Supplementary Information for

A DNA Crystal Designed to Contain Two Molecules per Asymmetric Unit

Tong Wang, Ruojie Sha, Jens Birktoft, Jianping Zheng, Chengde Mao, Nadrian C. Seeman*

Synthesis, purification and crystallization

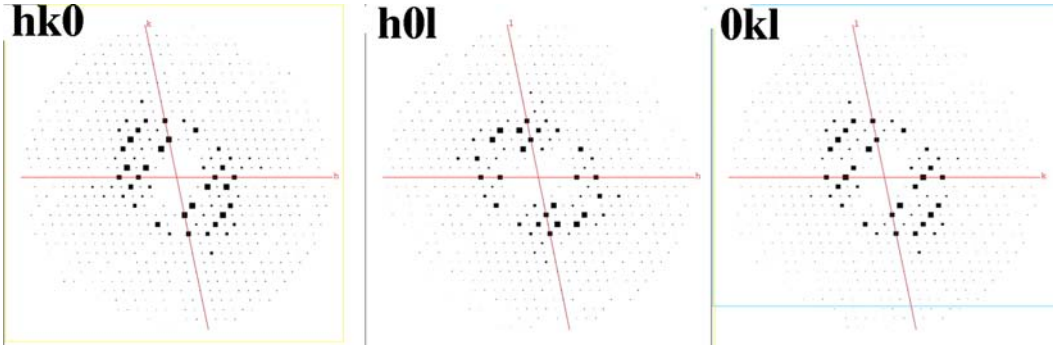
DNA sequences were designed using the program *SEQUIN*.[1] DNA strands, including Cy3 and Cy5 (Glen Research) derivatives, were synthesized by standard phosphoramidite techniques on an Applied Biosystems 394 DNA synthesizer. Strands were doubly purified by reverse-phase high-performance liquid chromatography using a C-18 column (Waters). Triangles A and B were separately annealed at 6 μM concentration in 1X TAE/Mg buffer by the established annealing procedure (5 min at 90 °C, 15 min at 65 °C, 15 min at 37 °C and 20 min at RT). 30 μL of each of the annealed Triangle A and B complexes were mixed with 40 μL buffer containing 125 mM magnesium acetate, 50 mM HEPES (pH 7.0) and 10% MPD to form a 100 μL sitting drop, equilibrated against a 0.5 ml reservoir of 1.4 M ammonium sulphate for two weeks, during which the volume of the drop diminished by about 90%. Native, Cy3 & Cy5-derivative crystals were all produced under identical conditions. In no instances were any macroscopic sized crystals without color observed in setups of A and B molecules derivatized with either Cy3 and/orCy5.

Data collection

Data Collection and Structure Solution: Crystals were flash frozen by quickly immersing them into liquid nitrogen. A complete sphere of X-ray diffraction data were collected at beamline X6A at the National Synchrotron Light Source, Brookhaven National Laboratory and were processed in space group R3 using HKL-2000.[2] Analysis by the Padilla-Yates method[3] indicates absence of twinning : $\langle|L|\rangle = 0.487$, $\langle L^*L\rangle = 0.315$. Molecular Replacement (MR) yielded an initial, experimental map and the major and minor groves of DNA could be resolved. The search model was the previously determined tensegrity structure (PDB code 3GBI). Nucleotides were mutated in *COOT*[4] to match those in molecules A and B, respectively. The resulting model from MR was subsequently refined against the data set including all observed data extending to 5.0 Å using the *PHENIX* program package.[5] Initially, the model was refined as one rigid body, using data in the full resolution range (35.0 -5.0 Å). Next, the eight DNA single strands were treated as individual rigid bodies. After two cycles of rigid body refinement, the R-free/R-work – factors were 0.268/0.291. Subsequently refinement using individual sites followed with separate atomic displacement parameter (ADP) refinement and with TLS (translation/libration/screw) rigid-body motion refinement[6] resulted in R-free/R-work-factors of 0.208/0.252. Weak restraints were placed on the hydrogen bonds in the presumed helical base pairs. Finally two cycles of simulated annealing resulted in R-free/R-work-factors of 0.182/0.239. The corresponding R.m.s.d. of angle and bonds are 1.221°/0.005Å The estimated coordinate error is 0.55Å, while the rmsd between equivalent main chain atoms in molecule A and molecule B is 0.954 Å. No pseudo-NCS restraints were incorporated at any stage during the refinement. It should be pointed out that at the resolution employed here (5.0Å) is not possible to distinguish between the different base pairs. The distinction between molecules A and B is very small, particularly at this resolution. . The two molecules each contain 63 base pairs, 30 of which are identical and another 18 pairs differ

