals with no terminal phosphates and duplexes with terminal phosphates in-solution. All crosslink trials shown in this figure were subjected to a 30 mg/mL EDC solution for 20 hours. Each sample was dissolved and run on a TBE-urea gel. (A) A schematic of the co-crystal DNA-DNA junction with no terminal phosphates. The terminal hydroxyls cannot react with EDC to ligate the flanking DNA duplexes. R_1 is the nucleobase and R_2 is the phosphate backbone. (B) We hypothesize DNA duplexes in-solution are randomly ordered, such that the DNA ends do not frequently stack as in the co-crystals. Therefore, the blunt ended duplexes cannot be ligated with EDC crosslinking as shown in the gels C and D. (C) CC1 co-crystal with no terminal phosphate and CC1 DNA duplexes 2.5 ng/µL (192nM) in-solution with terminal 3' phosphates or 5' phosphates. (D) CC2 co-crystal with no terminal phosphates and CC2 DNA duplexes 2.5 ng/µL (192nM) in-solution with terminal 3' phosphates or 5' phosphates.

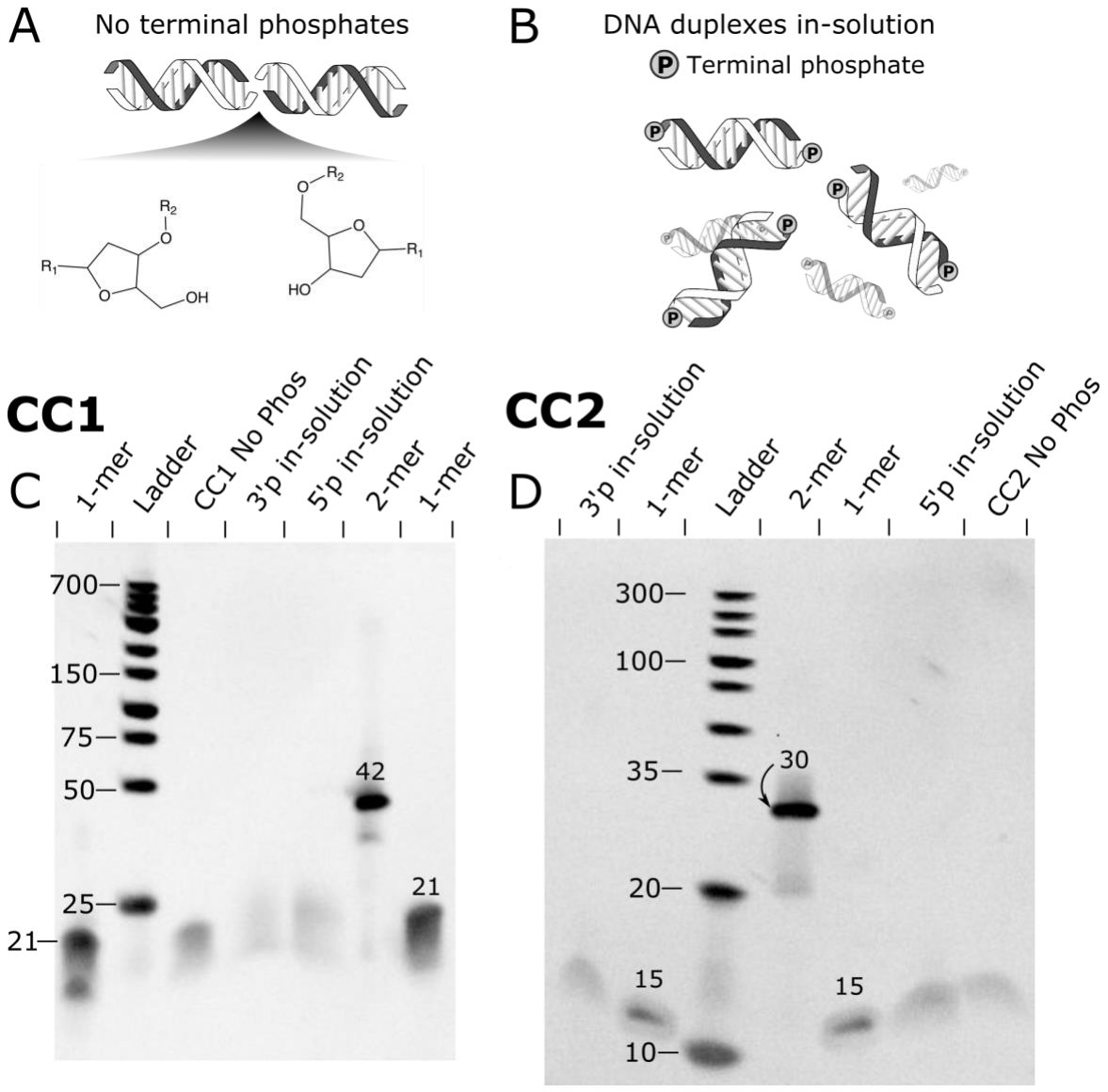


Figure S11. Magnesium chloride's effect on the EDC crosslinking of CC1 crystals. All crosslink trials shown in this figure were subjected to a 30 mg/mL EDC solution for 20 hours. Each CC1 sample was dissolved and run on a TBE-urea gel. The magnesium chloride was replaced with sodium chloride and the resulting crosslinks were compared.

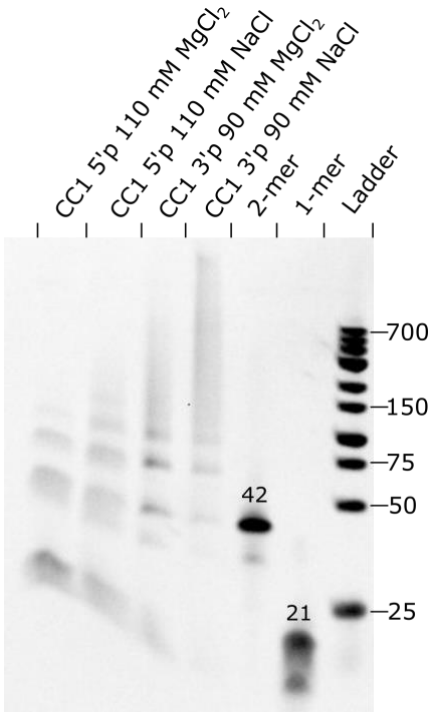
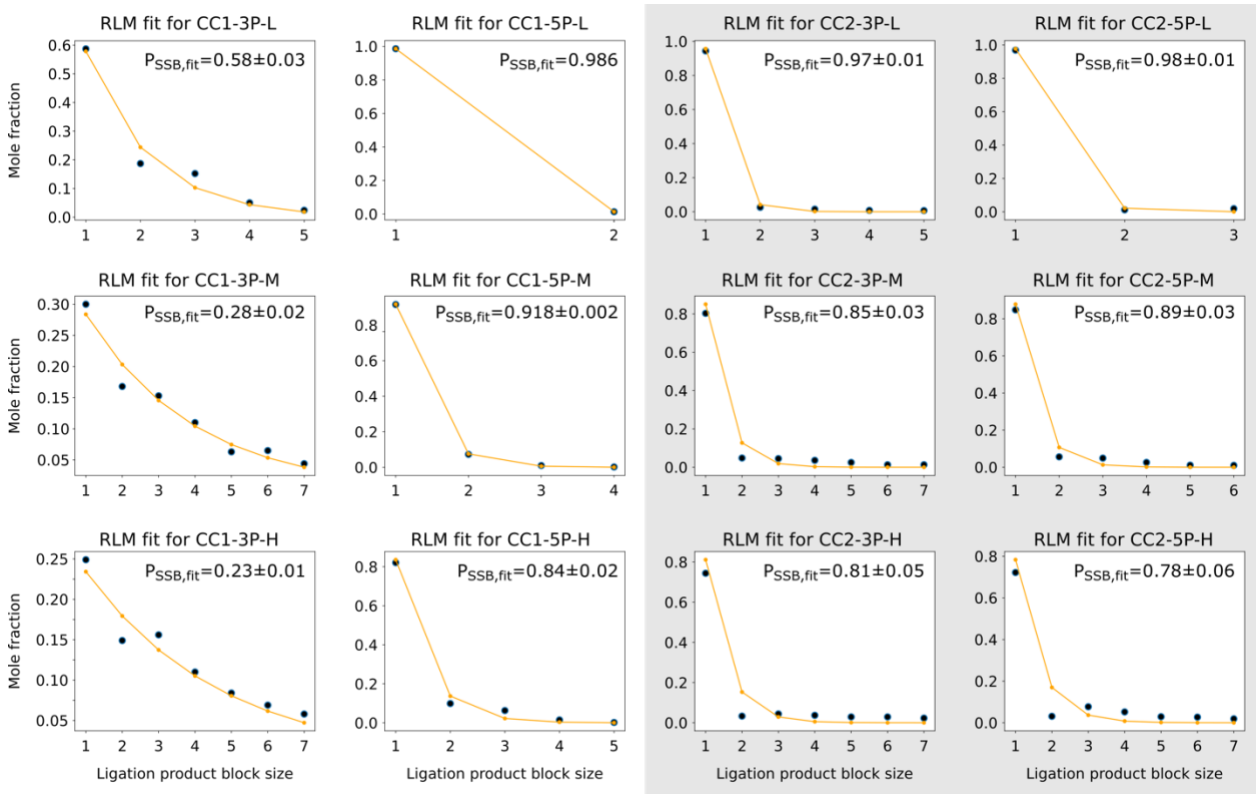


