

LIQUID CRYSTALS FORMED BY SHORT DNA OLIGOMERS  
AND THE ORIGIN OF LIFE

By

GREGORY PATRICK SMITH

B.S., Colorado State University, 2000

M.S., Colorado State University, 2003

B.A., University of Colorado, 2008

M.S., University of Colorado, 2014

A thesis submitted to the  
Faculty of the Graduate School of the  
University of Colorado in partial fulfillment  
of the requirement for the degree of  
Doctor of Philosophy  
Department of Physics

2018

This thesis entitled:  
Liquid Crystals Formed By Short DNA Oligomers and The Origin Of Life  
written by Gregory Patrick Smith  
has been approved for the Department of Physics

---

(Noel A. Clark, Chair)

---

(Joseph E. Maclennan, Representative)

Date 4/3/2018

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Smith, Gregory Patrick (PhD, Physics)

Liquid Crystals Formed By Short DNA Oligomers and The Origin Of Life

Thesis directed by Professor Noel A. Clark

When dissolved in water, base paired DNA oligomers form double helices with sufficient structural rigidity that, if they are at high enough concentration, can undergo a phase transition into chiral nematic or hexagonal columnar liquid crystalline (LC) order. Within these LC phases, constrained orientation allows these rods to stack more efficiently by hydrophobic forces than they would otherwise, building them into long double helical aggregates that can be chemically glued together (ligated) to further increase their lengths. Even in absence of chemical ligation, this stacking effect is strong enough that short DNA oligomers, which are otherwise too short to form phases, can stack reversibly with one another into aggregates with sufficient length to force the creation of LC phases. If these stacked aggregates are then ligated within an LC phase, the lengthened rods become able to form LC phases at lower concentrations than they could have previously, given their improved aspect ratio, making it easier for them to form liquid crystals later. This effect forms a feedback loop where self-assembly of short oligomers into aggregates and chemical ligation of these aggregates within LC phases to form longer DNA double helices enhances later rounds of assembly and ligation, leading to the hypothesis that LC phases could have helped to provide a feedstock of long, complementary oligonucleotide strands as a basis for biology and helped to bootstrap the origin of Life. This thesis presents research exploring the limits of this effect, detailing examination and discovery of LC phases with shorter and more basic DNA oligomers and ending with the discovery of LC phases by base paired DNA monomers, which has never been previously seen.

## DEDICATIONS

I dedicate this work to my family, my parents and my brother and sister-in-law for being there, but especially to my wife Min and my little Sophie.

I would dwell in darkness without you.

## ACKNOWLEDGMENTS

This work was supported by NSF Biomaterials Grant DMR-1611272 and by the Soft Materials Research Center under NSF MRSEC Grant DMR-1420736. Advanced Light Source Beamline 7.3.3 is supported by U.S. DoE under contract No. DE-AC02-05CH11231.

Much of the methods described in Appendix A are non-original and are attributable to the like of Professor Marv Caruthers and to many others in the extensive industry that has arisen around the chemical synthesis of DNA, including the good people at GE Healthcare Life sciences, who engineered the Akta Oligopilot, and Thermofisher, who produced the nucleosidal phosphoramidites I frequently used in my syntheses, and Glen Research, whose website I often visited in an effort to glean details for new projects, as well as many others. Given the depth of the DNA synthesis field, should I misattribute any proper credit, I am very sorry. While this is not original research by me, the usage of these techniques was mastered over the course of seven years in the lab by myself, and my predecessors, Dr. Mark Moran and Professor Ethan Tsai. I would acknowledge Ken Hill at Agilent Technologies as well as Kurt Andersson and John Brucclieri at GE Healthcare Life Sciences for their help hammering out problems that plagued my early efforts at synthesis.

## CONTENTS

### SECTION

1.0 Introduction: Linking Origin of Life to DNA mesophases.....	1
2.0 Revisitation of Dodecamer nanoDNA .....	18
2.1 Updated Dodecamer Phase diagram .....	19
2.2 X-ray Diffraction studies of Reverse Dickerson Dodecamer .....	24
2.2.1 Chiral Nematic of Reverse Dickerson Dodecamer .....	29
2.2.2 Uniaxial Columnar Phase of Reverse Dickerson Dodecamer ...	41
2.2.3 WAXS of Higher Order Columnar Phases of Reverse Dickerson Dodecamer .....	64
2.3 Comparison of rDD Mesophases by Concentration .....	105
2.3.1 Theoretical Relationship Between $q_{hex}$ and Concentration ...	106
2.3.2 Comparison of rDD Hexagonal-Spacing to Theory .....	110
2.4 Instability in the rDD Phase Diagram .....	112
3.0 4mer nanoDNA Constructs .....	116
3.1 Blunt-End 4mer $5'GTAC^3'$ .....	117
3.1.1 $5'GTAC^3'$ LC Phase Textures .....	119
3.1.2 $5'GTAC^3'$ X-ray Diffraction .....	129
3.1.3 Comparison of $5'GTAC^3'$ Mesophases .....	146
3.2 Sticky-End 4mer $5'GCCG^3'$ .....	149

3.2.1	5'GCCG3' Optical Textures and Phase Diagram .....	151
3.2.2	5'GCCG3' X-ray Diffraction Studies .....	158
3.2.3	5'GCCG3' Mesophase Hysteresis .....	174
3.2.4	G-quartet Participation in 5'GCCG3' Mesophase Formation ...	182
4.0	Liquid Crystals Made of Nucleosidal Triphosphates .....	186
4.1	Additional Materials and Methods Used for Deoxynucleosidal Triphosphate (dNTP) Experimentation .....	192
4.2	Observations of dATP/dTTP .....	194
4.3	dNTP X-ray Data and Analysis .....	200
4.4	dNTP Phase Diagrams .....	206
4.5	Variation of dNTP Ratio in Solutions .....	210
4.5.1	COL-ISO Phase Transition Temperature Versus NTP Molar Ratio .....	210
4.5.2	Model of Entropy of Mixing in the COL-ISO Phase Transition .....	211
4.5.3	NTP Entropy of Mixing –Role of Base Pairing .....	213
4.6	dATP/dCTP/dGTP/dTTP COL-ISO Texture Observations .....	215
4.7	The Importance of dNTP Liquid Crystals to the Origin of Life .....	217
Appendix A:	Oligonucleotide Synthesis and Purification .....	220
A.1	Overview .....	220
A.2	DNA Synthesis Chemistry .....	222
A.3	DNA Synthesis Automation .....	230
A.3.1	Solid Support .....	235

A.4	Synthesis Reagents .....	237
A.4.1	Preparation of Activated Molecular Sieves .....	238
A.4.2	Preparation of Capping Solutions .....	241
A.4.3	Preparation of Oxidizer Solution .....	242
A.4.4	Preparation of Deblocking Solution .....	244
A.4.5	Preparation of DEA Wash Solution .....	246
A.4.6	Preparation of Activator Solution .....	248
A.4.7	Preparation of dN-PA Solutions .....	249
A.4.8	An Example Planning Calculation for a Conventional Synthesis Program .....	251
A.5	Synthesis Process .....	253
A.5.1	Overview of Unicorn .....	254
A.5.2	Preparing the Oligopilot One Week Ahead of Synthesis .....	255
A.5.3	Preparation on the Morning of Synthesis and Use of the Method Editor .....	261
A.5.4	Running Automated Synthesis on the Oligopilot Synthesizer	265
A.5.5	Cleaning the Oligopilot .....	269
A.6	Purification Methods .....	271
A.6.1	Post Synthesis Processing .....	272
A.6.2	Desalting by Isopropanol Precipitation .....	274
A.6.3.1	High Performance Liquid Chromatography (HPLC) of LC Forming Oligonucleotides (DMT-off Method) .....	280
A.6.3.2	High Performance Liquid Chromatography (HPLC) of LC Forming Oligonucleotides (DMT-off Method) .....	285

A.6.4	PEG Crashing of Self-Complementary Oligomers .....	289
A.7	Characterization Methods .....	291
A.7.1	Mass Spectrometry of Synthetic Oligomers .....	291
A.7.2	UV Absorbance of Oligonucleotides .....	298
A.8	Specific Projects .....	303
A.8.1	Adding a Phosphate at the 3'-Terminal .....	303
A.8.2	Synthesis of Randomer Oligonucleotides .....	306
A.8.3	Usage of Universal Linker Solid Support .....	308
A.8.4	5'-Terminal Triphosphate Modification .....	309
A.9	RNA Oligonucleotide Synthesis .....	312
Appendix B:	Generalized nanoDNA LC Preparation Methods .....	315
B.1	nanoDNA LC Cells .....	319
B.1.1	Evaporative Cells .....	320
B.1.2	Oil Sealed Cells .....	321
B.1.3	Double Pane Oil Sealed Cells .....	324
B.1.4	Flame-Sealed Capillaries .....	326
B.1.5	Oil Sealed Free Surface Cells .....	331
B.2	Concentration Calculations .....	336
B.2.1	Legacy Concentration Calculations .....	337
B.2.2	Higher Precision Concentration Calculations .....	339
Appendix C:	X-ray Diffraction Theory and Methods .....	344

C.1	Useful Basic Theory of X-ray Diffraction in Liquid Crystals .....	344
C.2	Analysis of X-ray Scattering Peak Shapes .....	358
C.3	Experimental Methods for nanoDNA X-ray Diffraction .....	361
C.4.1	Analysis of Two Dimensional Hexagonal Lattices .....	362
C.4.2	Analysis of Flattened Hexagonal Lattices .....	364
C.4.3	Sheared Hexagonal Lattice .....	369
C.4.4	Face Centered Cubic Lattice .....	373
	REFERENCES.....	378

## FIGURES

- 1.1: Model of B-form DNA double helix...8
- 1.2: Onsager criterion for cylindrical systems...9
- 1.3: Onsager line...10
- 1.4: DNA LC phases with increasing concentration...10
- 2.1: Phase diagram of nanoDNA of lengths 6 -20 bp...19
- 2.2: Combined phase diagram of rDD...21
- 2.3: Dodecamer mesophase textures...22
- 2.4: Collected rDD WAXS from room temperature...25
- 2.5: Collected WAXS of rDD, base pairing region...26
- 2.6: Collected WAXS of rDD in small angle...27
- 2.7: WAXS of NEM\* phase of rDD...31
- 2.8: WAXS of 274 mg/mL rDD...32
- 2.9: Collected scans during temperature ramp experiment of 274 mg/mL rDD...36
- 2.10: Wide angle region, collected scans of 274 mg/mL rDD...37
- 2.11: Peak fitting versus temperature of rDD NEM\* mesophase...38
- 2.12: NEM\* spacing and structural coherence versus temperature...39
- 2.13: rDD COL mesophase typical WAXS...43
- 2.14: Collected rDD COL mesophase samples in WAXS...44
- 2.15: Scattering feature used for Hexagonal confirmation in COL phase...45
- 2.16: WAXS scattering of rDD COL+ columnar phase...47
- 2.17: rDD 373.6 mg/mL temperature ramping experiment...48
- 2.18: Peak fitting parameters for rDD 373.6 mg/mL during temperature ramp...49
- 2.19: Lattice Spacing and Coherence vs. Temperature for rDD 373.6 mg/mL...50
- 2.20: rDD 545.6 mg/mL temperature ramping experiment...52
- 2.21: Peak fitting parameters for rDD 545.6 mg/mL during temperature ramp...53
- 2.22: Lattice Spacing and Coherence vs. Temperature for rDD 545.6 mg/mL...54
- 2.23: rDD 627.1 mg/mL temperature ramping experiment...55
- 2.24: Peak fitting parameters for rDD 627.1 mg/mL during temperature ramp...56
- 2.25: Lattice Spacing and Coherence vs. Temperature for rDD 627.1 mg/mL...57
- 2.26: Comparison of COL phase Lattice spacing v. Temperature...58
- 2.27: Structural Coherence Length v. Temperature Comparison...60
- 2.28: Separation length in COL melting phase versus temperature...61
- 2.29: Coherence length in melting phase versus temperature...62
- 2.30: rDD COL lattice spacing versus concentration at room temperature...63
- 2.31: WAXS of COL2 from 781.2 mg/mL rDD...67
- 2.32: Collected circular averages of COL2 samples at room temperature...68
- 2.33: Temperature ramp experiment of rDD 764.5 mg/mL...69
- 2.34: WAXS of mesophase packing peak of rDD 764.5 mg/mL at 82.5° C...70
- 2.35: Peak position and FWHM versus Temperature for rDD 764.5 mg/mL...71
- 2.36: rDD base-stacking COL v. COL2...72
- 2.37: Flattened lattice degeneracy problem...75

- 2.38: Annotated peaks of room temp rDD 781.2 mg/mL...76
- 2.39: Central peak of COL2 mesophase fitted to Lorentzians...78
- 2.40: Temperature ramp of rDD 781.2 mg/mL...79-80
- 2.41: Peak center and FWHM v. temperature for rDD 781.2 mg/mL...81
- 2.42: Sheared Hexagonal Lattice...82
- 2.43: Comparison of Unit Cell Area for each lattice model...84
- 2.44: Lattice parameters of the Case #1 FHL...85
- 2.45: rDD 764.5 mg/mL Lattice Spacing versus Temperature...86
- 2.46: rDD 781.2 mg/mL Lattice Spacing versus Temperature...87
- 2.47: Structural Coherence Length versus Temperature for rDD 764.5 mg/mL...88
- 2.48: Structural Coherence Length versus Temperature for rDD 781.2 mg/mL...89
- 2.49: Unit cell scale models...93
- 2.50: Unit Cell Area variation...94
- 2.51: 781.2 mg/mL rDD Case 1 FHL scale model...95
- 2.52: Duplex Footprint Overlap...96
- 2.53: Simplified DNA helix diagram...98
- 2.54: Collected COL and COL2 lattices drawn in common scale...102
- 2.55:  $q_{hex}^2$  versus Concentration...110
- 2.56: Melting temperature versus Concentration...113
- 3.1: GTAC concentration phase diagram at constant temperature...120
- 3.2: Cholesteric fingerprint diagram...122
- 3.3: Metastable COL texture versus time...124
- 3.4: GTAC mosaic texture at ~560 mg/mL and 5° C...125
- 3.5: GTAC room temperature mosaic texture with ~50  $\mu$ m path length...127
- 3.6: NEM\* in 400 mg/mL GTAC with temperature ramp...128
- 3.7: GTAC parabolic focal conic texture...129
- 3.8: GTAC WAXS, circular average, full range...130
- 3.9: GTAC WAXS, circular average, small angle region...131
- 3.10: GTAC WAXS, circular average, small angle region at room temperature...132
- 3.11: GTAC circular average, small angle region...134
- 3.12: Room temperature 2D WAXS of 517 mg/mL GTAC...135
- 3.13: GTAC COL1/2 phase seen at room temperature...136
- 3.14: GTAC COL1/2 rapid annealing...138
- 3.15: GTAC COL2 at room temperature...143
- 3.16: GTAC COL2 prior to temperature cycle...144
- 3.17: GTAC high concentration 3D crystal...145
- 3.18: GTAC Phase Diagram...146
- 3.19: GTAC concentration versus  $q_{hex}$  compared to theory...148
- 3.20: Comparison of Blunt and Sticky assembly routes...150
- 3.21: G-Quartet Mesophases...151
- 3.22: Reproduction of GCCG-P results...152
- 3.23: Out of concentration sequence phases of GCCG...154
- 3.24: Block phase diagram for GCCG...155
- 3.25: Unusual textures of GCCG mesophases...158

3.26:	WAXS of 450 mg/mL GCCG at room temperature...	159
3.27:	GCCG room temperature WAXS before temperature cycle (Full)...	160
3.28:	GCCG room temperature WAXS before temperature cycle (SAXS only)...	161
3.29:	GCCG 419 mg/mL NEM* WAXS with temperature ramp...	164
3.30:	GCCG 450 mg/mL WAXS with temperature ramp...	165
3.31:	GCCG 641 mg/mL temperature ramp...	167
3.32:	GCCG 789 mg/mL ISO2 phase...	168
3.33:	Annotation of Phase 1 crystal peaks...	169
3.34:	GCCG Phase 1 FCC crystal lattice. All lengths are to scale...	171
3.35:	G-quartet mediated GCCG lattice...	173
3.36:	Hypothetical GCCG G-quartet mediated square Lattice...	173
3.37:	Temp cycle profile for hysteresis experiment with GCCG 450 mg/mL...	175
3.38:	Hysteresis Experiment part 1...	176
3.39:	Hysteresis Experiment part 2...	177
3.40:	Hysteresis Experiment part 3...	178
3.41:	Hysteresis Experiment part 4...	179
3.42:	GCCG 450 mg/mL phase diagram depending on hysteresis...	181
3.43:	Phases of GCCG-Li+...	184
3.44:	Mixed mode aggregate assembly...	185
4.1:	dATP and dTTP Molecular models...	188
4.2:	First observation of dATP/dTTP LC phases...	189
4.3:	dATP-dTTP compensator images...	190
4.4:	Interpretation of birefringence compensation...	191
4.5:	Development of dATP/dTTP mixture in free-surface cell...	195
4.6:	Chemical structures of dATP and dTTP...	196
4.7:	Flame Sealed Capillary cell containing dATP/dTTP (1:1) mixture...	198
4.8:	ACGT Matrix summarizing the search for LC phases...	199
4.9:	X-ray structure factors for dATP-dTTP...	203
4.10:	dATP-dTTP Plots of $q_h^2$ versus $c$ ...	205
4.11:	dATP/dTTP mixture phase diagram in $T$ versus $c$ ...	206
4.12:	dGTP/CTP mixture phase diagram in $T$ versus $c$ ...	207
4.13:	Compilation of relevant DNA liquid crystal phase diagrams...	209
4.14:	Gibbs free energies for a first-order ISO-COL phase transition...	212
4.15:	ISO-COL transition temp shift resulting from free energy change...	212
4.16:	The shift of transition temp for COL LC in NTP binary mixture...	213
4.17:	Observed shift of the transition temp of 840 mg/ml dATP/dTTP...	214
4.18:	Birefringent textures of 4-component dATP/dCTP/dGTP/dTTP mixture...	216
A.1:	Oligonucleotide chemical models...	221
A.2:	Schematic overview of oligonucleotide synthesis process...	223
A.3:	Adenine Deoxynucleoside phosphoramidite (dA-PA) structure...	224
A.4:	The main synthetic cycle...	225
A.5:	Capping reaction...	226
A.6:	Activation of dA-PA phosphoramidite during the base coupling step...	227
A.7:	Deprotection reactions to produce DNA after synthesis...	228

A.8:	Protected DNA bases...	230
A.9:	Äkta Oligopilot 100...	232
A.10:	Simplified schematic of DNA synthesizer plumbing...	233
A.11:	Oligopilot reactor columns...	234
A.12:	Molecular sieves...	239
A.13:	Molecular sieves in a solvent...	240
A.14:	dN-PA preparation and transfer technique...	251
A.15:	System Control Screen...	256
A.16:	Purge values on the pump heads...	258
A.17:	Method editor...	263
A.18:	HPLC trace of 16mer nanoDNA...	283
A.19:	DMT-On Purification strategy...	286
A.20:	MALDI-TOF on RNA purification fractions...	298
A.21:	Example of Spectrophotometry on rDD...	301
A.22:	Carbodiimide ligation by EDC...	304
A.23:	Triphosphate addition at 5'-terminal of DNA...	310
A.24:	RNA phosphoramidite...	312
B.1:	Labeled optical slow axis diagram...	317
B.2:	Polarized light microscope diagram...	318
B.3:	Michel-Levy Chart...	319
B.4:	Evaporative cell...	320
B.5:	Oil sealed cell...	321
B.6:	Double pane oil sealed cell...	325
B.7:	Index matching device for capillaries...	331
B.8:	Preparation and development of oil sealed free surface cell...	335
C.1:	X-ray wave scattered from two scatterers...	344

## TABLES

2.1:	Comparison Case 1 and Case 2 flattened hexagonal lattice...76
2.2:	Low angle peak and harmonics...77
2.3:	Sheared hexagonal lattice parameters...83
2.4:	COL2 flattened hexagonal lattice...99-100
3.1:	COL1/2 peak indexing...139
3.2:	COL1/2 Case 2 FHL lattice characteristics...140
3.3:	Alternative GTAC lattice assignments...141-142
3.4:	GCCG face centered cubic lattice indexing...170
A.1:	Cap A recipe...242
A.2:	Cap B recipe...242
A.3:	Oxidizer solution recipe...243
A.4:	Deblocking solution recipe...246
A.5:	DEA wash solution recipe...247
A.6:	Activator solution recipe...249
A.7:	DNA-phosphoramidite solution recipes...250
A.8:	Triethylammonium acetate (TEAA) buffer recipe...281
A.9:	DMT removal solution (for DMT-on method)...287
A.10:	MALDI matrix ammonium citrate stock recipe...293
A.11:	MALDI-TOF matrix recipe (HPA type)...293
C.1:	Face Centered Cubic lattice expected reflections...375

## Section1.0

### Introduction: Linking Origin of Life to DNA mesophases

There are few natural systems as independently complex and simultaneously baffling as Life. It was not many years ago that Life was considered so singular that it had to have arisen solely from a divine hand, and even today, it remains attributed as a product of divinity by many people, regardless of what new insights science has offered. The central tenet of modern biology is Darwinian Evolution, which tells how living species came to be as a result of the drifting genetic adaptation of organisms to inhabit steadily changing living niches available in the biosphere due to their selection by some criterion of fitness<sup>1</sup>.

For all of the great insights provided by Darwinian evolution, the theory and its aggregates only offer a route to explain how one form arose from another<sup>1</sup>. The theory fundamentally depends on the clay of a preexisting organism in order to render the entire library of species. As heritability requires the intact unit of the cellular genetic apparatus, one of the remaining great mysteries of Life is how the “original seed” of life, the Last Universal Common Ancestor (LUCA), came to be. Some cellular organism, called LUCA, was required as a precondition for Darwinian evolution to craft all known life, though whether this was a singular organism or

connected parallels is unknown<sup>2</sup>. Figuring out how this precursor organism came to be is the purview of an arena of study now called Abiogenesis<sup>2</sup>.

It is scientifically accepted that Abiogenesis is some physical process by which the prevailing environmental conditions on or around Earth conspired with natural physical processes to give rise to the LUCA<sup>3</sup>. The study of Abiogenesis has been perhaps one of the most challenging scientific endeavors humans have ever tackled: we understand that life can occur, but the lack of evidence about how it occurred makes it profoundly difficult to study. Fossil evidence suggests that life is 3.8 billion years old, attaining a cellular form so very long ago that the simple passage of time has destroyed most evidence of the process that gave rise to it<sup>4</sup>. Further, modern life has great interdependence<sup>1</sup>, where life fuels itself from the products and remains of other life, assuring that viable relics of life do not survive long in the environment. What parts of life that do get fixed into the fossil record are inevitably chemically altered from their original viable forms to such an extent that this evidence becomes able to survive the passage of the eons, thus assuring that we know what Life left behind, while simultaneously destroying any chance of knowing with certainty how true primordial life operated on a chemical level. The existence of life is well understood to have so chemically altered Earth, including giving the planet its reactive oxygen atmosphere 2 billion years ago<sup>5</sup>, that the conditions which produced life originally may not exist anywhere on the planet anymore.

As it is, we will likely never know exactly how life arose on Earth, except for the possibility of observing life originate somewhere else.

The elucidation of the chemical architecture of life has fleshed out at least some of the players. Life is bound within and dependent upon membranes of lipids, using these structures to support the creation of chemical potential gradients and to

help protect essential molecules from spurious degradation when exposed to the environment<sup>6</sup>. The polymer DNA in a cell stores genetic information in order to convey it to offspring cells<sup>6</sup>. Polymer RNA mobilizes genetic information from DNA and catalytically operates a translation apparatus to convert that information into protein catalysts<sup>6</sup>. Protein catalysts support the polymerization of DNA and drive all the metabolic chemistry to generate the precursor feedstock for the fabrication of all biological polymers<sup>6,7</sup>. Modern biological metabolisms derive chemical energy from an array of protein mediated reactions converting chemical species among a vast network of sugars, lipids and other small molecules<sup>6</sup>. Membranes in eukaryotes combine with specialized proteins to create organelle structures, like chloroplasts or mitochondria, which couple the production of chemical potentials in the form of gradients across membranes to absorption of light energy and to sugar chemistry<sup>6</sup>. Proteins pair with membranes in both eukaryotes and prokaryotes to form structures like flagella, cilia or pseudopods to facilitate cellular motility or specialized cytoskeletal structures in multicellular organisms<sup>6</sup>. All of these systems link together into the unit of the cell to play some role in cell division, creating the foundation of biological evolution by virtue of the simple fact that DNA genetic information storage and self-replication is “good enough” to produce viable daughter cells during division, but just imperfect enough that those daughters contain hereditary mistakes compared to the parent.

LUCA likely contained some unified parcel of membranes, proteins and nucleic acids<sup>4</sup>. The interdependence of the different systems seems to form a house of cards that makes it difficult in modern life to remove any one process without collapsing the rest. Without proteins and RNA, replication of DNA chromatin would be impossible, while without DNA and RNA, production of protein would be impossible. Membrane structures are understood to be spontaneous as a result of

entropy<sup>6</sup>, but they have significantly less capacity for chemistry without DNA, RNA and protein.

One idea favoring a pre-modern form for LUCA is the notion of the “RNA world<sup>4,8,9</sup>.” Of all the systems present in the cell, only RNA has shown simultaneous capacity to be both information storage, in the form of mRNA, and catalytic effector, such as the transpeptidase activity of the Ribosome<sup>6</sup>. DNA can be regarded as a close cousin of RNA, capable of most of the same chemistry, but subtly modified to extend stability<sup>6</sup>. Since DNA replication depends on an RNA primer for its initiation<sup>7</sup> and since RNA has been found in the chromosomes of certain viruses instead of DNA suggests RNA could have been first<sup>10</sup>. Further, protein production is dependent on RNA catalysts: mRNA, tRNA and ribosomes. The genetic code that regulates translation of DNA gene transcripts into protein effectors is implemented in RNA machinery that is so essential to modern cellular behavior that it persists today across all existing species. It is not too bold to say that modern life began when the ribosome translation system was achieved, though admittedly ribosome sequence can be dissected further to determine some order of assembly<sup>11,12</sup>. The “RNA world” postulates that a time existed when RNA ran all necessary biological functions, catalytic effector and information storage, sufficient to produce the ribosome system which ultimately defined the LUCA<sup>8</sup>.

The conditions leading up to the RNA world are still mysterious. Moreover, while it is clear that some unique conditions existed to give rise to the RNA translation system that ties life together, it is unclear what the exact nature of this chemical environment happened to be<sup>9</sup>. It is widely accepted that water is necessary for life and that a hydrosphere helped in life’s origination<sup>6</sup>. Research on the spontaneous appearance of biological molecules by seminal workers like Orgel<sup>13</sup> suggests that the pieces necessary to give rise to life, peptides and nucleotides,

could all have originated from the environment as a result of natural interactions between physical processes, light, temperature fluctuation, violent impact and simple aging. It is generally accepted that Life represents some subset of a broader chemical library that is now no longer accessed on Earth, opening up the possibility of unknown progenitor polymers that preceded RNA and DNA<sup>14</sup>. Despite the plausibility of this idea no candidate progenitor polymers are definitively understood as natural and no non-biological “extended” chemistries<sup>15</sup> can be absolutely ascribed to conditions present early Earth. While the polynucleotides are considered the more primordial polymer, proteins come from simpler building blocks, though proteins have no direct role in inheritance<sup>6</sup>. It has been argued that the nucleosides are fine-tuned by modern life and that the modern precursor of polynucleotides, the nucleosidal triphosphate, was unfit for prebiotic environments, and yet they exist now and life uses them<sup>16</sup>. Critical uncertainty exists in the natural source of the selected biomolecule chirality during the origination of life, though mechanisms have been proposed<sup>17,18</sup>. The simpler monomer units of peptides are found in meteorites<sup>19</sup>, suggesting that some essential chemicals could have come into the biosphere from extraplanetary sources and that chirality of these molecules could be enriched by cosmic mechanisms<sup>17</sup>. The dramatic spontaneity of membranes formed from lipid-like compounds shows that membranes can be readily achieved without life, though lipids used by life are considered fine-tuned molecules in their modern forms. It seems likely that autofeedback of cells selecting for more viable forms has resulted in alterations to the primordial systems, obscuring at least some of what existed earlier.

Is there a possibility that LUCA is merely one ortholog of life, where other variants existed marking different selections of chirality or organizing paradigms, and that LUCA was the only one to survive some extinction event? It is possible

that the path between origination and the LUCA of modern life is made so curvy by unknown selection and extinction that no linear path can be extracted through what we regard as the “RNA world” to justify any of the modern aspects of biology as truly primordial. For lack of being able to approach these more complicated issues, and given that they may exceed the intended scope, this thesis will take the Occam’s Razor approach and consider that the origination of life in the RNA world is best hypothesized as a “warm little pond” where a minimum of the necessary precursors subsisted under weathering conditions to give rise ultimately to a self-replicating system as a progenitor cell.

Forces of self-organization are already well understood for their contributions to the functioning of modern life<sup>6</sup>. Lamellar structures of lipids in aqueous solutions occur by spontaneous phase separation, forming membrane leaflets by hiding their hydrophobic portions from water<sup>6</sup>. Similar interactions play a role in the folding of proteins into enzymes, with hydrophilic and hydrophobic side-groups of a polypeptide seeking appropriate environments and driving the polymer to search through an ensemble of folded configurations until it reaches a free energy minimum in a functional enzymatic form<sup>6</sup>. DNA and RNA polymers base pair spontaneously, searching randomly through best-fit stem loop structures until they reach ordered double helices<sup>6</sup>. It would be easy to imagine that an entry route into the RNA world would also depend on this sort of self-assembly behavior.

The study of self-assembly and self-ordering is well represented in soft condensed matter by the study of liquid crystals (LC)<sup>20</sup>. Lamellar leaflets of lipid are the largest and most obvious LC structure known to biology: literally a two-dimensional fluid with uniform orientational order and only one degree of positional order across the layer’s thickness. In another example of a biological LC, Rosalind Franklin’s pulled DNA fibers created a hydrated solid with DNA strands out of

register along the axis of the fiber, but in a hexagonal array across the fiber's cross-section<sup>21</sup>. During X-ray diffraction, this assembly permitted vivid appearance of the duplex's helical structure factor without distortion from crystal reflections and allowed Watson and Crick to propose the double helical model that won them a Nobel prize<sup>22</sup> (Figure 1.1). Similar two-dimensional DNA lattices are observed in chromatin packing of virus particles<sup>23</sup> and are likely the most efficient thermodynamic outcome of cramming double helical DNA into the smallest possible space. LCs of macromolecular DNA polymers have been extensively studied for decades<sup>24</sup>.

Liquid crystalline forms of polynucleic acid, RNA and DNA both, seem a directly compatible outcome of the "warm little pond" model. Polynucleotides arising in primordial conditions dissolved in water almost certainly would have faced conditions of wet and dry cycling, which would result in spontaneous formation of liquid crystalline ordering at typical ambient temperature and pressure conditions. This mode of self-organization is accessible within conceivable biological ranges as a free-standing molecule in the absence of other biology. Since lamellar membrane envelopes are essential to life, do LC forms of polynucleotides also offer some advantages which might facilitate the origin of life?

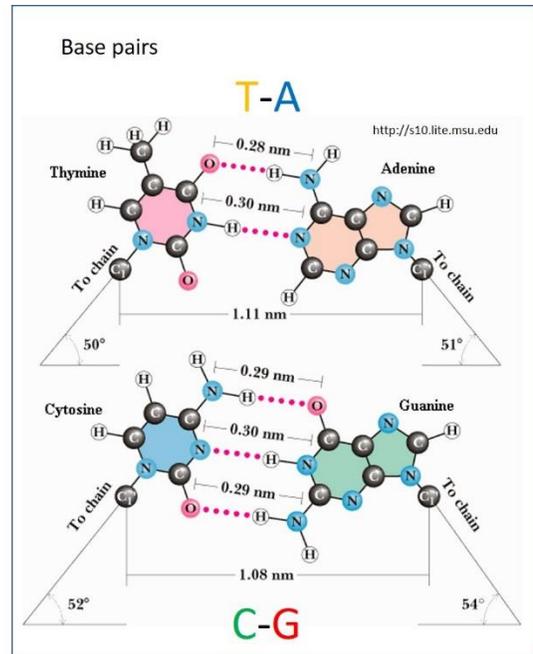
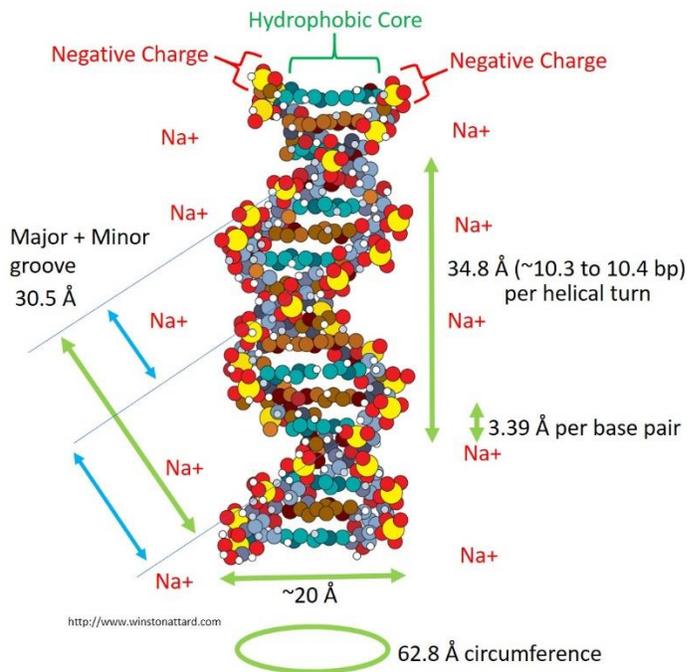


Figure 1.1: Model of B-form DNA double helix with size parameters and base pairs shown.

LC phases formed by polynucleic acids are a consequence of excluded volume interactions resulting within a collection of elongated bodies<sup>20</sup>. The double-stranded DNA duplex has a bend persistence length of 50 nm and a B-form duplex diameter of about 2 nm (Figure 1.1), giving an aspect ratio of at least 25 for long, double-stranded polymers<sup>24</sup>. For duplexed DNA dissolved in water at sufficiently low concentration, the rod-like molecule can sample any tumbling orientation; if water is gradually removed, the tumbling volumes sampled by individual DNA duplexes will tend to clash, forming a difference in distance of closest approach from neighboring molecules depending on how those molecules are oriented with respect to each other. If molecules bump end-to-side, their centers of mass are forced to remain widely separated, reducing the number of possible positions in space that both DNA duplexes might occupy. But, if the duplexes are co-oriented with parallel long axes, neighboring molecules are less obstructive to the occupation of many positions in space (add Onsager criterion here). This can be summarized by the

work of Lars Onsager in the form of the Onsager criterion for cylindrical systems (Figure 1.2)<sup>20</sup>.

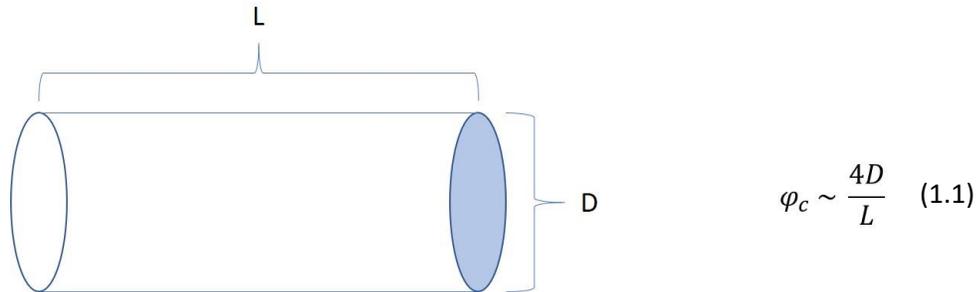


Figure 1.2: Onsager criterion for cylindrical systems.

Onsager criterion relates the approximate volume fraction  $\varphi_c$  (portion of occupied volume per total volume available, equivalent to concentration) describing the critical cut-off for formation of nematic order instead of isotropic order to the aspect ratio,  $L/D$  length per diameter, of the object being ordered, here a rigid cylinder. This relationship predicts a basic phase diagram for transitioning from disorder to order relative to increasing concentration (Figure 1.3)<sup>20</sup>.

For gradual increase in concentration of an aqueous solution of DNA duplexes, the mixture will undergo a phase transition to a nematic (co-oriented LC phase, NEM) at a volume fraction dependent on the duplex length –the loss of orientational entropy is offset by a gain of translational entropy. As concentration is further increased, neighboring duplexes eventually freeze into a tightly packed two-dimensional hexagonal lattice called a “columnar” phase because the DNA duplexes stack their long axes into columns along the same lattice points while remaining fluid in the third dimension. Polynucleotide duplexes can access more complicated variations of columnar phase at higher concentrations still, including an orthorhombic columnar phase and fully crystalline phases (Figure 1.4).

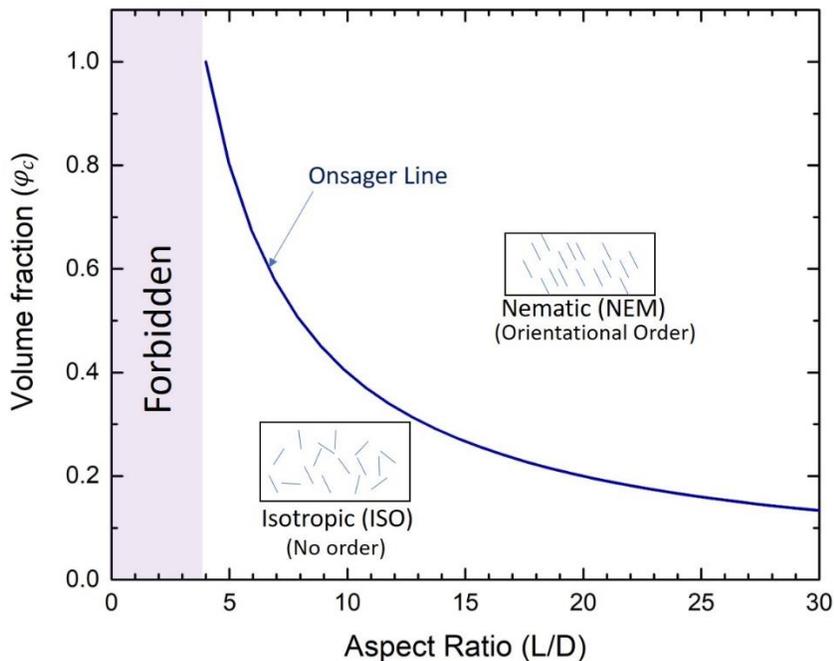


Figure 1.3: Onsager line. Volume fraction versus aspect ratio showing the cut-off concentration for establishment of nematic order. Forbidden region is where  $\varphi_c > 1$  is required for a phase transition, which is impossible.

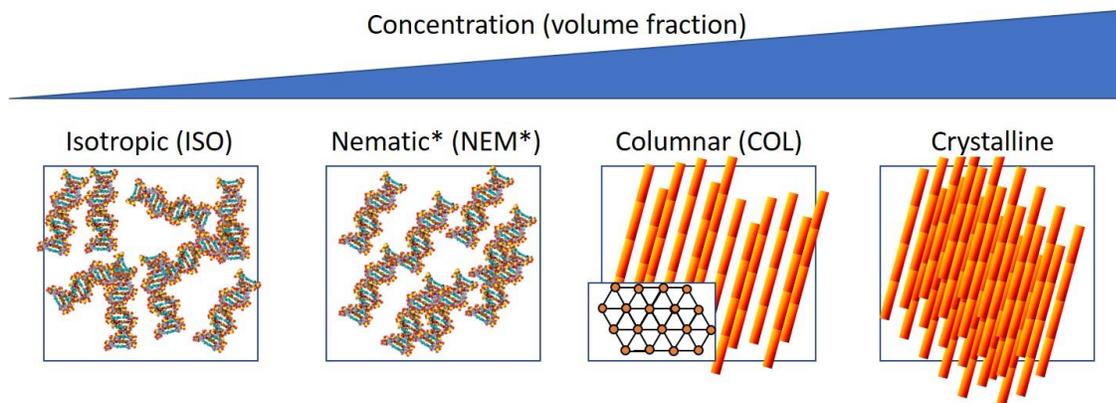


Figure 1.4: DNA LC phases with increasing concentration.

One possible immediate benefit of this sort of LC to an abiogenesis context is the potential for chemical self-protection. For very dry polynucleotide LC phases, very large molecules are packed into a remarkably restricted space, offering a

reversible monolithic domain that can occlude larger adversely reacting environmental chemicals from contacting deeply buried DNA duplexes. This advantage is shared by the formation of membrane leaflets, helping to protect the contents of a cell from the environment<sup>6</sup>. The permeability of a polynucleotide LC phase decreases with the increasing concentration of the LC phase where columnar (and higher order columnar) structures would be expected to be the most highly protected by virtue of being the most tightly packed during a period of dehydration in the environment.

Nematic LC phases of polynucleotide also exhibit an effect based on the chirality of the duplex (chiral nematic denoted by an asterix: NEM\*)<sup>24</sup>. Chiral helices prefer to co-orient with a slight skew so that the ridges of one helix fit the grooves of its neighbor, much like the packing of chiral objects like drill bits. Within a bulk material phase, this tendency to skew propagates across the width of an LC domain as a helical screw axis in the local orientation director field. This LC phase is known as a cholesteric<sup>20</sup>. Right-handed polynucleotide duplexes tend to give rise to cholesterics with left-handed twist<sup>24</sup>. While this might be expected to help generate a molecular chirality sorting mechanism, which is of obvious importance to the field of abiogenesis, DNA duplexes of opposing chirality are now known to share NEM\* phases, where mixed chirality is seen to result in unwinding of the cholesteric helix<sup>18</sup>.

On the basis of pure duplex aspect ratio (figure 1.2), theory predicts that the polynucleotide concentration which can result in an ISO to NEM\* phase transition is dependent on length of the duplex. The longer the duplex, where the aspect ratio is higher, the lower the concentration that duplex will tend to need in order to form a co-oriented phase. Very long DNA duplexes need much lower concentrations to form LC phases than shorter ones. This behavior offers some interest to abiogenesis

because it might suggest a way in which DNA duplexes could sort from one another by virtue of length. Whether there is an accompanying phase transition of short isotropic duplexes separating from longer NEM\* duplexes is unclear given a behavior which will be elaborated shortly.

One of these more complicated polynucleotide behaviors has emerged recently as an additional source of interest to abiogenesis research. On the basis of pure molecular aspect ratio, it would be expected that there is an oligonucleotide length at which liquid crystals can no longer be obtained. The reason for this expectation is that shorter and shorter polynucleotides gradually lose the aspect ratio necessary to drive the phase transition, in and of themselves. As seen in figure 1.2, the pure theoretical limit is an aspect ratio of 4, or 8 nm in length for a duplex with a 2 nm diameter --keeping in mind that this would be for concentrations that are nearly absolutely space-filling. A mesophase occurring at a relatively attainable volume fraction of 0.5 would be possible only with duplexes longer than 16nm, about 4 helical turns (~44 bases). At shorter than these lengths, the aspect ratio is so short that volume fractions of greater than one, which are unattainable, would be needed to drive the phase transition. As such, LC phases are theoretically forbidden for duplexes shorter than about two helical turns. For a duplex of twenty bases, about a turn and a half of the double helix, there is no expected concentration that can produce LC phases.

Despite this expectation, polynucleotide duplexes of this length and shorter *do* form LC<sup>25</sup>. LC phases emerged for duplexes as short as 6 base pairs, well below the expected limit<sup>25,26</sup>. This material is since referred to as nanoDNA or nanoRNA because the nucleotide oligomers forming these mesophases are on the single nanometer size scale.

Construction of these mesophases arises from a pattern of hierarchic self-assembly of oligonucleotides duplexes<sup>25</sup>.

While the nucleobase sequence which encodes genetic information is the result of hydrogen bonding between bases, dA to dT and dG to dC, this interaction is not solely responsible for providing the duplex with sufficient structural rigidity to give the persistence length for LC phase formation. A second driver is the hydrophobicity of the bonded pairs of nucleobases. After two single-stranded oligomers have base paired, the resulting structure undergoes a conformation shift to hide the hydrophobic, hydrogen-bonded nucleobases from the surrounding solvent. This interaction is further stabilized by a  $\pi$ -stacking interaction between the nucleobases where the aromatic ring systems in the bases lie atop one-another in a column at the core of the duplex. The double helix conformation is caused by an imbalance of distances between different parts of each monomer unit in the nucleic acid polymer after it has been base paired and stacked: the  $\pi$ -stacking imposes a 3.4 Å spacing between stacked nucleobases, whereas 7 Å of the sugar and phosphodiester polymer backbone must somehow be accommodated between each base, leading the excess length of backbone to wrap around the cylindrical perimeter of the duplex for each 3.4 Å of rise along the column<sup>22</sup>. The choice of helix direction is made by the uniform chirality of the sugars, leading every unit in the polymer to twist around in the same direction<sup>18</sup>. DNA is seen in three different forms of double helix, right-handed A-form and B-form at physiological solvation conditions and left-handed Z-form, which is not known to be biological<sup>6</sup>. DNA duplexes are most frequently seen to be B-form under aqueous conditions of interest<sup>6</sup>, while RNA tends toward A-form because of the extra steric hindrance from its 2'-hydroxyl. A-form and B-form can be imposed by protein interactions and

the helix is flexible enough that these can be attained in different stretches of the same duplex.

Hydrophobic effects are a large cause of duplex polynucleotide rigidity and the associated rigid persistence length which enables them to form LC phases. Duplex self-assembly proceeds first from a dimer association as a result of hydrogen bonding and then twists to a helical conformation based on hydrophobic effects.

For nanoDNA LC mesophases, self-assembly of an oligonucleotide duplex is required as a precursor for mesophase formation since single-stranded oligonucleotides do not form LC phases<sup>25</sup>. Hydrophobic stacking of base pairs is also understood as a chief driver since oligomers designed to contain flexible, non-interacting overhang regions at their duplex terminals are suppressed from forming mesophases<sup>25,26</sup>.  $\Pi$ -stacking in a duplex appears insensitive to backbone linkage between successive base pairs along an oligomer, meaning that base pairs may stack regardless of whether they are physically chained by a backbone or not. Blunt-ended duplexes from paired 6-base oligonucleotides have at either end an exposed aromatic surface that prefers to be hidden from the solvent as much as any base pair inside the stack, permitting such duplexes to stack end-to-end as an aggregate<sup>25</sup>. And, since the rigid persistence length of the aggregate is due more to hydrophobic stacking than to covalent backbone linkage, the resulting association behaves as a functionally longer object. Repeated stacking of these short units produces an aggregate long enough to result in LC phase ordering by excluded volume interactions<sup>25</sup>. Disruption of terminal stacking between short duplexes would be expected to prevent mesophase formation<sup>25,26</sup>. And, conversely, designed duplexes that contain flexible terminal overhangs that are mutually complementary in sequence such that they base pair, also form mesophases, showing that any route

leading to an elongated aggregate by some terminal-mediated assembly mode will favor mesophases<sup>27</sup>.

As it turns out, the formation of LC phases by these composite objects has a bootstrapping effect. Within liquid crystal phases, the coorientation of aggregates greatly increases the effective concentration of accessible hydrophobic surfaces. So, the formation of an LC phase by self-assembled aggregates favors greater stability of aggregation, leading the LC phase to cause an increased length of the aggregates inside it<sup>28</sup>.

If one were to envision a chemical mechanism by which duplexes can covalently crosslink after they have been stacked, this introduces a feedback loop that might be encountered in the “warm pond” model for the origin of life. Stacked short duplexes create LC phases during environmental drying, greatly increasing the susceptibility of such duplexes to chemical cross-linking and creating effectively longer duplexes. Those longer duplexes are then capable of forming LC at lower concentrations, increasing the degree to which they self-protect and leading them to resist dissolution to lower concentrations, then enabling them to form LC phases at those same lower concentrations upon encountering drying conditions again, potentially expediting the process of trapping and ligating further material<sup>25</sup>. This would lead to the production of gradually lengthening complementary oligonucleotide sequences in the environment decoupled from a biological source. It is only a short jump of the imagination that this could be a feed mechanism dumping raw material into the environment in support of an early RNA world.

A further facet of oligonucleotide behavior which enhances this picture is the potential for depletion effects. Depletion is a spontaneous chemical phenomenon where co-dissolved solutes tend to separate from one another based on their flexibility. A population of polymer chains with very short persistence length will

spontaneously phase separate from a mixture of objects with longer, more rigid persistence length in order to achieve a greater configuration space by not being near the rigid objects<sup>29</sup>. Oligonucleotides have the singular property that they can exist in both a flexible single-stranded form and in a substantially more rigid duplex form. Duplexed and single-stranded will therefore tend to phase-separate from one another, increasing the volume fractions of both within their independent partitions<sup>30</sup>. Thus, the partition containing the duplexed form may be forced above concentrations that result in LC formation and, for the single-stranded partition, decrease the volume through which oligomers would need to sift in order to find a suitable base pairing partner and increasing the chances that they would ultimately form a rigid duplex and cross into the LC fraction themselves.

In the context of abiogenesis, depletion combined with LC formation and ligation would have a selective effect, sorting and partitioning oligonucleotides based on length and sequence complementarity. The observation of LC phases in nanoDNA mixtures with random sequence content at lengths between 12 and 20 bases would seem to bear this out, suggesting that single-stranded oligonucleotides can indeed spontaneously sort into duplexes that can form mesophases and that such phases do not require their constituent oligonucleotides to be of identical sequence<sup>31</sup>. Experience with random sequence nanoDNA further suggests that the self-assembly process is tolerant to base pair mismatches if they do not disrupt terminal pairing or stacking and if they don't form significant or frequent enough out-looping structures to upset the spacing of the phase<sup>31</sup>. Random sequence nanoDNA typically produces only COL phases and not NEM\*.

For uncertainty of an oligonucleotide ligation chemistry that is consistent with reagents available at the beginnings of life, ligation experiments with nanoDNA mesophases turned to a reaction promoted by carbodiimide, (as described

by von Kiedrowski<sup>32</sup>) which can form phosphodiester linkages between a nanoDNA-linked terminal phosphate and a free hydroxyl at the expense of producing urea (see Figure 2.21 for more information). Formation of LC phases substantially enhanced ligation of nanoDNA over disordered phases, much as postulated above<sup>33</sup>. In order to provide sufficient carbodiimide to drive the ligation, this experiment also utilized deletion-driven phase separation in order to place a soluble carbodiimide source in contact with, but phase-separated from, an ordered LC, showing how chemical potential gradients similar to those used by modern cellular life might be established in absence of a cell within a barebones phase-separated system to achieve prebiotic chemistry<sup>33</sup>.

Much of the work performed on nanoOligonucleotide LC phases has been carried out using DNA because of the convenience and availability of this material, but some investigation has shown that such phases are also available in nanoRNA<sup>26</sup>. This suggests that RNA oligonucleotides are subject to similar forces and that these same kinds of effects could easily be extended to encompass an as-yet-unknown RNA progenitor.

## Section 2.0

### Revisitation of Dodecamer nanoDNA

A great complexity exists in nanoDNA mesophases that is superficially innocuous. These lyotropic LC phases are at least three component mixtures, constituted minimally of oligonucleotide, counter ion and water. This provides them with a wealth of behavior in a multidimensional phase diagram, first and foremost including temperature and concentration, and to a lesser extent pH and spectator ion content. The most similar type of lyotropic LC system, the chromonic liquid crystals, is distinct from nanoDNA because the molecular stacks from chromonic LC are typically monomeric whereas nanoDNA is dimeric and the aggregation physics is not merely isodesmic, but must incorporate some description of the association between single-stranded nanoDNA oligomers in order to arrive at a stackable complex. The result is that nanoDNA phases have several competing melting temperatures: a melting temperature for when the aggregates locked inside a mesophase tend to melt into isotropic or into lower order phases and also a melting temperature for when the two strands of a duplex separate from one another. Because of these competing behaviors, some of which prove difficult to separate from one another experimentally, previous descriptions of 6-20 base pair nanoDNA phase diagrams have been incomplete.

This section deals with advancement of work intended to directly reproduce and add precision to previous nanoDNA characterization efforts with a focus specifically on characterization of dodecamer.

## 2.1 Updated Dodecamer Phase diagram

Concentration vs. oligomer length phase diagrams were first reported for nanoDNA of length 6 to 20 bp in 2007 by Nakata<sup>25</sup>. This phase diagram is reproduced below (Figure 2.1). Among these, the stereotypical oligomer was a self-complementary, blunt-ended palindrome with two stretches of G-C base-pairs bracketing a core of A-T. This is typified by the Drew Dickerson Dodecamer CGCGAATTCGCG (DD).

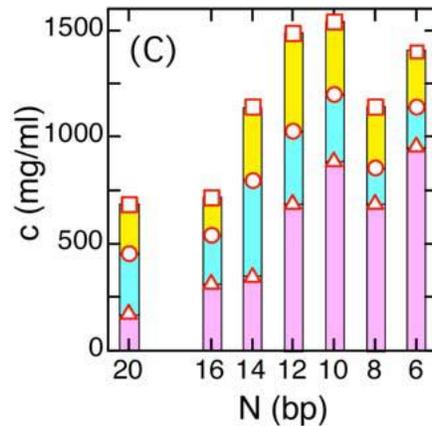


Figure 2.1: Uniform temperature phase diagram of nanoDNA lengths 6 -20 bp from Nakata 2007 (Supplementary Information<sup>25</sup>). DD is 12 bp. Phase diagram for this Oligomer is reported at a temperature of 20°C (except for 6 bp, located at 10°C). Phases reported are Isotropic (ISO) in purple, Nematic\* (NEM\*) in light blue and hexagonal Uniaxial Columnar (COL) in yellow.

This phase diagram is held as a prototype to most of the work to be reported here. The original work was with DD, while the new work was performed with a 12mer regarded to be equivalent; GCGCGTTAAGCGC which is also a self-complementary palindrome with a blunt-end, but with reversed sequence. This

oligomer is called Reverse Drew Dickerson Dodcamer (rDD). It is reasoned that if the original reported phase diagram is dependent solely on the self-assembly pathways to produce LC forming aggregates, by end-stacking of rigid duplexes, then the phase diagram should be insensitive to internal polymer sequence ordering if the base content is otherwise indistinguishable.

The original phase diagram reports ISO-NEM\* transition at  $\sim 700$  mg/mL with coexistence beginning at about  $\sim 500$  mg/mL. NEM\*-COL transition is then reported at above 1000 mg/mL. The COL phase in this work was determined to be hexagonal by X-ray diffraction through microdomains. Stacking energy ( $E_s$ ) is reported to be  $4 K_bT < E_s < 8 K_bT$ .

Work with rDD as a standard oligomer since the 2007 publication and work on other nanoDNA sequences was found to be irregular and not very equivalent to the phase diagram above. Observations on 12mer sequences between labs showed significant deviation from one another and great self-inconsistency, which will also be addressed (Section 2.4).

Phase diagram replication was approached by use of high stability capillary samples where concentrations were calculated using equation B.29. The phase diagram (Figure 2.2) reported here was exhaustively determined by a combined microscopic examination of phase textures in these capillaries and observation with X-ray diffraction (to be discussed, Section 2.2).

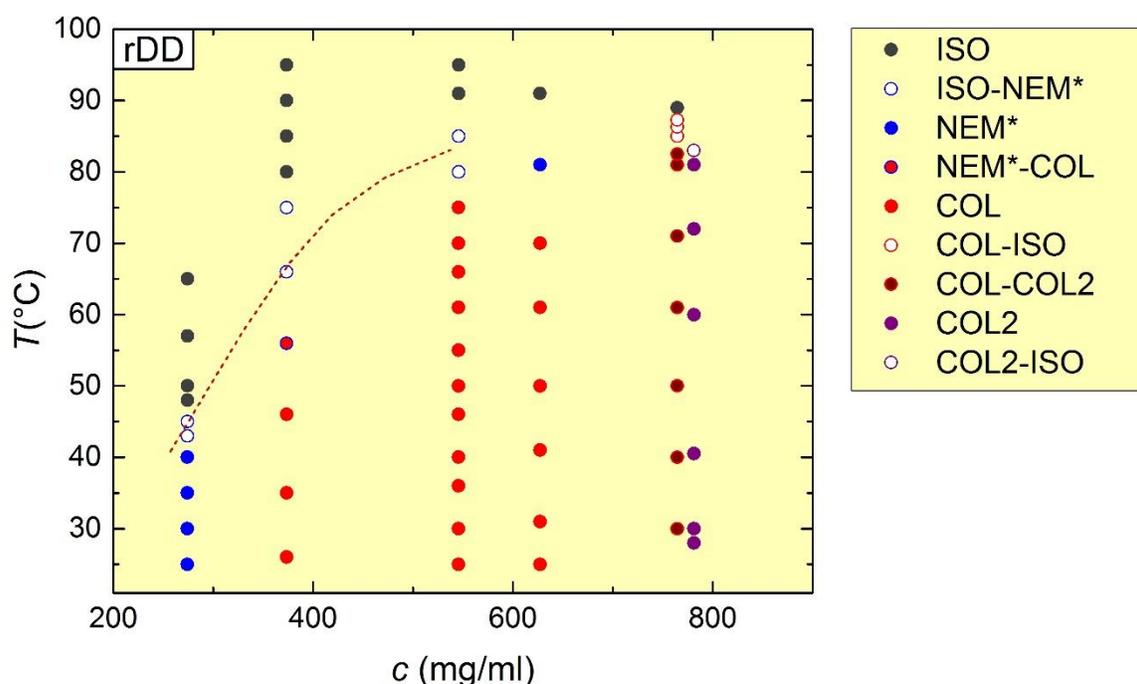


Figure 2.2: Combined phase diagram of rDD. Phases noted here include a ISO, NEM\*, COL and COL2. The dotted line marks the coexistence line of ordered phases with ISO.

rDD NEM\* phases were observed at concentrations of 270 mg/mL and COL texture phases appeared at concentrations no higher than 400 mg/mL; less than half the concentration of where they were previously expected. Greater than two-fold discrepancies at concentrations of  $\sim 500$  mg/mL are in the range of 25%-50% volume fraction ( $\phi$ ) error for an absolute dry DNA density taken to be 1687 mg/mL ( $\phi=1$ ).

Capillary phase textures are shown in Figure 2.3 in comparison to flat cell texture images. Flat cell textures frequently display large concentration discrepancies from capillary observations of the same phase textures, and flat cells drift in concentration over time at a non-uniform rate. The examination of capillary samples adds significant precision to the phase diagram by controlling potential errors due to evaporation of water or due to concentration gradients.

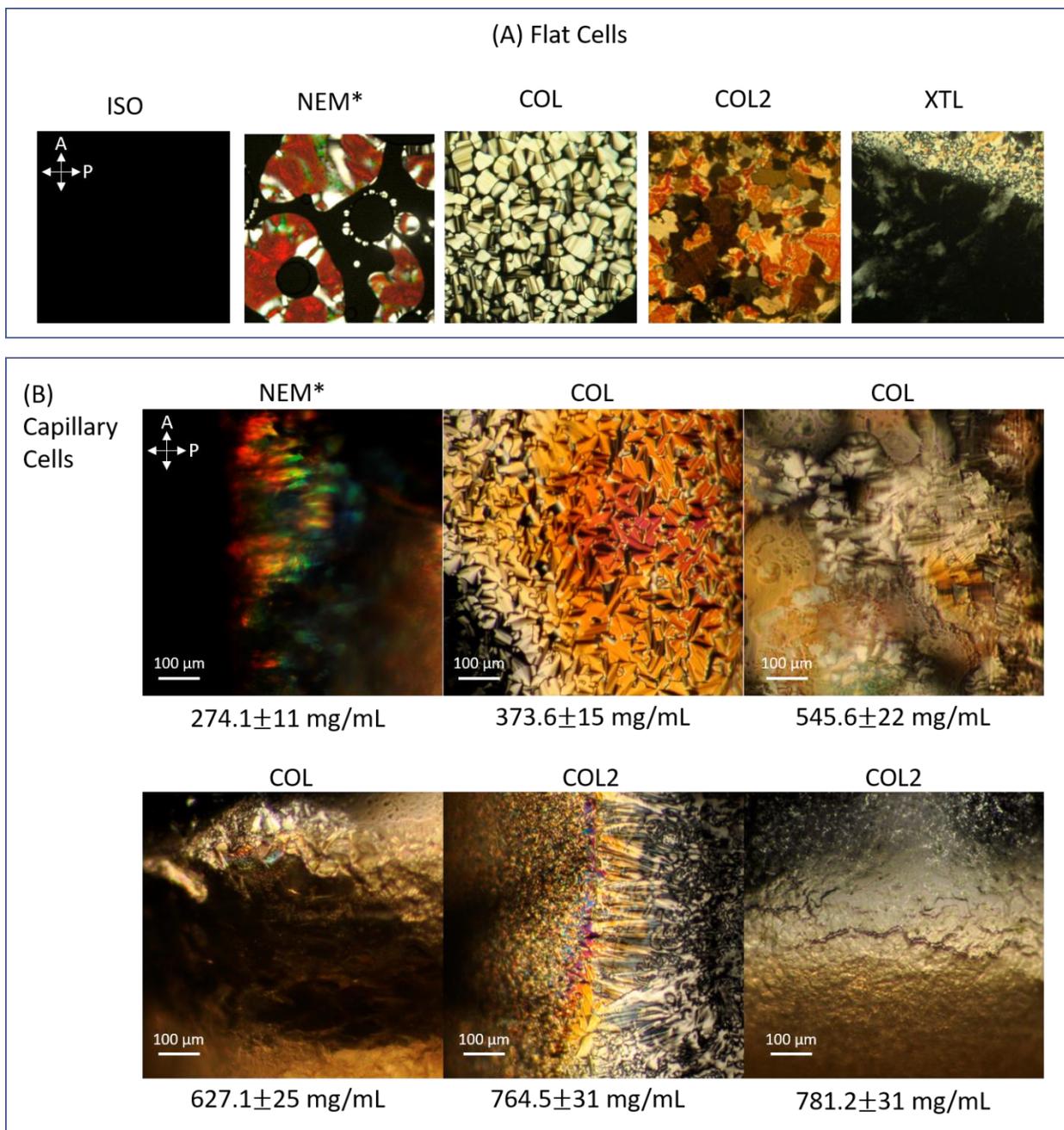


Figure 2.3: Dodecamer mesophase textures. A.) Textures in flat, oil sealed cells (from ref); nanoDNA sequence is DD. B.) Textures in equilibrated, flame-sealed round-profile capillaries as viewed with index matching, nanoDNA sequence is rDD. Concentrations are as calculated and error is normalized from percent error in weight measurements.

Textures examined in capillary show strong homology to phases seen in flat cell in most cases where they can be discerned. Microscopy of textures in capillary cells is difficult in some cases without additional means of assigning the identity of

the texture since the thickness of the 3D volume prohibits visualization of individual domains. This can be circumvented in some cases by allowing textures to equilibrate while spread up the walls of a capillary so that optical path lengths through this material are short. NEM\* capillaries show colored optical-pitch implying helical order, as exhibited in Figure 2.3B, though the fingerprint and Grandjean textures typical of capillaries cannot be easily discerned. Phase transitions occur simultaneously across the entire volume of these capillaries, as would be expected from the existence of minimal concentration gradients. COL phases in capillary show conformal focal conic textures identical to those seen in homologous flat cells. Higher order textures seen at higher concentrations become difficult to examine since domain sizes tend to shrink with increasing concentration. In cases where visualization was difficult, X-ray diffraction proved the nature of the phase conclusively (to be discussed in Section 2.2). Flame-sealed capillaries offered the additional benefit of stability during temperature cycling, allowing the phase transition temperatures to be clearly and repeatably identified both during X-ray diffraction and microscopy. X-ray diffraction was performed with minimal exposure durations in order to limit mesogen damage and this was confirmed optically by the observation of minimal variance in phase transition temperature occurring in cells prior and post X-ray exposure. Hysteresis of  $T_m$  between phase melting and phase condensation processes was not observed in these samples.

The phase diagram in Figure 2.2 benefitted from X-ray diffraction observations which allowed the assignment of transient NEM\*-like phases occurring at coexistence boundaries between ISO and COL phases, suggesting that melting is a dynamic unpacking process with overall density changes in the domains as phases come apart. COL2 phases and COL-COL2 phase coexistence

were predominantly identified in the capillary cells of Figure 2.3 from X-ray diffraction, neither of which was possible to confirm optically.

COL phases are understood to be hexagonal uniaxial columnar phases based on X-ray data (see Section 2.2.2). The exact nature of COL2 is not perfectly clear given the existence of more complicated lattice correlations, but appears to be a flattened hexagonal, orthorhombic-like form which would hold as common to longer oligonucleotide polymers (see Section 2.2.3).

## 2.2 X-ray Diffraction studies of Reverse Dickerson Dodecamer

This section will cover the extensive X-ray diffraction studies carried out on the full phase diagram of rDD nanoDNA. Characterization of each mesophase will be tackled in turn and properties of all such phases will be examined comparatively.

As an overview of the more specific data tackled in succeeding sections, collected room temperature WAXS scans of all the rDD mesophases can be seen in Figure 2.4. This series of plots gives a generalized taste of how diffraction data appears for this variety of mesophase. Generalized processing of this data is described throughout Appendix C. Sample dimensions include scattering  $q$  (in  $\text{\AA}^{-1}$ ) of a structural feature versus concentration (in mg/mL) and will be expanded to include the temperature dimension in following sections.

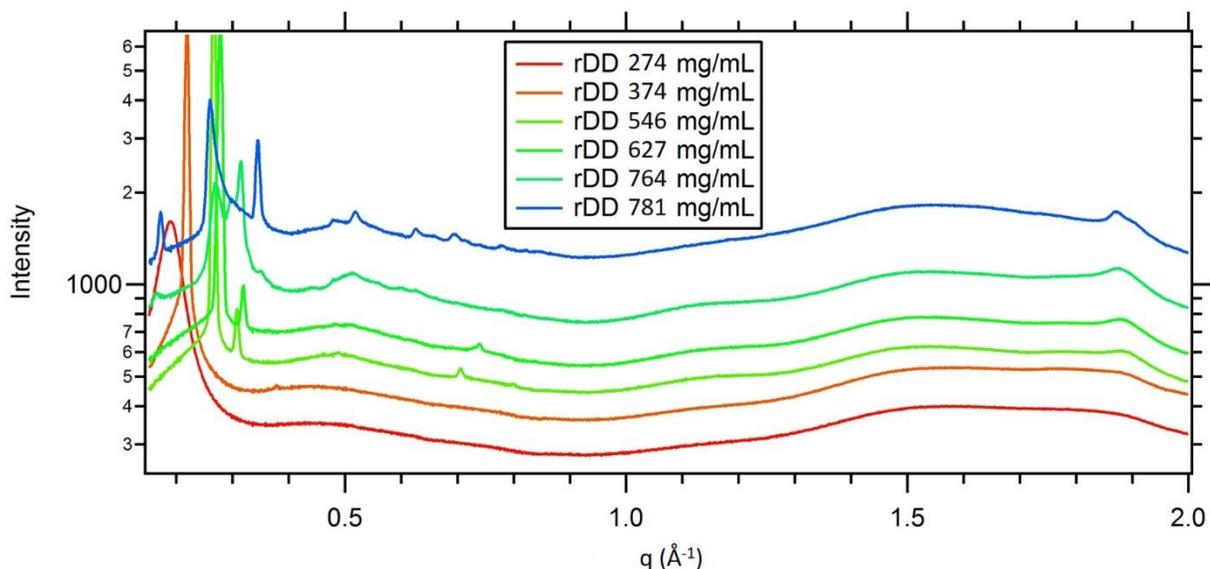


Figure 2.4 Collected rDD WAXS from room temperature. Concentrations are as depicted. All traces were subject to circular averaging and are plotted together with an offset. Reflections due to mesophase packing correlations appear in the range of  $0.2 \text{ \AA}^{-1}$ , while DNA base stacking correlations provide the reflection at  $\sim 1.85 \text{ \AA}^{-1}$ .

These data can be collected into two groups, X-ray reflections that are related to the structure of the mesophase, which appear in the small angle region (small  $q$ ) and the reflection due to scattering from the DNA duplex base stacking, which appears at wide angle (large  $q$ ). The base pair scattering peak is the strongest reflection due directly to the structure factor of the DNA duplex while the other reflections are from semicrystalline ordering. The red curve, for 274 mg/mL, is confirmed to be NEM\* by PLM (see Figure 2.3). The orange through green curves, 374 mg/mL through 627 mg/mL, are confirmed optically to contain COL textures (see Figure 2.3). Light blue and blue curves are of unambiguously higher order phases by WAXS, but the textures are unclear during microscopy of the capillary and are mainly taken to be representative of higher order textures found in flat cells as correlated by matching concentration (see Figure 2.3). Higher order phases seen in these samples are understood qualitatively to be LC phases rather than full crystals because of the lack of X-ray reflections seen in the region of  $q$  between 1.0

and 1.8 since highly structured crystal forms tend to generate higher harmonic reflections very easily.

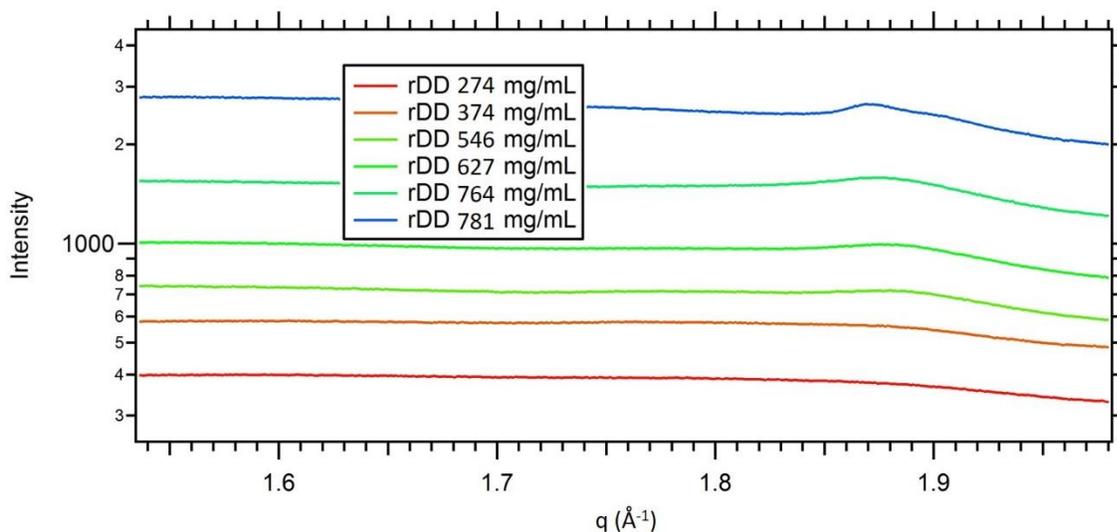


Figure 2.5 Collected WAXS of rDD, base pairing region in wide angle. Concentrations are as labeled in this plot focusing on the wide-angle reflection seen at room temperature. Base stacking correlations become stronger as the concentration increases. For a sense of scale  $1.85 \text{ \AA}^{-1}$  is about  $3.4 \text{ \AA}$ .

Evidence of base stacking grows stronger in more structured phases. It has been seen previously that a lack of base pairing along with a lack of stacking results in an inability to form mesophases altogether<sup>25,27,30</sup>, so while poorly structured phases do not have strong X-ray evidence of base stacking, stacking must still be present in those phases if they are observed to form mesophases. This introduces some ambiguity in the potential for WAXS in capillaries with unaligned nanoDNA mesophases to be able to monitor alterations in stacking or base pairing. The signal is usually not strong enough to do more than make measurements of the stacking periodicity. On the other hand, the high concentration phases show bimodal or multimodal stacking peaks in the single best trial, suggesting that WAXS may be able to detect the difference between various modes of stacking at high concentration: G-C and A-T type base-pairs may stack at a slightly different spacing

from a neighbor depending on whether their neighbors are of the same type, different type or even reversed in orientation in the double helix (an A-T pair stacked with an A-T pair, where one polymer backbone is 5'-AA-3' and the other is 5'-TT-3', versus A-T stacked with a T-A, where each polymer backbone contains 5'-AT-3'). No experiments were made to check this notion and the details involving this did not appear clearly across all experimentation. In Section 2.2.3, some thought is given to whether the bimodal structure seen here could in fact be due to aggregate distortion in the COL2 mesophase, but this is not clear from additional data presented in that section.

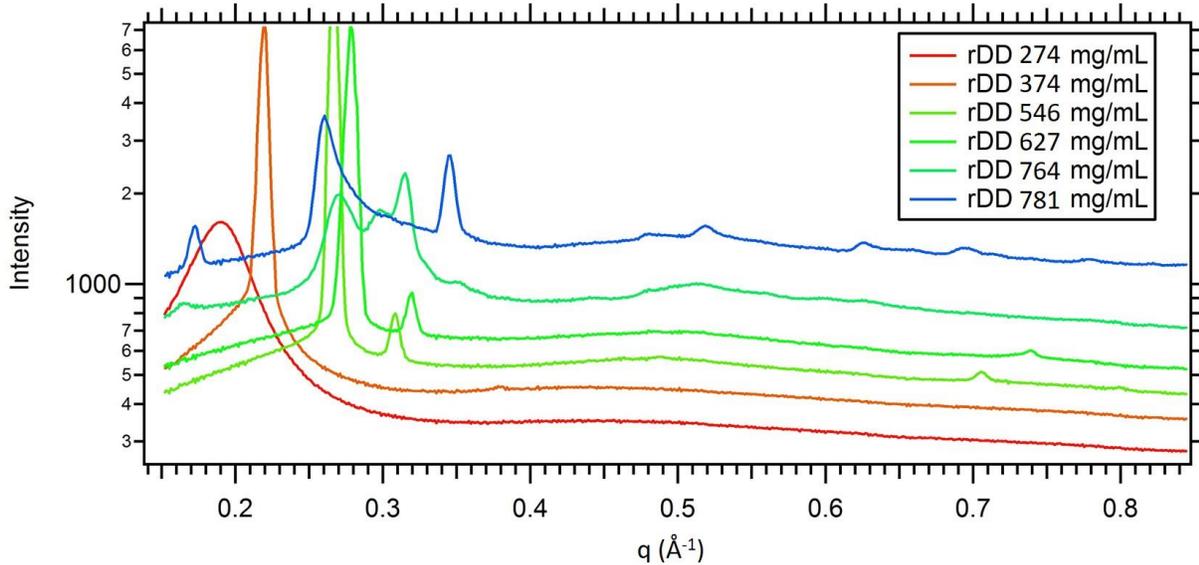


Figure 2.6 Collected WAXS of rDD in small angle. Concentrations are as depicted with all shots taken at room temperature. Smaller values of  $q$  correspond to larger spacing in real space. For a sense of scale, 2.0 nm is about  $0.31 \text{ \AA}^{-1}$ . Shifts toward larger  $q$  with increasing concentration imply compression of mesophase packing; the distance between columns is decreasing.

rDD mesophase structure has a surprising and somewhat unexpected wealth when examined by these means. NEM\* phases at very low concentration have a detectable positional correlation. COL phases in orange and green traces actually appear to show some distinct structure given the inclusion of a second scattering

peak that emerges at higher concentrations which does not appear to correspond to any hexagonal or square crystalline periodicities, even though it shifts at increasing concentrations along with the COL spacing peak –this will be distinguished from simple COL as COL+ with the extra peak. Higher order COL2 columnar phases begin to show a family of other peaks which beg additional interpretation given the concern that COL2 may actually be a collection of similar but distinct phases. At a qualitative level, this initial set of observations appears to show something like five different types of mesophase ordering. At the final interpretation (as appears in Figure 2.2) this has been collected into just three phases due to the potential for phase coexistence and the ambiguity of the COL+ correlation, which seems to appear freely in COL phases without an apparent change in mesophase texture.

Some of the analyses to be presented here must be noted to suffer from difficulty in the ability to collect a distinct, universally applicable baseline. Many efforts were made at acquiring a distinct baseline across multiple trips to the Advanced Light Source. First, shots through sections of the sample capillary lacking mesophase showed significant background inhomogeneity which resulted in negative baseline values that may be due to scattering contributions by water in the mesophase. Second, shots through attempted standardization capillaries containing pure water showed significant dissimilarity in baseline which may be due to differing thicknesses of the capillary wall or the lack of counter ion content. Finally, usage of the scattering collected from mesophases melted into the isotropic by increased temperature, ostensibly a correlation free version of the sample, shows significant dissimilarity in baseline from the low temperature baseline perhaps due to transient single strand fluctuation type interactions that are not present at lower temperatures. Baseline has been addressed as well as possible wherever it could be made consistent.

### 2.2.1 Chiral Nematic of Reverse Dickerson Dodecamer

Typical textures of the phase can be seen in Figure 2.2. NEM\* texture acquired with rDD shows the same sorts of characteristics seen with previous experimentation. Microscopically, visible textures include Grandjean textures that lack Schlieren lines seen in more typical unstructured nematic phases which often appear dark or poorly birefringent and tend to brighten slightly when decrossing the analyzer to the left, hinting at chirality. The Grandjean textures can be colored from red to blue apparently dependent on the nanoDNA concentration of the NEM\* domain, on the counter ion or spectator ion content and on the ambient temperature. The helical order of the chiral phase has been seen to wind more tightly toward blue colors on heated and unwind toward red colors when cooled. Colored Grandjean textures often appear to exhibit what seems to be a lamellar structure that will tend to move about on heating or cooling.

When examined perpendicular to the twist axis of the chiral nematic, the phase exhibits bright “fingerprint” line textures that show extinction brushes when perpendicular to analyzer or polarizer, particularly when the Grandjean is dark, typically suggesting that the helical axis of the cholesteric has a pitch length of microns or even tens of microns. Cholesteric pitch is highly changeable and multivariate. Fingerprint lines often appear at the edges of a bulk phase or at gaps inside the phase on the borders of voids.

The PLM observations of rDD are completely consistent with observations previously reported by Nakata<sup>25</sup> for DD. The main difference appears to be that the NEM\* phase is found at lower than 370 or 350 mg/mL, which is far lower than the original observations. Part of this is certainly due to the usage of more stable

sampling formats here, but part may also depend on differences in how the current materials were purified as compared to those materials previously used.

X-ray data were collected in WAXS format as described in Section C.3. The 274 mg/mL capillary sample examined previously by PLM was placed in a specially modified hot stage on the X-ray beamline and exposed at selected temperatures while heating the hot stage in order to observe the temperature dependence of the X-ray scattering features. Multiple data sets were collected over the course of ramping temperature from 26° C to 95° C.

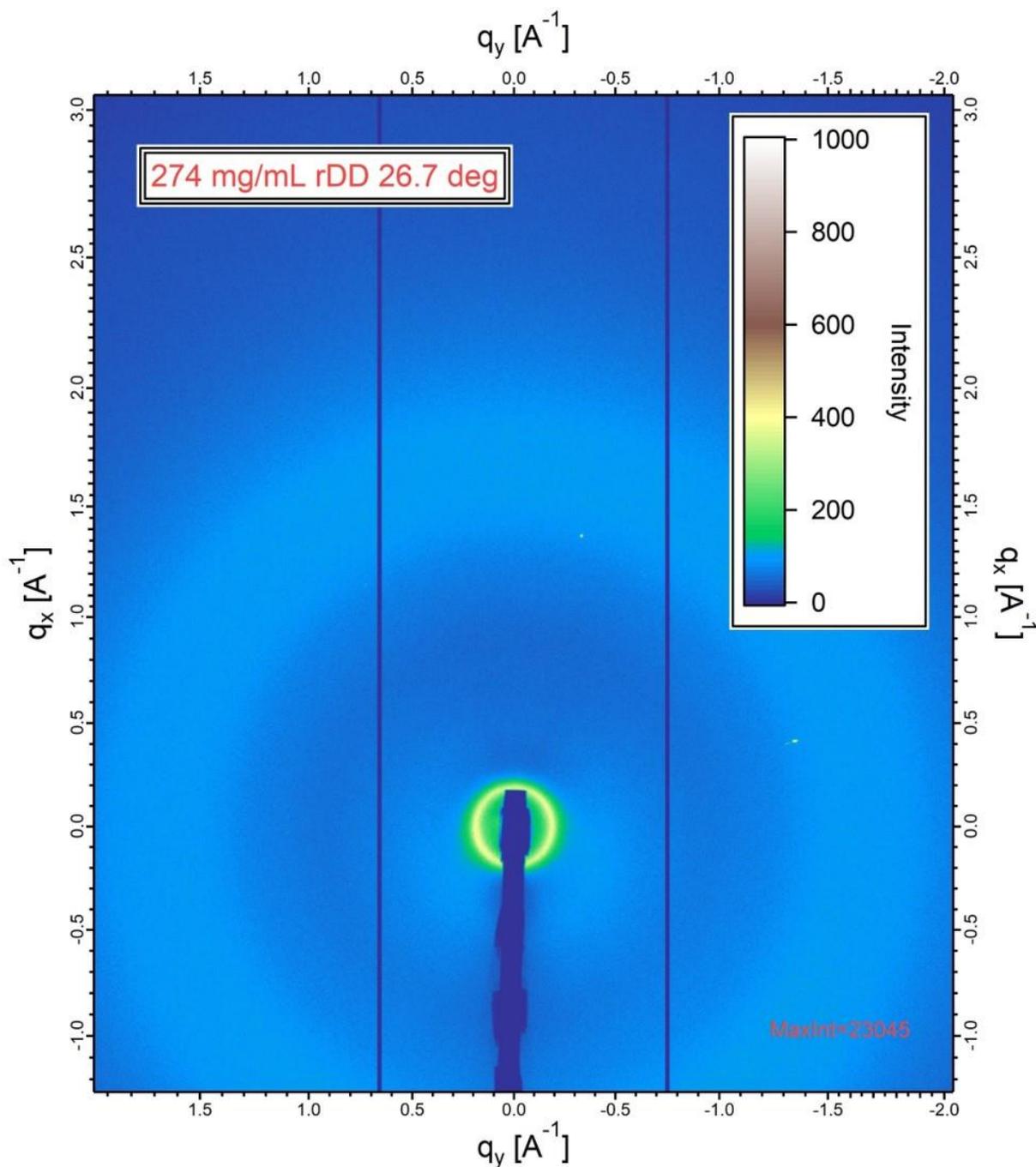


Figure 2.7: WAXS of NEM\* phase of rDD. 2D X-ray powder scattering from the rDD NEM\* phase at 274 mg/mL at room temperature; no phase orientation is attained in this sample. The two dark vertical lines across the 2D image are horizontal CCD panel seams while the shadowed bar at the center of the scattering pattern is the beam stop used to protect the detector from direct exposure to the unreflected synchrotron X-ray beam; all of these features are processed to remove them from the circular average. X-ray intensity is represented with a colored altitude map going from blue at low levels through green and blue to white where white represents intensities of higher than 1000 counts. The

maximum intensity is 23,000 counts and is due to reflections from a mica window present on the hot stage [reflections at  $(q_x, q_y) = (0.4, -1.3)$  and  $(1.4, -0.3)$ ]. The 2D space of the detector is marked in units of  $q$  as  $\text{\AA}^{-1}$ . The yellow-green ring is the NEM\* packing reflection while the base-pair stacking reflection is not qualitatively visible in the color scale above scattering background.

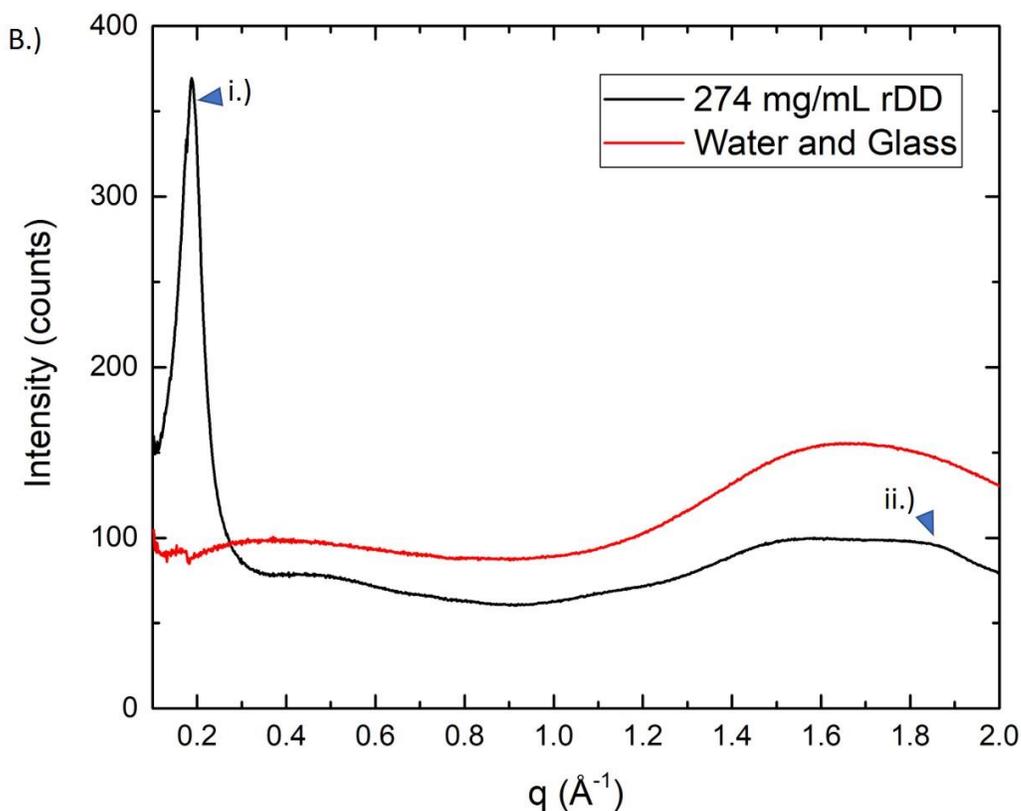


Figure 2.8: WAXS of 274 mg/mL rDD, 1D circular average of the 2D data set with the  $q$ -spacing converted to the radial distance from the center of the scattering pattern. A circular average of a water-glass control sample is added for comparison of the baseline. i.) NEM\* packing peak at  $0.18 \text{ \AA}^{-1}$ , which is expected to represent side-side mesogen density correlation in the phase. ii.) The stacking reflection at  $1.85 \text{ \AA}^{-1}$  represents the periodicity of stacked base-pairs in the structure factor of the nanoDNA aggregate.

Figures 2.7 and 2.8 shows a representative of the data sets typically gathered during the examination of NEM\* phase in WAXS, the first in the unaltered 2D data set and the second set in 1D after circular averaging. The side-side correlation is clearly visible as a ring at  $0.18 \text{ \AA}^{-1}$ , giving a  $d$ -spacing that is calculated to be  $34.9 \text{ \AA}$

in real space. No reflections suggesting a lattice in mesophase ordering are unambiguously visible at larger values of  $q$  above the background. The stacking peak is visible as a tiny shoulder at  $1.85 \text{ \AA}^{-1}$ , giving a stacking periodicity of  $3.4 \text{ \AA}$ .

Because there is only the one scattering reflection assigned to the packing correlation and no orientation data available to place it in  $q$ -space at some angle relative to the base-pair stacking peak, there is ambiguity as to whether these reflections are  $90^\circ$  perpendicular to one another, as would be expected if the packing correlation were a side-side feature. There is some small chance this peak is actually related to the periodicity of the duplex nanoDNA cylinders stacked up into an aggregate column, but the associated length would be expected to be  $40.9 \text{ \AA}$  rather than  $34.9 \text{ \AA}$ . There is also a chance the packing correlation is some self-width of the nanoDNA duplex, but again, such a width would be expected at about  $20 \text{ \AA}$ . Further, given the overall weakness of the base-pair stacking peak, other such features of the nanoDNA duplex structure factor would actually be expected to be weak as well, rather than strong. Another line of evidence places this peak as a side-side correlation which will be mentioned in examination of the collected data from all the rDD mesophases in Section 2.3.2.

Simple LC phase modeling also suggests that this peak can only be a side-side correlation. Since the phase is known to be a nematic, it must be at a greater concentration than the critical concentration implied by the Onsager criterion, which is dependent on nanoDNA the length of rDD undergoing self-assembly to produce an aggregate of sufficient aspect ratio to generate a nematic. This constrains the dimensions of the aggregate given that the diameter of a DNA duplex is known to be  $2 \text{ nm}$ . Using the Onsager criterion (eqn 1.1), the volume fraction of this NEM\* at  $274 \text{ mg/mL}$  is related to the aspect ratio of the aggregate in a known manner:

$$\varphi > 4 \frac{d}{L} \quad (2.1)$$

Here,  $\varphi$  is the volume fraction of the 274 mg/mL phase, which is greater than the critical volume fraction for forming a nematic phase  $\varphi_c$ ,  $d$  is the diameter of the aggregate and  $L$  is the length.  $\varphi$  is a fraction of the volume of the object  $V_o$  per the total volume available for it to occupy  $V_T$ , such that  $\varphi = V_o/V_T$ . Concentration  $c$ , as explained in Sections B.2.1 and B.2.2, is mass  $m$  of the object per total volume  $V_T$ , or  $c = m/V_T$ , while density  $\rho$  is the mass of the object per the volume  $V_o$  occupied by that mass,  $\rho = m/V_o$ . By eliminating mass,  $V_o/V_T = c/\rho = \varphi$ .

$$\frac{c}{\rho} > 4 \frac{d}{L} \quad (2.2)$$

Or, rearranged:

$$L > 4 \frac{d\rho}{c} \quad (2.3)$$

As the concentration is known to be about 274 mg/mL, the diameter of DNA is about 2 nm and the absolute density of DNA is 1687 mg/mL (see Section B.2.2), this requires  $L > 49.3$  nm. If the 3.5 nm periodicity is a characteristic spacing of objects in the NEM\* phase, it must be side-side since 50 nm objects can't stack end-end or side to end with a periodicity of 3.5 nm.

The packing peak demands some special processing because the X-ray beam stop clips into the scattering ring fractionally along one edge; as temperature is changed in the sample, this produces a small dip in the baseline which makes the peak hard to fit by disconnecting the scattering intensity statistics between neighboring radii during a full circular average. Data was secondarily processed by a sector average over an angular block of the ring that avoids this dip. This issue did not exist in other data sets where the peak position was typically further from the beam stop and the width of the peak was considerably smaller.

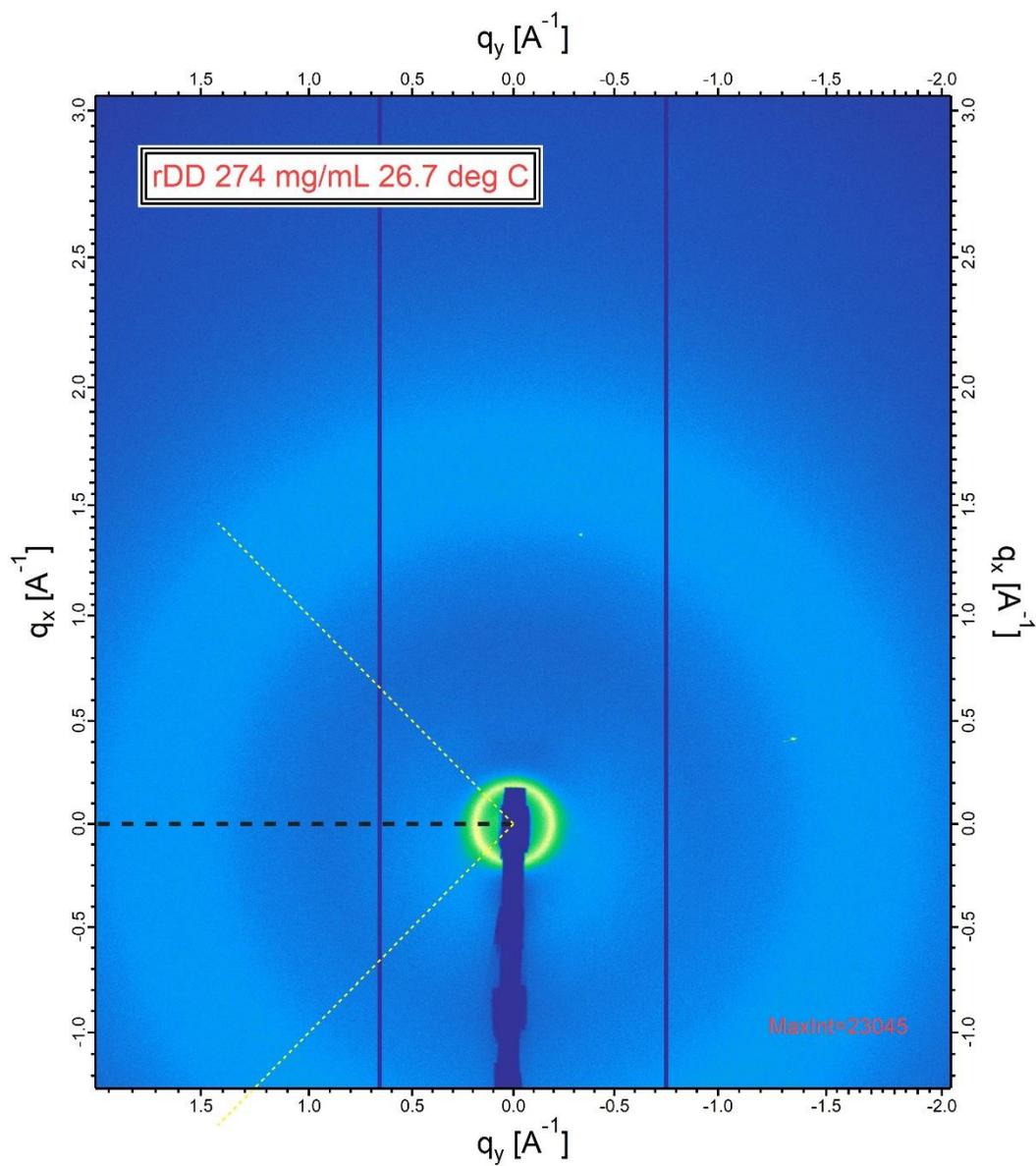


Figure 2.8: Sector Average angular block. Centered on the black dashed line is the angular sector used to avoid distortions in the averaging as a result of data clipping to the ring by the beam stop. Again, this issue existed only in the NEM\* data collection.