conservatively with a purine:pyrimidine pair replacing another purine:pyrimidine pair. In the remaining 15 pairs a purine:pyrimidine pair is replaced with a purine:pyrimidine pair. A refinement identical to the one described above was performed with the A and B molecule (the BA-dimer) interchanged and R-free/Rwork was 0.205/0.251. The corresponding r.m.s.d. of angle and bonds are 1.240°/0.0056Å Å. These data suggest the correct ascription of identities to the two molecules.

The central zones support the notion that this is a pseudo-centered cell by the presence of alternating weak and strong layer lines.

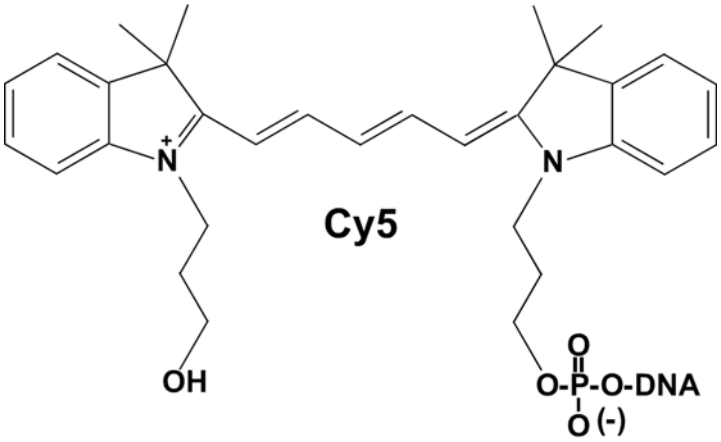


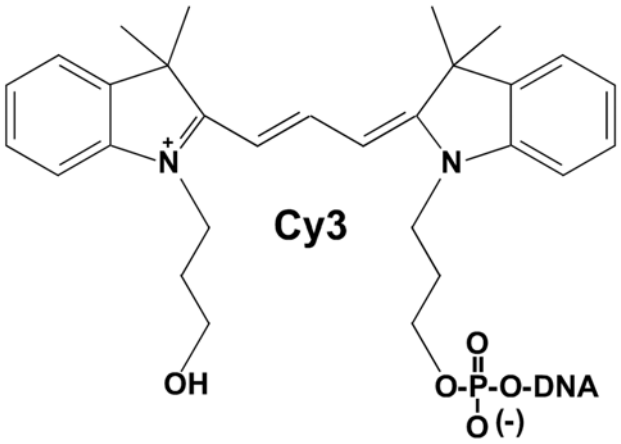
REFERENCES

1. Seeman, N.C., *De novo* design of sequences for nucleic acid structure engineering, *J. Biomol. Str. & Dyns.* **1990** *8*, 573-581.
2. Otwinowski, Z.; Minor, W., Processing of X-ray diffraction data collected in oscillation mode, *Methods Enzymol.* **1997** *276*, 307-326.
3. Padilla, J.; Yeates, T.O., A statistic for local intensity differences: robustness to anisotropy and pseudo-centereing and utility for detecting twinning. *Acta Cryst.* **2003** *D59*, 1124-1130.
4. Emsley P; Cowtan K., COOT: model-building tools for molecular graphics, *Acta Cryst.* **2004** *D60*, 2126-2132.
5. Adams P.D.; Grosse-Kunstleve R.W.; Hung L.-W.; Ioerger T.R.; McCoy A.J.; Moriarty N.W.; Read R.J.; Sacchettini J.C.; Sauter N.K.; Terwilliger T.C., PHENIX: building new software for automated crystallographic structure determination. *Acta Cryst.* **2002** *D58*, 1948-1954.
6. Painter J.; Merritt E. A., Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Cryst*. **2006** *D62*, 439-450.

STRUCTURAL FORMULAS

The structural formulas of the dyes Cy3 and Cy5 are shown below.


Cy5
OH
O
O-P-O-DNA
O (-)


Cy3
OH
O
O-P-O-DNA
O (-)