Figure S12. Random ligation model fit to mole fractions derived from gel densitometry
Experimental gel densitometry data (black discs), with mole fractions corrected for DNA length (Table 2), were fit to the geometric decay predicted by the random ligation model (RLM, orange line, Protocol S2) using non-linear least squares via the scipy.optimize.curve_fit function. The CC1 ligation product distribution for 3' phosphates (column 1) and 5' phosphates (column 2) was fit well with the RLM, albeit with a small systematic deficit of 2-mer and surfeit of 3-mer. The CC1 fit P_{SSB} values (inset, along with the standard error) also closely matched the P_{SSB} values calculated from the mole fractions. In contrast, the RLM provided an inferior fit for CC2 data (higher standard errors), overestimating the 1-mer and 2-mer fractions and underestimating higher-order product mole fractions. While CC2 fit P_{SSB} values only somewhat exceed the measured 1-mer mole fractions, underlined entries significantly deviate from P_{SSB} values derived from main text Equation 1 (reprised below). The biased ligation model (Protocol S3), or asymmetry in the ligation yield for one of the two nick sites, could address this CC2 discrepancy.



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

Figure S13. Terminal phosphate positions and symmetry

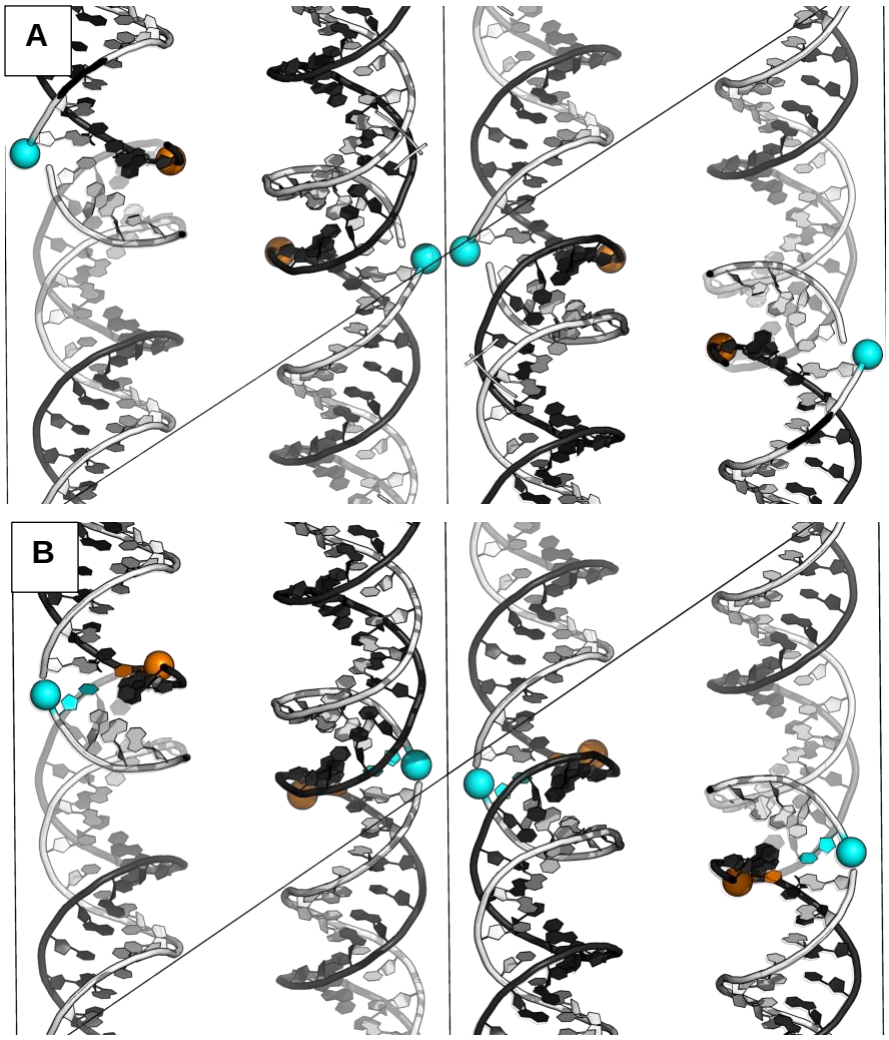


Figure S13. Terminal phosphate position. **(A)** CC1-3'P in our new PDB entry 7sdp has 3' terminal phosphates on chain A (orange spheres) and chain B (cyan spheres). Due to crystallographic symmetry, the chain B 3'phosphates are quite close (5.65Å) which might contribute to increased ligation propensity for the chain B nick compared to the chain A nick. Protein is hidden for clarity. **(B)** CC1-5'P in our new PDB entry 7sgc has 5' terminal phosphates on chain A (orange spheres) and chain B (cyan spheres). Notably, the chain B 5'phosphates are farther apart (9.29Å) than the 3' phosphate case above, which may decrease the electrostatic repulsion driving force for ligation. Protein is hidden for clarity.

SUPPLEMENTAL INFORMATION - TABLES

Table S1. DNA oligonucleotide sequences used in this study.

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

Table S2. Ligation percentages from gel densitometry (unweighted)

The data below corresponds with Table 2 in the text. Shown here is the unweighted distribution where the length of the DNA and resulting dye binding quantity is not incorporated in the percentages.

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

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

Table S3. Full version of densitometry output Table 2.
This version of Table 4.2 includes the small mole fractions for higher-order peaks.

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

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

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

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

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

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

Table S4. Absolute electron density values for the Figure 5 electron density maps
The electron density contours in the Figure 5 meshes are set to 3.0 to match COOT 3.0 rmsd contours. By loading the same maps into COOT, and setting the contours to 3.0 rmsd, we can also obtain the contour value for the electron density on an absolute scale.

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

SUPPLEMENTAL INFORMATION - PROTOCOLS

Protocol S1. Protein sequences for cloning and overexpression in *E. coli*.