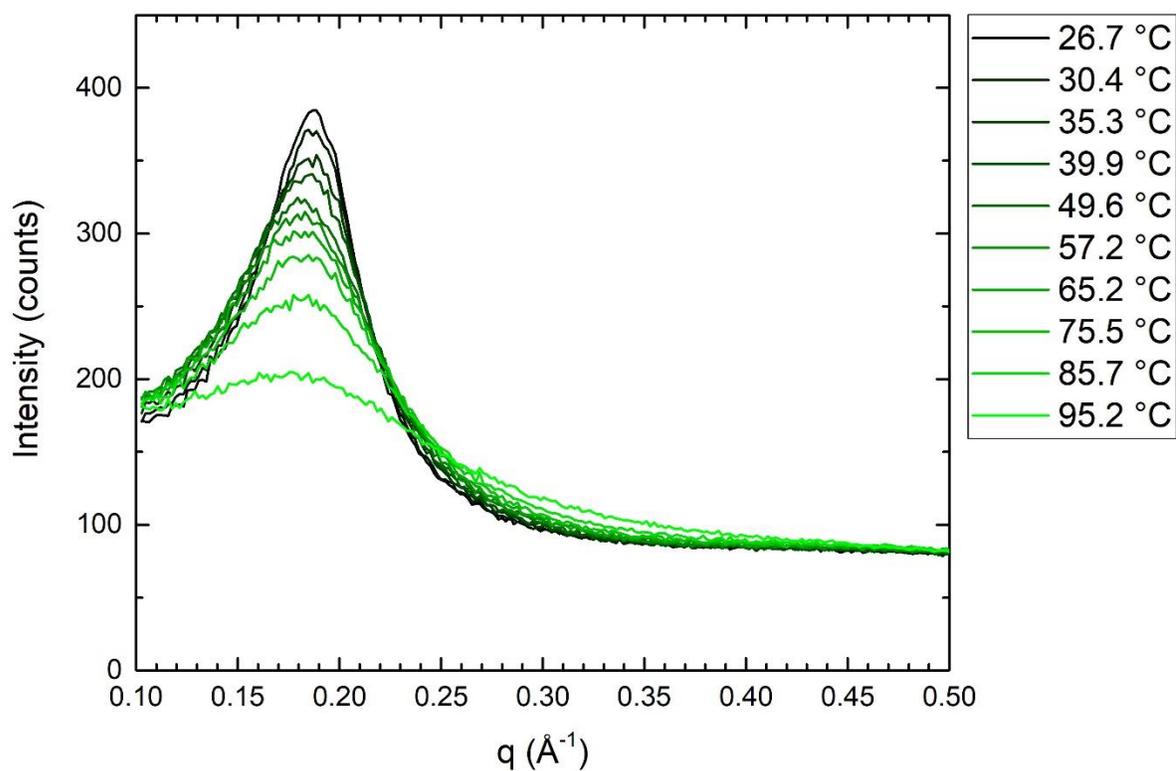


Figure 2.9: Small angle region, collected scans during temperature ramp experiment of 274 mg/mL rDD. This window highlights the small angle peak collected by sector average of the 2D scans where each line represents one 2D data set. The progression of color between scans from black to bright green follows the increase of temperature. The baseline varies somewhat between temperatures in the space between the peak and the beam stop, owed perhaps to the appearance of spurious interactions in the small angle with elevated temperature as the mesophase unpacks in order to melt.

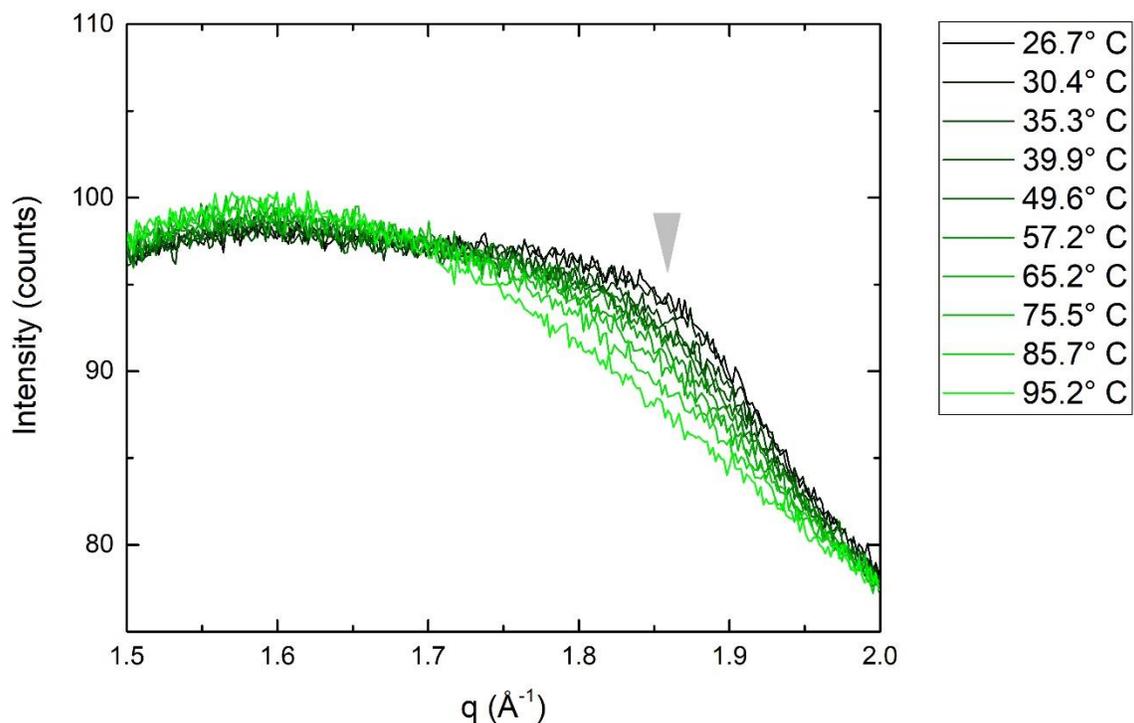


Figure 2.10: Wide angle region, collected scans during temperature ramp of 274 mg/mL rDD. This window highlights all scans taken in the region of the DNA stacking peak, denoted with the gray triangle. Data compression from 2D to 1D is by circular average in this sequence given the absence of instrumental clipping in this portion of the detector. Progression of trace color from black to bright green follows the increase of temperature. As all shots are on the same sample for the same duration of exposure, the variation of the baseline away from the peak reflects real variations of scattering by the sample which may be associated with spurious, wider-spaced base pairing interactions in melted or melting DNA duplexes. No fitting was attempted with this peak because of baseline variation between scans and the overall weakness of the peak.

Figures 2.9 and 2.10 represent all data collected on this mesophase during the temperature ramp. Figure 2.9 is the peak representing the mesophase structure while Figure 2.10 is the base pairing peak. No attempts were made to fit the base pairing peak because of variation of the baseline from one scan to the next, even in the same sample. The exact nature of this variation is uncertain but may be due to the appearance of poorly stacked base pairing interactions encountered while

heating and melting this mesophase; the shift always appears to increase the overall intensity at smaller angles, but not in any coherent scattering peaks or across the entire baseline uniformly.

Peak fitting was attempted with the small angle peak using both circular averaging and sector averaging on the block illustrated in Figure 2.8.

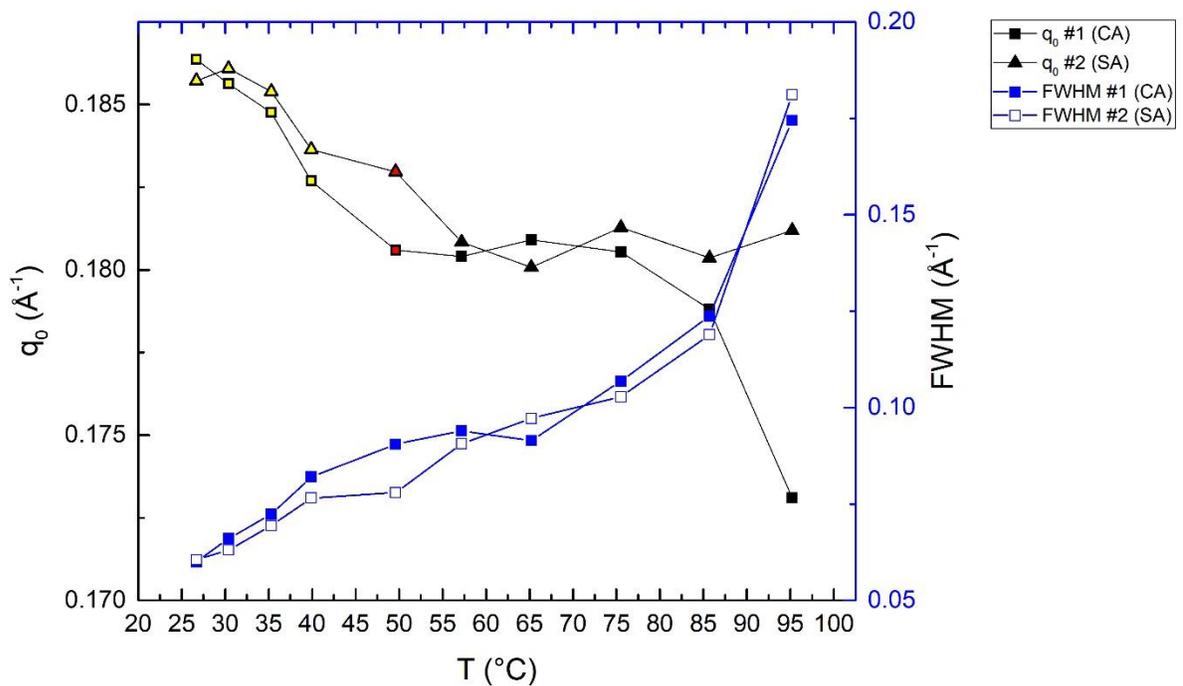


Figure 2.11 Peak fitting versus temperature of rDD NEM\* mesophase. Shown here are the collective Lorentzian peak fitting parameters for the peaks gathered in Figure 2.9. Processing was carried out on both clipped, circular averaged data (#1, CA) and unclipped sector averaged data (#2, SA). The black lines are for the fitted peak center  $q_0$  and the blue lines are for the peak width (FWHM). The vertical axis on the left is associated with the peak center measurement while the blue axis on the right gives the measured peak width. Peak center lines are augmented by markers filled with color to show the PLM observation of birefringence at a particular temperature: birefringent NEM\* phase (Yellow), phase coexistence during melting (Red), and non-birefringent ISO (Black).

Figure 2.11 shows the general trend of accuracy in the peak fitting using the different processing strategies available to try to avoid distortions present in the

original data set. The general trend in peak center appears to be a shift toward smaller angle with increasing temperature until the mesophase melts, at which point the peak center appears to stabilize at about  $0.18 \text{ \AA}^{-1}$  with large apparent fluctuations. The peak width appears to increase across the entire temperature window.

The data collected in Figure 2.11 can be further processed to directly produce one dimensional spacing periodicity and structural coherence length by methods detailed in Section C.1 and C.2.

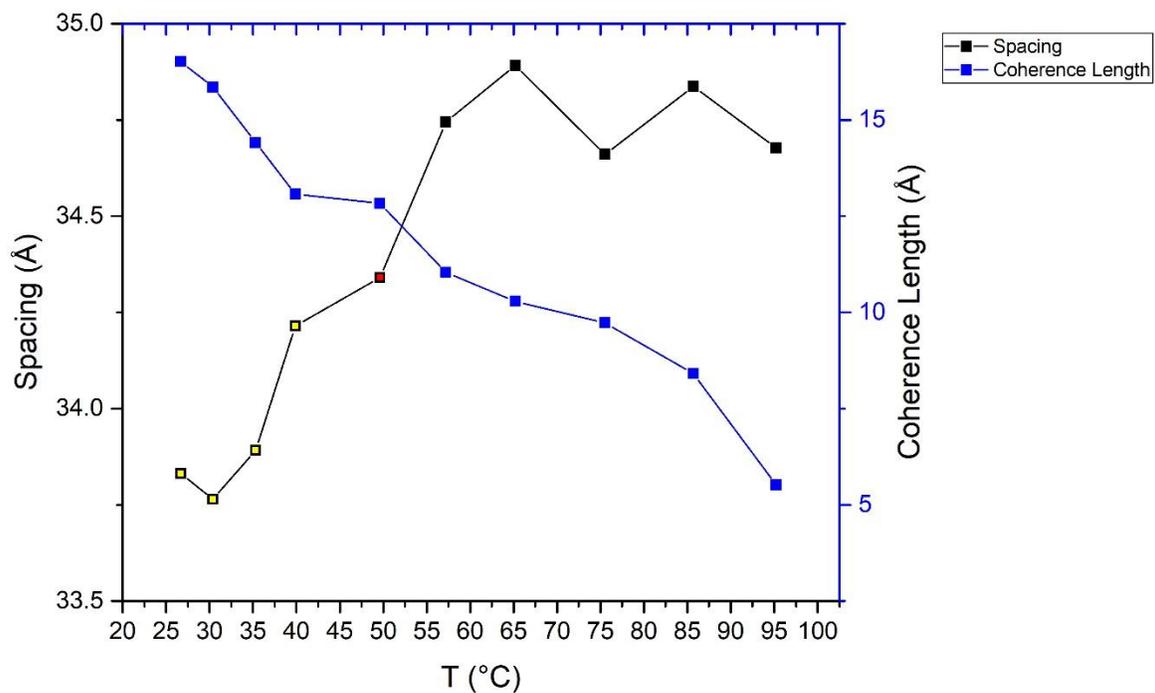


Figure 2.12: Spacing and structural coherence versus temperature. Here, the black line reflects the spacing peak as measured by the black vertical axis on the left, while the blue line represents the structural coherence as measured by the blue vertical axis on the right. Again, peak spacing is augmented by color in the markers to show birefringent NEM\* (Yellow), phase coexistence (Red) and ISO (Black) as observed with PLM.

Figure 2.12 tells much of the story of this mesophase. The approximate structural coherence length in NEM\* is always smaller than the spacing parameter,

suggesting that there is no coherent lattice, which is to be expected from a nematic phase that lacks positional order by definition. The packing correlation can only be a side-side pair-wise correlation reflecting the average side-by-side spacing of mesogen aggregates from one another. Average spacing is seen to grow by about an Angstrom between room temperature and melting temperature, which would certainly be characteristic of swelling of the packing of this phase with increasing temperature. The base-pair stacking is seen at very close to 3.4 Å, the expected value, and since the stacking peak does not move by more than a tiny fraction of an angstrom prior to melting, the swelling is only with the packing and does not involve the stacking. While the packing peak shows apparently discontinuous behavior relative to the phase transition, the coherence length increases continuously across the entire temperature spectrum. Whether this increase is linear is uncertain given the lack of statistical weight, but it appears to be a consistent, nearly linear trend from this one series of experiments.

For this phase, packing spacing would also be expected to shift with concentration, with a higher concentration resulting in a greater mass per unit volume, thereby compressing the spacing between neighboring aggregates. For the NEM\* phase, this would suggest that the packing  $q$  would increase with increasing concentration. This will be visited in greater detail later given that NEM\*/COL coexistence was observed with a low concentration columnar phase sample, but no samples of pure NEM\* were prepared at concentrations other than 274 mg/mL for examination by X-ray given the generalized expectation that NEM\* should not have interesting X-ray scattering to begin with. As NEM\* is fairly well understood from prior published work, this work serves merely to fill in details.

There is no way to observe the cholesteric pitch of NEM\* in the WAXS format since the helical pitch periodicity is expected to be in the hundreds of nanometers to

as large as microns. This might be observable in a format like RSoXS (Resonant X-ray scattering) using a soft X-ray to excite at the carbon edge since base-pair stacking of nanoDNA aggregates would be expected to effectively return to the same orientation on 1/2 the helical pitch period of the cholesteric. While such an observation would be interesting, a sampling format for this material which confines water sufficiently well against a vacuum while dispensing with glass barriers, as would be necessary for the appropriate instrument, has yet to be constructed.

### 2.2.2 Uniaxial Columnar Phase of Reverse Dickerson Dodecamer

Prior X-ray diffraction experimentation on DD has provided compelling evidence that the COL phase is a uniaxial columnar with a 2D hexagonal lattice<sup>25</sup>.

More current observations of rDD show that this nanoDNA produces the same sorts of confluent, splay-free textures in the form of highly ordered focal conic domains that possess negative birefringence. This has always suggested that nanoDNA aggregates circle around a focal conic texture with the plane of their base-pairs parallel to the texture's radius of curvature (as demonstrated in Figure 4.4). COL texture is usually seen at concentrations of between about 370 mg/mL up to greater than 650 mg/mL (see Figure 2.2). As before, the most major difference between the current rDD observations of COL phase and the originally reported observations with DD is in the huge decrease in concentration where the mesophase appears 370 mg/mL as opposed to 1000 mg/mL (see Figure 2.1). Based on the concentrations and the implicit volume fractions associated with them, the aggregates must be long, perhaps even approaching or exceeding the rigid

persistence length of the DNA duplex. And, they are expected to be in a 2D hexagonal lattice perpendicular to the long axis of the aggregate.

In the detailed experimentation performed here, three samples were subjected to repetitious WAXS with increasing temperature. As with the NEM\* phase (Section 2.2.1), these samples were not oriented domains and gave powder scattering (see Figure 2.13). Typical room temperature 1D circular averages of these samples appear in Figure 2.14. The loss of data by the powder averaging complicates the identification of hexagonal lattice packing because it obliterates evidence of the angle between the first order reflections, making it impossible to directly see if they are on a hexagonal lattice –there are six such reflections, but all six are at the same radius from beam center, allowing them to wash each other out in the powder average. As with the NEM\* phase, such a solitary 1D reflection may actually only imply a pairwise correlation between neighboring aggregates and not set them on a lattice.

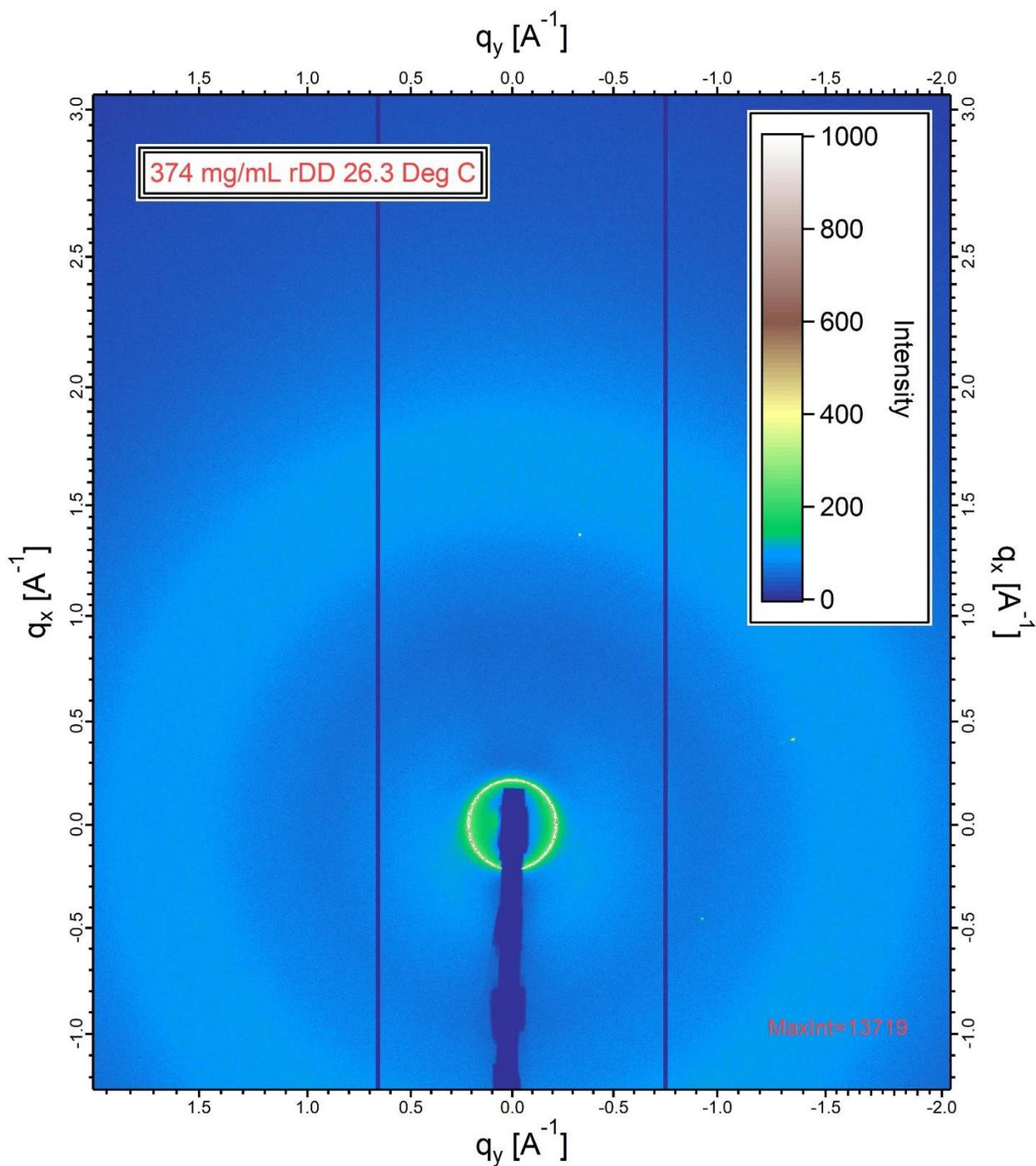


Figure 2.13: rDD COL mesophase typical WAXS. Taken from 373.6 mg/mL rDD experiments, this 2D image exemplifies data obtained from these samples prior to 1D data reduction by circular average. Shadowed features are again detector seams and the X-ray beam stop. Scattered intensity counts are depicted by an altitude map, as shown in the upper right. As with NEM\* data (Figure 2.7), the 13,000 count bright spots are from scattering by the mica window in the hot stage.

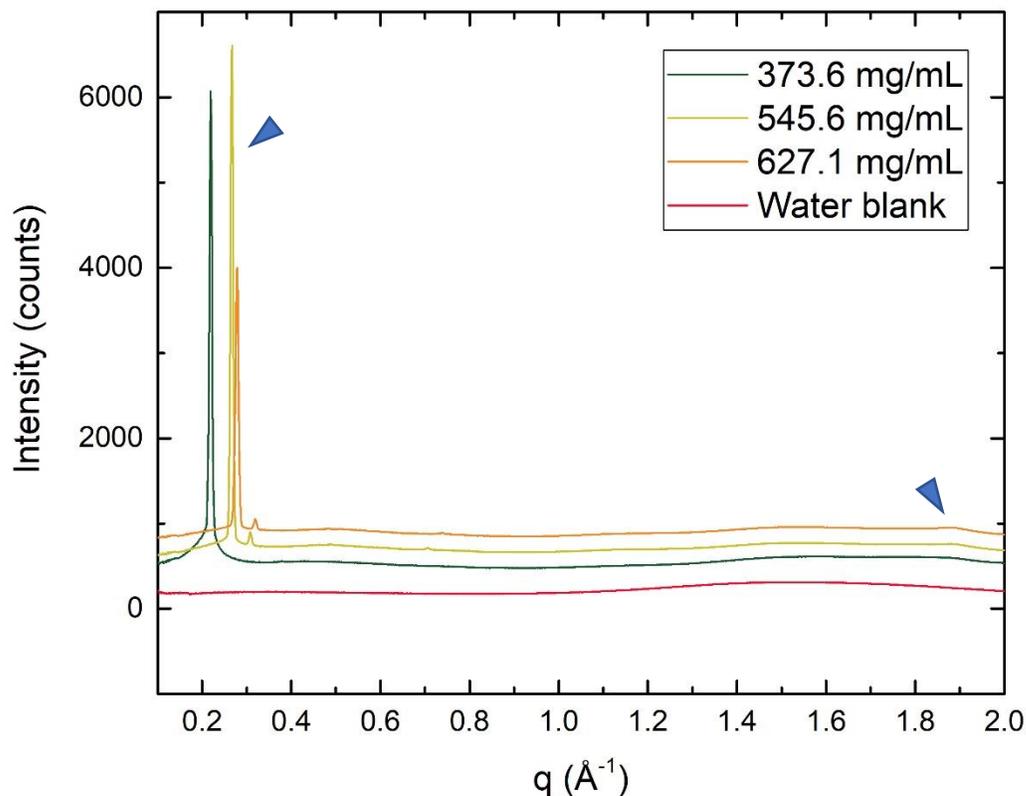


Figure 2.14: Collected rDD COL mesophase samples. This plot shows a comparison of all unambiguous COL samples as labeled by concentration; water is included to show the approximate baseline. The blue triangles are added to draw attention to the most major COL features, the mesophase packing peaks located between 0.2 and 0.35  $\text{\AA}^{-1}$  and the base-pair stacking peak at 1.85  $\text{\AA}^{-1}$ .

Figure 2.14 shows typical behavior of the COL phase in WAXS. In some of these samples, the packing peak was accompanied by a small, slightly higher  $q$  reflection of ambiguous provenance. This additional peak does not seem to accompany alterations of the COL texture and even vanished in later observation of the 627.1 mg/mL rDD sample, apparently placing higher order mesophase structure at a lower concentration some of the time since 545.6 mg/mL rDD retained the peak even though 627.1 mg/mL lost it. COL phase demonstrating this extra reflection peak can be distinguished as COL+. Because of the required length of the rDD

aggregates in this mesophase, as this is even more concentrated than the NEM\* phase, the COL packing reflection periodicity is again from side-side correlations at the very least without considering other evidence. The base-stacking peak is once more just a shoulder at  $1.85 \text{ \AA}^{-1}$  which appears to strengthen with increasing concentration while not appearing to shift in  $q$ .

Despite appearing initially as barren as the NEM\* plots, the pure COL phase has some additional scattering features on closer examination.

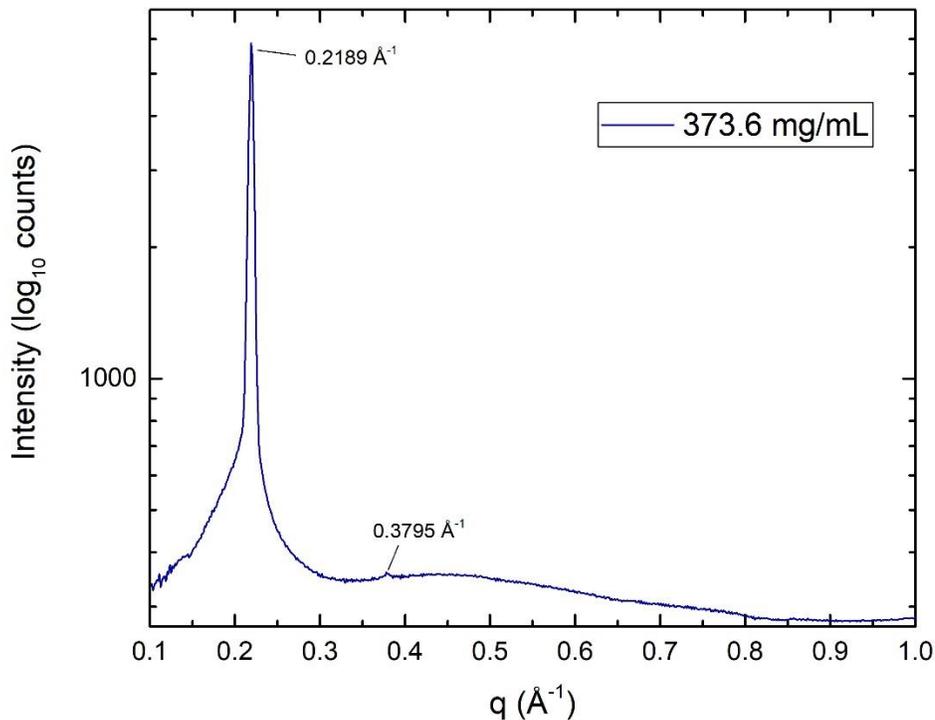


Figure 2.15: Scattering feature used for Hexagonal confirmation in COL phase. For the pure columnar phase (without the COL+ feature), a weak higher order reflection does appear in the circular average.

The pure COL phase X-ray scattering appears to contain a single weak higher order peak that can be used to support the hexagonal packing hypothesis of the columnar lattice given some modeling of the scattering pattern using concepts

put forward in Section C.2.1. These peaks can be assigned respectively as the hexagonal  $q_{11}$  and  $q_{31}$  peaks (see Section C.4.1).

Using the technique introduced in equation C.55 on the peaks extracted from Figure 2.15, one quickly discovers that  $0.2189/0.3795 = 0.5768 \cong 1/\sqrt{3}$  or hexagonal to within the noise of the peak selection. rDD COL phase is consistent with a hexagonal phase given that there are no other peaks beyond the base-stacking peak. Using one or the other equations from Section C.4.1, the real hexagonal spacing parameter of the 373.6 mg/mL rDD COL phase as determined from the initial room temperature scan in Figure 2.15 is 33.1 Å.

A similar technique used to address square lattices shows that the COL+ feature is not a square periodicity ( $1/\sqrt{2}$ ); both 545.6 mg/mL and 627.1 mg/mL rDD with the COL+ reflection are not some square lattice polymorphism of columnar phase. Both such phases possess very weak hexagonal reflections behind several other weak higher order peaks (see Figure 2.16).

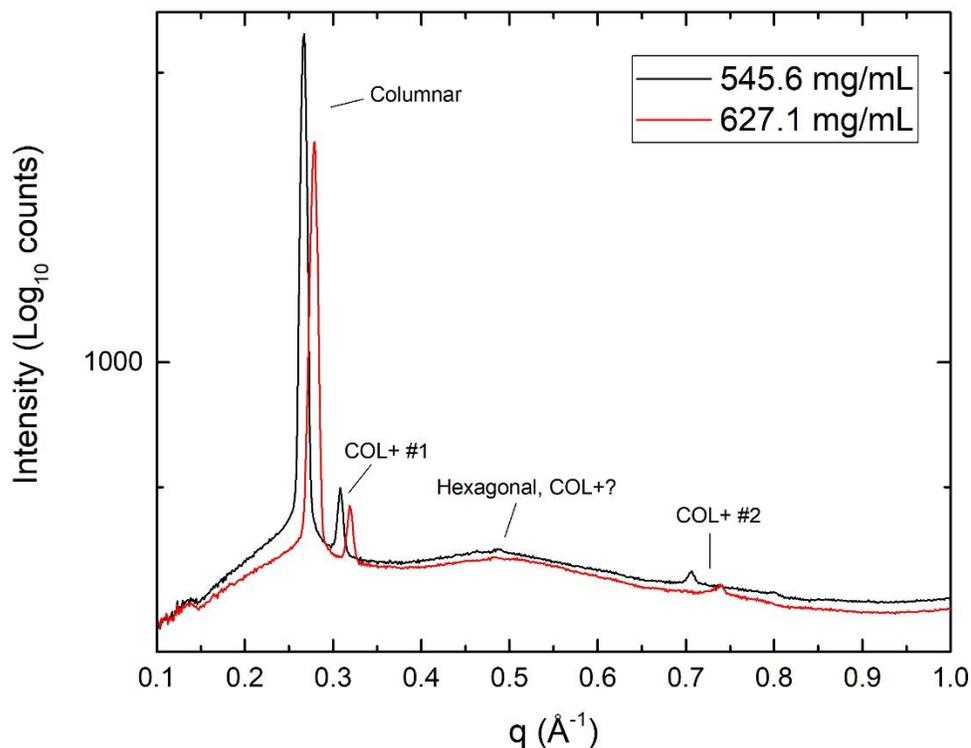


Figure 2.16: WAXS scattering of rDD COL+ columnar phase. COL+ phase demonstrates several additional peaks, of which COL+ peak #1 and peak #2 are the clearest. A third peak may sit near the second order hexagonal reflection and several other peaks may trail to higher angle  $q$ -values. COL+ peaks appear to compress with increasing rDD concentrations in concert with the central columnar periodicity peak, suggesting that they are in the 2D lattice. (zz, adjust nomenclature of this figure)

COL+ additional peaks do not follow any clear lattice periodicities by themselves and do not represent hexagonal or square periodicities of the main COL peak. It's possible that these peaks are higher order reflections for a lattice at very small  $q$ -values, representing some large scale in-plane density undulation or systematic omission, but there is no evidence to back up these hypotheses. The fact that rDD 627.1 mg/mL initially demonstrated COL+ scattering, but later demonstrated only COL, even while rDD 545.6 mg/mL continued to display COL+,

suggests that the energy barrier between COL and COL+ is small and that minimal perturbation to the sample can eliminate the COL+ feature.

Temperature ramping experiments of these mesophase samples reveal a series of other interesting behaviors.

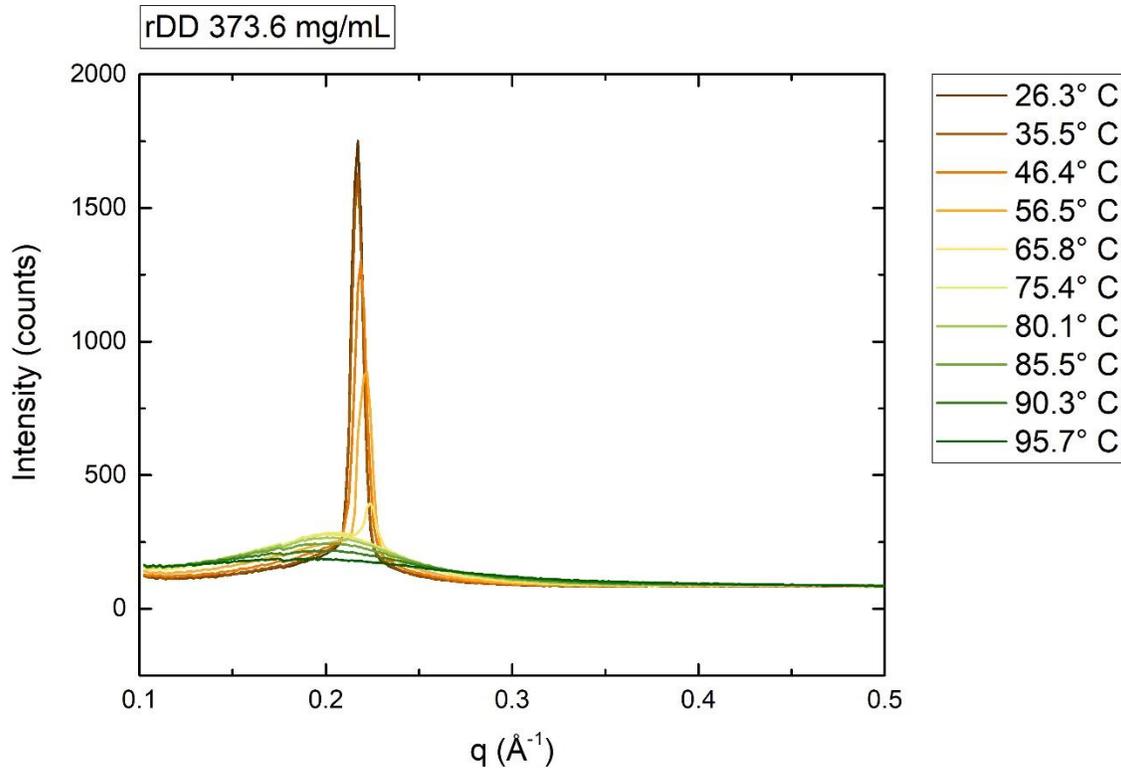


Figure 2.17: rDD 373.6 mg/mL temperature ramping experiment. Shown are the collected circular averages of WAXS from low concentration COL phase. Qualitatively, the COL peak shifts to higher  $q$  during the increase in temperature and then transitions to a broad, lower  $q$  peak.

During initial observations, 373.6 mg/mL rDD possesses a peak with a flared base, suggesting two overlapping peaks. Peak fits of Figure 2.15 give two Lorentzians: the sharp columnar peak at  $0.219 \text{ \AA}^{-1}$  and a weaker peak at  $0.1869 \text{ \AA}^{-1}$ , more in line with the NEM\* phase exhibited in the previous section but with a coherence length of  $28 \text{ \AA}$ . This initial fit is wrought with uncertainty due to the baseline difficulties near the beam stop, but 373.6 mg/mL rDD appears to be at or

near NEM\*/COL coexistence. In Figure 2.17, the COL peak diminishes with increasing temperature until a second, broader peak at smaller  $q$  rises up to overtake it. This second peak then diminishes as well. Of note, the phase retains birefringence by PLM across the period where this peak appears but is visibly melting.

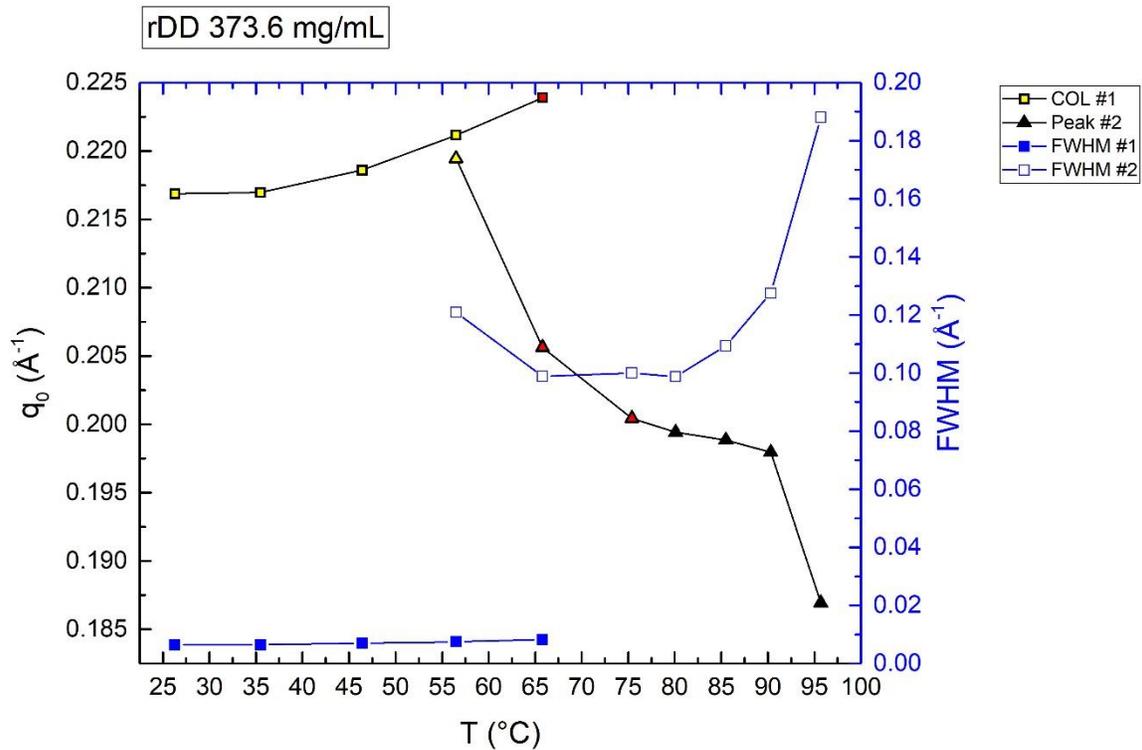


Figure 2.18: Peak fitting parameters for rDD 373.6 mg/mL during temperature ramp. Black lines denote measurements of peak center for peak #1 (Black squares), the COL spacing peak, and peak #2 (Black triangles), the melting peak; the black axis on the left measures the  $q$ -values for both. Blue lines represent the FWHM of both peaks and are measured by the blue axis on the right. Peak center plots are augmented by colored markers to depict the behavior of mesophase birefringence under PLM through the experiment: COL (Yellow), coexistence (Red) and ISO (Black).

Peak fitting parameters can again be converted to separation lengths in real space and structural coherence lengths by equations C.42, C.56 and C.50.

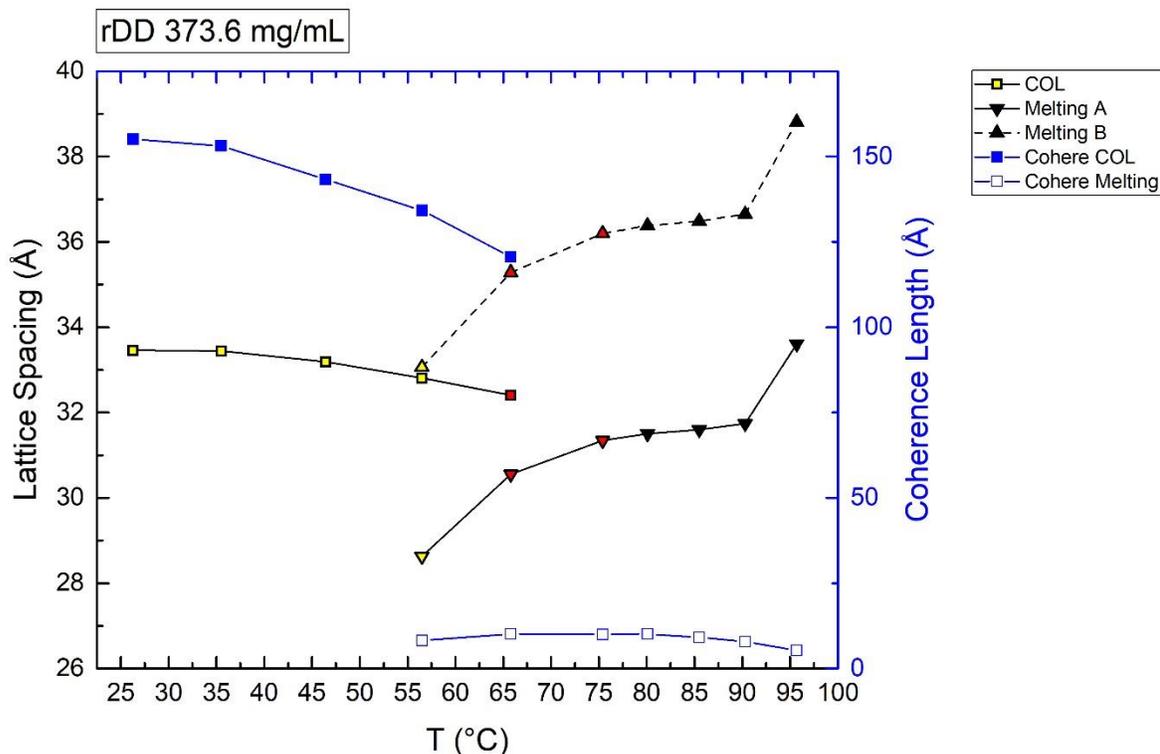


Figure 2.19: Lattice Spacing and Coherence Length vs. Temperature for rDD 373.6 mg/mL. This plot shows lattice spacing from the axis on the left (Black) and Coherence Length (Blue) from the axis on the right. The two peaks considered are the hexagonal lattice spacing parameter (equation C.56) (Black line, Squares) and the melted phase side-side packing (solid Black line Up Triangles and dashed Black Line Down triangles); the markers are filled with color to depict PLM birefringence observations at each given temperature, COL (Yellow), Coexistence (Red) and ISO (Black). Coherence length for COL (Filled Blue Squares) and coherence length for melted phase (Unfilled Blue Squares) are as shown. Spacing is calculated differently for the COL phase versus the Melting phase given the differences in coherence length: spacing for COL is the direct lattice spacing parameter from Figure 2.15, while Melting A phase (solid Black line with down triangles) is calculated as a 1D pairwise interaction (equation C.42) as seen with NEM\* in Section 2.2.1. Melting B (dashed Black line with up triangles) is included to show how the spacing of the melting phase appears if calculated as columns on a hexagonal lattice.

The COL phase has a lattice spacing close to 3.35 nm which gradually compresses with increasing temperature until it is 3.25 nm. This phase has a fairly large structural coherence length equal to a side-side distance of about 4.5 column widths (15 nm). The phase then melts beginning at 55° C to a transitory phase

which looks like a NEM\* in spacing that possesses a coherence length of about 1.2 nm, suggesting only side-side correlation, much like NEM\* seen in Section 2.2.1, and completely melts to ISO by 70° C.

Calculation of lattice spacings by these means creates an interesting conundrum. The coherence length of the melting phase is clearly on the order of or smaller than the average side-side spacing of aggregates in that phase, while the COL phase is larger. For the COL phase, the closest side-side spacing of aggregates in their hexagonal lattice is about 3.3 nm. The melting phase is clearly not coherent enough to be obviously on a lattice, but by calculating it as purely a side-side correlation (equation C.42), the average spacing between aggregates abruptly decreases to smaller than the COL phase spacing upon crossing the phase transition. If the melting phase continues to reside on a hexagonal lattice, albeit a lattice so disordered that its fluctuations in column spacing are on the size scale of a unit cell over the length of less than a unit cell, then the spacing is only ever larger than the spacing of the COL phase (dashed line, Figure 2.19). This seems intuitively self-contradictory. On the other hand, it may provide a hypothesis for understanding the entropic pressure toward forming COL phase: by dropping onto the hexagonal lattice, side-side spacing is assured to be larger between any two aggregates on average than it would be if the aggregates are randomly distributed, thus preserving some translational entropy upon adopting the lattice. For the melting condition, the reality is probably some mixture of the two cases, partially hexagonal at the beginning, but losing order as melting proceeds.

It is expected that other interesting behavior might emerge in the case where these observations are repeated with gradual cooling rather than gradual heating, but such measurements have not been made. Lyotropic LC has the complication of having to deal with a 2D phase space, where there's a concentration axis in addition

to a temperature axis, which majorly increases the time required for a given piece of work.

Similar experiments were performed with both COL phases that initially showed COL+ type scattering. As mentioned, 545.6 mg/mL rDD remained COL+ at the time of the experiment, but 627.1 mg/mL rDD appeared as only COL.

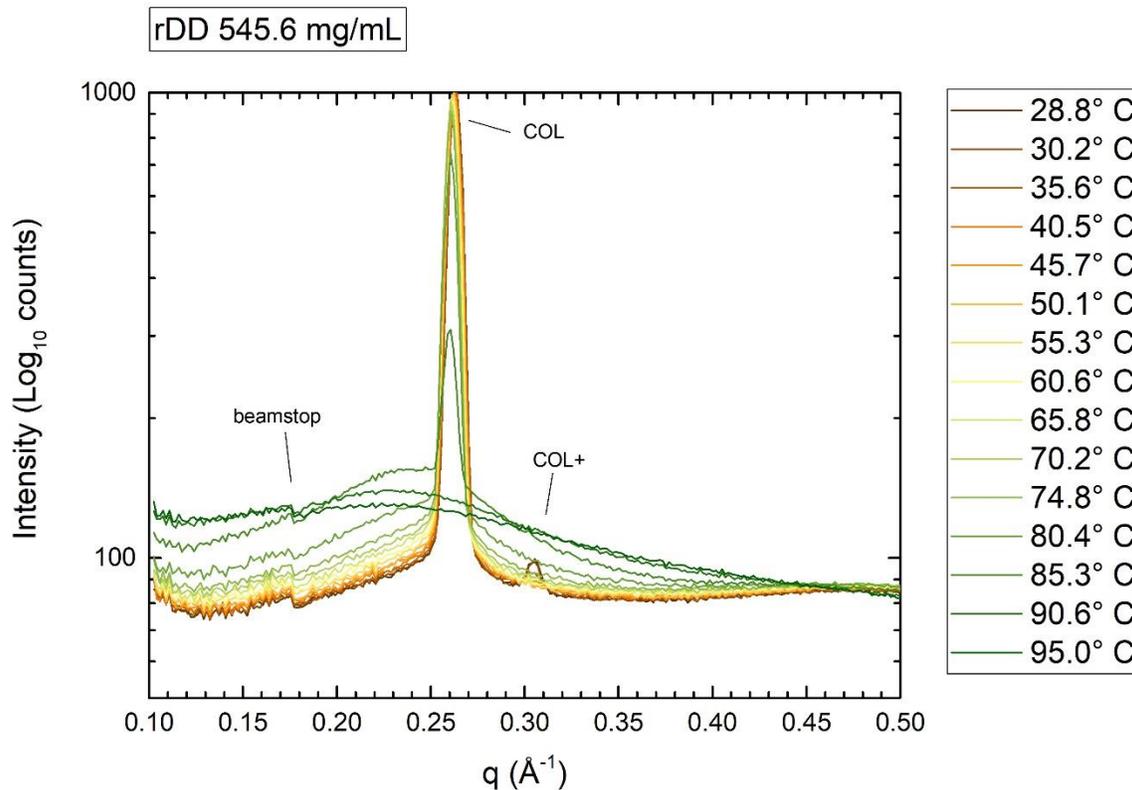


Figure 2.20: rDD 545.6 mg/mL temperature ramping experiment. Collected circular averages for all shots in intensity (Log<sub>10</sub> counts) vs.  $q$  (Å<sup>-1</sup>). COL peak and COL+ are noted as is a feature related to clipping of the data near the beam stop. COL peak appears to shift left in  $q$  with increasing temperature and ultimately drops into a broad peak at high temperature. COL+ vanishes above 35° C.

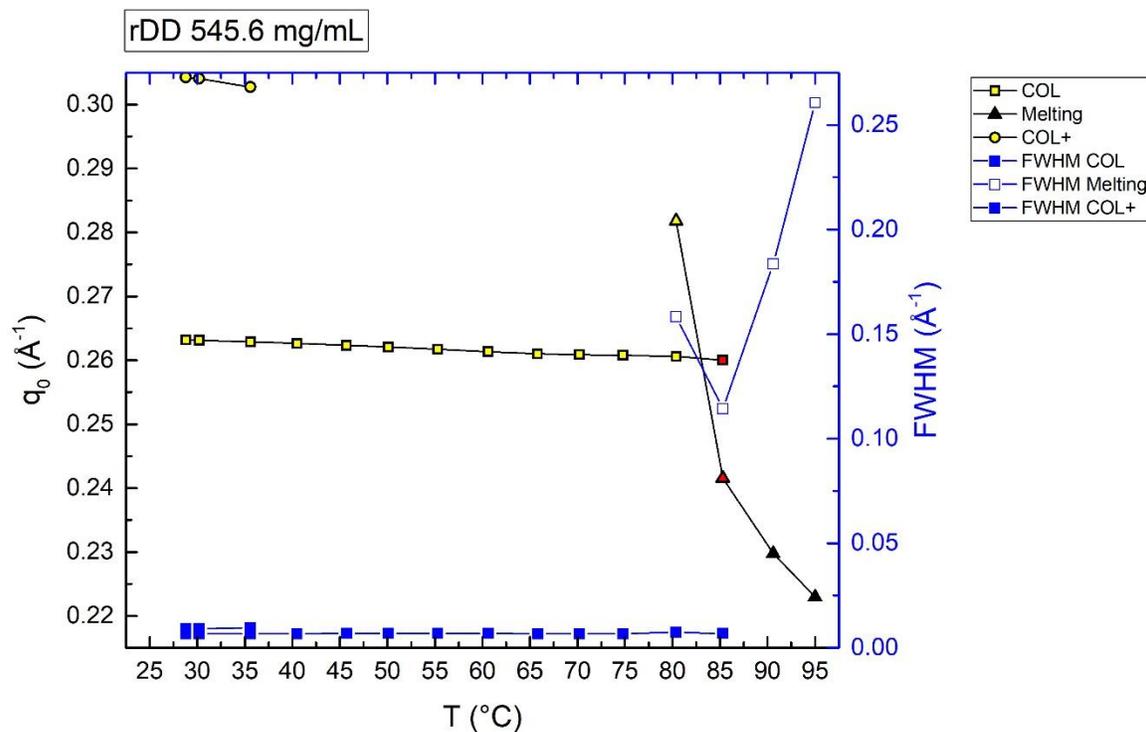


Figure 2.21: Peak fitting parameters for rDD 545.6 mg/mL during temperature ramp. Black lines denote measurements of peak center for COL peak (Squares), COL+ peak (Circles) and melting peak (Triangles); the black axis measures the  $q$ -values for all three. Blue lines represent the FWHM of all peaks and are measured by the blue axis on the right. Peak center plots are augmented by colored markers to depict the behavior of mesophase birefringence under PLM through the experiment: COL (Yellow), coexistence (Red) and ISO (Black).

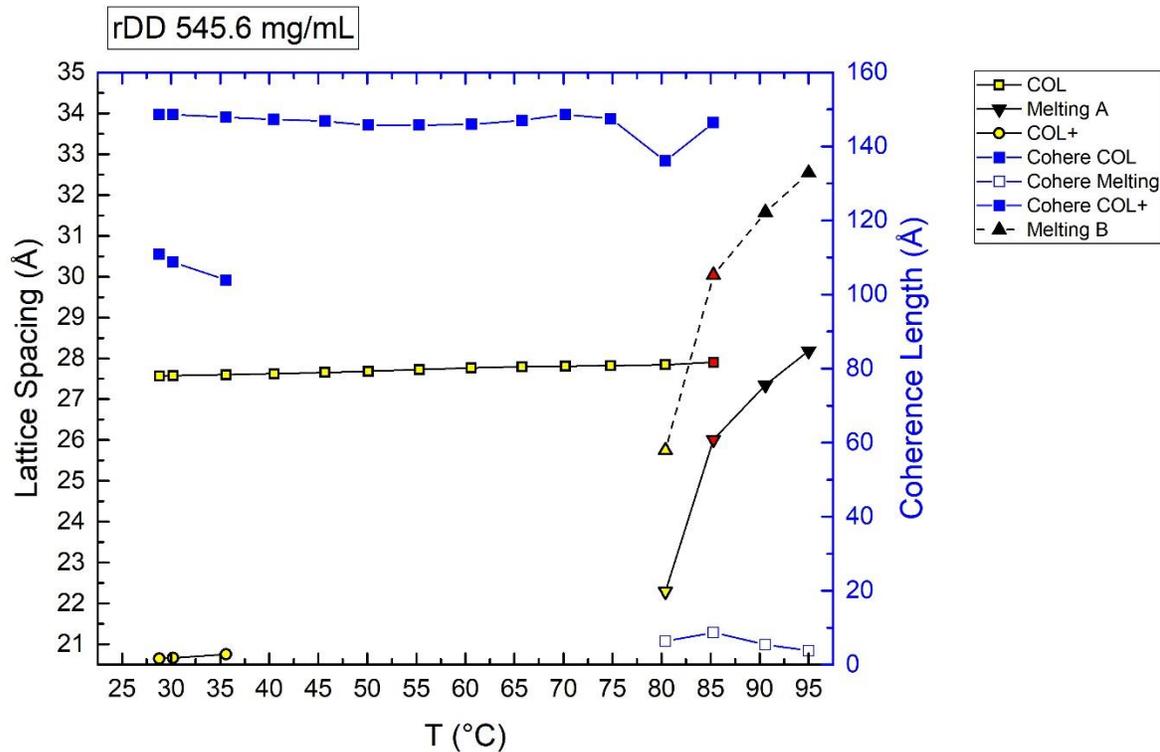


Figure 2.22: Lattice Spacing and Coherence Length vs. Temperature for rDD 545.6 mg/mL. This plot shows lattice spacing from the axis on the left (Black) and Coherence Length (Blue) from the axis on the right. The three peaks considered are the hexagonal lattice spacing parameter (see Section C.4.1) (Black line, Squares), the COL+ peak (Black line, Circles) and the melted phase side-side packing (solid Black line Up Triangles and dashed Black Line Down triangles); the markers are filled with color to depict PLM birefringence observations at each given temperature, COL (Yellow), Coexistence (Red) and ISO (Black). Coherence length for COL and COL+ (Filled Blue Squares) and coherence length for melted phase (Unfilled Blue Squares) are as shown. Spacing is calculated differently for the COL phase versus the COL+ and Melting phase: spacing for COL is the direct lattice spacing parameter from Figure 2.15, while COL+ and Melting A phase (solid Black line with down triangles) are calculated as 1D pairwise interaction as seen with NEM\* in Section 2.2.1 since COL+ is not known to be on a lattice and Melting phase has a short coherence length. Melting B (dashed Black line with up triangles) is included to show how the spacing of the melting phase appears if calculated as columns on a hexagonal lattice.

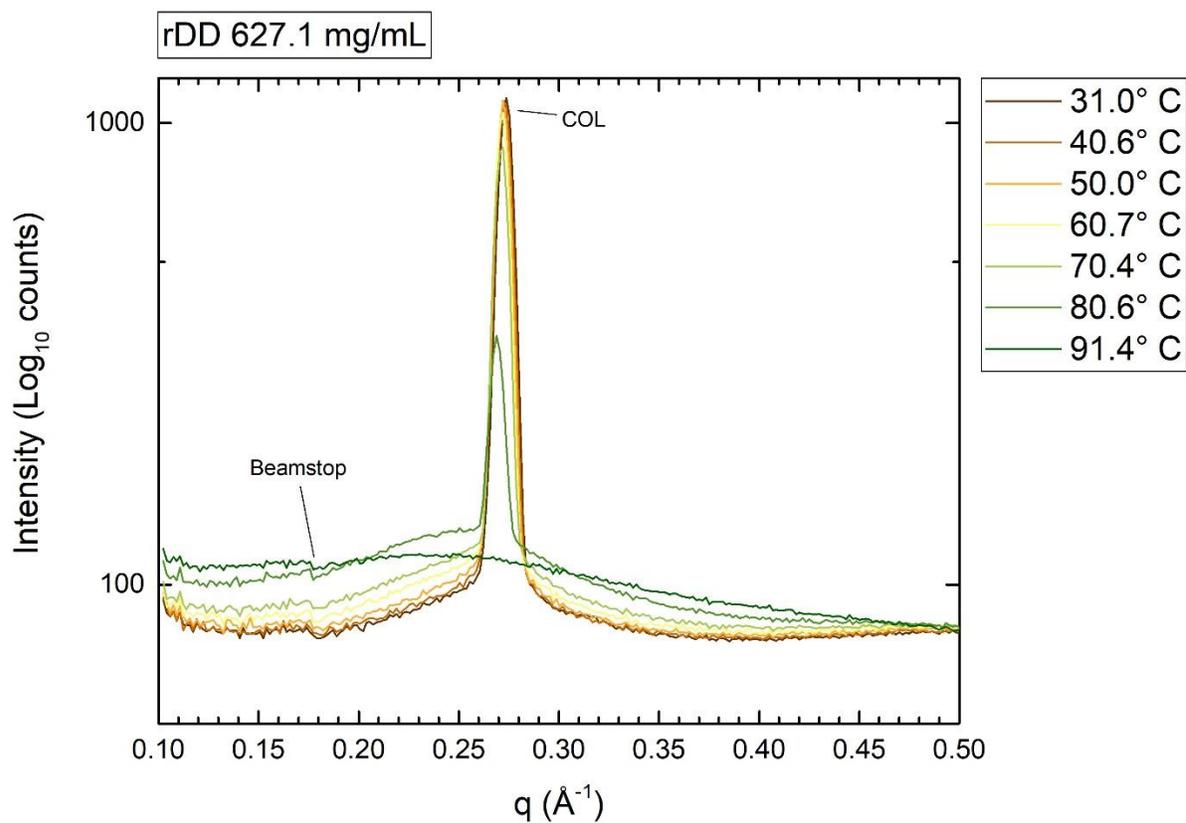


Figure 2.23: rDD 627.1 mg/mL temperature ramping experiment. Collected circular averages for all shots in intensity ( $\text{Log}_{10}$  counts) vs.  $q$  ( $\text{\AA}^{-1}$ ). COL peak is illustrated as a feature related to clipping of the data near the beam stop. COL peak appears to shift left in  $q$  with increasing temperature and ultimately drops into a broad peak at high temperature. COL+ does not appear in this sample during these observations; compare with Figure 2.16 for an earlier observation of the same sample.

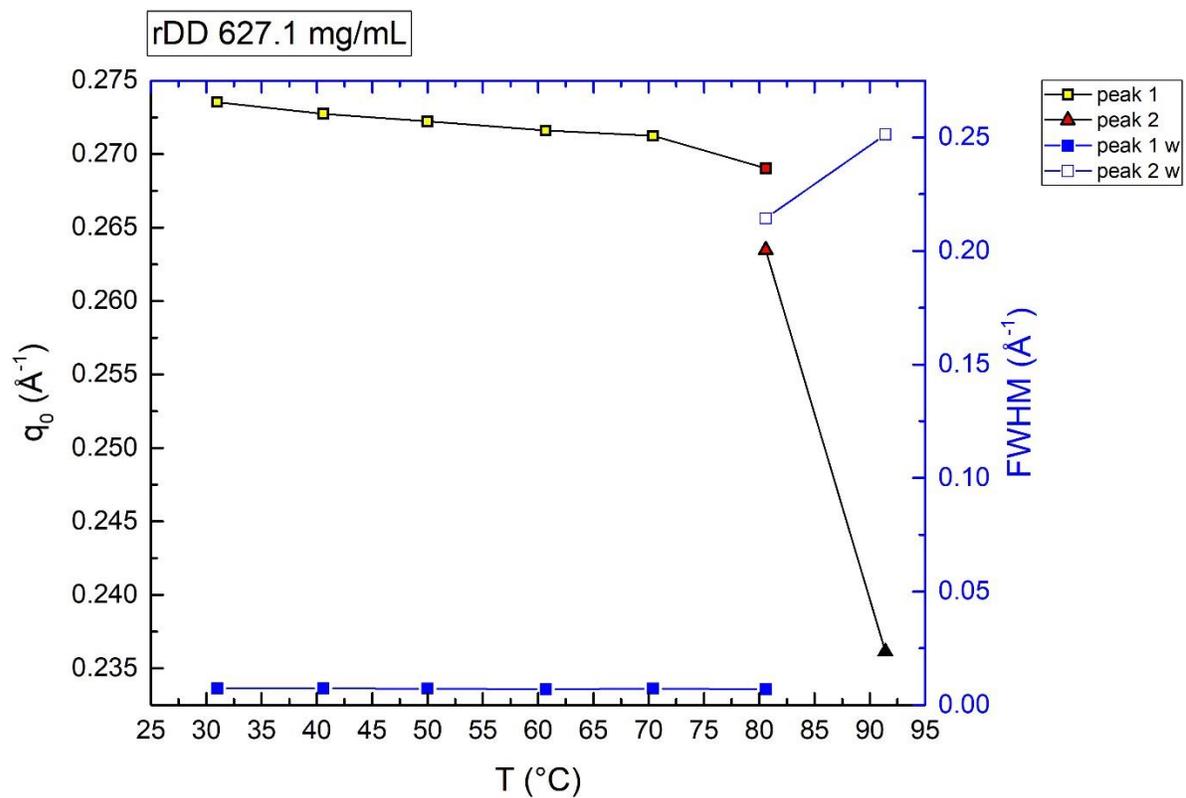


Figure 2.24: Peak fitting parameters for rDD 627.1 mg/mL during temperature ramp. Black lines denote measurements of peak center for COL peak (Squares) and melting peak (Triangles); the black axis measures the  $q$ -values for all three. Blue lines represent the FWHM of all peaks and are measured by the blue axis on the right. Peak center plots are augmented by colored markers to depict the behavior of mesophase birefringence under PLM through the experiment: COL (Yellow), coexistence (Red) and ISO (Black).

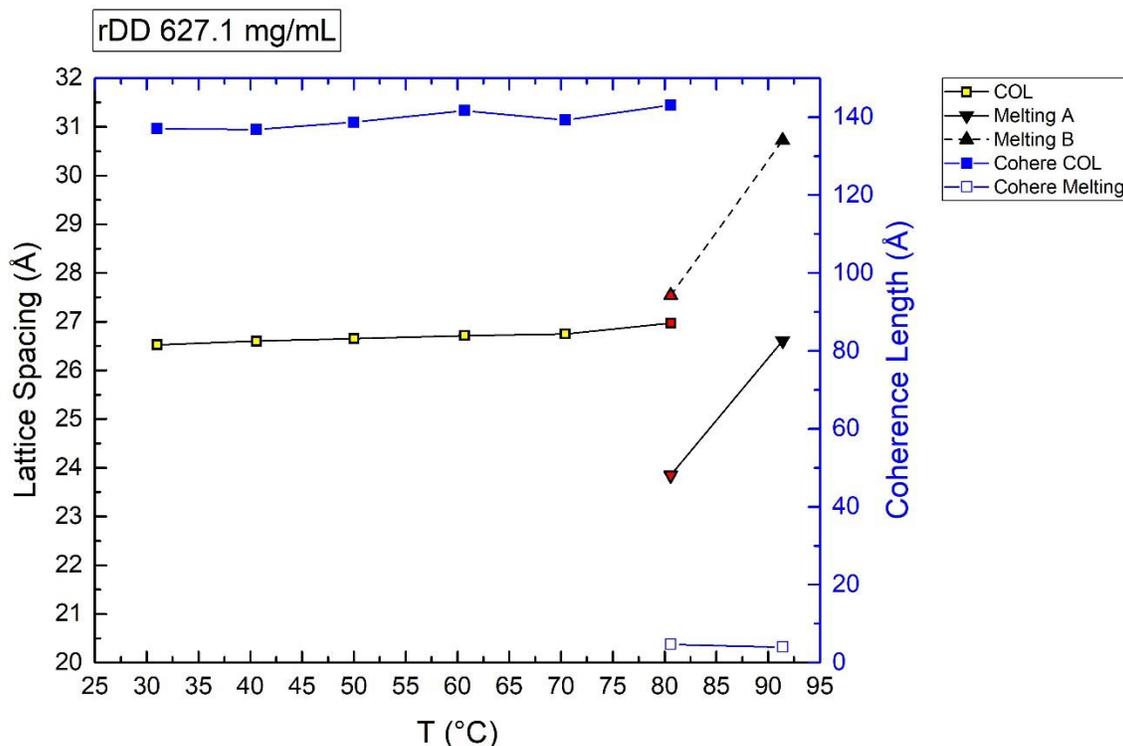


Figure 2.25: Lattice Spacing and Coherence Length vs. Temperature for rDD 627.1 mg/mL. This plot shows lattice spacing from the axis on the left (Black) and Coherence Length (Blue) from the axis on the right. The peaks considered are the hexagonal lattice spacing parameter (see Section C.4.1) (Black line, Squares), and the melted phase side-side packing (solid Black line Up Triangles and dashed Black Line Down triangles); the markers are filled with color to depict PLM birefringence observations at each given temperature, COL (Yellow), Coexistence (Red) and ISO (Black). Coherence length for COL (Filled Blue Squares) and coherence length for melted phase (Unfilled Blue Squares) are as shown. Spacing is calculated differently for the COL phase versus Melting phase: spacing for COL is the direct lattice spacing parameter from Figure 2.15, while Melting A phase (solid Black line with down triangles) is calculated as 1D pairwise interaction as seen with NEM\* in Section 2.2.1 since this phase a short coherence length. Melting B (dashed Black line with up triangles) is included to show how the spacing of the melting phase appears if calculated as columns on a hexagonal lattice.

Figures 2.20, 2.21 and 2.22 cover the analysis of rDD 545.6 mg/mL while Figure 2.23, 2.24 and 2.25 cover the same analysis of rDD 627.1 mg/mL. COL phases of rDD all have similar behaviors in that they appear to have a hexagonal columnar lattice that phase transitions to a transitory NEM\*-like phase at high temperatures before reaching ISO phase. The phase transition  $T_c$  pushes toward

higher temperatures with increasing concentration but seems to plateau above 545 mg/mL at between 80° and 85° C for this mesophase. 627 mg/mL is reported in the X-ray data to have a slightly lower  $T_c$  before going to ISO than 545, but this may simply be in the noise of the observation; the 545 mg/mL data set contains a better resolution due to time constraints at the synchrotron. On PLM, both samples phase transition at very close temperatures.

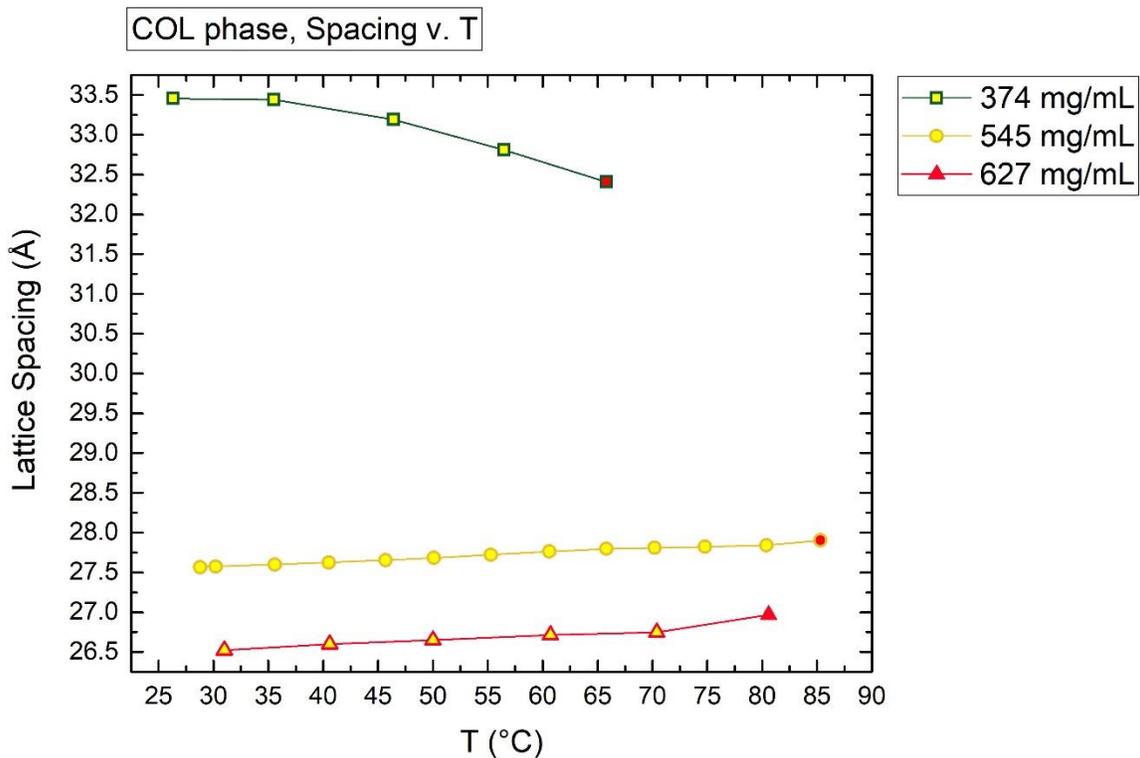


Figure 2.26: Comparison of COL phase Lattice spacing v. Temperature. Lattice spacing from 373.6 mg/mL (Squares), 545.6 mg/mL (Circles) and 627.1 mg/mL (Triangles) are plotted on a common axis. Data is taken from previous figures where the fill color reflects the observed birefringence behavior, COL texture (Yellow) and coexistence (Red). The call of coexistence in the 627.1 mg/mL data is in the noise of the set slightly higher than where it is depicted on the plot due to a lack of data points between coexistence and ISO.

COL phases of different concentrations have a different response to temperature increases. For low concentration COL, the increase in temperature results in a compression of the phase to smaller separations. High concentration

COL, on the other hand, tends to expand spacing with increasing temperature. This suggests a mismatch between changes in internal pressure of the mesophase depending on concentration; previous observations of COL mesophase showed a continuous COL-ISO coexistence where small domains of COL are immersed in a surrounding milieu of ISO such that the volume fraction of COL in the coexistence increases with increasing rDD concentration<sup>25</sup>. Here, for low concentrations of COL, the pressure of the mesophase increases more slowly than the pressure of the surrounding ISO, allowing the ISO to compress the spacing of the COL lattice as the temperature increases. At high concentrations, the pressure inside the mesophase increases more quickly with temperature than the pressure of the surrounding ISO, allowing the phase to resist compression and enabling it to expand. Across the temperature range examined, roughly room temperature to just below water's boiling point, this shift in volume amounts to about 6% or less for the assumed cylindrical packing. This particular observation is from a set of three concentrations; the details of the compression-expansion behavior and the position of the rollover from compression to expansion requires more examination.

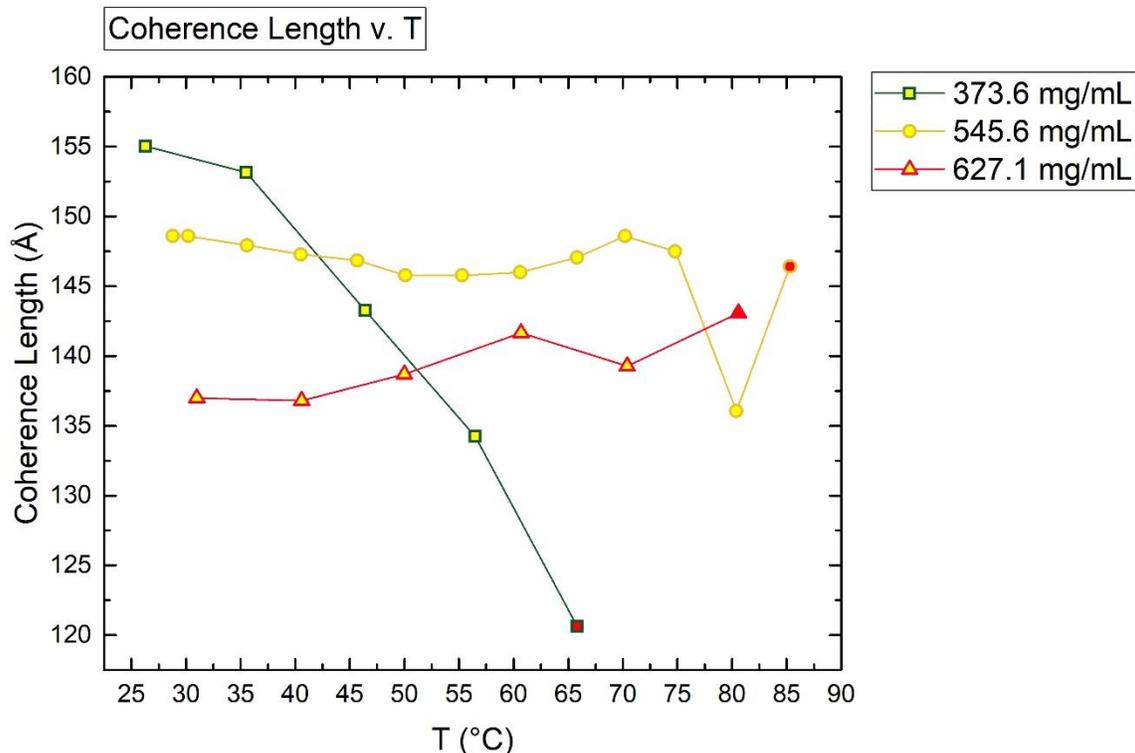


Figure 2.27: Structural Coherence Length v. Temperature Comparison. Coherence length from 373.6 mg/mL (Squares), 545.6 mg/mL (Circles) and 627.1 mg/mL (Triangles) are plotted on a common axis. Data is taken from previous figures where the fill color reflects the observed birefringence behavior, COL texture (Yellow) and coexistence (Red). The call of coexistence in the 627.1 mg/mL data is in the noise of the set slightly higher than where it is depicted on the plot due to a lack of data points between coexistence and ISO.

Coherence length (see Figure 2.27) also shows some interesting behavior across the range of COL concentrations. At low concentrations, coherence starts high and then decreases with increasing temperature. The opposite seems to happen at high concentrations, the coherence starts low and actually appears to increase with increasing temperature. The reasons for this are unclear. One possibility may be due to the viscosities of these mesophases: for rDD and other nanoDNA, high concentration mesophases tend to have very high viscosity which appears to increase with increasing concentration. This could affect order by restricting the rate at which these mesophases coarsen, decreasing the rate at

which aggregates in a phase move around with increasing concentration. For a low concentration COL, relatively lower viscosity would allow the phases to order more quickly, giving the chance to happen upon more uniformly ordered lattices more frequently. The opposite would hold for high concentration phases with relatively higher viscosity: ordering within a phase would happen more slowly, preventing the mesophase from finding uniformity as quickly. This effect could then cut the opposite direction as temperature is elevated: the low concentration COL is not stuck in place and rattles apart with little encouragement, while the high concentration COL undergoes an annealing of sorts by rattling aggregates into a more uniform lattice before breaking apart. “Breaking apart” in this case may be duplex DNA melting into single strands instead of the high concentration COL phase itself melting; this subtlety has not been addressed and is not clearly addressed from the base-stacking peak.

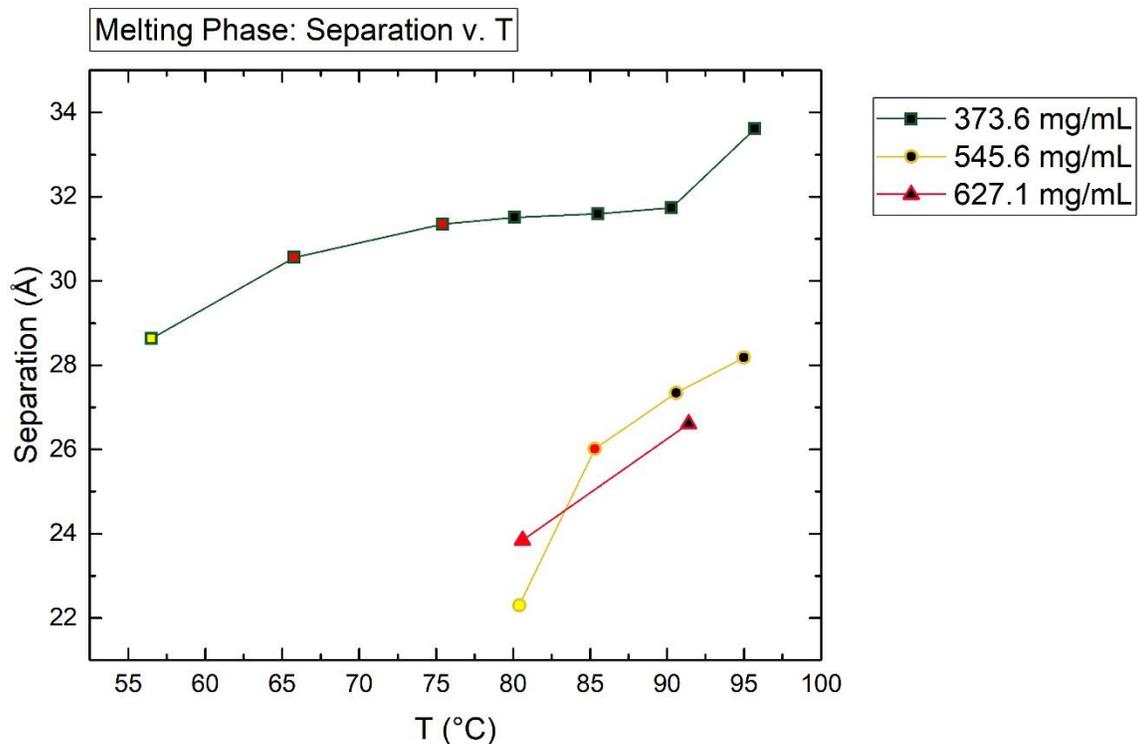


Figure 2.28: Separation length in melting phase versus temperature. Separation length in the melting phase from 373.6 mg/mL (Squares), 545.6 mg/mL (Circles) and 627.1 mg/mL (Triangles) are plotted on a common axis. Data is taken from previous figures where the fill color reflects the observed birefringence behavior, COL texture (Yellow) and coexistence (Red). The call of coexistence in the 627.1 mg/mL data is in the noise of the set slightly higher than where it is depicted on the plot due to a lack of data points between coexistence and ISO.

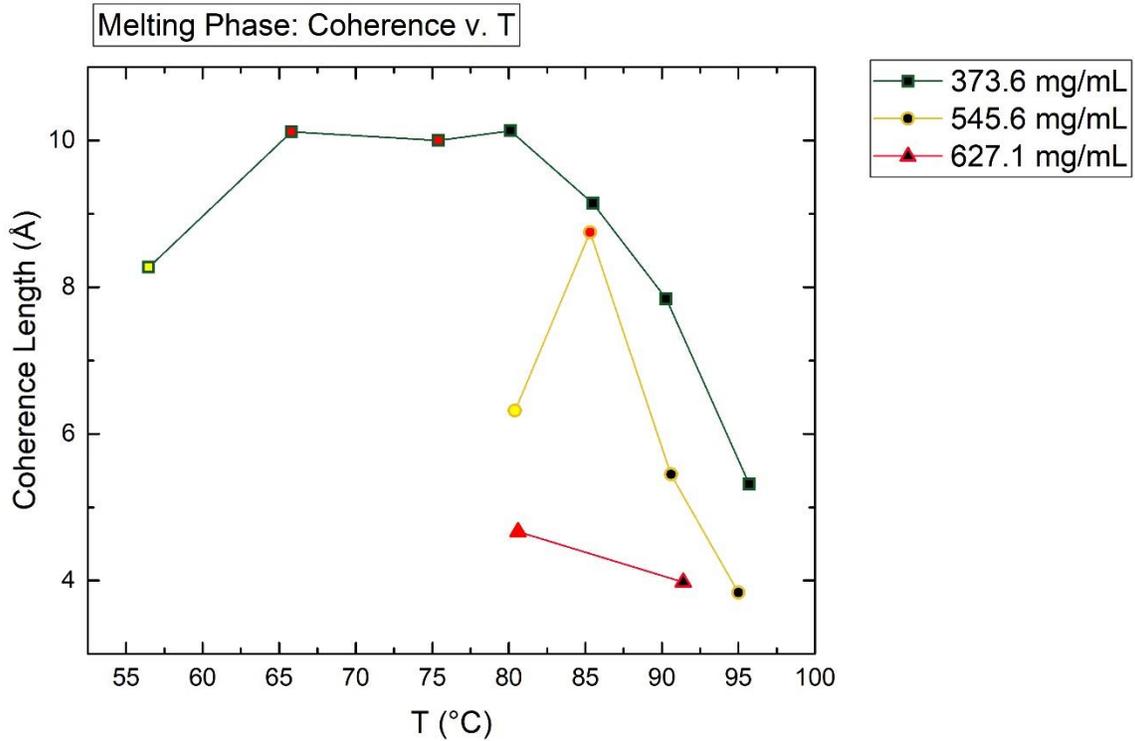


Figure 2.29: Coherence length in melting phase versus temperature. Coherence in the melting phase from 373.6 mg/mL (Squares), 545.6 mg/mL (Circles) and 627.1 mg/mL (Triangles) are plotted on a common axis. Data is taken from previous figures where the fill color reflects the observed birefringence behavior, COL texture (Yellow) and coexistence (Red). The call of coexistence in the 627.1 mg/mL data is in the noise of the set slightly higher than where it is depicted on the plot due to a lack of data points between coexistence and ISO.

Less data is available for the melting phase than for the intact columnar phase (see Figures 2.28 and 2.29) because it exists through much less of the observation space. The melting phase seems to have some characteristics that make it look structurally like a NEM\* and it is referred to as NEM\* in Figure 2.2, but this

could be uncertain. Separation distances for the COL melting phase are short and grow shorter with increasing concentration, but the phase appears in two samples from the curve fitting to jump from lower coherence length to higher coherence length before bleeding away. More data is needed to judge.

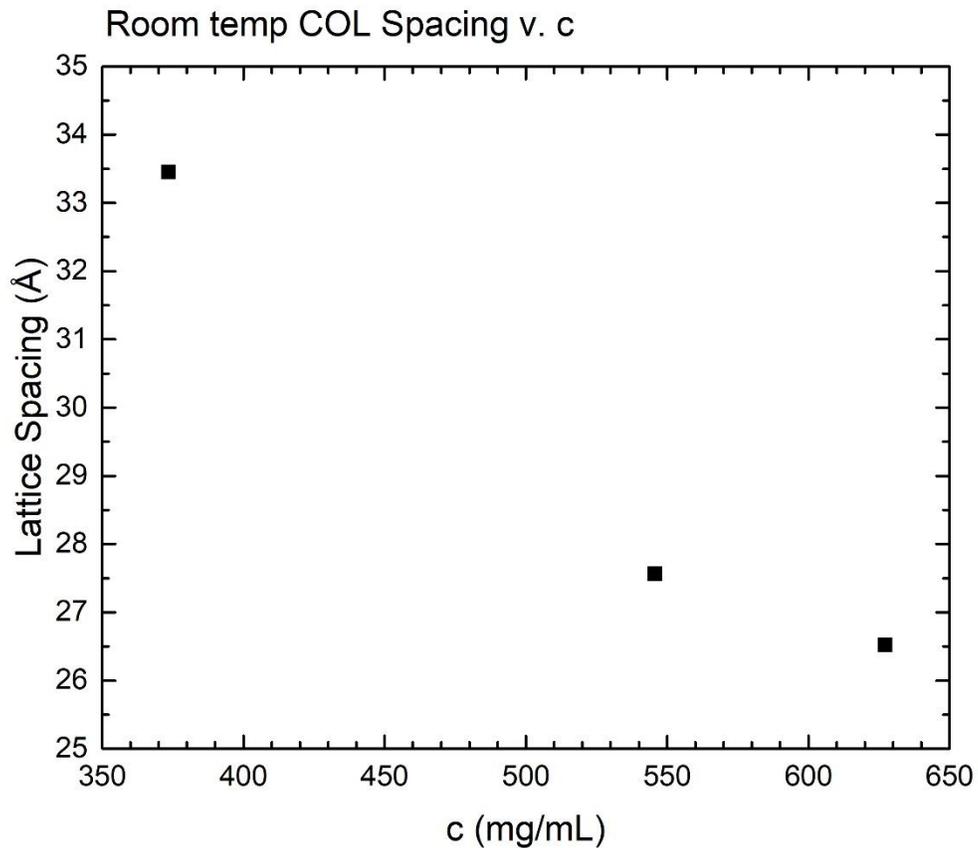


Figure 2.30: rDD COL lattice spacing versus concentration at room temperature. This sums up the relationship of rDD COL lattice spacings to each other given concentration.

The final figure in this section is a summary comparison of the three X-ray samples here to each other in the broadest stroke. This data will be revisited during comparison between all the mesophases of rDD.

### 2.2.3 WAXS of Higher Order Columnar Phases of Reverse Dickerson Dodecamer

The final mesophases examined in detail by WAXS of rDD have been collected under the name “higher order columnar” (COL2) and are the least well understood of the original LC phases reported in earlier work on this subject<sup>18,25,27,30,31,34</sup>. COL2 mesophases are difficult to examine in a stable manner by microscopic cells under PLM because of their high viscosity and high concentration. Prepared samples tend to drift rapidly in concentration because of their relative paucity of water, meaning that a very small shift in water content can lead to a very large shift in concentration. Coupled with the potential for temperature increases to drive water out of a sample, these phases were never reported to accurate concentrations in previous nanoDNA work and not temperature ramped in X-ray diffraction at all.

Higher order mesophase structure in samples of long-stranded DNA is found to be an orthorhombic phase with the unit cell of  $\vec{a} = 24.09 \text{ \AA}$  and  $\vec{b} = 39.33 \text{ \AA}$  or crystalline phases at higher concentration still<sup>24</sup>. For nanoDNA, the exact nature of these mesophases had not been examined by WAXS until this work.

Extreme concentration samples of DNA pose some interesting potential features based upon the detailed structure of the DNA duplex. Low concentration DNA mesophases are mainly explained from a model that maps the DNA duplex as an effective cylinder, when in reality, duplex DNA is very famously a double helix. Within COL mesophase, rDD aggregate columns are packed on a 2D hexagonal lattice where columns are repeated on a period no shorter than about 2.65 nm –if column aggregates are truly 1.9 to 2.0 nm in diameter, columns on this lattice do not approach closer than  $\sim 6.5 \text{ \AA}$ . Featureless columns would fit on this lattice as well as a double helical object of the same diameter. In its duplex form, double-stranded B-form DNA contains a great deal of open volume along the major and

minor grooves of the double helix, by no means a cylinder except in first approximation. It would be expected that if DNA duplexes are crowded within a lyotropic mesophase above some critical concentration, that molecules must press through the cylindrical envelopes of their neighbors into the voids left by major and minor grooves of nearby double helices in order to maximize packing. It seems plausible that such an arrangement could still be liquid crystalline since duplexes invading each other's grooves could be transformed by simultaneous translation and rotation arbitrarily along the screw axis of the helix and still manage the same fit. At what concentration would such a behavior become possible?

Using rDD with a deprotonated molecular weight of 3636 g/mol, a duplex weight of 7272 g/mol, a diameter of 2 nm and a stacking pitch of 0.334 nm per 12 bases, simple arithmetic gives a duplex mass of  $1.21 \cdot 10^{-17}$  mg/rDD molecule and a cylindrical duplex volume of  $1.26 \cdot 10^{-20}$  mL/rDD. This gives a concentration of about 960 mg/mL for the occupation of B-form DNA mass on the footprint of the cylinder occupied by the double helix. Given that the absolute density of DNA is 1687 mg/mL (see Section B.2.2), this is a volume fraction of about 0.56, which is a very high but attainable concentration.

High concentration samples at 764.5 mg/mL and 781.2 mg/mL (see Figure 2.3) were originally created in an effort to sample these very high concentrations and were initially calculated to be approaching 850 mg/mL prior to the revision of the concentration calculation methods. The capillary format prohibits clear observation of the mesophase textures at these concentrations and the domains in these particular samples appear very small and difficult to observe. The sample 764.5 mg/mL showed some birefringent domains that appear similar to COL in some sections of the capillary. It is unknown if these textures are the same as the COL2 textures as originally published. Typical COL2 textures in flat cells, as seen

by PLM, frequently appear as frond-shaped objects with an optic axis that does not vary in direction across the entire domain. Birefringence goes extinct in the whole domain simultaneously if the domain's director is parallel to the polarizer or the analyzer, hinting at very rigid, aligned structure. COL2 does not form focal conic domains in the manner of COL or COL+; the director is always very restricted.

Initial WAXS of these samples revealed scattering patterns that were similar, yet distinct, leading to questions as to whether COL2 might actually be several overlapping mesophases.

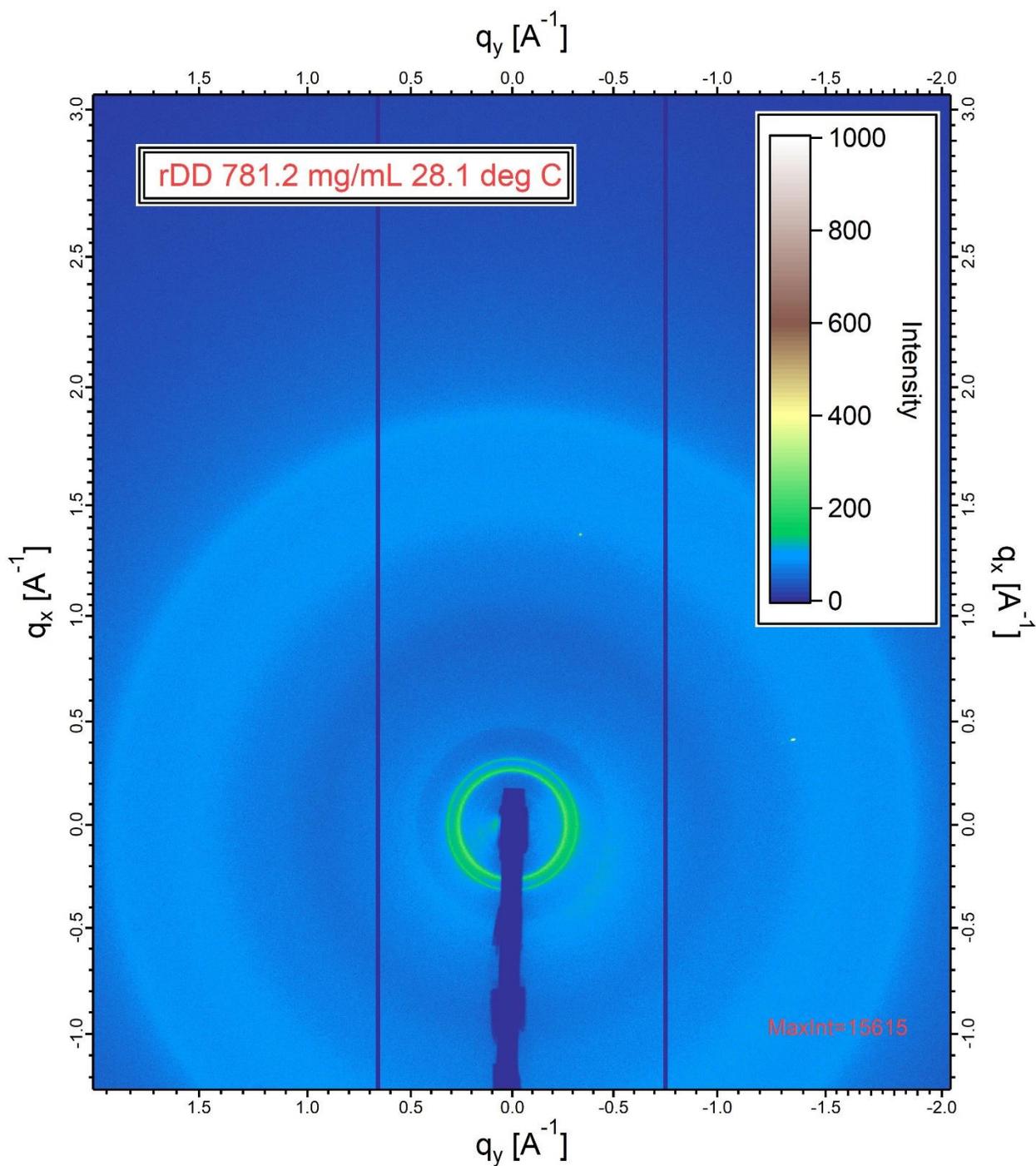


Figure 2.31: WAXS of COL2 from 781.2 mg/mL rDD. This 2D image exemplifies data obtained from similar samples prior to 1D data reduction by circular average. Shadowed features are again detector seams and the X-ray beam stop. Scattered intensity counts are depicted by an altitude map, as shown in the upper right. As with NEM\* and COL data (Figures 2.7 and 2.13), the 15,000 count bright spots are from scattering by the mica window in the hot stage. The DNA base-pair stacking peak is just visible as the outer light blue ring.

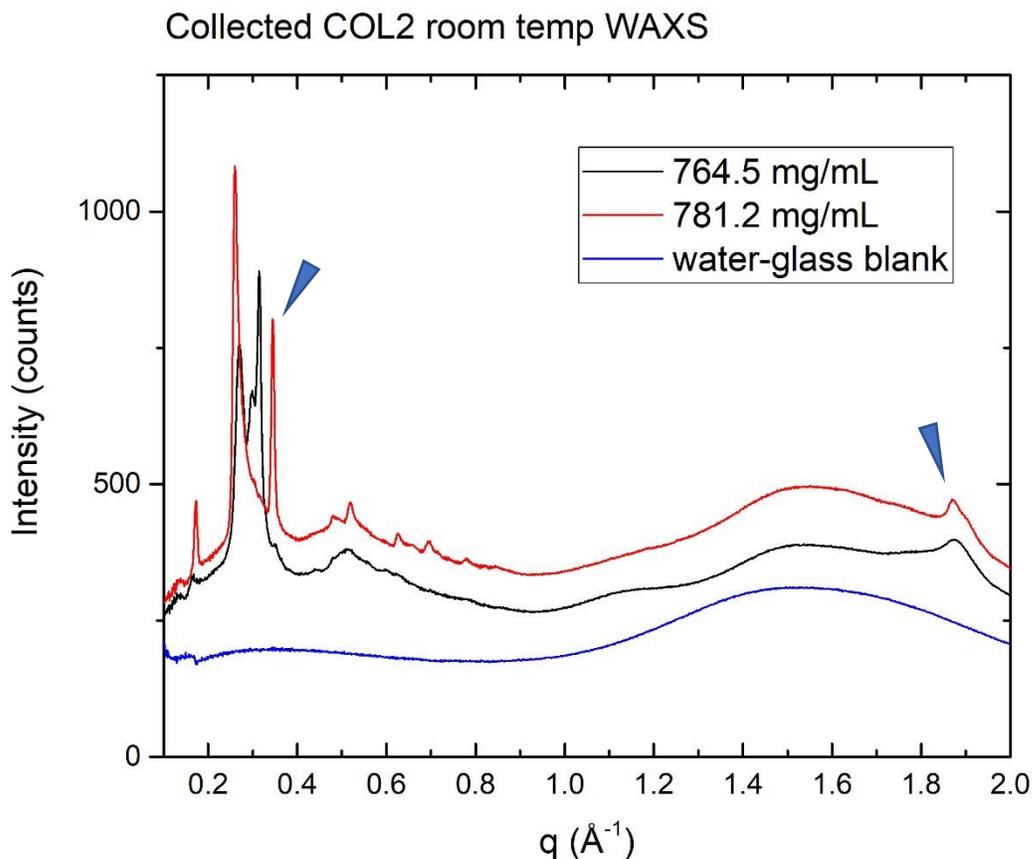


Figure 2.32: Collected circular averages of COL2 samples at room temperature. 764.5 mg/mL and 781.2 mg/mL rDD COL2 samples are shown with a water and glass blank as a near baseline comparison. The blue triangles mark the packing peaks around  $0.25 \text{ \AA}^{-1}$  and the base pairing peak at  $1.85 \text{ \AA}^{-1}$ .

Figures 2.31 and 2.32 show COL2 data first as the typical 2D data set, then collected in the circular averaged 1D plot. Both samples display a great deal of structure, though in a somewhat non-identical fashion; 764.5 mg/mL rDD appears to show a cluster of peaks at the side-side ordering length while 781.2 mg/mL rDD shows additional peaks at a slightly wider angle and only two peaks in the core region. There is also a smaller angle peak that has appeared below  $0.2 \text{ \AA}^{-1}$  that is not seen in previous samples. In both samples, the DNA stacking peak at  $1.85 \text{ \AA}^{-1}$  is relatively strong and initially appeared to show distinct structure that is dissimilar

between both COL2 samples. The lack of additional peaks in the region between  $0.9 \text{ \AA}^{-1}$  and  $1.85 \text{ \AA}^{-1}$  suggests qualitatively that these phases are liquid crystalline and not fully crystalline. The structure is clearly more complicated than the COL phases detailed in Section 2.2.2.

Insight into the structure of these phases benefits somewhat from an exhibition of the temperature ramp melting behavior of  $764.5 \text{ mg/mL}$  rDD.

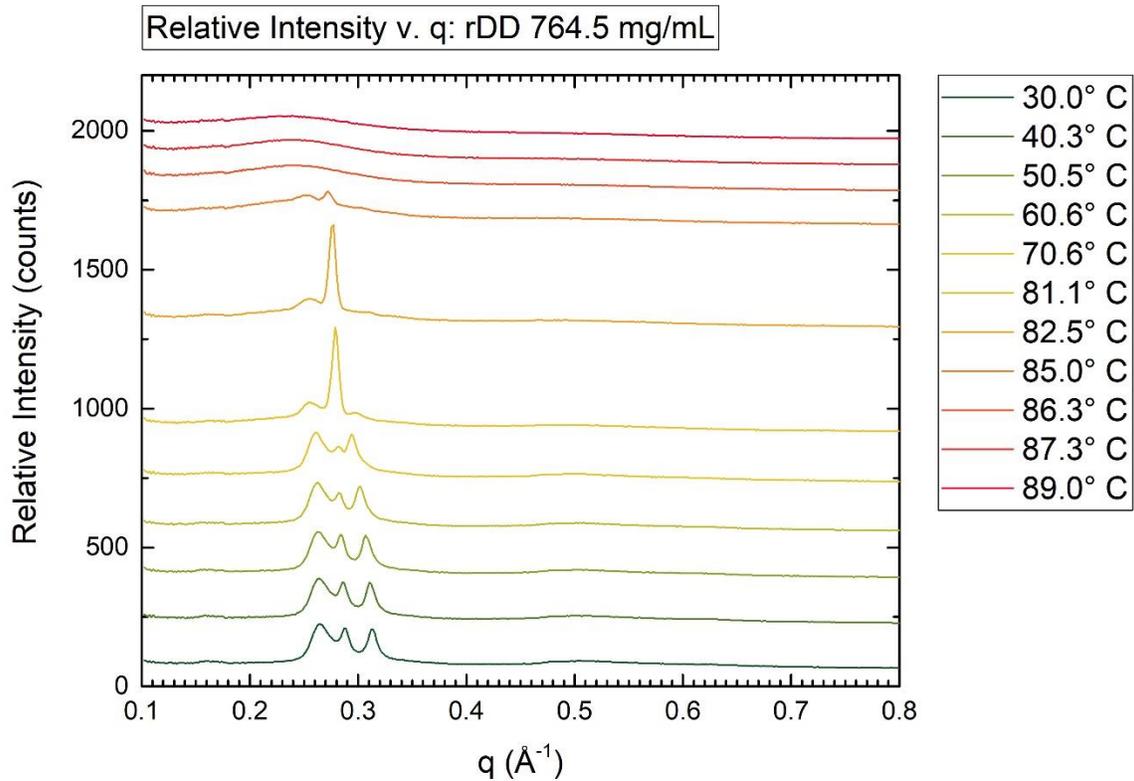


Figure 2.33: Temperature ramp experiment of rDD  $764.5 \text{ mg/mL}$ . Collected circular averages of all scans during the temperature ramp with temperatures marked, stacked on relative intensity to distinguish them.

As a function of increasing temperature, rDD  $764.5 \text{ mg/mL}$  undergoes a series of clear transformations (see Figure 2.33). Through the lower temperatures, the peaks steadily shift to lower  $q$  until about  $80^\circ \text{ C}$  when the central peak of the cluster suddenly explodes in intensity over the other two. At this point, the central

peak strongly resembles the  $q_{11}$  lattice spacing peak of the COL phase (see Section 2.2.2) possessing a periodicity of  $0.2771 \text{ \AA}^{-1}$  (see Figure 2.34). At that temperature, a subsidiary peak is also apparent at  $0.468 \text{ \AA}^{-1}$ , very close to the  $q_{31}$  hexagonal lattice periodicity expected for COL at that spacing.

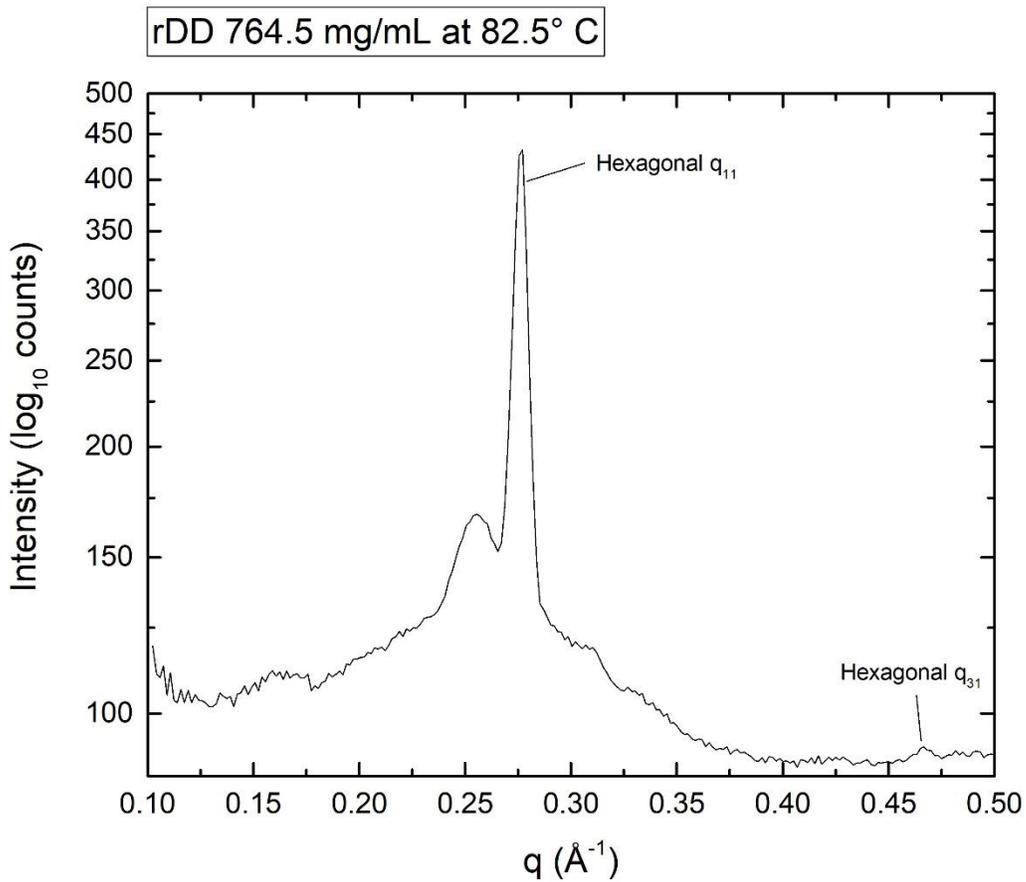


Figure 2.34: WAXS of mesophase packing peak of rDD 764.5 mg/mL at 82.5° C.

If the central peak of the mesophase packing cluster in rDD 764.5 mg/mL is from a COL phase, the two other side-side packing peaks of the 764.5 mg/mL COL2 phase begin to resemble those of 781.2 mg/mL. In addition, in Figure 2.33, all the packing peaks gradually move left in  $q$ -space, suggesting phase expansion with increasing temperature. If 781.2 mg/mL is simply a more compressed version of 764.5 mg/mL, the  $0.25 \text{ \AA}^{-1}$  and  $0.3 \text{ \AA}^{-1}$  peaks of 764.5 mg/mL would potentially

converge to the values seen with 781.2 mg/mL. The 764.5 mg/mL sample seems to be constant coexistence of COL and COL2 with 781.2 mg/mL sitting firmly in the COL2 region of the phase diagram (see Figure 2.2). The phase expansion in Figure 2.33 can be interpreted as a gradual unlimbering of COL2 that ultimately results in a conversion to COL as the volume fraction of the phase becomes low enough and the internal pressure becomes high enough at high temperature.

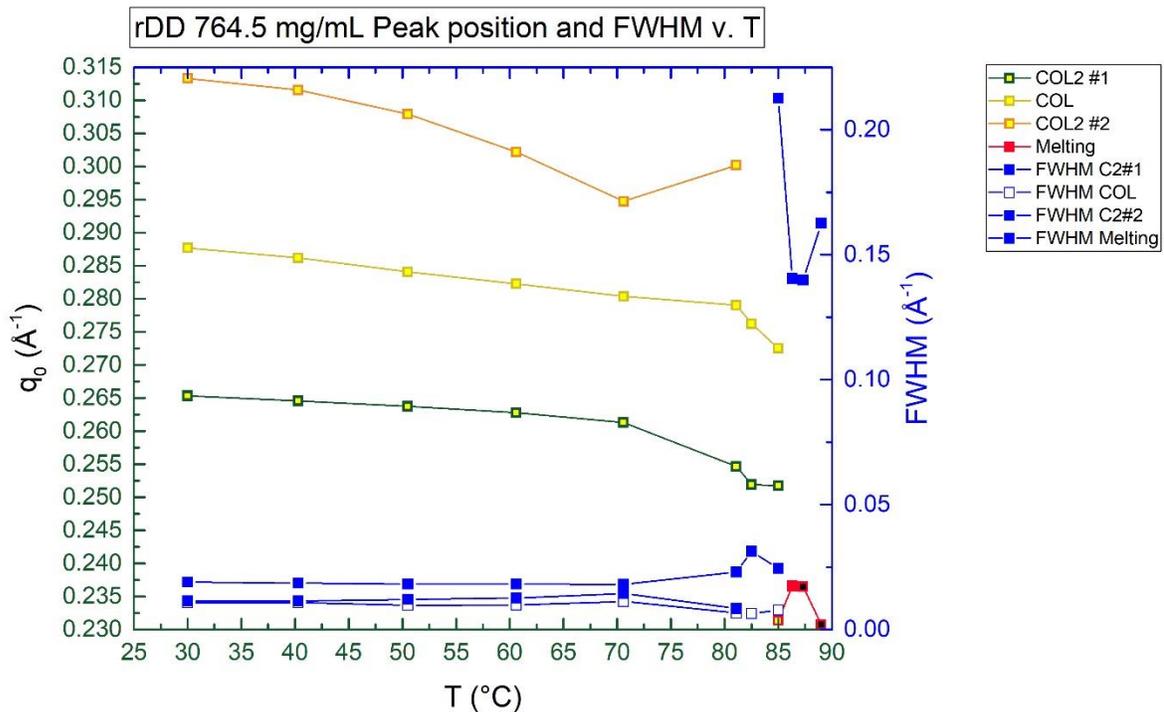


Figure 2.35: Peak position and FWHM versus Temperature for rDD 764.5 mg/mL. Peaks for COL (Tan line) and COL2 (#1 Green line, #2 Orange line) and for Melting phase (Red line) have square markers that are augmented with color to show optical birefringence observations at a given temperature, birefringent (Yellow), melting coexistence (Red) and ISO (Black) –the word “coexistence” here meaning between birefringent phases and isotropic given that this sample is already in coexistence between COL and COL2 with two to three phase coexistence throughout the Figure. Blue lines and markers all denote FWHM of these peaks as measured by the blue axis at right.

The complicated melting behavior of rDD 764.5 mg/mL is captured with peak fitting in Figure 2.35. The characteristic peaks of the phase slide toward lower  $q$

with increasing temperature until they undergo a melting catastrophe that results in a significant alteration of the phase which breaks the smoothness of the progression before converting into a NEM\*-like melting peak as seen with COL in Section 2.2.2. The phase ultimately melts above 85° C to ISO. This particular data collection was encountered by luck during shooting at the synchrotron and the data capturing frequency was increased relative to temperature change in order to successfully resolve the details of the behavior.

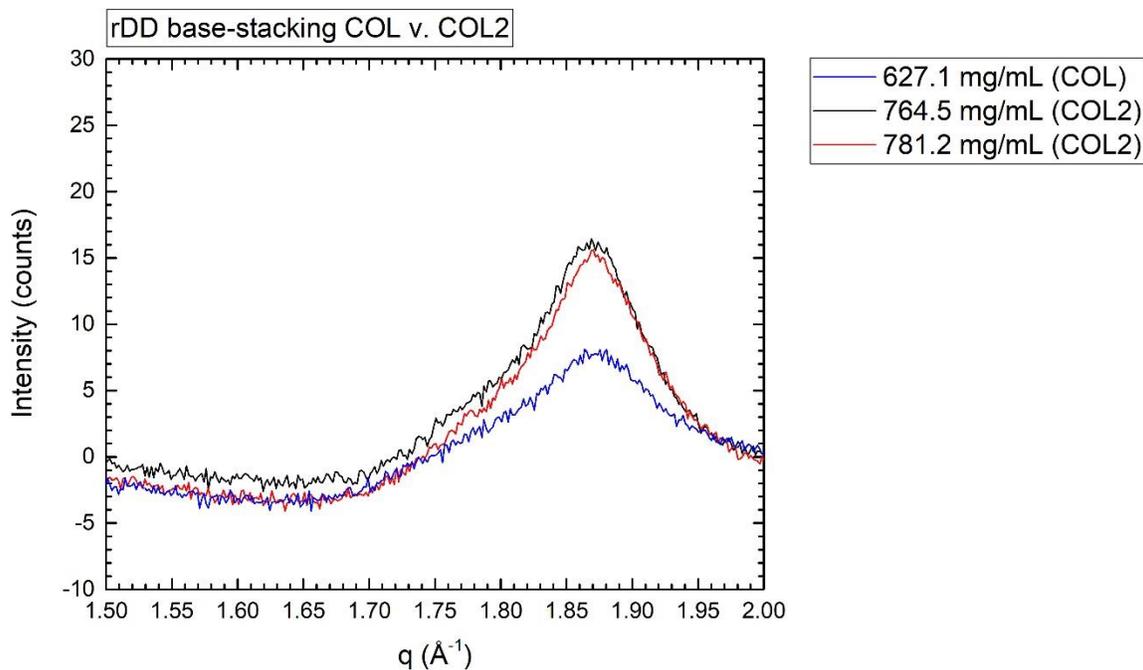


Figure 2.36: rDD base-stacking COL v. COL2. Comparison of base-stacking periodicities for known COL phase at 627.1 mg/mL (Blue) as compared to higher order COL2 at 764.5 mg/mL (Black) and 781.2 mg/mL (Red). The baseline here was attained by subtracting high temperature isotropic intensity, which results in a slight deviation of the base below zero intensity at  $q$  smaller than  $1.7 \text{ \AA}^{-1}$ .

In the intro at Section 2.2 it was noted that the COL2 base-stacking reflection appeared to have an altered shape between 764.5 mg/mL and 781.2 mg/mL, hinting that maybe this could imply distortion to the aggregate, but the phenomenon there did not persist between trials distinctively enough to provide

confidence. Base stacking does not appear significantly different between COL and COL2 according to Figure 2.36. If distortions to the duplex that perturb the stacking periodicity are required to form COL2, it would be expected that COL2 stack significantly differently from COL. This may or may not be indicative of any structural shifts to the mesophase simply because the base-stacking peak is reflective of the  $\pi$ -stacking period found for aromatic molecules. If the conformation of the stacking is radically altered, say by sliding base pairs past one another in the plane of the  $\pi$ -stacking interaction without changing the physical separation between them, WAXS might not be able to discern it. As an example, A-form DNA contains a large shift in helical period away from B-form DNA, where the A-form rise is 28.6 Å per helical turn versus 35.7 Å, all accommodated by torquing the normal vector of the base-pair plane away from the axis of the duplex without actually changing the packing distance. Despite this, the WAXS data contains no definitive evidence of perturbation to the long axis of the aggregate which would render COL2 significantly different from COL in that regard.

For COL phase, compression of the mesophase with increasing concentration seems to occur only to the 2D hexagonal lattice and not to the base-pair stacking which dominates the third dimension of the mesophase. Concentration forces the linear columns to fit together more closely side-to-side. From the similarity of the base stacking data (Figure 2.36) COL2 must also be a variation of this where the columns are most likely forced to alter packing without compressing length. Variation of concentration is therefore directly coupled to variation of the area of the 2D lattice that describes phase packing. As a result, the hexagonal 2D lattice of COL at some concentration must have the same area as the lattice describing the equivalent concentration of COL2. Phases at such a hypothetical coexistence have the same concentration and, correspondingly, the same cross-sectional area. The 2D

lattice of COL2 would logically be expected to be some variation of the hexagonal COL lattice.

From COL to COL2, the mesophase side-side packing reflection appears to split from one to two peaks. One possible way to account for this would be if the variation of hexagonal packing that exists in COL2 shifts from facets containing equilateral triangles to facets that have isosceles triangles (see Section C.4.2) by shortening only a single lateral. This deformation would create peak splitting that is only two-fold, as required by the data. If all three laterals are different, which would be encountered with other deformations, the scattering pattern would match that symmetry and have three fundamental peaks at different radii as well as higher order peaks matching the proposed basis set for that lattice. Any valid model would be expected to explain all the scattering seen in reciprocal space, as collapsed into 1D by powder averaging, making it possible to use the presence of higher order peaks to check the assignment of the fundamental peaks. Further, area conservation would be expected between COL and COL2 lattices at equivalent concentrations, given the apparent similarity of these mesophases in the dimension parallel to the long axis of the aggregate.

One feature that complicates analysis in this system is the degeneracy introduced by collapsing the 2D data set into 1D by powder averaging. In the FHL model introduced in Section C.4.2, the observed COL2 data could be assigned reversibly to the  $q_{11}$  and  $q_{20}$  reflections, leading to two different models with different unit cell areas (see Figure 2.37).

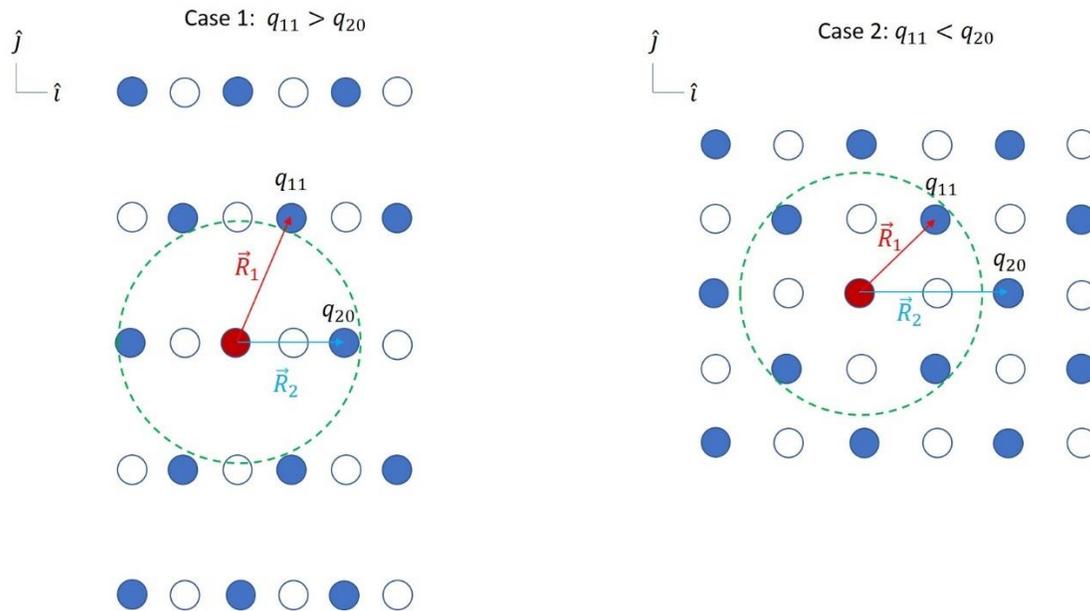


Figure 2.37: Flattened lattice degeneracy problem. Case 1 and Case 2 clearly produce lattices of differing area, but may be assigned from the same data simply by varying which reflection has the longer radius.

The FHL hypothesis can be tested by lifting COL2 #1 and #2  $q$ -values from the peaks in Figure 2.35 and using those to predict the COL  $q$  in the same figure. #1 =  $0.26535 \text{ \AA}^{-1}$  and #2 =  $0.31331 \text{ \AA}^{-1}$  at  $30^\circ \text{ C}$ . Using Case 1 ( $q_{11} > q_{20}$ ) with #1 as  $q_{20}$  and #2 as  $q_{11}$ , the predicted COL peak is  $0.295 \text{ \AA}^{-1}$ . Using Case 2 ( $q_{11} < q_{20}$ ) with #2 as  $q_{20}$  and #1 as  $q_{11}$ , the predicted COL peak is  $0.278 \text{ \AA}^{-1}$ . The observed value of the apparent COL peak in  $764.5 \text{ mg/mL rDD}$  is  $0.28771 \text{ \AA}^{-1}$ , which is fairly close to  $0.295$  but not a perfect hit. With this discrepancy in mind, one can begin to examine assignments of the higher order peaks.

The very best observation of higher order COL2 reflections comes from the initial room temperature WAXS of rDD  $781.2 \text{ mg/mL}$  (see Figure 2.32 and 2.38). The observations of rDD  $764.5 \text{ mg/mL}$  has a washed-out example of the similar peaks, but  $781.2 \text{ mg/mL}$  gave very sharp peaks on that observation. Peaks extracted from this figure are in Table 2.1 with indexes generated by matching the appropriate FHL basis vectors.

Table 2.1:

Peak $q$ ( $\text{\AA}^{-1}$ )	Case 1 (h,k), $\vec{A} = 0.13 \text{\AA}^{-1}$ , $\vec{B} = 0.32 \text{\AA}^{-1}$	Case 2 (h,k), $\vec{A} = 0.17 \text{\AA}^{-1}$ , $\vec{B} = 0.19 \text{\AA}^{-1}$
0.259848	(2,0)	(1,1)
0.344697	(1,1)	(2,0)
0.479545	Not matching	Not matching
0.518939	(4,0) (predict 0.519696)	(2,2) (predict 0.51969)
0.625	Not matching	Not matching
0.694697	(2,2) (predict 0.689331)	(4,0) (predict 0.689394)
0.77803	Not matching	(3,3) (predict 0.779) (0,4) (0.7786)

Upon considering higher order peaks, there is a series of matches possible for both combinations of  $q_{11}$  and  $q_{20}$ . However, neither case accommodates all of the peaks.

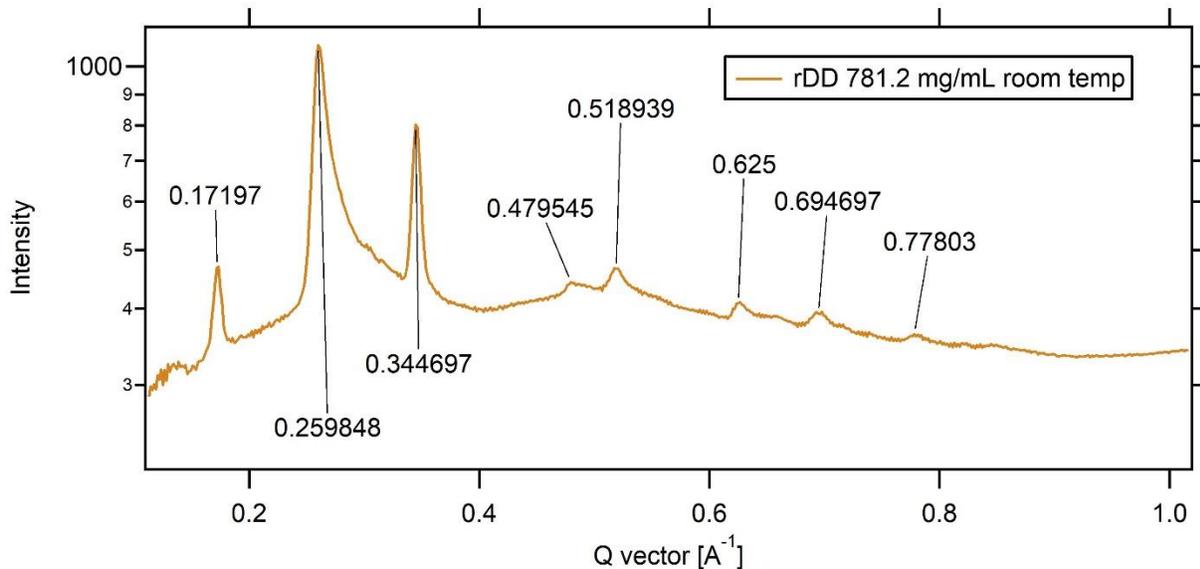


Figure 2.38: Annotated peaks of room temp rDD 781.2 mg/mL.

The peak at  $0.172 \text{\AA}^{-1}$  has some potential for contributing to the higher order reflection structure as well. For reasons that will be discussed in greater detail, this particular reflection might be considered a fundamental peak that is parallel to the long

axis of the aggregate, making the diffraction pattern in Figure 2.38 a 1D reduction of a three-dimensional pattern. The harmonics associated with this peak have potential implications to Figure 2.38 (see Table 2.2).

Table 2.2

Peak $\text{\AA}^{-1}$	Harmonic	Observed? ( $\text{\AA}^{-1}$ )
0.17197	1	Yes
0.34469	2	Yes (0.344697)
0.5159	3	Yes (0.518939)
0.68788	4	Yes (0.694697)
0.85985	5	Maybe (noisy)

From Table 2.2, most of the Case 1 matches are doubled. Case 2 matches fare slightly better, with  $0.778 \text{\AA}^{-1}$  not being duplicated. That the  $q_{11}$  or  $q_{20}$  at  $0.344 \text{\AA}^{-1}$  also sits on a multiple of  $0.172 \text{\AA}^{-1}$  is curious and bothersome, but might be argued as coincidence given that the analog of the  $0.344 \text{\AA}^{-1}$  peak seen elsewhere is frequently much stronger than the  $0.172 \text{\AA}^{-1}$  peak (see Figure 2.39 for more typical strength of  $0.17 \text{\AA}^{-1}$  peak.)

An unusual insight about this mesophase at  $781.2 \text{ mg/mL}$  comes from the bizarre shape of the  $0.26 \text{\AA}^{-1}$  peak in the central pair of peaks. In every instance of this pattern, seen both with rDD and with blunt-end 4mer  $5\text{GTAC}3$ , this clustered peak pattern gives the  $0.26 \text{\AA}^{-1}$  peak (of Figure 2.38) an odd “horn-like” appearance. With peak fitting, this distinctive shape turns out to contain a third peak sitting in the valley between the two obvious peaks (see Figure 2.39).

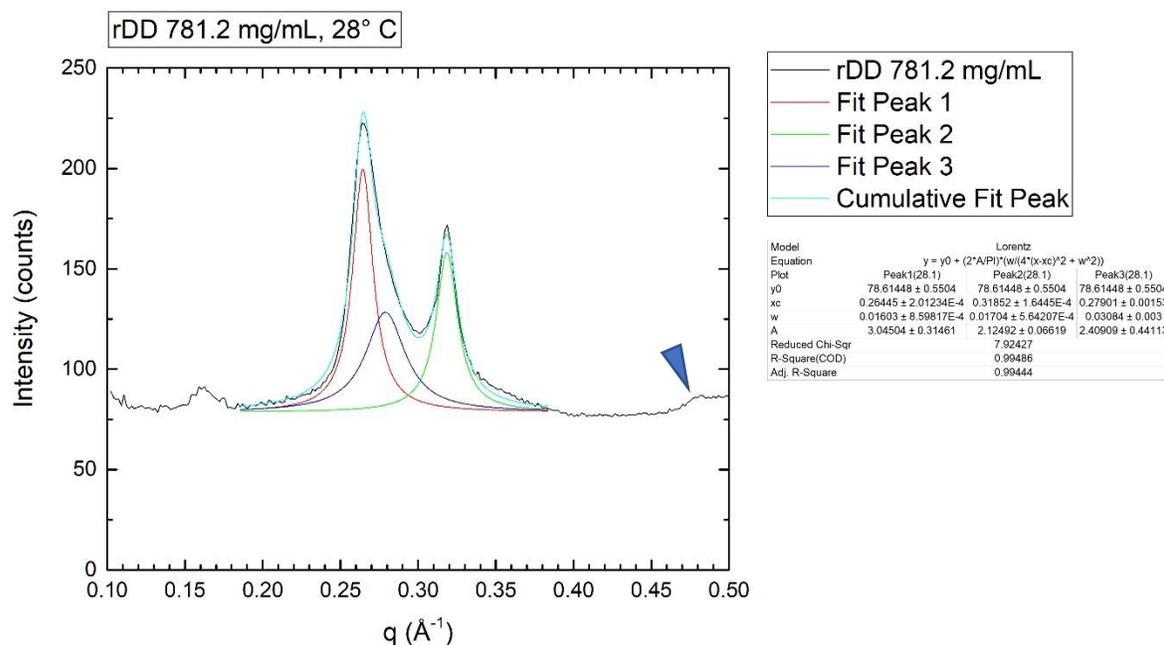
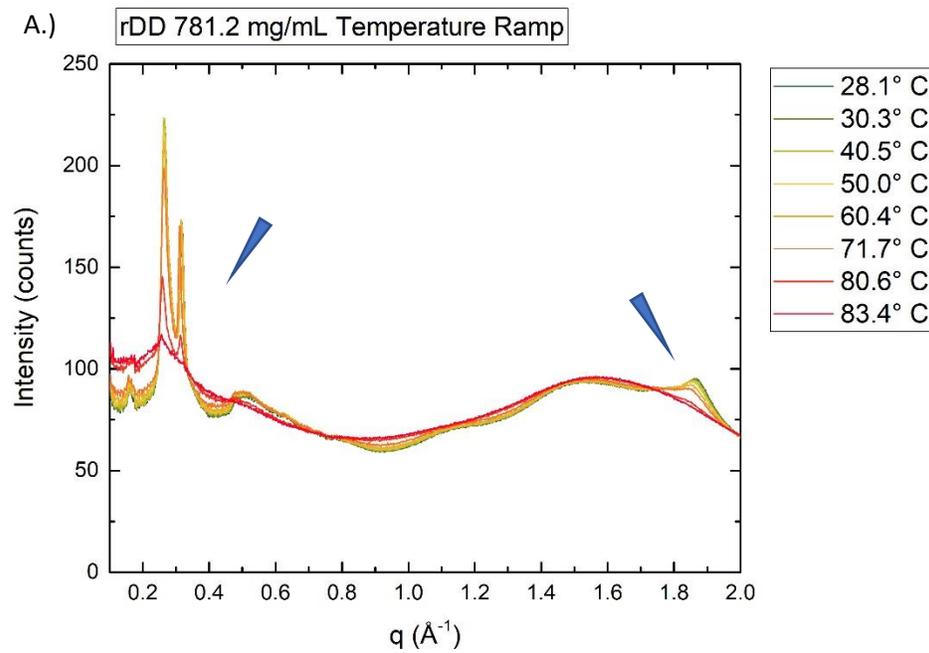


Figure 2.39: Central peak of COL2 mesophase fitted to Lorentzians. The basic WAXS of 781.2 mg/mL rDD (Black Line) fitted to three Lorentzians, Peak #1 (Red Line), Peak #2 (Green Line) and underlying peak (Purple Line). The cumulative fit is bright blue. Peak 3 is fit only based on the expectation that it fits somewhere between Peaks #1 and #2; the centering of this peak comes exclusively from the fit. Fitting parameters are noted in the details beneath the legend. The dark blue triangle picks out a higher order peak at  $q = 0.481 \text{ \AA}^{-1}$ .

The third peak depicted by the purple line in Figure 2.39 is centered at  $0.27901 \text{ \AA}^{-1}$ . The peak noted by the blue triangle at a position of  $0.481 \text{ \AA}^{-1}$  would match a  $q_{31}$  spacing for a hexagonal lattice if the buried peak is taken to be the hexagonal  $q_{11}$  ( $q_{hex}$ ). This position corresponds to the  $0.4795 \text{ \AA}^{-1}$  peak which is noted “not matching” in Table 2.1. If peaks #1 and #2 in Figure 2.39 are taken as FHL  $q_{11}$  and  $q_{20}$  respectively, eqn C.70 picks a hexagonal peak at  $q_{hex} = 0.2782$ , very close to the center of the curve fit peak obtained in Figure 2.39. If the FHL Case 2 is the proper assignment, its equivalent COL reflection is always present in the shape of the peak, and this buried peak always has a hexagonal higher order reflection associated with it. This suggests either a coexistence with COL in COL2 at the concentration of COL2 or an *unflattened* 2D hexagonal lattice present somehow in

the 3D structure of the mesophase. This makes the story of COL2 somewhat more complicated because it blurs the line between COL and COL2. Later in this section, references to COL are sometimes also referred to as “the hexagonal lattice” on the understanding that apparently two 2D lattices are coexistent at all observed concentrations of COL2.

For the temperature ramping experiment performed for 764.5 mg/mL rDD as depicted in Figures 2.33 and 2.35, a similar experiment was carried out for 781.2 mg/mL rDD (see Figure 2.40).



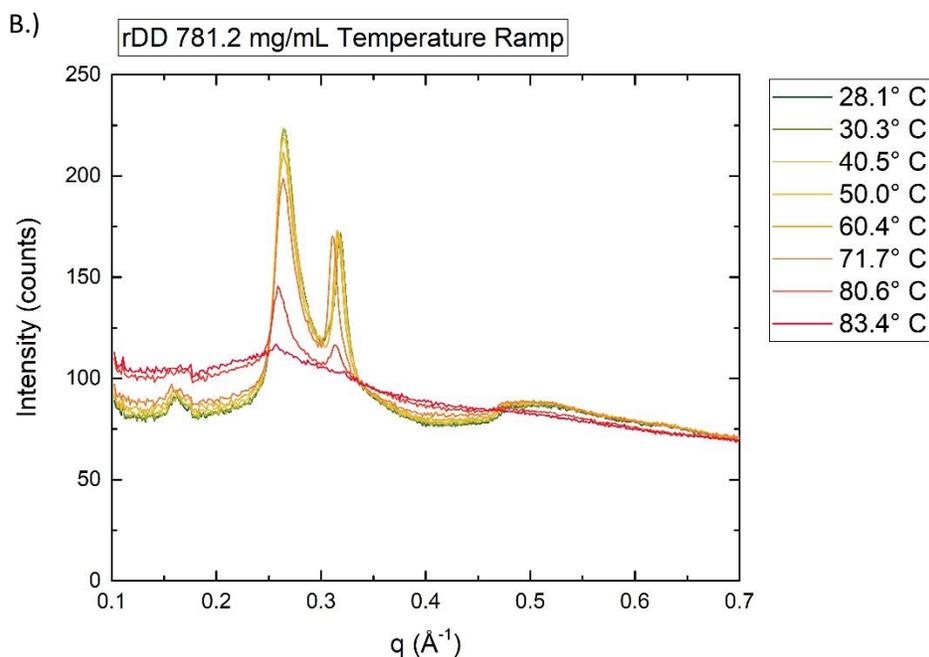


Figure 2.40: Temperature ramp of rDD 781.2 mg/mL. A.) Full 1D reciprocal space data reduction of all WAXS images, blue triangles point out the side-side packing region and the base stacking region. B.) 1D reciprocal space focused on the column packing region. During the time of this experiment, rDD 781.2 mg/mL did not show as much clear structure as seen in initial examination (see Figure 2.38). Beam stop clipping occurs to the data just outside the  $0.172 \text{ \AA}^{-1}$  peak.

rDD 781.2 mg/mL did not undergo as complicated a melting behavior as that seen with rDD 764.5 mg/mL. There is not enough data to pick out a melting phase peak as can be seen with all lower concentrations of rDD except in the single highest WAXS image. At least as seen in Figure 2.40, rDD 781.2 mg/mL appears to phase transition at a slightly lower temperature than the other COL2 sample. Whether this is a recording error at the synchrotron or a real phenomenon is uncertain; microscopically, it appears to transition to ISO at a similar temperature to rDD 764.5 mg/mL. Except for the peak at  $0.172 \text{ \AA}^{-1}$ , higher order peaks do not appear in this experiment as clearly as they did in Figure 2.38, despite being from the same sample; it is possible that this is an effect of aging the sample or of

synchrotron radiation damage. Higher order peaks are clearly present in the extended feature between  $0.47 \text{ \AA}^{-1}$  and  $1.0 \text{ \AA}^{-1}$  but are blurred together and difficult to sort. Peak fitting reveals the buried peak in all data sets and up until the very highest temperatures, this peak is located at the predicted  $q_{hex}$  for the Case #2 FHL.

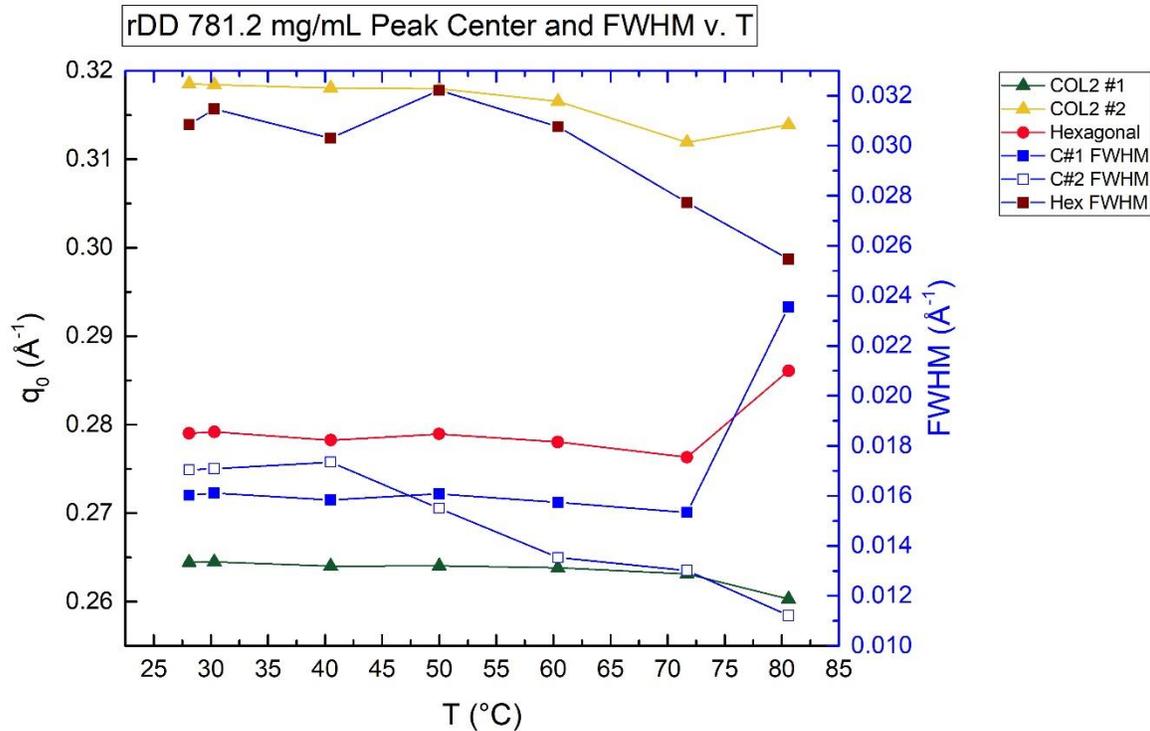


Figure 2.41: Peak center and FWHM v. temperature for rDD 781.2 mg/mL. Peak center is measured from the black axis at the left of COL2 #1 (Dark Green triangles), COL2 #2 (Yellow triangles) and the buried Hexagonal  $q_{11}$  peak (Red circles). FWHM is measured from the blue axis at the right for COL2 #1 (filled blue squares), COL2 #2 (empty blue squares) and the buried Hexagonal peak (Purple squares).

From Figure 2.41, the positioning of all the mesophase packing peaks are held quite rigidly until above  $70^\circ \text{ C}$ . The peak width for COL2 #1 varies differently than that for COL2 #2 while the width of the buried hexagonal  $q_{11}$  peak shifts in a manner similar to COL2 #2.

The existence of this third “COL-like” peak in both concentrations of COL2 casts some additional doubt on the FHL hypothesis because both COL2 examples appear to have evidence for three peaks rather than two. Three-way peak splitting would be indicative of a different style of hexagonal deformation, exemplified particularly by a sheared hexagonal lattice (SHL) where all the parameters of a hexagonal lattice are different from one another (see Figure 2.42). As depicted in Figure 2.39, there is evidence that it would be appropriate to assign this third peak to a separate hexagonal lattice in coexistence with the COL2 phase, but it cannot be discounted that maybe these peaks somehow fit an SHL instead (see Section C.4.3). Invoking the SHL hypothesis complicates matters once again because of peak degeneracy: from the 1D data, the three peaks can be assigned three different ways, giving three different SHL type lattices.

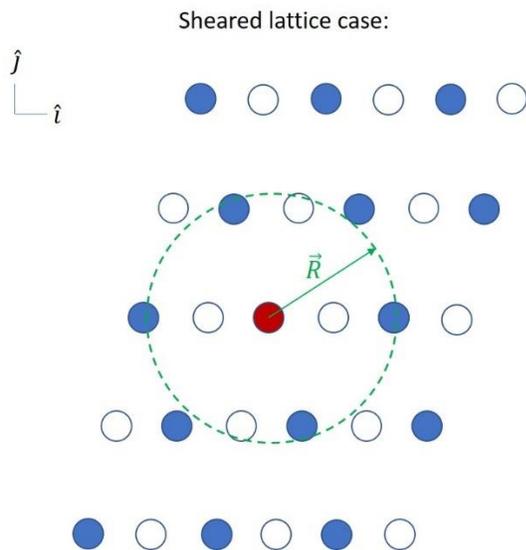


Figure 2.42: Sheared Hexagonal Lattice. Note that all three of the fundamental side-side packing reflections are now displaced from the radius of the hexagonal  $q_{11}$  reflection.

To help distinguish among the different cases, it is helpful to generate the parameters of all three possible SHL versions for data sets of both 764.5 mg/mL and 781.2 mg/mL rDD at room temperature. The parameters calculated here are as described in Section C.4.3.

Table 2.3A

764.5 mg/mL	0.26535	0.31331	0.28771	$ \vec{a} $	$ \vec{b} $	$\sin\theta$
Case 1	$q_{11}$	$q_{-11}$	$q_{02}$	25.38 Å	44.5 Å	0.98154 (for 79.0°)
Case 2	$q_{-11}$	$q_{02}$	$q_{11}$	23.41 Å	47.63 Å	0.994216 (for 83.8°)
Case 3	$q_{02}$	$q_{11}$	$q_{-11}$	27.64 Å	40.26 Å	0.996252 (for 85.0°)

764.5 mg/mL	$q_{hex}$	Area ( $A_{SHL}$ )
Case 1	0.28676 Å <sup>-1</sup>	1109.4 Å <sup>2</sup>
Case 2	0.28676 Å <sup>-1</sup>	1109.4 Å <sup>2</sup>
Case 3	0.28676 Å <sup>-1</sup>	1109.4 Å <sup>2</sup>

Table 2.3B

781.2 mg/mL	0.26445	0.31852	0.27901	$ \vec{a} $	$ \vec{b} $	$\sin\theta$
Case 1	$q_{11}$	$q_{-11}$	$q_{02}$	25.02 Å	46.16 Å	0.97615 (for 77.3°)
Case 2	$q_{-11}$	$q_{02}$	$q_{11}$	23.72 Å	48.19 Å	0.986099 (for 80°)
Case 3	$q_{02}$	$q_{11}$	$q_{-11}$	28.57 Å	39.52 Å	0.998409 (for 86.7°)

781.2 mg/mL	$q_{hex}$	Area ( $A_{SHL}$ )
Case 1	0.28441 Å <sup>-1</sup>	1127.1 Å <sup>2</sup>
Case 2	0.28441 Å <sup>-1</sup>	1127.1 Å <sup>2</sup>
Case 3	0.28441 Å <sup>-1</sup>	1127.1 Å <sup>2</sup>

For the SHL hypothesis models, all three cases tend to generate both the same area and the same  $q_{hex}$ ; they are otherwise indistinguishable from one another. Aligned mesophase samples with 2D data not suffering from a powder average would be required to distinguish among them.

This leaves multiple ways to interpret the data. A.) A Case #1 FHL with a coexistent hexagonal lattice at the predicted periodicity for the equivalent Case #2 FHL. B.) A Case #2 FHL with a coexistent hexagonal lattice at the predicted periodicity. C.) One of three SHL forms where the assigned hexagonal  $q_{31}$  reflection is actually something else. The fourth obvious interpretation is that none of the examined cases fit. Interpretation B.) seems most reasonable from this level of examination.

An important test to distinguish among these possibilities is to see if they obey the expected behavior between the two samples. As mentioned, a chief trend taken from the hexagonal COL seen in Section 2.2.2 is that the area of the lattice is compressed with increasing concentration. A suitable COL2 lattice would be expected to decrease in unit cell area with increasing concentration. This particular test generates an extremely surprising result (see Figure 2.43).

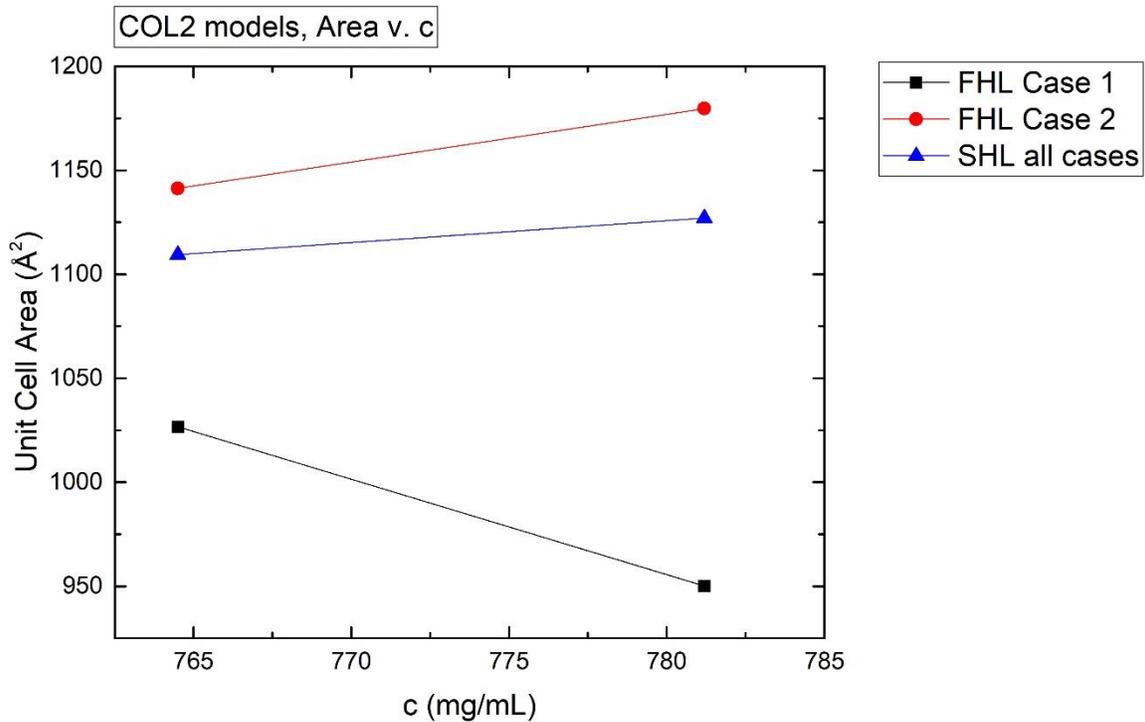
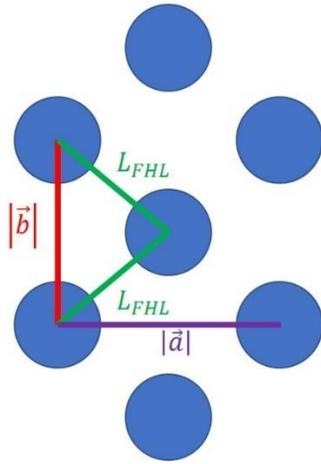


Figure 2.43: Comparison of Unit Cell Area for each lattice model with increasing concentration. All models are included, FHL Case #1 (Black Squares), FHL Case #2 (Red Circles) and all cases of SHL (Blue Triangles). Unit cell area is calculated from  $A_{FHL}$  and  $A_{SHL}$  found in Sections C.4.2 and C.4.3 respectively.

From Figure 2.43, it is clear that no deformed hexagonal lattice models obey the expected concentration trend except for FHL Case #1. All other cases, FHL Case #2 and the three different SHL cases, fail to compress in area with increasing concentration. FHL Case #1 is the only model that can work.

From this choice of assignment of  $q_{11}$  and  $q_{20}$  in the Case #1 FHL model, the lattice parameters  $L_{FHL}$ ,  $|\vec{a}|$  and  $|\vec{b}|$ , can be calculated in terms of these reflections as defined in Section C.4.2 (equations C.74, C.75 and C.76) repeated here for convenience as Figure 2.44).

Flattened Hexagonal Lattice:



$$|\vec{a}| = \frac{4\pi}{q_{20}}$$

$$L_{FHL} = 2\pi \sqrt{\frac{1}{(q_{20})^2} + \frac{1}{4(q_{11})^2 - (q_{20})^2}}$$

$$|\vec{b}| = \frac{2\pi}{\sqrt{(q_{11})^2 - \left(\frac{q_{20}}{2}\right)^2}}$$

Figure 2.44: Lattice parameters of the Case #1 FHL. The lengths in this diagram are not drawn to match the scale of the actual lengths seen in the real mesophase 2D spacing lattices.

With these lattice parameter definitions, the peak data shown in Figure 2.41 can be converted to show how the lattice dimensions vary with temperature. For the sake of comparison, one should note that  $L_{hex}$  corresponds to the same lateral in the hexagonal lattice that  $L_{FHL}$  does in the FHL model in Figure 2.44, but also that  $|\vec{b}| = L_{hex}$  in the hexagonal case.

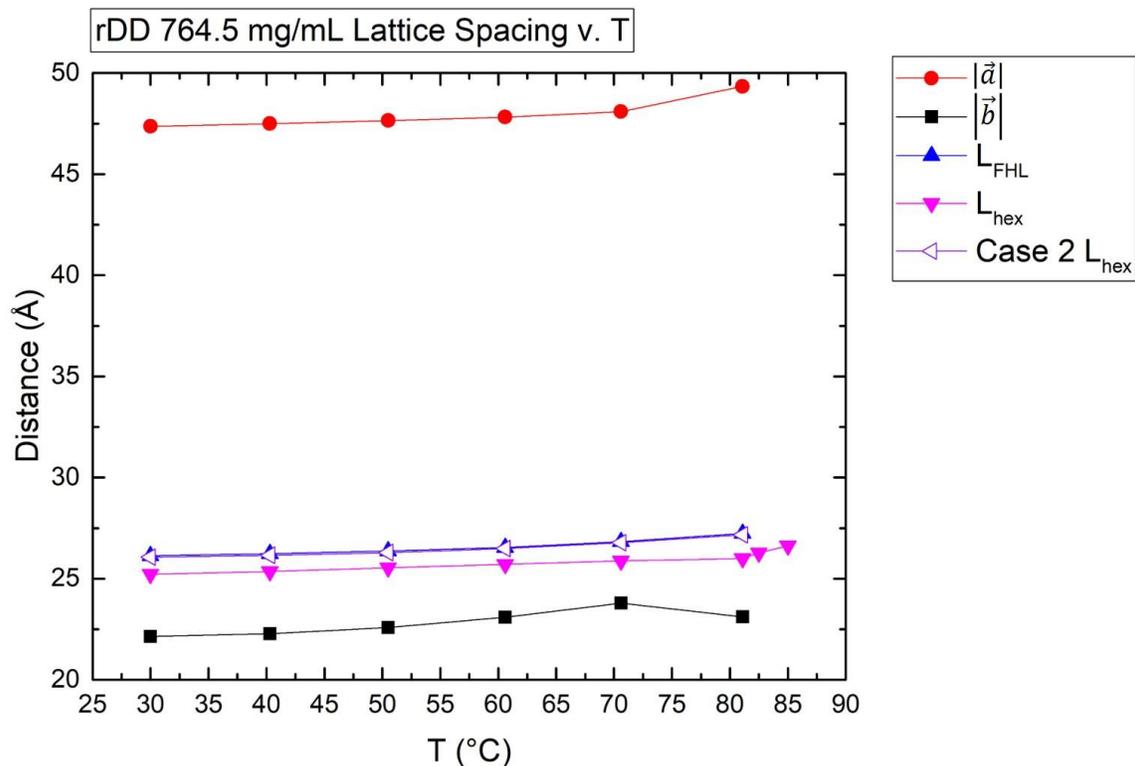


Figure 2.45: rDD 764.5 mg/mL Lattice Spacing versus Temperature. Variation of lattice parameters, including  $|\vec{a}|$  (Red Circles),  $|\vec{b}|$  (Black Squares), Case 1  $L_{FHL}$  (Blue Triangles),  $L_{hex}$  (Magenta Triangles) and Case 2  $L_{hex}$  (White Triangles).  $L_{FHL}$  is calculated from the COL2 reflections assigned as Case 1 FHL.  $L_{hex}$  is calculated from the hexagonal  $q_{11}$  reflection identified in Figure 2.34. Case 2  $L_{hex}$  is calculated from the FHL  $q_{11}$  and  $q_{20}$  reflections first identified as COL2 in Figures 2.33 and 2.34 (though these peaks are now identified as Case 1 based on Figure 2.43).

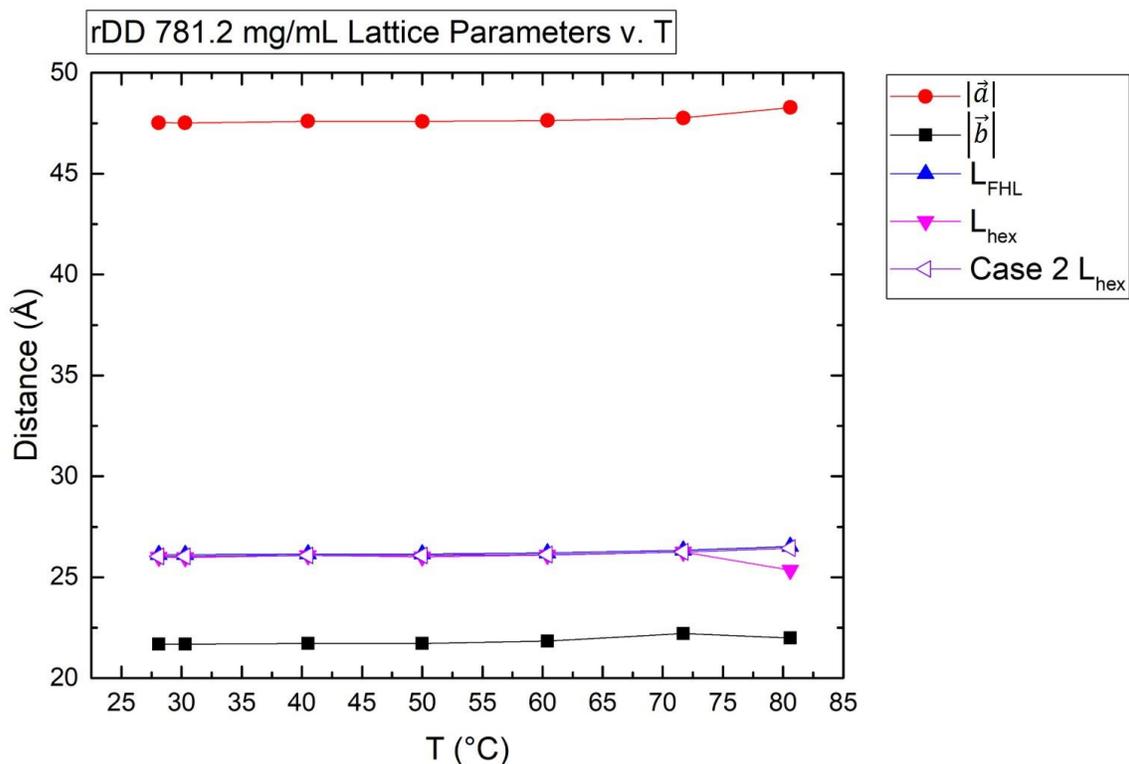


Figure 2.46: rDD 781.2 mg/mL Lattice Spacing versus Temperature. Variation of lattice parameters, including  $|\vec{a}|$  (Red Circles),  $|\vec{b}|$  (Black Squares), Case 1  $L_{FHL}$  (Blue Triangles),  $L_{hex}$  (Magenta Triangles) and Case 2  $L_{hex}$  (White Triangles).  $L_{FHL}$  is calculated from COL2 reflections assigned as Case 1 (Table 2.1).  $L_{hex}$  is calculated from the buried peak identified as a hexagonal  $q_{11}$  reflection in Figure 2.39. Case 2  $L_{hex}$  is calculated from the FHL  $q_{11}$  and  $q_{20}$  reflections first identified in Table 2.1 and formally taken as Case 1 based on lattice area behavior in Figure 2.43.

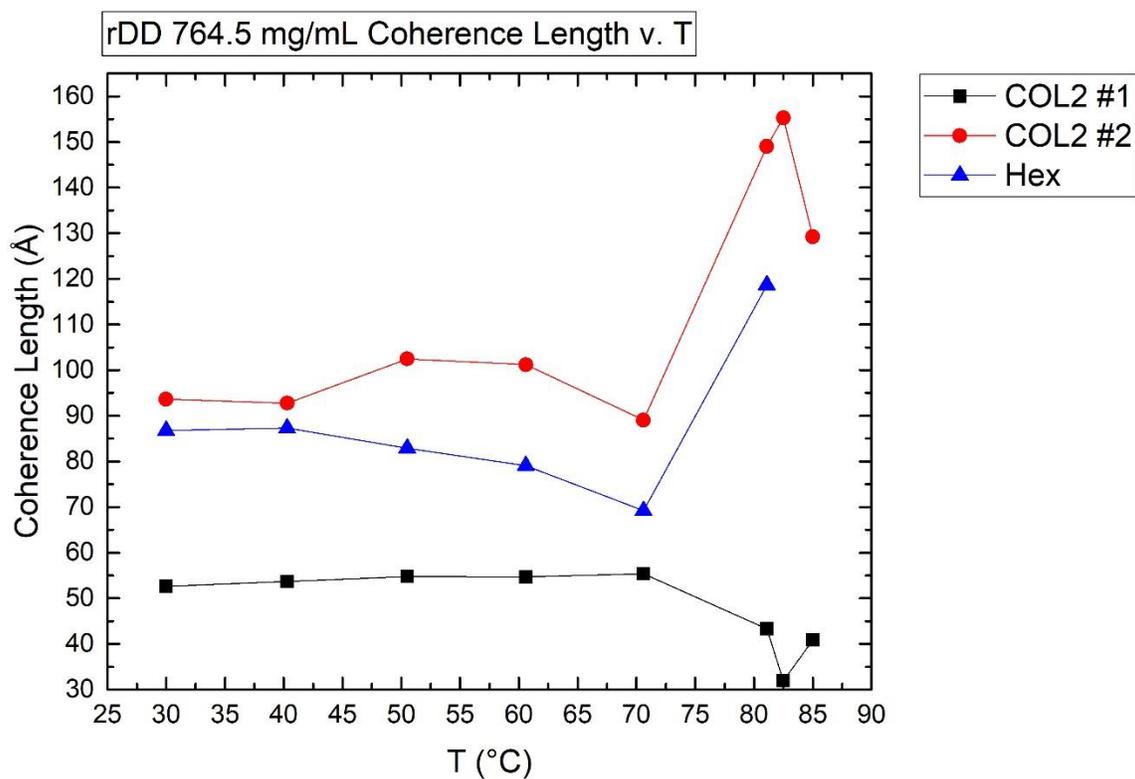


Figure 2.47: Structural Coherence Length versus Temperature for rDD 764.5 mg/mL. Reciprocal space peak widths summarized in Figure 2.35 are converted to COL (Hex, Blue Triangles) and COL2 (Peak #1 in Black Squares and Peak #2 in Red Circles). Melting phase is not included here. The hexagonal lattice (Hex) is associated with the  $L_{hex}$  parameter identified in Figure 2.45. The COL2 peaks #1 and #2 are associated with the FHL parameters identified in Figure 2.44 and 2.45.

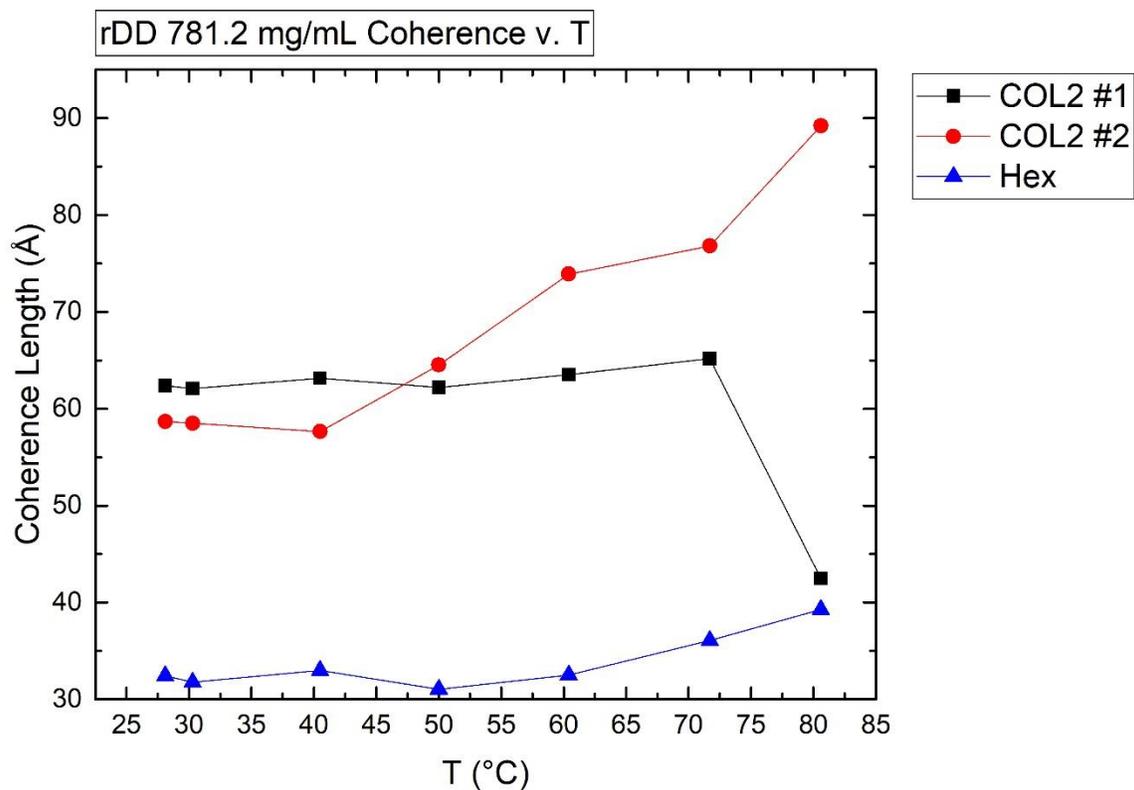


Figure 2.48: Structural Coherence Length versus Temperature for rDD 781.2 mg/mL. Reciprocal space peak widths summarized in Figure 2.35 are converted to COL (Hex, Blue Triangles) and COL2 (Peak #1 in Black Squares and Peak #2 in Red Circles). Melting phase is not included here. Hexagonal lattice (Hex) is associated with the buried peak giving  $L_{hex}$  in Figure 2.46. COL2 peaks #1 and #2 are associated with the FHL parameters identified in 2.44 and 2.46.

Assigning the lattice as the Case 1 FHL introduces an incredible conundrum in trying to explain the existence of the buried lattice peak in the rDD 781.2 mg/mL sample, whose  $q_{hex}$  spacing was predicted by the FHL Case #2 style peak assignment. The two lattices that seem to exist in the data cannot be related by conserved unit cell area, and, paradoxically, the difference between those areas increases with increasing concentration (Figure 2.43). Either the models are wrong, requiring some additional insight, or interpretation A.) from above is the correct interpretation where a Case #1 FHL is coexisting with a hexagonal lattice whose

spacing is coincidentally, but not meaningfully, predicted by the FHL Case #2 model. Although upsetting, accidental correlation is permissible.

In the rDD 764.5 mg/mL sample, Figure 2.45 shows that the identified hexagonal lattice (taken to be COL coexistence with COL2) having spacing  $L_{hex}$ , does not quite correspond with any of the COL2 dimensions.  $L_{hex}$  and  $L_{FHL}$  are similar here, but not identical. The calculated hexagonal lattice, Case 2  $L_{hex}$ , always has a direct correspondence with the Case 1  $L_{FHL}$  parameter; they do not vary from each other despite being calculated through different routes.

This is contrasted by Figure 2.46, where  $L_{FHL}$ , the buried hexagonal  $q_{11}$  reflection  $L_{hex}$  and the calculated Case 2  $L_{hex}$  are all *exactly* on the same line except at very high temperatures. In this Figure,  $L_{FHL}$  is obtained from a calculation depending on the Case 1 FHL assignment of  $q_{11}$  and  $q_{20}$ ;  $L_{hex}$  comes from a peak extracted from the data by peak fitting; and Case 2  $L_{hex}$  comes from calculation of  $q_{hex}$  based on inverted assignment of COL2 Case 1 FHL  $q_{11}$  and  $q_{20}$ . All three are obtained by different means, and yet lie on the same trend. The reasons for this correspondence are not entirely clear.

From Figure 2.45, it could be assessed that the rDD sample is in some form of COL/COL2 coexistence where COL and COL2 can each have different concentrations based on some imbalance in osmotic pressure between mesophase domains at 764.5 mg/mL. In Figure 2.46, on the other hand, it would seem likely that the peaks are all somehow embedded in the same mesophase, so that COL2 contains both FHL and conventional hexagonal type lattices in the same domain. These two lattices cannot be achieved by requiring equality of their areas, even though they seem to coexist at the same concentration. One possible escape may lie in the strong correspondence of  $L_{FHL}$  with  $L_{hex}$  in this mesophase. Requiring that these two lateral distance parameters be the same between a hexagonal and an

FHL type lattice results in a formula which can predict  $q_{hex}$  from  $q_{11}$  and  $q_{20}$  where  $q_{11} > q_{20}$ .

$$q_{hex} = \frac{2q_{20}}{\sqrt{3}q_{11}} \sqrt{(q_{11})^2 - \left(\frac{q_{20}}{2}\right)^2} \quad (2.4)$$

This must be contrasted with the form that originally predicted the peak while trying to distinguish Case 1 FHL from Case 2:

$$q_{hex} = \left( \frac{2q_{11}}{\sqrt{3}} \sqrt{(q_{20})^2 - \left(\frac{q_{11}}{2}\right)^2} \right)^{\frac{1}{2}} \quad (C.70)$$

With the peaks obtained from the data set, these formulae give equivalent values for  $q_{hex}$  to several decimal places, but no algebraic manipulation has been identified capable of transforming the first into the second. Area conservation is clearly not required to produce the Case 2 hexagonal lattice from the Case 1 FHL.

Figures 2.47 and 2.48 hold some suggestion as to why it may be possible for these lattices to coexist despite the fact that they do not share the same unit cell area. In both Figures, for 764.5 mg/mL and 781.2 mg/mL, the coherence lengths of all measured peaks do not appear to be amazingly large. The B-form DNA duplex has a footprint nearly 20 Å in diameter, while coherence length in both of these samples never appears to be higher than about 100 Å at lower temperatures, or not greater than about five DNA diameters (or four column spacings where  $L = 26$  Å). This means that the typical scattering domain deviates from the FHL lattice by as much as a lattice parameter length over a distance of about four such unit cells. The rapidity of this disordering increases with concentration, where 781.2 mg/mL has shorter coherence length overall than 764.5 mg/mL. The hexagonal lattice in the 764.5 mg/mL rDD has a coherence of about four unit cells, while the one in 781.2 mg/mL has a coherence comparable to the hexagonal spacing, which is to say about

the size of the unit cell. Between these two concentrations, the coherence of the FHL spacing roughly decreases for peak #2, but remains nearly the same for peak #1, suggesting that the lattice disorders preferentially in one direction over another.

The trend of the hexagonal lattice coherence in 764.5 mg/mL rDD also tells a rather bizarre story. In Figure 2.47, COL2 peak #2 and the hexagonal COL peak remain in fairly similar coherence until the mesophase begins to melt, at which point the packing seems to anneal and grow more coherent by several column spacings until the mesophase vanishes. From this, while the sample is taken to be COL/COL2 coexistence at lower temperature, it is still possible that it's somehow entirely COL2 until it tries to snap into plain old COL at higher temperatures (as seen in Figure 2.33). Again, it is completely clear that there is no required area correspondence between FHL and hexagonal lattices actually seen in these phases, so the absence of correspondence at this concentration would not violate the observation even though the spacing of the dominant hexagonal lattice is not predicted from the spacing of the FHL lattice. Moreover, in Figure 2.45, the trends of  $L_{hex}$  and  $L_{FHL}$  with increasing temperature remain in lockstep --even though the measurements are not the same-- until the mesophase begins to melt, at which point  $L_{hex}$  shifts more freely. These observations could be taken to mean that the hexagonal lattice seen at 764.5 mg/mL is embedded with FHL in COL2 and not merely an instance of coexisting hexagonal COL. Why a hexagonal lattice whose area is not constrained by concentration is allowed to coexist in COL2 with an FHL is not clear, but may imply a more complicated order in the mesophase. Some possibilities will be explored later in this section.

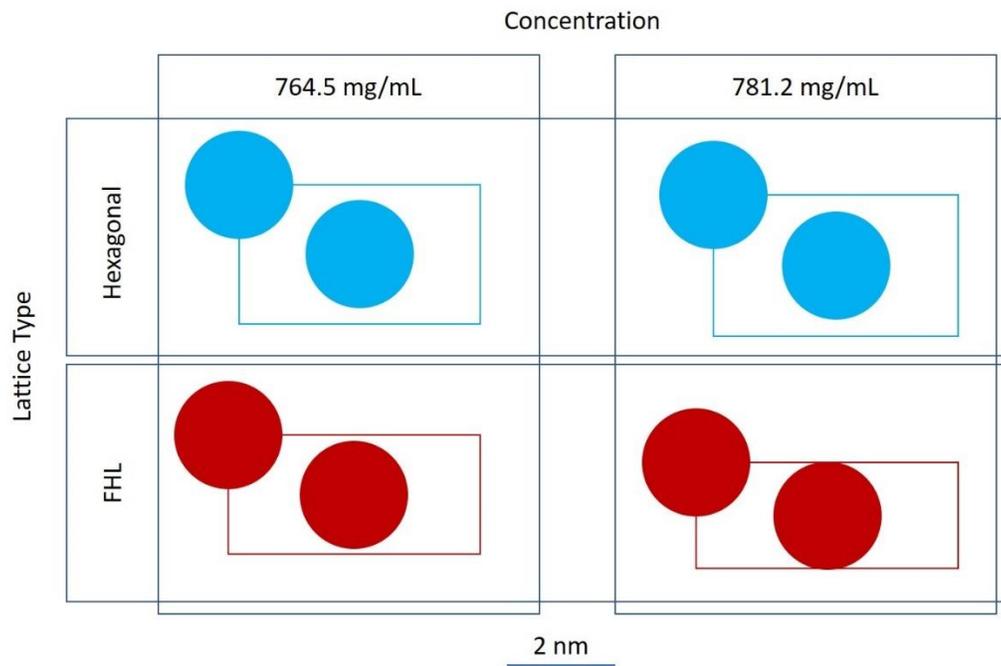


Figure 2.49: Unit cell scale models. Apparent 2D lattices observed at 764.5 mg/mL and 781.2 mg/mL, hexagonal (Blue) and FHL (Red). DNA duplex footprints are 2 nm in diameter matching the scale bar at bottom.

These 2D lattices (Figure 2.49) can be used to render measurements of unit cell area for the FHL with increasing temperature from the data in Figures 2.45 and 2.46.

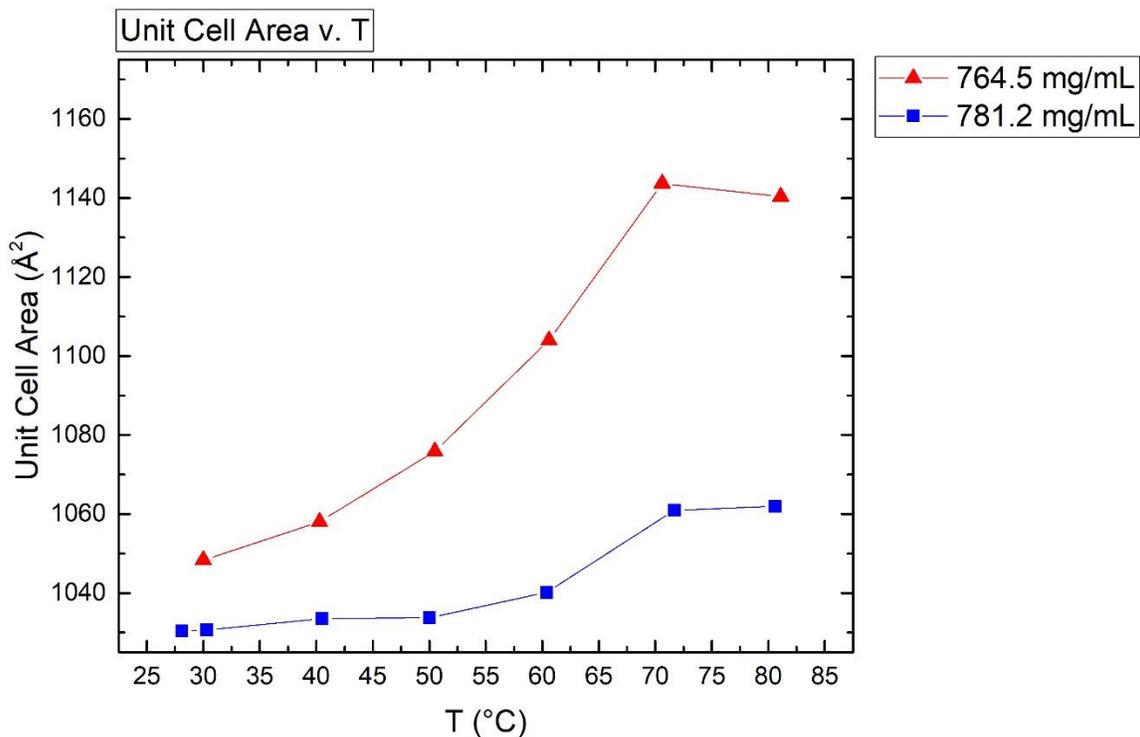


Figure 2.50: Unit Cell Area variation. Cell area in  $\text{\AA}^2$  for the Case 1 FHL model is plotted against temperature for rDD at 764.5 mg/mL (Red Triangles) and 781.2 mg/mL (Blue Squares).

From Figure 2.50, the area of the unit cell very clearly seems to expand with increasing temperature in both cases. Expansion appears to occur to a greater degree at the lower COL2 concentration: about 8% for 764.5 mg/mL and only 3% for 781.2 mg/mL. The reasons for this are not clear, but may be due to differences in viscosity between these concentrations. Expansion data seen here appear to follow a very regular trend and may hint at an underlying law. That is, assuming the structure model ultimately holds up.

In Figure 2.49, at 781.2 mg/mL, it can be seen from the FHL model that neighboring unit cells tend to overlap their DNA footprints. The unit cells in Figure 2.49 can be elaborated into a full FHL lattice to illustrate this overlap (Figure 2.51).

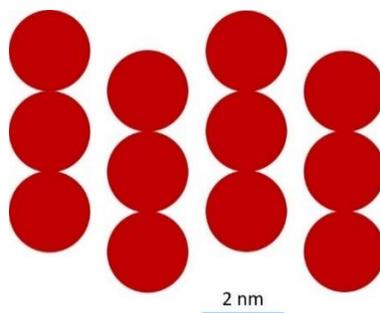


Figure 2.51: 781.2 mg/mL rDD Case 1 FHL scale model. Six unit cells constructed into a scale lattice.

In Figure 2.51, the footprints of neighboring rDD columns appear to brush against one another. Earlier in this section, it was speculated that this sort of situation would become inevitable if mesophases are prepared at concentrations approaching ~960 mg/mL. Provided the lattice assignment is without fault, the Case 1 FHL at 781.2 mg/mL would appear to require some form of aggregate column overlap if the DNA is B-form (and more so if the DNA is A-form since A-form has a larger footprint). Aggregates grazing against one another would be required to fit in a register that optimizes helical packing, with the backbone of an aggregate falling into the major or minor groove voids of a neighbor (see Figure 2.52).

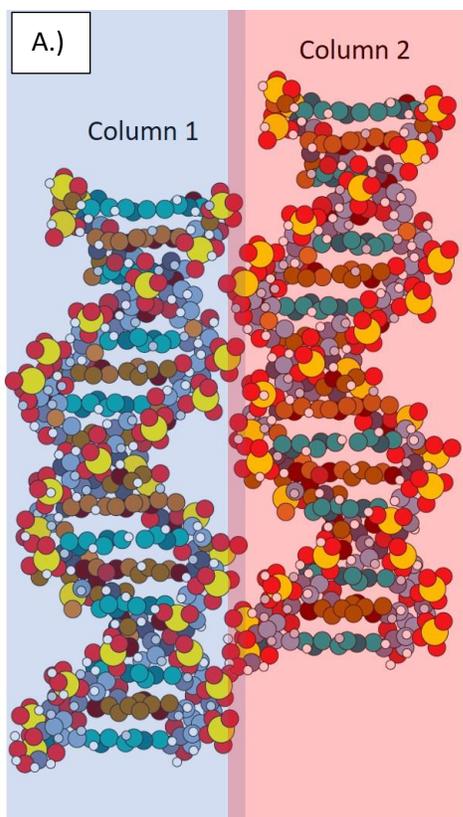
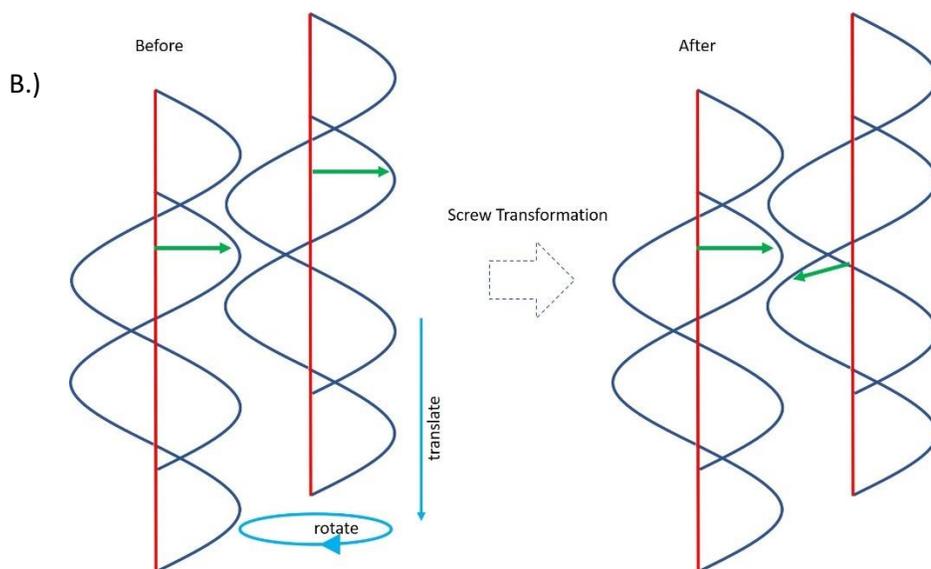


Figure 2.52: Duplex Footprint Overlap A.) Speculative interaction between two columns containing B-form DNA where the footprints of the columns are allowed to mesh. B.) The duplex inside column 2 could be translated along the axis of the column if it is also rotated, in a screw translation, allowing the meshing to continue undisturbed while prohibiting the base stacks between the columns from falling into consistent register from one column to the next. This translation freedom would provide a basis for 3D positional order without actually becoming a fully a crystal.



The kind of meshing shown here would allow the column footprints to overlap, but without forming a full crystal, as detailed in Figure 2.52. This vision is appealing because it could be used to explain the  $0.172 \text{ \AA}^{-1}$  reflection present in both 764.5 mg/mL and 781.2 mg/mL. If DNA duplexes sitting on a Case 1 FHL lattice are

falling into register in order to facilitate an overlap, they would produce a consistent and repeatable register between double helices in neighboring columns, perhaps simply a continual repeat of the pair shown in Figure 2.52. For X-ray waves incident perpendicular to the column axis, this sort of register would produce a lengthwise density undulation that is repeated the length of the column at the period of the DNA helical turn.  $0.172 \text{ \AA}^{-1}$  is very close to the expected  $0.178 \text{ \AA}^{-1}$  period of a B-form DNA helix with 10.4 base pairs per turn and  $3.39 \text{ \AA}$  of rise per base pair. In the hexagonal lattice of COL, without the apparent overlap, there is no set register between neighboring columns, allowing electron density to average along the column axis so that no density undulation appears.

The only other feature of similar size which could compete with this interpretation is the period of the fundamental rDD duplex. rDD is a 12 base polymer: in duplex form, 12 bases at a period of  $\sim 3.4 \text{ \AA}$  per base gives a total length of  $40.8 \text{ \AA}$  for each duplex in the aggregate stack. If these duplexes were somehow falling into register in the stack, this would produce a reflection at  $0.154 \text{ \AA}^{-1}$  in reciprocal space. This is close to  $0.172 \text{ \AA}^{-1}$ , but not as close as  $0.178 \text{ \AA}^{-1}$ . Data clipping near the beam stop may distort peak position measurements in this region, but the period of the DNA double helix is more likely the culprit for this reflection than the rDD stacking period in the column aggregate.

One can envision a simple model for B-form DNA to explain why the double helix period could relax from  $0.178 \text{ \AA}^{-1}$  to  $0.172 \text{ \AA}^{-1}$  in this system.

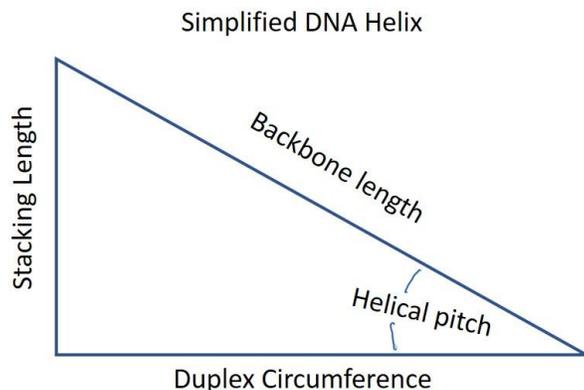


Figure 2.53: Simplified DNA helix diagram. Imagine that DNA is a single helix and that this helix is unwound to place the cylindrical circumference of one turn all in a line. This maps to a right triangle with the stacking length as the vertical side and the backbone length as the hypotenuse with the helical pitch as the depicted base angle.

With regard to the helical pitch (see Figure 2.53), consider that an aggregate column of stacked duplexes is composed of many 4 nm segments of rDD, similar to a longer DNA, but where the backbone of the DNA polymer is discontinuous between rDD segments. Contrary to the cylindrical approximation, DNA base pairs are not disc-like, but rather plank-like and though these planks stack nearly parallel, they turn at angles to each other as the helix winds around. The mismatch of lengths between the backbone connecting each base ( $\sim 7 \text{ \AA}$ ) and the base-stacking period ( $3.4 \text{ \AA}$ ) is responsible for forcing DNA into a helix; the backbone must wind around the base-pairs into order to burn its extra  $3.6 \text{ \AA}$  of length for every  $3.4 \text{ \AA}$  that the base-pair stack rises. The skewed packing between two neighboring base-pairs was forced by the helical twist as a compromise and the hydrophobic surfaces of the base stack would have a chance to pack more efficiently and optimize the  $\pi$ -stacking if this skew is removed. In rDD, every twelfth backbone segment is disconnected, allowing every twelfth base-pair set to pack as it wishes –if the bases pack such that they more completely hide their aromatic surfaces in these locations, the net effect is to untwist the helix by a twelfth. Taking the figure of 10.4 base-pairs per turn as cardinal with a 2 nm diameter duplex, the duplex will have a circumference of 62.83

Å, which corrects the backbone length to a more exact value of 6.93 Å per base in the polymer. Each base would then provide 6.041 Å toward the circumference of each turn. If every twelfth segment of backbone fails to force the twist of the helix, the rise now includes the extra rise over the position that does not twist and is instead, on average,  $(1+1/12)*3.4$  Å for each 6.93 Å segment of backbone, meaning that each backbone segment now provides 5.87 Å per base to the circumference of the duplex turn. For 62.83 Å of circumference, 10.7 bases are now required to complete the turn. 10.7 bases are a stacking rise of 36.38 Å or 0.173 Å<sup>-1</sup> in reciprocal space. In this way, the disconnected nature of the rDD aggregate duplex could operate to expand the apparent helical period of the aggregate.

The rationale offered here takes aim at linking the observation that the Case 1 FHL tends to exhibit a column footprint overlap and how that overlap could be used to explain the 0.172 Å<sup>-1</sup> reflection seen in Figure 2.38. The 0.172 Å<sup>-1</sup> reflection has a very strong impact on the reflections actually seen in reciprocal space given the seeming observation of its many harmonics as in Table 2.2. If the 0.172 Å<sup>-1</sup> reflection is along the column axis, perpendicular to the Case 1 FHL structure, the 1D WAXS data seen throughout this section is a reduction of a 3D diffraction pattern rather than a 2D pattern. To this end, the 0.172 Å<sup>-1</sup> reflection could be taken as a third basis vector in reciprocal space perpendicular to the plane defined by the two basis vectors constructing the Case 1 FHL. Combined with the Case 1 FHL 2D lattice, this proposition identifies the previously unidentified 0.625 Å<sup>-1</sup> and 0.77803 Å<sup>-1</sup> reflections in Table 2.1.

Table 2.4

Peak $q$ (Å <sup>-1</sup> )	$(h,k,l) \vec{A} = 0.13 \text{ Å}^{-1}, \vec{B} = 0.32 \text{ Å}^{-1}, \vec{C} = 0.172 \text{ Å}^{-1}$ (all perpendicular)
0.259848	(2,0,0)
0.278	2D Hexagonal (1,1)

0.344697	(1,1,0)
0.479545	2D Hexagonal (3,1)
0.518939	(0,0,3)
0.625	(2,0,2)
0.694697	(0,0,4)
0.77803	(0,2,2)

The assignments depicted in Table 2.4 should be regarded with a note of caution; the reduction of a 3D diffraction pattern into 1D significantly eliminates information about the diffraction. From previous tables, it must be noted that some indexing solutions that used other basis vector sets have similar peaks. The fit offered here appears to be the most comprehensive and it produces the expected higher order peaks with these three basis vectors. Other assignments for 0.518939  $\text{\AA}^{-1}$  and 0.694697  $\text{\AA}^{-1}$  are regarded to be less likely because of the overall strength of the 0.172  $\text{\AA}^{-1}$  reflection in the 781.2 mg/mL data set that was used for peak indexing. In data sets where 0.172  $\text{\AA}^{-1}$  is weaker, the higher order peaks picked out for this table mostly have dropped into the background, as seen with all data sets of 764.5 mg/mL rDD. A true confirmation of this assignment would be to obtain an aligned COL2 monodomain and to subject it to WAXS with sample rotation in order to collect the entire Ewald sphere. This is the best that can be done with a powder averaged sample.

The remaining loose end in this story is explaining exactly how a hexagonal lattice might be incorporated with the FHL. A natural explanation would be that rDD samples at 764.5 mg/mL and 781.2 mg/mL are both actually coexistence between a hexagonal COL and an FHL COL2 where the differences in lattice area can be explained away by some imbalance in osmotic pressure between COL and

COL2 or some local concentration gradients in the sample. The only hitch in this explanation is that the observed hexagonal lattices increase in unit cell area between 764.5 mg/mL and 781.2 mg/mL, which is backward from the reasonable trend. It has been noted that the poor structural coherence observed in COL2, growing larger with increasing concentration, might suggest a means by which these separate lattices could be neighbors in the same domain. The coherence simply allows a poorly ordered domain to shift back and forth between lattices across its breadth with the hexagonal lattice growing more poorly represented with increasing concentration.