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

MRGSHHHHHHGSMAETAVINHKKRKNSPRIVQSNDLTEAAYSLSRDQKRMLYLFVDQIRKSDGTLQEHDGI
CEIHVAKYAEIFGLTSAEASKDIRQALKSFAGKEVVFYRPEEDAGDEKGYESFPWFIKPAHSPSRGLYSVHINPY
LIPFFIGLQNRFTQFRLSETKEITNPYAMRLYESLCQYRKPDGSGIVSLKIDWIIERYQLPQSYQRMPDFRRRFL
QVCVNEINSRTPMRLSYIEKKKGRQTTHIVFSFRDITSMTTG**

E2F8 Transcription Factor (CC2) protein sequence
The expression plasmid (pETG20A-SBP) containing E2F8 Transcription Factor was donated by the Taipale Lab. The protein sequence contained an N-terminal His-tag and Thioredoxin with a **TEV cleavage site** before the remainder of the E2F8 Transcription Factor.

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

Protocol S2. Random Ligation Model: Simulation and Mathematics.
To calculate the ligation product size distribution expected for random ligation throughout the crystal we use both simulation and mathematical arguments, which agree to high precision.

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

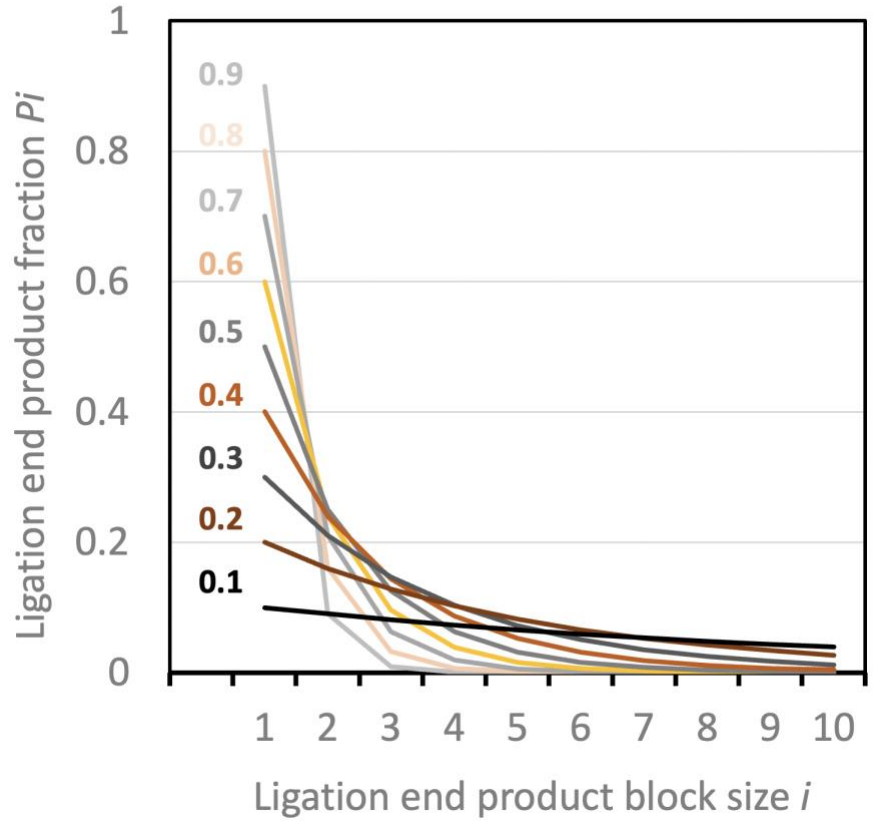


Figure Protocol S2. Product distributions from simulated random ligation. Each trace is labeled on the left with the P_{SSB} threshold value used to halt the ligation simulation.