The shared length of  $L_{FHL}$  with  $L_{hex}$  between the observed FHL and hexagonal lattices also suggests another interesting hypothesis. Observed FHL and hexagonal lattices can be superposed with a common z-axis and compared by concentration (see Figure 2.54).

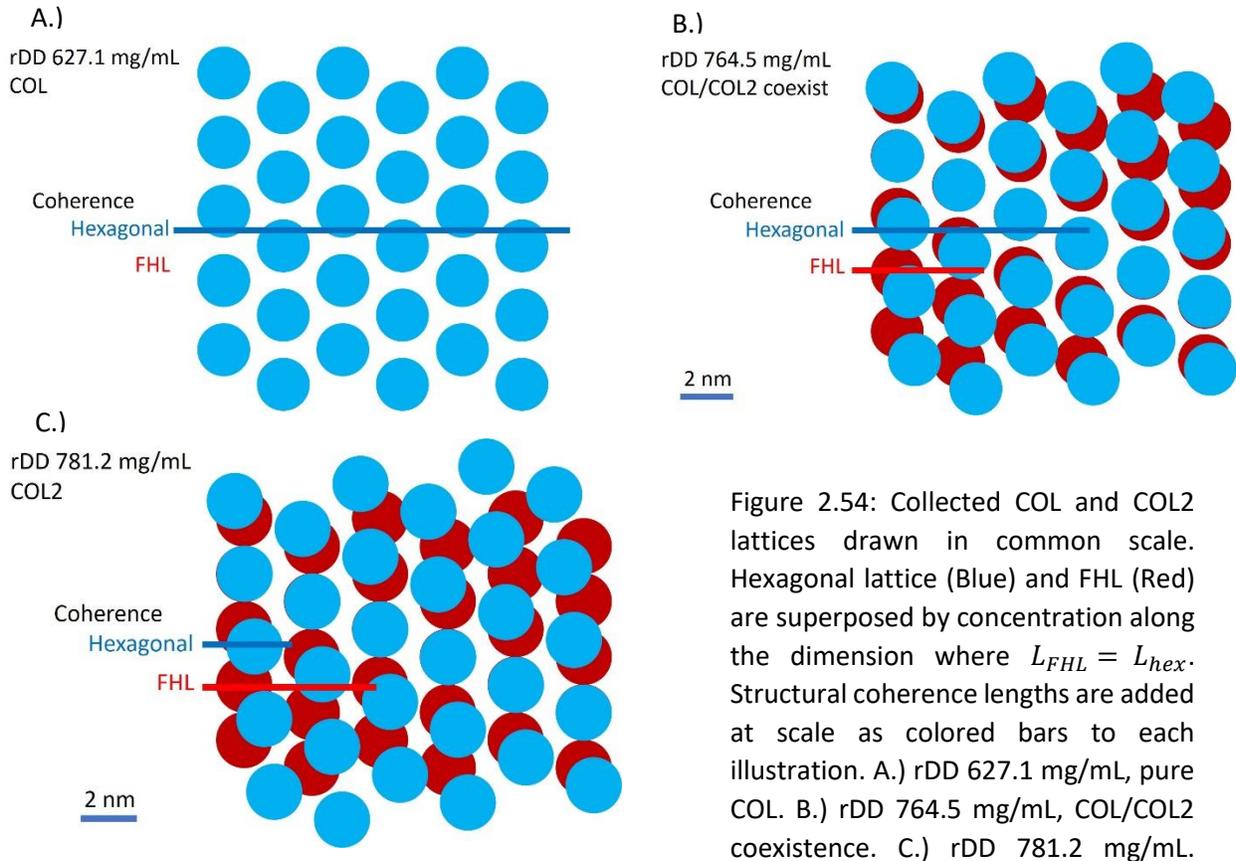


Figure 2.54: Collected COL and COL2 lattices drawn in common scale. Hexagonal lattice (Blue) and FHL (Red) are superposed by concentration along the dimension where  $L_{FHL} = L_{hex}$ . Structural coherence lengths are added at scale as colored bars to each illustration. A.) rDD 627.1 mg/mL, pure COL. B.) rDD 764.5 mg/mL, COL/COL2 coexistence. C.) rDD 781.2 mg/mL.

Figure 2.54 is constructed to collect all the lattices together in the continuum of concentration, using a pure COL concentration as seen in Section 2.2.2 at 627.1 mg/mL as a basic comparison. Each illustration in Figure 2.54 attempts to represent the coherence lengths seen for each lattice at that concentration, showing clearly how the hexagonal lattice possesses a larger coherence length in the pure COL mesophase, which then steadily decreases with increasing concentration. In contrast, the FHL coherence is seen to steadily increase, though never to a large degree, beginning where there is no FHL in COL and progressing through the COL2 samples. The blue footprints are the hexagonal lattice at that concentration while the red footprints are the FHL of the same concentration. Superposition was

picked along a dimension of both lattices, so that  $L_{hex}$  of the hexagonal lattice is matched to an equivalent length represented by  $L_{FHL}$  in the FHL, simply by rotating the blue lattice until a line of footprints could be superimposed. There is no set z-dimension to place a blue lattice in relation to the red below it, but the effect appears to resemble a slow cholesteric-like twist if one visualizes aggregate columns passing through both layers into the page such that they match footprints: pure COL at 627.1 mg/mL is untwisted with aggregates projected straight into the page, while 764.5 mg/mL gains a twist with aggregates passing at angles to one another through both layers, which is then intensified by increasing concentration to 781.2 mg/mL. This apparent deformation is balanced by the structural coherence at a given concentration, suggesting that the way any existent lattices match is on a size scale no greater than the coherence length.

The existence of a twist present in COL2 would not be an unexpected behavior for DNA, which is known to not pack colinearly in its nematic mesophase, giving rise instead to the twisted NEM\* detailed by optical microscopy in Figure 2.3. Chiral DNA aggregates typically pack with a twist when they are permitted to do so. On the other hand, COL2 textures seen in flat cell have a uniform optic axis, which would not be expected in the presence of such a twist that could completely reorient the optic axis over large distances. If the twist exists, and the textures seen in the flat cell are the same mesophase as was detected by WAXS in the capillary, it seems inevitable that the structural coherence length be incredibly short and the mesophase highly disordered for the optic axis to appear to average out as roughly uniform in single domain. The attempt to twist would have to disorder the domain. This may be completely reasonable in light of what is demonstrated in Figures 2.51 and 2.52, where columns appear to be forced into a position where they would

structurally interact side-to-side; in other circumstances, chiral DNA aggregates would almost certainly attempt to escape this by twisting.

The  $0.172 \text{ \AA}^{-1}$  peak is interpreted earlier with fair success as an undulation along the z-axis at the period of the DNA double helix. Some thought was applied to considering how it might also be a beat frequency from X-rays passing through structurally crossing duplexes in a twisted configuration and that the fundamental is not actually at  $0.172 \text{ \AA}^{-1}$ , but  $0.086 \text{ \AA}^{-1}$  or something less in order to give sufficient room for the FHL and hexagonal lattices to actually be coplanar and separated by a consistent register. Unfortunately, the available peaks do not support this view and the peaks seen above  $0.35 \text{ \AA}^{-1}$  in the 3D FHL would still need some means of account. The interpretation that the hexagonal lattice is present only to a tiny degree and that the mesophase is mainly a significantly disordered Case 1 FHL which periodically escapes to hexagonal order would seem more reasonable.

Attempts were also made to fit the hexagonal and FHL lattices together on the assumption that they are not in the same plane and are instead on planes that intersect at a skewed angle. Unfortunately, they do not fit together successfully with a physically reasonable rotation.

There has been some doubt that the flat cell COL2 as noted in Figure 2.3 is the same texture as the 781.2 mg/mL capillary noted in the same Figure. It seems likely that COL+ and COL are almost identical in Section 2.2.2 given the apparently low energy barrier between them and would be anticipated to have the same texture, but the frond-like mesophases in the flat cell called COL2 in Figure 2.3 may actually be the structure noted as COL+. If so, the mostly untextured capillary at 781.2 mg/mL may truly have a highly disordered optic axis due to twisting and might be a completely different texture than the flat cell. Some 4mer nanoDNA samples have given a poorly birefringent texture in this concentration

range and it is possible rDD does the same under the right conditions (see Section 3.1.1). As such, there is the possibility that the texture noted as COL2 in flat cells is actually COL+ and COL2 seen in the capillary is really a COL3. This ambiguity could be remedied with WAXS of samples in flat cells, which has proven challenging due to issues with stably trapping water. The mesophase assignment put forward with the phase diagram is retained because there is not currently enough evidence to justify changing it.

As mentioned previously, the prospect of an oriented sample would also substantially benefit further examination of this mesophase.

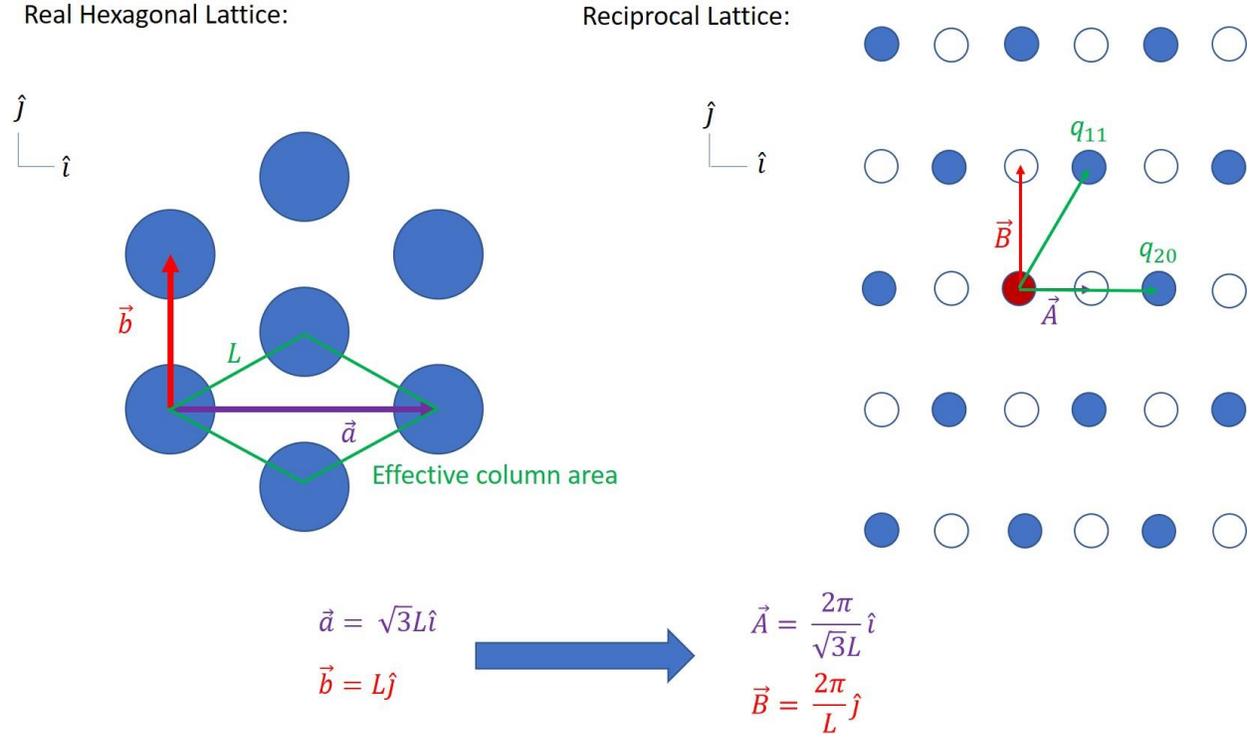
### 2.3 Comparison of rDD Mesophases by Concentration

The three major rDD mesophases contain a large amount of diverse behavior and a few remaining mysteries, including the nature of the COL+ reflection and the exact structural details of COL2. Unlike with thermotropic liquid crystals, this intricate structure emerges from variation of the confinement volume of the mesogen: even different concentrations of the same mesophase are subtly different from one another in structure given shifts of lattice spacing. Similar sorts of behaviors admittedly occur in thermotropic LC, but a lyotropic system contains many more variables in a multidimensional phase diagram covering a space including both temperature and concentration. For the previous three sections, structural details in each phase along the temperature axis were explored. Collecting these phases together, what about the continuum of concentrations crossing lines between the mesophases? Is there a way to generally process the data obtained from these different mesophases and track the variation by concentration? Thus far, this has only been directly attempted once in Figure 2.30 and never

between different mesophases. Available data suggests possible comparisons of structural coherence and lattice spacing both by concentration. COL and COL2 structures can even be seen as commonly amenable given the conservation of lattice area that links them.

### 2.3.1 Theoretical Relationship Between $q_{hex}$ and Concentration

COL and COL2 at least can both be written in terms of the  $q_{11}$  hexagonal lattice reflection  $q_{hex}$ . A simple theory model can be derived from the assumption that columnar aggregates in either of these mesophases do not compress along their long axes with respect to changes in concentration: the DNA stacking periodicity remains constant or very nearly constant in all mesophases and one dimension of the 3D volume occupied by these mesophases is dominated by this invariance. The hexagonal reciprocal lattice reflection  $q_{hex}$  arises from either the  $q_{11}$  or  $q_{20}$  reflection in the structure of the face centered cell, where the variation of the  $L$  parameter that defines lattice spacing is directly related to changes in concentration. Concentration is taken as in Section B.2.2 to be defined as mass/volume, or in this case aggregate mass per total volume occupied by that aggregate where the length of that volume is tied to the invariant length of the aggregate and the 2D cross section of that volume can vary with shifts in concentration. Iterating again the relation between real space and reciprocal space for a hexagonal lattice:



Within the hexagonal lattice of a COL phase, each aggregate occupies the 2D area of one parallelogram facet such that there are as many parallelograms in the lattice as there are aggregates. This area  $A$  can be obtained using the lengths of the characteristics vectors  $\vec{a}$  and  $\vec{b}$ .

$$A = \frac{|\vec{a}||\vec{b}|}{2} = \frac{\sqrt{3}L^2}{2} \quad (2.5)$$

In turn, this area can be directly written in terms of  $q_{hex} = 2|\vec{A}| = \frac{4\pi}{\sqrt{3}L}$

$$A = \frac{\sqrt{3}}{2} \left( \frac{16\pi^2}{3q_{hex}^2} \right) = \frac{8\pi^2}{\sqrt{3}q_{hex}^2} \quad (2.6)$$

If the aggregates are considered to be segmented objects, much like rDD is a stack of 12 linked base-pairs constructed from two antiparallel polymers, concentration  $c$  in a hexagonal mesophase can be considered the mass  $m_u$  of one base-pair occupying the volume  $V_u$  available to that pair.

$$c = \frac{m_u}{V_u} \quad (2.7)$$

Where  $V_u$  is determined by the stacking period of each base-pair  $l_u$  and the Area  $A$  the column occupies in the lattice.

$$c = \frac{m_u}{l_u A} = \frac{m_u \sqrt{3} q_{hex}^2}{l_u 8\pi^2} \quad (2.8)$$

This gives a linear relationship between the square of the hexagonal reflection and the concentration.

$$c = \left( \frac{m_u \sqrt{3}}{l_u 8\pi^2} \right) q_{hex}^2 \quad (2.9)$$

or

$$q_{hex}^2 = Kc \quad (2.10)$$

The constant of proportionality in this equation  $K$  is a function of the stacking period  $l_u$ , which is known to be about 3.4 Å with DNA and a unit mass  $m_u$  that can be calculated by case.

The unit mass can be specialized for rDD, which has a molecular weight  $M$  of 3636 g/mol in its fully deprotonated form. Each unit of rDD duplex stacked in a liquid crystalline aggregate is made of two polymers where the number of polymers is  $N_p$  and where there are 12 units  $N_u$  in a stack. The molar mass can be converted to actual mass using Avagadro's number  $N_{av} = 6.02 \times 10^{23}$  molecules/mol where the number of molecule complexes in the volume of interest is assured to be one. This leads to a form for  $m_u$  which can be adjusted by case beyond just nanoDNA.

$$m_u = \frac{N_p M}{N_u N_{av}} \quad (2.11)$$

For rDD, one mass unit  $m_u = 1.0066 \times 10^{-21}$  g.

As  $c$  is usually taken to be in mg/mL for this work, a necessary unit conversion is to force  $l_u$  and  $q_{hex}$  from Å to cm and use  $1 \text{ cm}^3 = 1 \text{ mL}$  and take g to mg. For this  $1 \text{ Å} = 10^{-8} \text{ cm}$ .

$$N_c \frac{g}{\text{Å}^3} = \frac{mg}{mL} \rightarrow N_c = \left( \frac{\text{Å}}{cm} \right)^3 \frac{mg}{g} = 10^3 10^{24} = 10^{27} \quad (2.12)$$

And from this, the constant of proportionality can be calculated.

$$K = \frac{8\pi^2 l_u}{\sqrt{3} m_u N_c} \quad (2.13)$$

For rDD, as specified,  $K = 0.00015397 \text{ mL}/(\text{mg} \cdot \text{Å}^2)$ .

With  $K$  in hand, the relation eqn 2.10 can be used to calculate a theoretical absolute end-point ( $c_{max}, q_{max}^2$ ) by taking the maximum concentration equal to the DNA density  $c_{max} = \rho_{DNA}$  and calculating an appropriate  $q_{max}$ .

For rDD, this theoretical maximum point is (1687 mg/mL,  $0.2597 \text{ Å}^{-2}$ )

This theory can be extended across concentrations of COL and COL2 both by using equation C.70 to calculate a  $q_{hex}$  reflection from the appropriate COL2 lattice.

The theory line acquired here can be regarded as a perfect condition, absolute maximum line that specifies what  $q$  values can possibly be seen for ideal samples of double-stranded rDD nanoDNA on a hexagonal lattice. This line can be used to check cases where the aggregate complex in the hexagonal lattice is not necessarily a duplex, like DNA, but possibly a triplex or a quadruplex (by varying  $N_p$  to 3 or 4) (as in Sections 4.1 and 4.3 and Figure 4.10 with dNTP) or some other complexation on this lattice, or also modifying to incorporate other nanoDNA polymers by varying polymer length  $N_u$  or  $M$  to other values (as in Section 3.1.3, Figure 3.19 for GTAC).

### 2.3.2 Comparison of rDD Hexagonal-Spacing to Theory

Using the theory introduced in Section 2.3.1, data obtained for rDD mesophases can be compared against the theory for incompressible aggregates on a 2D hexagonal lattice (equation 2.10). For this data, a  $q_{hex}$  is calculated for COL2 from the model of the Case 1 FHL.

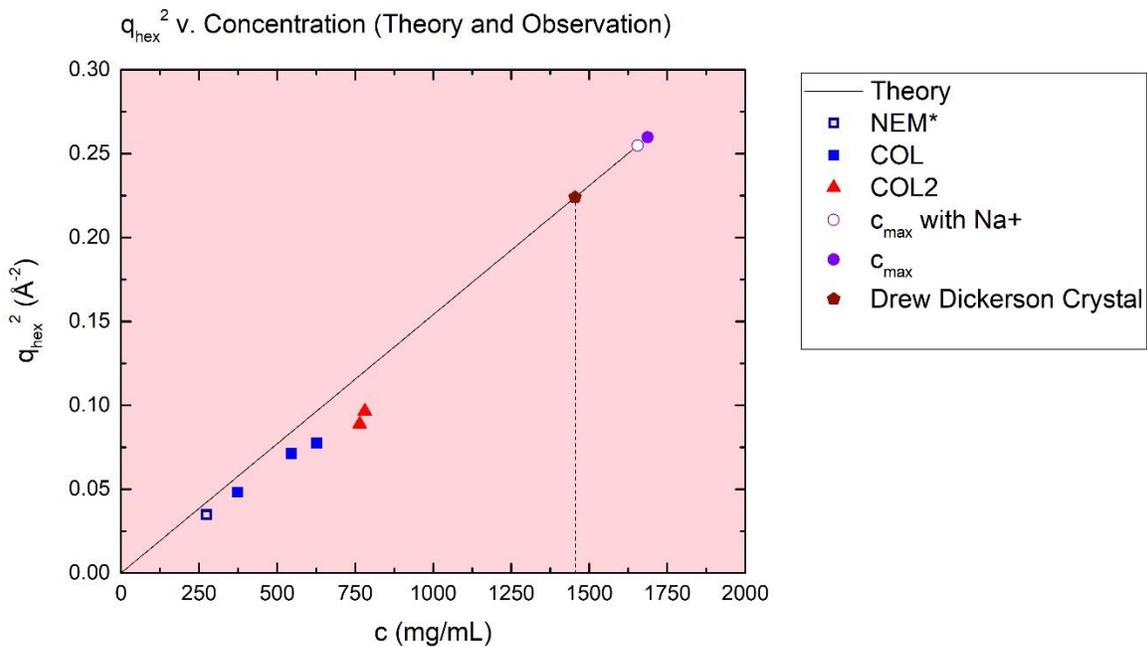


Figure 2.55:  $q_{hex}^2$  versus Concentration. Relation of NEM\* (Unfilled Blue Square), COL (Blue Squares) and COL2 (Red Triangles) as compared to the theory (Black Line) from Section 2.3.1. Several maximum possible values are added:  $c_{max}$  (Violet Circle) is the absolute maximum density possible for duplex DNA if the DNA fills all space with no gaps;  $c_{max}$  with Na+ (Violet Circle with White fill) is the maximum realistic concentration where the DNA is present with a sodium counter ion (equation B.29 with  $G_2=0$ ). Also added is the calculated concentration for the Drew Dickerson crystal structure (Burgundy Pentagon with the dotted line)<sup>35</sup>. Note- NEM\* is not on a hexagonal lattice; room temperature side-side  $q$  is simply squared and plotted.

Figure 2.55 is a good comparison of NEM\*, COL and COL2 data presented previously. Data for NEM\* is not understood to be on a hexagonal lattice, so the

point plotted in the figure is simply the  $q$  for the side-side correlation squared.  $c_{max}$  sets the upper limits on the theory line, showing where the absolute highest concentration exists.  $c_{max}$  with  $\text{Na}^+$  adjusts the maximum concentration point to accommodate sufficient sodium in the phase to fully cancel the charge of the rDD phosphodiester backbone. The line and point designated “Drew Dickerson crystal” calculates the concentration of a crystal cell from Wing *et al.*<sup>35</sup> which details the X-ray crystal structure associated with  $5'\text{CGCGAATTCGCG}^3$ , the prototypical 12mer DNA which served as the original nanoDNA LC (called elsewhere DD). The Drew Dickerson crystal is an absolute concentration mesophase which is locked in fully crystalline order, from which an X-ray structure for B-form DNA was originally solved –the crystal structure is not hexagonal, but contains 3 two-fold axes on a primitive cell  $p2_12_12_1$  and is added here only to compare current data to the concentration where that structure was found<sup>35</sup>.

The theory line offered here is the ideal hexagonal packing situation for completely accurate concentration. The real data falls below the theory line due to inaccuracy in the concentration; the concentration for these samples is calculated based on the assumption that the mesophase contains only water, DNA and sufficient sodium ions to completely neutralize the DNA charge, which does not accommodate the prospect of spectator contamination. For real DNA, purification is an act of trying to minimize excess salts that inevitably end up in the preparation and minimize the presence of damaged DNA, but removal of these requires an acrobatic balance since removing too much salt can result directly in DNA damage. For the mesophase samples above, the actual weight of the DNA sample includes a small unknown amount of these contaminants ( $G_{con} \neq 0$  in equation B.28), meaning that the concentration always overestimates the DNA mass by assuming that all of the measured mass is mesophase forming DNA, when it actually is not. Clearly, the

WAXS experiment could be used to calibrate for sample contamination by running this whole process backward, but confirmation would require an independent measure of structure.

The technique here appears a useful comparison and has been applied elsewhere using rDD measurements as the standard case.

#### 2.4 Instability in the rDD Phase Diagram

A large case is made for the notion that the original DD phase diagram (Figure 2.1) is inaccurate due to difficulties in measuring concentration in the experimental formats used previously<sup>25</sup>. Given that the concentration measures offered in the updated phase diagram (Figure 2.2) are quite possibly overestimates as well, from Figure 2.55, the original measurements would most likely also fall below the theory line. The truth, however, remains that this might be a somewhat simplistic view. Phase diagrams for DD and rDD have proven quite challenging to reproduce experimentally.

One important reason for this appears to be that nanoDNA mesophase are not due completely to aspect ratio and excluded volumes as in the Onsager criterion (equation 1.1). One very important facet of the mesophase which is still somewhat undefined is the contribution from charge and counter ion content. Attempts to completely validate the identity between rDD and DD samples, which would be expected to have identical phase diagrams on the strict basis of aspect ratio, initially failed. One phase diagram pushed the concentrations where phases appeared to inexplicably higher values, with NEM\* and COL phases appearing at much higher concentration. After some significant leg work, the culprit of this shift was identified to be a dialysis treatment of one of the samples (see Figure 2.56).

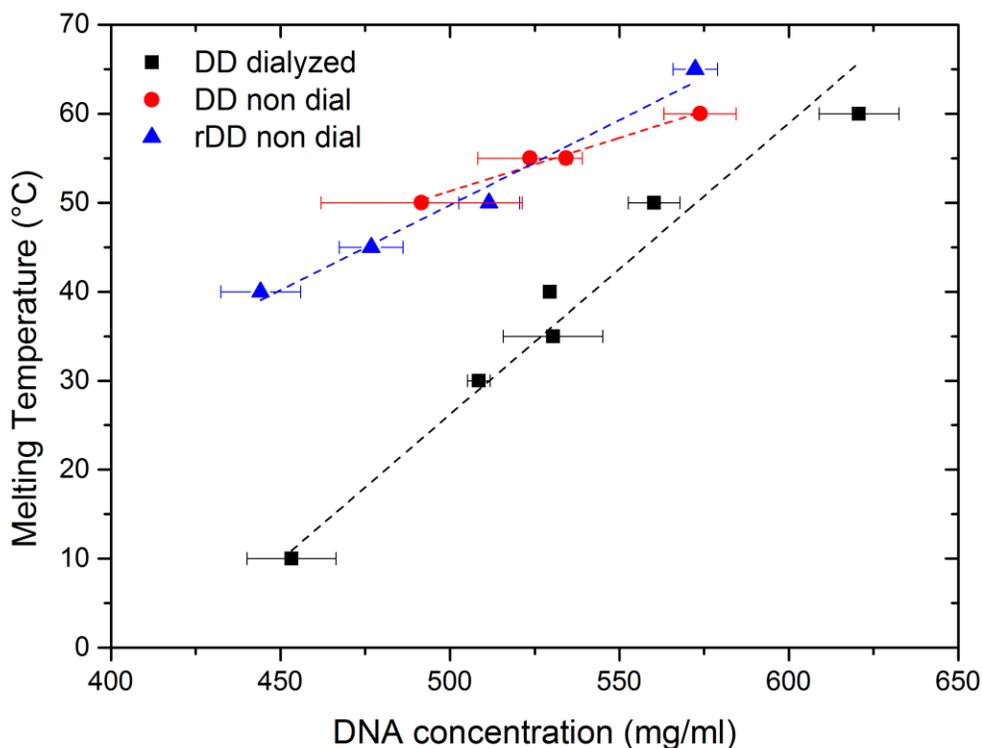


Figure 2.56: Melting temperature versus Concentration. Used with permission from Tommaso Fraccia. Comparison of melting temperatures of dialyzed DD (Black Squares) and undialyzed DD (Red Circles) with undialyzed rDD (Blue Triangles).

In dialysis, a large molecular weight sample is partitioned within a semipermeable membrane with a pore size smaller than the sample of interest. This permeable partition is then placed in contact with a buffer reservoir and molecules cross through the membrane due to chemical potential based on whether or not they fit through the pores. In the case of nanoDNA, oligomers remain trapped within the dialysis cell while water, salts and counter ions can cross out of the cell. The requirement that a solution be electrically neutral does not strictly mean that equal charges are isolated exactly where they are expected in the situation where the solvent is polar like water and can act as a dielectric, or even create charges by deprotonation as an acid. This means that counter ions for nanoDNA backbone

charge can cross out of a dialysis cell into the surrounding reservoir, leaving the DNA behind, and while the solution appears electrically neutral, different ions end up located in different places. When the DNA is then collected to make mesophases, the tuning of the counter ion and salt content result in shifts to the phase diagram (see Figure 2.56).

The upshot is that the polarizability or acidification of pure water is not equal to the presence of counter ions. Backbone charges of the nanoDNA aggregates are not suppressed in highly concentrated LCs when those charges are now forced into close proximity with one another. If counter ions are removed from the sample, the nanoDNA phase diagram tends to shift to the right in concentration space, with lower order mesophases occurring at higher concentrations, as if aggregates are less prone to sit next to one another due to insufficiently suppressed backbone charge forces. So, dialysis treatment pushes the phase diagram toward higher concentrations. In the extreme, removal of counter ions --in particular, by a high-quality size exclusion chromatography column-- can totally eliminate nanoDNA mesophases. It seems likely that insufficiently suppressed charge in the phosphate backbone can, in the extreme, overwhelm base pairing and stacking to rip the DNA duplex apart, which would also appear as isotropic phase occurring at higher concentrations than expected.

As such, the nanoDNA phase diagrams are polyvariate, depending not only on concentration and temperature, but also on pH, counter ion identity and concentration and on salt content, to name only a few. As salt and counter ion content hinge on purification, shifts in purification technique can strongly impact the phase diagram of nanoDNA samples. Further, preparation of nanoDNA samples strictly by measurement of component weight assumes intrinsically that the densities of all of the components remains constant, which may not be true if

electrostatic interactions influence the volumes any particular components might tend to occupy, which could also explain deviations in Figure 2.55.

## Section 3.0

### 4mer nanoDNA Constructs

It is generally expected in Abiogenesis research that oligonucleotides of DNA or RNA cannot be truly basic molecules since they must come from somewhere, even if they are only 12 nucleobases in length. Such an oligonucleotide is already a complicated structure. The capacity for LC to template crosslinking of oligomer aggregate columns can make longer sequences from shorter ones, but the absolute minimum possible length at which this affect occurs was not found to be zero nucleobases or simple monomers. The original investigations of mesophases formed by sub-Onsager length nucleic acids found LC at lengths down to 6 nucleotides. Early work with deoxynucleoside monophosphates, both as plain monophosphate and cyclic monophosphate, found no conditional evidence of LC with all bases but dG, which has been known to form mesophases based on G-quartet or G-quadruplex structures in an independent, non-Watson Crick variety. If LC bootstrapping cannot be executed with oligomers shorter than 6 bases, where did those 6 base objects come from? A longstanding direction of research has therefore been to push the envelope and examine oligomers shorter than the original 6 nucleotides to see if conditions exist where the previously understood limits become flexible. Is 6 bases really the limit, or can shorter oligomers form mesophases as well?

Continual adjustment of the DNA synthesis process eventually revealed that oligomers purchased from commercial vendors appear optimized for use in a manner that is detrimental to materials science applications. The most common users of synthetic DNA are molecular biologists whose typical methods involve diluting DNA oligomer samples to  $\mu\text{M}$  or  $\text{nM}$  concentrations, giving them the opportunity to dilute away spectator ions present in the sample. From common purification techniques, we already know that at least one possible spectator, triethylammonium, is able to disrupt nanoDNA mesophase formation. This ion is an ubiquitous HPLC counter ion known not to interfere with molecular biological experimentation. That it appears detrimental to LC phases is happenstance. With the revelation that commercial materials are not necessarily friendly to our applications, we developed a handful of techniques to stabilize nanoDNA samples for the application of mesophase formation. Altering cleanliness of our samples revealed nanoDNA LC phases in 4-base, 3-base and 2-base systems.

The 3-base and 2-base systems are not well characterized as yet, but the 4-base systems have matured. This section will be focused on work in two of the 4-base systems. The existence of these results has brought evidence of nanoDNA LC phases from previous endpoint of 6-bases down to mere dimers, spanning nearly the entire short-oligomer spectrum.

### 3.1 Blunt-End 4mer $5'\text{GTAC}3'$

The first system that will be detailed in this section is work performed on the Blunt-end 4mer,  $5'\text{GTAC}3'$  (GTAC). GTAC was designed as a minimum version of the self-complementary palindrome design that is exemplified by the Drew Dickerson Dodecamer mentioned in Section 2.0. This design typically contains a bracket of

triple hydrogen bonded G-C base pairs surrounding a core of double hydrogen bond A-T pairs; the idea being to try to enforce stability at the exposed ends of the duplex with the maximum number of hydrogen bonds possible to prevent the duplex from unzipping at the ends. This may be a retroactive consideration on our part since the original DD construct was designed around an EcoR1 restriction endonuclease enzyme target site with perhaps minimal consideration for structure otherwise<sup>35</sup>.

The reader should know that while a great deal of work has now been done with GTAC, there are still some abiding mysteries that appeared with this system that have yet to be resolved. GTAC was initially synthesized at a time when our in-house DNA synthesis was finally beginning to show some success after chasing down some pesky flaws persisting in the system. The original quality of this oligomer was poor and early examination showed LC phases that did not form without refrigeration to 5° C. GTAC was a helpful synthetic target because it not only makes LC phases, but is a short enough oligomer that it can be produced without consuming large quantities of synthesis reagents, making it possible to test alterations to the synthesis and purification methods without wasting enormous amounts of money. That GTAC was an LC producing nanoDNA formerly in a forbidden part of sequence space was a helpful extra at the time which made it feel as if the ultimate research goal was progressing forward despite the synthetic bottleneck. A later batch of GTAC, upon which much characterization was performed, increased the ISO phase transition temperature to 40° C, more than 35° C higher than where it was initially observed, giving room temperature LC phases. This shift in the phase diagram has not been reproduced.

GTAC is also the source of the observation characterized in Section 2.4. At that time, the hang-ups of the salt exchange purification technique (Section A.6.2) and the difficulties with triethylammonium (Section A.6.3.1) were known and other

strategies were being explored. Usage of a commercial size exclusion column had shown some potential for producing LC forming material after HPLC purification, helping to remedy the triethylammonium problem without losses present in rotovapping, but that the material purified by size exclusion also faced large losses given the small size of the column being used. The column was simply not long enough to separate a small oligomer like GTAC from contaminants present in the column void. When the column was scaled up to produce a longer path length, definitively separating the GTAC band from the void volume, the GTAC totally ceased to produce LC. It turned out that this situation could be reversed by equilibrating the same size exclusion column to 10 mM NaCl and the dead GTAC passed back through it. GTAC previously incapable of making LC began to make LC phases again. It was reasoned that the initial condition of the size exclusion column, equilibrated against deionized water, was sufficient to suck Na<sup>+</sup> and triethylammonium counter ions both away from the nanoDNA, allowing electrostatic charges in the GTAC phosphodiester backbone to rip the duplex hydrogen bonding apart, prohibiting LC phases, and that equilibration with a salt restored sufficient counter ions to screen the charge on the oligomers enough to base pair and form LC again. A softer version of this result was replicated with dialysis in Figure 2.55, helping to nail down a source of variability in nanoDNA systems. It has been reasoned that something similar to this was the source of the earlier jump in LC-ISO phase transition temperature in the GTAC system, but this is not confirmed.

### 3.1.1 5'GTAC3' LC Phase Textures

GTAC has a complicated history of LC textures. The initial observations performed in oil-sealed flat cells offered not only approximate concentrations based

upon preparation, but poor resolution of the temperature dimension (see Figure 3.1). Concentrations produced by this method are typically more accurate for lower order mesophases because evaporation shifts the contents of the cells more quickly when there is less water to lose: as such, error will tend to favor calling a higher order phase at lower concentrations than where it might actually be encountered. Moreover, the early materials used to examine these textures had an apparent purity-based deficiency which resulted in an ISO transition temperature that was very low, typically between 5° C and 10° C. Despite this, the textures encountered strongly resemble those seen previously (see Figure 2.2).

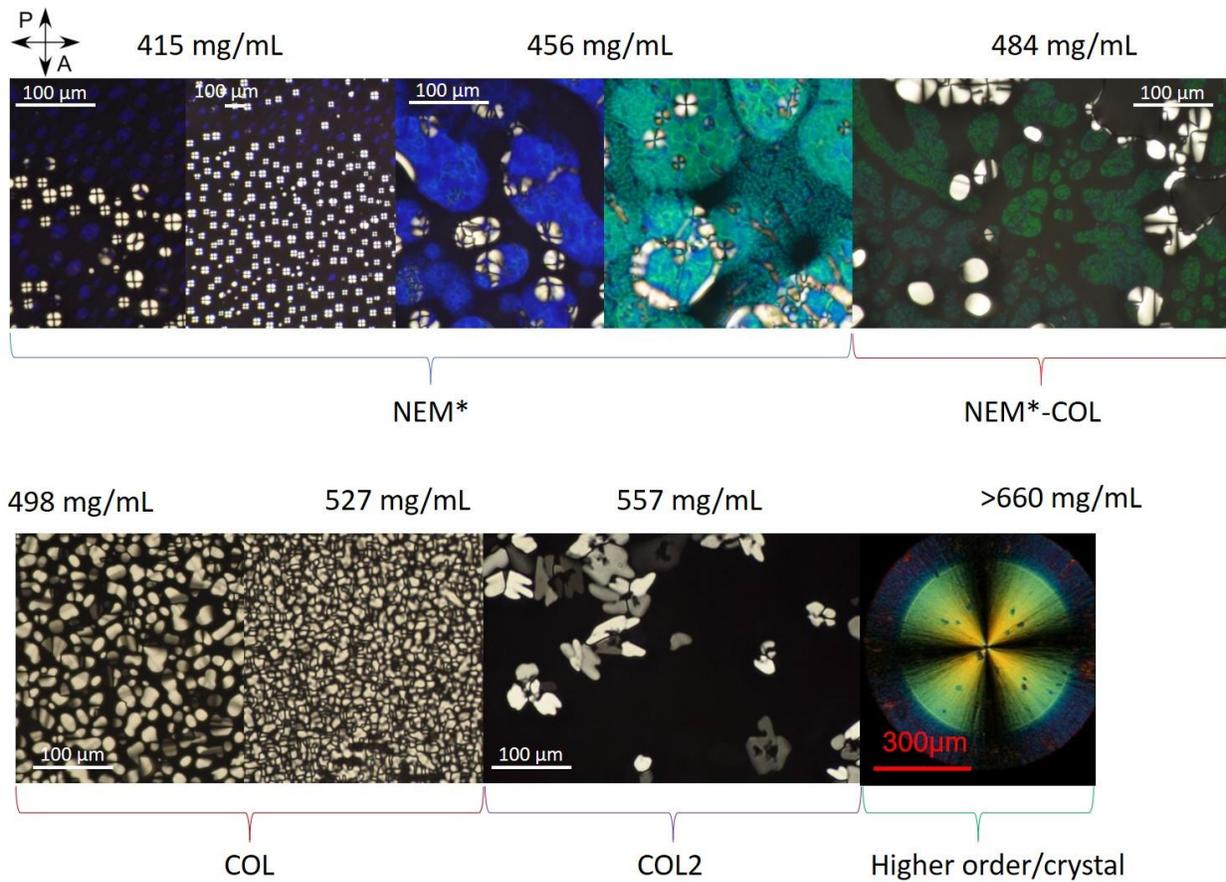


Figure 3.1: GTAC concentration phase diagram at constant temperature. Mesophase textures illustrated in this image are marked with increasing (approximate) concentration proceeding from left to right, showing at lowest concentration ISO-NEM\* coexistence, NEM\*, NEM\*-COL coexistence, COL, COL2 and Crystal-like phases at the highest

concentrations. Error in calculated concentration increases with increasing concentration. Scale is 100  $\mu\text{m}$  where shown and textures are identified as written.

As identified in these earliest images (Figure 3.1), NEM\* is understood to be a chiral nematic based on the appearance of the colored cholesteric. At its lowest concentrations, NEM\* showed focal conic brushes on the small domain droplets that are not conformal, suggesting the existence of twist or splay deformation. Schlieren are not clearly apparent in these phases beyond the existence of the focal conics. Typically, NEM\* mesophases changed from focal conic to colored after a domain droplet coarsened to fill the entire thickness of the cell. Since the domains are colored, the cholesteric twist must have a pitch of several hundred nanometers (matching the wavelength of the color), meaning that any fingerprint textures seen in this phase would be too fine to resolve optically. This suggests that what appears to be a focal conic texture is actually very fine cholesteric fingerprints where the axis of the helix is parallel to the sloping surface of the droplet (Figure 3.2), enabling the microscope to see across the axis of the cholesteric. As the droplet fills in and the slope becomes vertical with respect to the microscope, when the domain stretches from the front to the back of the cell and the droplet edges become parallel to the direction of observation, the resulting colored domain is some form of Grandjean texture with the axis of the cholesteric helix going straight into the image. In these early observations, the color of the cholesteric appeared to depend somewhat on the concentration, violet at lowest concentrations and shifting through blue to green as the concentration was increased. This trend did not necessarily hold when later samples of GTAC with room temperature mesophases were produced. The colored phase typically had a mottled appearance, differing somewhat from later observations.

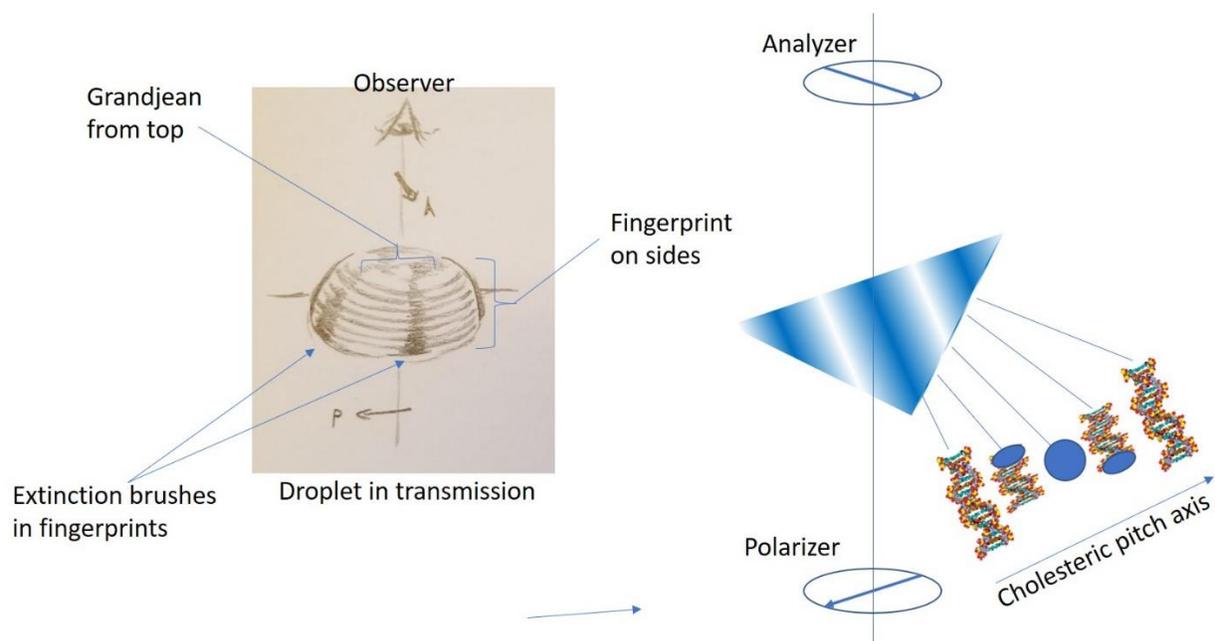


Figure 3.2: Cholesteric fingerprint diagram. Hemispherical droplet sits on a plane with light passing through normal to that plane. Fingerprints can become visible as dark and bright fingerprint-like bands if the axis of the cholesteric helix is parallel to the curved surface of the droplet, such as when the droplet is a convex or lens-like surface, as drawn, where the local orientation of molecules along the surface turn around the helical pitch of the mesophase, becoming dark if the nanoDNA is homeotropic to the surface or bright if oriented planar to the same surface one-quarter helical pitch later. If cholesteric pitch axis is turned normal to the surface, the DNA aggregate orientations average out to eliminate a preferred fast axis and the texture becomes dark or grayish Grandjean. The texture can become colored Grandjean if the cholesteric helix has a pitch on the size scale of visible light wavelengths, enabling it to reflect light interferometrically. The droplet can have extinction brushes where the bands of planar aligned DNA in the helix are oriented with their base pairs along the axis of either the polarizer or the analyzer, allowing the droplet to look like a focal conic if the fingerprint has a pitch too tight to resolve.

COL phase seen in these earliest observations is indistinguishable from other nanoDNA COL textures (see Section 2.0). From the observed extinction brushes, GTAC produced conformal, bend-only focal conics. Further, these textures were the same between early observations of GTAC which had only low temperature phases and later versions that were room temperature.

COL2 texture also appeared to be very similar at first examination to COL2 textures seen elsewhere, having leaf-like or frond-like shapes. These domains

typically had a uniform director so that the entire domain underwent extinction at the same orientation relative to the microscope polarizers, implying very rigid unidirectional order.

At higher concentrations still, GTAC showed some highly ordered crystal-like mesophases that could display focal conic or fan-like structure, but appeared to have much more rigid texture than COL.

These mostly familiar results in the early observations were balanced by some more mysterious behaviors that remain incompletely resolved due to an inability to convincingly reproduce them with later GTAC samples.

The first peculiar behavior encountered was an apparent metastability of the COL phase where the COL was seen to fill in, but then spontaneously transition to an ambiguous phase with very poor birefringence (Figure 3.3).

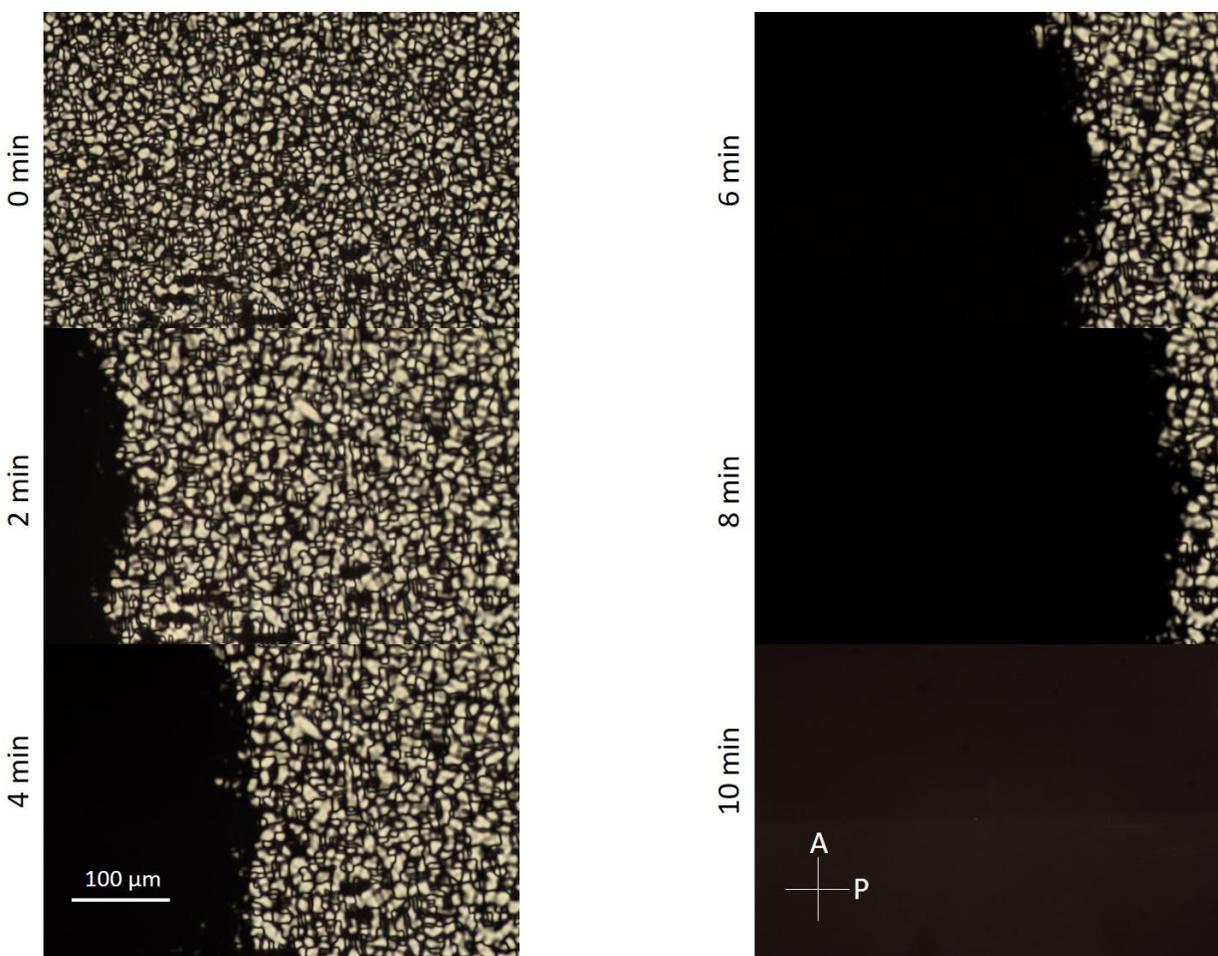


Figure 3.3: Metastable COL texture versus time.  $\sim 530$  mg/mL GTAC observed by PLM at  $5^\circ$  C every 2 minutes after the COL mesophase has filled in. Scale is  $100 \mu\text{m}$  and polarizers are as shown. Dark phase can be seen to fill in from the left, converting COL to the dark texture.

The form of metastability seen in Figure 3.3 was not reproduced in GTAC samples that had room temperature phases. The poorly birefringent mesophase that appeared here did not possess discernable textural features. Within these samples, the appearance of COL metastability was a reproducible effect, suggesting that some quality of the synthetic preparation was influencing the LC phase stability. There was some speculation that COL textures could be converting spontaneously to homeotropic anchoring in Figure 3.3, but further evidence to this

end did not bear out. Evidence from other cells suggests a monolithic, poorly birefringent mesophase.

A second mystery phase, which may be related to the one that appeared in Figure 3.3 is implied by the COL2 image in Figure 3.1. In that image, the COL2 frond textures do not completely fill the area of the cell; the dark portion of the cell initially appears to be a large amount of ISO coexistence. Closer examination of this dark texture revealed a mosaic pattern (Figure 3.4).

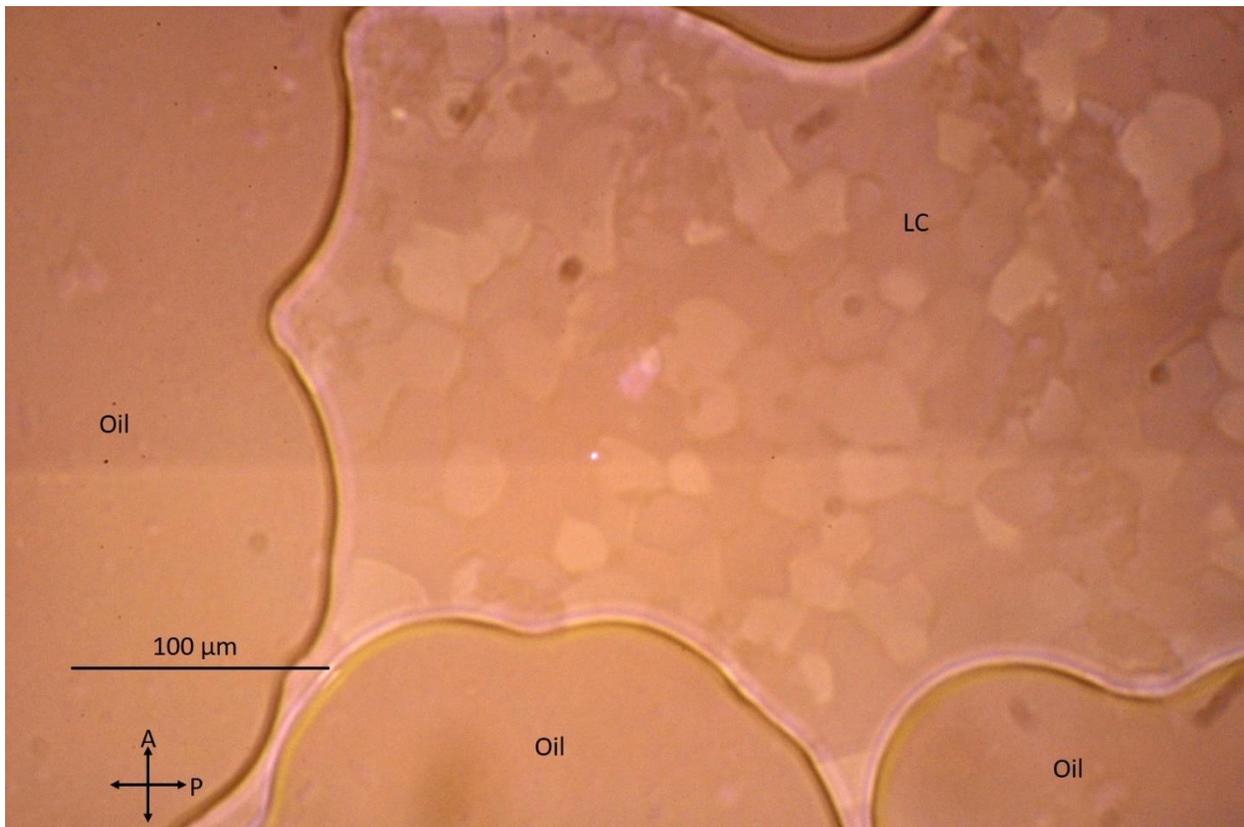


Figure 3.4: GTAC mosaic texture at  $\sim 560$  mg/mL and  $5^\circ$  C. In this image, the sample is highly overexposed so that the isotropic background (at the oil) now appears an orange color. The image was then manipulated for brightness and contrast to make the mosaic tile pattern more apparent. Polarizers and scale are as shown.

For GTAC at cold temperatures, the mosaic pattern shown in Figure 3.4 is coexistent with the conventional COL2 textures (Figure 3.1). While this texture first appeared in GTAC possessing phases only near the freezing point of water, it was

later reproduced in GTAC that has room temperature mesophases. Early speculation about this mesophase included that it was a Blue phase of some sort, which is similar in some ways with its helical order to a cholesteric and would not be completely implausible for nanoDNA, which is known to make chiral structures. Despite significant efforts, further evidence identifying this mesophase as a Blue phase has not been obtained. One complication preventing this goal is the attenuation of concentration due to evaporation between flat cells and capillaries that could be used to examine nanostructure by X-ray diffraction.

Once GTAC with room temperature mesophases was acquired, most of the textures mentioned above in Figure 3.1 were reproduced. The single exception was the metastable COL seen in Figure 3.3. GTAC with room temperature mesophases never showed metastable behavior. Specific effort was directed at finding the mosaic texture in Figure 3.4, particularly in cells with a long enough path length to facilitate X-ray scattering experiments. With small pathlengths (1-4  $\mu\text{m}$ ), the mesophase was poorly birefringent (as in Figure 3.4), but when the path length was scaled up to  $\sim 50$   $\mu\text{m}$ , brighter texture with better feature contrast appeared (Figure 3.5).

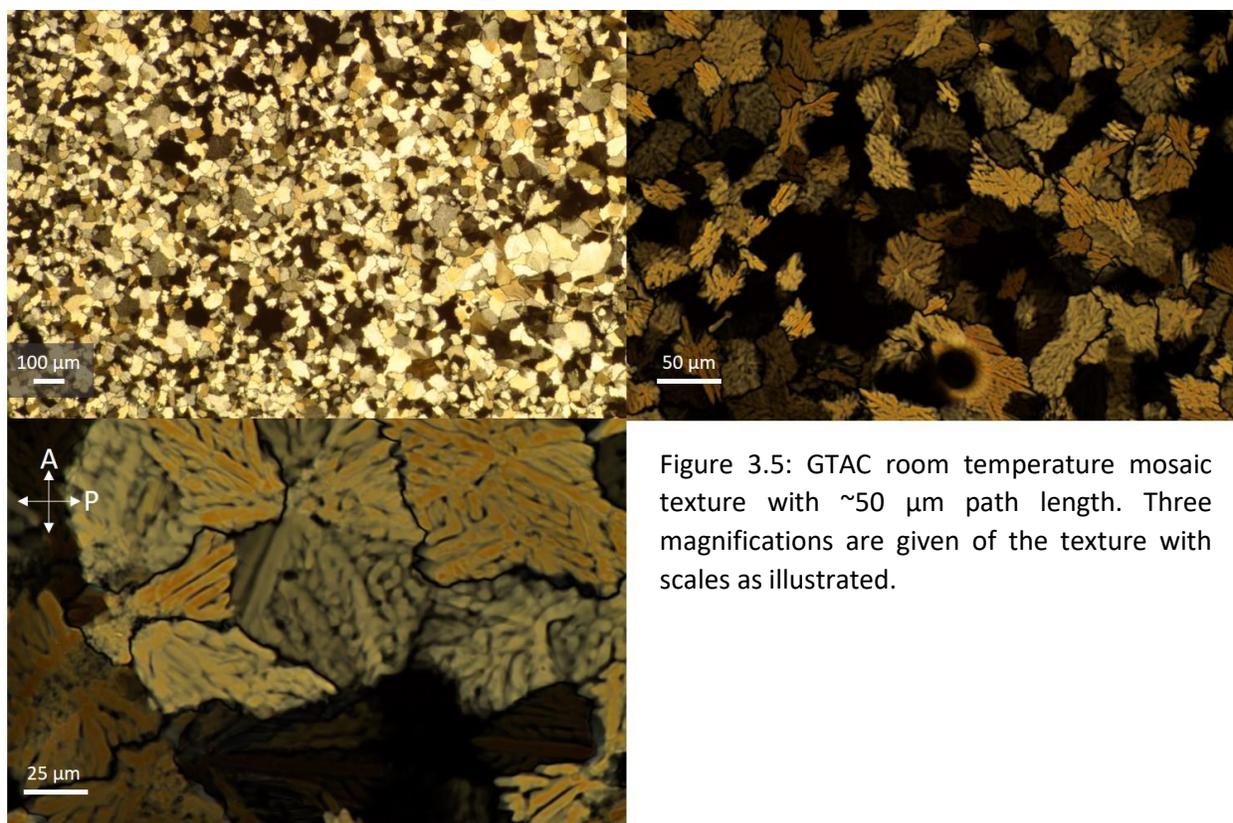


Figure 3.5: GTAC room temperature mosaic texture with  $\sim 50 \mu\text{m}$  path length. Three magnifications are given of the texture with scales as illustrated.

In the case of GTAC with room temperature mesophases, the dark texture seen before is still dark when seen with a short pathlength. With a scaled up pathlength, the texture has weak birefringence (as in Figure 3.5) given that the birefringence color of the phases still seems to be in the first order by the Michel-Levy Chart despite the huge pathlength increase. The locking structure of the domains clearly appear as if they are mosaic tiles at low magnification, but the morphology of the domains resembles some forms of COL2 seen previously, with branching, frond-like shapes that apparently contain a uniform director structure that extinguishes across the entire domain simultaneously with respect to polarizer or analyzer. This hints at highly restricted order. Whether this texture is exactly the same as that seen previously at low temperature and in cells with short pathlength is not completely certain but the similarities are compelling.

Several intriguing properties emerged in NEM\* in GTAC possessing room temperature mesophases. It seems likely that these observations were not possible previously given the extremely tight ( $\sim 3^\circ\text{C}$ ) temperature range where the low temperature version of NEM\* was observed.

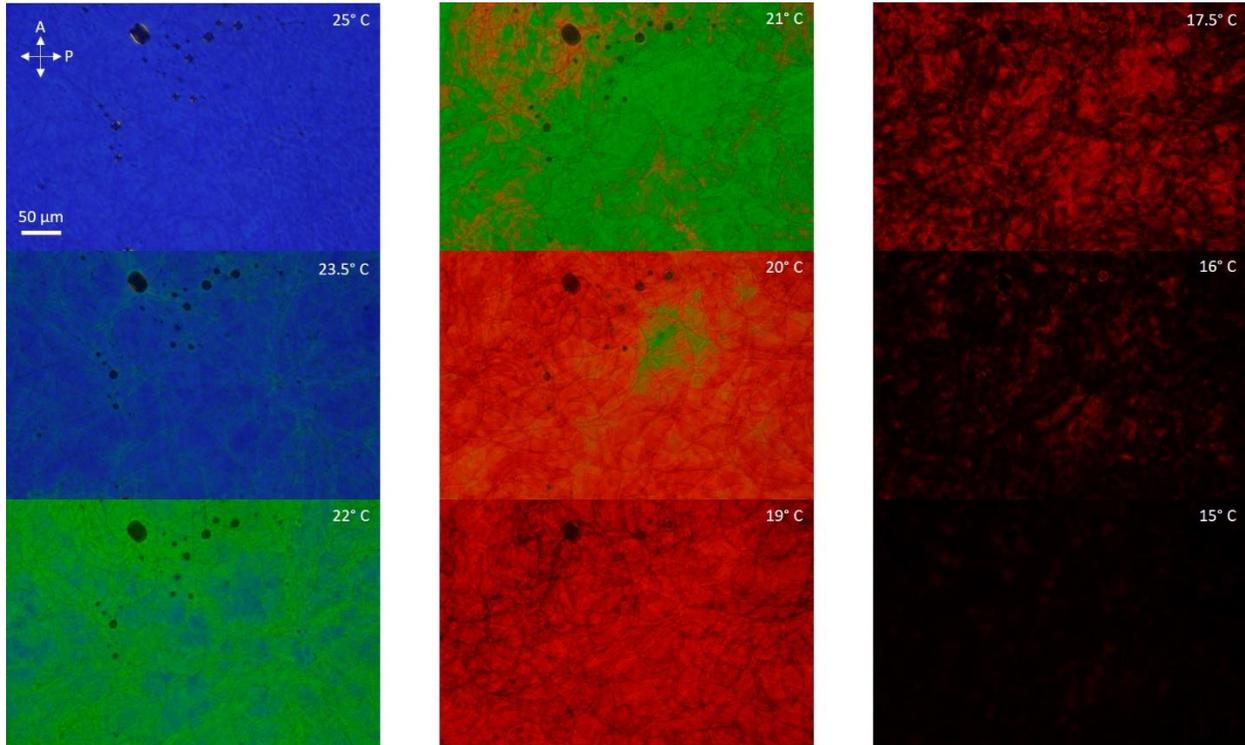


Figure 3.6: NEM\* in 400 mg/mL GTAC with temperature ramp. With the microscope in transmission mode, temperatures are as recorded for an observation that began at  $25^\circ\text{C}$  and ramped to  $15^\circ\text{C}$ . Scale bar is  $50\ \mu\text{m}$  and polarizers are as depicted.

When subjected to temperature ramping, GTAC at 400 mg/mL in a Grandjean texture showed pitch variation to its cholesteric helix (Figure 3.6). Transmitted light shifted from blue color at  $25^\circ\text{C}$  through green, orange, red and finally into infrared (presumably) at  $15^\circ\text{C}$ , suggesting helix expansion. Temperature dependent unwinding of the chiral pitch would have clear application as a thermometer. This specific behavior has not been significantly examined for concentration dependence across NEM\* concentrations but seems plausible. Shifts to cholesteric helix pitch seen here proved somewhat unidirectional since the ramp

returning from 15° to 25° C caused the Grandjean texture to shift into an odd structure that appears to contain parabolic focal conic defects before returning to a plain old Grandjean (see Figure 3.7). There has been some variability to the appearance of GTAC NEM\* textures that seems to depend on variations in the purification protocol, so significant work is still needed to clarify complexities present here.

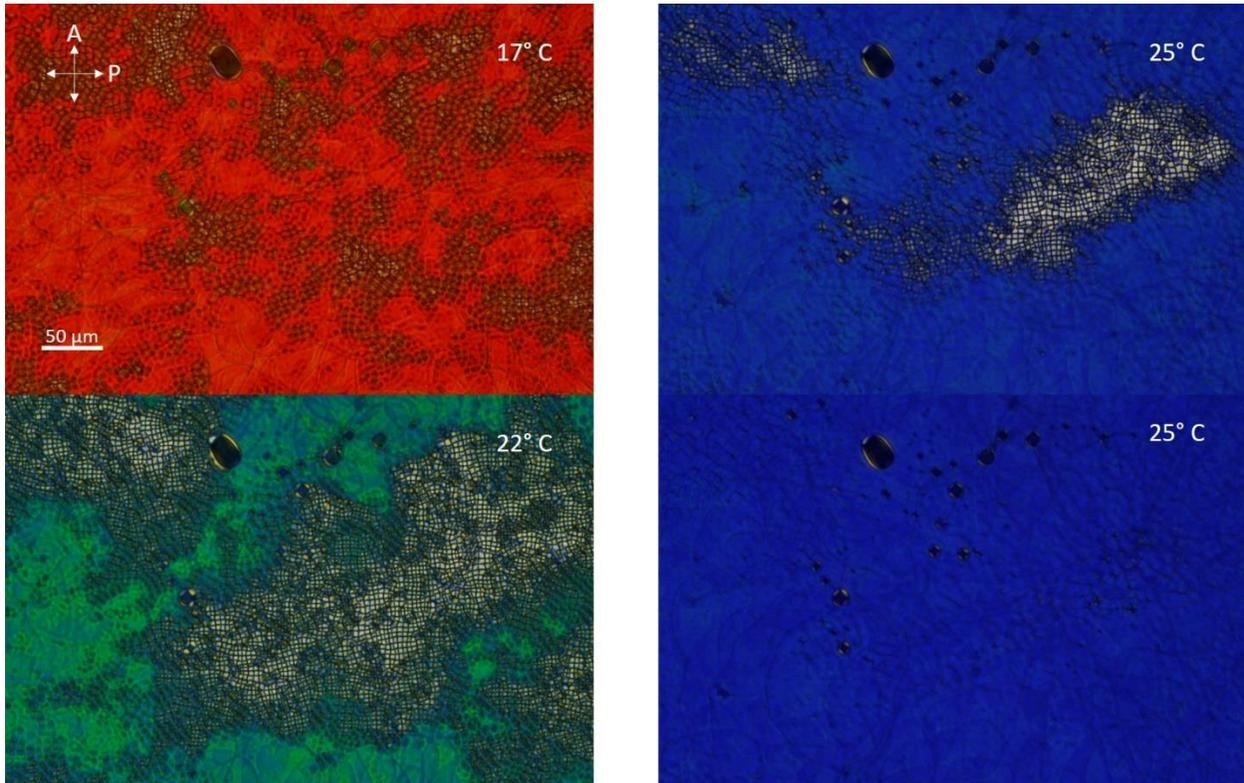


Figure 3.7: GTAC parabolic focal conic texture. Again, with microscope in transmission mode, NEM\* parabolic focal conic textures are seen on a temperature ramp from 15° to 25° C. The background Grandjean appears to wind from a red pitch to a blue pitch simultaneously.

### 3.1.2 <sup>5</sup>GTAC<sup>3</sup> X-ray Diffraction

GTAC was subjected to extensive X-ray diffraction in experimental schemes not different from those outlined with rDD (Section 2.2). GTAC diffraction data

turned out to be substantially more complicated than what was seen with rDD and elucidation of GTAC X-ray results turned out to depend somewhat on interpretation of rDD X-ray results given their similarities. What is presented here assumes a familiarity in the reader of the results in Section 2.0.

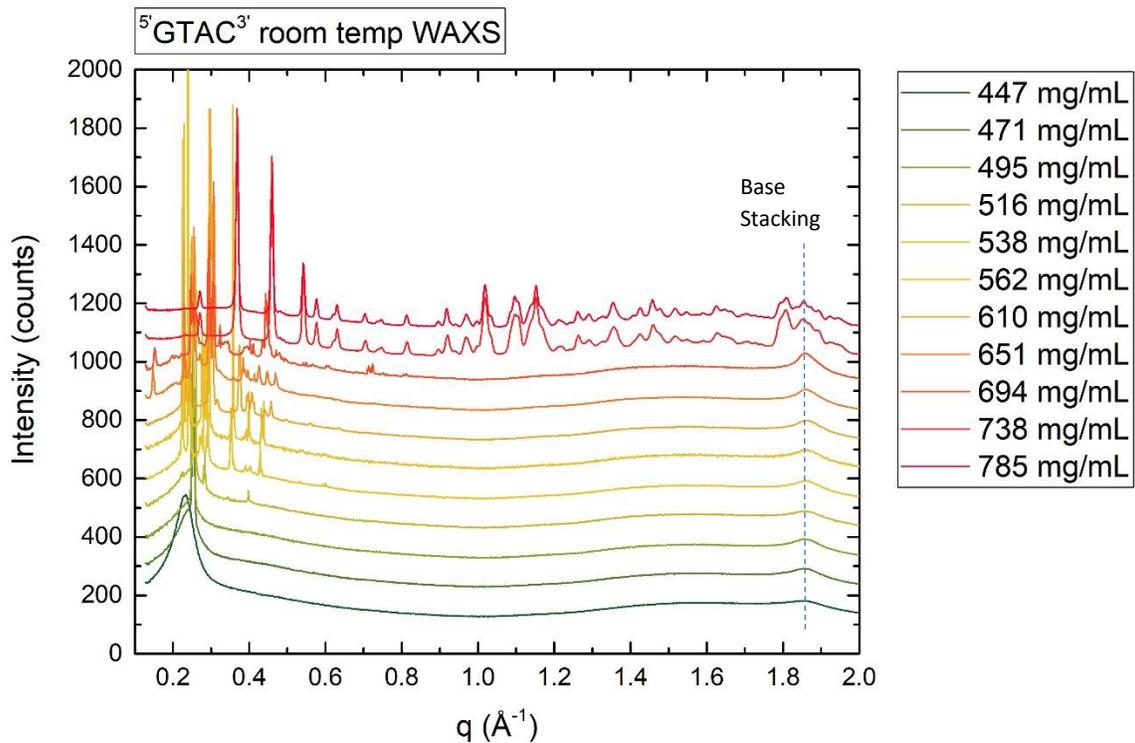


Figure 3.8: GTAC WAXS, circular average, full range. Capillary samples with corrected concentrations using the methods of Section B.2.2 are displayed together. No temperature cycling was executed prior to collecting this X-ray scattering at room temperature. The DNA base stacking peak is clearly visible along the dotted line.

Initial WAXS of GTAC samples showed a great deal of structure (Figure 3.8). The scattering features prior to temperature cycle appear to be in clusters in the small angle rather than at specific  $q$  values so that trends are muddled except in samples of low concentration. The small angle region with the mesophase packing features is expanded in Figure 3.9.

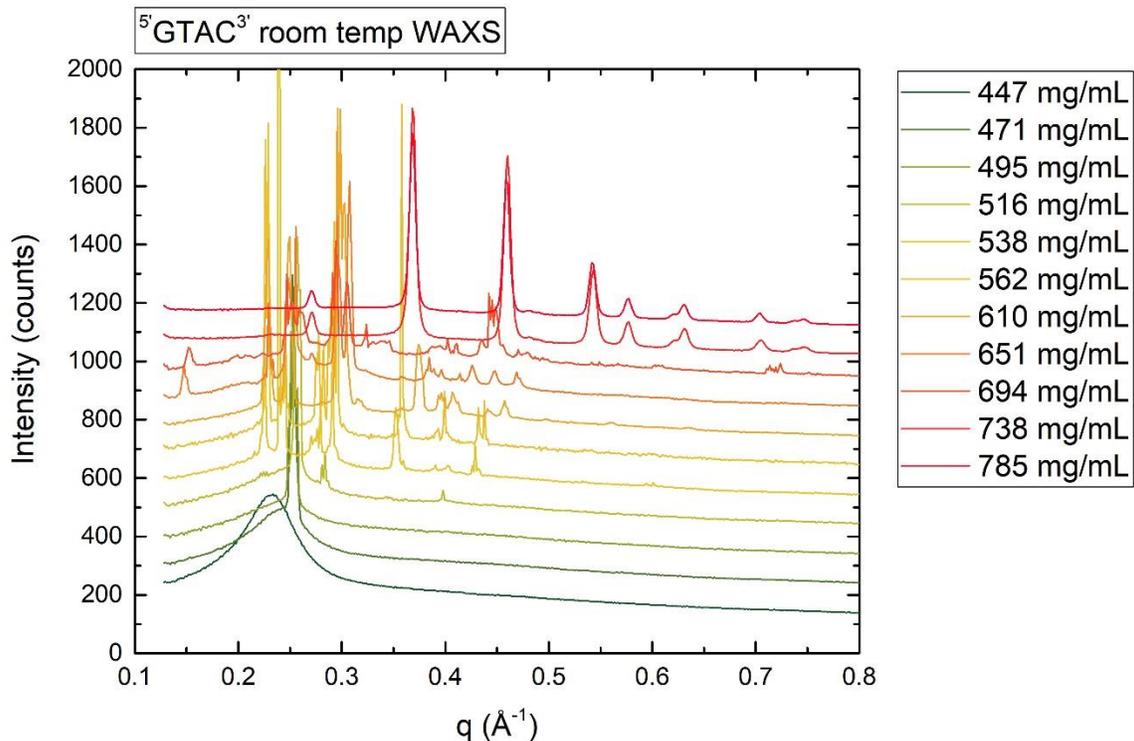


Figure 3.9: GTAC WAXS, circular average, small angle region. Capillary samples with corrected concentrations using the methods of Section B.2.2 are displayed together. No temperature cycling was executed prior to collecting this X-ray scattering at room temperature. This image focuses only on the small angle region. Some of the features seen here, particularly around  $\sim 540$  mg/mL, are very sharp features in contrast to much of the WAXS shown in Section 2.0, hinting at extremely long coherence lengths where very sharp features are not conserved from one sample to the next and appear in clusters in the same sample. Rings in 2D diffraction had granular scattering from clear monodomains (for example, see Figure 3.12), as opposed to the smooth powder averaging seen with rDD. GTAC appears to coarsen to very sharp but inhomogeneous order.

With a temperature cycle, the small angle features became more clarified, if lower in structural coherence length (Figure 3.10).

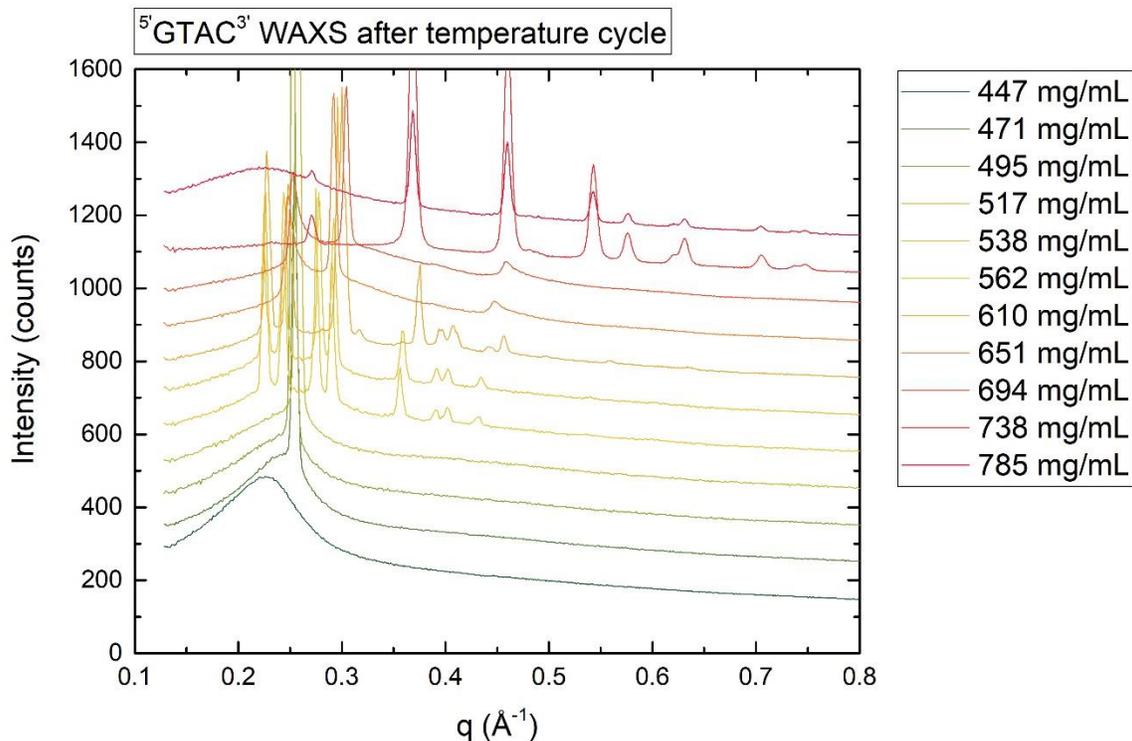


Figure 3.10: GTAC WAXS, circular average, small angle region at room temperature. Capillary samples with corrected concentrations using the methods of Section B.2.2 are displayed together. WAXS is shown of samples following a temperature cycle carrying them to ISO (except for the high concentration crystals, which did not melt). This image focuses only on the small angle region.

Figure 3.10 divides the concentration space into no fewer than five different mesophases proceeding from low concentration to high: NEM\*, COL, COL1/2, COL2 and Crystal. NEM\*, COL and COL2 all appear similar to the mesophases seen in rDD (Section 2.2), but COL1/2 is a complicated diffraction pattern not previously seen. This new phase is characterized by a cluster of four peaks in the packing region and apparently four more peaks at what seem to be a second harmonic. Figure 3.10 is following a temperature cycle that pushed most of the samples to melt to ISO phase and the WAXS was taken immediately on cool down. It is noteworthy that these phases are perhaps somewhat strained when the data above was taken and positional deviations of peaks in reciprocal space may in some cases

not reflect room temperature equilibrium positions. Data shown for COL in the GTAC comparison to theory (Figure 3.19) were taken from the samples prior to the temperature ramp (Figure 3.9) as COL samples following the temperature cycle (see Figure 3.11 for detail) did not give the expected positional trend with lower concentrations having the widest spacing. Prior to the temperature ramp, they do show the expected trend. GTAC samples can apparently have relaxation times longer than the observation times afforded at the synchrotron. Importantly, except for the crystal phases seen at very high concentration, GTAC mesophases seen here had a uniform melting temperature going to ISO phase: nearly everything melted at 40° C (see phase diagram Figure 3.18).

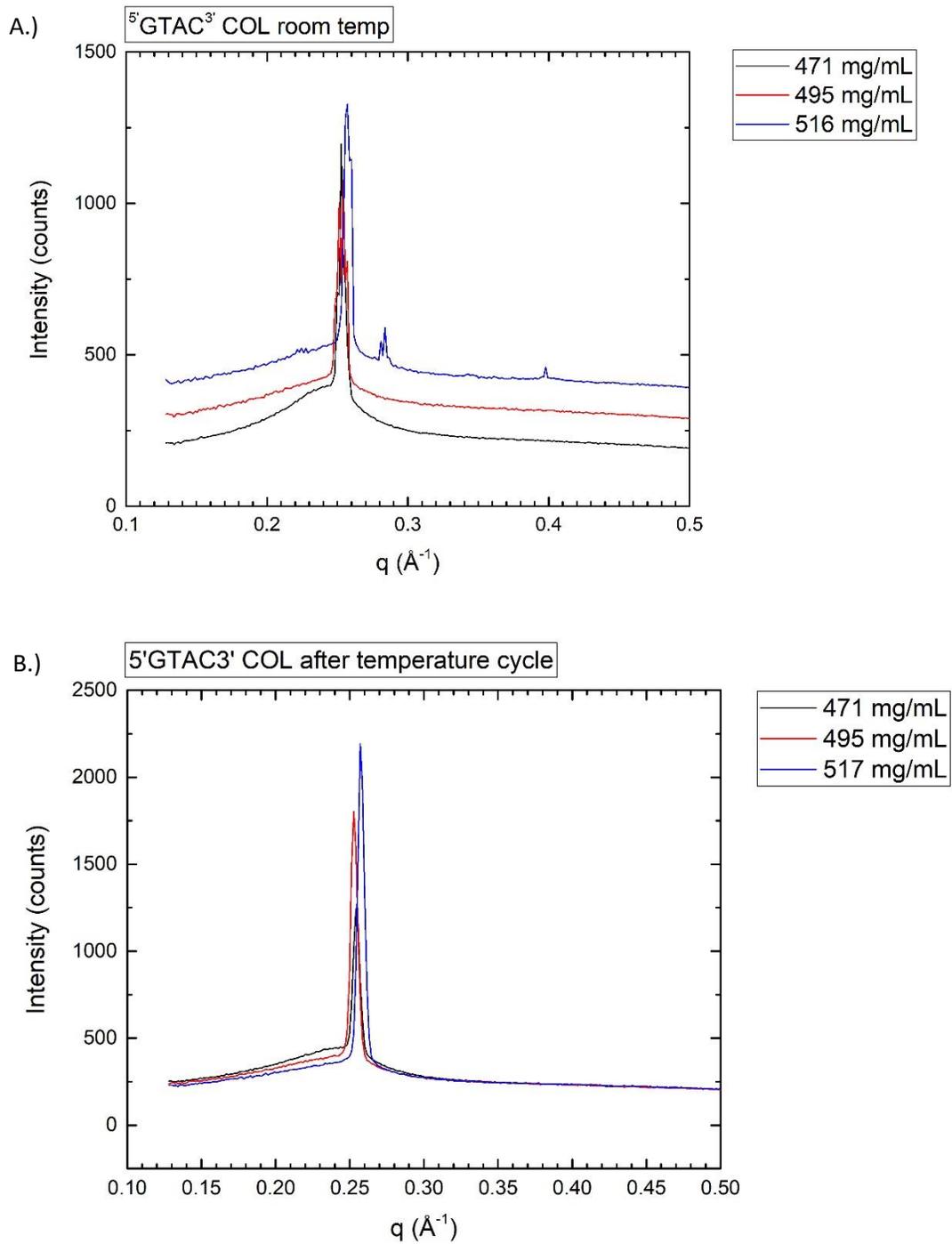


Figure 3.11: GTAC circular average, small angle region. Samples shown here are COL exclusively, 471 mg/mL (Black), 495 mg/mL (Red) and 517 mg/mL (Blue). A.) At room temperature prior to temperature cycle. B.) At room temperature after temperature cycle. Sample 517 mg/mL shows the COL+ feature (see Section 2.2.2) and a hexagonal  $q_{31}$  reflection prior to the temperature cycle.

Figure 3.11 shows the WAXS characteristic of GTAC COL both before and after temperature cycling. This mesophase appears to be a 2D hexagonal phase, where the sample at 517 mg/mL gives the characteristic  $q_{31}$  reflection for hexagonal order. This phase is clearly very similar to the rDD mesophase of the same order since 517 mg/mL also showed the COL+ reflection seen in rDD. The exact nature of the COL+ reflection remains unclear in GTAC. Figure 3.11 includes images from both before and after the temperature cycle to show how the data goes from notchy in the highly coarsened sample to smooth and more diffuse after thermal annealing. As mentioned, the notchiness is a result of the samples aging into discrete monodomains with each domain having very high structural coherence, but where the domains are apparently not identical –this was known from the granularity of the 2D scattering (see Figure 3.12). The reasons for this inhomogeneity are not entirely clear but would imply microscopic concentration gradients throughout the sample. The extent to which this same sort of observation might hold in rDD (Section 2.2) is not clear since similar coarsening was not seen with that material.

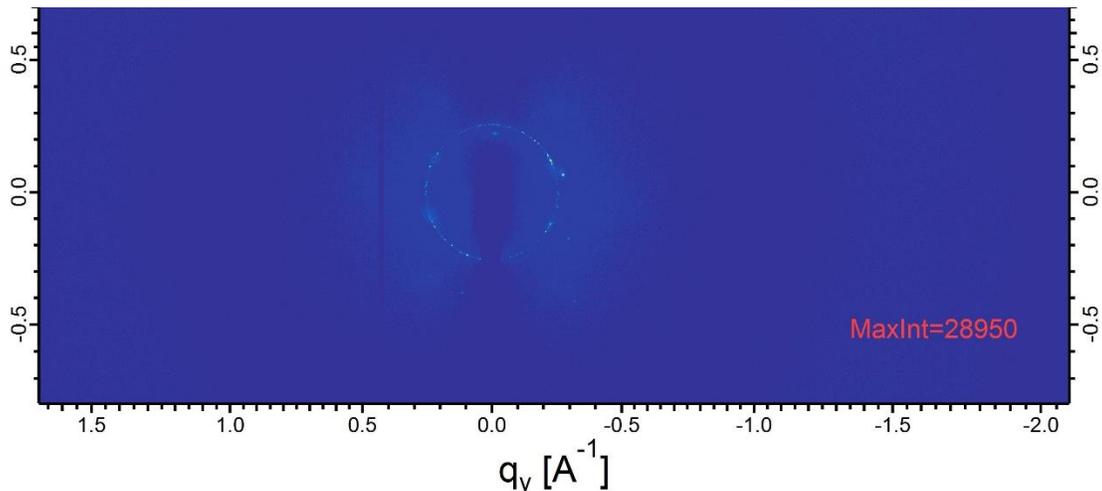


Figure 3.12: Room temperature 2D WAXS of 517 mg/mL GTAC. Example of the granular monodomain X-ray scattering seen frequently from GTAC as exemplified by a COL sample. The discontinuous ring is the COL packing peak. The color scale here is a terrain scale with blues giving low counts, greens and browns giving higher counts and whites saturating to max intensity. Except for NEM\*, all GTAC samples showed this sort of coarsening.

The second highly ordered mesophase seen with GTAC is the new phase, called here COL1/2 because of its apparent occurrence between COL and COL2. rDD has never shown an equivalent diffraction pattern, suggesting that this mesophase might well be exclusive to GTAC. As mentioned previously, this phase has a very complicated characteristic and diffraction pattern (see Figure 3.13) when seen as a 1D reduction.

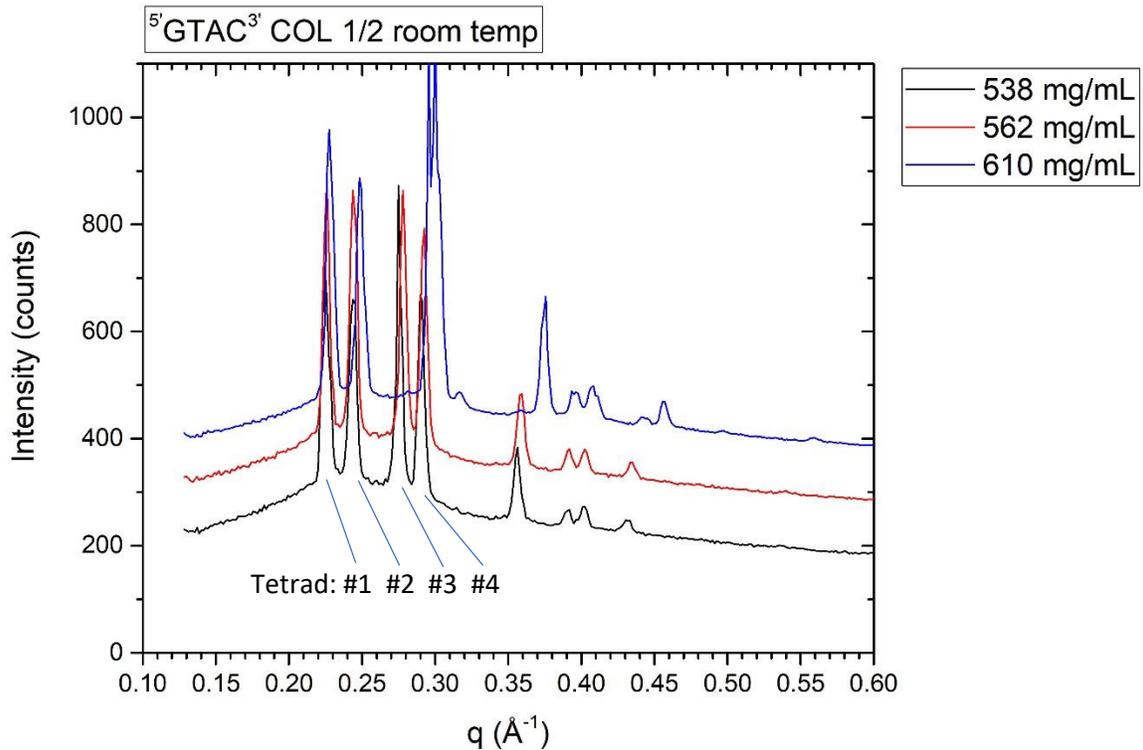


Figure 3.13: GTAC COL1/2 phase seen at room temperature following an annealing thermocycle. Concentrations demonstrating this phase were 538 mg/mL (Black), 562 mg/mL (Red) and 610 mg/mL (Blue). 610 mg/mL is an aberration which appears to be in slight coexistence with COL2 given the additional peak splittings that have appeared. Peaks of the central tetrad are labeled #1, #2, #3 and #4 from smallest  $q$  to largest.

The collected peaks in Figure 3.13 are the best controlled examples obtained of COL1/2 scattering; which was seen but without clear understanding of the relatedness between samples in three different synchrotron trips given the complexity of the scattering and its potential for variability among different

concentrations. All diffraction spots in COL1/2 appear to shift fractionally toward higher  $q$  with increasing concentration, suggesting strongly that the reflections are related to compression in packing of the mesophase and are expected to be a 2D lattice form perpendicular to the aggregate stacks. It was initially considered possible that scattering seen in this phase is somehow a 3D lattice much like rDD COL2, but none of the reflections fits a reasonable period that would be found along the long axis of a GTAC aggregate stack. The central four peaks, the tetrad, cannot be understood as a 3D scattering feature given their remarkable self-symmetry and the fact that they are four peaks instead of three, as would be needed to define three basis vectors for a 3D primitive lattice. As well, the similar symmetry present in the four second harmonic reflections is confusing.

One possible set of hints about this lattice came from the GTAC 610 mg/mL sample seen in Figure 3.13. This sample underwent a clear set of phase transitions while under WAXS observation as it was annealed from ISO into LC phases by rapidly dropping the temperature (Figure 3.14).

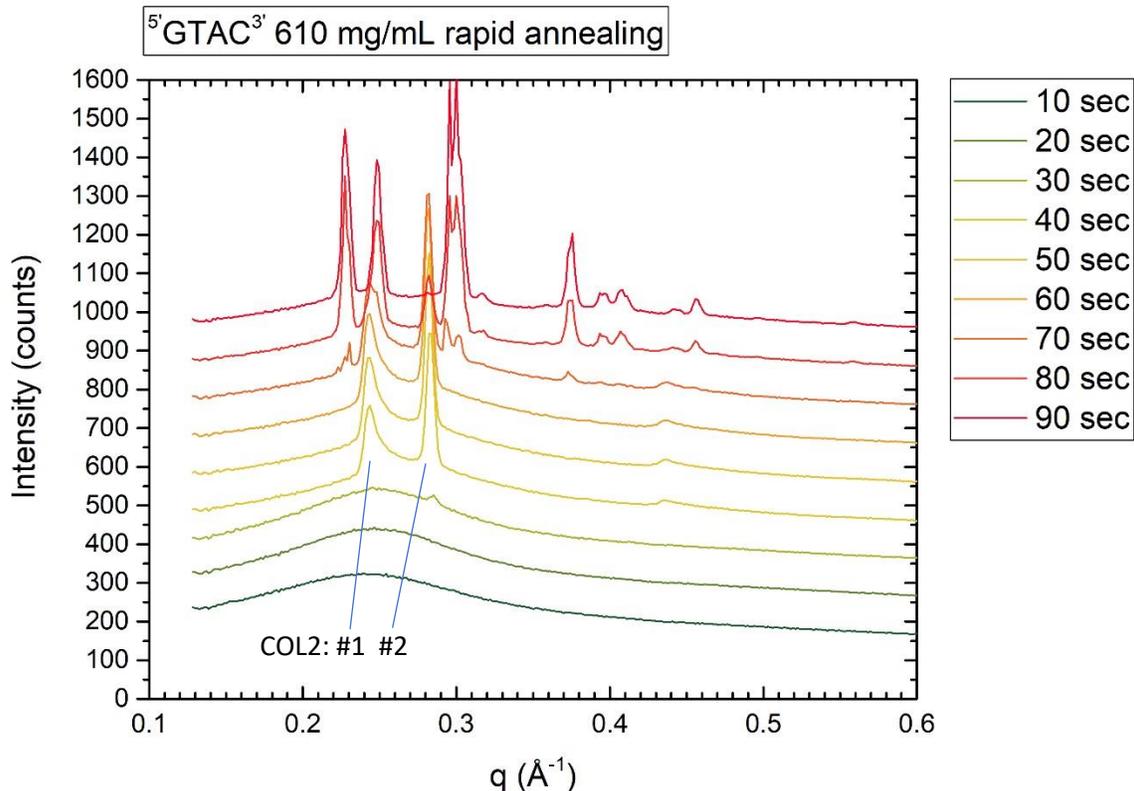


Figure 3.14: GTAC COL1/2 rapid annealing. Sample melted to ISO phase was subjected to rapid cooling to 24° C while under examination with WAXS. As the temperature reached room temperature, WAXS shots were taken at the phase about every ten to fifteen seconds for no less than 90 seconds (approximate times shown). The peaks of COL2 are labeled for convenience, #1 and #2.

In Figure 3.14, GTAC undergoes a phase transition first from ISO to COL2, and then to COL1/2 after a short lag. COL2 peak #1 (Section 2.2.2) shifts subtly to become the second peak in the COL1/2 central tetrad while COL2 peak #2 vanishes and is replaced with tetrad peaks #3 and #4, which appear in this sample to be squeezed together. Tetrad peak #1, the peak at smallest  $q$ , emerges from the baseline approximately as peaks #3 and #4 supersede COL2 peak #2. The horn shape of COL2 peak #1 vanishes just before peak #1 shifts to a slightly larger  $q$ . COL1/2 and COL2 are clearly not the same structure, but they seem to be connected by the motion of COL2 peak #1.

To best estimation, COL1/2 appears to be another polymorphism of the FHL (Flattened Hexagonal Lattice) offered in Section C.4.2 and used extensively to analyze COL2 in Section 2.2.3. This structure is, however, unlike the one seen in rDD. First, the lattice appears to be a Case 2 FHL rather than Case 1, where  $q_{11} < q_{02}$ , and second, it is a twinned lattice with two examples of the same type of lattice but with a slightly different dimension lying on top of one another. This can be regarded as plausible since the COL2 #2 reflection vanished when COL1/2 tetrad reflections #3 and #4 appeared, meaning that these peaks are not smooth transformations into one another by lattice deformation, but wholesale reorganization of the lattice, making it possible to move from Case 1 to Case 2. Further, as Case 2 FHLs, the areas of these lattices decrease with increasing concentration, the expected trend, unlike the Case 2 assignment of COL2.

Tetrad peaks can be grouped into pairs in order to match them as Case 2 FHLs to the available higher order peaks. Tetrad peaks #1 and #3 group together into a couplet, while peaks #2 and #4 also group together. The best available indexing of this structure can be seen in Table 3.1 using  $q$  values drawn from GTAC 538 mg/mL.

Table 3.1: COL1/2 peak indexing

Large Case 2 FHL	Peak ( $\text{\AA}^{-1}$ )	Indexing (h,k)
$\vec{A}=0.17791 \text{\AA}^{-1} \vec{B}=0.13764 \text{\AA}^{-1}$	0.22494	(1,1)
	0.27528	(0,2)
	0.3558	(2,0)
	0.41292	(3,0) (forbidden)
Small Case 2 FHL		
$\vec{A}=0.195946 \text{\AA}^{-1} \vec{B}=0.145215 \text{\AA}^{-1}$	0.24389	(1,1)
	0.29043	(0,2)

	0.3918	(2,0)
	0.4356	(3,0) (forbidden)

This lattice assignment is not without some significant misgivings. The peak indexed as (3,0) is an expected omission in an FHL; it should not appear. With only these peaks available, and no aligned mesophase data, this set of assignments has unfortunately been the only set to successfully predict higher order reflections from an assignment of the fundamental basis vectors.

Table 3.2: COL1/2 Case 2 FHL lattice characteristics

Concentration	Large lattice	Small lattice	Large lattice area	Small lattice area
538 mg/mL	$\vec{a}=35.31 \text{ \AA}$ $\vec{b}=45.65 \text{ \AA}$	$\vec{a}=32.07 \text{ \AA}$ $\vec{b}=43.27 \text{ \AA}$	1611.9 $\text{\AA}^2$	1387.7 $\text{\AA}^2$
562 mg/mL	$\vec{a}=35.35 \text{ \AA}$ $\vec{b}=45.19 \text{ \AA}$	$\vec{a}=32.09 \text{ \AA}$ $\vec{b}=42.99 \text{ \AA}$	1597.3 $\text{\AA}^2$	1379.5 $\text{\AA}^2$

The coexistence of two lattices would not be completely impossible in a GTAC sample given the routine appearance of coexisting textures at high concentration: in Figure 3.1 and 3.3, it was noted that the poorly birefringent dark mosaic phase routinely occurred along with brightly birefringent COL2-like frond textures. The scattering could be from two separate crystal systems with similar lattices that are present at the same time. The only ambiguity from this is the lack of area conservation. If COL1/2 is again a 2D mesophase where distance between columns and the arrangement of those columns on a 2D lattice is directly set by the GTAC concentration, how is it possible for a common COL2 diffraction pattern to break down (as it did in Figure 3.14) into a COL1/2 diffraction pattern? One would expect concentration homogeneity in ISO phase, but the large and small lattices emerge simultaneously from homogeneous COL2. The COL1/2 large lattice at all

concentrations is larger in 2D area than the hexagonal lattices of GTAC COL phases at lower concentrations, which seems impossible.

Another possibility here is that the COL1/2 large and small lattices are somehow interleaved so that they occur simultaneously, allowing one lattice to be an artifact of the other. The exact means by which this could be accomplished is ambiguous, but it must be noted that the assignment of COL1/2 as overlapping Case 2 FHLs is dependent on a scattering pattern that does not contain all the expected omissions: the (0,3) reflection in both large and small lattices is a forbidden reflection that appeared anyway. The model for COL1/2 as presented is intrinsically broken, meaning that it must be seriously considered that all the proposed assignments are wrong and some, as-yet-envisioned lattice system is a better answer. The one bright light in all of this is that the area of the small COL1/2 lattice is fairly closely set to the trend established by COL samples (see Figure 3.19), enabling the explanation that the large lattice is somehow an artifact to at least be possible.

The (0,3) reflection is a forbidden reflection in a face centered lattice. How this reflection can appear is confusing since it should not be achievable. One conclusion from this deficiency is that COL1/2 is not actually any sort of face centered lattice, calling into question all of the lattice indexing efforts offered above. Again, unfortunately, the closest second option to explain this diffraction pattern is a series of overlapping square and hexagonal lattices for which there is insufficient data support. Each of the four higher order reflections can be taken as related to one member in the central tetrad as either a second order square or second order hexagonal lattice reflection and one reflection overlaps as either square or hexagonal for two different members of the central tetrad (Table 3.3).

Table 3.3: Square and hexagonal lattice assignments for 538 mg/mL GTAC

Central Tetrad	Related higher order reflection	Predicted lattice
.22494	.3918	.389 hexagonal $q_{31}$
.24389	.3558	.3449 square $q_{11}$
.24389	.4356	.4225 hexagonal $q_{31}$
.27528	.3918	.389 square $q_{11}$
.29043	.41292	.4107 square $q_{11}$

Without data from an oriented sample, Table 3.3 illuminates much of the difficulty with the available data sets. A complicated piece of 1D data has the potential to significantly overlap between many different models. By fixating on correlations to one model in lack of sufficient data, another closer match might be overlooked. Non-degenerate 2D data is needed to resolve the overlaps.

The lattice indexing in Table 3.1 is confoundingly consistent. It seems possible that COL1/2 is a pair of interleaved rectangular lattices containing some unforeseeably complicated structure factor producing weird omissions which is able to somehow convert into the COL2 Case 1 FHL given some simple manipulation. The relatedness between COL peak #1 and tetrad peak #2 even suggests that the second couplet defining the basis for the small lattice is a real observation while the first couplet is artefactual. The vision offered in Table 3.3 might also be true: COL1/2 is multiple different overlapping structures that happen to have similar free energies. And, it seems frustratingly possible that COL1/2 is some as-yet-unknown structure, either 2D or 3D.

The second to last phase seen in GTAC is a more usual structure: a form of COL2 much like that seen with rDD (Section 2.2.3). The two GTAC samples showing COL2 are in Figure 3.15.

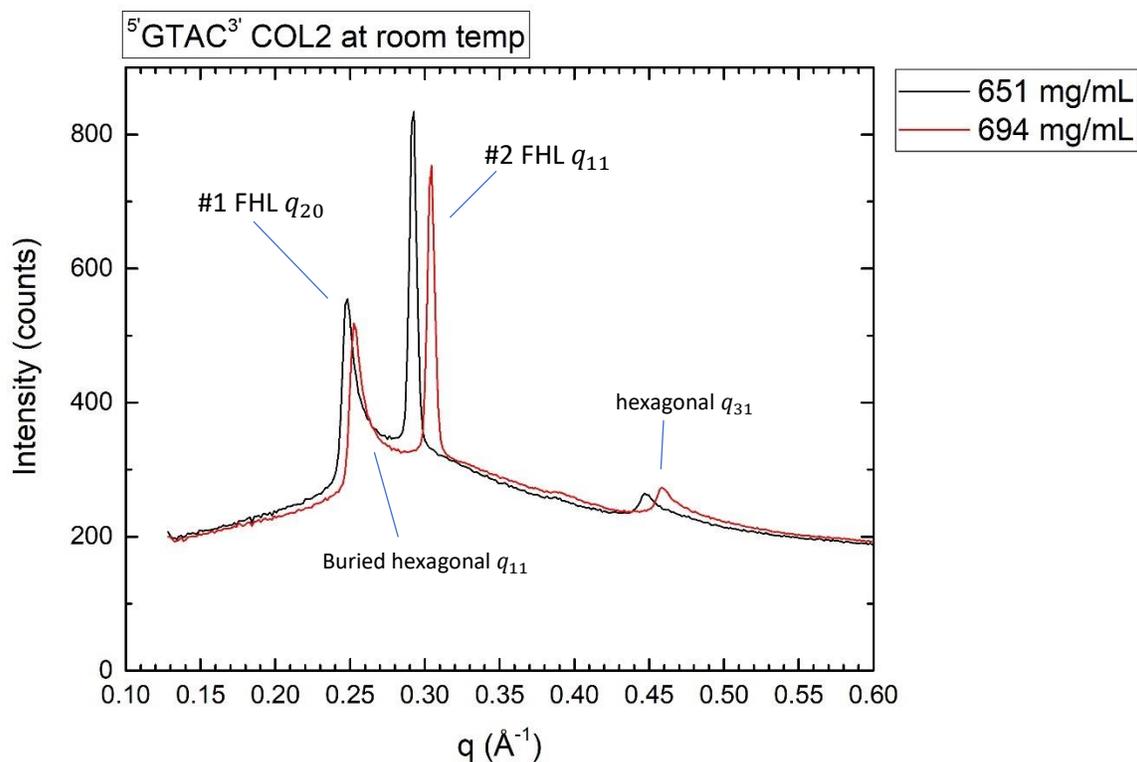


Figure 3.15: GTAC COL2 at room temperature. Both samples are shown, 651 mg/mL (Black) and 694 mg/mL (Red). COL2 peaks #1 and #2, assigned as Case 1 FHL  $q_{20}$  and  $q_{11}$  respectively. The buried hexagonal  $q_{11}$  peak is noted from the shape of the saddle, as is the prominent hexagonal  $q_{31}$  which accompanies it.

GTAC COL2 as a Case 1 FHL is very similar to that seen with rDD. In Figure 3.15, the sample is freshly annealed after a temperature ramp that melted it into ISO phase. The peak typically seen in rDD associated with the DNA duplex helical pitch is not present in Figure 3.15 as are none of the higher order peaks analyzed in Section 2.2.3. The helical pitch peak does make an appearance before the temperature cycle that melts the phase (Figure 3.16).

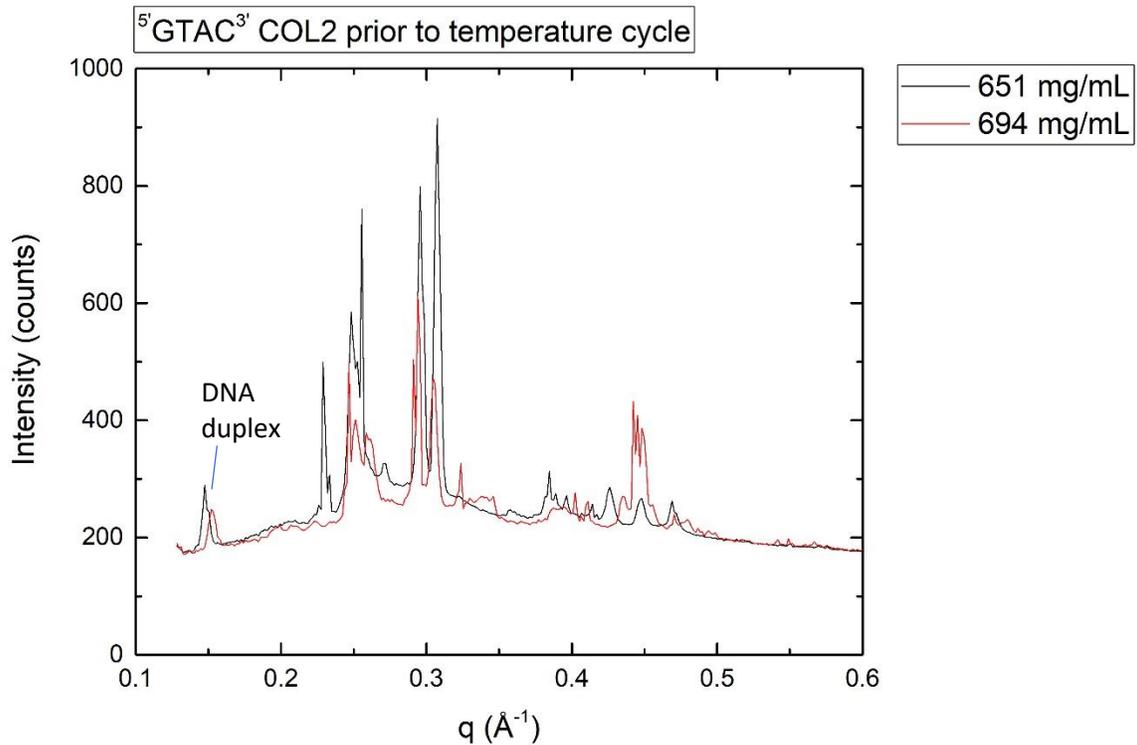


Figure 3.16: GTAC COL2 prior to temperature cycle. Concentrations 651 mg/mL (Black) and 694 mg/mL (Red) are as labeled. The peak related to the DNA duplex is noted at  $q = 0.15 \text{ \AA}^{-1}$ .

Prior to cycling the temperature to melt the phase to ISO, COL2 for GTAC is almost completely unrecognizable due to the polydispersity of the scattering pattern, except for the peak at  $q = 0.15 \text{ \AA}^{-1}$  which is suspiciously similar to the one seen with rDD at  $0.172 \text{ \AA}^{-1}$  but a smaller  $q$ . This peak was previously associated with the B-form DNA helical pitch in rDD. The argument in Section 2.2.3 was that the defected structure of rDD helped to relax the DNA helical pitch from  $0.178 \text{ \AA}^{-1}$  to  $0.172 \text{ \AA}^{-1}$ . A GTAC aggregate is roughly three times as discontinuous as an rDD aggregate, lending to a similar argument. Using the same set up and numbers as seen in Section 2.2.3, every fourth phosphodiester in GTAC is broken. So, GTAC has  $(1+1/4)*3.4 \text{ \AA}$  of rise for each  $6.93 \text{ \AA}$  segment of backbone, which adds  $5.24 \text{ \AA}$  of circular circumference for each nucleobase worth of rise in the double helical turn.

For 62.83 Å of circumference, 11.98 bases are now required to complete the turn. 11.98 bases is a stacking rise of 40.7 Å or 0.154 Å<sup>-1</sup> in reciprocal space. This relaxation is comparable to the peak seen in Figure 3.16. It is possible that the temperature cycle disordered the common register of the duplex pitch in Figure 3.15, causing it to wash out in that scattering pattern. Given more relaxation time during shooting at the synchrotron, the pitch register may well have reemerged.

The final mesophase examined in GTAC, however briefly, is unquestionably a 3D crystal (Figure 3.17).

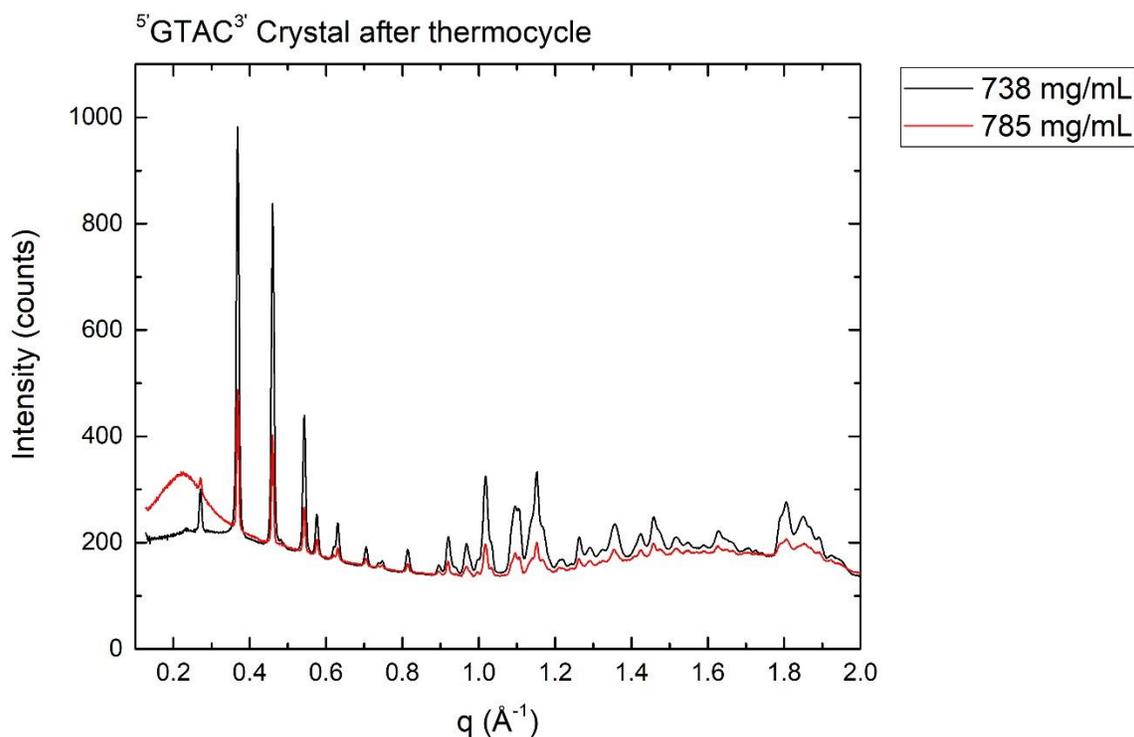


Figure 3.17: GTAC high concentration 3D crystal. Seen following a temperature cycle that failed to melt the phase to ISO are GTAC at 738 mg/mL (Black) and 785 mg/mL (Red).

No efforts have been made to index this crystal structure. The only special feature of note is the presence of a broad hump at  $q = 0.2 \text{ \AA}^{-1}$  in the highest

concentration sample that appeared after the temperature cycle: this may well be a sign of reemergent ISO phase present as some form of disordered glass.

### 3.1.3 Comparison of $5^{\prime}\text{GTAC}3^{\prime}$ Mesophases

GTAC has an extremely complicated phase diagram for the relative simplicity of the construct. An unexpectedly large number of different mesophases emerged from the WAXS experimentation and a great deal of behavior was observed microscopically. With melting temperatures acquired for these mesophases during X-ray diffraction a phase diagram for GTAC becomes possible (Figure 3.18).

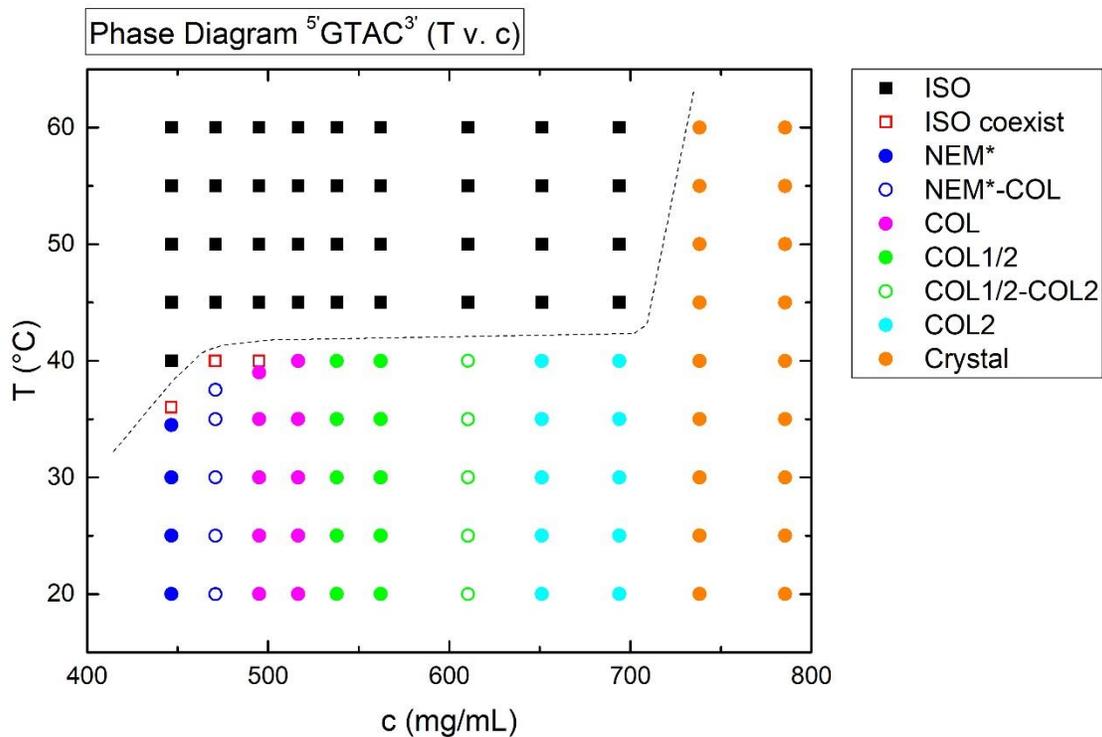


Figure 3.18: GTAC Phase Diagram. All phases are as denoted ISO (Black Squares), ISO Coexistence (Red Empty Squares), NEM\* (Blue Circles), NEM\*-COL coexistence (Blue Empty Circles), COL (Magenta Circles), COL1/2 (Green Circles), COL1/2-COL2 coexistence (Green Empty Circles), COL2 (Cyan Circles) and Crystal (Orange Circles). The dotted line denotes the apparent ISO interface.

One very interesting quality of the GTAC phase diagram is the sharp plateau that occurs for the melting critical temperature of mesophases at 40° C in COL and above. At this temperature, without any sign of coexistence, birefringence for the mesophase simply vanished. This is corroborated in X-ray by the loss of sharp diffraction peaks at that temperature. The author speculates that, under the salt and counter ion conditions of the GTAC samples used for these investigations, 40° C is the temperature at which a GTAC duplex unzips. An online oligomer  $T_m$  calculator available from IDTDNA suggests a melting point of ~27° C for concentrations of 50 mM GTAC in 1 M Na<sup>+</sup>; the calculator would not reach the 300 mM to 400 mM concentrations attained with ~500mg/mL GTAC samples. This is likely one example of a nanoDNA where the upper temperature limit of the mesophases is strongly dependent on the  $T_m$  of the DNA duplex as opposed to the critical temperature of the mesophase. GTAC duplexes appeared stabilized in the crystal phase, which did not melt at examined temperatures within the time periods of observation.

GTAC can also be compared by concentration to the theory developed in Section 2.3.1 aimed at matching concentration  $c$  to the square of the hexagonal lattice reflection  $q_{11}$  (called  $q_{hex}$ ) on the assumption that the aggregate lies on a 2D hexagonal lattice which is incompressible in its third dimension (see Figure 3.19).

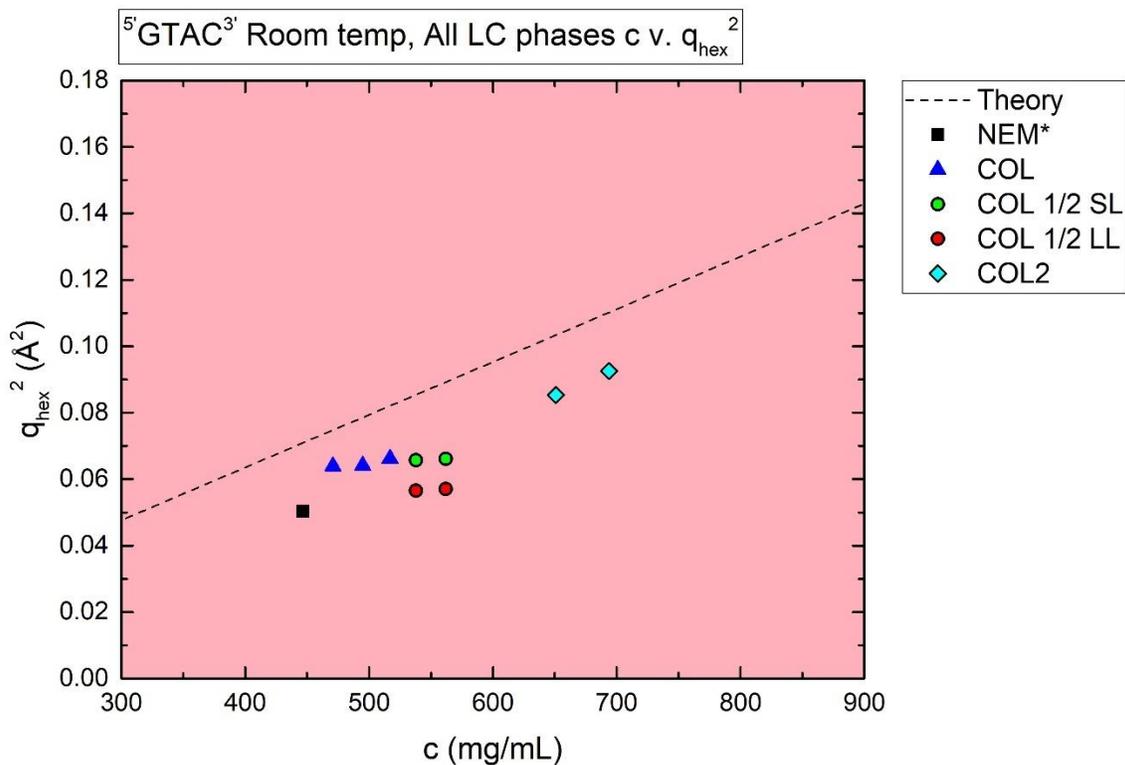


Figure 3.19: GTAC concentration versus  $q_{hex}$  compared to theory (Section 2.3.1). Mesophases added are NEM\* (Black Square), COL (Blue Triangle), COL1/2 Small Lattice (SL, Green filled circles), COL1/2 Large Lattice (LL, Red filled circles) and COL2 (Cyan Diamonds). NEM\* is directly the nematic side-side  $q$  value and is not understood to be on a hexagonal lattice. COL1/2 SL and LL are calculated from the areas offered in Table 3.2, which were then used to back calculate the appropriate  $q_{hex}$  from equation C.57 For GTAC, theory line (equation 2.10) with  $K = 0.000158778 \text{ mL}/(\text{mg} \cdot \text{\AA}^2)$ , which compares to rDD,  $K = 0.00015397 \text{ mL}/(\text{mg} \cdot \text{\AA}^2)$ . The crystal phase is not added given a lack of a useful comparison metric.

The trends seen in Figure 3.19 are generally similar to what appeared for rDD (Figure 2.55) in how the distribution lies below the theory line; as would be expected for unaccounted salt contamination. The COL and COL1/2 small lattice follow what appears to be a consistent trend with a smaller slope value than the theory slope while the COL2 is consistent with the theory slope. NEM\* is not particularly expected to sit on the same line given that it lacks a lattice and the COL1/2 large lattice, of course, suggests an unrealistically large area by lying lower

than everything. If the structure of COL1/2 is truly something different, it would be expected to have a similar surface area to the small lattice, or possibly pack a slightly larger amount of mass into the area in question. The assignments of COL1/2 as the Case 2 small lattice FHL may be flawed for the reasons already mentioned, but compares well here to the COL values.

The reason that the slope appears to be too low for the COL set in Figure 3.19 may be associated with the preparation methods used for the capillaries in this experiment. These samples were left for exceptionally long times at elevated temperature (70° to 80° C) in an effort to make the concentrations of each homogenous. It has been observed since that time that this form of treatment can damage DNA over time, causing the sample to take on a brownish color. If these samples have been heat-damaged, they may follow a slightly dysfunctional trend that artificially increases the lattice spacing at a given concentration by depleting the amount of LC forming oligomer. The samples of rDD used in Section 2.2 were not treated quite as harshly. More recent versions of the capillary preparation methods result in highly uniform concentrations without the same level of abuse.

### 3.2 Sticky-End 4mer 5'GCCG<sup>3'</sup>

The second 4mer construct to be examined in this section is the so-called “sticky-end” 4mer 5'GCCG<sup>3'</sup> (GCCG). This particular sticky-end 4mer is one of three variations that have been explored by close collaborators, including also 5'GCTA<sup>3'</sup> and 5'ATTA<sup>3'</sup>, and has been published<sup>34</sup>. We began examining GCCG in particular after a collaborator complained that this construct had some inexplicable behavior. The outcomes of this examination will be detailed in this section.

LC forming sticky-end type oligomers differ from blunt-end forms by the manner in which the nanoDNA aggregate is assembled. Blunt-end oligomers assemble into linear aggregates by use of hydrophobic stacking forces between the exposed ends of neighboring duplexes. Sticky-end nanoDNA oligomers instead use sequence directed hydrogen bond templating from exposed complementary single strands to mediate assembly (Figure 3.20).

A.) Blunt-end assembly

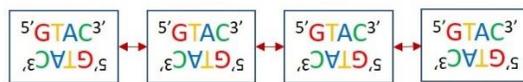
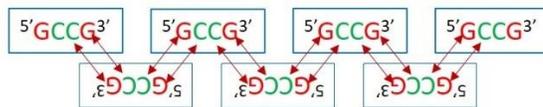


Figure 3.20: Comparison of Blunt and Sticky assembly routes. A.) GTAC hydrophobic blunt end assembly. B.) GCCG base pair mediated sticky end assembly.

B.) Sticky-end assembly



In the 4mer versions, sufficient oligomer length exists to construct a structural core in the blunt-end version only. Blunt-end oligomers can form short duplexes that subsist in solution base paired without necessarily associating by hydrophobic interactions into stacks with one another. A sticky-end 4mer contains only complementary overlap sequences that must pair with a neighbor in order to form a structural core and if they are not associated with a neighbor, they are only floppy single strands. Qualitatively, sticky-end 4mers would be expected to be more fragile, about equivalent to stacked dimers that are singly crosslinked to neighboring dimers at their ends.

GCCG is particular in having a very high guanine content. Guanine is specially renowned for having more complicated behavior than simple G-C Watson-Crick pairing, adding in not only G-C Hoogsteen<sup>36</sup> pairing (which is also possible with A and T) but including the potential for G-quartets, a form of self-association

that can compete with mutual associations under certain conditions<sup>37–39</sup>. G-quartets are known to form chromonic LC mesophases in absence of other associations (Figure 3.21).

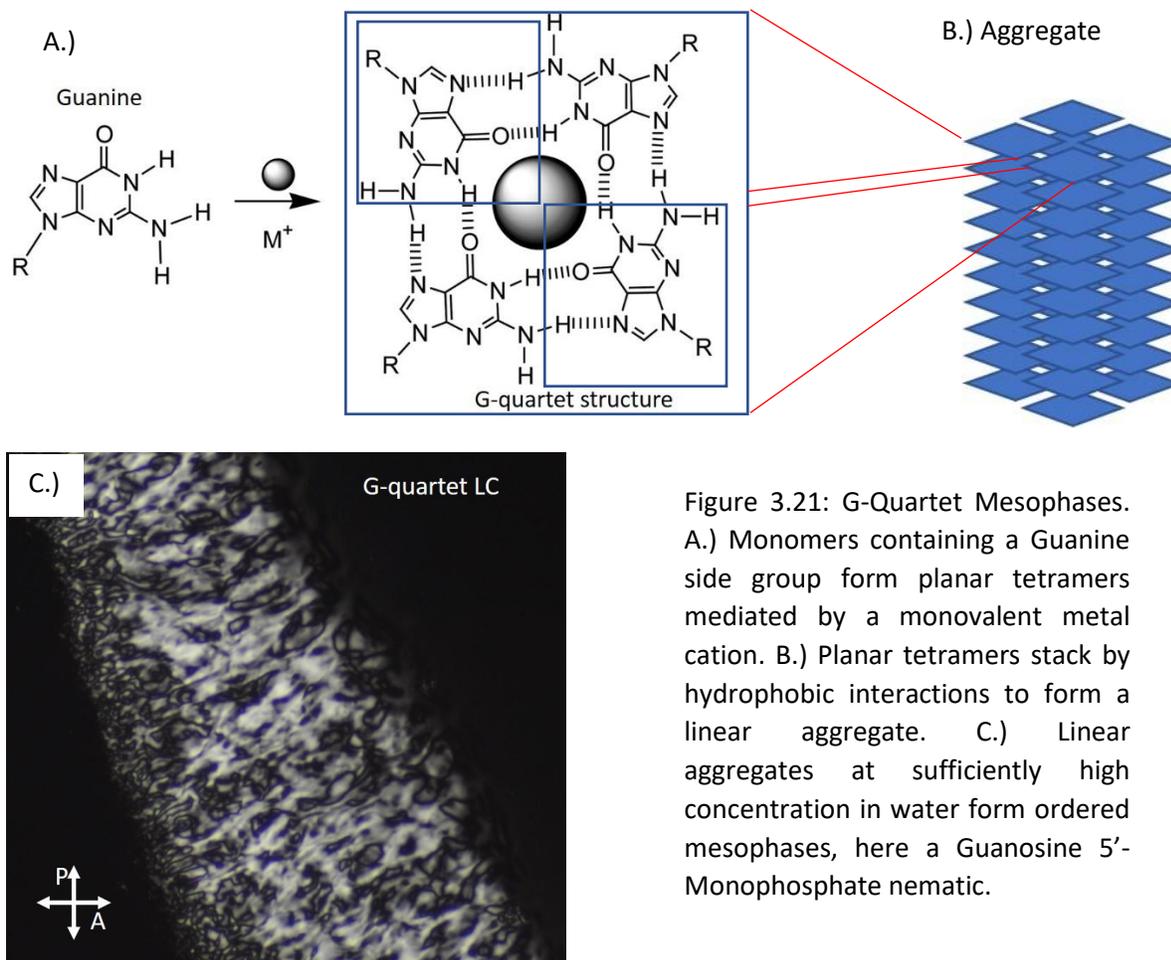


Figure 3.21: G-Quartet Mesophases. A.) Monomers containing a Guanine side group form planar tetramers mediated by a monovalent metal cation. B.) Planar tetramers stack by hydrophobic interactions to form a linear aggregate. C.) Linear aggregates at sufficiently high concentration in water form ordered mesophases, here a Guanosine 5'-Monophosphate nematic.

### 3.2.1 <sup>5</sup>GCCG<sup>3</sup> Optical Textures and Phase Diagram

GCCG phase diagram reported to the literature came from a version engineered as <sup>5</sup>GCCG<sup>3</sup>-P (GCCG-P) giving it a phosphate group at its 3'-terminal. This material possesses a phase diagram that is similar in many ways to most of the nanoDNA examined previously (see Figure 3.22) with some unusual extra features that provided the original motivation to expand our examination.

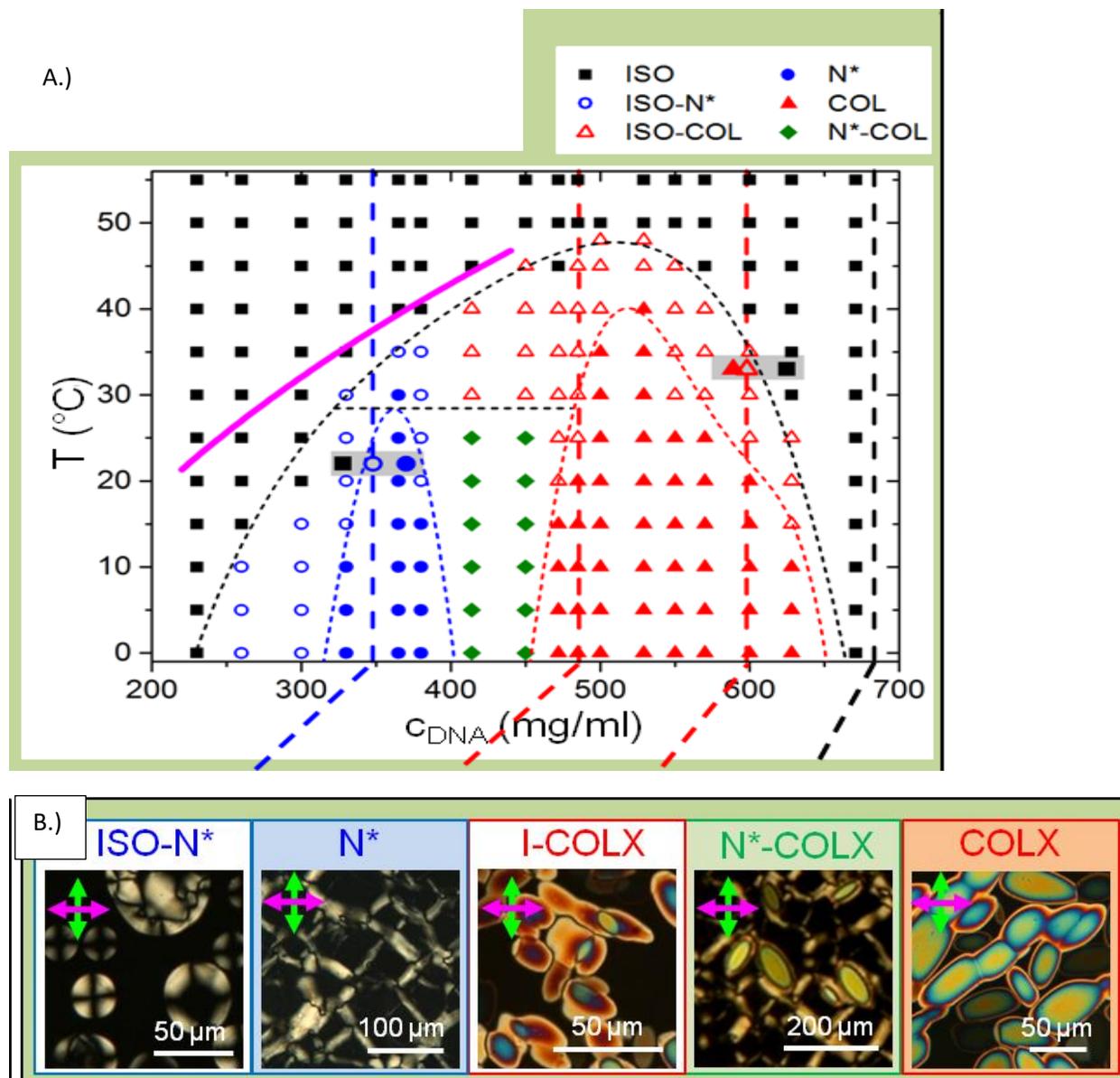


Figure 3.22: Reproduction of GCCG-P results from Fraccia *et al.*<sup>34</sup> A.) GCCG-P phase diagram includes ISO (Black Squares), ISO-NEM\* coexistence (Empty Blue Circles), NEM\* (Blue Circles), ISO-COL coexistence (Empty Red Triangles), NEM\*-COL coexistence (Green Diamonds) and COL (Red Triangles). In this paper NEM\* is labeled as N\* B.) Observed mesophase textures for GCCG-P, including previously seen textures and COLX. COLX is higher order columnar with an unknown structure, lacking frond-like shapes, but possessing a uniform optic axis.

As with GTAC, GCCG in Figure 3.22 does not appear to have phases at temperatures above  $\sim 40^\circ\text{C}$ . The particularly baffling behaviors observed in this LC system include the “reemergent ISO phase.” This can be seen in the phase diagram

from the reversing trend of the ISO melting  $T_c$  initially increasing with concentration, and then reversing at about 500 mg/mL where ISO grows back in until it comes to dominate again at all temperatures. The ISO phase therefore “reemerges.” In addition, COLX seen here has a morphology dissimilar from previously observed forms of COL, only vaguely reminiscent of COL2 (compare with Figure 2.3).

We entered this system with the intention of reproducing and helping elucidate these two baffling behaviors. A single pivotal initial difference may have lead to a variety of additional bizarre behaviors: the observations based on Figure 3.22 were from GCCG-P, possessing a terminal phosphate group to facilitate von Kiedrowski carbodiimide chemistry, while we ended up synthesizing GCCG in-house, with no terminal phosphate.

GCCG initially totally failed to produce the expected phase diagram. Samples were observed immediately to have atypically high viscosity, resisting pipetting even when at ostensibly nematic concentrations. Moreover, samples failed to produce the expected phase sequence, repeatedly: NEM\* textures were appearing at higher concentrations than COL or COL at lower concentrations than NEM\* (see Figure 3.23).

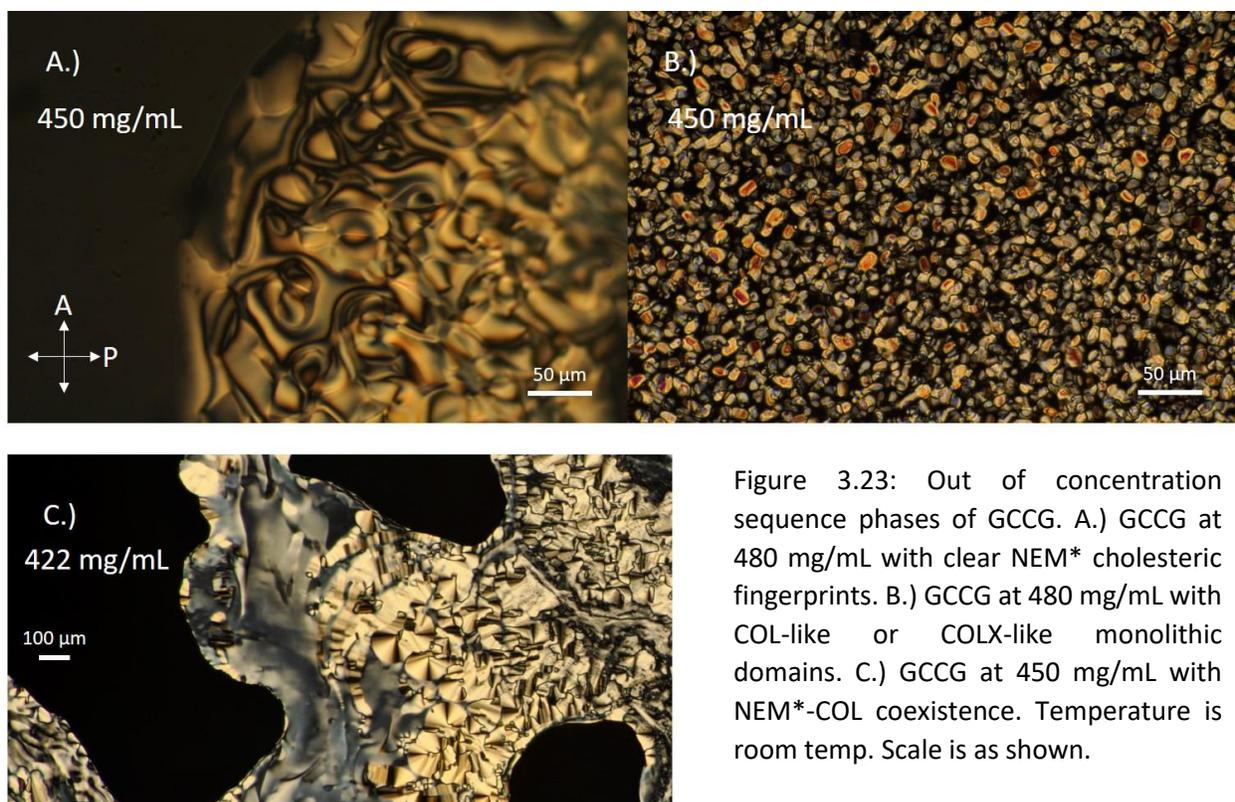


Figure 3.23: Out of concentration sequence phases of GCCG. A.) GCCG at 480 mg/mL with clear NEM\* cholesteric fingerprints. B.) GCCG at 480 mg/mL with COL-like or COLX-like monolithic domains. C.) GCCG at 450 mg/mL with NEM\*-COL coexistence. Temperature is room temp. Scale is as shown.

In Figure 3.23, these three varieties of mesophase are all observed clearly outside of the typical phase sequence with a higher order columnar-like phase observed at the same concentration as a lower order NEM\*, whose fluidity was clearly recorded with visible decompression of its cholesteric helix (the helix here being of ambiguous directionality, and perhaps disordered across the width of the texture). Normal COL and NEM\* focal conic and Grandjean textures were then seen to coexist at a lower concentration. One might argue that this sequence of textures could be seen at the same concentration without specifying the temperature dimension, except that all of these observations occurred at the same temperature. One might also argue that these sorts of observations could occur in a sample cell with large concentration gradients, which is also true, except that these observations occurred over and over again regardless of efforts to control for them. With the complexity of the disrupted mesophase sequence, there was also substantial difficulty determining even a clear ISO melting temperature in the

same microscopic cell; melting temperatures could vary widely from one day to the next in an initially unpredictable fashion. A generalized phase diagram attempting to capture the apparent steady state behavior was ultimately constructed using a combination of flat cells and flame sealed capillaries (see Figure 3.24).

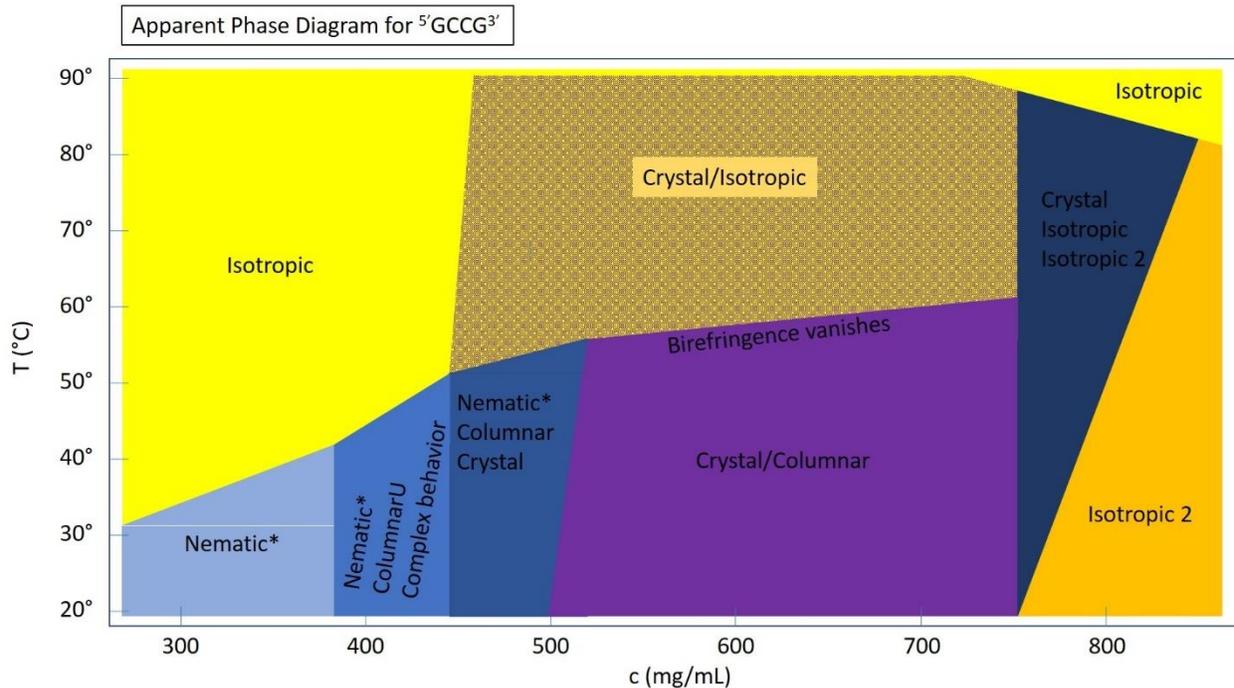
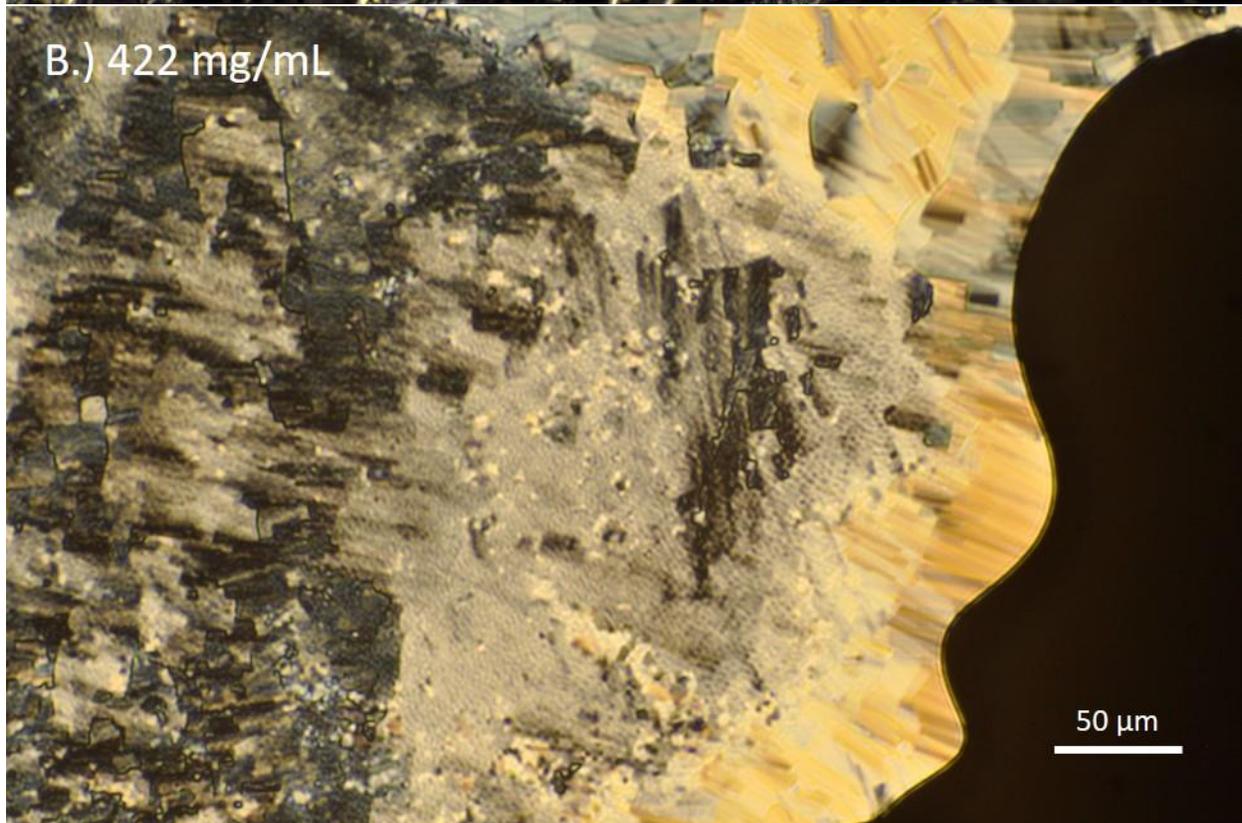
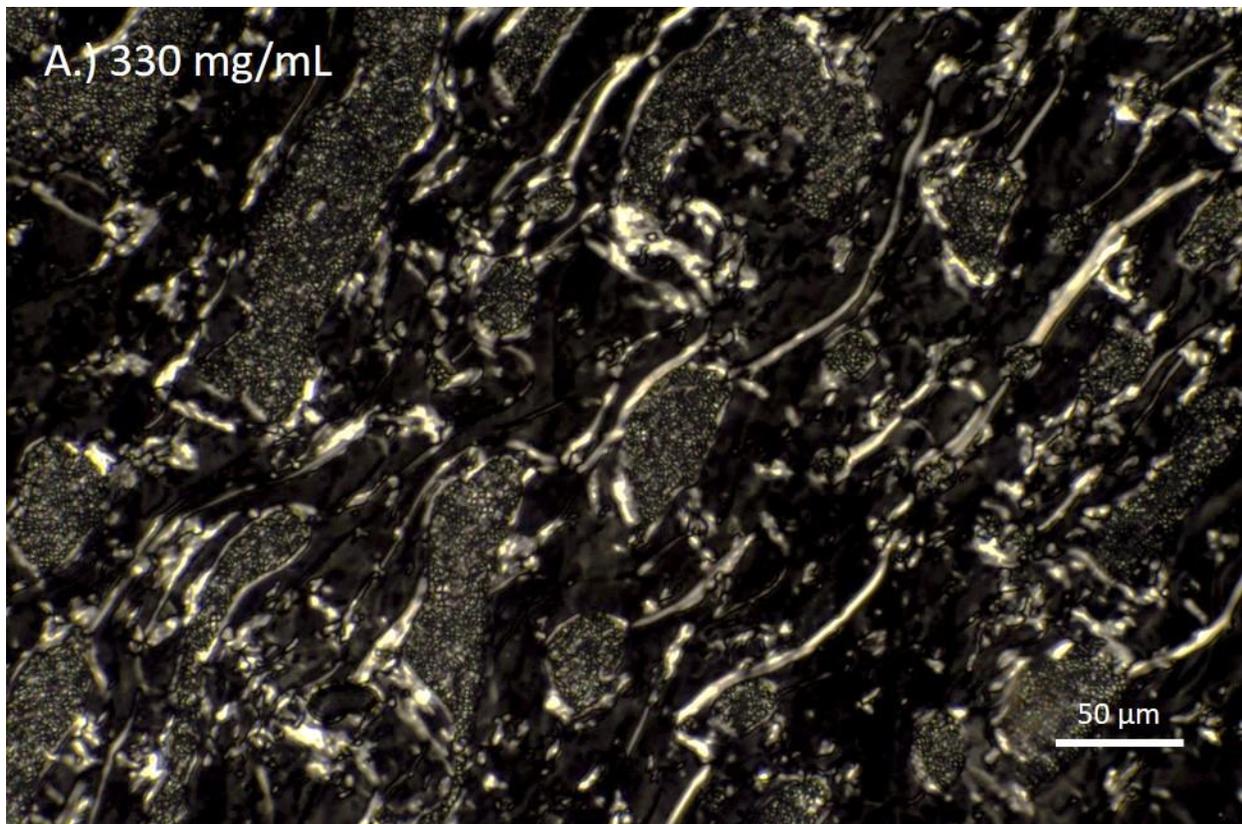


Figure 3.24: Block phase diagram for GCCG.

The phase diagram in Figure 3.24 attempts to capture the bizarre disorder seen in GCCG samples. In the region between 400 and 500 mg/mL, the phase sequence could fluctuate placing lower order phases at higher concentration than high order phases, or higher order at lower concentrations, all very unpredictably in the initial examinations. Additionally, an apparent crystal phase frequently appeared above 450 mg/mL that had a ISO  $T_c$  as high as 85° to 90° C. Birefringence almost always disappeared by around 50° C despite coexisting monolithic, poorly birefringent crystal blocks. In later observations, these crystals appeared after latency, making it difficult to determine the  $T_c$ . Some doubt existed as to whether the crystal might be a salt contamination: WAXS (Section 3.2.2) suggests that the dimensions of the

crystal cell are much larger than would be expected for a salt crystal. The one observed behavior similar to what was seen with GCCG-P was the apparent emergence of an ISO phase (called here Isotropic 2) at high concentrations.

GCCG mesophases in the region where NEM\* and COL type textures were observed, between 300 and 500 mg/mL, also showed some other textures which are unusual to nanoDNA (see Figure 3.25).



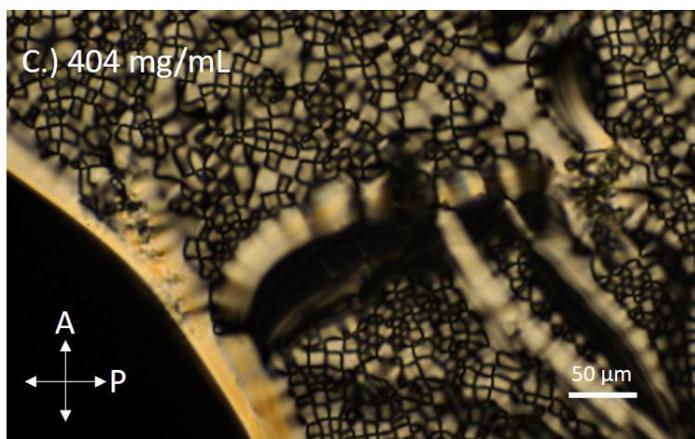


Figure 3.25: Unusual textures of GCCG mesophases. Pictures are included in large size so that the atypical features within the textures are visible. A.) 330mg/mL, stippled rafts in a NEM\* background. B.) 422 mg/mL, stippling with COL texture. C.) 404 mg/mL, parabolic focal conic-like snake skin texture. Scale is as indicated and polarizer orientations are the same in all images.

The textures shown in Figure 3.25 are enlarged so that the fine structure present in the mesophase is visible. The textures depicted in Figure 3.25 A.) and B.) are similar in appearing as tiny, frog egg-like clusters in masses or in rafts floating with more familiar NEM\* Grandjean or focal conic COL. GCCG also showed a parabolic focal conic like texture at NEM\* concentrations that almost always accompanied shifts in temperature, much like with GTAC but seemingly with greater chance of appearing. The exact mechanics of the parabolic focal conic textures was not noted in as great of detail as presented with GTAC (Figure 3.7).

The mesophases showing the tiny “stipple” (pointillist, frog-egg like clusters) are abnormally viscous. These structures so viscous that they can barely be pipetted despite their apparent fluidity, and the stipple appears to persist without coarsening into other structures for days to weeks. These textures can be melted with a temperature ramp and are clearly not permanent.

### 3.2.2 <sup>5</sup>GCCG<sup>3</sup> X-ray Diffraction Studies

GCCG samples contained in flame-sealed capillaries to preserve water content were subjected to the same sort of temperature varying WAXS study as carried out on GTAC and rDD (Section 3.1.2 and Section 2.2). As in previous cases,

none of the samples showed appreciable alignment and contained only powder averaged diffraction. Scattering patterns for these samples turned out frequently to be a complicated mixture of diffuse rings and sharp peaks from monolithic domains, much like with GTAC.

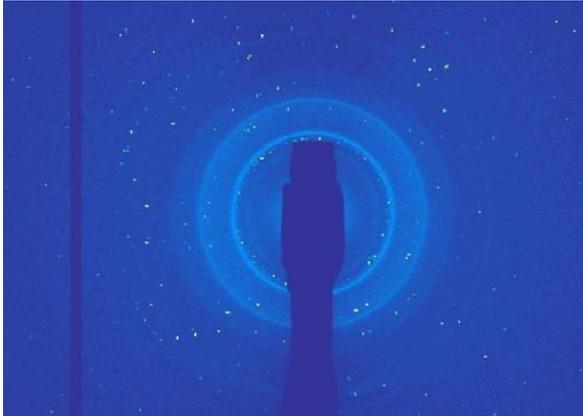


Figure 3.26: WAXS of 450 mg/mL GCCG at room temperature. Color again matches a terrain map with blue marking lowest counts and white marking highest counts. Shadowed features are a detector seam and the X-ray beam stop. The diffuse rings and the white spots are all due to scattering from GCCG. Region displayed here are  $q$  values approximately from  $0.1 \text{ \AA}^{-1}$  to  $0.5 \text{ \AA}^{-1}$ . See Figure 3.30 for more detail.

Results obtained before and after a temperature ramp melting the samples to ISO showed very different scattering properties. These observations will be first outlined and then analyzed.

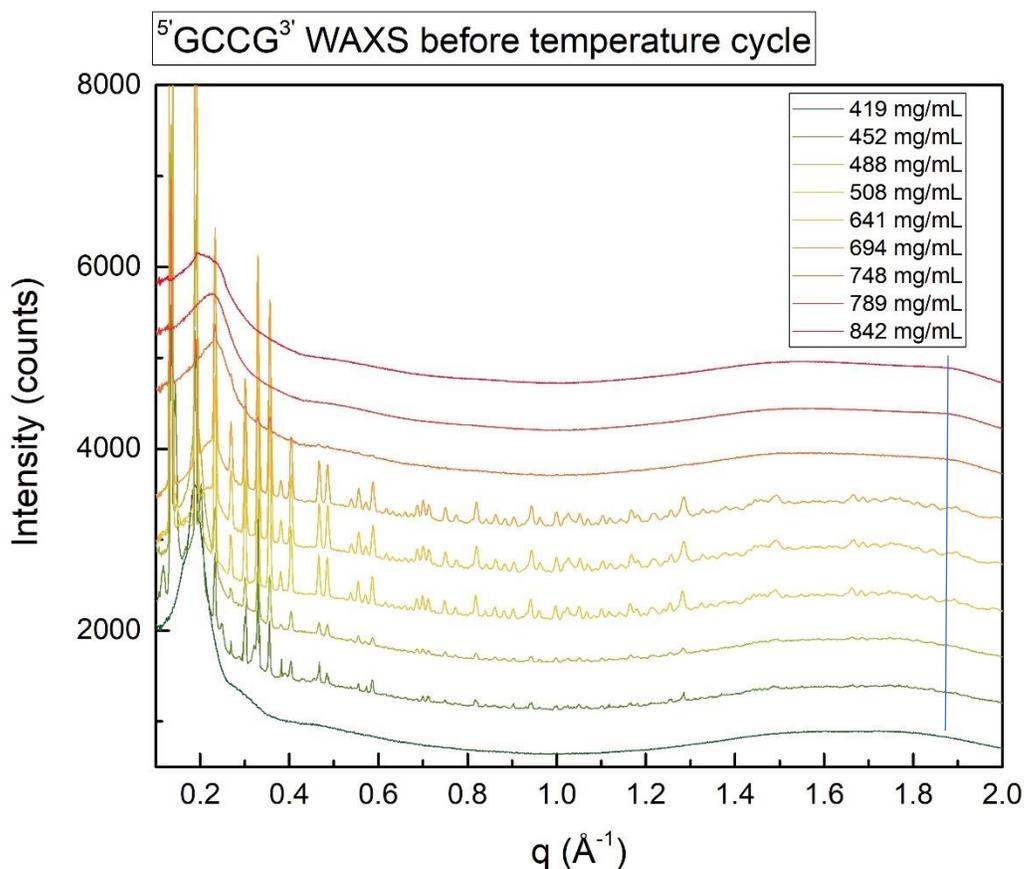


Figure 3.27: GCCG circularly averaged room temperature WAXS before temperature cycle. The full scattering range from 0.1 to 2.0 Å<sup>-1</sup> is shown. All samples are labeled by concentration. The blue line at ~1.86 Å<sup>-1</sup> marks the base stacking periodicity.

Almost all of the samples examined here showed complicated scattering patterns. Most of the visible scattering features are very sharp, containing small half-widths, and appear invariant in  $q$  with concentration, suggesting that they are incompressible. The base stacking peak is difficult to isolate from the clutter.

Focusing in on the small angle region gives a better feel for the complexity in Figure 3.28.

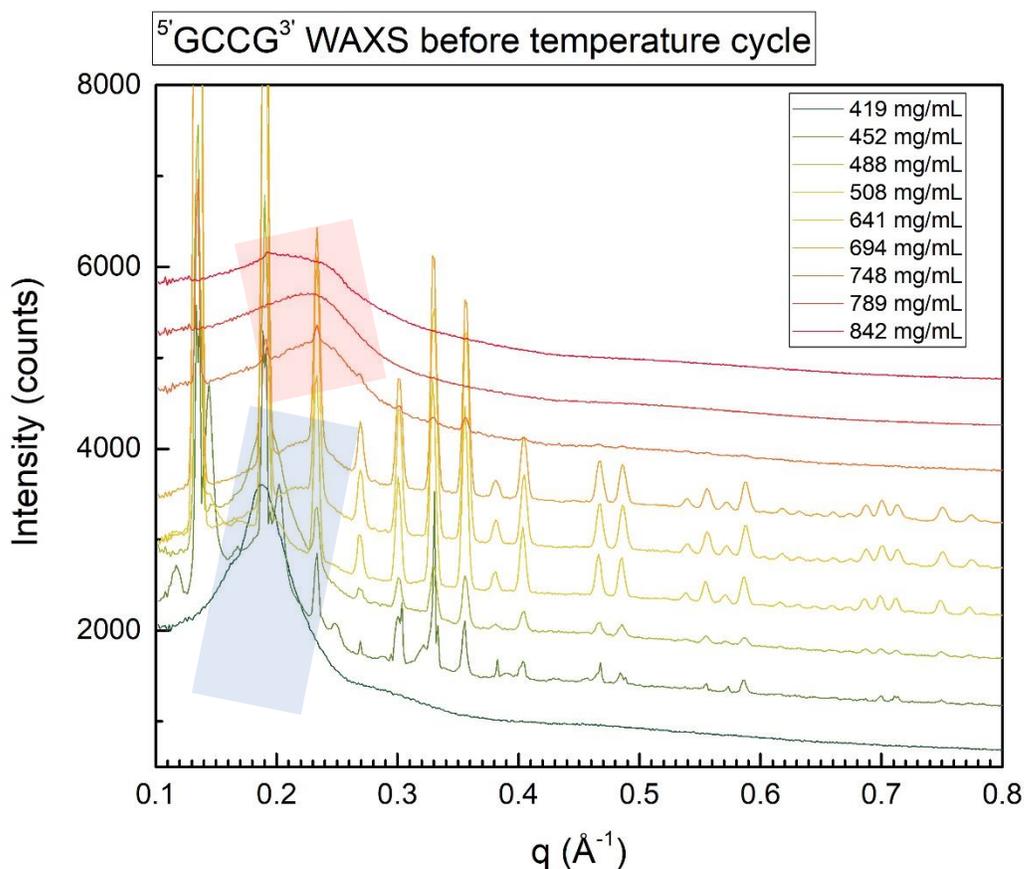


Figure 3.28: GCCG circularly averaged room temperature WAXS before temperature cycle. Small angle scattering range is shown. All samples are labeled by concentration. The transparent blue box draws attention to a series of broad underlying peaks coexistent with the crystal that appear to vary with concentration pushing generally to larger  $q$ . The transparent red box denotes a second set of underlying peaks that vary by concentration pushing to smaller  $q$ .

In Figure 3.28, the lowest concentration visible is understood to be NEM\* from its microscopic texture, but the next five higher concentration samples appear at first look to be dominated by the crystal form. That said, all five crystal dominated samples contain broad underlying peaks at  $\sim 0.2 \text{ \AA}^{-1}$  that appear to vary from one sample to the next. The three highest concentrations 748 mg/mL, 789 mg/mL and 842 mg/mL contain visible base-stacking peaks at  $\sim 1.85 \text{ \AA}^{-1}$  and, in 789 mg/mL and 842 mg/mL, only a broad NEM\* or ISO-like peak at  $\sim 0.2$  to  $0.25 \text{ \AA}^{-1}$ . 748

mg/mL GCCG contains signs of the crystal, but the peaks are weak compared to lower concentrations. None of these samples contain birefringence at temperatures higher than 40° to 50° C, but the crystal phase resists melting to as high as 85° C. The lack of birefringence of the crystal phase suggests strongly that the unit cell must be symmetric in such a way as to have no preferred fast axis. The broad, intermittent peaks present around  $0.2 \text{ \AA}^{-1}$  (blue and red boxes in Figure 3.28) do not persist to high temperature, suggesting that they are associated with the birefringent phases and these peaks do move from one sample to the next, suggesting compressibility. From this cursory overview, it seems likely that most samples shown here are mixtures of multiple mesophases.

The WAXS seen after the temperature ramp is greatly simplified (see Figure 3.29).

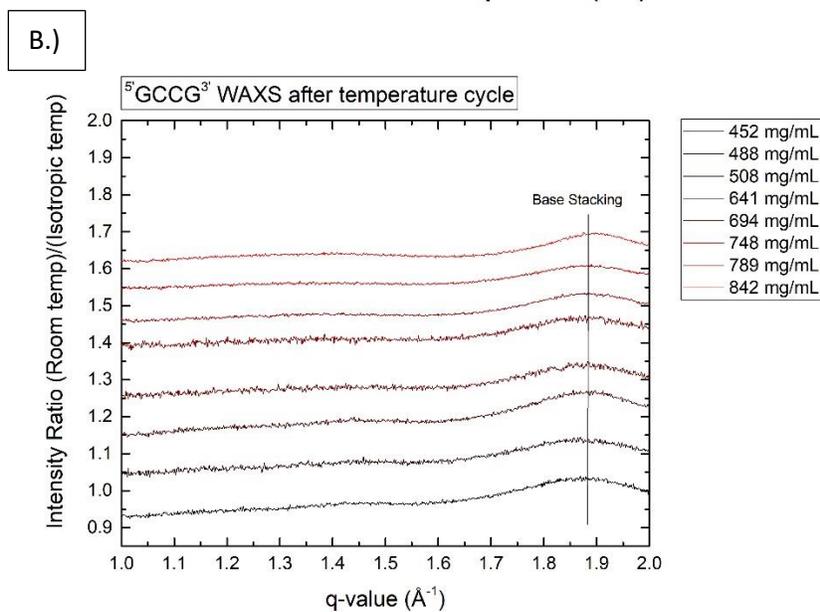
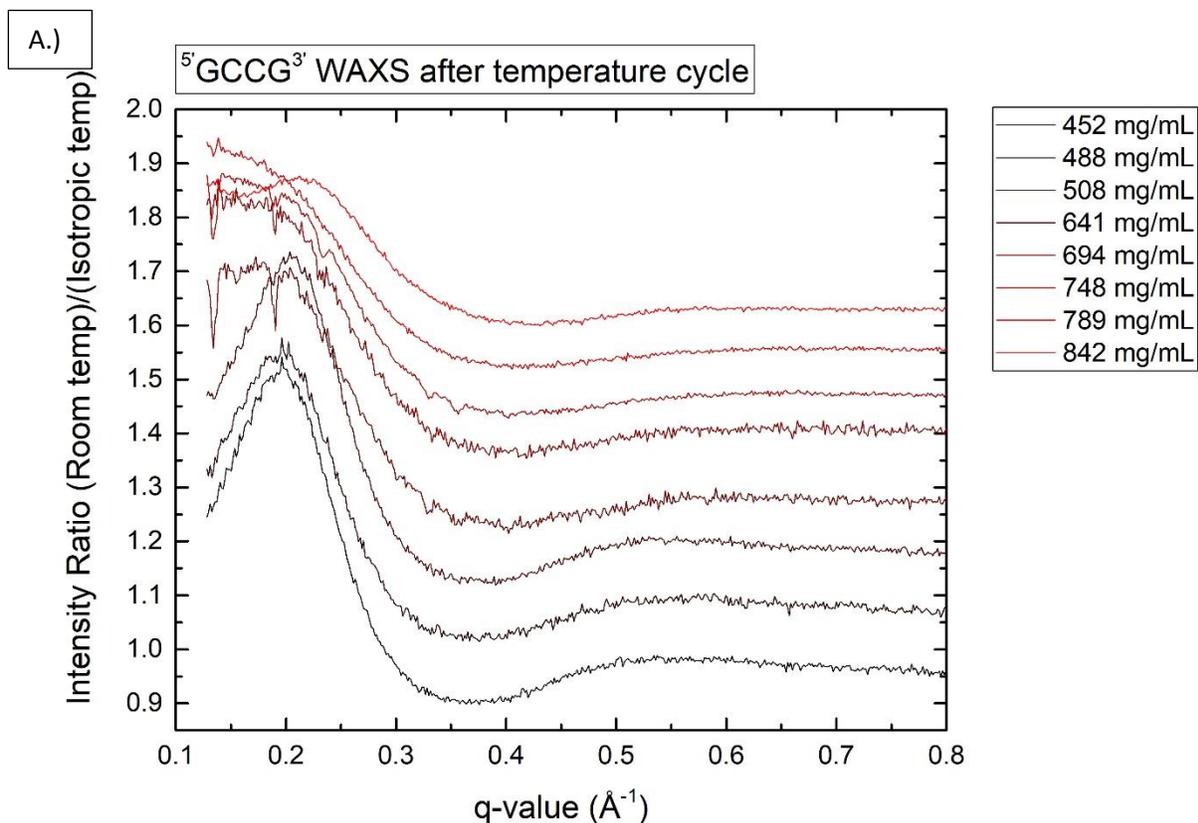


Figure 3.29: GCCG WAXS after temperature cycle. Concentrations are as depicted for A.) the small angle region and B.) the base-stacking region. The base-stacking periodicity is shown with a line. Intensity is normalized to show a ratio between the scattering at the ISO phase transition temperature and room temp.

The complicated scattering seen in Figures 3.28 and 3.29 does not recover from the temperature cycle. After the temperature cycle, all scattering is nearly featureless and appeared this way for the duration of observation available at the

synchrotron. This outcome hints strongly at mesophase hysteresis, where the phases observed are dependent on the thermal history of the sample. Low concentration phases responded somewhat differently to the temperature ramp from the high concentration phases. At low concentrations, when the temperature is dropped back to room temp, the phase annealed in a peak that looks very much like the NEM\* peaks seen for rDD and GTAC. At high concentrations, the side-side correlation peak shifts toward smaller angle and does not revert: this can be seen by comparing Figure 3.28 to 3.29 where the top four concentrations slide to smaller  $q$  after the temperature ramp (compare Figure 3.29 to red box in Figure 3.28).

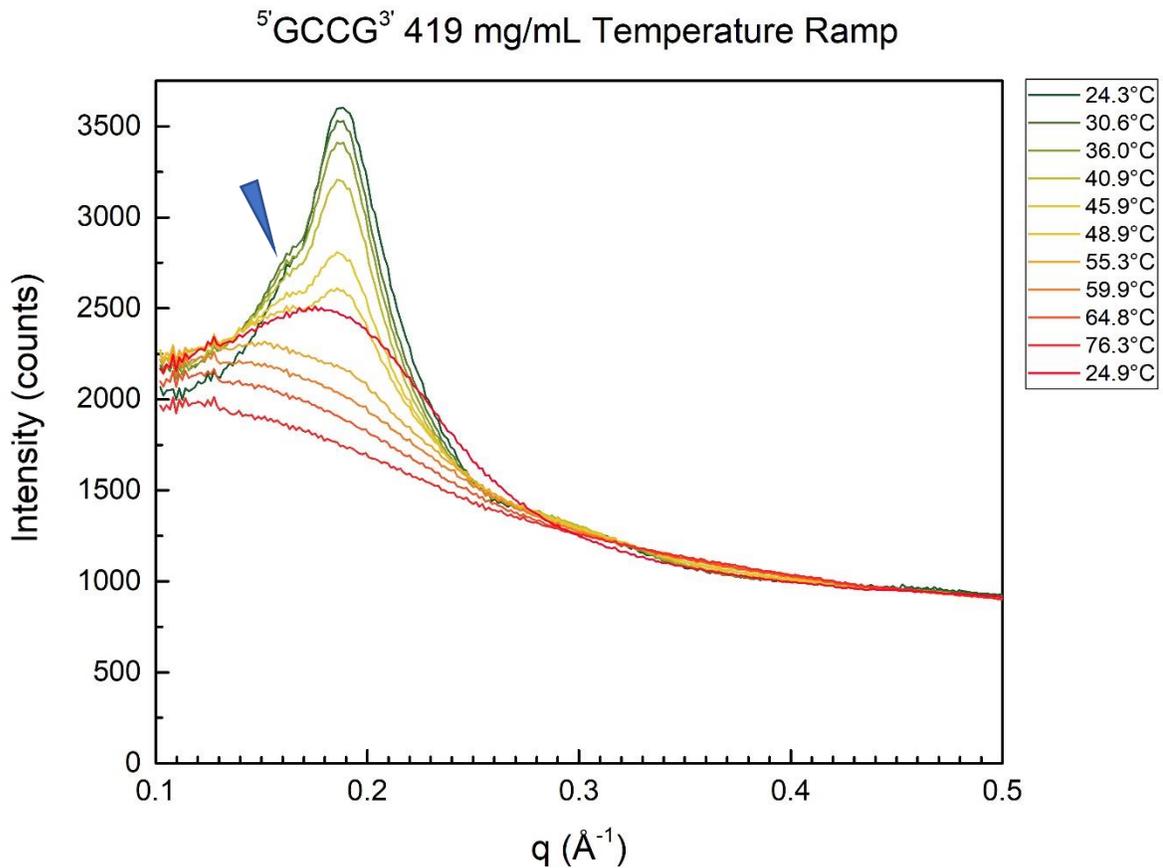


Figure 3.29: GCCG 419 mg/mL NEM\* WAXS with temperature ramp. Temperature ramp depicts a cycle where temperature is elevated until the mesophase has melted to ISO and then dropped rapidly to room temperature again in the last step. The pre-temperature

cycle peak. The blue triangle indicates a shoulder to the NEM\* peak which is not present in other nanoDNA.

For low concentrations, it is clear that the NEM\* present in GCCG is distinctly different from those seen in other nanoDNA. The GCCG NEM\* contains a peak shoulder indicating some deeper structure to the mesophase which is not present with rDD or GTAC. The reason for this peak is not known, but it suggests a secondary side-side correlation that is at a slightly wider spacing than the normal NEM\* spacing. This shoulder does not immediately recover from the temperature ramp, but has been seen on a later observation of the same mesophase, suggesting that it does ultimately recover.

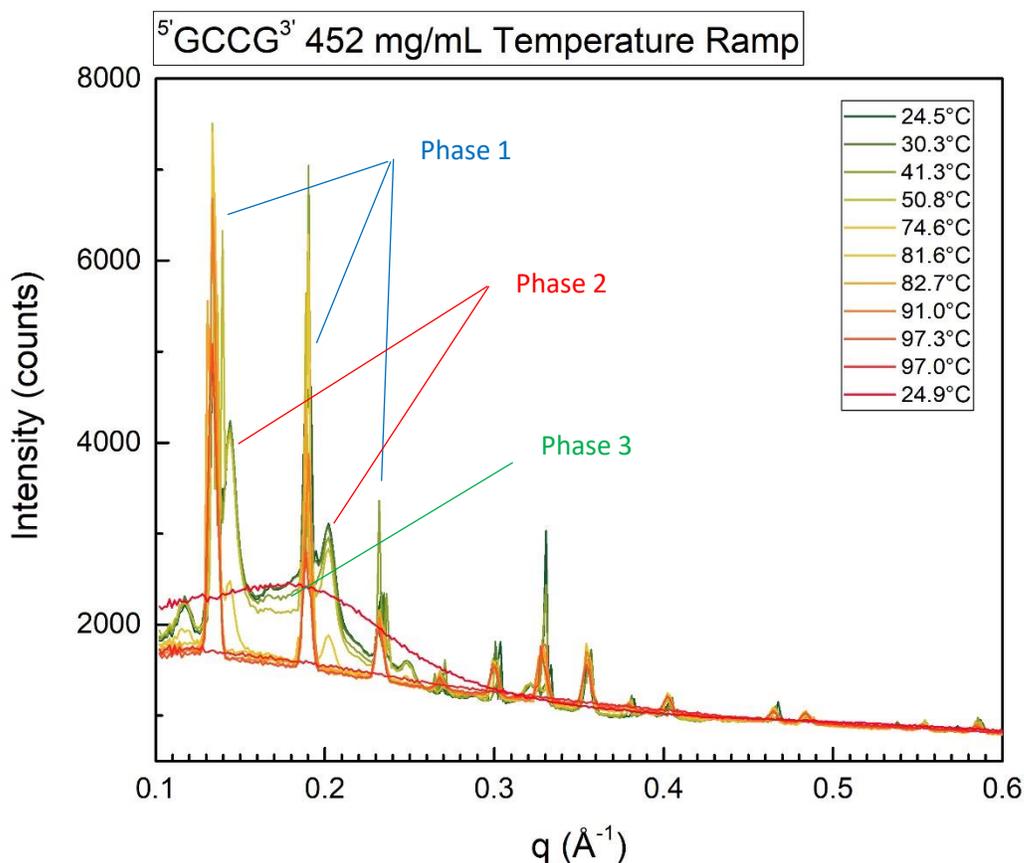


Figure 3.30: GCCG 450 mg/mL WAXS with temperature ramp. Temperatures during scans are as indicated. Three mesophases are picked out of the clutter based on peak

similarities to lower and higher samples. Phase 1 (Blue) is crystal, Phase 2 (Red) is not quite crystal but not COL, Phase 3 (Green) appears to be NEM\*.

Figure 3.30 handily illustrates the difficulty with evaluating GCCG X-ray diffraction: there appear to be not fewer than three mesophases present. Phase 1 is quite clearly the incompressible crystal seen in most of the higher concentrations, Phase 2 appears to be pseudocrystalline but is not the same as Phase 1, despite the clear similarity. Phase 3 is quite obviously NEM\* and clearly melts between 40° and 50° C. Phase 2 melts by about 70° C and Phase 1 finally melts at 90° C. Phase 2 and Phase 3 have diffuse rings in the 2D WAXS pattern while Phase 1 has highly ordered monodomains.

All of the samples where the crystalline Phase 1 dominates during the temperature ramp are fairly monotonous and can be summed up by a representative set (see Figure 3.31).

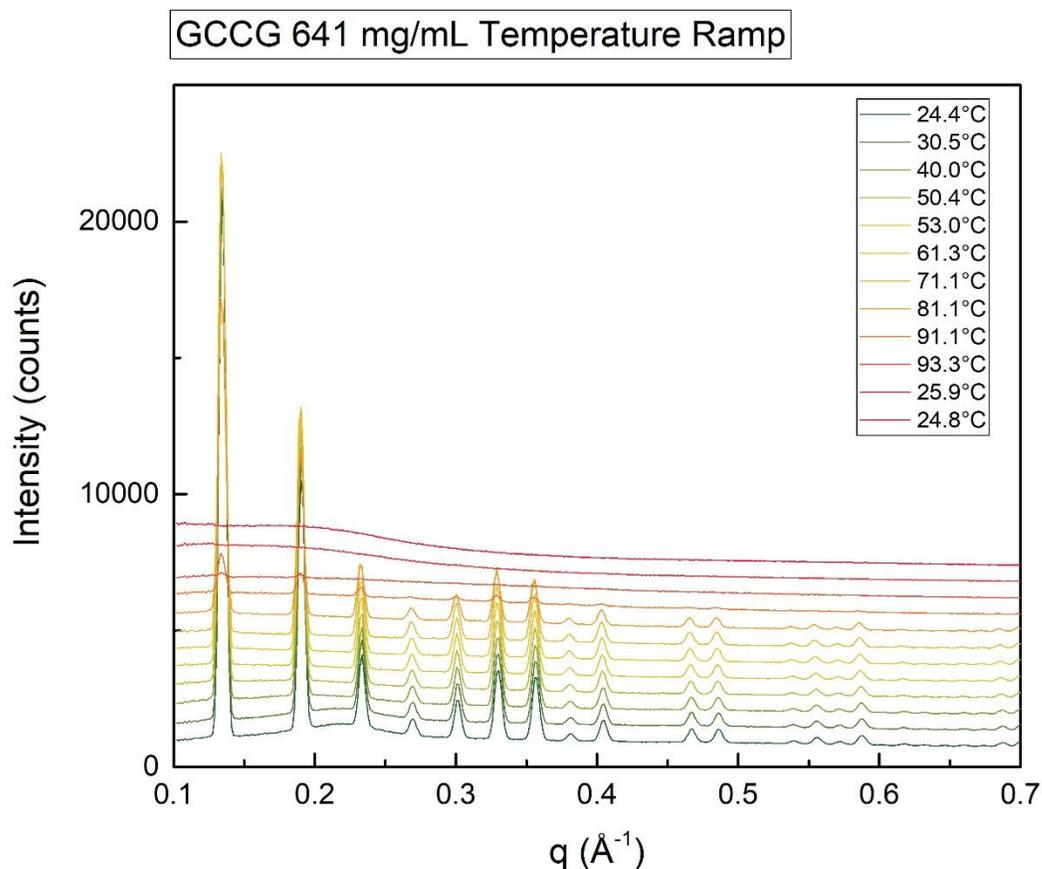


Figure 3.31: GCCG 641 mg/mL temperature ramp. WAXS scans are offset vertically in order to distinguish them.

Crystal phase 1 peak positions do not vary between concentrations and one data set can be used to extract peak positions for crystal indexing, which will be discussed shortly. Little more can be learned about this mesophase without an oriented sample.

On the other hand, the mesophase seen at higher concentrations than the region dominated by Phase 1 also can be understood somewhat in terms of Phase 1 (see Figure 3.32). One example occurred where the highest concentration mesophase snapped as an intermediate into Phase 1.

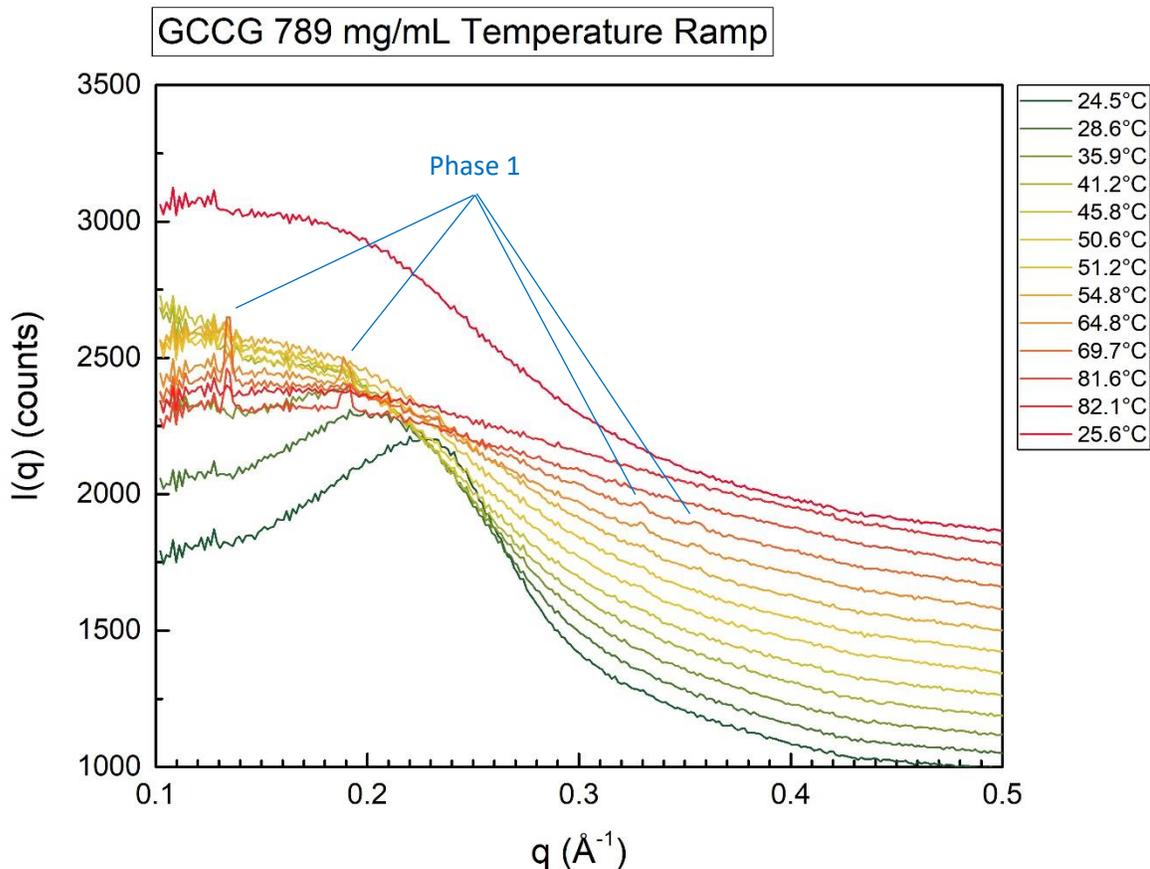


Figure 3.32: GCCG 789 mg/mL ISO2 phase. WAXS scans are stacked with an offset to help with clarity and peaks identified as Phase 1 are indicated in blue.

One of the initial objectives in examining GCCG was to address the phenomenon of a reemergent isotropic state, called here ISO2. In Figure 3.32, the capillary was understood to be poorly birefringent before examination and was expected to be a demonstration of ISO2. As the temperature was elevated in this capillary, the side-side correlation starts out looking vaguely nematic and shifts leftward with increasing temperature. At about 70° C, the mesophase undergoes a reorganization and peaks consistent with the Phase 1 crystal suddenly emerge before melting above 80° C. This appears to suggest that GCCG 789 mg/mL wants to be in the Phase 1 crystal structure but is somehow unable to do so until it undergoes a temperature annealing. At a sufficiently high temperature, the phase

snaps into the crystal structure. Since it is already known that these mesophases are abnormally viscous, it seems probable that ISO2 is a kinetically locked structure, like a glass, which can become mobile enough to order into the Phase 1 crystal if enough energy is supplied to break down the viscosity. Without any further appearances of the Phase 1 crystal, higher concentration samples than 789 mg/mL also have similar features to Figure 3.32 where they start with a correlation that looks nematic-like at just over  $0.2 \text{ \AA}^{-1}$  that shifts to lower  $q$  with increasing temperature before ultimately flattening. When these samples are brought back to room temperature, they reform a broad peak as seen in the final trace in Figure 3.32 which presumably ages into the narrower peak beginning the sequence in the same figure, a process which was not directly witnessed during observations at the synchrotron. All of this would be consistent with high viscosity blocking shifts in the organization of these mesophases.

The Phase 1 crystal is further characterized by annotating its peaks and indexing them to a crystal lattice system.

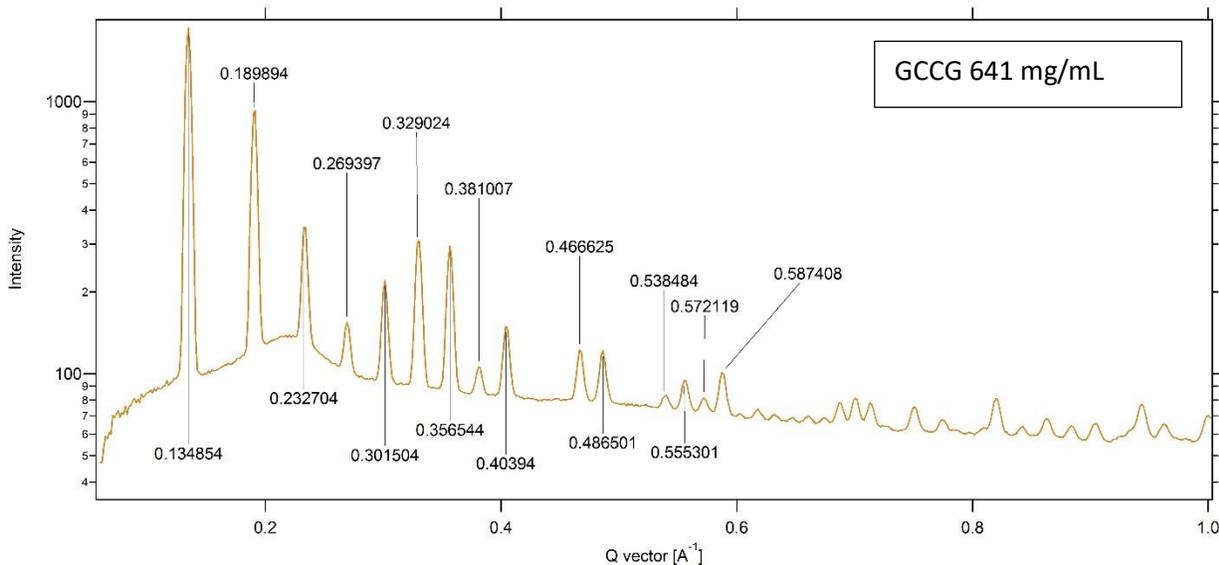


Figure 3.33: Annotation of Phase 1 crystal peaks.

The Phase 1 crystal system indexes to a Face Center Cubic (FCC) crystal lattice (see Section C.4.4) with reciprocal lattice  $\vec{A} = \vec{B} = \vec{C} = 0.067176 \text{ \AA}^{-1}$  where all basis vectors are orthogonal (see Table 3.4).

Table 3.4

Reflection $q$ -value	Index
0.13854	(2,0,0)
0.189894	(2,2,0)
0.232704	(2,2,2)
0.269397	(4,0,0)
0.301504	(4,2,0)
0.329024	(4,2,2)
0.356544	(3,3,3)
0.381007	(4,4,0)
0.40394	(4,4,2), (6,0,0)
0.466625	(4,4,4)
0.486501	(6,4,0)
0.538484	(5,5,4)
0.555301	(6,4,4)
0.572119	(6,6,0)
0.587408	(6,6,2)

The reflections for the FCC are too cluttered in 1D to efficiently index above  $\sim 0.5 \text{ \AA}^{-1}$ . The FCC determined here would be expected to have a fundamental reflection for its (1,1,1) index at  $0.116 \text{ \AA}^{-1}$  which seems to have fallen behind the beam stop.

The assignment of FCC agrees well with the microscopic observations of this mesophase given that it lacks birefringence: since the symmetry of FCC is identical in all three directions, the crystal system cannot have a preferred fast axis. This sort of crystal symmetry is frequently seen in salt crystals and simple ionic bonded structures, but the dimensions of such a crystal unit cell would be expected to be in the angstroms instead of the nanometers, arguing strongly that the assembly is dominated by a larger molecule like GCCG.