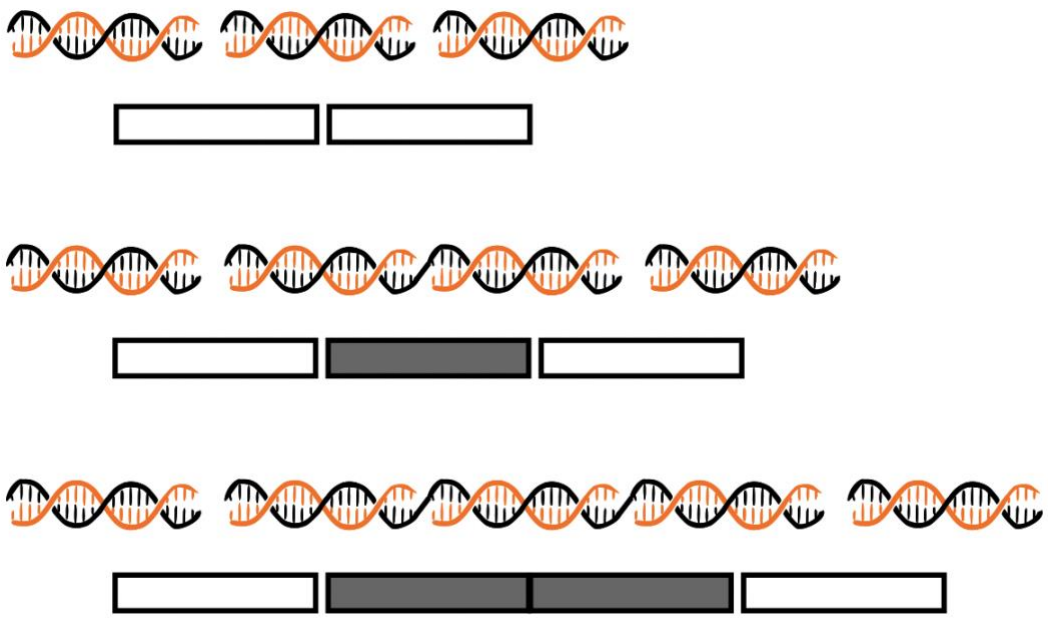
Mathematics:

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

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

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

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



To finish the derivation, we can use the properties of this geometric series.
Starting from the generic formula for $r < 1$:
$\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$

We can replace r with $(1 - P_{SSB})$:
$\sum_{i=0}^{\infty}(1 - P_{SSB})^i = \frac{1}{1-(1-P_{SSB})} = \frac{1}{P_{SSB}}$ *Equation 2*

The 1-mer, 2-mer, and higher order ligation product mole fractions should sum to 1:
$1 = \sum_{i=0}^{\infty} P_{i+1}$ *Equation 3*

Substituting Equation 1 into Equation 3:
$1 = \sum_{i=0}^{\infty} C \cdot {P_{SSB}}^2(1 - P_{SSB})^i = C \cdot {P_{SSB}}^2 \sum_{i=0}^{\infty}(1 - P_{SSB})^i$ *Equation 4*

Substituting Equation 2 into Equation 4 we can solve for the proportionality constant C:
$1 = C \cdot {P_{SSB}}^2 \left(\frac{1}{P_{SSB}}\right) = C \cdot P_{SSB} \qquad \therefore \qquad C = \frac{1}{P_{SSB}}$

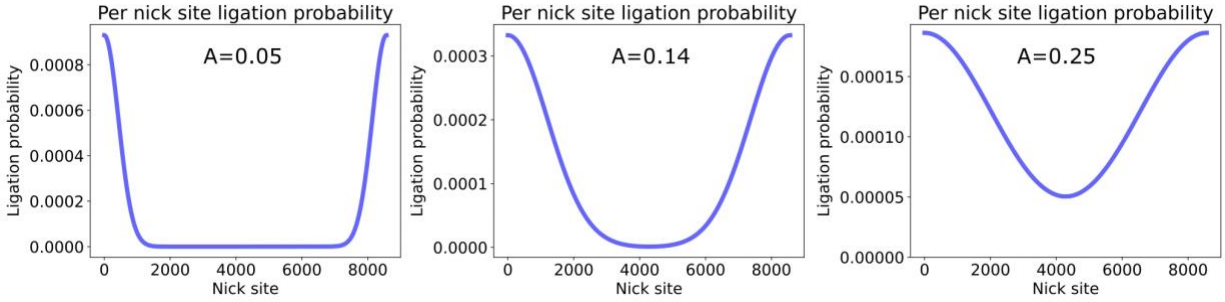
Therefore, we can simplify Equation 1:
$P_{i+1} = P_{SSB}(1 - P_{SSB})^i$ *Equation 5*