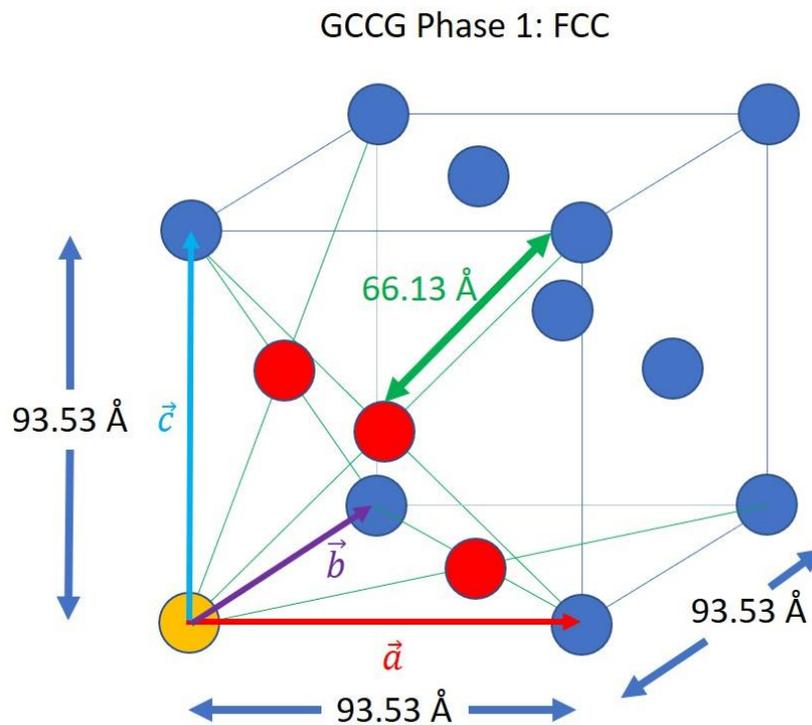


Figure 3.34: GCCG Phase 1 FCC crystal lattice. All lengths are to scale.

The occurrence of a 3D crystal with a 9.3 nm lattice spacing in a 1D data set is almost completely uninterpretable for an object the size of a nanoDNA 4mer. GCCG would be expected to have a backbone length of  $\sim 21 \text{ \AA}$  which could be extended by as much as 1.5 nm on either end to give  $\sim 5 \text{ nm}$  if the lengths of the nucleobases are stretched so that they enhance the length of the backbone. This stretches to cover only half of the edge dimension of the FCC, which leaves an

enormous additional amount of volume to fill. It can be anticipated that the crystal unit cell contains many asymmetric units. Each element of the FCC is separated from its neighbors by a distance of 6.6 nm, a length that could be readily bridged by two 4 nm lengths of GCCG nanoDNA.

Consider that the highest concentration where the Phase 1 crystal has been seen is something around 700 mg/mL with 9.3 nm of length for each of its three cubic axes. A back-of-the-envelope calculation suggests that if the Phase 1 crystal is 700 mg/mL, then  $5.727 \times 10^{-16}$  mg of material is present in one FCC crystalline unit cell. When GCCG has a molecular weight of 1172.8 g/mol, each of the individual GCCG oligomers weighs  $1.95 \times 10^{-18}$  mg. This allows 294 GCCG oligomers to be packed into the unit cell and split with 74 oligomers in each of the four FCC units. Conversely, if the crystal concentration is only 450 mg/mL, the minimum concentration where crystal is seen, 188.8 GCCG oligomers fit into the unit cell and 47 each in the four face centered units. The structure present would be difficult to forecast given the available information. This is somewhat baffling given the overall simplicity of the GCCG oligomer: with only a simple construction paradigm, how can GCCG consistently assemble into cubic symmetry objects containing such a large number of oligomers? One would expect an aggregate-type object to be of some average length or size. How are these so specific? Specifically structured assembly would be expected more for a longer DNA polymer.

One element that might fit into this puzzle is the G-quartet structure introduced in Figure 3.21. G-quartets have a square symmetry and could be anticipated to form consistently square structures. In turn, these square objects stack by hydrophobic forces into quadruplexes. Each GCCG oligomer possesses the capacity to participate in two quartets, allowing the bridging of two quadruplexes. Combining the Watson-Crick base pairing of G-C in the overhangs of the oligomer

to the possibility of G-quartet and G-quadruplex type structures, this combines a large variety of assembly characteristics into a relatively simple molecule. Bundled and bridged quadruplexes could easily be folded up into assemblies that are generally cubic. The whole object could easily be a form of G-quartet fractal.

Phase 2 seen in Figure 3.30 is a structure that is seemingly related to the Phase 1 FCC, but with less coherence and fewer obvious higher order peaks. The WAXS indexing for this mesophase is at least square with a 5.2 nm lattice. 5.2 nm is on the expected size scale for the stretched-out length of one GCCG oligomer. One could anticipate the capacity for this object to form a bridged lattice structure (Figure 3.35 and Figure 3.36).

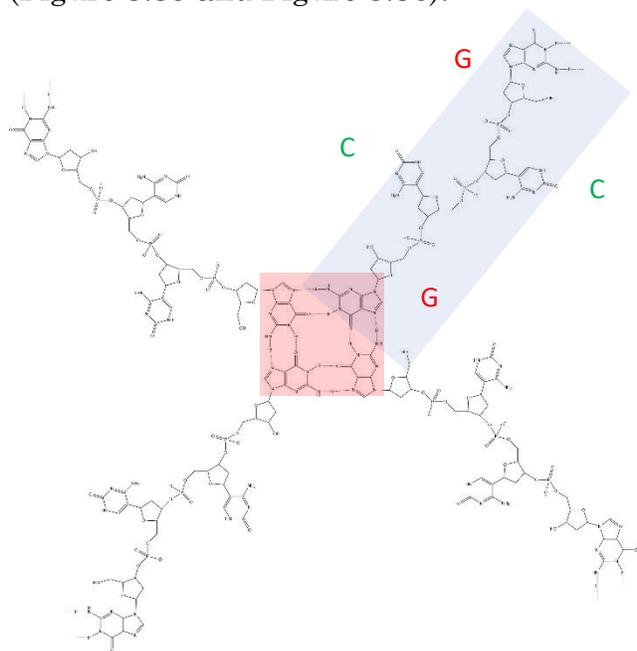


Figure 3.35: G-quartet mediated GCCG lattice. Four GCCG oligomers are shown; the red box outlines the G-quartet and the blue box outlines one GCCG oligomer. dG and dC side-groups on the GCCG oligomer are illustrated with colored letters.

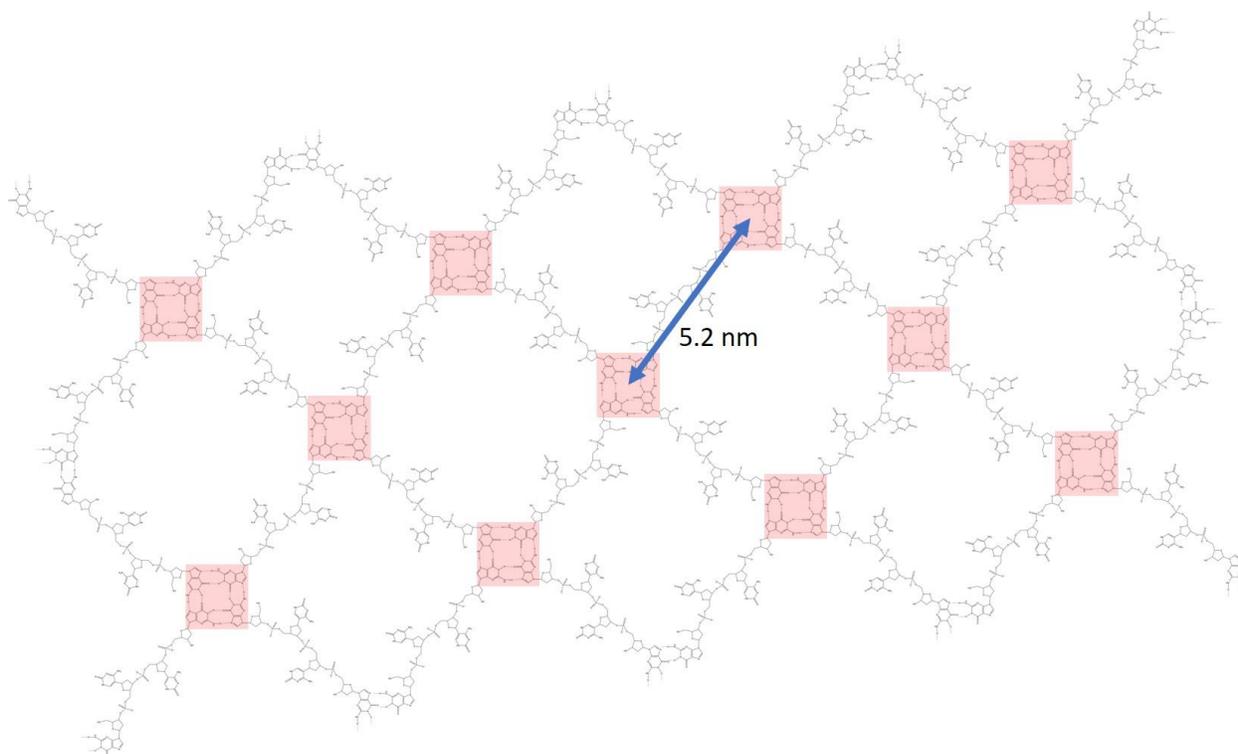


Figure 3.36: Hypothetical GCCG G-quartet mediated square Lattice. Each GCCG participates in two G-quartets. Of note, each G-quartet can stack into a quadruplex and the backbone portion of oligomer containing dC is free to flex in any direction, allowing this lamellar structure to fold up.

The existence of the 2D GCCG lattice depicted in Figure 3.35 is somewhat questionable given the overall lack of information about Phase 2. Phase 2 could also index as a slightly compressed and highly disordered version of the Phase 1 FCC.

Of note, the inclusion of the G-quartet as a part of the assembly scheme for GCCG mesophases may help to explain the odd shoulder peak seen in Figure 3.29 for the GCCG NEM\* phase. If GCCG can selectively undergo either Watson-Crick base pairing assembly or G-quartet association with stacking, the NEM\* might be expected to contain both the more slender G-C base paired aggregates that sometimes cross-link at their ends by G-quartet, creating larger diameter complexes part of the time and giving rise to the double peak. Heavy cross-linking by G-

quartet assembly would also explain the unusually high viscosity of the GCCG NEM\* phase.

### 3.2.3 <sup>5</sup>GCCG<sup>3</sup> Mesophase Hysteresis

The huge difference between in X-ray scattering between before melting GCCG mesophases to ISO and after they have been melted and reannealed (Figures 3.28 and 3.29) suggests an explanation for the unprecedented instability seen with the GCCG phase diagram (Figure 3.24). GCCG mesophases are subject to a rather extreme thermal hysteresis and exactly which phase appears during a given observation condition is due to how that sample was treated in the days or even weeks leading up to observation. In order to unequivocally confirm the existence of this hysteresis, the 450 mg/mL GCCG capillary sample examined by WAXS in Figure 3.30 was subjected to a repetitious observation regimen with carefully controlled temperature conditions over a long period.

The sample initially sat at room temperature in the lab for several weeks before being placed in the hotstage. In the hotstage, the sample was subjected to temperature cycling to specific temperatures with long term holds at those temperatures lasting for a period of days. Experience from this series of temperature cycles is collected together into a diagram for brevity which will guide discussion of the experiment (Figure 3.37).

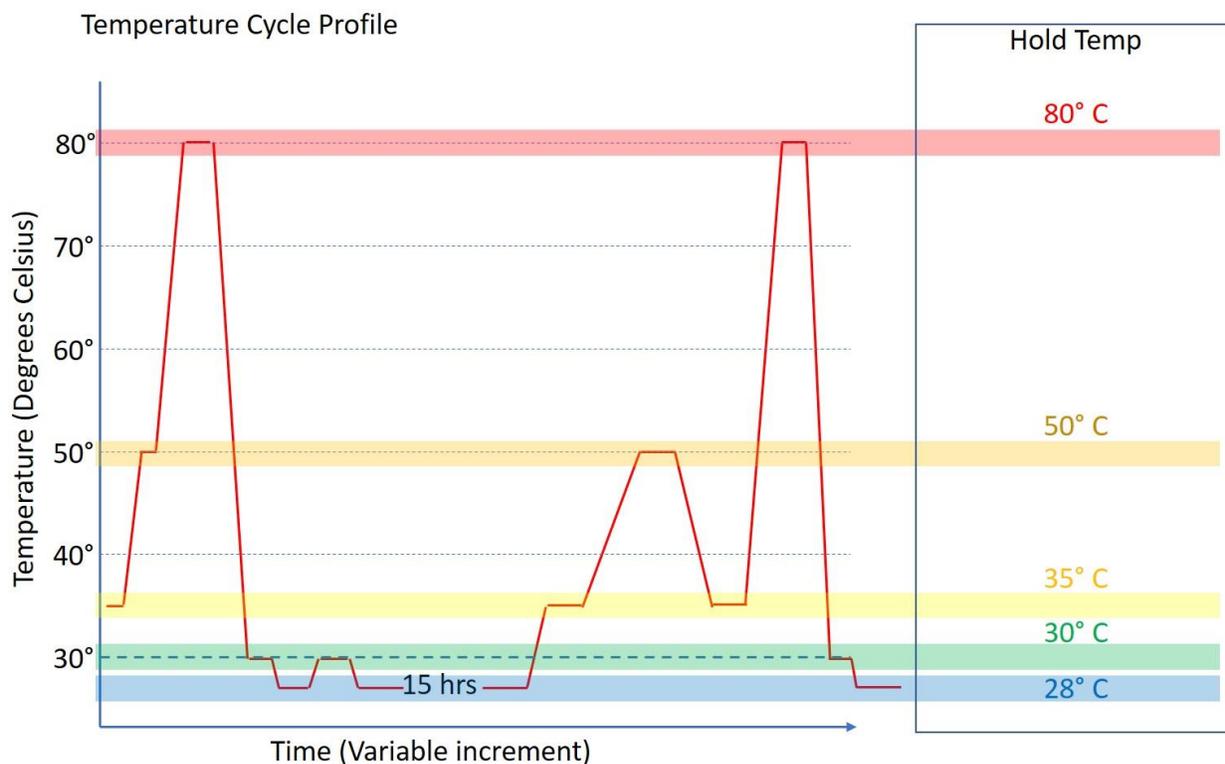


Figure 3.37: Temperature cycle profile for hysteresis experiment with GCCG 450 mg/mL. The various hold levels are outlined in color. 80° C (Red) was selected because it was known from the WAXS that all structure melts to ISO when held at that temperature. 50° C (Orange) was selected because all comparable 4mer oligomers (such as GTAC) melt to ISO at this temperature. 35° C (Yellow) was selected because this temperature is where 4mers typically become dynamic, not necessarily melting, but sometimes in ISO coexistence. 30° C was selected because this is a stable temperature for 4mers and 28° C is even more stable. The experimentation was intended to ramp between these levels while starting with a particular phase from a particular history and examine how the sample behaves with this treatment. Each different section of the experiment will be handled as a separate figure.

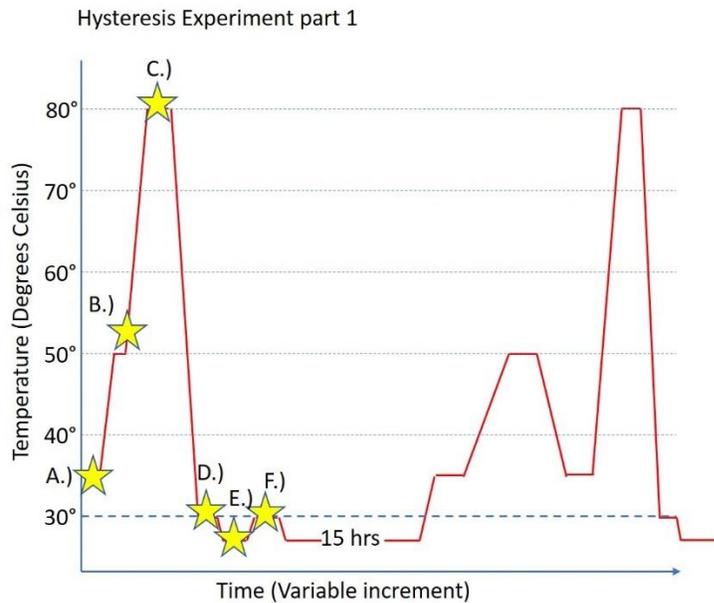
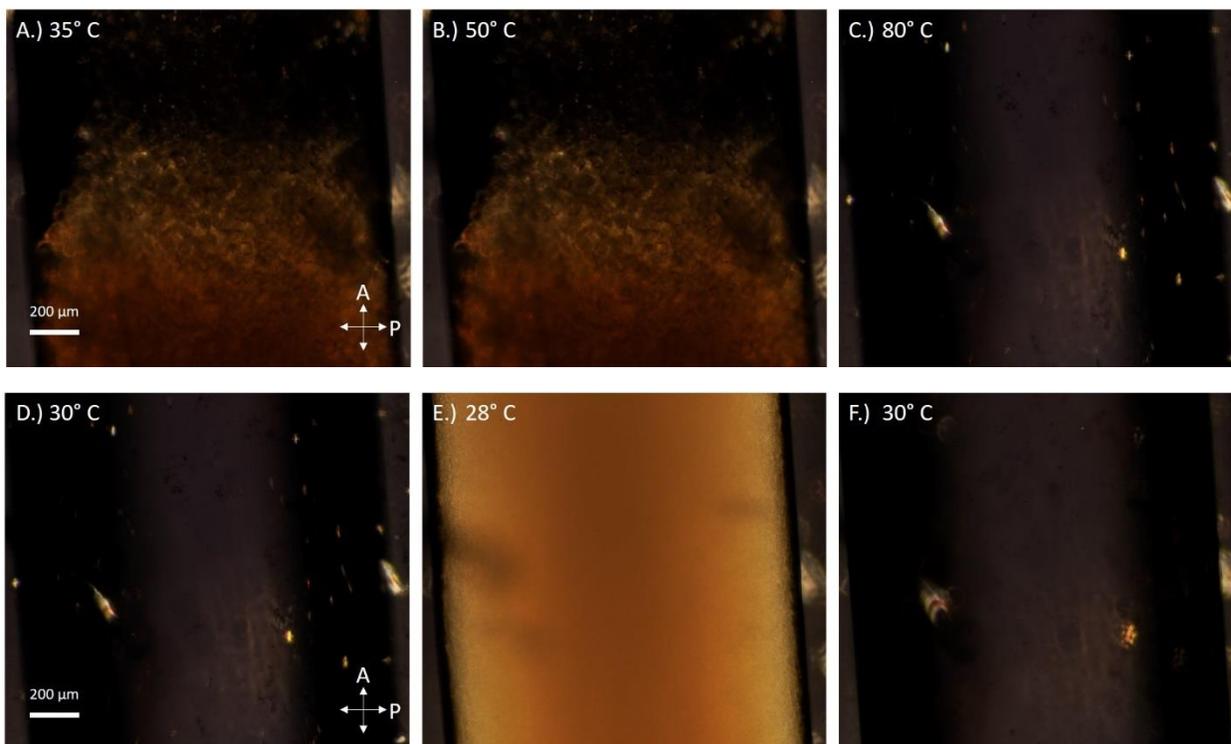


Figure 3.38: Hysteresis Experiment part 1. The location for the image in the temperature sequence is illustrated at left with a star and the letter for the image. All images are the same scale and their temperatures are marked. A.) Crystal at 35° C. B.) 50° C Crystal. C.) 80° C ISO after crystal has melted. D.) 30° C ISO, stable. E.) 28° C NEM\*. F.) 30° C ISO, after NEM\* melts.



The figure above (Figure 3.38) shows the first part of the hysteresis experiment. Within the figure, the 450 mg/mL sample begins the experiment in the Phase 1 crystal after weeks of room temperature incubation. Phase 1 persists to 80° C before melting. Upon returning to 30° C, ISO is now perpetually stable. NEM\*

forms if the temperature is dropped just 2° to 28° C, but NEM\* melts back to ISO if temperature is reverted immediately to 30° C.

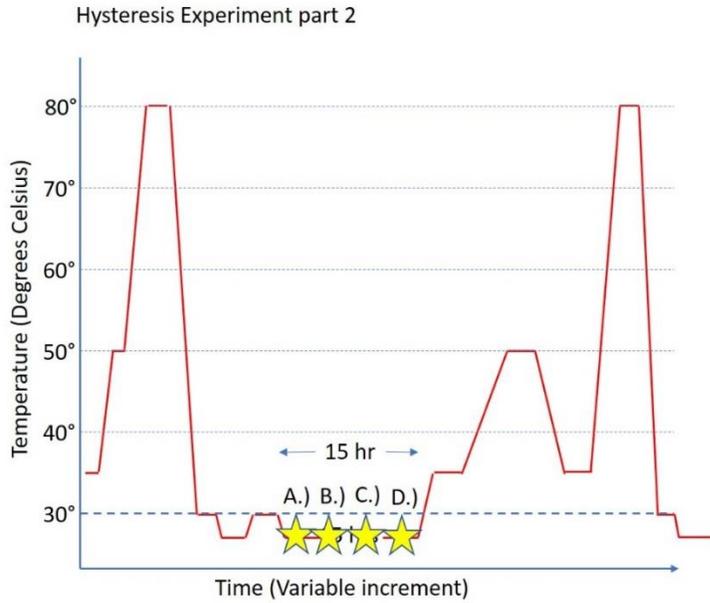
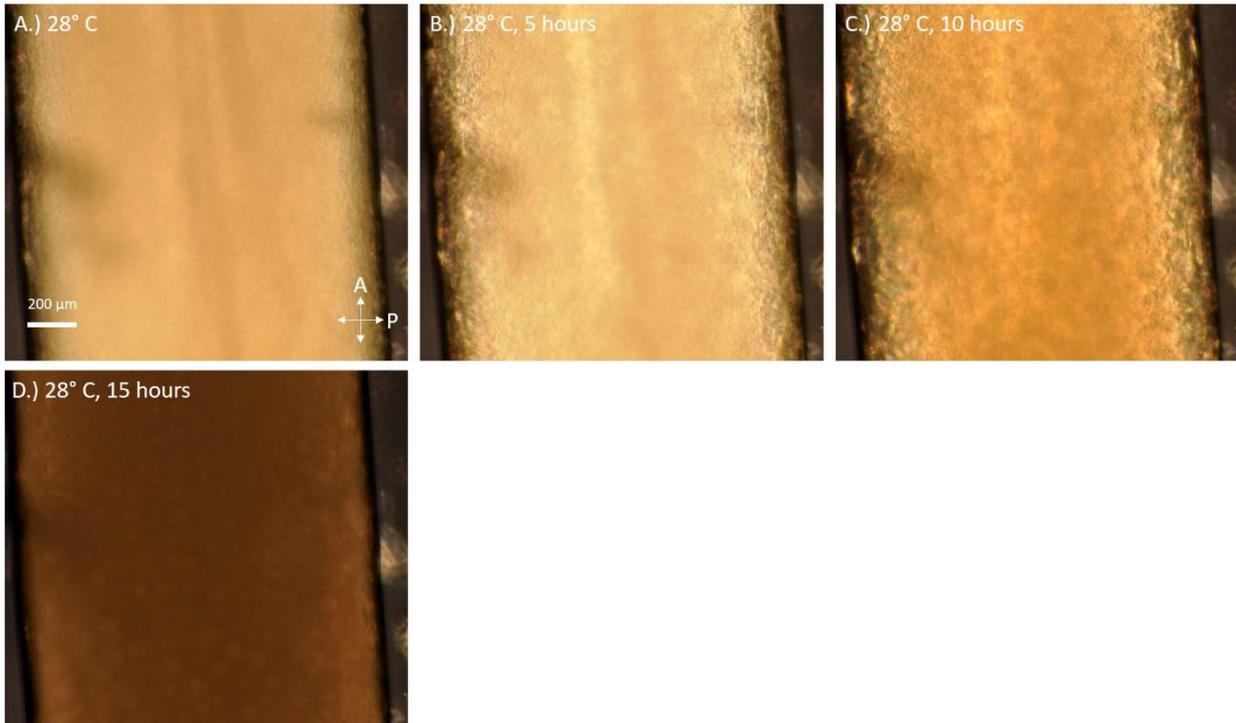


Figure 3.39: Hysteresis Experiment part 2. The location for the image in the temperature sequence is illustrated at left with a star and the letter for the image. All images are the same scale and their temperatures are marked, as is the duration of time at a step. A.) 28° C NEM\* immediate. B.) 28° C NEM\* after 5 hours. C.) 28° C NEM\* after 10 hours. D.) 28° C NEM\* after 15 hours.



In the figure above (Figure 3.39), the sample is held at 28° C and is allowed to coarsen for 15 hours. The phase undergoes a steady evolution in texture until the

quality of its birefringence appears to shift. Nothing more was done but to leave the sample at a temperature and incubate. The sequence depicted here has been observed multiple times to verify its repeatability.

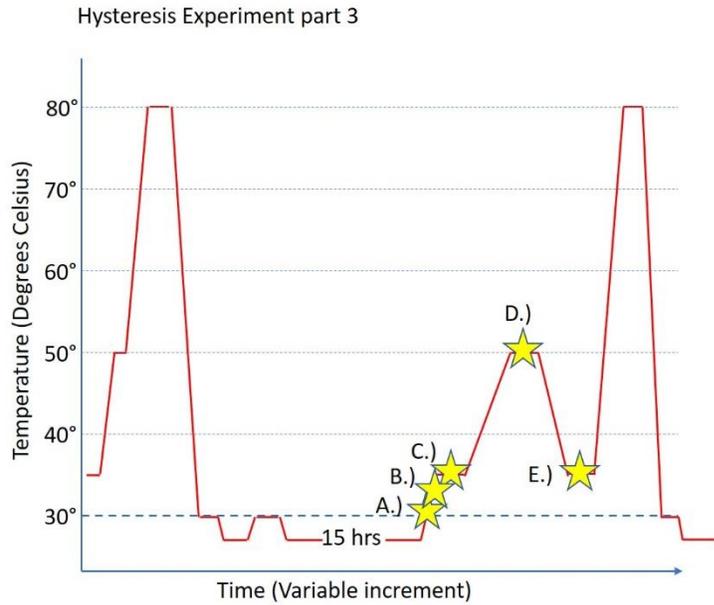
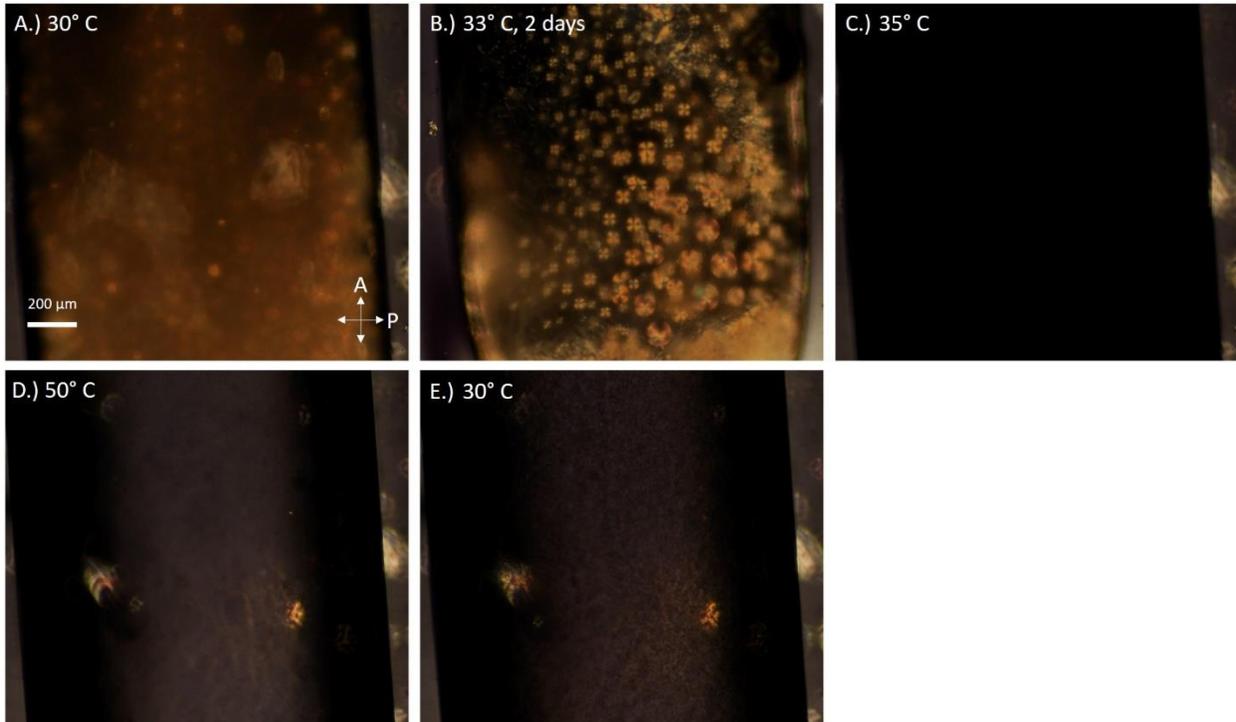


Figure 3.40: Hysteresis Experiment part 3. The location for the image in the temperature sequence is illustrated at left with a star and the letter for the image. All images are the same scale and their temperatures are marked, as is the duration of time at a step. A.) 30° C NEM\*-Dark. B.) 33° C NEM\*-COL-Dark after 2 days. C.) 35° C Dark phase. D.) 50° C Dark-ISO. E.) 30° Dark-ISO.



The figure above (Figure 3.40) introduces the existence of a dark phase seen just above the NEM\* after the sample has been incubated for 15 hours. If temperature is elevated to 30° C after the 15 hours incubation, it no longer melts to ISO, but instead displays a poorly birefringent texture that appears to be coexistent with the NEM\*, called here Dark phase. When allowed to sit at 33° C for 2 days, this coexistence generates columnar-like focal conic textures.

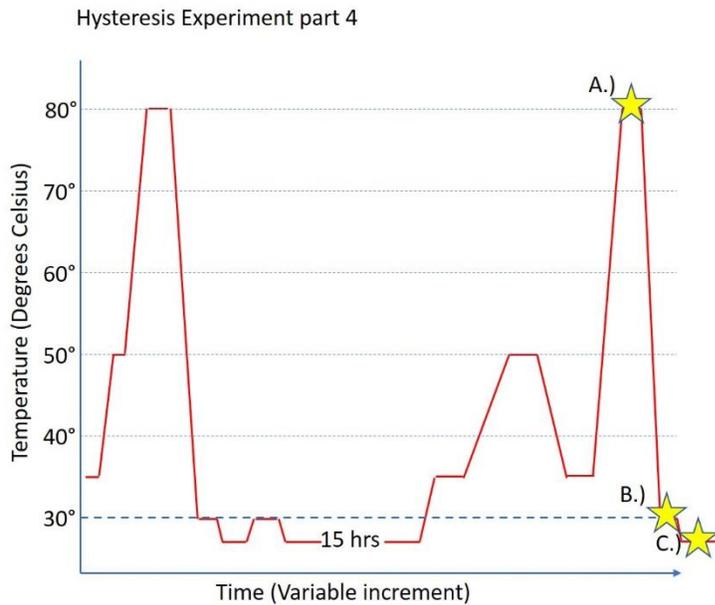


Figure 3.41: Hysteresis Experiment part 4. The location for the image in the temperature sequence is illustrated at left with a star and the letter for the image. All images are the same scale and their temperatures are marked. A.) 80° C ISO. B.) 30° C ISO. C.) 28° C NEM\*. This section of the experiment brings the hysteresis cycle full circle.



The section of the hysteresis cycle containing the dark textures is partly reversible and partly not reversible. Once the 15-hour coarsening step has been carried out, as in Figure 3.39, the mesophase is clearly stabilized so that it has birefringent textures in a region of the phase diagram that was formerly dominated

by ISO (in Figure 3.38). For the section of temperatures 30° to 35° C a variety of textures appear that could be interpreted as higher order COL (as seen in Figure 3.40 B.). The dark texture that appears coexistent with the birefringent textures was initially interpreted as ISO until it was noticed at 35° C that absolutely no light was being transmitted through the capillary (Figure 3.40 C.). This dark texture can be reversed to NEM\* coexistence by simply lowering the temperature to 30° C and it is apparently stable for long periods of time given that it can persist for at least days –all in a region of the phase diagram that contained only ISO in Figure 3.38. The Dark phase demonstrates an irreversible shift in texture if the temperature is elevated briefly to 50° C, where the texture thins and becomes translucent. If the temperature is lowered to 30° C after the brief bump to 50° C, the phase recovers as a granular texture (Figure 3.40 E.) devoid of birefringence.

The phase sequence can be recovered to ISO at 30° C by simply ramping to 80° C. The sample returns to demonstrating a phase transition to NEM\* at 28° C, which is again a sharp cut-off where elevating back to 30° C causes melting to ISO. It seems likely that extending the coarsening step in Figure 3.39 from 15 hours to a duration of weeks permits access to the temperature resistant crystal seen in Figure 3.38, tying the whole pattern together. Regarded naively, at a single concentration of 450 mg/mL, this system would appear to have three distinct phase diagrams (Figure 3.42).

Phase diagrams for 450 mg/mL 5'GCCG3' by hysteresis

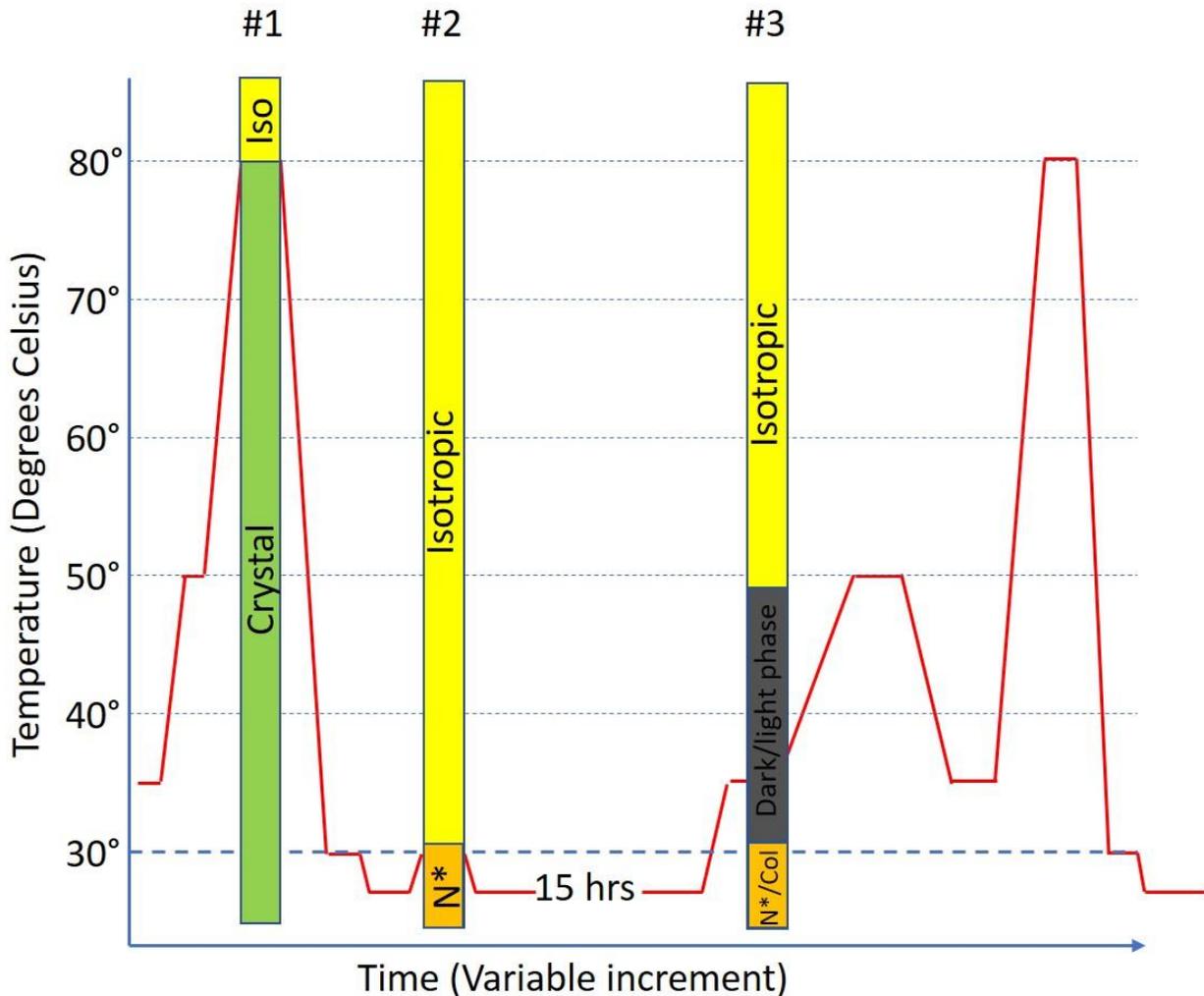


Figure 3.42: GCCG 450 mg/mL phase diagram depending on hysteresis. There appear to be three different, distinct phase diagrams, #1, #2 and #3, for this material at only one concentration.

From the WAXS data presented in Section 3.2.2, it seems likely that the crystal phase seen here is the Phase 1 FCC. Phase 2, either compressed FCC or square 2D lattice with no consistent orientation, may well be the dark phase on bar #3 in Figure 3.42. It would be expected that Dark phase would become the crystal with significant incubation. The appearance of COL textures and NEM\* textures depending on hysteresis would appear to solve the riddle of the confounding

mesophase variability brought to light in the initial GCCG phase diagram (Figure 3.24).

The hysteresis would seem to imply that there are several overlapping self-assembly paradigms, one that is accessible very quickly, on short time scales, and one that begins to take over after longer periods of time. Likely, huge variability occurs in the material when both paradigms are coincident, before one has completely overwhelmed the other.

#### 3.2.4 G-quartet Participation in $5'GCCG3'$ Mesophase Formation

Circumstantial evidence suggests very strongly that G-quartet formation plays a major role in the rich mesophase variability seen with this relatively simple nanoDNA sequence. With only four nucleobases, this sequence has unquestionably the most confounding complex behavior of any nanoDNA yet observed. The overlap of a short size with multiple possible assembly modes could easily give this oligomer its multimodal character.

Directly establishing the role of G-quartets in this phase diagram offers another challenge given the lack of a direct probe for G-quartets to this point. One possible way of accessing this information lies in a savvy examination of the nature of the G-quartet. Unlike Watson-Crick base pairing, which is mediated solely by hydrogen bond interactions and hydrophobic base stacking, G-quartet assembly depends on hydrogen bond associations and on the participation of a monovalent cation as a ligand in the core of the quartet. The G-quartet does not necessarily need to make hydrophobic stacks to be stable. On the other hand, the cation ligand depends strongly on the identity of the ion, with Sodium and Potassium both able to mediate quartets, while larger and smaller ions (Cesium and Lithium respectively)

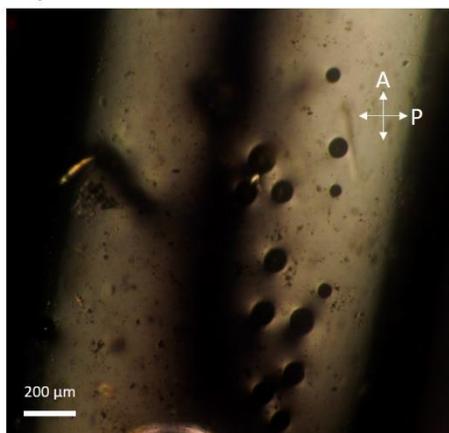
become incapable<sup>40</sup>. This liability makes it possible to perform PCR on G-rich sequences --where G-quartets can interfere with base pairing associations in the experiment and significantly increasing the potential for mutations-- by use of dNTPs that omit sodium or potassium as a counterion in favor of Lithium.

We reasoned that a similar trick might be possible with nanoDNA LC. We have previously established that the presence of Sodium counterions is required for formation of LC phases by quenching repulsive forces from the phosphodiester backbone (Sections 2.4). This would mean that Sodium we have deliberately introduced to stabilize the phases might also be available in the case of GCCG to form G-quartets at the expense of Watson-Crick G-C pairs. So, with GCCG, if G-quartets have complicated the self-assembly routes, behavior might be simplified by substituting Lithium in place of Sodium, suppressing G-quartet formation and focusing solely on G-C pairing.

A sample of GCCG was taken through the desalting method (Section A.6.2) but with LiCl in place of NaCl. The identical process was carried out on a sample of rDD as a control since the 12mer does not demonstrate the same sorts of bizarre behaviors.

The outcome of this experiment was extremely surprising: the entire phase diagram of GCCG disappeared! None of the phases reported in Section 3.2.1 were observed at all. Only one birefringent texture appeared and this was a very weak texture seen at 5° C (see Figure 3.43). On the other hand, rDD substituted with Lithium still demonstrated its entire phase diagram, including ISO, NEM\* and COL.

A.) GCCG 5° C



B.) Control rDD Room temperature

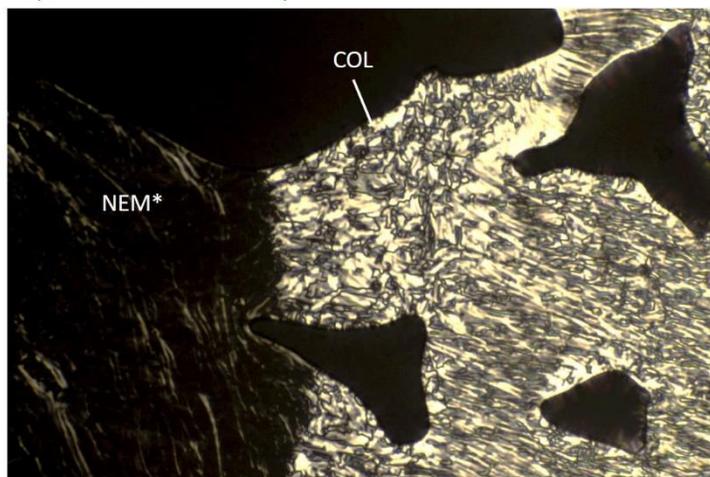


Figure 3.43: Phases of GCCG-Li+. A.) The only birefringent textures seen with lithium substituted GCCG, at 5° C and completely distinct from previously seen textures. B.) Birefringent textures seen with lithiated rDD, including a long pitch NEM\* in Grandjean texture and typical COL.

To reiterate, substituting Lithium for Sodium in GCCG completely destroyed the GCCG phase diagram. Essentially no phase behavior existed in the material afterward save a very weak birefringence seen at 5° C and in a texture that is not familiar. 12mer DNA rDD did not lose its phase diagram in the same way. There is some small doubt as to whether or not Lithium was fully substituted for Sodium in the rDD sample which would demand some additional experimental replication, but the result is pretty conclusive.

What this means is slightly uncertain: either Lithium kills GCCG for a reason that is not fully understood, or G-quartets are essential for this 4mer to have mesophases at all.

GCCG has an assembly hierarchy that is dependent on the interplay between G-quartet and G-C base pair assembly where G-quartets can eventually come to dominate (Figure 3.44).

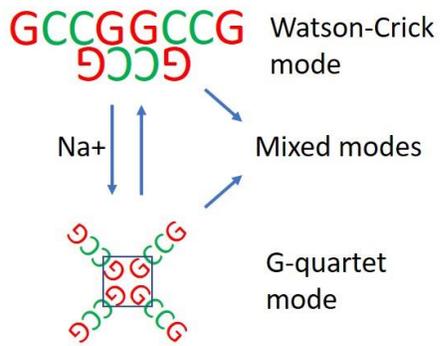


Figure 3.44: Mixed mode aggregate assembly. GCCG mesophases demonstrate unprecedented complexity because of the ability to mix assembly modes between G-quartet and Watson-Crick pairing.

## Section 4.0

### Liquid Crystals Made of Nucleosidal Triphosphates

Those first steps in the origin of life depend on the manner in which less complicated or more fundamental molecular species can be organized in such a way as to give rise to the familiar biomolecules and biological structures. We have offered nanoDNA liquid crystal assembly (Section 1.0) as a route by which the natural properties of mesophase self-assembly can give rise to a feedback loop which enables longer complementary DNA to be made from shorter fragments in absence of the conventional biological mechanisms of catalysis. The original lower limit of nanoDNA LC assembly was placed at 6 base pairs using species similar to DD and rDD (Section 2.0). This lower limit was pushed to include shorter and shorter assemblies, including GTAC and GCCG (Section 3.0), as well as incomplete work on 3mers and 2mers. The holy grail objective has been to link the processes seen with longer DNA to more fundamental molecules: in this case, extend the hypothesis to include a single-base system, which is envisioned to be as simple and stripped-down an extreme as is achievable, literally the direct DNA homolog of the chromonic liquid crystal. Logically, single nucleotide base-pairs should be able to stack in a similar manner to chromonic dye molecules.

Initial work on mononucleotides was carried out at room temperature with nucleosidal 5'-monophosphates focusing on Adenosine 5'-monophosphate (AMP) and

Thymidine 5'-monophosphate (TMP) as well as some work on the 3'-5' cyclic monophosphate versions of these, cAMP and cTMP. We avoided examining Guanosine 5'-monophosphate (GMP) with Cytosine 5'-monophosphate (CMP) upon noting that GMP can form LC by itself (Figure 3.21), complicating the ability to judge the behavior observable in mixtures. In this system, it was originally noted that TMP was soluble enough in water to achieve concentrations where LC might be observed, but the same was not true with AMP. AMP turned out to be soluble in water, but not able to achieve concentrations of hundreds of mg/mL when it was expected that any LC seen with the AMP-TMP combination would require concentrations higher than those needed for nanoDNA, which are already higher than those needed for long-strand DNA. No LC was seen with AMP-TMP in those earliest examinations.

Several years passed with this objective waiting in the wings. Results were obtained that pushed the prior limits from 6 bases to 4 bases and lower, which turned out to be intellectually important because some of these results required temperatures lower than room temp to achieve. Prior work with mononucleotides had not considered that the phase diagram for these species might not be observable at room temperature or above. Secondly, the nanoDNA phase diagram has a strong dependence on charge, particularly in sensitivity to counter ions (see Figure 2.56). In particular, the solubility of AMP is dependent on a balance between the charge present in the monophosphate, promoting solubility, with the hydrophobicity of the purine, promoting insolubility.

The biomolecule Adenosine 5'-triphosphate (ATP or dATP) (Figure 4.1 A.) is a very important part of central metabolism which serves not only as energy currency for the cell, but as the monomer source of Adenine used to polymerize DNA. This molecule became interesting to us firstly for its biological importance, but also

because --in its neutral pH form-- it contains four negative charges, which is much greater charge than AMP. Armed with the knowledge that we might need to examine the system at lower than room temperature to find interesting mesophases, we went after ultra-high concentration mixtures of dATP with Thymidine 5'-triphosphate (dTTP) (Figure 4.1 B.) that might promote LC phases.

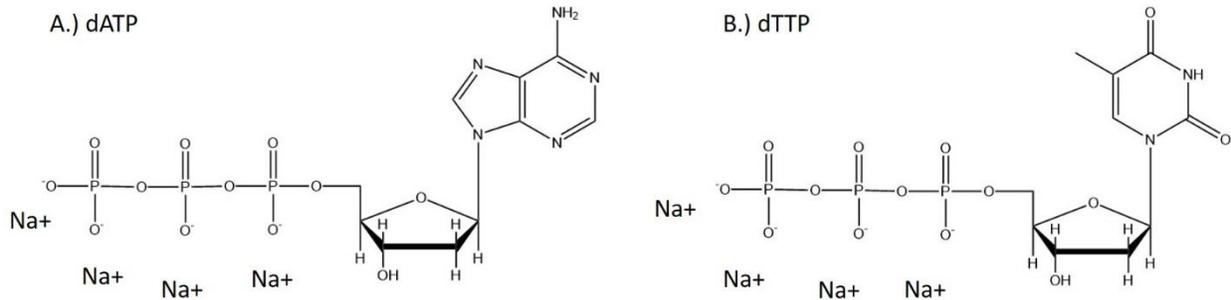


Figure 4.1: Molecular models. A.) dATP at neutral pH with sodium counter ion. B.) dTTP at neutral pH with sodium counter ion.

With dATP and dTTP placed together at a 1:1 molar ratio in water and cooled to 5° C, the dATP/dTTP mixture produced birefringent phases in an evaporative cell (see Figure 4.2). Phases do not exist for either substance alone.

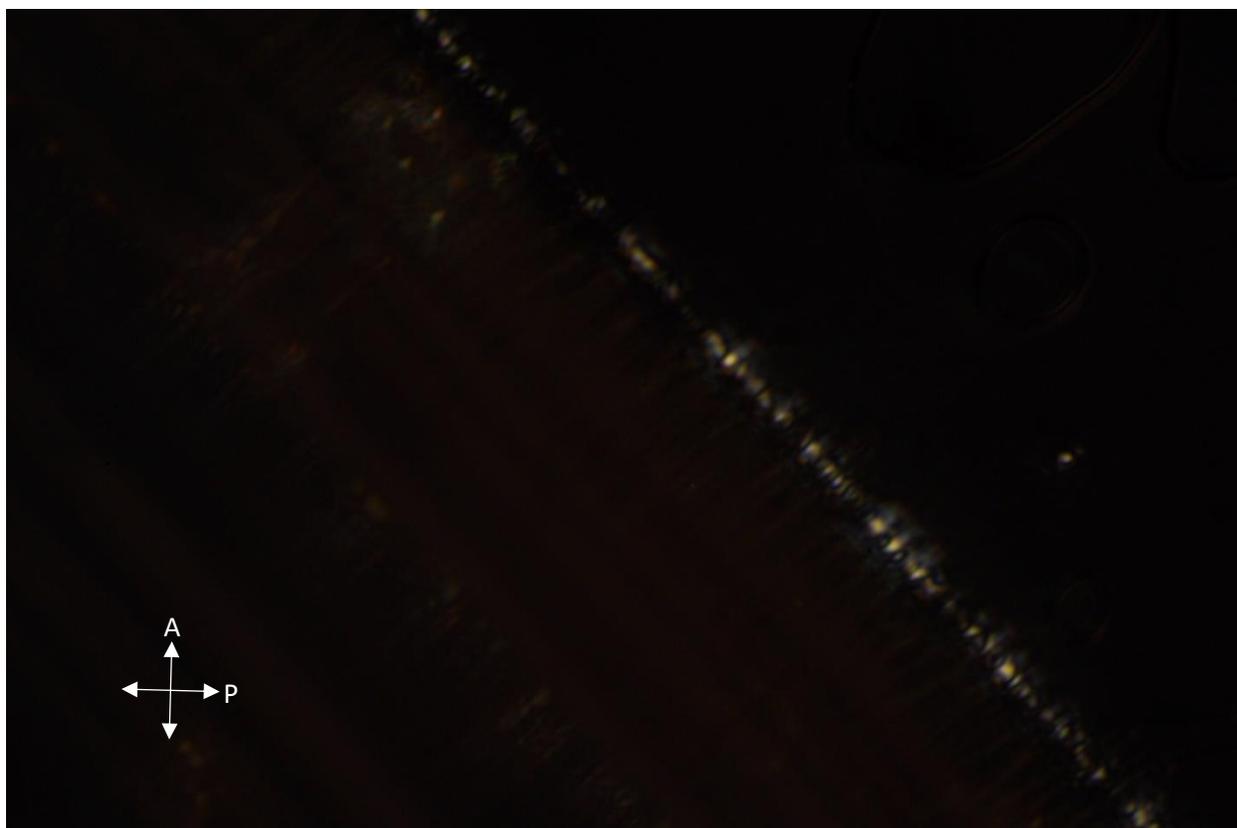


Figure 4.2: First observation of dATP/dTTP LC phases.

The phases seen here were poor, undefined textures that did not occur at higher temperatures, impossible in all regards to distinguish morphology or to identify a director field. This behavior was improved by use of an oil-sealed free surface cell (see Figure B.8) which started by drying the dATP/dTTP mixtures to a solid upon a glass substrate and then adding water back to the sample through an oil layer. When these samples were cooled to 5° C and allowed to age for five to six days, they produced significantly more distinct mesophases (Figure 4.3).

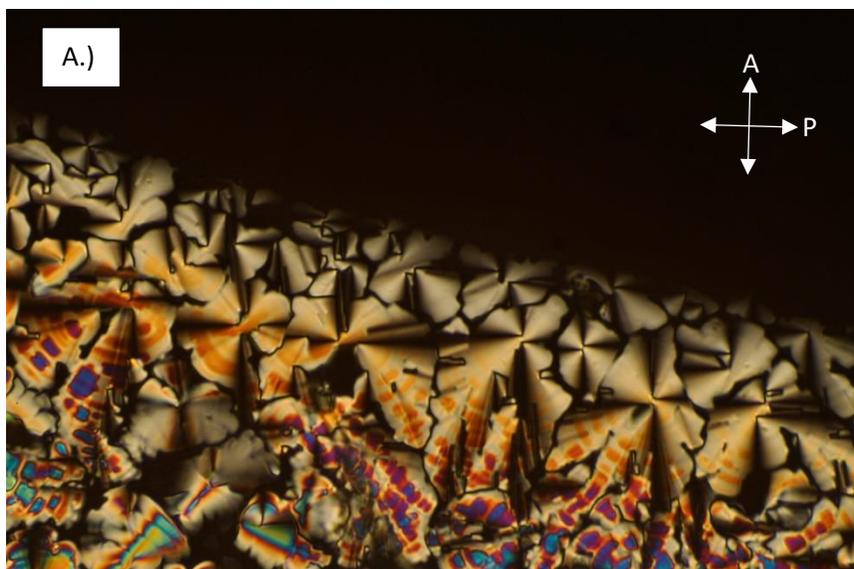
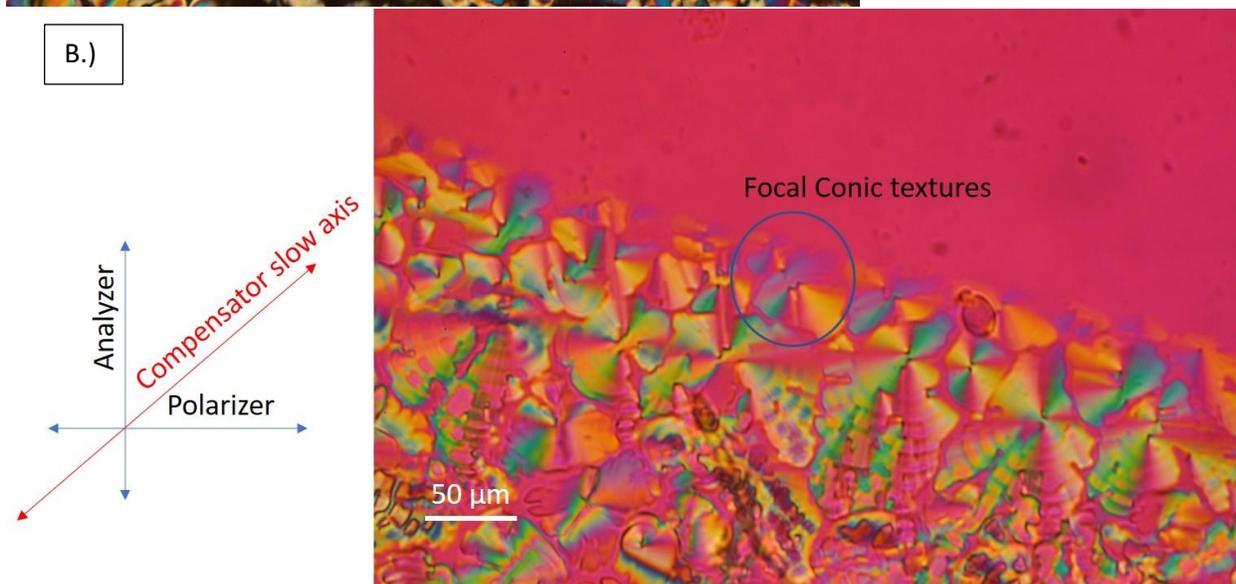


Figure 4.3: Compensator images. A.) dATP/dTTP COL focal conic textures seen at 5° C after days of incubation for coarsening. B.) Same region as seen through a birefringence compensator.



When the birefringence compensator is added with its slow axis at 45° to both the Analyzer and Polarizer and set to place the background color as purple, at the interval between the first and second order colors (see Figure B.3), the focal conic domains show red perpendicular to the compensator slow axis and blue parallel to it (see Figure 4.3 B.).

This compensation arrangement suggests that the slow axes of the molecules arranged in these focal conic structures behave very similarly to how nanoDNA aggregates behave in similar textures (see Figure 4.4).

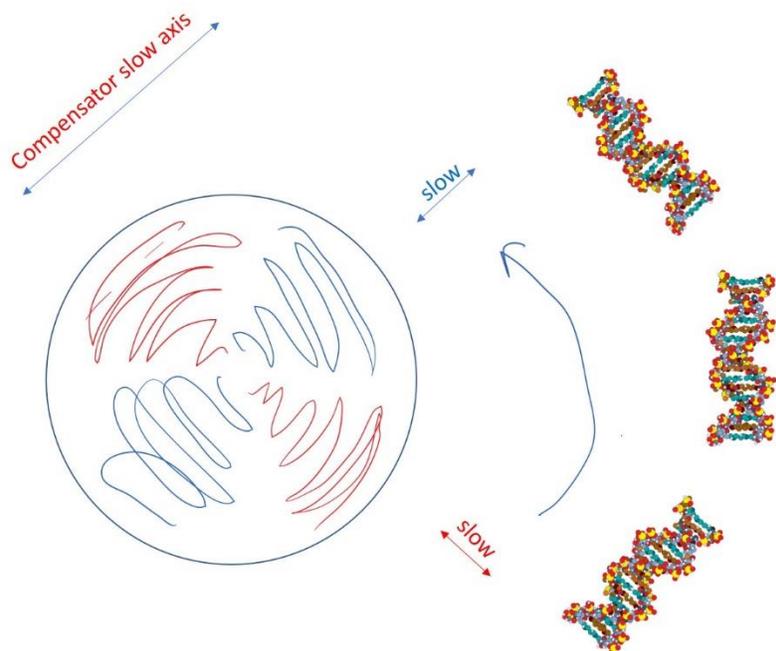


Figure 4.4: Interpretation of birefringence compensation.

In nanoDNA focal conics, base pairs are aligned parallel to the radius of the circular texture. Where the compensator is parallel to the local slow axis, the birefringence values add and push the observed color to a larger birefringence (into the blue). Where the compensator and the local slow axis are perpendicular, the birefringence is subtractive and gives a lower order color (into the red). This makes the texture negatively birefringent (slow axis perpendicular to optic axis), exactly like nanoDNA, and suggests that the dATP/dTTP 1:1 mixture stacks in a similar manner to intact sections of DNA or a chromonic dye like sunset yellow<sup>41-43</sup>. Whether the stacking was some bizarre form of high concentration nematic or a columnar phase was not initially known, though the overall similarity of the texture to nanoDNA focal conics suggested COL (see Section 4.2 for greater detail).

Given that dATP and dTTP both have considerable chemical potential energy, sufficient to crosslink and form intact DNA polymers, this series of observations connects well to the abiogenesis framework (Section 1.0). The

remainder of this section will draw heavily from an academic paper produced examining these materials to this end.

#### 4.1 Additional Materials and Methods Used for Deoxynucleosidal Triphosphate (dNTP) Experimentation

Liquid crystalline phases of the dNTPs (dATP, dCTP, dGTP, dTTP) were produced from PCR grade DNA nucleoside triphosphates acquired in 250 uL x 100 mM predissolved fractions from Qiagen (cat# 201913). The dNTPs detailed above had been pre-adjusted prior to acquisition by NaOH treatment to render their pH effectively neutral, placing approximately three sodium cations into solution for each dNTP anion. These samples contained no additional spectator ions because the endogenous triphosphate pKa values provide substantial buffering capacity and prevent the solution from varying in pH. This is in contrast to nanoDNA, which is frequently contaminated with spectator ions and salts.

Determination of concentration in dNTP samples used the same methods put forward in Section B.2.2. Equation B.28 was used where  $G_{con}$  is taken to be zero,  $\rho_{dna}$  becomes instead  $\rho_{dNTP}$ , giving equation 4.1.

$$c = \left( \frac{1}{\rho_{dNTP}} + \frac{k_1}{\rho_{Na}} + G_1 \left( \frac{k_2}{\rho_w} + G_2 \right) \right)^{-1} \quad (4.1)$$

Here, constant  $k_2$  is measured to determine excess water present in the sample prior to usage in experimentation. Unlike samples of nanoDNA, which were invariably prepared with lyophilization to remove all water, dNTP samples were only treated with rough vacuum to dry them, leaving some excess of water. The intent was to try to minimize the preparation time given the chemical instability of the molecules in the sample.  $\rho_{dNTP}$  was taken to be 1403 mg/mL<sup>44</sup>.

Samples examined for this work were prepared in cells as detailed in Section B.0, with particular emphasis on the free surface cell (Section B.1.5) as already mentioned in Section 4.0. Samples typically were prepared by mixing 100 mM concentration fractions of dNTP at a desired ratio, then drying them to surface with a rough vacuum before adding water back to these materials in order to attain a desired concentration.

The methods put forward in Section 2.3.1 were also used to determine the stacking parity of the dATP/dTTP mixtures in X-ray diffraction on the assumption that they are locked in a hexagonal COL lattice much like nanoDNA. Theory lines were generated for duplex, triplex and quadruplex structures. Using equation 2.11, molecular weight  $M$  is taken to be the average molecular weight  $\langle M \rangle$  between dATP (491.2 g/mol) and dTTP (482.2 g/mol), where  $\langle M \rangle = 486.7$  g/mol.  $N_p$  can be varied to give  $m_u$  for duplex ( $1.602 \times 10^{-21}$  g), triplex ( $2.403 \times 10^{-21}$  g) or quadruplex ( $3.204 \times 10^{-21}$  g). This gives theoretical slope  $K$  values of 0.00009646 mL/(mg Å<sup>2</sup>) for duplex, 0.00006431 mL/(mg Å<sup>2</sup>) for triplex and 0.00004823 mL/(mg Å<sup>2</sup>) for quadruplex using equation 2.13 (see also Figure 4.10).

Data collection for X-ray diffraction was also subtly different from previous experiences with nanoDNA and can be considered one of the great coups of the experimental work presented in this thesis. As with nanoDNA, stable samples for X-ray diffraction were obtained using flame sealed capillaries. Unlike with nanoDNA, because birefringent phases occur at below room temperature and sometimes required long incubation periods to establish, they were transported to the synchrotron as ISO specimens and refrigerated on site to produce suitable LC phases, then observed cold. This proved especially challenging because the available experimental apparatuses on site were predominantly focused on attaining extreme conditions, much more extreme than the 0° C required for this experimentation: the

conditions needed for examination of the dNTP phases existed in a capability gap. X-ray scattering data for liquid crystalline dATP/dTTP mixtures were obtained by improvisation using an ice-water slurry produced from ice obtained from a hotel ice machine, then observed at 0° to 5° C on a hot stage with cold water circulation. Samples containing dGTP and dCTP were significantly easier since some of these had birefringent phases at room temperature at the very high concentrations examined.

#### 4.2 Observations of dATP/dTTP

Aqueous dATP/dTTP mixtures form birefringent phases at sufficiently high concentrations and sufficiently low temperatures, whereas solutions of dATP or dTTP alone do not. If aqueous dATP/dTTP (1:1) mixtures are submitted to evaporation, birefringence appears at temperatures of 5°C particularly in regions of the cell where water is departing most rapidly and the highest dNTP concentrations are achieved. In a strictly evaporative cell, the structures of the birefringent phases were not distinct enough to judge the texture or determine the order of the phase. After samples dried, birefringence vanished, leaving only a solid material. If water is added back to the dried substance, birefringence can reappear if the sample is held at sufficiently low temperature. Under these conditions, the textures were quite solid, but too disorganized to judge a type. If water is re-added to a dried sample according to the Free Surface Contact Cell preparation of Section B.1.5 for samples topped with heavy mineral oil, the development time is extended by dampening the influence of evaporation.

Therefore, during PLM studies of aqueous dNTPs, the samples were studied in cells in which controlled amounts of water and vacuum-dried dNTP were brought

into physical contact at a free surface under an oil seal (Figure B.8) and allowed to develop (Figure 4.5).

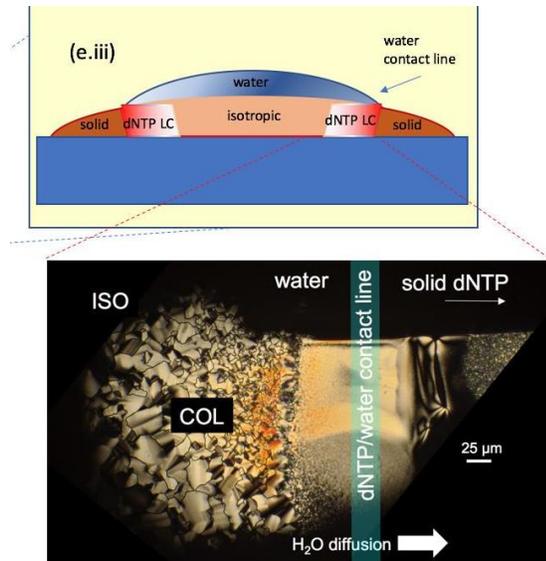


Figure 4.5: Development of dATP/dTTP mixture in free-surface cell. Adapted from Figure B.8 e.iii). Water diffusing from the water droplet into solid dNTP forms a contact line by action of diffusion, allowing solid dNTP to migrate out into the water, forming a gradient of concentrations from ISO to COL and higher order.

In this configuration, we could also measure the weight of the dNTP substance and approximate the concentration by regulating the amount of water added. Water was added as a droplet by micropipettor to dried dNTP beneath the oil layer and, for dATP/dTTP (1:1), allowed to sit at 5°C for a protracted time. The target average dNTP anion/water concentrations when mixed,  $c$ (mg/ml), were in the range  $400 \text{ mg/ml} < c < 1000 \text{ mg/ml}$ , which can be compared to the low-hydration limit of dATP/dTTP (1:1) of  $c_{lim} = 1,350 \text{ mg/ml}$  (equation B.28,  $G_2 = G_{con} = 0$ ). At target average concentrations  $c > 700 \text{ mg/ml}$ , this produces a ring of birefringence that gradually expands across the face of the dried layer where the water droplet contacts the dried substance. This ring effect is seen for samples up to  $c > 1100 \text{ mg/ml}$ . At  $c = 900 \text{ mg/ml}$ , the ring expands slowly enough over the course of four days that the last fringes of the sample develop into predominantly disorganized, defect-rich birefringent fan structures that bear strong resemblance to COL phase LC. Some of the fans form complete focal conics that exhibit negative birefringence (see Figures 4.3 and 4.4), like oligomeric NA LCs (see Figure 4.6 (e),(f)). After

several days of diffusive mixing, the samples with target average  $c \sim 1000$  mg/ml and  $T = 5$  °C, in the band running parallel to the dNTP contact lines where the dNTP concentration was largest and the phase fluid, spawned the defect-rich birefringent fan textures, shown in Figure 4.6 (h).

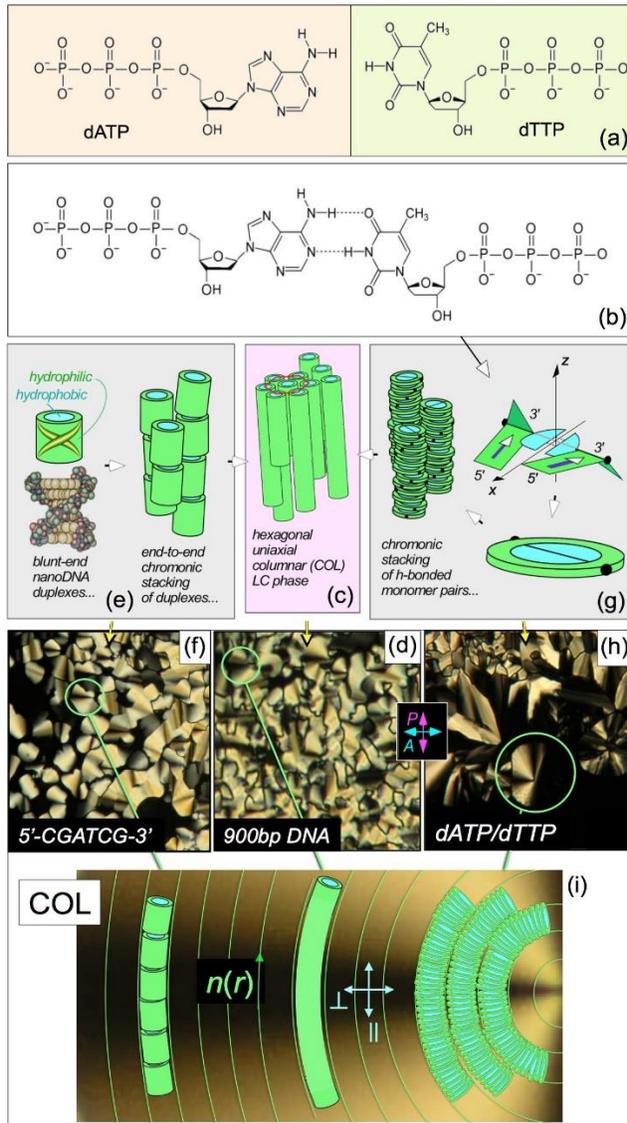


Figure 4.6: (a) Chemical structures of dATP and dTTP. (b) Watson-Crick (W-C) hydrogen bonded dATP/dTTP dimer showing its hydrophobic and hydrophilic domains. The hydrogen bonding reduces the solubility of the base pair, promoting the chromonic stacking of bases. (c) Uniaxial hexagonal columnar (COL) LC phase of polymeric B-DNA. (e) COL LC phase formed by the self-assembly of anisotropic aggregates of duplex paired molecules. Such “chromonic” aggregation is driven by base pairing and stacking to shield flat hydrophobic molecular components from contact with water. (g) COL LC ordering of duplex NTP stacks, sketched for W-C duplexing, as indicated by the experiments. The sketch of the base pair indicates its chiral symmetry. The sub-columns of single bases have opposite  $5' \rightarrow 3'$  directions, as in W-C NAs. The black dots indicate the helical stacking in the columns, forced by the bulky triphosphate groups. Polarized transmission optical microscopic textures of the characteristic conformal domains of the COL liquid crystal phase of: (d) long (900bp) B-DNA (concentration,  $c \sim 500$  mg/ml); (f) the self-complementary DNA

hexamer  $5'$ -CGACG- $3'$  ( $c \sim 800$  mg/ml); and (h) the dATP/dTTP mixture ( $c \sim 900$  mg/ml). (i) Schematic illustration of a conformal domain of the COL phase showing the local geometry of columns of stacks of oligomeric NA, long NA, and NTP duplexed bases. The column packing permits bend deformation of the director field  $\mathbf{n}(r)$ , the local column axis direction, but not splay, the key feature of columnar ordering that produces its characteristic conformal domains. The COL birefringence,  $\Delta n = n_{||} - n_{\perp}$  is negative because of the larger optical polarizability parallel to the planes of the bases.

Taking the LC director field  $\mathbf{n}(\mathbf{r})$ , indicated by the green arrows in Figure 4.6 (i), to be along the chain axis, the principal characteristic structural feature in all of the textures in Figure 4.6 (d),(f) and (h), is that they are conformal: only bend deformation of  $\mathbf{n}(\mathbf{r})$  is allowed (Figure 4.6 (i)). Thus, the absence of splay of  $\mathbf{n}(\mathbf{r})$  is direct evidence for the constraint imposed on molecular orientation by the splay-free packing of molecular stacks required in a columnar phase, as sketched in Figure 4.6 (i). Additionally, the fact that  $\mathbf{n}(\mathbf{r})$  in the textures is readily bent shows that the phase is fluid and that there is free sliding of the columns parallel to their axes, i.e. that the phase is a liquid crystal (green lines indicating  $\mathbf{n}(\mathbf{r})$  in Figure 4.6 (i)). The absence of bands parallel to  $\mathbf{n}(\mathbf{r})$  of differing birefringence indicates that the phase is uniaxial about  $\mathbf{n}(\mathbf{r})$ , *i.e.* is a uniaxial columnar LC phase (COL), which is typically the first columnar phase to appear in NA oligomer solutions, from either the NEM or ISO as  $c$  is increased. Measurement of the sign of the visible light birefringence shows that it is negative (see Figure 4.3): the effective refractive index is largest where the optical polarization is parallel to the smooth dark brushes in the texture, the same as for DNA (Figure 4.6 (i)). This sign is due to the larger polarizability of the NA bases for light polarized parallel to their planes and thus normal to the local DNA duplex axis.

As with all of the duplex oligomeric NA columnar phases we have observed to date, the dATP/dTTP (1:1) LC phase melts into the isotropic (ISO) phase via a first order transition characterized by a temperature range of phase coexistence of the LC and ISO domains. Since the limits of this range increase in  $T$  with increasing concentration, they are difficult to measure in contact cells because of the inherent gradients in  $c$ . In the sealed capillaries, LC formation is marked by the appearance of COL droplets, as shown in Figure 4.7. The LC phases readily melt on heating,

but reformation on cooling is slower, sometimes taking hours for domains to nucleate.

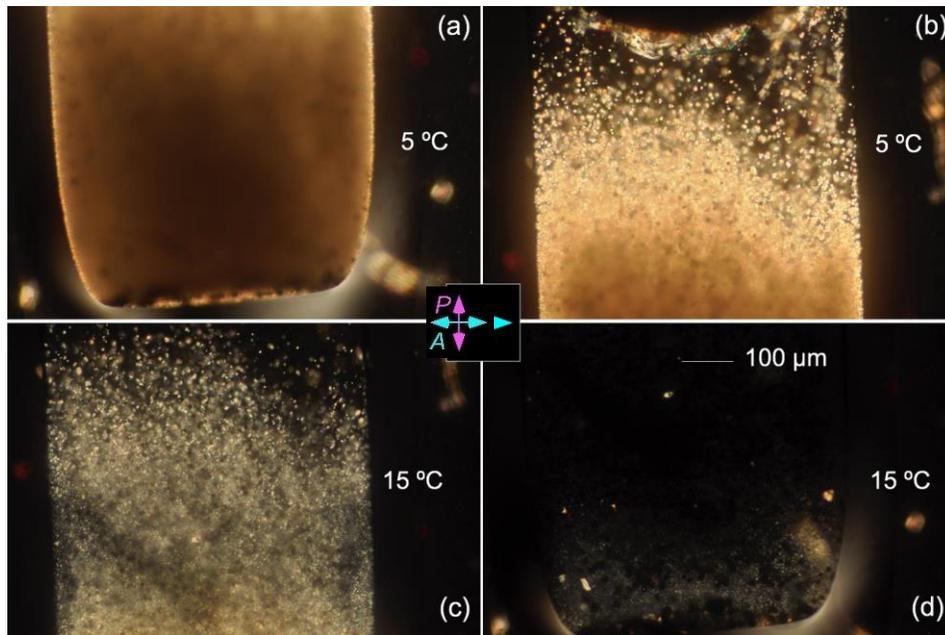


Figure 4.7: Flame Sealed Capillary cell containing dATP/dTTP (1:1) mixture at  $c = 810$  mg/ml,  $\sim 60\%$  of the limiting NTP concentration,  $c_{lim} = 1350$  mg/ml. (a) Bottom end of capillary held at  $5^\circ\text{C}$  for 24 hours, COL phase. (b) Meniscus of capillary held at  $5^\circ\text{C}$  for 24 hours, ISO portion of phase coexistence is visible. (c) Capillary temperature increased to  $15^\circ\text{C}$ , birefringence of phases near the meniscus begins to vanish as phases melt to isotropic. (d) Bottom end of capillary at  $15^\circ\text{C}$  as domains complete their transition to isotropic phase. Phase transitions occur nearly uniformly across the cell, implying minimal concentration heterogeneity.

To check whether dNTP mesophases are dependent on A-T or G-C base pairing, trials were run of the various different possible dNTP mixtures to see LC can occur from a dNTP by itself, or in a mixture with a non-Watson-Crick (or Hoogsteen) base pair partner (see Figure 4.8).

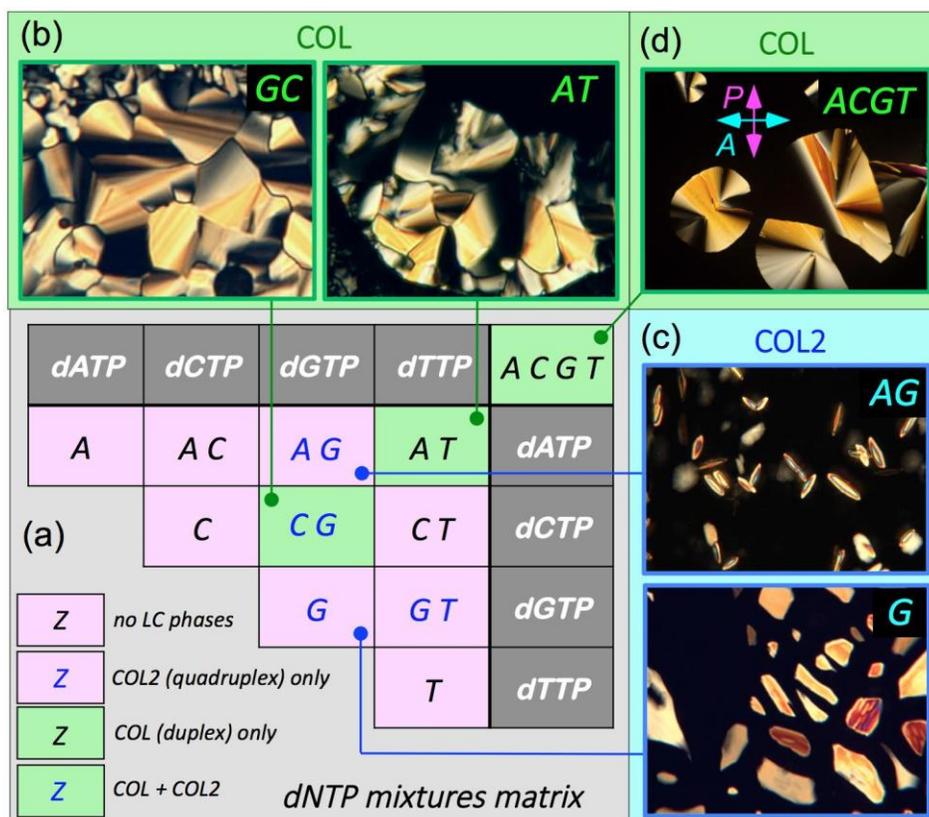


Figure 4.8: (a) ACGT Matrix summarizing the search for LC phases in aqueous solutions of dNTPs and their mixtures, showing a significant dependence on choice of base. The matrix colors indicate the most ordered phase to appear for each base pair combination for  $T > -20^\circ\text{C}$ . The COL phase appears in NTP solutions at high concentration only for pairs exhibiting W-C complementarity. Macroscopic phase ordering was evaluated by PTOM and internal structure by XRD. The general pattern of COL ordering is the self-assembly of columnar chromonic stacks of H-bonded bases and the packing of the columns into two-dimensionally hexagonally ordered arrays, where the distinction between duplex and quadruplex stacking is made using XRD (Figures 4.9 and 4.10). (a,b) Green shading indicates the binary solutions dATP/dTTP (AT) and dCTP/dGTP (CG) which exhibited uniaxial hexagonal duplex columnar (COL) LC phases of chromonic stacks of base pairs. (a,b) Red shading indicates the binary solutions AC, AG, CT, and GT, and the single dNTP solutions A, C, G, and T, none of which exhibited the COL phase.

The four-component mixtures, dATP/dCTP/dGTP/dTTP (ACGT), also exhibited COL domains. (c) Blue font indicates solutions containing dGTP, including G only, that exhibited, at higher concentrations of dGTP, domains of uniform internal orientation of a higher-ordered columnar phase of chromonic stacks of H-bonded base quadruplexes (the COL2 label in the Figure is not known to be the same as the mesophases examined in Sections 2.0 or 6.0). The ACGT mixtures exhibited only the COL phase at the

concentrations studied (see Figure 4.13), probably due to CG phase separation and ordering.

The appearance of duplex COL phases only for the binary mixtures AT and CG, which are the only mixtures of polymeric NA columns in solution to exhibit complementary duplexing, is evidence for the Watson-Crick or Hoogsteen base-pairing motif in the COL LC phase.

Of special note: nanoDNA structures seem to be of B-form DNA duplexes given the appearance of the B-form duplex periodicity in higher order COL2 (both for rDD and GTAC as seen in Section 2.0 and 6.1). These are assumed to predominantly persist as Watson-Crick base paired forms, but Hoogsteen pairing is transiently possible in long DNA or short DNA<sup>36</sup>. The existence of Hoogsteen pairing does not preclude abiogenesis related complementarity arguments following from nanoDNA templating and stacking enhancement as a side effect of mesophase formation –if the duplex has Hoogsteen pairs, it is still complementary pairing. The same is true here; while it is not immediately clear whether the structure is A-form or B-form or even truly duplex, LC dependence is on expected base pair partners, suggesting duplexing, and which are complementary regardless of whether stacking is Watson-Crick or Hoogsteen.

### 4.3 dNTP X-ray Data and Analysis

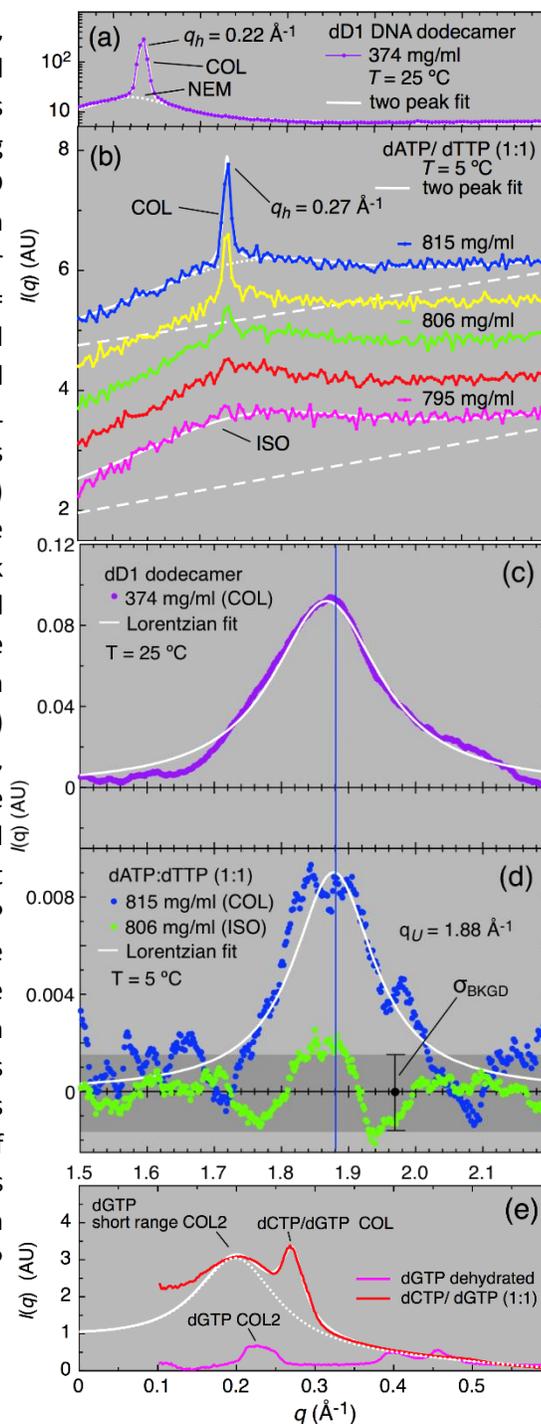
As noted in Section 3.2, columnar LC ordering can be expected for the dGTP-containing mixtures as a result of the H-bonding of the G's into planar quadruplexes and their resulting strong tendency to make chromonic columnar stacks, and is manifested in the dNTPs as the ambiguous higher order columnar phase. However, observation of the COL LC domains only for the base-pairing AT and CG solutions among all of the monomer and binary dNTP mixtures in Figure 4.8 suggests a role for duplex pairing and stacking in the formation of the COL LC

phase in the NTPs. This role was quantitatively established by X-ray diffraction which, by characterizing the packing of columns and stacking of molecules in samples with known  $c$ , can be used to unambiguously determine the multiplicity of bases at each level in the chromonic stacks of a columnar LC phase according to equation 2.7 from Section 2.3.1, including theory lines suggested in Section 4.1. X-ray study of the NTP solutions of Figure 4.8 were carried out on randomly oriented LC multidomains in Flame Sealed Capillaries vs. temperature and concentrations, with results summarized in Figures 4.9 and 4.10. Included in these figures for comparison purposes are results from capillaries of similarly prepared solutions of rDD (from Section 2.0). At sufficiently high  $c$ , rDD forms duplexes which stack by association of the duplex blunt ends into linear aggregates, which then phase separate into hexagonal columnar (COL) and nematic (NEM) LC phases. Figure 4.9 shows the coexistence of broad and narrow scattering peaks in a wavevector scan for  $c = 374$  mg/ml, near where NEM/COL phase coexistence can be expected in rDD, and at  $q \sim 0.22 \text{ \AA}^{-1}$ , the wavevector range expected for DNA hexagonal columnar ordering. This latter expectation can be quantified with the help of Figure 4.10, which plots the square of the measured position of the sharp peak,  $q_h^2$  as a function of  $c$ . For the case that this sharp peak is from an ordered NA hexagonal columnar phase of stacked bases, this function is linear and given by equation 2.9. Base stacking period is confirmed from the data in Figure 4.9 (c), the high- $q$  region of the scans in Figure 4.9 (a). These  $q_h^2$  vs.  $c$  lines terminate where the maximum possible concentration equals the internal mass density of the base unit, 1,687 mg/ml for DNA, and 1,403 mg/ml for the NTPs. The indicated limiting (dehydrated),  $c_{lim} = 1,656$  mg/ml for DNA, and  $c_{lim} = 1,356$  mg/ml for the dNTPs, are where all of the water has been removed but the inter-columnar space contains the remaining sodium counter ions.

The columnar phases of solutions quantitatively diluted from a known mass of [DNA anions + sodium counterions only] will, in the COL phase, produce  $q_h^2$  vs.  $c$  on the blue line in Figure 4.10. As mentioned in Section 2.3.2, the rDD measurements ( $\square$ ) fall slightly below this line, due to remnant NaCl contamination in the starting sample. In such a case the net volume of DNA anions is lower than as weighed, and the spacing between DNA columns therefore larger, resulting in a peak at smaller  $q_h$ . So, measured peak positions can fall below the blue line because of extra ions in solution, but can never be above the line for chemically correct DNA, facts of key importance if the x-ray scattering peak position is being used to assess the nature of the unit at each base level. For example, a COL phase of DNA triplexes will have a  $q_h^2/c$  slope roughly 2/3 smaller than the duplex value in Figure 4.10. The blue circle data, being only a few percent below the duplex DNA line, eliminate the possibility that the rDD is making a phase of triplexes.

The corresponding XRD peaks for the dATP/dTTP (1:1) mixtures are shown in Figure 4.9 (b),(d) and (e), with Figure 4.9 (b) indicating the phase coexistence of ISO and COL phases with the sample cooled to 5° C at different concentrations. The ISO phase exhibits a broad peak at  $q \sim 0.26 \text{ \AA}^{-1}$ , which is absent in solutions of only dATP or dTTP, that provides evidence for short-ranged correlation of pairs in the ISO phase, and which is also found in the ISO phase of chromonic dyes<sup>45</sup>. Figure 4.9 (d) and (e) show that the COL LC base-stacking peak position  $q_U = 1.88 \text{ \AA}^{-1}$  is the same in the dATP/dTTP monomers as in the DNA oligomers. The resulting  $q_h^2/c$  model for the dNTPs is plotted in Figure 4.10 for dNTP base unit stackings of duplexes, triplexes and quadraplexes having the average dNTP unit mass.

Figure 4.9: X-ray structure factors,  $I(q)$ , showing the periodicities in the COL liquid crystal phase due to the packing of columns of chromonic aggregates and base stacking within the columns. (a) The rDD dodecamer, included for comparison purposes, forms chromonic columnar stacks of blunt-end duplexes, which, at  $c = 374$  mg/ml order into coexisting uniaxial hexagonal columnar (COL) and uniaxial nematic (NEM) phases. (b) Coexisting COL and isotropic (ISO) phases versus concentration,  $c$ , in the dATP/dTTP (1:1) mixture. These data are taken along the black box in Figure 4.11. The diffuse peak in the ISO phase indicates short-ranged aggregation and packing of columns. The ISO peak is similar to that found in chromonic phases of organic dyes<sup>45</sup>. (c-e) X-ray structure factors,  $I(q)$ , at large  $q$ , showing the base-stacking periodicity in the chromonic aggregates forming COL liquid crystal phases. The base stacking peak is at  $q_U = 1.88 \text{ \AA}^{-1}$  in both rDD and the NTP mixtures, corresponding to  $L_U = 3.34 \text{ \AA}$ . The base-stacking peak is not observable in the ISO phase (green). (e) X-ray scattering from dehydrated dNTPs. Plotted peak positions in Figure 4.10 show that dehydrated dGTP's exhibit hexagonal uniaxial COL2 ordering of G-quadruplex columns, whereas dCTP/dGTP (1:1) mixtures exhibit both dCTP/dGTP duplex COL and dGTP quadruplex COL2 peaks.



XRD measurements of  $q_h$  were made for the dNTP mixtures that exhibited LC ordering.  $q_h^2$  vs.  $c$  for the dATP/dTTP (1:1) ( $\circ$ ) and dCTP/dGTP (1:1) ( $\diamond$ ) mixtures fall close to the dNTP model duplex line, indicating that the dATP/dTTP and dCTP/dGTP base units are duplexes, and are mixtures with only NTP anions,

Na<sup>+</sup> counterions and water. These data, corresponding to the black boxes in the phase diagrams of Figures 4.11 and 4.12 respectively, are taken close to the COL/ISO phase transition where equilibration of the 2D hexagonal COL lattice can be achieved. At higher concentrations the ISO phase becomes more glassy and the kinetic barriers to nucleation of the COL phase higher, leading to quenching of the ISO phase.

Among the four solo-base dNTP aqueous mixtures, only dGTP showed XRD evidence of LC ordering, exhibiting a G-quadruplex columnar phase, even when completely dehydrated (Figure 4.9f). Dehydrated mixtures of dCTP/dGTP (1:1) show two distinct diffuse XRD peaks (Figure 4.9f), having peak positions with  $q_b^2$  values in the ratio of  $\sim 0.5$ , than expected for coexisting duplex [dCTP/dGTP (1:1)] and quadruplex [dGTP] phases of the same  $c$ . The formation of G-quadruplex columns in a 1:1 mixture forces dCTP into the solvent phase, so that the effective concentration of both phases decreases and their column spacing increases, with the percent,  $P_G$ , of G's in G-quadruplex stacks. For  $P_G = 100\%$ , i.e., all of the G in quadruplexes and all of the C in between, we would have  $q_b^2 = 0.033\text{\AA}^{-2}$ . The actual G-quadruplex peak position in Figure 4.10,  $q_b^2 = 0.04\text{\AA}^{-2}$  is close to this, indicating that most ( $P_G \sim 70\%$ ) of the dGTP is in G quadruplexes in the dehydrated sample.

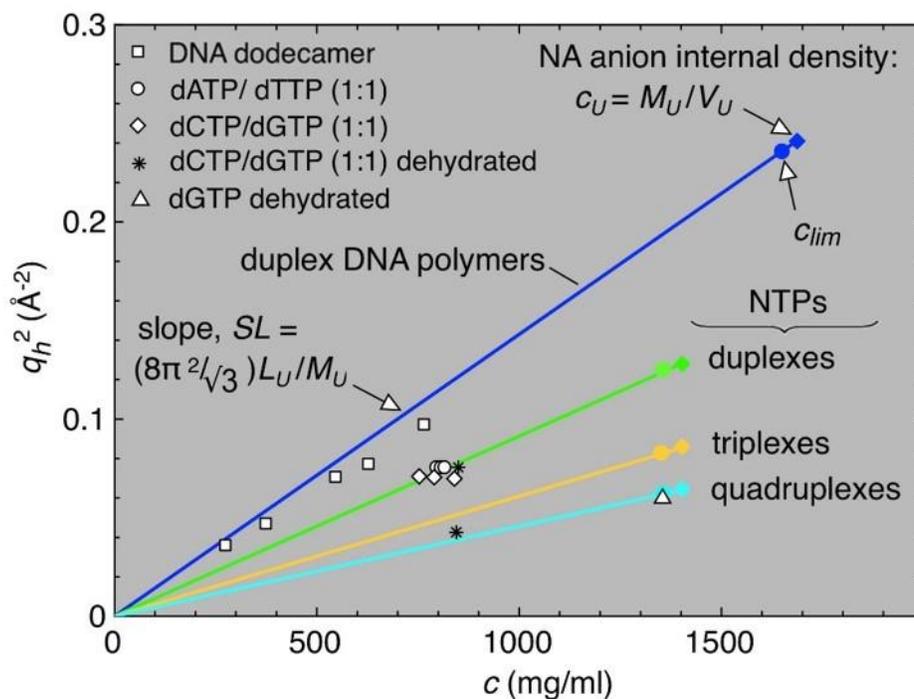


Figure 4.10: Plots of  $q_h^2$  versus  $c$ , where  $q_h$  is the fundamental columnar packing peak position for uniaxial hexagonal columnar ordering of rDD and dNTPs, and where  $c$  is the NA anion concentration. The solid lines are from equation 2.10 with base-stacking periodicity  $L_U = 3.34\text{\AA}$ , from Figure 4.9. These lines terminate at  $c_U = M_U/V_U$ , the internal density of the NA anion base repeat unit,  $U$ . Data of  $q_h^2$  versus  $c$  for mixtures having only NA ions and native Na counterions must lie on such lines, enabling x-ray based distinction of duplex, triplex, and quadruplex dNTP COL phases. Chemically correct DNA cannot give data above the line but added salt in the starting dehydrated sample can give data below the line, as for rDD ( $\square$ ). The data for the dATP/dTTP ( $\circ$ ) and dCTP/dTTP ( $\diamond$ ), collected in the black outline box regions in Figures 4.11 and 4.12, respectively, indicate that the mixtures form COL phases of duplex aggregates. The limiting concentration,  $c_{lim}$ , given in equation B.29 (with  $G_2=0$ ) is the concentration of dehydrated samples, the largest achievable, and slightly smaller than  $c_U$  because of the  $\text{Na}^+$  counterions remaining in the solution partition. Dehydrated dGTP is fully quadruplex with COL short ranged order ( $\Delta$ ), whereas dehydrated dCTP/dGTP (1:1) forms a mixture of duplex and quadruplex COL phase, with the unpaired dCTP acting as solvent, providing spacing between the columns (\*).

#### 4.4 dNTP Phase Diagrams

X-ray data was combined with extensive observations of flame sealed capillaries under PLM to determine phase diagrams for the base pair combinations of dNTP mixtures, 1:1 dATP/dTTP (Figure 4.11) and 1:1 dGTP/dCTP (Figure 4.12).

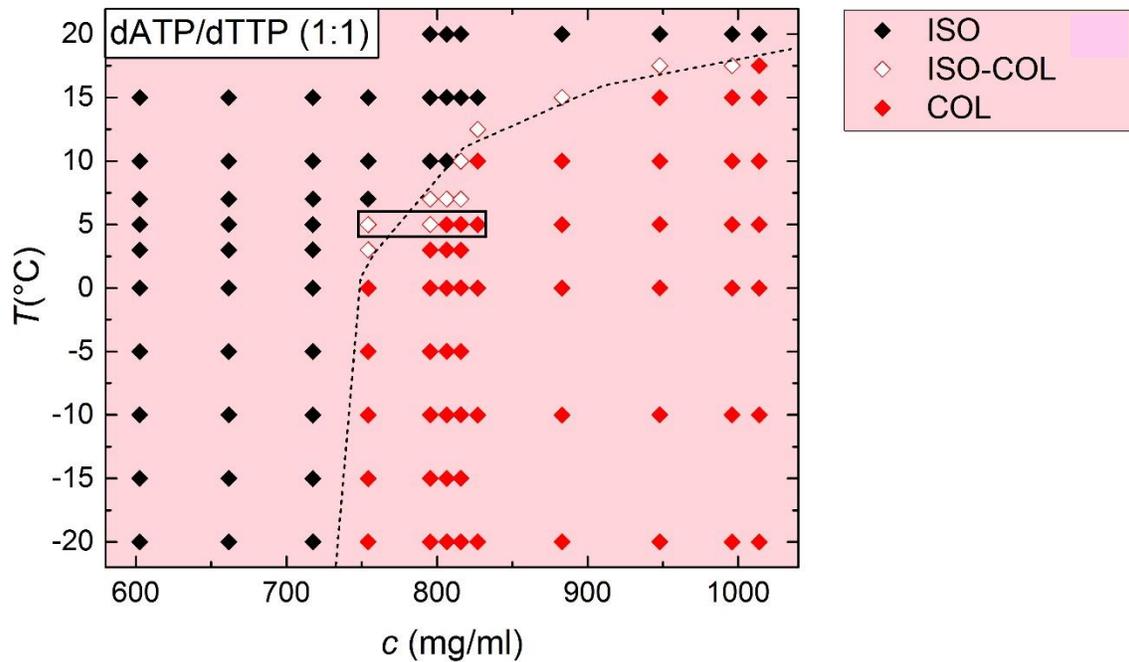


Figure 4.11: dATP/dTTP mixture phase diagram in  $T$  versus  $c$ . Three phase combinations are shown, including ISO (Black Diamonds), ISO-COL (White Diamonds) and COL (Red Diamonds). Dotted line roughly follows the COL-ISO phase boundary. Black box depicts data points where X-ray diffraction was collected (Figures 4.9 and 4.10).

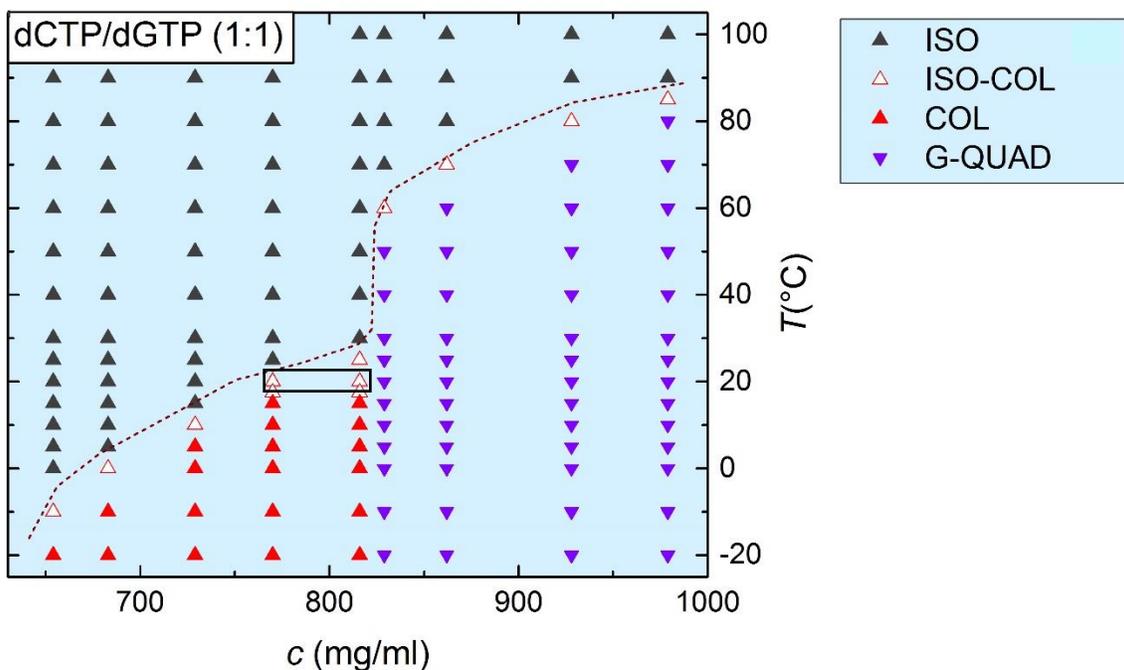


Figure 4.12: dGTP/CTP mixture phase diagram in  $T$  versus  $c$ . Four phase combinations are shown, including ISO (Black Triangles), ISO-COL (White Triangles), COL (Red Triangles) and apparent G-quadruplexes (G-QUAD, Purple Triangles). Dotted line roughly follows the COL-ISO phase boundary. Black box depicts data points where X-ray diffraction was collected (Figures 4.9 and 4.10).

Both dNTP combinations show phase diagrams generally at lower temperatures than observed nanoDNA. The one possible exception to this is the occurrence of temperature stability in dGTP/dCTP mixtures that appears to correlate with the presence of G-quartet structures, which have already been well documented as highly stabilizing in nanoDNA mixtures (Section 3.2). One additional behavior in dNTP mesophases which is not well touched by the phase diagrams here is the appearances of thermostability with longer incubations; dNTP mixtures sometimes showed a proclivity toward becoming more stable and resilient against melting if they sat for longer times at temperatures where they had birefringent phases. In particular, mixtures containing all four dNTPs showed significant stabilization over time which elevated the melting point from between 0°

and 10° C to as high as 30° C, though this particular effect may be due to slow phase separation of dATP/dTTP base pairs from dGTP/dCTP base pairs. Given the difficulty of reproducing these longer-term results, clear trends have not yet been uncovered. It is also noteworthy that several capillaries incubated for multiple months showed the appearance of birefringent objects that were super stable to temperature ramping. These objects do not melt with temperatures elevated to 90° C, but their molecular constitution has not been determined. Such variation in behavior could be explained by the diminished chemical stability of dNTPs relative to DNA since no homologous behavior has been seen with nanoDNA. There is the outside chance that this behavior is due to dNTP polymerization within these LC phases over time, but mass spectrometry thus far probing of these samples has not supported this hypothesis.

A summary phase diagram of dNTP behavior can be collected in comparison to currently explored nanoDNA (see Figure 4.13).

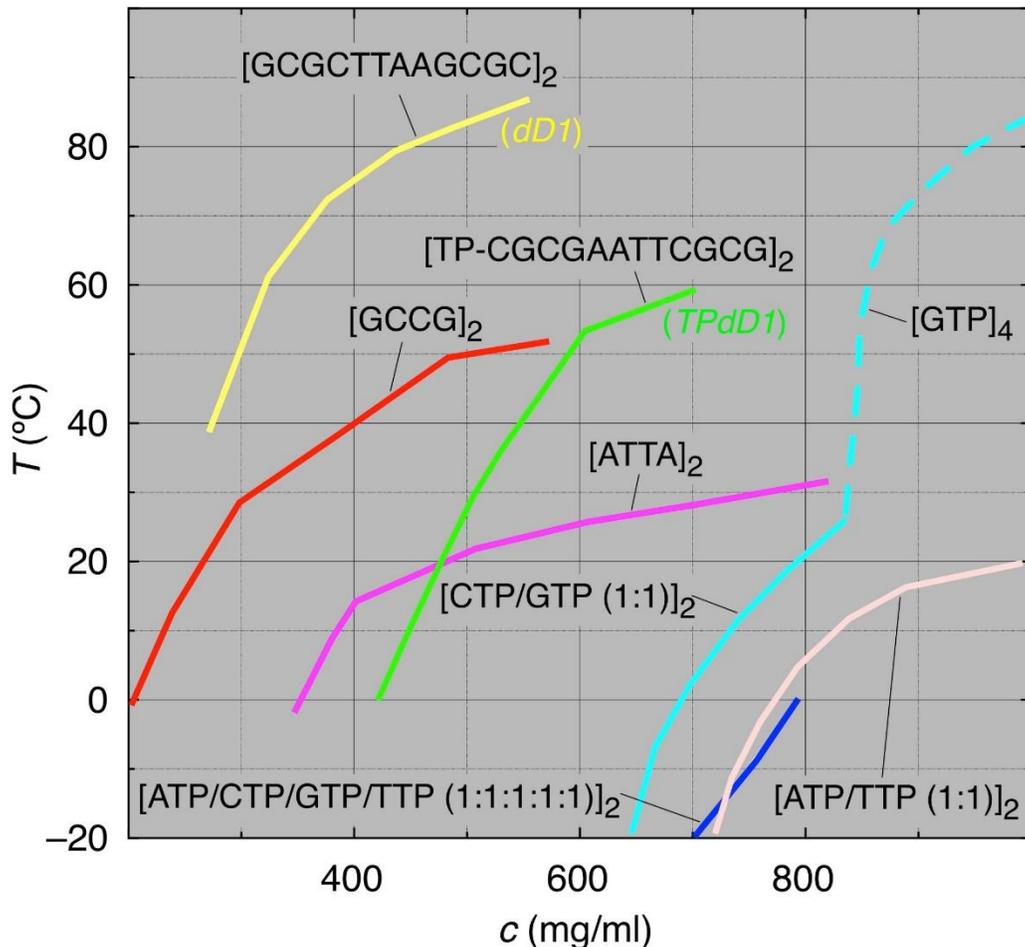


Figure 4.13: Compilation of relevant DNA liquid crystal phase diagrams, determined by PTOM and XRD, showing the high-temperature extent of LC ordering in: blunt-end self-complementary duplexes (dD1, called also rDD); triphosphate-terminated self-complementary duplexes (TPdD1); the shortest duplex forming NA oligomers previously reported (dATTA, dGCCG)<sup>34</sup>; and several NTP solutions. The blue line is the ISO-COL transition temperature in the dATP/dCTP/dGTP/dTTP (1:1:1:1) mixture. These LC phases are all from columns of duplex aggregates, except for the dCTP/dGTP (1:1) mixtures: at  $c \sim 800$  mg/ml where LC ordering first appears, the COL phase is an assembly of duplex C-G columns, whereas for  $c > 850$  mg/ml coexisting G-quadruplex columns appear, exhibiting stronger higher order columnar LC phase thermal stability.

The general trends evident in these data are: (i) Larger H-bonding and stacking free energy makes the COL phase more stable (CG vs. AT). Solutions of single bases do not appear because they do not exhibit LC ordering. (ii) Extending the oligomer length by native NA polymerization strongly stabilizes the COL LC

phase. (iii) Triphosphate on the end of duplex chains destabilizes the COL phase. (iv) H-bonding and stacking free energy is substantially larger in the G-quartet columns than for the duplex columns. The general behavior of the phase boundaries of Figure 4.13 is movement to higher  $c$  and lower  $T$  with decreasing oligomer length, with the trends,  $(\partial T_{ILC}/\partial N)_c \sim -6^\circ\text{C}$  and  $(\partial T_{ILD}/\partial c)_N \sim 0.06^\circ\text{C}/(\text{mg/ml})$ , found for blunt end self-complementary DNA in the length range  $6 < N < 20$  duplexes, roughly in force. The stability of the C-G mixtures relative to that of the A-T mixtures is evident for the shorter molecules. Taking the phase transition temperature ( $T_{CI}$ ) to be the upper  $T$  limit of the coexistence, we find that  $T_{CI} \sim 20^\circ\text{C}$  for the dATP/dTTP (1:1) mixtures at  $c > 900$  mg/ml. These COL textures show little change upon cooling in the range  $20^\circ\text{C} > T > -20^\circ\text{C}$ . For  $c \sim 900$  mg/ml, the sodium ion concentration in the water is  $[\text{C}_{\text{Na}}] \sim 5\text{M}$ , strongly depressing freezing of the solvent. In the dCTP/dGTP (1:1) mixtures at  $c \sim 650$  mg/ml where LC ordering first appears, the COL phase is an assembly of duplex C-G columns. For  $c > 800$  mg/ml G- quadruplex columns appear, strongly enhancing LC COL phase stability.

## 4.5 Variation of dNTP Ratio in Solutions

### 4.5.1 COL-ISO Phase Transition Temperature Versus NTP Molar Ratio

Under the assumption that bases move as individuals in the ISO and as pairs in the COL phase, the net free energy change associated 1:1 pair-wise association of, for example, dATPs and dTTPs into such aggregates has a contribution from the loss of the entropy of mixing of the ISO phase, a loss that is minimized if the molar ratio of the two bases in the ISO is also 1:1 (see Section 4.5.2). This ordering step

forms aggregates of pairs which may or may not simultaneously achieve long-ranged columnar LC order, the latter being the first-order phase transition step which renders the aggregates visible in PTOM and XRD. However, in oligomers observed to date the temperature between LC melting and (higher-temperature) duplex unbinding decreases with oligomer length<sup>25–27,31</sup> and disappears completely in, for example, the 4-mers dATTA and dGCCG<sup>34</sup>, where aggregation is via simultaneous duplexing and base stacking. So we assume that the latter is also the case in the dATP/dTTP mixtures, with H-bonding into duplexes, base stacking, and COL LC formation essentially coincident events, meaning that it is at the observable ISO-COL phase transition that the entropy of mixing is lost.

#### 4.5.2 Model of Entropy of Mixing in the COL-ISO Phase Transition

In order to calculate the entropy of mixing effect, we consider the transition to be first order, with  $\Delta S$  the unbinding transition entropy per A-T base pair, taken to be  $\Delta S_{IC} \approx 20$  cal/molK<sup>46,47</sup>. Consider the case of an ATP/TTP mixed solution, with an ATP molar ratio,  $x$ :

$$\text{molar ratio} = x = \frac{[ATP]}{[ATP]+[TTP]} \quad 0 \leq x \leq 1 \quad (4.2)$$

The contribution of ATP/NTP mixing to the entropy of the ISO phase is  $\delta S_{ISO}$ .

$$\delta S_{ISO} = -k_B [x \ln x + (1 - x) \ln(1 - x)] \quad (4.3)$$

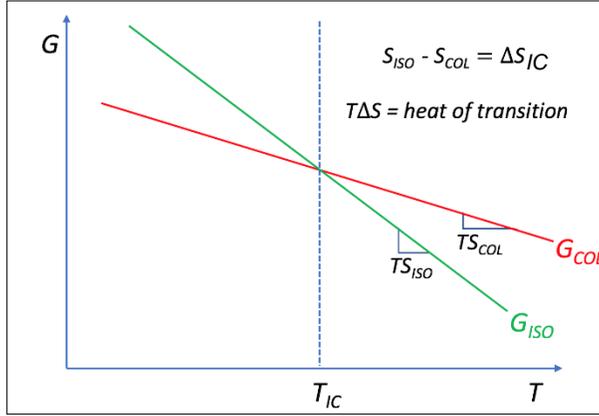


Figure 4.14: Gibbs free energies for a first-order ISO-COL phase transition. The entropy difference between the ISO & COL phases is obtained from the heat of transition  $T_{IC}\Delta S_{IC}$ <sup>47</sup>.

The Gibbs free energy of the ISO-COL transition can be plotted as in Figure 4.14 under the assumption that it is first-order. Suppose that the change in  $\delta G_{ISO}$  is produced by a change in its entropy  $\delta G_{ISO} = -T_{IC}\delta S_{ISO}$ . Then, from Figure 4.15 with  $\delta x = \frac{1}{2} - x$ ,

$$\frac{\delta T_{IC}(\delta x)\Delta S_{IC}}{k_B T_{IC}} = f(\delta x) = -\left[\left(\frac{1}{2} + \delta x\right) \ln\left(\frac{1}{2} + \delta x\right) + \left(\frac{1}{2} - \delta x\right) \ln\left(\frac{1}{2} - \delta x\right)\right] \quad (4.4)$$

plotted as  $f(\delta x) - f(0)$  in Figure 4.16. For  $\delta x$  small:

$$\frac{\delta T_{IC}}{T_{IC}} \approx \left(-\frac{k_B}{\Delta S_{IC}}\right) [2\delta x^2]. \quad (4.5)$$

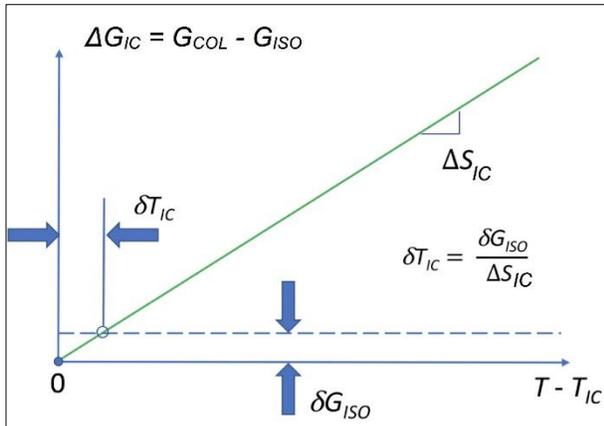


Figure 4.15: ISO-COL transition temperature shift resulting from a change in the free energy of the ISO phase,  $\delta G_{ISO}$ . In the case of mixing,  $\delta G_{ISO} = T_{IC}\delta S_{ISO}$ .

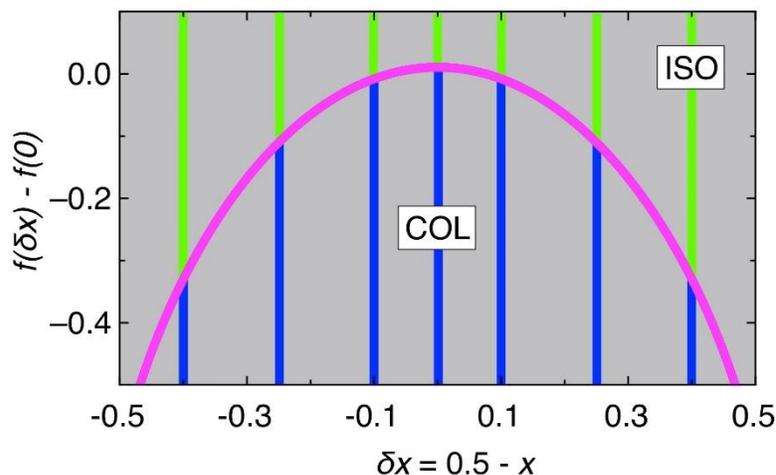


Figure 4.16: Model of equation 4.4 for the shift of the transition temperature range for COL LC ordering of an NTP binary mixture with increasing unbalance of molar ratio, as given by  $\delta x = 0.5 - x$ . Equation 4.4 calculates the shift in  $T_{IC}$ , the ISO-COL transition temperature due to  $\delta S_{iso}(x)$ , the entropy of mixing of the NTPs in the ISO phase (magenta curve).

#### 4.5.3 NTP Entropy of Mixing –Role of Base Pairing

The specific role of base-pairing in the stabilization of the LC phase in NTP mixtures was probed by measuring the dependence of the LC phase behavior on the relative concentrations of dATP and dTTP in dATP/dTTP mixtures of fixed overall concentration,  $c = 840$  mg/ml, prepared in a series of flame sealed capillaries. The transition to a COL phase is not observed for temperatures down to  $T = -20$  C in solutions containing only dATP or dTTP. If a COL phase appears, the essential ordering step is simultaneous H-bonding and aggregation that nucleates and/or extends the length of duplex stacks. The data of Figure 4.13, for mixtures with  $c = 840$  mg/ml and  $x = 0.10, 0.25, 0.40, 0.50, 0.60, 0.75,$  and  $0.90$ , show a clear depression of  $T_{IC}$  relative to the mixture of maximum  $T_{IC}$ , which is for  $x = 0.4$ . Although the entropy of mixing effect is symmetric about  $x = \frac{1}{2}$ , the overall solution

thermodynamic properties are generally not. For example, 60% dTTP-rich and 60% dATTP-rich solutions at the same overall  $c$  will have different values of the free energy difference  $\Delta G_{IC}$  between their ISO and COL phases, since these mixtures have different chemical compositions. Apparently  $\Delta G_{IC}$  is smaller for dTTP-rich mixtures.

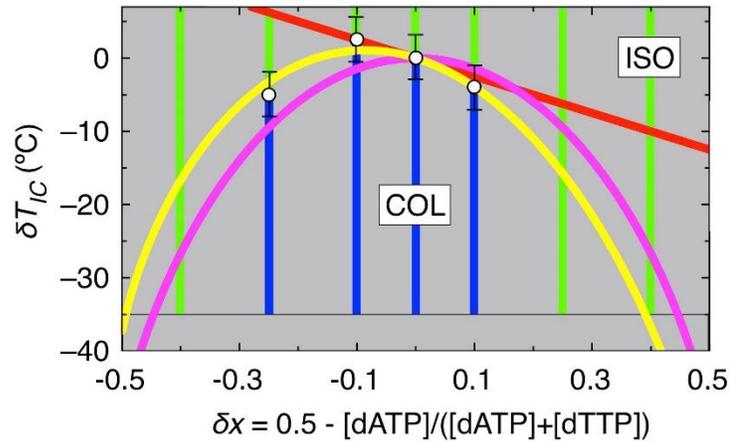


Figure 4.17: The points show the observed shift of the transition temperature  $\delta T_{IC}$  for ISO/COL LC ordering of 840 mg/ml dATP/dTTP mixtures with varying degrees of unbalance of their molar ratio as given by  $\delta x = 0.5 - [dATP]/\{[dATP]+[dTTP]\}$ . Here  $\delta T_{IC} = T_{IC}(\delta x) - T_{IC}(\delta x=0)$ , where  $T_{IC}(\delta x=0) = 15^\circ\text{C}$ . The green lines are the observed range of the ISO phase, and the blue lines the observed range of the COL phase at the various  $x$ -values. The green (blue) lines end at  $T = -20^\circ\text{C}$ , the lowest temperature available. Equation 4.4 calculates the shift in  $T_{IC}$ , the dATP/dTTP ISO-COL transition temperature due to  $\delta S_{iso}(x)$ , the entropy of mixing of the dATP and dTTP in the ISO phase (magenta curve). This contribution is symmetric about, and peaked at,  $\delta x = 0$ . The observed boundary is peaked  $\delta x = -0.1$ , indicating that COL order is stabilized by increasing the overall concentration of dATPs, a behavior modeled by assuming a contribution to  $\delta T_{IC}(\delta x)$  linear in  $\delta x$  ( $l(\delta x)$  red line). The  $\delta T_{IC}(\delta x)$  data are then fit to the function  $\delta T_{IC}(\delta x) = \alpha l(\delta x) + \beta f(\delta x)$ , where  $f(\delta x)$  is in Equation 4.4 and  $\alpha$  and  $\beta$  are parameters. The yellow curve is a fit of this function to the data.

The observed  $T_{IC}(\delta x)$  data are plotted in Figure 4.17 as  $\delta T_{IC} = T_{IC}(\delta x) - T_{IC}(\delta x=0)$ , where  $T_{IC}(\delta x=0) = 15^\circ\text{C}$ . In order to model  $T_{IC}(\delta x)$  we assume that the A-rich stabilization of the COL phase can be described by a contribution,  $l(\delta x) = a\delta x$ , to  $\delta T_{IC}(\delta x)$  that is linear in  $\delta x$ , added to the entropy of mixing contribution in equation 4.4 as  $\delta T_{IC}(\delta x) = a\delta x + \beta f(\delta x)$ , where  $a$  and  $\beta$  are adjustable constants. This function

gives a reasonably good fit (yellow curve), enabling the determination of  $\beta = 75\text{K} = k_B T_{IC} / \Delta S_{IC}$ , giving, in turn,  $\Delta S_{IC} = 7.6 \text{ cal/molK}$ . The literature values have  $\Delta S_{IC}$  of poly[d(A)-d(T)] or poly [d(A-T)] duplexes in the range  $17 < \Delta S_{IC} < 25 \text{ cal/molK}$ , with the lower  $\Delta S_{IC}$  values for the larger salt concentrations<sup>46,47</sup>. The counterion salt concentrations employed here are substantially larger, so that the lower  $\Delta S_{IC}$  might be expected on this basis. The absence of the COL phase for large  $|\delta x|$  may be a result of an enhanced barrier for nucleation, due to the lower  $T$  and dilution of one of the components.

#### 4.6 dATP/dCTP/dGTP/dTTP COL-ISO Texture Observations

Samples of dATP/dCTP/dGTP/dTTP (1:1:1:1) were prepared and observed as for the single and binary NTP mixtures described above: aqueous 100 mM dNTP solutions were added together and mixed, dried onto a glass substrate by rough vacuum, then sealed with oil and hydrated with a droplet of water to promote formation of phases in an oil sealed free surface cell (Figure B.8). These samples were then held at  $5^\circ \text{C}$  and examined by PTOM for the formation of phases along the contact line. Samples containing all four dNTPs formed birefringent structures at  $5^\circ \text{C}$  much more slowly than either the dATP/dTTP or dGTP/dCTP mixtures, but formation accelerated upon reducing  $T$  to  $T \sim -5^\circ \text{C}$ . When subsequently incubated at  $2^\circ \text{C}$  for several days, these samples showed well organized textures characteristic of COL phase (Figures 4.8 and 4.18), as observed with the AT and CG binary dNTP mixtures. After incubation for several days textures show an enhancement of thermal stability, the COL phase persisting up to  $T \sim 25^\circ \text{C}$  before completely melting, behavior which indicates a slow partitioning of the dNTPs between the COL and ISO phases. Upon heating to produce complete melting, such a sample must be lowered to  $2^\circ$ - $5^\circ \text{C}$  to reestablish the COL, i.e., significant

hysteresis is observed. In oil sealed flat cells the COL domains melt and reform with change of  $T$  as indicated by the blue curve in Figure 4.13. Such low  $T_{IC}$  values indicate that the COL phase is either dATP/dTTP or dCTP/dGTP duplexes. In either case  $T_{IC}$  in the four-base mixture would be shifted lower relative to that in an equimolar two-base mixture by an amount  $\delta T_{IC} = -(k_B T_{IC} / \Delta S_{IC}) \ln(4/3)$ , as a result of excess entropy of mixing in the four-base ISO phase. Taking the unbinding entropy per base pair to be that obtained in Section 4.5.3  $\Delta S_{IC} \sim 7.6$  cal/molK, gives  $\delta T_{IC} \sim 20^\circ\text{C}$ , an upward shift of the dATP/dCTP/dGTP/dTTP (blue) curve in Figure 4.13 to the range of the dCTP/dGTP 1:1 mixtures, suggesting that the COL phase in the four-base mixture in Figures 4.8 and 4.18 is the ordering of dCTP/dGTP duplex columns.

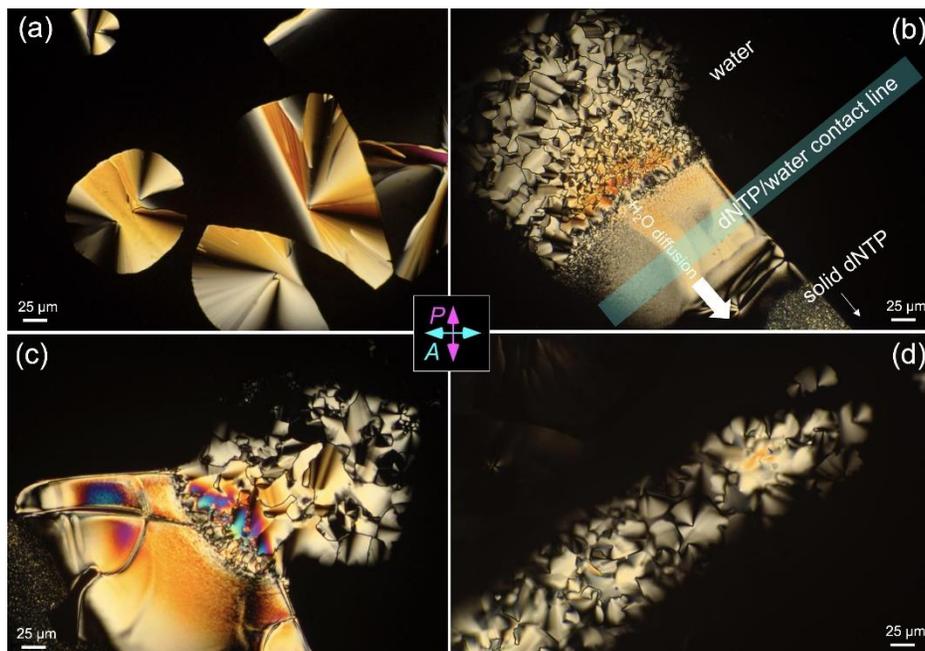


Figure 4.18: (a-d) Uniaxial columnar (COL) domains observed after four days of incubation of an oil sealed free surface cell of the four component equimolar dNTP mixture dATP/dCTP/dGTP/dTTP. (b,c) Emergence of the COL domains from the solid dNTP/water interface mixture in the free surface cell. The COL textures indicate that the domains are of duplex stacks and the melting temperature  $T_{IC} \sim 0^\circ\text{C}$  indicates that the domains are most likely from duplex dCTP/dGTP columns.

#### 4.7 The Importance of dNTP Liquid Crystals to the Origin of Life

Many lines of evidence presented suggest quite strongly that the dNTP liquid crystals, particularly the dATP-dTTP mixture, are strongly homologous to the mesophases produced by duplexes formed of intact oligonucleotides. dNTP phases are apparently columnar and uniaxial given the appearance of textures that appear very similar to COL (Figure 4.6). These phases occur in circumstances where a given dNTP is present with its Watson-Crick (or Hoogsteen) base pairing partner, suggesting strong dependence on the act of pairing (Figure 4.8). They possess negative birefringence, which implies that the material slow axis is perpendicular to the local director in the texture, much like nanoDNA aggregates, further suggesting COL (Figure 4.3). WAXS data indicates that dNTP mixtures sit on a COL-like 2D lattice, where a hexagonal model for duplex packing matches very closely to the expected reciprocal space periodicity at a given concentration (Figure 4.10). As well, dATP with dTTP appears to fit to the thermodynamics of two-component mixtures in the appearance of a characteristic melting point depression when the mixture is shifted away from the 1:1 mixing ratio (Figure 4.17). Finally, dNTPs can form phases in four component dA, dT, dG and dC mixtures, but may be subject somewhat to demixing (Figure 4.18).

The exact nature of the dNTP COL phase is still somewhat uncertain in its nanostructure. The dNTPs appear to stack in columnar aggregates, but it is currently unknown whether they order as a double helix or simply maintain some average cylindrical distribution in the stack where the orientation is not fixed in any direction. Early atomistic simulation studies would seem to favor an A-form double helix for this material instead of B-form and it is still not completely certain if base pairing is by the Watson-Crick mode (syn-pairing) or the Hoogsteen mode

(anti-pairing) or even some noisy mixture of the two. Still, that the mesophase is dependent on pairing at any level seems completely certain.

Expansion of this research into RNA NTPs in order to direct efforts to the origin of life have borne fruit but remain somewhat incomplete. RNA appears to be only passingly homologous to DNA; in particular, the substitution of Uracil for Thymine appears to prohibit LC formation, though whether or not it is fully the combination of the 2'-hydroxyl with the removal of the Thymine methyl to make Uracil is still being explored since partial combinations of these two features can make mesophases.

In all of this, one of the most critical questions remains open. The special biological feature of dNTP is the chemical reactivity of its triphosphate group. DNA and RNA are made from dNTP (and NTP, respectively) where a polymerase enzyme uses the chemical potential energy from the leaving group of pyrophosphate to drive a hydroxyl mediated crosslinking reaction between the 3'-terminal of a growing polynucleotide and the properly positioned triphosphate of a dNTP<sup>6</sup>. Column-stacked dNTPs would seem passingly to be the definition of "properly oriented" to help facilitate this reaction without a polymerase. This has been one of the most tantalizing potential features of this system: ATP is the chemical energy backbone of life, connecting energy metabolism to protein chemistry to nucleotide polymerization. Connecting DNA liquid crystal formation in some way to the main thoroughfare of biochemistry, suggesting how the RNA world could be coupled to energy metabolism, would significantly expand the scope of the hypothesis.

On the other hand, we have not yet seen conclusively that polymerization occurs spontaneously at times when liquid crystals are possible. There have been a few hints that dNTP stability is imperfect and that the chemical populations in a sample change gradually over time but we have had no success probing those

populations to find polymerized DNA. On concluding this thesis, a lack of confirmation of this last point seems to be one of the greatest gaps of the research presented here. While dNTP is renowned for its stability, it is also known in some contexts to polymerize spontaneously. Would it be too much to ask that such a context be here?

There are some other pathways open which hold promise. In particular, there has not been significant exploration yet on the effect of polyvalent cations on the mixture, particularly magnesium. Magnesium is well understood to play a role in natural DNA polymerization and some usage of this cation could reveal a few things we have not yet seen.

It's also true that the abiogenesis community as a whole seems to disfavor triphosphates for their complexity and dismisses them as “too tuned and too modern.” In the author's opinion, this neglects the fact that dNTP is seen in nature even while compounds apparently more favored by the community have not been seen in the same context—can an artificial compound ever be considered “more natural” when it hasn't come out of the nature? Still, there is the potential that LC mesophases similar to those seen with dNTP can be found with compounds more favored by the abiogenesis community, such as the nucleosidal phosphorimidazolides. An NTP-like molecule possessed of sufficient charge, base pairing valence and hydrophobic stacking should be able to duplicate the NTP assembly mode. Finding a suitable homolog is simply a matter of time and effort.

## Appendix A

### Oligonucleotide Synthesis and Purification

The section of thesis that follows is a detailed report of methods outlining how to synthesize and purify oligonucleotides with an eye for application to nanoDNA liquid crystals. Much of the original research presented elsewhere in this thesis can be reproduced by other members of the lab, while what follows here cannot be recreated from scratch without a vast timestep despite the fact that it is only indirectly important to research –as this section is a giant cookbook, I switch frequently to direct personal pronouns in my efforts to be instructive. What follows in this section, I feel, is a necessary attempt to safeguard these techniques for the sake of the lab against the future. It is not sexy research, but the intricacy renders it completely irreplaceable.

#### A.1 Overview

Deoxyribonucleic acid (DNA) and Ribonucleic acid (RNA) are heteropolymers composed of a repeated sugar and phosphodiester backbone where each sugar is decorated with one of four nitrogenous base compounds (Figure A.1)<sup>48</sup>. RNA and DNA monomers are constructed around the core of a five-carbon ribose sugar, the

carbons numbered 1' through 5' for identification purposes, are cyclized into a 5-membered hemiacetal ring with the hydroxyl at the 4' position attacking the aldehyde at 1' (numbering of the carbons is diagrammed in Figure A.3). This core differs between DNA and RNA by replacing the 2' hydroxyl found in RNA ribose with a simple proton in DNA deoxyribose. The nitrogenous base adorns the 1' carbon and the substituents of the sugar core define the chirality of the polymer. The phosphodiester linkages that string together the backbones of both DNA and RNA are established between the 3' hydroxyl of one sugar and the 5' hydroxyl of the next sugar in the backbone. For DNA, the bases are Adenine (A), Guanine (G), Thymine (T) and Cytosine (C), while RNA is much the same, but replaces T with Uracil (U). Reference to DNA versions of base monomers will contain a prefix of “d” (dA, dG, dC) while RNA forms will be labeled with an “r” (rA, rG, rC). Oligonucleotides with sequences given in caps will be DNA (e.g. Dickerson Dodecamer  $5'CGCGAATTCGCG^3'$ ) while sequences in lower case will be RNA (e.g. RNA Dickerson Dodecamer  $5'cgcgauuucg^3'$ ).

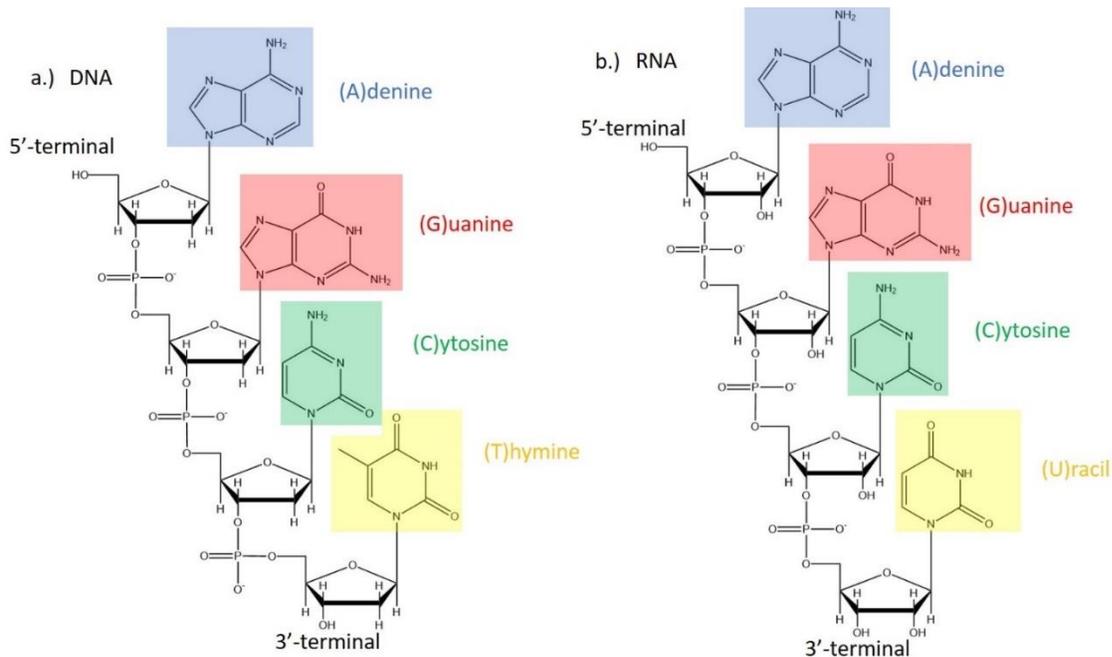


Figure A.1: a.) Deoxyribonucleic acid. b.) Ribonucleic acid.

In conventional nomenclature, including only the sugar and base, the monomer unit is called a nucleoside; including sugar, base and phosphate, the monomer is a nucleotide. As produced by the synthetic process to be described, oligomers contain  $n-1$  nucleotides and a single nucleoside. Polarity of the oligomer is described relative to exposed 5' and 3' hydroxyls on the sugars at the terminal ends of the polymer sequence, as marked in Figure A.1, where the 5'-terminal is the end of the polymer with an exposed 5'-hydroxyl while the 3'-terminal is the exposed 3'-hydroxyl at the opposite end of the polymer. 5' is typically regarded as the "front" of the oligonucleotide while the 3' is the "back." This choice of polarity derives from the direction of polymerization in living cells where oligonucleotides are synthesized beginning at the 5' end and traveling toward the 3' end. A final phosphate group may be added at the 3' and/or 5' terminals by specialized methods, also to be described.

## A.2 DNA Synthesis Chemistry

Automated synthesis of DNA and RNA is achieved by running a simple series of reactions in a repeatable cycle to expand a chemical functionality bound to a solid phase into a full length heteropolymer with a precise sequence of monomer units. The general scheme of the method, as pioneered by Caruthers et al<sup>49,50</sup>, is outlined for brevity in Figure A.2.

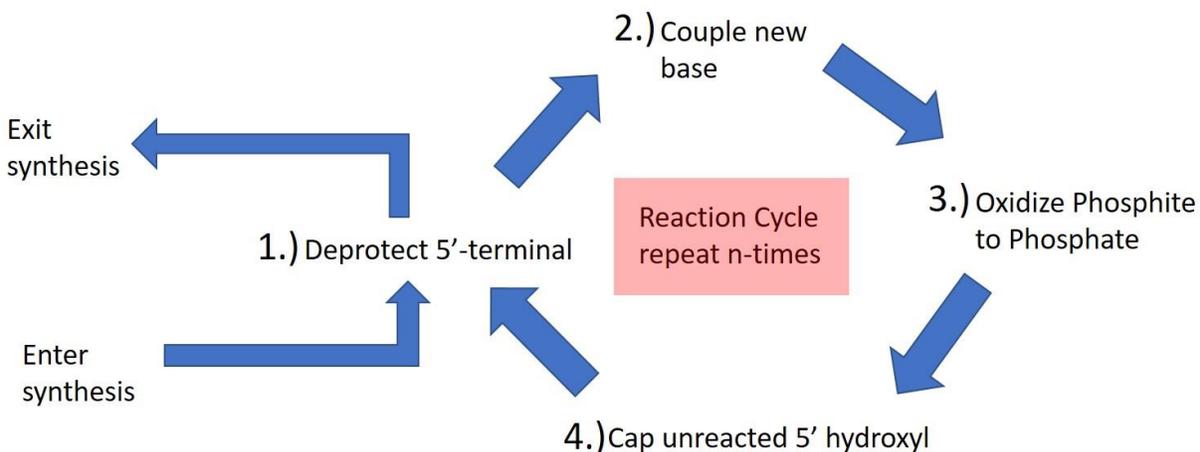


Figure A.2: Schematic overview of oligonucleotide synthesis process.

Synthesis is achieved by performance of only four reactions in a loop. This loop can be repeated many times at high yield to produce polymers of more than 50 nucleotides. The first reaction of synthesis is the 1.) terminal deprotection reaction, followed by 2.) attachment of a new base, 3.) formation of the phosphodiester linker, then 4.) block unreacted terminals. Each loop of the synthetic cycle is restricted to addition of only one base at a time by the inclusion of protecting groups that prohibit base monomers from reacting with each other and forward progress into a successive loop cannot occur until a chemical protecting group blocking the 5'-terminal is removed. Given typical methods, repetition of the reaction cycle  $n$ -times produces an oligomer of length  $n+1$ . Conclusion of the synthesis also typically exits from step 1.) as the protecting group sitting on the 5'-terminal is not present in natural DNA. All DNA (and RNA) synthesis is built around some version of this method. Wherever necessary, I will detail variations that have been used in my work.

The Caruthers synthetic method is driven by phosphoramidite chemistry<sup>49</sup>. This method revolves around a synthon called a DNA (or RNA) phosphoramidite (Figure A.3). The acid labile DMT-protecting group at 5' assures that base additions

only occur once per synthetic loop, while decomposition of the phosphoramidite at the 3' position permits creation of natural phosphodiester linkages at very high yield. Other protecting groups prevent undesired cross reactions.

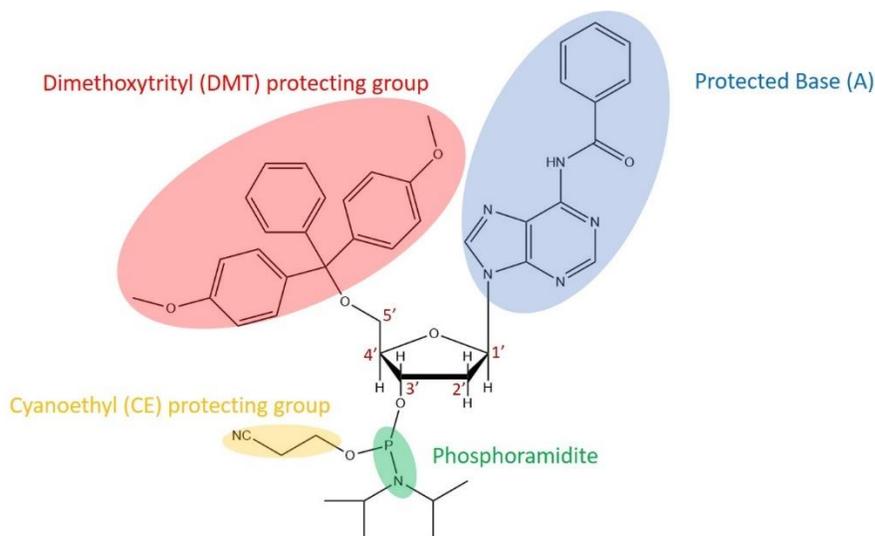


Figure A.3: Adenine Deoxynucleoside phosphoramidite (dA-PA). Other relevant phosphoramidites swap the base. Numbering of the sugar carbons is as shown. Deoxynucleoside phosphoramidite contains an acid labile DMT protecting group on the 5' carbon (red) and base labile CE protecting group helping block the phosphate (yellow). It may also contain whatever other groups are protecting the base (blue). The reactive phosphoramidite on the 3' carbon is shown in green. Going forward, use of the term “phosphoramidite” referring to this entire compound and its four A,T,G and C variants will be abbreviated as “PA,” while reference specifically to the reactive phosphoramidite functionality will not be abbreviated.

A more detailed schematic for the chemistry of the main synthetic cycle for the Caruthers method is presented in Figure A.4. One loop of this figure depicts production of a protected DNA dimer from a monomer; on return to the starting state of the synthetic loop, one can imagine moving one base in the 5'-direction along the growing polymer and repeating the same steps each time a new base is added. Again, since the coupling reaction can only occur between an activated phosphoramidite and an exposed hydroxyl, the blockage of all hydroxyls except for the 5' terminal of the nascent polymer assures the directionality of the synthesis and restricts addition to one base per cycle at that location.

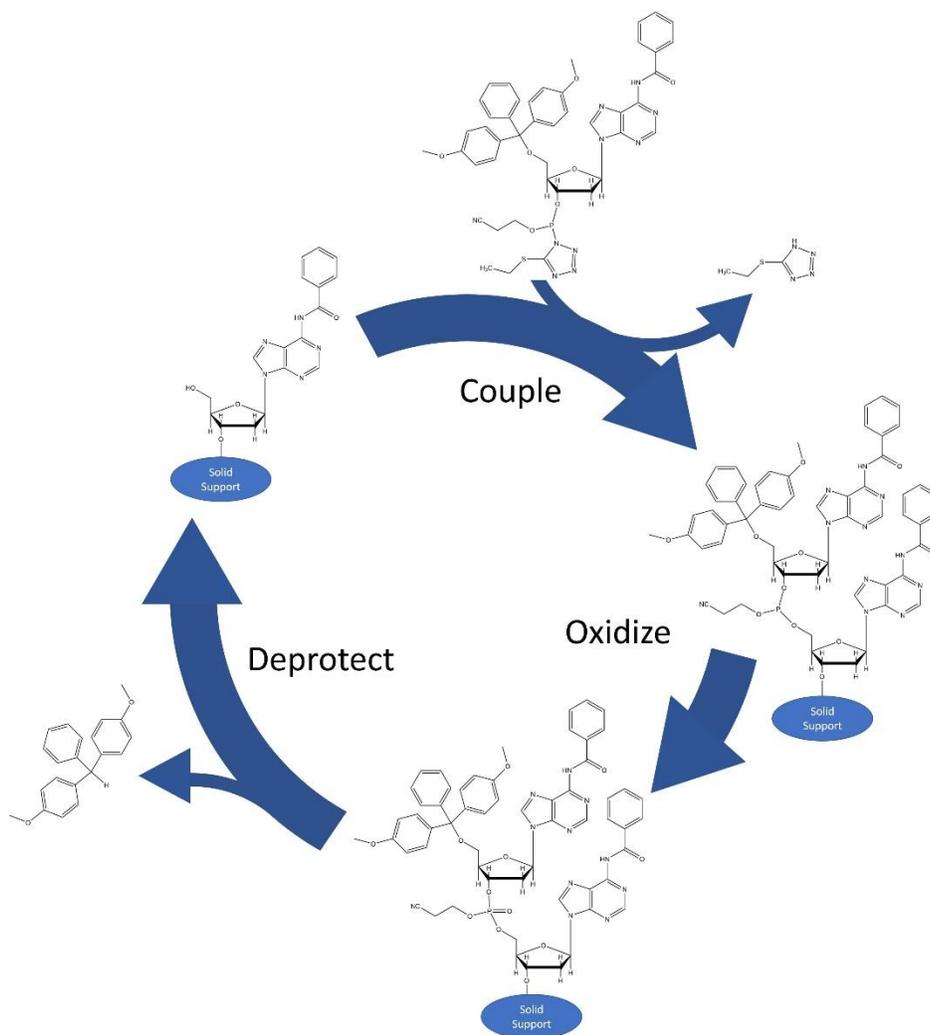


Figure A.4: The main synthetic cycle. Illustrated is the production of an Adenine nucleotide dimer left open ended to proceed into a second base addition cycle where the final deprotection step would bring the concluding dimer into the position occupied by the starting monomer. Bases between the active 5'-terminal dA and the solid support may be regarded as an inert linker where one simply climbs up the chain by one base at each deprotect step. As presented here, the phosphoramidite of the PA is in its tetrazole activated form rather than the storage form shown in Figure A.3.

The part of the main synthetic cycle presented in Figure A.4 includes only steps 1-3 depicted in the block diagram in Figure A.2. This is the “successful” side of the reaction progression: steps 3 and 1 follow in a synthetic program on the assumption that step 2, the coupling reaction, occurred properly. Phosphoramidite chemistry is of high enough fidelity that a vast majority of polymers in the mixture

follows this path. But, coupling is not perfect and reaction 4.) is included in the synthetic loop in order to prohibit polymers that have failed coupling at their exposed 5'-hydroxyl from progressing through successive synthetic loops and accumulating internal base omissions in the oligomer. The capping reaction is shown in Figure A.5.

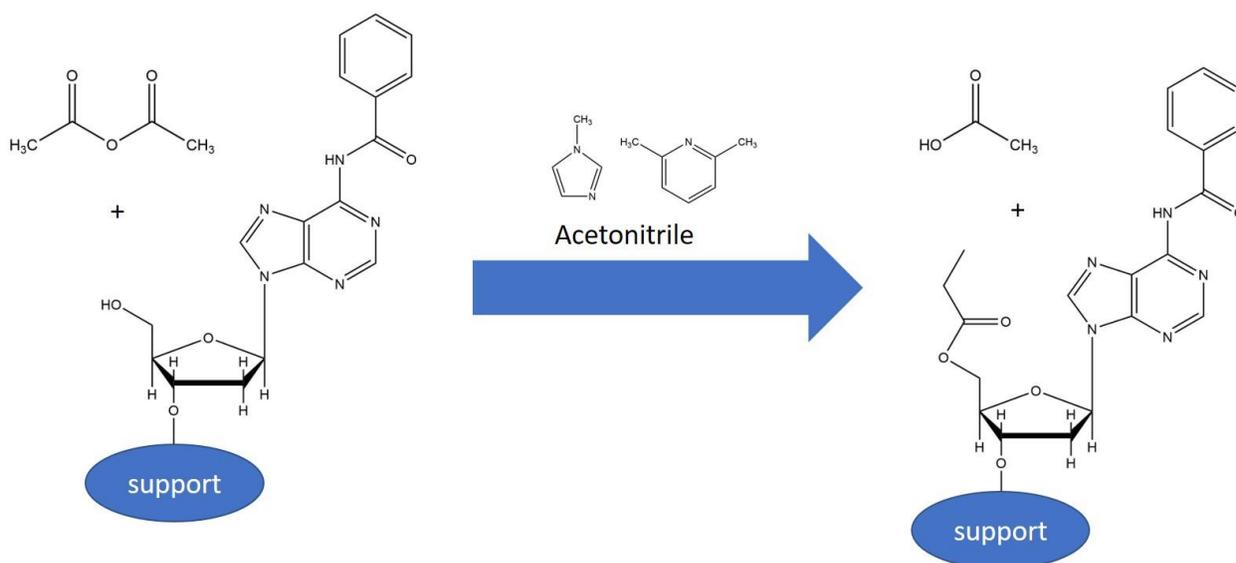


Figure A.5: Capping reaction. The capping reaction is implemented during the synthesis program following the oxidation step, but prior to the deprotection step that begins the next cycle of base addition. During the capping reaction, unprotected 5'-hydroxyl attacks acetic anhydride in the presence of 2,6-Lutidine and 1-Methylimidazole in acetonitrile solvent to produce acetate and 5'-acetylated deoxynucleoside (here protected adenine deoxynucleoside). The acetylation cannot be reversed in the conditions of the synthesis cycle and acetylated sequences do not participate in later reactions, making them unable to elongate further in synthesis and making it easier to separate them from successful synthetic sequences during subsequent purification. We frequently refer to the material generated here as “failure sequences” since they failed to elongate appropriately as a result of missing the coupling reaction.

An overview of the chemical steps producing the tetrazole activated form of phosphoramidite during synthesis is presented in Figure A.6. The chemistry depicted here is spurred to take place during synthesis by mixing ethylthiotetrazole with dN-PA during the coupling step<sup>51</sup>. This series of reactions takes the dA-PA

compound from the relatively stable storage form seen in Figure A.3 to the transient, highly-reactive “activated” form seen in Figure A.4 by exchanging a weaker leaving group for a stronger one.

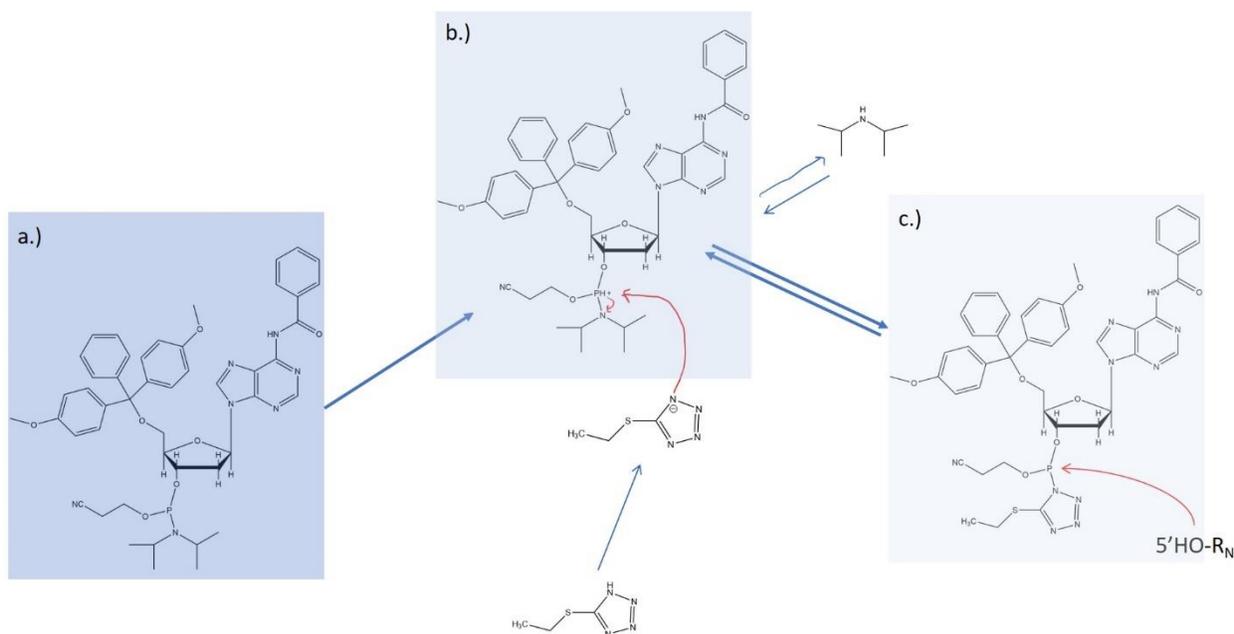


Figure A.6: Activation of dA-PA phosphoramidite during the base coupling step. a.) Unactivated dA-PA. b.) 1H-ethylthiotetrazole mixed in during the base coupling step attacks the phosphoramidite and displaces diisopropylamine. c.) Ethylthiotetrazole activated phosphoramidite is then subjected to an attack by available hydroxyl groups to form a phosphite linkage as seen in Figure A.4, displacing and regenerating the tetrazole.

A broad variety of related tetrazole compounds are regularly used in DNA synthesis, including the 1H-ethylthiotetrazole shown in Figure A.6. The tetrazole-diisopropylamine exchange step is a slow, reversible step, as indicated, and forward reaction is assured by the rapidity of the hydroxyl attack. Of note, any hydroxyl can participate in Figure A.6c.), including water, if it is present.

The deprotection steps typically encountered to produce the final DNA oligomer at the end of synthesis are presented in Figure A.7. This sequence of steps unblocks all amines and hydroxyls present in native oligonucleotide and frees oligomer from its anchoring to the solid support.

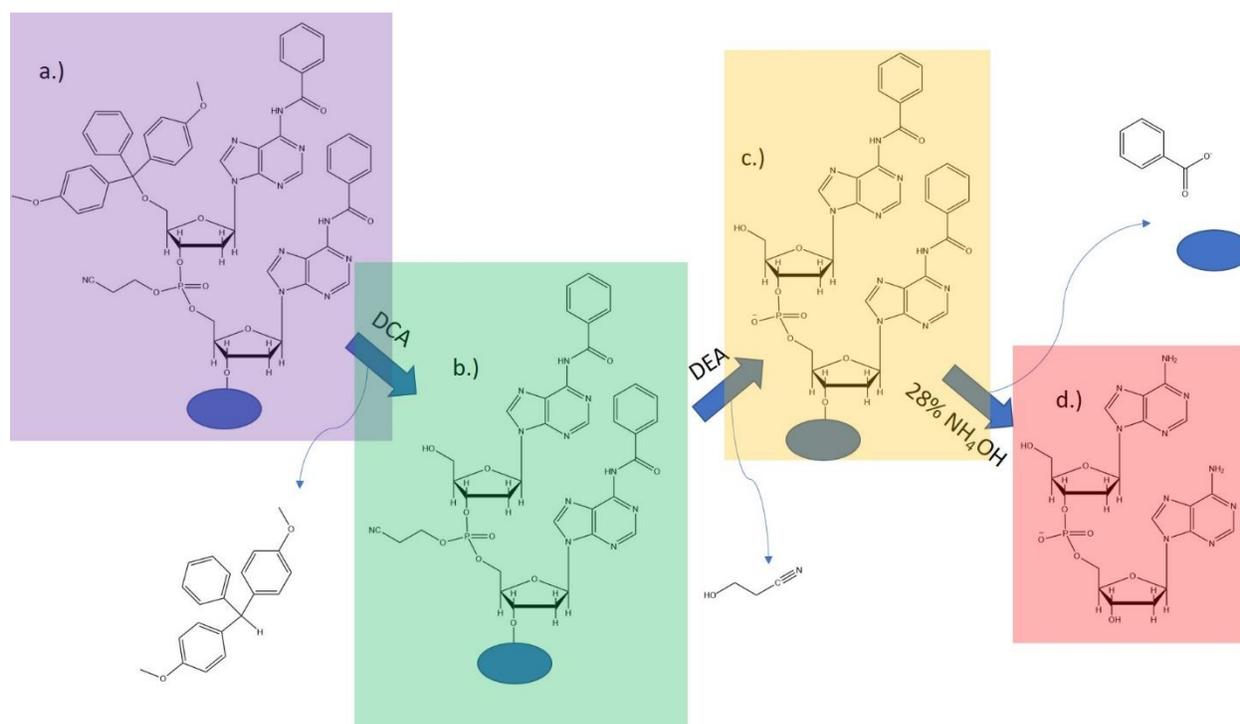


Figure A.7: Deprotection reactions to produce DNA after synthesis as exemplified by a dA dimer. a.) Acid wash (with Dichloroacetic acid, DCA) removes DMT protecting group. b.) Gentle base wash with Diethylamine (DEA) removes cyanoethyl protecting group from the phosphate backbone by  $\beta$ -elimination. c.) Strong base wash with 28% Ammonium hydroxide removes benzoic acid protecting groups and solid support from dA dimer (other DNA base protections are removed by the same treatment.) d.) Unprotected dA dimer, ready for purification.

These deprotection steps are specific to DNA synthesis. RNA synthesis is highly analogous but includes one additional protecting group to block the 2'-hydroxyl and will be tackled elsewhere. Removal of oligomer protection is the end of synthesis, but only the beginning of the process for obtaining purified DNA; at this point in the process, fractions containing synthetic DNA also include failure sequences produced by the process outlined in Figure A.5 as well as particulate matter in the form of solid support grains and residual protecting groups which may also be insoluble matter.

The requirements for purification are highly dependent on the quality of the coupling step during synthesis and on the overall length of the desired oligomer. A low-quality coupling step produces abundant failure sequences, requiring more involved purification, while a longer oligomer (>25 bases) tends naturally to accumulate a larger proportion of failure sequences over its run even with good coupling, also requiring more involved purification. If the coupling step is high quality and the oligomer sequence is relatively short, purification is in its simplest form. It should be understood that while debris from failure sequences is most minimized during synthesis of small oligomers, conventional DNA purification techniques begin to fail with sequences shorter than ~3 bases, requiring other techniques. In addition, longer DNA sequences, while tending to be more contaminated than short sequences, are typically better behaved in conventional DNA purification techniques, making purification yields higher. These and other considerations cause synthetic and purification yields to seesaw. The details of oligonucleotide purification and characterization will be addressed more completely in Section A.6 and A.7.

For completeness, Figure A.8 shows all four of the nucleic acid bases in their protected forms as would be appended to protected deoxyribophosphoramidite.

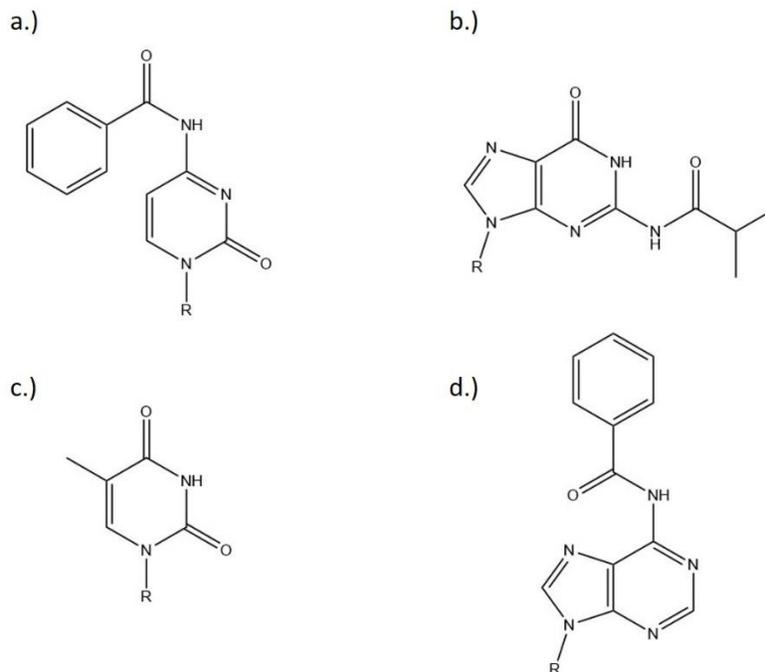


Figure A.8: Protected DNA bases. a.) Benzamide protected dC. b.) Isobutylamide protected dG. c.) Unprotected T; this base does not need protection during synthesis. d.) Benzamide protected dA. R-group shown here is the protected deoxyribosephosphoramidite of Figure A.3. All protecting groups shown here are labile only to stronger alkaline treatment (e.g. they resist a gentle base like DEA.)

### A.3 DNA Synthesis Automation

A discussion about the Caruthers synthetic method would be incomplete without some introduction to the associated automation. The DNA synthesis chemistries discussed in Section A.2 largely revolutionized molecular biology and have helped to support parallel advances due in large part to PCR technology, but they depend somewhat on a layer of advances in other technologies as well.

In absence of automation, the DNA synthesis workload is really quite big: unlike conventional plastic polymers, where identical monomers are strung together *en masse* to produce gigadalton sized molecules, DNA and RNA are copolymer type molecules where controlling the sequence order in which four disparate monomers

are strung together is of central interest. Attention must be placed on each and every monomer addition included in the oligomer. The result is that the polymerization process cannot be performed in a single pot. Fabrication of a discrete 12 base nucleotide requires, at minimum, 50 distinct chemical reactions implemented in a specific order, this including the deprotection steps necessary to produce the final oligonucleotide. Scaling this up to larger oligonucleotides quickly expands the number of reactions into the hundreds. A single person can perform this task by hand, but it would require specialized tooling at least and probably many manhours of work. Given that the Caruthers method boils down to four reactions performed in a cycle, the utility of automation is completely obvious.

With a robotic DNA synthesizer, the labor to perform a long string of chemical reactions is compressed to a computer script which can be executed in a single afternoon. If such a synthesizer is working perfectly, the human lab worker need only focus on the proper preparation of the program and chemical reagents and then decide how to manage the materials after the machine has concluded its cycle. This can reduce a laborious, demanding process to a merely onerous one where the user need only know the proper set-up without having to understand all the details of the chemistry being implemented.

On the other hand, because DNA synthesizers manage a comparatively intricate activity, they are necessarily complicated pieces of equipment and small mistakes on the part of an unprepared user can compound into serious faults in the material being produced. This sort of fault can be incredibly difficult to overcome. My first great recommendation about successfully using a DNA synthesizer is to always assume that the root cause of any problem is first and foremost the fault of the user before trying to lay blame on the machine. The machine will usually give a dutiful performance of any instructions provided by the user with the material

reagents prepared by that user; do not underestimate your own ability to give completely faulty instructions!

The chemical synthesizer used for experimentation described in this thesis is an Akta Oligopilot 100 originating from Amersham Pharmacia Biotech under the auspices of GE Healthcare Life Sciences company.

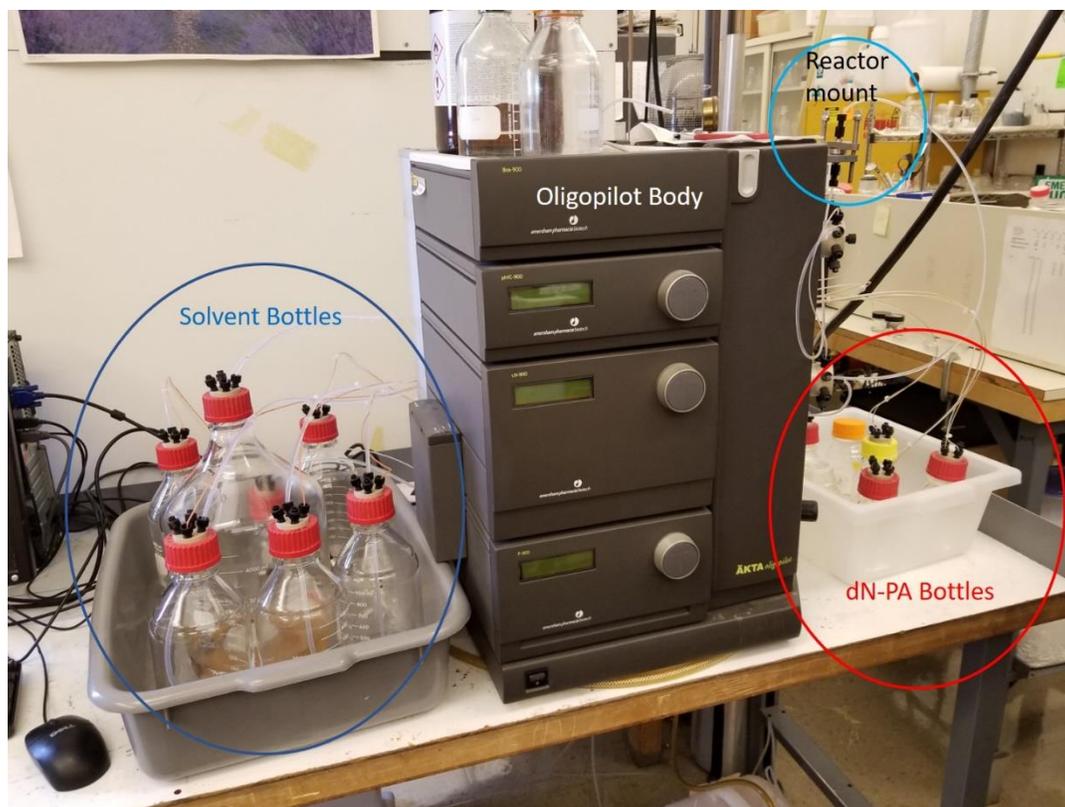


Figure A.9: Äkta Oligopilot 100. The main body of the robot is encased in the dark gray box. Solvent and reagent bottles sit in the gray bin on the left while smaller bottles for the dN-PA sit in the white bin to the right. At the upper right-hand corner of the Oligopilot body is the mounting bracket for the reactor column. Not pictured, the instrument is run by a Dell computer and is connected to a purified argon tank and a five gallon waste outlet reservoir.

This equipment amounts to nothing more than a collection of well-timed pumps and valves with an integrated gas manifold pressurizing the reservoir bottles with an inert atmosphere. It pulls liquids from the bottles, pumps them

through the reactor column, where the DNA is synthesized, and ultimately collects the outflow in a large waste reservoir to await disposal (Figure A.10).

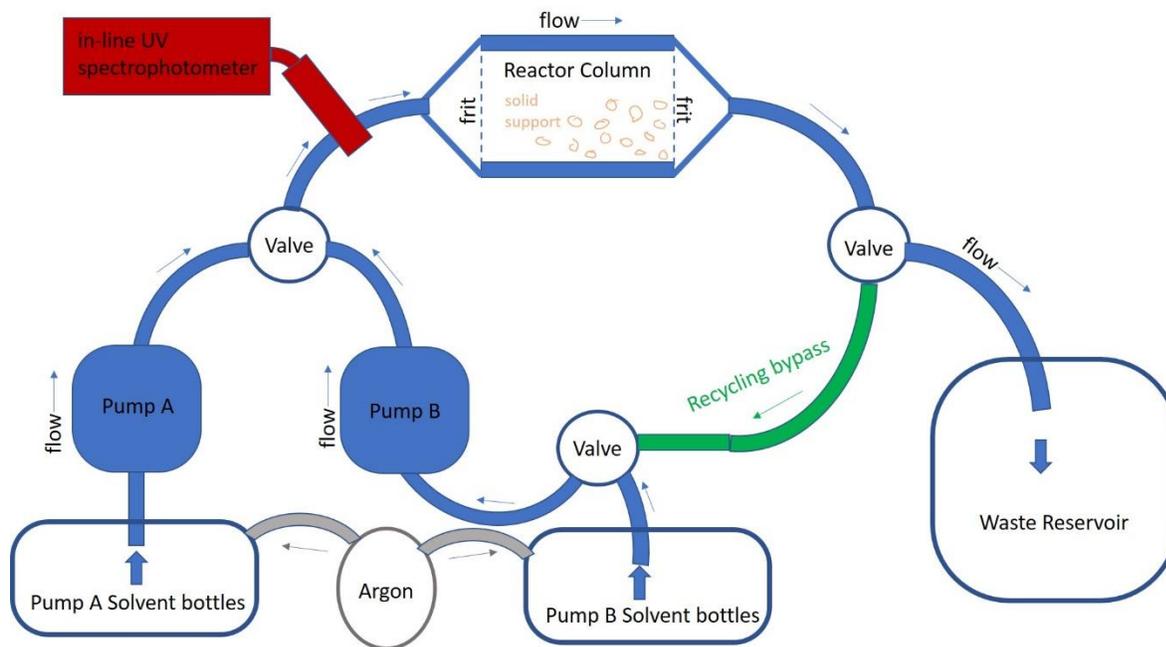


Figure A.10: Simplified schematic of DNA synthesizer plumbing.

Found in Figure A.10 is a cartoon schematic of the plumbing of a DNA synthesizer based loosely on the layout of the Oligopilot 100. The robot has two solvent pumps that work in tandem upstream from the reactor column. Each of these pumps serves a small collection of reagent bottles that are drawn from one-by-one using a rotating valve. These reservoir bottles are universally pressurized by purified argon to help exclude oxygen and water vapor from the system. The pumps draw from the solvent bottles and push reagents into a steel reactor column filled with solid support, upon which DNA oligomers are grown. The frits of the reactor column have a mesh size smaller than the solid support grain size, sequestering growing DNA in this portion of the system without need of a more rigid or mechanically complex partition. The two-pump design facilitates the mixing of chemicals within the solvent lines and can push the resultant mixture into the reactor column with no delay, or chance exposure to the external atmosphere. After

a reaction has been completed, the pumps can switch to an underlying solvent and wash spent chemicals out of the reactor and down line to the waste reservoir. The plumbing also contains a bypass line which can form a closed loop to connect the downstream end of the reactor column to the upstream end of Pump B, enabling recirculation within the system. Finally, the system adds various sensors, including a UV wavelength spectrophotometer, in order to keep track of solvent pressure, conductivity and absorbance during reactions.



Figure A.11: Oligopilot reactor columns. The 6.3 mL column is fully assembled while the 12 mL column has one end screwed off and the porous frit disassembled to exhibit the inside of the reactor.

The two columns in Figure A.11 are the available reactors for the Oligopilot used in this work. Solid support is sealed inside the crucible of the column and all reagents flow through the porous steel frits capping both ends. As mentioned, this device sequesters DNA within the reactor by simply making the frit pores smaller than the grain size of the solid support, allowing the growing oligonucleotides to be

exposed to the rest of the system through the frit without requiring any sort of active mechanism to contain them, and thereby prohibiting the grains from moving with the solvent flow. This general strategy is referred to commonly as Solid phase synthesis. The available reactor columns are 6.3 mL and 12 mL in volume and the combination allows for some scaling of the synthesis batch size. The Oligopilot is referred to as a “packed-bed” synthesizer because it does not use a commercially prepacked reactor column.

### A.3.1 Solid Support

Planning usage of solid support is of special importance in oligonucleotide synthesis. If the oligonucleotide is to be chemically functionalized at the 3'-terminal, this functionality is often added by use of specialized solid support. Functionalities of interest can include a phosphate group attached to the 3' hydroxyl, or structural linkers or specialized reactive groups, like sulfide groups suitable for azide Click chemistry. Any desired modifications at the 3'-terminal are most efficiently generated by special chemistry in the solid support.

As mentioned in Section A.3, solid support enables the techniques of solid phase synthesis, which serves to hugely improve the speed of synthesis and enables automation of the reaction cycle detailed in Section A.2. The main advantage of solid support is its ability to confine nascent oligonucleotides within the synthesizer while still enabling the desired chemistry to be brought to and removed from the reactor by a mobile solvent phase.

Typical solid support for DNA synthesis is fabricated from porous polystyrene or glass beads functionalized by a succinate linker that is then attached to the 3' position of an otherwise protected DNA monomer (see Figure A.4), which will serve

as the 3'-terminal nucleobase of a planned oligomer. This nucleobase is functionalized by the usual chemical protecting groups and protected at its 5'-terminal by DMT.

To be completely clear: one should always assume that any bases preloaded to solid support are thus protected unless documentation with the materials explicitly says otherwise and that synthesis *must* always begin with a clearly successful deprotection step. Omitting this initial deprotection step will automatically create conditions ripe for uniform n-1 failure sequences with an internal nucleotide omission at the second to last base from the 3'-end of the sequence. The assumption that solid-support preloaded bases are deprotected from the outset is a recipe for disaster that can take literally years to tease out!

For the most basic synthesis process there are four equivalent types of solid support, each bearing one of the four protected DNA bases. A typical choice of solid support is Primer Support 5G from GE Healthcare Life Sciences (Cat# 28-9964-25, 28-9964-26, 28-9964-27, 28-9964-28).

For a packed-bed synthesizer, an important consideration for planning synthesis is the desired scale of the process. Scale is selected by choosing the reactor column volume and solid support combination. Solid support is typically fabricated with a particular  $\mu$ mole quantity of seed-monomer per unit mass of the support: e.g. 350  $\mu$ mole of preloaded monomer in 1 gram of support using the solid support written above. The amount of support loaded in the reactor sets the absolute maximum  $\mu$ mole quantity of DNA oligomer that can be synthesized in a given synthesis run. Additionally, the volume of the reaction vessel itself sets limits on the polymer length that can be attained during synthesis; as successive monomers are added to the growing polymer, the volume occupied by the solid phase gradually swells until the reactor column is no longer large enough to contain the amount of

solid being confined there. From this, backpressure on the synthesizer pumps eventually becomes unmanageable and solvent flow rates through the reactor diminish. For the recommended reactor column loading density, this limit is about 25 nucleotides in an oligomer. For Primer Support 5G, the recommended loading is 111 mg of solid support to 1 mL of reactor volume (or 9 mL/gram of support). This variable can be obtained from solid support manufacturer specs. Polymers longer than 25 bases can be produced by this type of solid support, but it is recommended that the quantity of solid support be significantly reduced in order to make suitable room in the reactor for the final product.

#### A.4 Synthesis Reagents

For synthetic DNA described in this thesis, synthesis reagents were mixed from suitable chemicals acquired from typical vendors rather than purchased ready-made. For self-mixed DNA synthesis reagents, the theme governing success is “Dry.” As noted in Section A.2, water is a strong competitor with hydroxyl-groups for activated phosphoramidites, potentially outcompeting the solid support for the coupling reaction. Except for the oxidation reaction, which requires water, other reagents must be dry. We used acetonitrile for the main universal solvent and, as per the theme of dryness, suitable acetonitrile is dry to less than 10 ppm of water. Vendors frequently advertise this acetonitrile as “DNA synthesis grade” or “biosynthesis grade.”

The main tools for keeping the synthesis reagents out of contact with water are the presence of an inert purified argon atmosphere taking up the dead space in the chemical reservoir bottles of the synthesizer (prepurified, compressed Argon obtained from Airgas, cat# UN1006), and activated water-scavenging 4Å molecular

sieves (obtained from Acros Organics, Molecular sieves 4Å, 8 to 12 mesh, code# 197275000). These molecular sieves are a very hygroscopic material which is stable to many organic solvent environments which chemically sequesters water and effectively dries a solvent in the process. Molecular sieves were added to all synthesis reagents except the oxidizer.

For the most successful synthesis runs, the reagents were prepared one week in advance and allowed to sit on the synthesizer with molecular sieves added, under an argon atmosphere. Without this latency time to dry the chemicals, synthesis reactions were typically poorly efficient, at best <90% yield per cycle, conceding massive losses which can be observed by the time an oligomer has been grown to only six bases.

I will detail preparation for all the reagents and describe what these reagents do during synthesis wherever pertinent. Should this instruction be necessary, I recommend performing reagent preparation in a hood wherever possible and transporting solutions across the lab to their synthesizer bottles with lids screwed into place in order to minimize exposure to fumes –chemical MSDS documentation should be examined to note any additional pertinent safety information.

#### A.4.1 Preparation of Activated Molecular Sieves

Molecular sieves should be verified as active prior to adding them to any solvents. If they are not active, they should be activated a day prior to use in mixing reagent solutions.

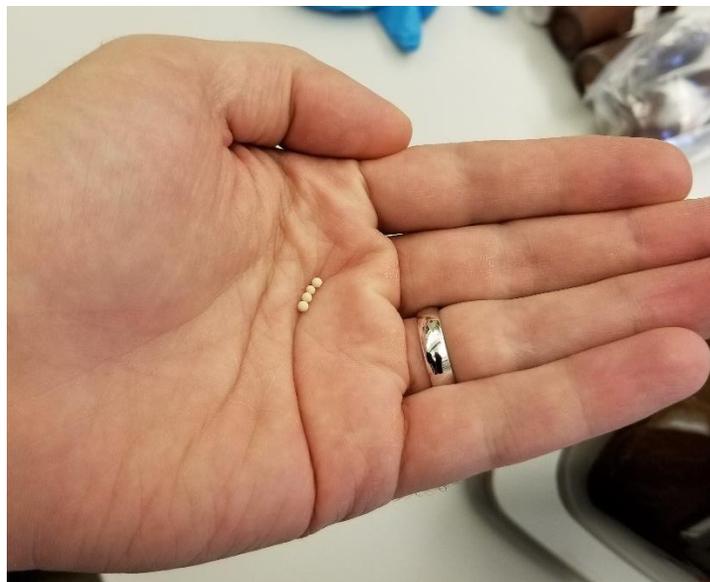


Figure A.12: Molecular sieves held in the hand prior to adding drops of water to check for the heat reaction consistent with activated molecular sieves.

To verify molecular sieve activity, take a small number of sieve beads in the palm of your hand and pour several drops of cold water onto them (Figure A.12). If they emit sufficient heat that you can feel it in the palm of your hand, they're active. One should take caution not to use too many beads at once to perform this check: they can cause thermal burns to your skin when they heat up. Through experience, you can notice that the heat reaction described here gradually dulls while sieves are stored, as they steadily take up humidity from the air. If you're uncomfortable with touching sieves, the heat reaction can be felt through nitrile or latex gloves adding only a few more beads.

Molecular sieve beads can be activated in bulk by cooking them overnight at 120°C and then allowing them to cool just prior to use. For the sake of being thorough, we cooked the beads at 120°C in a vacuum oven connected to a roughing pump and then allowing them to cool while still under vacuum to keep them out of contact with humidity during the activation process. The activation process is merely vaporizing water bound to the sieve material. As such, sieves which have

gradually inactivated through long term storage can be reactivated by cooking them again.

It should be noted that sieves are offered in porosity of either 3Å or 4Å. Do not use 3Å sieves because this mesh size is too small to admit water, making it difficult for 3Å sieves to dry solutions as efficiently as 4Å sieves.

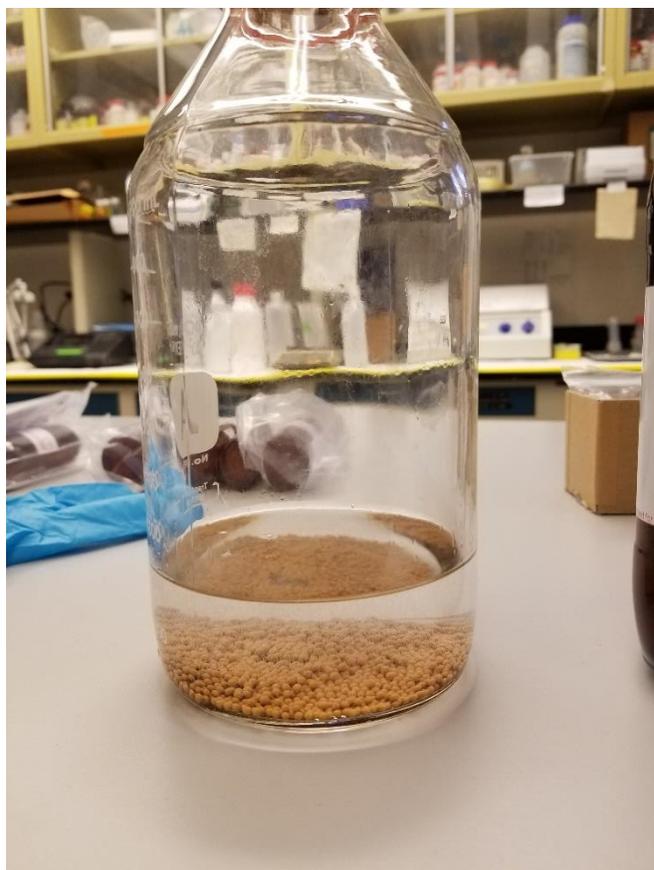


Figure A.13: Molecular sieves added in a layer to the bottom of a reagent bottle.

When sieves are added to solvents and reagents (Figure A.13), one will note an emission of tiny bubbles as the dry space of the sieves takes up solvent. It is typical to see a cloudiness or murkiness gradually appear in the solution from sieve dust. It has not been my observation that this cloudiness in any way inhibits DNA synthesis, but it does sometimes necessitate the cleaning of bottles during routine maintenance of the synthesizer system. I have also observed that the molecular

sieves apparently cross react with the acidic deblocking solution, resulting in the appearance of crystalline objects on the sieves.

#### A.4.2 Preparation of Capping Solutions

As noted in Section A.2, the DNA or RNA synthesis process is designed to try to inhibit continued growth of failure sequences by capping oligomers that fail their base coupling reaction. This is executed in the capping reaction (step 4 in Figure A.2). The capping reaction acetylates any exposed 5'-hydroxyl group prior to the DMT- deblocking step of successive synthesis cycles.

Capping is carried out with a binary chemical mixture which is combined immediately during its injection to the reactor column. The synthesizer brings solution Cap A, containing 1-methylimidazole, via pump A into contact with solution Cap B, containing acetic anhydride, on pump B, mixing them in-line while pushing both into the reactor column and onto the solid support. The mixture flows through the reactor, blocking exposed hydroxyls and is washed away afterward by a flush with empty solvent.

The synthesizer uses capping solutions at a relatively sparring rate of only a few milliliters each per base addition cycle. The usage rate is determined by the recorded reactor column volume. For a preferred synthesis program, the amount of capping solution added is one column volume per capping reaction, meaning that when the 6.3 mL reactor column is in use, each base addition will involve 3 mL of Cap A and 3 mL of Cap B per base. This means that 100 mL each of both capping solutions is sufficient to support the complete synthesis of several dozen bases. The amount used is doubled for the 12 mL reactor column and it can also be increased for certain solid support methods that I will outline later.

Table A.1

Cap A	100 mL
N-methylimidazole (Aldrich Cat# M50834-500G)	20 mL
Dry Acetonitrile	80 mL
Add molecular sieves to solvent bottle on synthesizer, a layer on the bottom	

Table A.2

Cap B	100 mL
2,6-Lutidine (Sigma-Aldrich Cat# L3900-500ML-A)	30 mL
Acetic Anhydride (Sigma-Aldrich Cat# 320102-500ML)	20 mL
Dry Acetonitrile	50 mL
Add molecular sieves to solvent bottle on synthesizer, a layer on the bottom	

#### A.4.3 Preparation of Oxidizer Solution

After the coupling reaction connecting the next dN-PA to the growing oligonucleotide, the newly added base is present at the 5'-terminal of the polymer linked by a phosphite triester (see Figure A.4), which is an unstable chemical functionality. The polymer must be stabilized by treatment with a strong oxidizer in order to form a phosphodiester bond.

This is achieved by adding an oxygen atom to the phosphate, as supplied by water, using Iodine in the form of  $I_2$ .

Since water is required in this reaction step, molecular sieves should *not* be added to the oxidizer solution. Also, given that water is circulated through the system in the process of running the oxidation reaction, special care is not absolutely required to dry down either the solvent lines of the synthesizer or the reactor column. The reactor column and attached solvent lines and valves can be sufficiently dried by circulating dry acetonitrile through them during washing steps in the synthesis program.

The oxidizer solution also contains pyridine in order to reverse phosphite adducts formed during the coupling reaction at undesired locations in the molecule.

Of special note, this particular solution has a very powerful scent provided by the pyridine and should absolutely be prepared in the hood and transferred to the synthesizer with a closed bottle. Pyridine is a hazardous material known to cause infertility and is carcinogenic, so it should be handled with care and breathed minimally. Also, because water is present in the oxidizer, this solution is not subject to the drying latency period used on other reagents chemicals and can be prepared any time prior to synthesis –given the noxious nature of the pyridine, immediately prior to synthesis is acceptable.

The synthesizer typically uses more oxidizer during a synthetic program than it does capping solutions, meaning that usually more should be prepared. The amount of oxidizer added by the synthesis program is calculated as a molar equivalent based on the number of moles of DNA seed present in the solid support. This number is automatically calculated by the system and does not particularly vary between programs, and so does not require special consideration. A preferred program uses 2.5 molar equivalents for every mole of DNA oligomer and applies a 2 minute contact time before washing the oxidizer out. 400 mL of oxidizer is more than sufficient for a typical DNA 12mer made in either 6.3 mL or 12 mL reactor column scale formats.

Table A.3

Oxidizer Solution	400mL
Solid Iodine (dissolve on addition of pyridine) (Fisher Chemical Cat# I37-100)	5.08 g
Pyridine (Fisher Chemical Catalog# P368-500)	360 mL
Water (to 10% /v)	40mL

#### A.4.4 Preparation of Deblocking Solution

Deblocking solution, also called “Detrit” for its function as a “detritylating solution,” uses a gentle acid treatment to remove the DMT-group from the 5'-hydroxyl of the growing oligonucleotide in preparation for the coupling reaction (see Figure A.4). This is the “deprotect” step leading every base addition cycle.

Because DMT is colored in deblocking solution, absorbing strongly at 350 nm and lending it a vivid red color, it can serve as a useful spectroscopic tag. When it is released from the growing oligomers, the amount detectable in solution is directly proportional to the number of 5'-hydroxyls that have been exposed. So, DMT can be used to judge the instantaneous oligonucleotide yield just prior to adding a new base. The synthesizer runs an automated spectrophotometer sampling of red-colored reactor column outflow during the deprotection reaction and integrates the area under the trace to determine the size of the DMT peak released. Changes to the area of this peak between samplings, which are performed at the beginning of every base addition cycle, can reflect changes in the oligonucleotide population bound to the solid support inside the reactor. This is one mechanism in the system to monitor the progress of the synthesis.

As a result, deblocking solution is used in relatively large quantities during synthesis and at a slightly variable rate. An amount of deblock is expended at the beginning of all synthesis programs in order to establish a spectroscopic baseline and then an amount is used during each base addition cycle, totaling as much as 15 column volumes during each deprotect step. This amount is variable because the flow triggering in the automation is dependent on the intensity of the spectroscopic trace, which does not stop the flow of deblock until DMT absorbance drops below a particular value. Typically, the release rate of DMT from solid support during the deprotection step decreases as the synthesis progresses and growing

oligonucleotides begin to hamper DMT outflow, causing the triggering block to increase in length over the course of synthesis –the absorbance peak area should ideally stay the same, but its time dimension will increase. The exact amount of deblock solution needed is not completely predictable, but an excess is required. The first handful of bases will require about 50 mL per base with this volume increasing gradually. During long synthesis programs (>20 bases) more than 1 L will be needed overall and the reservoir will need refilling at some point.

As a related note; the assertion that the area of the 350 nm absorption peak observed during deprotection does not change depends on the assumption that the absorbance does not saturate at any point. For large scale synthesis programs, early in the program when the DMT release rate is at its highest, the absorbance peak can saturate. Under these circumstances, unless the area is dropping precipitously from one cycle to the next, efficiency of coupling can be difficult to judge. If the area appears to decrease, but the height remains the same, this usually means only that the DMT release rate has increased rather than decreasing, which can happen in the first five or six base addition cycles. If the absolute height of the 350 nm peak begins to decrease while the time duration of the reaction remains the same, this means something is seriously wrong with the coupling efficiency.

The requirement that the Deblocking solution be especially dry is debatable. We typically add molecular sieves to the Deblocking solution reservoir, but efficiency of the oligomer yield does not seem to be notably improved by adding in a drying latency for this solution prior to use. Deblocking solution can be made within a day or so of synthesis and the reservoir can be topped off during the synthesis process on need if it is about to run dry, all without any apparent effect on yield. For long synthesis programs --like synthesizing a DNA 30mer-- one can expect to top-off the Deblocking reservoir at some point during the program since the bottle on the

machine is just not large enough to contain sufficient Deblocking solution to facilitate all possible scenarios. That it need not be especially dry is perhaps convenient. As previously mentioned, there is a minor cross reaction between Deblocking solution and molecular sieves where the sieves will tend to accumulate a precipitate crust while sitting in the Deblock for a prolonged time. This observation suggests a dubious utility in adding molecular sieves to the Deblocking solution at all.

The combination of acids and volatile organics present in the Deblocking solution can be observed to make aerosols or vapor condensates when pouring new solution into the argon-containing environment of the reservoir bottle. Appearance of “mist” when topping off solutions is to be expected on this reservoir. In addition, the gas distribution manifold of the synthesizer will tend to show signs of contamination after prolonged use, likely from vapor back-filling of this particular chemical. Finally, when preparing this solution, be aware that residual pyridine on glassware used to prepare the oxidizer solution can apparently react with the acids in this solution, causing aerosols; I recommend using glassware that has been actively cleaned, rather than merely solvent rinsed, between preparing these two reagents.