Equation 5 explains why the observed probability for observing 1-mers is equivalent to the single strand break probability:
$P_1 = P_{0+1} = P_{SSB}(1 - P_{SSB})^0 = P_{SSB}$

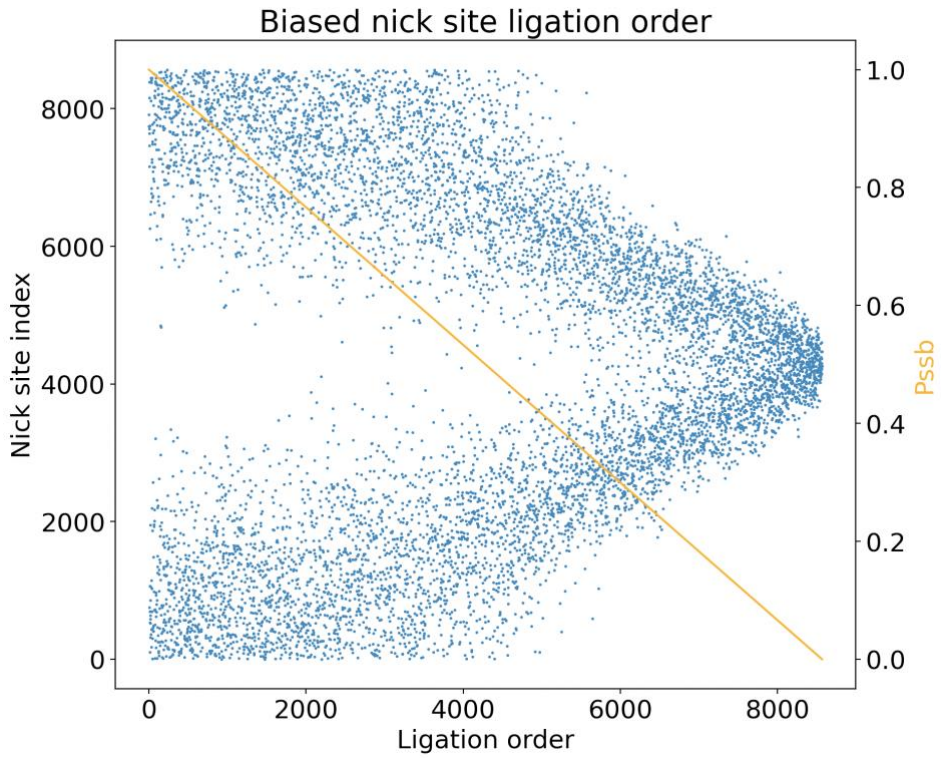
Notably, the random ligation model suggests that the 1-mer population fraction is equivalent to the single-strand-break probability, $P_1 = P_{SSB}$ and that a constant factor equivalent to $1 - P_{SSB}$ describes the population decrease with incrementing block size. Despite the approximations of the random ligation model, we can use these relationships to supply two additional estimates of the single-strand-break probability: $P'_{SSB} = P_1$ and, if κ is the best-fit value for the geometric decline in the population of higher-order ligation products with increasing block size, $P''_{SSB} = 1 - \kappa$. In the case of CC1, P'_{SSB} estimates were notably similar to the P_{SSB} estimates calculated from all the mole fractions (Table 2). In contrast, P'_{SSB} estimates for CC2 were systematically lower than P_{SSB}. P''_{SSB} estimates appear to be a less accurate estimation method because higher-order block populations do not cleanly decline with a reliable geometric factor either when normalized (Table 2) or prior to normalization (Table S2). Presumably these small mole fractions are sensitive to noise.

Protocol S3. Biased Ligation Model.

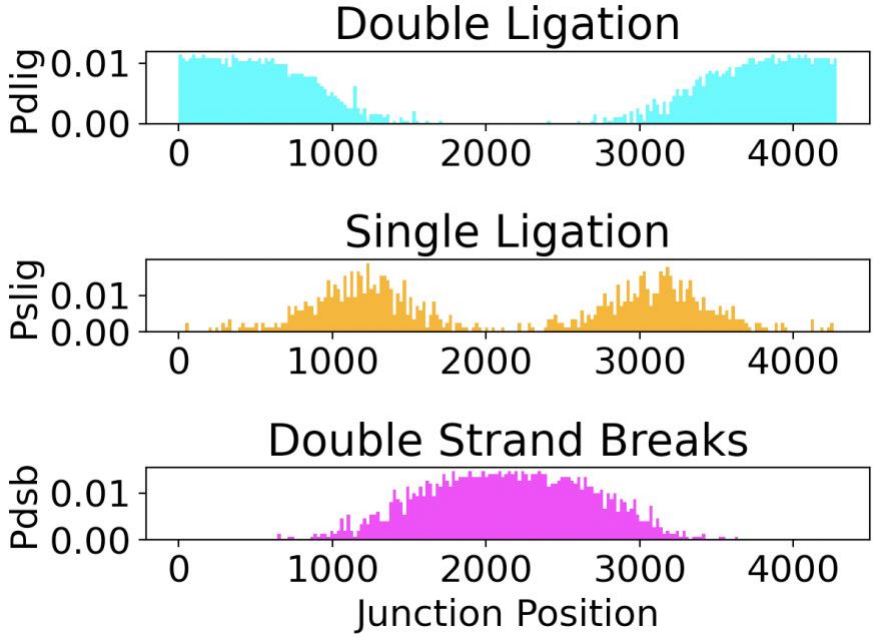
To determine whether transport limitations would be a reasonable explanation for the observed significant deviations from the RLM for CC2 crystals, we ran simulations of ligation where nick sites closer to the surface of the crystal were more likely to be ligated. We used a simple empirical model where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



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



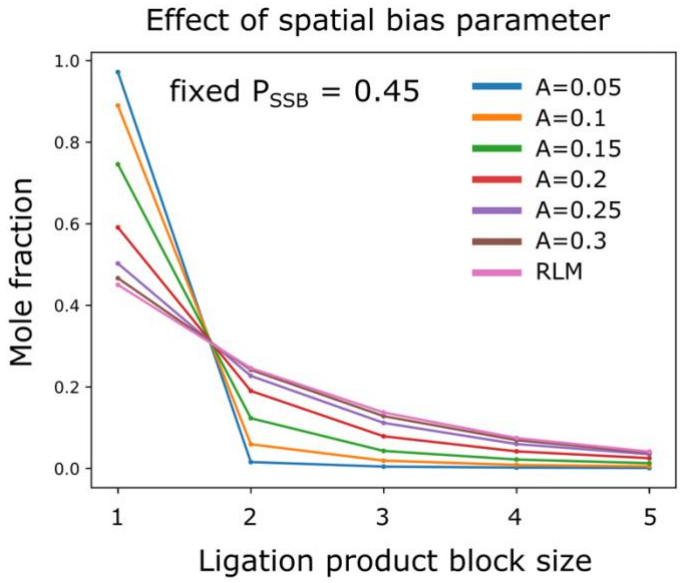
If we halt the ligation process when total P_{SSB} matches an experimentally derived target from the densitometry data (e.g. P_{SSB} = 0.45 for CC2-3'P High in Table 4.2), this model suggests a possible spatial distribution of doubly-ligated junctions (P_{DLIG}), singly-ligated junctions (P_{SLIG}),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



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

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

In the case of CC2-3'P under the strongest ligation conditions, the densitometry analysis (Table 4) suggested that $P_{SSB} = 0.45$, but the mole fraction for the 1-mer was estimated to be 74%, which is inconsistent with the RLM. The analysis here suggests that spatial bias in the ligation may resolve this discrepancy. The best fit biased ligation model was for A=0.14 (the spatial bias case illustrated above).



Protocol S4. Crystal measurements

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

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