Table A.4

Deblocking Solution (Detrit)	1L
Dichloromethane (Fisher Chemical Catalog# D37-4)	970 mL
Dichloroacetic acid (Sigma-Aldrich Catalog# D54702-2.5L)	30 mL
Add sieves to reservoir, but understand that it's probably not absolutely needed	

#### A.4.5 Preparation of DEA Wash Solution

The Diethylamine (DEA) wash is added as a one-time treatment at the end of the synthesis process following the final deblocking step to remove the DMT

protecting the last 5'-hydroxyl at the end of the oligomer (see Figure A.7). This gentle alkaline treatment removes the CE protecting group esterified to the phosphate of all phosphodiester in the oligomer backbone, rendering the backbone unprotected. The DEA wash is followed by a thorough acetonitrile flush prior to the end of the synthesis program in order to make certain that CE has been cleaned away.

DEA solution is another volatile solution which should be prepared in a hood. We add molecular sieves to this solution more or less out of habit, but I don't believe that it is essential here since no DEA is ever present at any time when it could decrease the efficiency of the dN-PA coupling reaction and since no successive chemistry after the DEA treatment has to worry about water content. The only reason to add sieves to the DEA reservoir is the outside chance that this helps to decrease the threat of water vapor present in the argon atmosphere of the synthesizer. No drying latency is needed for this solution; it can be made within a day of use.

DEA solution is not used in an especially large quantity, but there is a lengthy priming step for the bottle prior to delivering it to the reactor column. Around 60 to 100 mL of DEA will be used for one synthesis process. One batch of DEA solution will support three synthesis cycles with absolutely no dependence on the length of the oligomer.

Table A.5

DEA wash solution	200 mL
Diethylamine (Sigma-Aldrich Catalog# 471216-500ML)	40 mL
Acetonitrile	160 mL
Add sieve to reservoir, but not completely essential	

#### A.4.6 Preparation of Activator Solution

Activator solution is essential to the dN-PA coupling reaction and is responsible for making base monomers able to couple to the 5'-hydroxyl of a growing oligomer (see Figure A.6). As such, this reagent is extremely sensitive to water contamination.

Activator is added simultaneously to the dN-PA in the synthesis process with activator drawn by pump B and dN-PA delivered by pump A. The chemicals are mixed in-line and pushed into the reactor column, then recycled for 4 to 6 minutes across the reactor column in a closed plumbing loop to facilitate reaction. During this recycling period, one can notice a  $1 - e^{-\frac{t}{\tau}}$  type shift in the maximum pump backpressure as the free DNA 5'-hydroxyl reacts with activated phosphoramidite to form the trivalent phosphite and extends the oligomer off the solid support, making it more occlusive to circulating solvents.

Activator must be prepared in sufficient quantities to support every base addition reaction. For the 6.3 mL reactor column, this reagent is used at a rate of nearly 7 mL per base addition and the quantity used is automatically calculated as an excess over the molar equivalent chosen for the dN-PA in order to force the activation of the dN-PA. 150 mL to 200 mL is usually amply sufficient for most synthesis programs.

As noted above, this chemical is very sensitive to water contamination and must be especially dry. I'm convinced that a molecular sieve drying latency period involving this one chemical all by itself almost single-handedly improved synthetic persistence in my hands. Consider this chemical to be a major route for contaminant water and treat it appropriately. Any activator to be used in synthesis

should be mixed a minimum of one week in advance and dried on the synthesizer under argon. As there is a potential for this chemical reservoir to run dry during long synthesis programs (>20 bases) using the 12 mL reactor column, plan to make and dry excess of Activator under argon one week ahead for such programs.

Table A.6

Activator	100 mL
1H-Ethylthiotetrazole	7.74 g
Acetonitrile	100 mL
Add molecular sieves on the synthesizer (use an excess)	

#### A.4.7 Preparation of dN-PA Solutions

I've already detailed the importance of the dN-PA chemicals in Section A.2 and will support that here only with sufficient information to prepare the solutions for the synthesizer while planning a synthesis program.

As mentioned, there are four types of dN-PA for the conventional DNA bases and so four different possible reagent solutions. The usage of any particular dN-PA is dependent on the sequence of the DNA oligomer to be synthesized: if the sequence is a 4mer composed only of TTTT, one can prepare solution for the T reservoir and leave all of the other dN-PA bottles empty. Because this chemical is expensive overall, we must prepare solutions that are sufficient for a given synthetic program: at our particular scale of production, dN-PA is typically best acquired as a collection of bottled 2 g fractions where we dissolve the entirety of enough bottled fractions at any given time to cover only our short-term needs. After a dN-PA fraction has been dissolved in acetonitrile, you can't return it to storage; use it for synthesis and then dispose of the remains afterward.

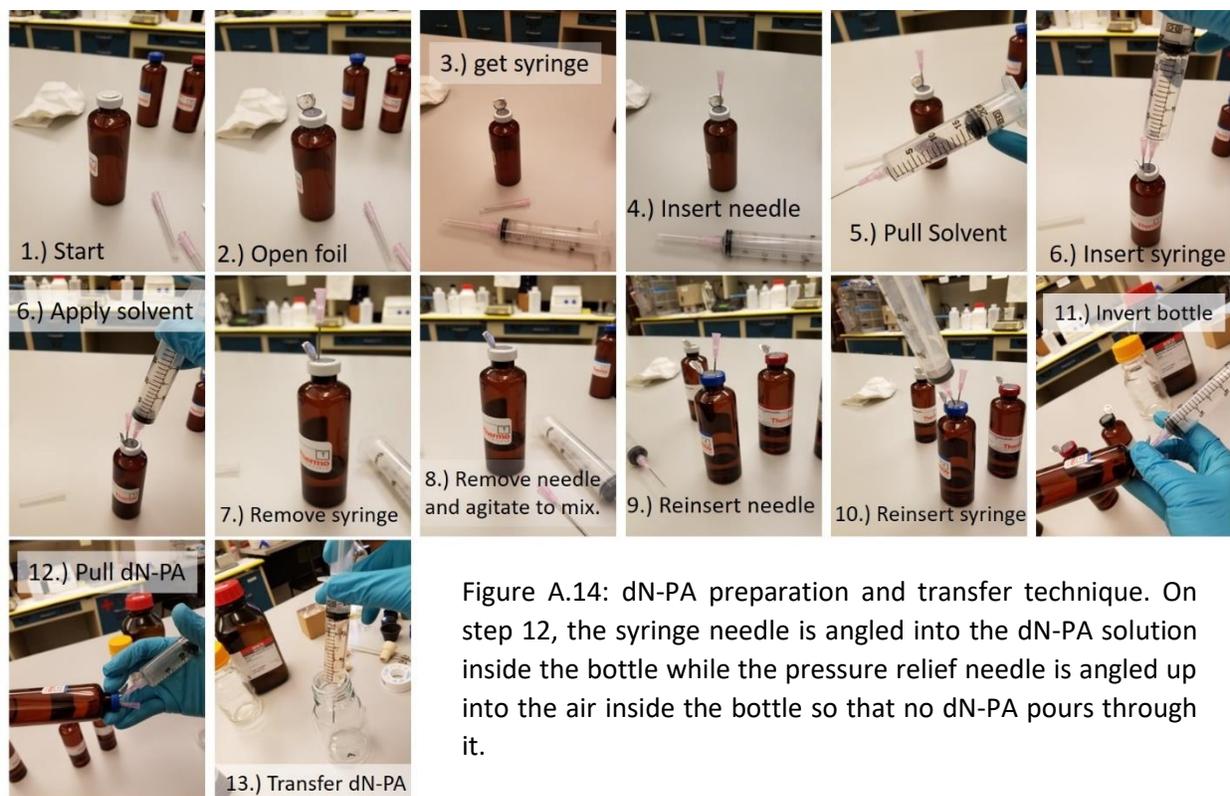
For our synthesis programs, we prepare the dN-PA to 150 mM. The following table contains the figures for preparing one 2g bottled fraction of each different conventional dN-PA.

Table A.7

iBu-Deoxyguanosine Phosphoramidite (dG) 2g fraction	150 mM
Dry Acetonitrile	15.9 mL
iBz-Deoxyadenosine Phosphoramidite (dA) 2g fraction	150 mM
Dry Acetonitrile	15.5 mL
iBz-Deoxycytidine Phosphoramidite (dC) 2g fraction	150 mM
Dry Acetonitrile	16.0 mL
Thymidine Phosphoramidite (T) 2g fraction	150 mM
Dry Acetonitrile	17.9 mL

Mixing acetonitrile with the dN-PA is its own procedure (Figure A.14). Each 2g fraction of dN-PA is typically contained in a sealed glass bottle with a rubber septum in the lid. Use a 20G needle with a 10 mL or 20 mL syringe to inject the acetonitrile through the rubber septum directly into the glass bottle. A second needle without a syringe should be simultaneously pricked through the lid in order to relieve gas pressure during addition of the volume of liquid. After the acetonitrile is injected onto the dry dN-PA, remove the needles and gently agitate the bottle (swirling is best) to dissolve the contents. Once the solid has been completely suspended, insert both needles back through the septum and carefully invert the bottle in order to draw out the dissolved reagent with the syringe while being careful not to spill liquid through the pressure relief needle. The mixed dN-PA is then squirted directly into the synthesizer reservoir and molecular sieves are added to remove any water that might be present. The dN-PA solutions should be mixed one week prior to their date of intended use as a drying latency. This chemical is not stable in the long term, but we have kept it at room temperature, dry with molecular sieves, under argon on the synthesizer for three or four weeks at a time

without seeing significant decrease in synthesis efficiency. As mentioned, mixed dN-PA cannot be restored to the freezer and must be disposed at the end of use.



#### A.4.8 An Example Planning Calculation for a Conventional Synthesis Program

I will demonstrate the planning of a typical synthesis program. The target sequence will be a typical liquid crystal forming nanoDNA self-complementary 12mer oligomer: Dickerson Dodecamer,  $5'CGCGAATTCGCG3'$ .

The first decision to make is the scale of the synthesis. As previously mentioned, we have two different reactor vessels available with about a 2-fold difference in size between them; 6.3 mL and 12 mL. The 12 mL reactor column can produce nearly a gram of DNA 12mer in a single synthesis batch with perfect synthetic yield and purification efficiency but uses twice as much dN-PA to do it. The 6.3 mL reactor can produce upward to 500 mg of DNA in a batch and uses half

the material, meaning that it can synthesize a DNA oligomer of around twice the length for the same quantity of dN-PA expended. We will suppose a smaller batch using the 6.3 mL reactor column.

The first thing to notice about the sequence of the polymer we intend to synthesize is that the first nucleotide added will be the 3' guanine. This is important because it determines our intended solid support. We must use a solid support decorated with guanine. As described in Section A.3.1, a suitable solid support is GE Healthcare Life Science's Primer Support 5G, dG version.

For this solid support, the loading density of dG is 350  $\mu\text{mol/g}$  and the support should be packed at a density of no greater than 111 mg of support per 1 mL of reactor volume. For the 6.3 mL reactor column:

$$6.3 \text{ mL} \times \left( \frac{111 \text{ mg}}{1 \text{ mL}} \right) = 699.3 \text{ mg} \approx 0.7 \text{ g}$$

From this, we would need about 0.7 g of solid support loaded into the reactor.

This choice indirectly establishes the quantity of dN-PA needed to complete the synthesis program. Most programs use 1.5 to 1.8 molar equivalents of soluble dN-PA for each mole of solid phase seed. The amount of seed is calculated from the loading density:

$$0.6993 \text{ g} \times \left( \frac{350 \mu\text{mol}}{\text{g}} \right) = 244.8 \mu\text{mol}$$

This is the absolute maximum possible yield of the synthesis. If 1.8 equivalents of dN-PA is used for each base addition step, the following amount is used at each base addition:

$$1.8 \times 244.8 \mu\text{mol} = 440.6 \mu\text{mol}$$

As dN-PA is prepared to 150 mM in our synthesizer preparations, each base will require:

$$\frac{150 \text{ mmol}}{L} \rightarrow \frac{150 \text{ } \mu\text{mol}}{\text{mL}} \quad 440.6 \text{ } \mu\text{mol} \times \frac{\text{mL}}{150 \text{ } \mu\text{mol}} = 2.94 \text{ mL per base addition}$$

So, each base addition cycle will use about 3 mL of dN-PA.

Neglecting the most 3'-G base, the sequence contains 3xG, 4xC, 2xA and 2xT. With a quick comparison to the volumes of dN-PA prepared for each different 2 g fraction (Section A.4.7), we can be fairly assured that one bottle each should be prepared for every base. The largest fraction needed will be the dC-PA, at 4\*3mL, or 12 mL total: since 2 g prepares 16 mL, there is 4 mL of headroom. It is a fair estimate to assume that 2 mL of dN-PA will be used by the synthesizer in priming each bottle, so 2 mL of excess gives sufficient room for the entire program. This is on the assessment that all extra dN-PA will be wasted following synthesis.

As a comparison, suppose that I had instead decided to synthesize a 16mer containing only T: sequence being 5'TTTTTTTTTTTTTTTTTT3'. We use 0.7 g of T-decorated solid support and 15\*3= 45 mL of T-phosphoramidite. As the T bottle prepares only 17.9 mL of solution, 45/17.9 = 2.5 says that 2.5 bottles are now needed, meaning that 3 bottles should be prepared.

The 12 mL reactor column roughly doubles these figures. For Dickerson Dodecamer, this would mean preparing 2 bottles each for G and C and 1 bottle each for A and T.

## A.5 Synthesis Process

This section is intended as a first approximation overview to the preparation and use of the in-house Oligopilot synthesizer to execute oligonucleotide synthesis.

Much greater detail about the synthesizer and the Unicorn program can be found from technical sources. If you're aiming to set up or run the synthesizer from the following instructions, be sure to fully read them first and re-read as necessary during preparation.

### A.5.1 Overview of Unicorn

Unicorn 5.11 is the program used to run the oligopilot synthesizer on our in-house system. To awaken the oligopilot, turn on the synthesizer by the switch at its base and its accompanying computer. As both boot up, instruct the computer not to error check its network and then open the program "Unicorn 5.11" and select "default" in the immediate "log-on" pop-up. The computer and the synthesizer should automatically sync after the synthesizer has run through a series of self-tests.

Unicorn will open four window tabs: Unicorn Manager, Method Editor (see Figure A.17), System Control (see Figure A.15) and Evaluation. Unicorn Manager is a file manager for the program system which is largely unneeded in day-to-day use. Method Editor opens and creates synthesizer scripts and is very important for planning how the synthesizer will execute its automation. System Control is the window which overviews the current operation of the synthesizer equipment and contains status icons for the equipment, a window for UV-absorbance read-out and a schematic of the oligopilot's plumbing to show what the synthesizer is doing at any given time –the bottles are all labeled, the valves are all numbered (V1 through V8) and the pumps are labeled "A" and "B." Finally, the evaluation window loads records of synthesis runs and allows analysis of what the instrument did after the fact. Method Editor is "Beginning," System Control is "Middle," Evaluation is "End."

When the computer and the synthesizer sync, the “Instruments” indicator in System Control (see Figure A.15) should switch from “Scanning” to say “Ready” and “Connection” should say “YES.”

#### A.5.2 Preparing the Oligopilot One Week Ahead of Synthesis

Advanced preparation begins roughly one week ahead in order to incorporate a drying latency for the synthesizer chemical reagents. In this step of preparation, reagent reservoirs must be cleaned and filled, placed under argon and set to drying by molecular sieves. Most operation of the Oligopilot is handled in System Control during this process.

The first preparative task is to check the argon tank and replace it as necessary. Open the valve on the argon tank and check that the pressure gauge on the door of the synthesizer jumps to between 0.2 and 0.25 bar. If the argon tank is empty, replace it with purified argon of the same type. If you unscrew the lid of any solvent reservoir bottle, it should make a soft, pressurized “pop!” and the argon tube attached to the bottle lid may “hiss.”

It is also useful to check the 20 L carboy waste reservoir sitting on the floor by the synthesizer to be certain it can catch the 5 L or 6 L of liquid waste that will be produced during a synthesis run.

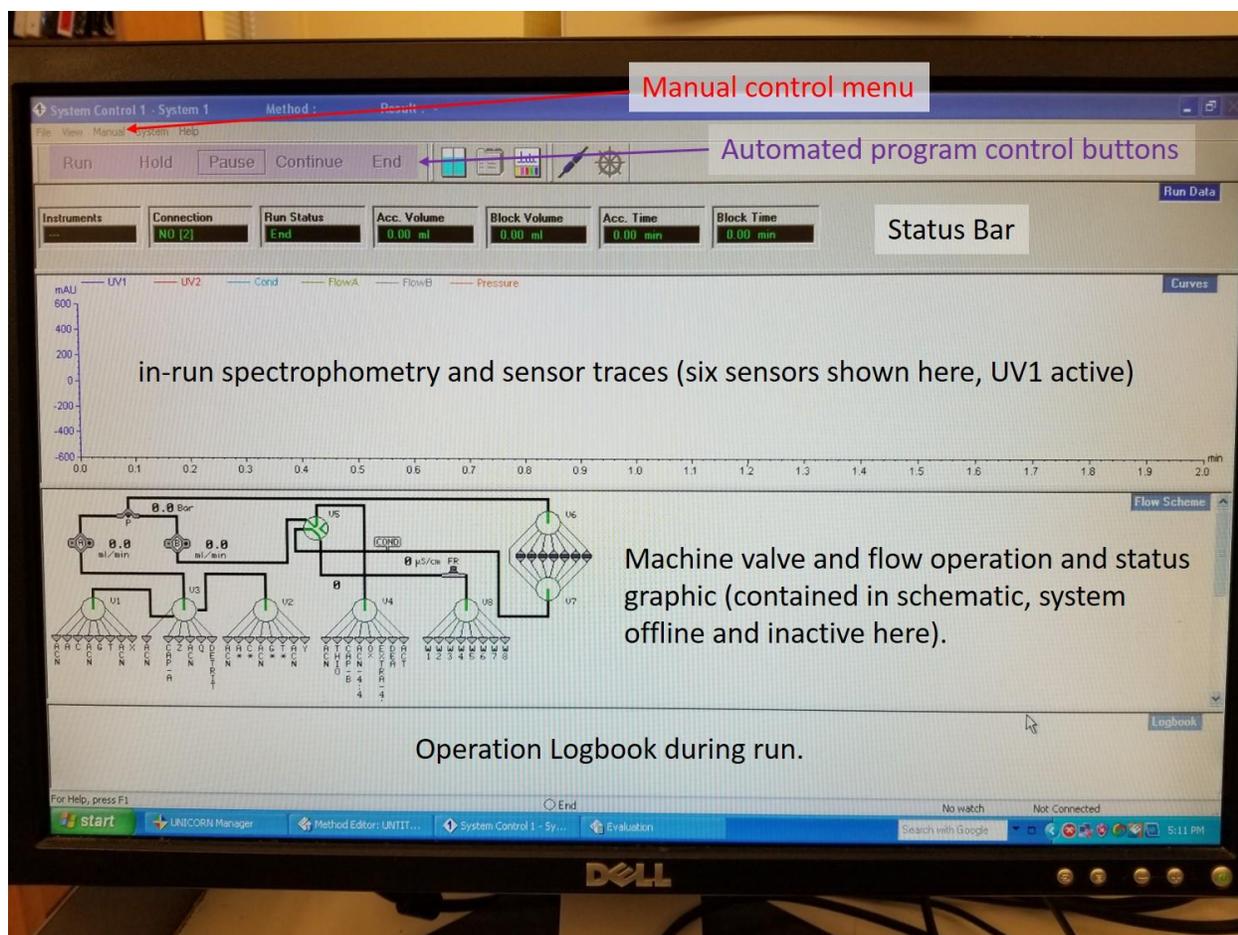


Figure A.15: System Control Screen. Shown in inactive state with major windows and features marked.

After the argon is added, the main solvent reservoir should be cleaned and filled with fresh dry acetonitrile up to 1 L or 1.5 L. Go to System Control (see Figure A.15) and select “pump” from the “Manual” menu. This will open the manual operation window for running the Oligopilot and it will be open to its pump instruction window –the pumps are controlled by “Pump” and the valves are controlled by “Flowpath” in the same menu. This instruction interface can be used to implement short program scripts as needed during manual preparation: a script is written by highlighting an instrument instruction in the available scrolling menu list, filling in its accompanying parameters and clicking “Insert” to place the instruction into the script display

window, clicking “Execute” will start the synthesizer running at those parameters. Priming the main solvent will require washing acetonitrile through much of the synthesizer system. For this action, one needs to give the pumps an instruction to run and set the valves to positions that deliver solvent to the desired lines: what follows can be considered a template for any priming or washing action on any line by switching flow rates and valve settings. For setting the pump flow, pick the Flow\_AB instruction, set both pumps to 10 mL/min, then Insert. To set the valves, click the Flowpath instructions, click “Solvent\_A” then select “ACN\_Amidites\_3.1” in the parameters box and select Insert, then select and Insert “ACN\_Reag\_4.1” from “Solvent\_B” in the same manner (“Amidite” contains all the valve settings for the dN-PA reservoir bottles which are on pump A, “Solvent\_A” contains valve settings for all the chemical reagent bottles connected to pump A and Solvent\_B contains settings all the remaining reagent bottles, which are connected to pump B). Assure that the valve path is set to the appropriate waste outflow by picking Waste, then Inserting a flowpath of “Waste\_1” (the “Waste” menu contains valve settings for possible additional waste paths, of which we only really use one, while the “Column” menu contains settings for selecting a reactor column flow path: Waste and Column are both down the line after solvents are mixed and near the end where outflow from the synthesizer occurs.) With these commands, pumps A and B will each begin running at 10 mL/min, pulling from the acetonitrile reservoir and pumping it through the main acetonitrile lines for valves associated with both pumps. Upon execution, the solvent will flow automatically from the reservoir bottles, through the column bypass (the default reactor column position) and to the waste\_1 receptacle, which is the 20 L carboy on the floor.

For the initial start-up, as the pumps are running, open the door on the synthesizer and, one-by-one, open and close the plastic purging valves on the sides

of each of the pump heads in order to vent air bubbles in the system (Figure A.16). If the synthesizer has sat empty for a long time, the lines may be full of air bubbles. The pressure trace should be a somewhat stable sinusoid echoing the pump-head strokes and it will contain a low pressure gap if the pump has a big air bubble in it and can't prime enough to move solvent –briefly opening the screw valve should release the air and be met with a small splash of solvent when the pump is running. After this initial wash, the pumps can be used independently as needed by running the command Flow\_AB with the pump speed parameter set to 0 mL/min on the pump that is desired to be turned off.

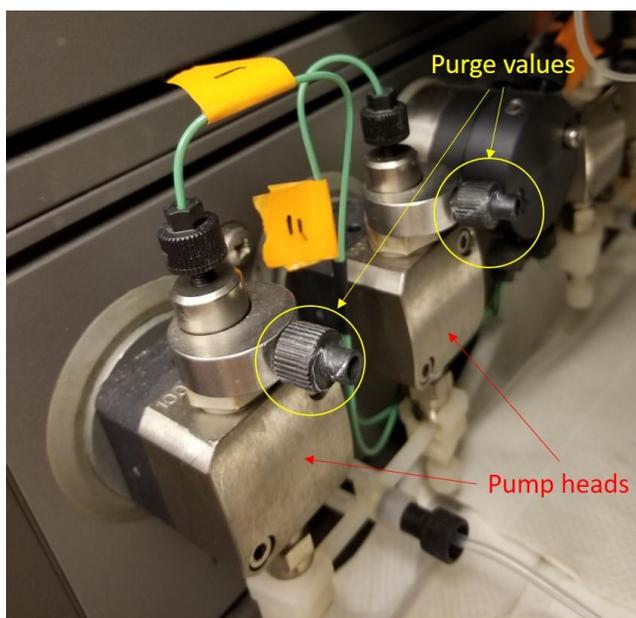


Figure A.16: Purge valves on the pump heads. Both pumps each have two pump heads where each pump head has a purge valve.

Wash each of the acetonitrile lines by instructing the flow to run through each different acetonitrile port on all the needed reagent bottle valves: mostly just ports on V1 and V4, which are found with Amidite, Solvent\_A and Solvent\_B instruction menus. Acetonitrile lines can be located on the system schematic by looking for “ACN” among the bottle labels on screen where at two to three ACN lines are located on each valve. This many acetonitrile ports on each valve allows the system to switch the valve in such a way while running synthesis that a solvent

acetonitrile wash can always occur between when one chemical on the system can come into contact with another in the same valve.

While the pumps are running, the “Curves” block displays spectroscopic traces of the solvent. Washing often results in absorbance peaks propagating through the system as contaminants let loose from the lines: you should wash until the absorbance trace on a particular line has stabilized, typically after 3 or 4 minutes. The synthesizer can be instructed to stop running by clicking the large “End” button up on top of the System Control page and the system will reset to its resting state.

After the ACN lines have been washed, it's time to fill the chemical reservoirs with the appropriate reagents. The process for filling a chemical is to pick a bottle and unscrew it from its line on the synthesizer, empty any remaining storage solvent from it, dry that bottle by evaporation with house pressurized air, mix the appropriate chemical for that bottle from Section A.4 and fill that chemical into the bottle, transport the bottle to the synthesizer (capped to keep from breathing the chemical), open the bottle and immerse the solvent line in the bottle, add any needed molecular sieves to dry the chemical, then screw the bottle cap closed and prime the line. The bottle is primed by running the pump on that chemical line until the spectroscopic trace spikes as the chemical reaches the valve and ultimately progresses down line to pass through the spectrophotometer sensor.

There is a special priming operation for dN-PA bottles (which I will call “amidite bottles” in the language of the synthesizer) because the lines are somewhat lower volume. To prime an amidite bottle after filling it, immerse the line and screw the bottle cap closed. Set pump A to run at 10 mL per minute on the amidite line of choice; when you execute, the pump will begin to run... count six seconds and click End to stop the run... the pump should stop. Then, set the pump A to run again at

10 mL per minute, but position the valve on the Acetonitrile port *next to* the amidite line you're trying to prime (in the Amidite menu, these ports are called ACN\_A/X for dA-PA, ACN\_C/G for dG-PA and dC-PA and ACN\_T/X for dT-PA.) After executing, within about 1 min, you should see an absorbance peak pass through the spectrophotometer –this is the priming flush of the dN-PA you tried to prime. The reason for running this procedure is to prime a minimal volume of dN-PA from the amidite bottle in order to not waste any dN-PA prior to synthesis; the spectrophotometer is a considerable volume downline from the valve, making it possible to pump out most of the chemical from the amidite bottle prior to seeing the trace.

Be especially careful when adding dN-PA to the amidite bottles. If the bottle lid is not completely screwed closed, acetonitrile will gradually evaporate out of the bottle and the dN-PA can crystallize inside the bottle and line, plugging it.

With the chemical reservoirs all cleaned and primed, the synthesizer can be shut down for the next few days to allow for the molecular sieves to do their work and dry the chemicals. Shut down the Unicorn program from the Unicorn Manager screen. The System Control tab will flash orange and prevent the program from closing. Open the System Control screen and answer the question about locking the synthesizer to simply keep it unlocked. The four tabs of the Unicorn program will then close. The Oligopilot can be turned off now by flipping the switch at its base and the computer can be turned off by normal means.

During the time between setting up the chemicals and the intended day of synthesis, the argon gas supply should be checked routinely to see that excessive gas is not being leaked by the system. The Oligopilot will remain under argon as long as dry chemicals are present in the system.

It is sometimes useful to clean the reactor column ahead, though this is not time sensitive.

### A.5.3 Preparation on the Morning of Synthesis and Use of the Method Editor

If all the chemicals have been filled into the synthesizer, primed and allowed to dry, very little work must be done the morning of synthesis to prepare the equipment. It should just be ready. The only new task prior to setting up the automation is to prepare the solid support. At this time, the planned scale of the synthesis should have all been calculated as in Section A.4.8 in order to determine the quantity of solid support to load into the reactor column. Load the reactor column with the desired quantity of solid support and screw the reactor closed while being careful to keep the screw threads clean. Weigh the reactor to check its initial mass; through the course of the synthesis, the reactor will gain mass and a difference in mass at the end of the experiment is a good coarse grain check that things are working. Top off the acetonitrile solvent reservoir: a good rule of thumb is 1 L per four bases of planned synthesis and while this is a little in excess of what is actually needed, it gives a safety buffer.

Mount the reactor column in its holder on the side of the oligopilot (see Figure A.9) with tubing running to it both to top and bottom. With an adjustable wrench, screw down the bars holding the line in place against the reactor (only 1/4 turn past finger tight is enough). You will need to check that the lines here do not leak in the next step. Depending on which reactor column is in use, the mount may have needed switching; these can be removed and screwed back on but make certain that the line going to the top remains the line on top.

Using the manual control in the System Control screen, as during the prior set-up in Section A.5.2, set both of the pumps to flowing at 10 mL/min from the main acetonitrile ports on the Solvent\_A and Solvent\_B menus. Use the “Column” menu to set the flow path through “Column\_1,” which is where the reactor was just installed, and on into “Waste\_1.” On execution, acetonitrile will flow from both pumps through the reactor column and to the waste. If this is the first synthesis of the day, I always briefly purge any air bubbles from the pump heads by briefly opening the purging valves on the pumps. As the pumps are running, they are now pushing acetonitrile into the reactor column with the solid support: a large air bubble will be visible pushing out of the line from the top of the reactor, but it should quickly be replaced by solvent. Press a paper towel against the seams and screw fittings attached to the reactor to make certain that solvent is not weeping out: the reactor should not leak. If the reactor is leaking, stop the pumps and check that all the fittings associated with the reactor are appropriately tightened.

A brief peak of absorbance should leave the reactor column while it is being washed. As before, wash until the absorbance stabilizes. Then, turn off the pumps. It's time to set up the automation.

The automation of the synthesis is set in the Method Editor (Figure A.17). Assuming the solid support of choice is a basic high-load support like Primer Support 5G 350, the method can be built by going to the File menu and clicking New. This will bring up the method wizard. All templates in the method wizard contain information about their date of inception and their intended uses and parameters. For basic synthesis on the solid support mentioned above, highlight the template “RECYCLE 6mL PS5G AKOPc121 Ed 005” and click “OK.” This should open the template in the “Run Setup” screen with the “Sequence” tab highlighted (see Figure A.17). The details of the protocol can be tweaked with the various tabs

available in the Run Setup screen. Some settings are annoyances that can be eliminated in future use, but a few are essential to getting the synthesizer to run. This will focus only on the essentials.

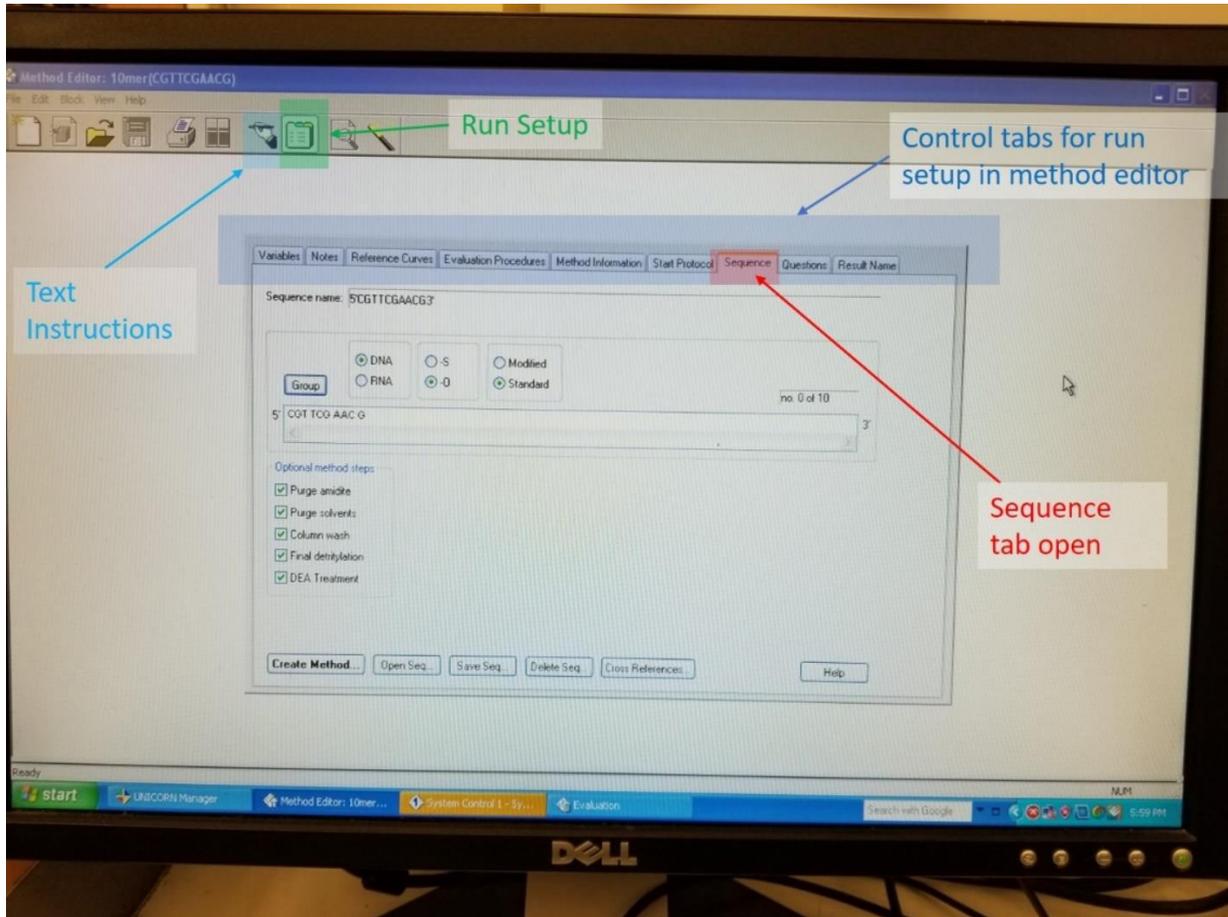


Figure A.17: Method editor: Important highlight tabs of the method editor.

On the “Sequence” tab, it should be possible to enter the sequence of the oligomer you intend to synthesize by replacing the sequence written in the available block. The sequence will have a dummy name up top, but ignore it. The box for “DNA” will be filled, as is the box “-O” for normal phosphodiester, and “Standard.” All optional steps below will be checked. The optional steps can be turned on and off depending on the synthesis plans. If this is the second synthesis of the day and the chemicals have already been purged and used, “Purge amidite” and “Purge solvents”

may be unchecked. If the intent is to use DMT-on purification (see Section A.6.3.2), uncheck “Final Detritylation.” If the intent is to perform chemistry with the final hydroxyl (as in Section A.8.4) uncheck “DEA Treatment.” The option “Column wash” should always be checked and can result in a polymer truncation if left unchecked.

Click the “Variables” tab at the far left of the run setup window. This will contain a spread sheet for all of the synthesizer script variables, controlling literally everything the synthesis script will do. Most of the variables need not be adjusted. Write-in the values for the Column volume, the weight of the support (from when you loaded the reactor), and the loading value of the support. The method automatically contains the expected parameters for the Primer support 5G with 350 umol/g loading in the 6.3 mL reactor column. Set the amidite concentration “Conc\_Amidite\_DNA” to 0.150 for 150 mM. For longer oligomers, the recycle time can be lengthened and the amidite equivalents “Eq\_Amidite\_DNA” can be adjusted.

Return to the sequence tab and click “Create Method.” The method editor will ask to save the sequence to the sequence library. I input the sequence bracketed by 5’ and 3’ as the name since most of our oligomers are short and click “OK.” The editor will then ask to save the method, which I always save under the “Mark” directory with a name that includes the length of the sequence and the content of the sequence [e.g. 10mer (CGTTCGAACG)] and then click “OK.”

After the method has been created and saved, check that the script for the automation is correct. This can be done by clicking the icon for “Text Instructions” which is shaped like a little hand with a pencil. This will open the script for the synthesizer program. The script will be a series of blocks that are to executed in

order from top to bottom. The first blocks will be a series bottle purges which ready the synthesizer for its run followed by a block called "Block Column\_wash." Always make certain this block is present. The blocks below that mark the addition of each nucleotide in the requested sequence followed finally by any requested deprotections. Nothing should need to be changed in this screen unless the method is a universal linker solid support method (see Section A.8.3).

At this point, the synthesizer is ready to run.

#### A.5.4 Running Automated Synthesis on the Oligopilot Synthesizer

All chemical reservoirs are filled and dried as needed and the automation has been constructed. What remains now is to run.

Go to the System Control screen (Figure A.15). Open the file menu and select "Run." The automation method created in Section A.5.3 will be located in the Mark directory. Highlight the method and click "OK." There will now be a last chance to make certain everything will run as intended. Check the variables and click "next," check the text method then advance by "next;" add start notes, then click next; ignore the request to enable printing the method in "Evaluation Procedures" then click "next;" check that the sequence is correct then click "next;" record any variables that seem pertinent in the "Questions" blanks then click "next;" Check the result name and change it as desired. (Many of the actions represented here can be suppressed in the construction of the method in Section A.5.3 by turning them on and off in "Start Protocol" under "Run Setup".) Finally, past the paperwork, click "START."

At this point, the synthesizer will take its own recognizance and begin to work through the program script. If everything has been set up to perfection and sufficient amounts of each reagent are present in the system, the operator really needs to do very little. This is the ideal. In reality, the equipment can stand a little babysitting to make certain everything is progressing as desired.

The synthesizer will self-zero its spectrophotometer and run through the included bottle purges. Following the last bottle purge, it will wash the column and begin with the first detritylation action to remove DMT- from the solid support. It monitors the outflow for a high absorbance peak at 350 nm and will pause synthesis and trigger an alarm if no peak is detected. If everything has been set as instructed, this alarm should not occur. While detritylation is taking place, the line exiting the reactor column will contain a strong red color which can be seen directly. The synthesizer will report the area and height of this absorbance peak in the activity log and I keep a written record of these numbers from one addition cycle to the next to monitor efficiency. After detritylation, the machine injects dN-PA with activator into the reactor and begins the recycling wash, recirculating dN-PA within the system for several minutes to promote the coupling reaction. Monitoring the pressure change during this time can help track the progress of the coupling. The machine then washes out the reactor column and performs the oxidation reaction to transform phosphite to phosphate and then caps failed couplings. After roughly 19 minutes with the basic settings, it is ready to return to the detritylation step and begin the next nucleobase coupling.

The instrument compares detritylation peak height and area between coupling steps and reports a coupling efficiency number based on this observation. My experience with coupling efficiency is that it can be masked by saturation of the

spectrophotometer and that sometimes the intensity of the detritylation reaction is more stable than the system reports simply because it can't see the actual shape of the peak. This is especially true for the first ten couplings. However, the shape of this peak will change as the synthesis advances and the growing oligonucleotides begin to affect the rate of the detritylation reaction, slowing it down. This will widen the profile of the peak and eventually bring its apex out of saturation. Through the first ten bases, the peak area will appear to decrease, then increase as the detritylation reaction lengthens and the apex of the peak comes out of saturation.

Given a typical 242  $\mu\text{mol}$  scale, if coupling efficiency is truly poor, it will be reflected most consistently by a decrease in the maximum height of the detritylation peak by about the 6<sup>th</sup> nucleobase addition. This implies that the peak has come out of saturation more quickly than usual and is accompanied by a gradual decrease in detritylation peak area. The strength of peak saturation is very strongly dependent on the scale of the synthesis program and this observation depends very much on reactor loading. Among the reasons for reduced efficiency that I've encountered are water contamination, underloading of solid support and underloading of dN-PA equivalents during the coupling reaction.

For longer synthesis runs, as the detritylation reaction lengthens, the termination trigger for the cessation of detrit flow can sometimes become too soon. Typical termination point in the program is set as 250 mAU on the 350 nm absorbance band. As the tail lengthens on this reaction, sometimes a significant portion of the peak can occur as the absorbance drops past 250 mAU. If this is occurring during a synthesis run for a longer oligomer, while the synthesizer is integrating this spectroscopic peak, click the "Hold" button to force the synthesizer to hold its action before the peak drops through 250 mAU and persist until the tail

decreases below 25 mAU, at which point pushing the “Continue” button will resume the normal program. By terminating prematurely, the synthesizer can introduce a portion of N-1 failures by simply not giving the population enough time to totally deprotect. This can be preempted in writing the script method (in Section A.5.3) by changing the trigger point from 250 mAU to 25 mAU before ever beginning synthesis (but this is unnecessary for synthesis of 12 bases or less).

Also for longer synthesis runs, the nucleobase coupling reaction may not go to completion during the duration for the recycling step. This can be judged by watching the backpressure during the recycle: the pressure will tend to increase to a saturation point. If the pressure is short of saturating when the recycle is due to terminate, simply click the “Hold” button to prolong the recycle, then “Continue” when the pressure has apparently saturated. Again, this can be preempted by increasing the recycle period when creating the synthesis method.

One of the greatest risks during synthesis is the possibility that some chemical reagent or the main solvent reservoir will run out before synthesis is complete. Monitoring the bottle levels can be of benefit through the course of the synthesis program in order to avoid running out. This is not usually a problem with shorter synthesis runs, but is almost inevitable for synthesis of longer oligomers. The only way to remedy this is to make more of the chemical which is about to run dry and to refill the bottle –if this ever happens to the activator, coupling efficiency will take an efficiency hit for lack of a drying latency. In order to refill a bottle while the synthesizer is running, wait for a wash routine to begin, click “Hold” to prolong the wash until no more absorbance is leaving the reactor column, then click “Pause.” The synthesizer will stop where it is without terminating the script. Open up the bottle, refill the chemical in question and add molecular sieves as necessary,

dry the lip of the bottle with a paper towel, then close it again. After the chemical is refilled, click the “Continue” button. The synthesizer will then resume its script from where it left off.

After the synthesizer has completed its coupling blocks and fully synthesized the oligomer, it will perform a final detritylation reaction (if requested) followed by a DEA treatment to remove the CE-groups protecting the phosphates (also if requested). These two deprotections will typically be the stopping point for a synthesis run. After completing the DEA treatment, the synthesizer will perform a final wash and then sing a little chime to announce its completion. The reactor column can then be removed and taken to dry: purification methods discussed in Section A.6.1 become valid. The reactor can be weighed after it has been vacuumed dried (as in Section A.6.1) to determine the change in weight during synthesis. For a 12mer at 242  $\mu\text{mol}$  scale, this should be ~500-700 mg.

The complete record of a synthesis run can be loaded and examined in detail in the Evaluation program of Unicorn. I won't detail operation of Evaluation because it is relatively straightforward to figure out.

#### A.5.5 Cleaning the Oligopilot

After synthesis is completed and services of the oligopilot are no longer required, it should be cleaned. Cleaning is a bit of a reversal of the steps detailed in Section A.5.2. All chemicals remaining on the synthesizer except for the acetonitrile solvent should be disposed of.

To clean a bottle, remove the bottle from the synthesizer and dump its contents directly into the waste carboy. The remaining molecular sieves should be disposed of as a solid waste and not in the carboy (don't mix waste streams!) The bottle should then be rinsed with methanol, and this also be dumped to the waste carboy. Finally, add a small amount of methanol back to the bottle, replace the bottle to its position on the synthesizer and manually operate the synthesizer to prime the line with methanol until the UV absorbance trace stabilizes, signifying that the chemical reagent remaining in a given line has been expelled to waste. Methanol can be left in the line for storage.

Of the chemicals present on the synthesizer, pay special attention to the dN-PA and the activator. Both of these chemicals were prepared from solid and cannot be evaporated away. These bottles should be rinsed with acetonitrile several times prior to adding methanol in order to assure that residual chemical has been cleaned away. These lines can also be briefly primed with pure acetonitrile to help remove residual chemical and help protect the system from either of these chemicals crystallizing and plugging lines. Some attention should also be paid to the oxidizer because the chemicals in this bottle do not mix well with methanol and several rinsing steps are needed to fully remove them.

After the bottles have been cleaned and chemicals, aside for solvents, fully removed from the system, the argon tank can be closed.

When the Oligopilot is in storage, a number of small parts can be cleaned in a more thorough fashion if desired. The pump heads can be removed and cleaned by sonication, or replaced if they are getting old and the system is showing signs of pumping far less than the sensor would claim. The valve liners can be removed to

be cleaned or replaced. The absorbance flow cell can be replaced. The argon gas manifolds can be removed from the system and cleaned: this is periodically necessary because chemical vapors can diffuse through the argon lines and react with each other to form buildups inside the argon manifold that will eventually affect gas distribution to the bottles or the pressure releases that enable the machine to hold the desired gas pressure.

## A.6 Purification Methods

I will assume a basic background in common purification methods for this section. What will be presented here are various methods used for the different DNA oligonucleotide projects I've attacked.

There are two broad classes of DNA oligonucleotide purification strategy that are dependent on a choice undertaken during synthesis. In Figure A.7, step 'a,' the final DMT protecting group covering the 5'-terminal may or may not be removed in the program. Because the DMT group is a big aromatic object, it can serve as a hydrophobic purification tag to help separate desired oligonucleotides from failure sequences. The choice of leaving DMT attached at 5' during synthesis does not effect subsequent deprotection steps. The class of purification methods which anticipate DMT remaining as a cap at the 5'-terminal of an oligonucleotide are called DMT-on purification. On the other hand, removal of DMT, as demonstrated in Figure A.7, are called DMT-off. My early work focused on the former, while my later work, as my DNA synthesis process became more efficient, turned to the latter. I will talk about both.

### A.6.1 Post Synthesis Processing

This method is common to nearly all oligonucleotide sequences with little or no variation. Where the DNA synthesizer concludes its programmed action is not quite at the end of the process shown in Figure A.7. The synthesizer automatically performs Figure A.7 steps 'a' through 'c,' while step 'd' is left to bench work.

Immediately after synthesis concludes, oligonucleotides remain bound to the solid support with base protections still in place. For DMT-on purification, the DMT protecting group remains bound to the 5'-terminal of the oligo, or is removed for DMT-off. The presence of DMT does not alter what follows. The reactor column is removed from the synthesizer and dried for ~30 minutes by vacuum pump at room temperature. Dryness is typically judged by watching that the weight of the reactor decreases until it finds a stable point. After dryness is reached, the reactor is screwed open and the solid support, still bearing oligonucleotide, is transferred to a 50 mL conical tube, where it can be temporarily stored. In the case of immediate storage, this conical is kept at -20°C until the next phase of purification. There is no clearly defined point when this storage must end, so it can be a convenient stopping point.

Purification continues by completing step 'd' in Figure A.7. 40 to 45 mL of 28% Ammonium hydroxide is added to the 50 mL conical tube and the solid phase is agitated to disrupt it. A small stir bar is added to the conical to help keep the support grains suspended and the conical is incubated 16 to 20 hours at 50°C with

stirring. This treatment deprotects the nucleic acid bases and separates the DNA from the solid support, rendering it totally soluble.

At the end of the incubation, the conical is cooled gradually to room temperature. If one opens the tube too hastily while the mixture remains warm, the ammonium hydroxide will boil. After the tube is cooled, the mixture is suspended by agitation and then poured into a medium grain glass filter frit, where a vacuum is then applied to pull the solution through into a sidearm flask. Solid support grains will not pass through the frit and are filtered out. We rinse the conical with 4x50 mL of 50% ethanol which is then washed through the layer of solid support grains trapped on the surface of the frit. We next wash the solid support in the frit with 6x50 mL of water. The volume in the sidearm flask will increase to around 540 mL. After the solid support has been completely washed, we sample 1 mL of the supernatant for yield characterization purposes.

After completion, the wash, which now contains soluble oligonucleotide, is transferred to a round bottom flask and mounted on a rotary evaporator. The heat bath on the rotovap is set at 50°C and the condensation coil at -4°C. The solvent is removed from the sample with spinning on the rotovap.

My experience has been that the rotovap is perhaps not the most suitable method to use here, but it is the available equipment. The difficulty with the rotovap is that if one pushes the oligonucleotide sample completely to dryness, some of the dried sample will not ever return to solution, harming overall purification yield. After pushing completely to dryness, when solvent is added to dissolve the precipitate, some will remain precipitated and must be filtered out later. My speculation is that the rotovap can bleed off counter-ions that are essential to DNA

solubility, where the counter ion present at this point is ammonium ion, which is not absolutely stable given its equilibrium with ammonia. In some rare cases, I've been able to restore solubility by adding a small amount of sodium hydroxide. The aim then is to rotovap to small volume. By not going quite to dryness, no insoluble material is observed. I typically rotovap from 540 mL down to ~5 mL.

The small amount of liquid obtained here is then passed to another purification step.

#### A.6.2 Desalting by Isopropanol Precipitation

If synthesis has been exceptionally clean, producing few failure sequences, it is possible to conclude purification with the isopropanol precipitation presented in this section immediately after the processing outlined in Section A.6.1.

The isopropanol precipitation has ended up being the final general purification step in all purification methods I've used because it invariably produces DNA oligonucleotides that are capable of forming liquid crystal (LC) if such is possible. On completion of many purification techniques, DNA which should otherwise make LC phases, such as Drew-Dickerson Dodecamer, fails to make phases, necessitating some additional step in order to produce material suitable for experimentation. The technique described here has been the most reliable at restoring LC phases.

It may seem contradictory, but this technique is not the most stringent or effective cleaning method; it always leaves some excess salt with the DNA oligomer.

In instances where I've tried to produce oligomers that are of better purity, I have frequently observed that LC formation becomes compromised. This particular technique produces DNA that is not quite pure so much as it is stable. The residual salt impurity produced by this technique has not been an impediment.

Be warned that this technique is not efficient. One can anticipate losses in purification yield as high as 50% on this one technique. The requirement that synthetic material be capable of LC formation, and therefore useful to experimentation, has so far outweighed the need for high yield in my work. Materials lost in this step can be recovered by other methods, but generally at an overall penalty of purity.

The isopropanol precipitation begins typically with around 5-30 mL of oligomer dissolved at a concentration of ~20-50 mg/mL. The volume of the solution is measured and sufficient crystalline sodium chloride is added to produce a salt concentration of 400 mM. Pure isopropanol is then gradually added until the solution precipitates; this usually occurs when the isopropanol concentration reaches ~50% (there is variation here which I will detail). The precipitated mixture is allowed to sit for ~20 min at room temperature to reach solubility equilibrium, and is then centrifuged at 14,000 rpm for 10 minutes in an SS-34 rotor to sediment. Supernatant is decanted, yielding a pellet at the bottom of the centrifuge tube. The pellet is washed with several milliliters of 70% ethanol and is centrifuged again at 14,000 rpm for 5 minutes to sediment before pouring off the ethanol. The pellet is resuspended in 1 - 2 mL of water. If this is the last step of preparation, the resulting solution may be dried, preferably by lyophilization.

There are some additional notes to relate about this purification technique.

First, oligonucleotides purified by isopropanol precipitation have shown signs of drift toward acidic pH upon later rehydration. This is likely due to incomplete counter ion substitution.

The idea of this technique is that sufficient sodium is introduced that the oligomer will pair with the predominantly available counter ion to neutralize charges that make the oligomer soluble and the resulting complex will precipitate when the polarity of the solution is adjusted by an alcohol. I calculate that the number densities of sodium to the starting DNA counter ion (triethylammonium in some cases involving hydrophobic HPLC as the previous purification step, or ammonium if the prior step is A.6.1) range between 5:1 and 1:1 for sodium chloride added at 400 mM depending entirely on the length of the DNA oligomer: a 4mer oligomer has 3 charges, while a 12mer has 11 charges –the same concentration of a 12mer has approximately 4 times the charge. This is significant because of observations that NaCl added at concentrations above about 500 mM can result in sufficient salt coprecipitation during the isopropanol addition step that salt crystals appear mixed with LC phases during later experimentation. To assure that ammonium or triethylammonium are completely numerically out-competed by sodium, an amount of salt would need to be added that rapidly proves a physical for longer oligomer sequences, requiring that sodium simply be the better ligand, which is only imperfectly true. The volume of solution where the isopropanol precipitation is enacted should be at least great enough that the amount of salt added to reach 400 mM is better than 1:1 molarity with the number of charges present in a given DNA oligonucleotide. While materials purified by the isopropanol precipitation method reliably show LC phases, this charge substitution uncertainty is likely behind observations of pH drift toward acidic values at high oligomer concentrations. pH can drift if incompletely substituted ammonium ion, acquired in

Section A.6.1, volatilizes off of dried DNA by exchanging a proton to the DNA backbone and converting to ammonia, and thus yield DNA that is acidic when introduced to water. Oligomers that have passed through both Section A.6.1 and the isopropanol precipitation method show more stable LC phase diagram behavior than those which have only passed through A.6.1, but the pH drift effect is likely a symptom of incomplete removal of ammonium. If pH drift has a significant influence on an experiment and this needs to be regulated, my recommendation is to either dialyze against salt at a salt content in the dialysis reservoir equivalent to the charge content of the oligomer in the dialysis bag or to directly add sufficient NaOH to the oligomer to match the presumed (former) ammonium content of that oligomer. (This content can be calculated by examining the ratio of the different counter ion species present during the original precipitation reaction and calculating true DNA concentration using X-ray methods to be discussed later and hence actual ion content)

The second additional note is that the isopropanol solubility of oligomers in a salt solution is heavily influenced by the hydrogen bonding state of a particular oligomer with the surrounding solute.

This observation has a huge influence on this purification method in a situation that is very particular to DNA liquid crystal science. A synthetic DNA LC oligomer construct of interest is the palindrome (e.g. 5'GCGCTTAAGCGC3') because this particular type of sequence is self-complementary. In a self-complementary sequence, only one oligomer must be synthesized in order to create a molecule that can produce a base-paired duplex with identical molecules. And, when single-stranded oligomers base pair, the nucleotide bases of both oligos in the duplex are brought out of contact with the surrounding solvent medium so that they only have

hydrogen bonding interactions with each other and not the solvent. At 400 mM NaCl content for the oligomer concentration typically used in the isopropanol precipitation process, precipitation will tend to occur for self-complementary oligomer at ~45% - 50% isopropanol content. On the other hand, if the oligomer being produced is not self-complementary the sequence will not precipitate until as high as 70%-75% isopropanol content.

This hints at an exploit that can boost purification yield with oligomers that are not self-complementary. One other interesting construct for DNA liquid crystal science is the mutually complementary oligomer set (e.g. 5'CCGCAAACGCC<sup>3'</sup> with 5'GGCGTTTTGCGG<sup>3'</sup>) where both oligomers are needed to form a duplex and neither make LC phases alone. These sequences would be synthesized and purified in isolation and their overall yields will be lower than would be experienced with a self-complementary palindrome because they would not precipitate until significantly larger overall volumes for the same set-up, putting them at more unfavorable K<sub>sp</sub> equilibrium during precipitation. If you save the unprecipitated supernatant fractions from both of these sequences during their isopropanol precipitation steps (each ~65% to 70% isopropanol content) and add these fractions to each other, heavy precipitation is observed and the resulting precipitate can be collected and cleaned and seen later to make LC phases. In this, mutually complementary sequences synthesized in isolation are converted into a base paired duplex, which is less soluble at high isopropanol content, thus boosting recovery when two poorly recovered samples are added together and increasing yield of experimentally viable material.

A third additional note is that precipitation with oligomers is not always a flocculent solid, but can appear as oils.

For sequences with high T content in particular or the combination of T and A, precipitation induced during the addition of isopropanol results not in a solid, but in an oil. In this case, the precipitation is clearly a liquid-liquid phase separation where the oily phase can be harvested and shown to make LC under whatever circumstances where it forms a rigid duplex. These phase-separated materials tend to collect into oily layers which do not require hard centrifuging in order to clarify a clear phase boundary. Separating supernatant from “pellet” in this case requires greater care because the oily phase can be easily disrupted. Sequences that have been specifically modified to place a phosphate group at the 3'-terminal also tend to display oily phase separation during isopropanol precipitation and require the same care.

A fourth additional note is that the isopropanol precipitation reaction can be used to some extent to trade the counter ion species present with an oligomer.

Lithium or potassium can be exchanged into the isopropanol precipitation by simply substituting LiCl or KCl for NaCl in the precipitation reaction. Unfortunately, numerical balances will require this to be a dilution of the original counter ion rather than a complete replacement since the original ion always ends up an appreciable proportion of the mixture. To assure a large substitution rate, the desired counter ion should be added at a commensurately large numerical excess. This sort of exchange can be more successful for short oligomer sequences than long ones because it's easier to attain higher excess ratios for an oligomer with only a few backbone charges.

As a fifth and final note on this technique, be aware that smaller oligomers are more difficult to precipitate by this method than larger ones. A minimum

successful oligomer size was 4 bases, but at high loss of yield. In a case where Trimer oligomers were synthesized, this technique produced catastrophically high losses of yield, resulting in very little material for experimentation. Purification of synthetic dimer oligomers was out-sourced, but can be successfully purified on a flash column using an organic solvent system.

#### A.6.3.1 High Performance Liquid Chromatography (HPLC) of LC Forming Oligonucleotides (DMT-off Method)

Sections A.6.1 and A.6.2 can be used in tandem to produce suitably purified oligonucleotides for LC experimentation in absence of extensive failure sequence production during synthesis. There are however some instances where the oligomer may contain failure sequences or other impurities that could be optimally removed by adding another purification step between A.6.1 and A.6.2.

The most commonly used intermediate technique in my hands has been HPLC. There are two types of HPLC which are suitable to this task: oligomers can be purified by HPLC using either reverse phase chromatography (hydrophobic medium) or by ion exchange chromatography. Given the lack of a suitably high capacity preparative scale ion exchange column, most of my work has been with reverse phase chromatography using a Sepax GP-C8 column with 21.2 mm diameter bore and a 250 mm length, which contains an eight carbon aliphatic substrate bound in 10  $\mu\text{m}$  grains with 120  $\text{\AA}$  porosity. I used a binary buffer system with a polar solvent (buffer A) as 50 mM Triethylammonium acetate pH 7 (TEAA) and the apolar solvent (buffer B) as either Acetonitrile or Methanol.

Table A.8

Triethylammonium acetate (TEAA) 2M stock	500 mL
Glacial Acetic Acid	57.2 mL
Triethylamine	139.4 mL
Water	303.4mL
	Filter on 0.22 um cellulose acetate filter

2 M TEAA stock solution is made more or less by adding an acid to a base, so some caution is necessary since the mixing reaction is significantly exothermic. In the hood, mix by adding Triethylamine to the water with a rapidly spinning stir bar, then slowly and carefully add the Glacial acetic acid. The mixture will steam and will become very hot. Acid and base are not mixed directly to each other since the water serves to dilute one beforehand. With this formula, the pH should automatically be very near to pH 7, but small amounts of triethylamine or glacial acetic acid can be added to either raise or lower the pH as desired.

The 50 mM TEAA buffer A solution is made by diluting the 2 M TEAA stock solution 20-fold in HPLC grade water. If the pH drifts away from pH 7 on dilution, it can be adjusted by adding either triethylamine or acetic acid to raise or lower it. The buffer should then be filtered on 0.22  $\mu\text{m}$  cellulose acetate to remove any particulates (never use buffers on the HPLC that are not either suitably filtered or purchased HPLC-grade prefiltered; particulate contamination can destroy the column.)

Buffer B can be either acetonitrile or methanol. Even though Acetonitrile is the buffer of choice, methanol is cheaper and more purification can be done with it using less money, meaning that methanol is more useful when doing purification at the (large) preparatory scale necessary to support oligonucleotide synthesis. As a warning, methanol eats cellulose acetate filters, so purchase it prefiltered.

Sample completing Section A.6.1 arrives in this method with a volume of ~10 mL. Volume should be filled out using 50 mM TEAA to around 40 mL in order to dilute the sample sufficiently so as not to overload the column during injections. It should then be filtered on 0.22  $\mu\text{m}$  cellulose acetate because TEAA added to oligonucleotide sample from Section A.6.1 can produce ammonium acetate when the solutions are mixed, causing a particulate precipitate that complicates the HPLC. No precipitate should enter the HPLC. If the sample was rotovapped to dryness during Section A.6.1, it will always solublize with some precipitate and this will need to be removed anyway.

The workings of this method are that triethylammonium ion introduced with TEAA will exchange with ammonium ion present with the oligonucleotide from Section A.6.1 and suppress the charge of the oligomer's backbone with a semi-hydrophobic counter ion, making it possible for the oligomer to associate with the hydrophobic packing of the column. Charge suppressed oligomer is injected into the hydrophobic column and will stick to the C8 resin packing while the solvent polarity remains high in the presence of buffer A. The solvent flowing through the column is then gradually shifted by mixing buffer B into buffer A and gradually decreasing the ambient solvent polarity. Molecular contaminants with the oligomer sample are all anticipated to possess differing hydrophobicity and so become soluble at different degrees of apolarity while gradually changing the solvent and these contaminants become separated from whatever remains bound to the column. Eventually, the oligomer of interest elutes from the column and is caught by fractionation using a UV spectrophotometer to see when fractions of high absorbance depart the column with the solvent (Figure A.18). If there are many contaminants, the elution can display many equivalent absorbance peaks and you may have to go searching

through these peaks using a characterization method in order to find when the material of interest exited the column.

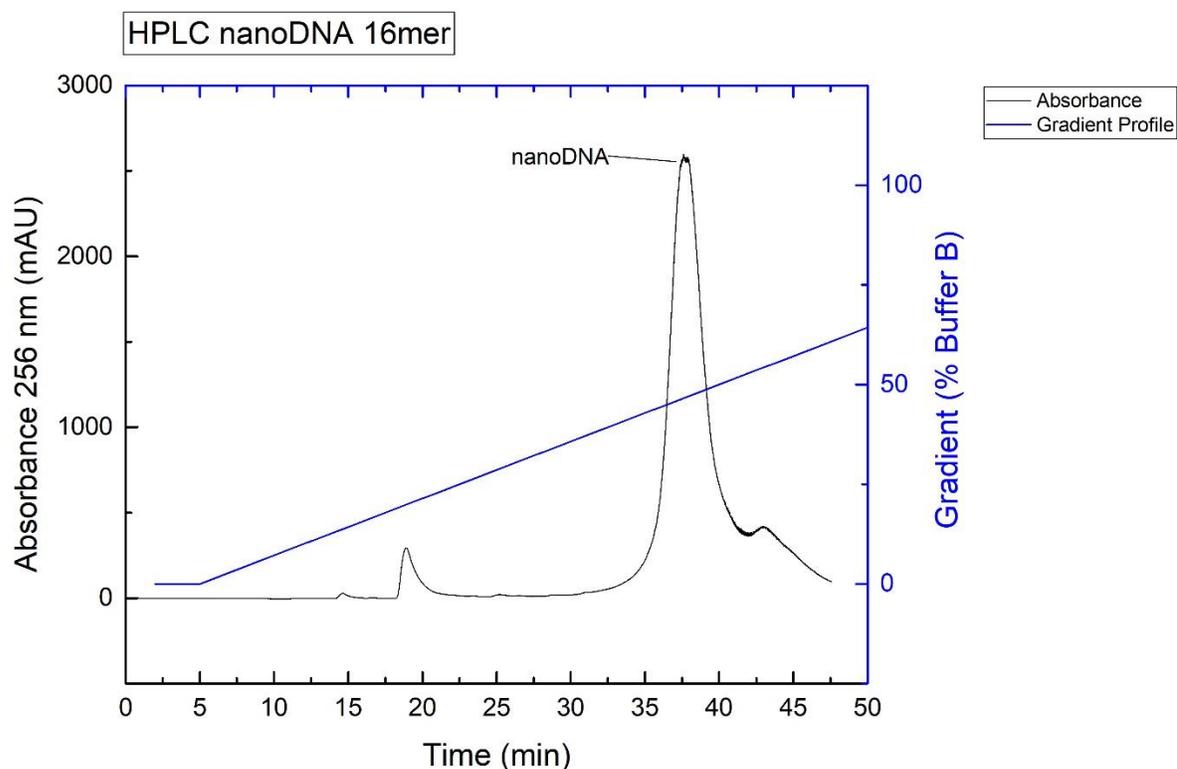


Figure A.18: HPLC trace of 16mer nanoDNA. UV absorbance of the hydrophobic column eluate (Black) relative to the elution gradient (Blue). The peak where DNA elutes is noted.

Because preparative scale HPLC column injections can overwhelm the spectrophotometer of the HPLC instrument, I frequently use a pilot injection in order to examine the contents of the sample without saturating the detector. This is more of a problem when using methanol as buffer B because it is more absorptive than acetonitrile in the 260 nm band where the oligomer absorbs strongly. The pilot injection is usually ~0.25 mL of sample where preparatory injections are ~8 mL. I perform the pilot injection to see where the sample elutes on the automated solvent A-B gradient, then I go to the preparatory injections to actually clean my sample knowing where to expect it to elute. As a second tale, column back-pressure will

tend to be high while running buffer A, then shift lower as the gradient increases buffer B content: if sample begins to elute from the column at some point on the gradient, the back-pressure can suddenly spike. I use a second spectral trace at 220 nm to help identify the oligomer peak since the 260nm:220nm absorbance ratio will tend to be different from the oligomer for molecules that are not oligomer –this ratio is automatically recorded by the instrument and can be examined independently using a clean oligomer as a standard.

My preferred gradient scheme pumps solvent at 20 mL/minute and runs 75 minutes; it runs 100% buffer A for the first 5 minutes during the void volume after the sample injection, then ramps gradually from 100% buffer A to 100% buffer B over the next 70 minutes. Oligomer samples typically elute on this gradient between 27 and 40 minutes depending on their length with longer oligomers eluting later. I frequently terminate the run after about 42 to 45 minutes to conserve solvent after the oligomer peak has eluted. Failure sequences typically lead the elution of the desired oligomer sequence as a series of gradually heightening peaks. Failure sequence peaks can look more significant than they actually are during preparative runs due to peak deformation at saturating absorbances and should be examined in a pilot run to judge their actual intensity.

Fractions containing the oligomer are pooled after they elute from the HPLC and are rotovapped using settings similar to Section A.6.1. These samples will never go to dryness, but exhibit insolubility when water is added to them if they are rotovapped aggressively. As before, I recommend not rotovapping to the absolute extent, but to continue on into the method outlined in Section A.6.2 after the sample is down to ~10 mL.

In using the methods discussed in Section A.6.2, I would recommend setting up a base volume that significantly favors NaCl (at least 5:1) over the

triethylammonium present with the oligomer after the run through the HPLC. Triethylammonium counter ion strongly prohibits the formation of LC phases and must be removed to the greatest extent possible in order to produce material useful in later experimental work. I had limited success trying to switch triethylammonium with ammonium in the rotovap by swamping the HPLC sample with 28% ammonium hydroxide, but I'm convinced at this point that such efforts are futility and that prolonging the rotovap treatment does little beyond damaging the oligomer. Using concentrations of NaCl in the Section A.6.2 method above 400 mM are worth the risk here if they are necessary.

#### A.6.3.2 High Performance Liquid Chromatography (HPLC) of LC Forming Oligonucleotides (DMT-on Method)

On the background of the more general HPLC reverse phase chromatography method, it becomes useful to talk now about the DMT-on method as first mentioned in Section A.6.

DMT-on oligomers are attained by omitting part a. of Figure A.7 in the final processing during oligonucleotide synthesis; DMT typically removed during this step is left in place while DEA and ammonium hydroxide are both used as usual to remove the alkaline labile protecting groups on the bases and backbone. This results in oligomers that are capped at their 5'-terminals with the bulky DMT-group. This treatment can be exploited as a purification handle: because DMT is so bulky and hydrophobic, it tends to stick much more strongly to the hydrophobic column as used in Section A.6.3.1. This substantially increases the resolution of the hydrophobic column by significantly increasing the percentage of buffer B required to cause the DMT-coupled oligomer to elute from the column as opposed to failure

sequences which intrinsically have no DMT. Using the gradient elution described in A.6.3.1, the temporal separation of failure sequence elution from elution of the DMT-linked oligomer can increase by ten or fifteen minutes. This makes DMT-on purification more stringent than the basic method illustrated in A.6.3.1.

The typical strategy (Figure A.19) for reverse phase DMT-on purification is to run one round of purification, as seen in Section A.6.3.1, in order to fish the DMT-linked oligomers out of a mixture containing extensive failure sequences, then use a gentle acid treatment to cleave the DMT from the 5'-terminal of the oligomer and then to repeat the reverse phase purification scheme of A.6.3.1 to separate the now-unlabeled oligomer from the formerly attached DMT.

The second chromatography step added here to separate the target oligomer from dissociated DMT after first separating it from failure sequences doubles the solvent needed for purification and more than doubles the time. The fidelity of fishing for sequences that are tagged with DMT over those that are not also significantly improves the separation of target oligomer from contaminants. One must judge which of these competing interests wins in any given case.

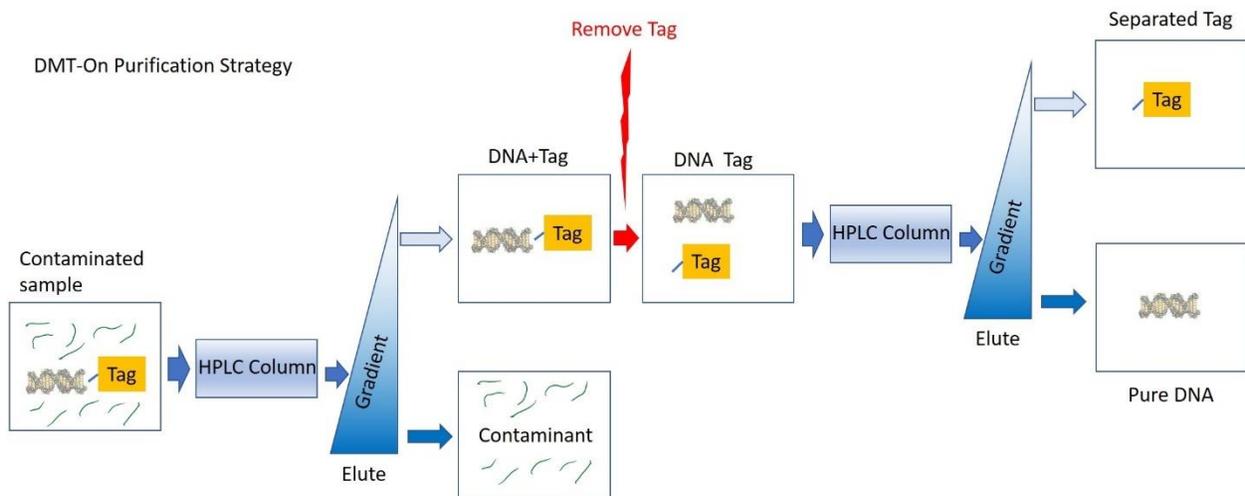


Figure A.19: DMT-On Purification strategy.

For the specifics of how this purification technique is run, all the details of A.6.3.1 remain applicable. The additional new elements reside in the intervening space between the first chromatography round and the second. After the first chromatography step is completed, sample from the column must be collected and concentrated into a smaller volume, which has typically been carried out by rotovap. When I was working with this technique, I was still unaware of how damaging drying the oligomer sample in rotovap could be, so I routinely concentrated all the way to dryness before running the reaction to dissociate DMT from the oligomer. In my current thinking, rotovap must be used gently and the sample should come out of rotovap still contained in a liquid phase. DMT is removed by a gentle acid treatment. It should be noted that oligomers are ultimately not stable to residing in acidic solution, but that they can tolerate some moderately low pH values for short times, so the DMT removal treatment should take place in a timely manner (~30 min acid). As I typically performed this removal on a solid phase, for better or worse, DMT removal solution (as documented below) was added straight to a pellet in a round bottom flask and allowed to react at room temperature for about 30 min with gentle shaking. The solution turns a faint pinkish color as DMT is released (as opposed to the bright red in the solvent conditions of the synthesizer). After the reaction is completed, the solution should be applied straight to the reverse phase column in order to separate the acid solution from the oligomer-DMT mixture. This leads naturally into the second column chromatography step, where the now-freed oligomer will elute separately from the DMT.

Table A.9

DMT removal solution	30 mL
Glacial acetic acid (80%)	24 mL
Water (20%)	6 mL

A second, more efficient way to run this technique, which avoids the rotovap in the intervening step between the two chromatography rounds, is to perform the DMT removal reaction on the chromatography column. This depends on the certainty that the reverse phase chromatography column packing material is stable to an acid treatment, which they sometimes aren't. If the column can take this sort of treatment, one retains the DMT-linked oligomer on the reverse phase column by terminating the buffer A-buffer B gradient program after the failure sequences have eluted from the column during the first chromatography round and then washing with 100% buffer A for several column volumes to restore the column to its starting solvent conditions. This should be supported by running the gradient program with a slow enough rate of buffer B content increase so as not to place the elution point of the DMT-linked oligomer within one column volume of when the last failure sequence has eluted since this would cause the DMT-linked oligomer to begin elution prior to the optimal termination point of the gradient program (one column volume is approximately 5 minutes at 20 mL/min, or 100 mL, on the Sepax column). After 100% buffer A conditions have been restored to the column, DMT removal solution is then injected using the injection loop and washed through slowly in several column volumes (~150 mL) at a pump rate that stretches the treatment out for 30 minutes, which will mean multiple injections. Post reaction, the column is washed with buffer A to remove the acid and the second chromatography round is undertaken. This technique was used successfully in my hands several times.

The inspiration for this second version of the technique was a single column commercial preparation product called Poly-Pak (Glen Research, Cat# 60-1100-10), which enacts the DMT-on purification strategy on a single disposable column in small format. The format here is unfortunately too small for the preparative scale

where I typically operate, but the technique adapts well to larger scale with an acid resistant reverse phase column.

After completion of the two chromatography rounds, purified oligomer is rotovapped to small volume and again passed through A.6.2 in order to produce material that forms liquid crystal. As typical, materials yielded directly from this method (Section A.6.3.2) do not form LC phases, even though they are “clean.”

#### A.6.4 PEG Crashing of Self-Complementary Oligomers

While A.6.2 has been a workhorse for producing LC forming oligomers, the technique presented there is not without a number of unfortunate caveats. The biggest of these is the major loss of yield on running the precipitation. A preoccupation of mine during much of my work has been trying to find a way around the losses of A.6.2 while preserving the capacity of that technique for generating LC forming materials.

The most successful of these endeavors exploits a characteristic of self-complementary LC forming oligomers. This is the capacity for these materials to respond to the Depletion Effect. In a nutshell, depletion is the phenomenon where flexible objects tend to demix from a collection of rigid objects in order to maximize the entropy of the flexible objects. An LC forming self-complementary oligomer in favorable salt conditions tends to be base-paired in a rigid duplex form which, further, tends to favor self-exclusion from flexible polyethylene glycol (PEG) polymers. If demixing is strong enough, as can be regulated by adjusting the PEG concentration, the oligomers can be forced to demix into a small enough partition space as to favor LC formation, producing large physical domains which can then be readily sequestered from the mixture by centrifugation.

Supernatant containing  $5^{\prime}\text{GCGCTTAAGCGC}3^{\prime}$  saved from the first centrifuging step in A.6.2 was added to an equal volume of 40% PEG with 500 mM NaCl at room temperature. The mixture immediately produced a heavy, flocculent precipitate which was removed to pellet form by centrifuging at 10,000 rpm with an SS-34 rotor. Direct sampling showed this material to be birefringent. UV-absorbance characterization (Section A.7.2) demonstrated that the optical density yield per volume recovered here equaled that recovered initially with A.6.2, roughly doubling the yield for the synthetic prep. The only drawback to the technique is that purity, as quantified by optical density per recovered mass, took a substantial hit: only about 75% as pure by mass compared to material recovered in the initial centrifugation in A.6.2. Whether the impurity is predominantly NaCl or PEG is unknown, but likely both.

Owed to a lack of instances where I could try this technique, since it requires a self-complementary oligomer, this method is non-optimized. Independent experimentation suggests that NaCl concentrations down to 80 mM to 100 mM (in the mixed, precipitating fraction) may be sufficient to produce this effect. PEG concentrations have not been optimized for this application at all. Of note, 100 mM NaCl is a quite minor relative contaminant since sodium counter ion in an LC phase is expected to be at concentrations of 800 mM or up to around 1.2 M depending on the phase. The presence of PEG as a contaminant directly inside LC domains is unlikely, but it can easily show up as a contaminant during centrifugation by being present in the interstitial spaces between small granular LC domains present in a centrifugal pellet.

Mutually complementary oligomers should also be recoverable by adding these to each other in order to generate a duplexed form, but no method has yet been found, or is particularly anticipated, for extracting these oligomers from each

other after they've been recovered. An unpaired oligomer, without any available complement, is not expected to be recoverable in this manner, but this experiment is untried due to a lack of synthetic need for such oligomers after this recovery technique was originated.

## A.7 Characterization Methods

One must characterize the materials produced by the Caruthers synthetic method in order to understand the quality of any given synthesis batch. Useful parameters for judging synthetic quality are the mass of the synthesized polymer as determined by mass spectrometry along with the homogeneity of that mass and the UV absorbance of the material in various solvent backgrounds encountered through the processing of the synthetic prep.

### A.7.1 Mass Spectrometry of Synthetic Oligomers

Mass spectrometry (mass spec) of an oligonucleotide target has been the source of confirmation of correct sequence synthesis during my work. This technique reveals the actual molecular weight of an oligomer which can be compared to the expected theoretical weight of that molecule as calculated by summing up the known weights of its component nucleotides. This method should be considered a confirmatory technique verifying correct synthesis and is not specifically sensitive to the exact sequence of the oligomer since it will report the same mass for every oligomer that possesses the same number of dA, dG, dC and T bases. So, mass spec cannot distinguish among oligomer sequences of identical weight but differing nucleotide order. What it can detect is sequences which omit or add mass, like sequence truncates or concatemers.

The specific technique used most commonly has been MALDI-TOF mass spectrometry, or matrix assisted laser-desorption initiated time-of-flight mass spectrometry. In this technique, the analyte is co-crystallized with a photo-reactive molecule called a matrix which will donate or abstract protons from the molecule when pulsed by laser light and allow that analyte to be ejected into a volatile form which can then be accelerated between electrically charged plates and timed to see how fast it flies.

$$qV = \frac{1}{2}mv^2 \quad (\text{A.1})$$

$$\frac{2V}{v^2} = \frac{m}{q} \quad (\text{A.2})$$

$$2V \left( \frac{\Delta t}{\Delta l} \right)^2 = \frac{m}{q} \quad (\text{A.3})$$

The amount of kinetic energy carried by the analyte (charge  $q$ ) after passing through the accelerating voltage is equal to the potential drop across that voltage ( $V$ ), as shown in equation A.1. Equation A.2 rearranges this to extract mass per charge and equation A.3 supplies the parameters owed to the mass spectrometer, namely the length of the flight path,  $\Delta l$ , and the time of flight,  $\Delta t$ . If the molecule contains only a single charge, mass per charge (typically called  $m/z$ ) directly reflects the mass of the molecule.

Oligonucleotides possess a strong intrinsic negative charge owed to the low pKa of the phosphates in the polymer backbone, giving one unit of negative charge per base, requiring the mass spec to be set in negative polarity mode in order to detect them. The matrix molecule is added to supply protons to the backbone during the desorption reaction to reduce the molecule's net charge to only -1 or -2 and to help eject the analyte in a volatile form. Two matrices are suitable, including THAP (2',4',6'-Trihydroxyactophenone monohydrate) and 3-HPA (3-Hydroxypicolinic acid). THAP is considered less suitable because it is linked to in-flight oligomer degradation<sup>52</sup>.

MALDI matrix preparations are as follows:

Table A.10

Ammonium citrate Stock solution (104.9 mg/mL)	
Ammonium Citrate	29.5 mg
Water	280.1 $\mu$ L

Table A.11

MALDI-TOF matrix	83.2 $\mu$ L
3-HPA (3-hydroxypicolinic acid, Fluka cat# 56197)	2.92 mg
Ammonium citrate (from 104.9 mg/mL stock)	5.66 $\mu$ L
Acetonitrile (to 10%)	8.32 $\mu$ L
Water	69.2 $\mu$ L

These were empirically optimized as taken originally from Shahgholi *et al.*<sup>53</sup> and Sauer<sup>52</sup>.

When co-crystallized with the analyte, there is not a specific analyte:matrix ratio that will immediately generate good a MALDI-TOF signal for every oligomer, requiring a number of ratios to be explored experimentally. Typically, solid oligomer is dissolved  $\sim$ 1 mg/ 20  $\mu$ L of water and is then spotted on a gold MALDI grid in a 1:1

serial dilution. Here, 1  $\mu\text{L}$  of analyte is mixed with 1  $\mu\text{L}$  of matrix on the grid spot to dilute the sample to 1 in 2 and 1  $\mu\text{L}$  of the resulting mixture is carried to the next spot, where it is mixed with another 1  $\mu\text{L}$  of matrix for a subsequent dilution of 1 in 4. This dilution pattern is repeated as many as ten or twelve times. In this process, the analyte content is reduced by a factor of 2 from one grid spot to the next while matrix always fills in the remainder. The spots are allowed to dry, then taken to MALDI-TOF, where they are screened until a strong, clean  $m/z$  is obtained.

If the oligomer is not in a solid form and the sample is from some intermediate phase after synthesis or during purification, the concentration of the oligomer is frequently immediately high enough in concentration that it can be plated straight to a MALDI plate and detected by mass spec. One should be aware that the solvent and ion content of some of these intermediate samples can distort mass spec results slightly.

Without including a precision calibrant, MALDI-TOF can be expected to report accurate  $m/z$  within about 10 Daltons --about the mass of a carbon or less-- making it possible to conclude the molecule is the correct species since such small discrepancies of mass are unlikely to be omitted from the intact molecule since, of the organic atoms present in the oligomer molecule, none but hydrogen are smaller than carbon. For the instrument used, oligomers typically flew with relatively strong laser power, ~65% or a little higher, and laser power was frequently adjusted gradually upward until signal appeared. Spots on the MALDI grid in the dilution series which typically began to give observable signal were between 1/4 to 1/8 or higher dilutions.

One should also be aware that the actual mass of an oligonucleotide will tend to be slightly lighter than the predicted theoretical values found from on-line mass spec tools. These tools predict a weight where all phosphates in the backbone are

protonated, despite the fact that these sites can be reliably anticipated as deprotonated in neutral pH aqueous solution due to the low pKa of a phosphate group. The predicted mass spec read-out value for a 12mer oligomer with no phosphates adorning either its 3' or 5' terminals will be 10 protons too heavy since the value anticipates only one exposed charge. This discrepancy does not show up in the mass spec data because the mass spec is only detecting singly or doubly charged oligomers.

Oligomers show some characteristic artifact peaks that one should watch for. At analyte:matrix ratios that most strongly favor the analyte, one can expect to see sodium (or ammonium) ladders. These will appear as one sharp peak followed by a series of exponentially lower peaks trailing to higher m/z values where the spacing between each peak in the train is around 23 Daltons (or 19 Daltons). The number of peaks in the train should not exceed the number of phosphates present in the oligomer's backbone. Sodium ladders will occur when the oligomer flies with some variable number of sodium ions substituted in place of protons on the backbone (one substitution gives a weight of +23 Daltons; two gives +46 Daltons and so on). A second artifact is an oligomer mass ladder. Oligomers tend to stick to one another and can give a widely spaced train of peaks at m/z values that are multiples of the expected single oligomer m/z. A third type of peak artifact appears at m/z values that are 1/2 that of the expected oligomer: this is indicative of oligomers flying doubly charged instead of singly charged. A final artifact effect that can sometimes occur is that the main analyte peak will appear to push toward fractionally lower values when the counter ion is unstable. This seems to occur when counter ions volatilize away from oligomers in flight, skewing the m/z mid-experiment and giving fractional effective charges. This can be counteracted by adjusting pH or reducing the ion content present with the analyte. The appearance of all of these artifacts can

be minimized by screening an entire series of dilutions since lower analyte concentrations frequently provide a better, more stable signal. Some of these artifacts can also be useful to check that a particular mass peak is oligonucleotide, especially the existence of counter ion trains.

The most common contaminant signal to expect is the presence of oligomers with lower mass than the expected species. This can occur either by fragmentation of the expected molecule, or by synthetic failure sequence contamination, both of which will appear as a train of masses leading toward lower  $m/z$  with a spacing between each successive peak of about 300 to 350 Dalton, or the approximate mass of a base mononucleotide. Searching for such an oligomer ladder prior to purification can be a good tail for the efficiency of the coupling reaction during synthesis since it permits direct visualization of failure sequence oligomers. Some failure sequence should always be expected, but the larger the ratio of (desired oligomer intensity)/( $n-1$  failure intensity), the better the synthesis. Ideally, the  $n-1$  failure should not be observable above noise. That said, longer oligomers (>20 bases) will often be accompanied by an observable failure sequence train because efficiency inevitably begins to fade after a number of base additions in synthesis. The appearance of an oligomer train after purification is helpful also in determining if some treatment has caused the molecule to fragment, which can occur with excessive rotovapping or prolonged acid exposure.

One catastrophic synthetic failure occurring early in my time with oligomer synthesis was noted with an apparent  $n-1$  failure being the only synthetic peak, when the desired oligomer was never observed. This failure was particularly confusing because the  $m/z$  did not appear to be a terminal truncate, where either the 3' or 5' base was omitted, but occurred sometimes at an intermediate value, as if it were a partial failure or a partial degradation. If a single base were being omitted

randomly in the sequence, we reasoned that the peak would actually reflect four different masses, which was not observed, suggesting uniformity of synthesis. The peak appeared to be a single mass. Teasing out the cause of this particular discrepancy took significant work since the MALDI-TOF results at the time were also questionable by themselves, making it difficult to determine whether the observed fragmentation was a MALDI artifact, a purification artifact or a synthetic problem. I would recommend any future student facing this situation to do your best to simplify simultaneous variables: buy a small amount of commercially produced oligomer and use it to hammer out your quality detection experiment (like your mass spec) before trying to address synthetic problems. We ultimately determined that this failure was an n-1 omission occurring at the second position along the backbone from the 3'-terminal –this omission occurred due to a skipped DMT removal of the 3'-seed base bound to the solid support, as noted in Section A.5.3, allowing addition of the second nucleotide to fail at the outset.

MALDI-TOF has been a relatively good tool for the purposes of checking oligomer synthesis and purification (see Figure A.20), but it suffers from an inability to be readily quantitative. It reflects only quantities of a prearranged solid that desorb due to laser illumination and fly in the mass spec, not actual concentrations. We attempted to use MALDI to quantitate chemical ligation in an LC-forming oligomer species, but were unable to separate this sort of signal from the intrinsic stickiness among oligomers.

A recommended additional method to complement or replace MALDI-TOF which I won't detail here is LCMS. This technique promises greater precision and accuracy than MALDI-TOF and is less likely to suffer the same sorts of failures. As of this writing my methods for LCMS have improved, but are not fully optimized or reliable to the purposes described in this thesis.

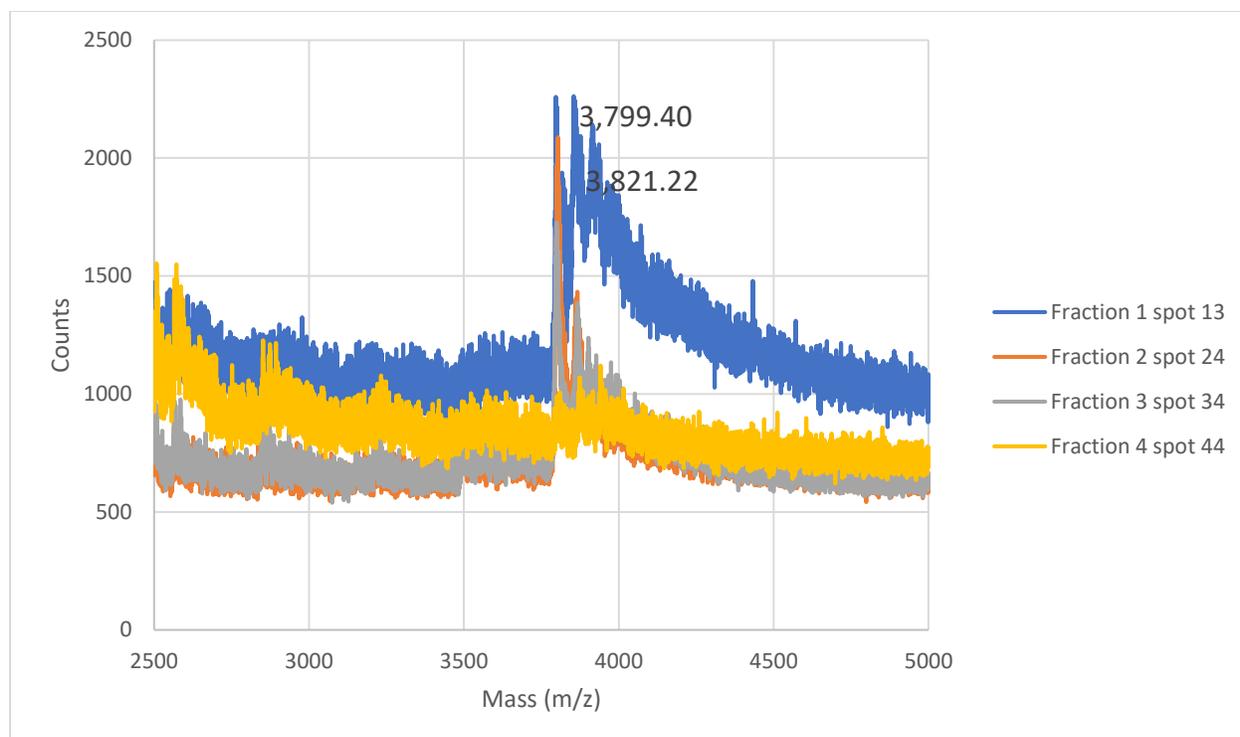


Figure A.20: MALDI-TOF on RNA purification fractions. Shown here are MALDI-TOF traces of purification fractions for a Drew Dickerson variant of RNA while attempting to check purification (Fractions 1,2,3 and 4). For sequence  $5'cgcgaaucgcg3'$  expected singly charged  $m/z$  was 3810.4 g/mol. Error here is less than a carbon without internal calibration.

### A.7.2 UV Absorbance of Oligonucleotides

The second major tool used for characterizing oligomers in my work was UV absorbance. The method depends on the Beer-Lambert Law.

$$A = \epsilon lc \quad (\text{A.4})$$

Based on the assumption that material absorbers are decoupled from each other, this famous law states that the absorbance ( $A$ ) of a sample is proportional to the concentration ( $c$ ) of that sample. This proportionality depends on the length of the light path through the sample ( $l$ ) and a material-dependent factor called the extinction coefficient ( $\epsilon$ ), which can be calculated theoretically or determined

empirically. I will refer to absorbance by the unit of OD (Optical density) which is calibrated in arbitrary units (AU) based on transmittance (ratio of incident light intensity to transmitted intensity). OD is logarithmic in base 10, where 1 AU is 10% transmittance, 2 AU is 1% and 3 AU is 0.1%<sup>54</sup>. Oligonucleotides absorb strongly in the UV at 260 nm, owed to the aromatic groups in the nucleotide bases.

Oligonucleotide extinction coefficients can be somewhat tricky to judge in my experience because one value is possible for single-stranded, unbound oligomer, while a different value is obtained for duplexed oligonucleotide, which is not exactly the sum of two oligomer values. For better or worse, most of my work in this avenue was not directed at the determining extinction coefficients.

I use two absorbance measurements for characterizing oligomer synthesis. The first is intended to judge oligomer yield per solid support loading and is intended to grade the quality of a synthetic program. The units for this are OD<sub>260</sub>/μmol loading, roughly oligomer UV absorbance per synthetic scale. The second is intended to grade the oligonucleotide purity as a final purification step in units of OD<sub>260</sub>/mg mass, which corresponds roughly to oligomer absorbance per unit mass of purified solid. The first measurement was implemented at the suggestion of an oligonucleotide synthesis expert from GE Healthcare bioscience as a means of troubleshooting synthesis issues. I implemented the second measurement as a way to compare quality across multiple batches of synthetic oligonucleotide when I realized that liquid crystal formation capacity is not directly proportional to purity and that my purifications were in fact not completely identical from one to the next –so OD<sub>260</sub>/mass gives a means of comparing variation across batches.

I also have begun more recently to use a standardization measurement in order to track variation of the lamp in the spectrophotometer at the wavelength used for oligonucleotide. The standard I chose is acetone, which has currently been

seen to possess  $\sim 213 \text{ OD}_{260}/\text{mL}$ . Given standard temperature and pressure conditions in the lab, I expect this value to only vary depending on the spectrophotometer. The measurement for this is exactly as for the other two values described above.

These numbers are obtained taking a trial sample and diluting it two-fold a series of ten or twelve times in water, then measuring the  $\text{OD}_{260}$  at each dilution in a 1 cm pathlength quartz cuvette. This data is plotted absorbance versus dilution and then linear regressed across only the region showing linear behavior (see example in Figure A.21). The resulting formula is used to back-calculate the  $\text{OD}_{260}$  in the undiluted sample in units of AU. Of note, formulas obtained this way can be used to directly calculate an extinction coefficient, where the extinction coefficients I typically use are in units of  $\text{AU}/(\text{cm} \cdot 2\text{-fold dilution})$  in comparison to the Beer-Lambert law. I typically did not convert to a coefficient expressing molarity of the oligonucleotide because the samples I use have some salt contamination as a side effect of the requirement that the sample be optimized for LC formation. Performing a 2-fold dilution series for this experiment has been particularly stable because the pipettors do not need to be adjusted at any point during the experiment, giving good equipment precision without having to error check unnecessary pieces of equipment along the way.

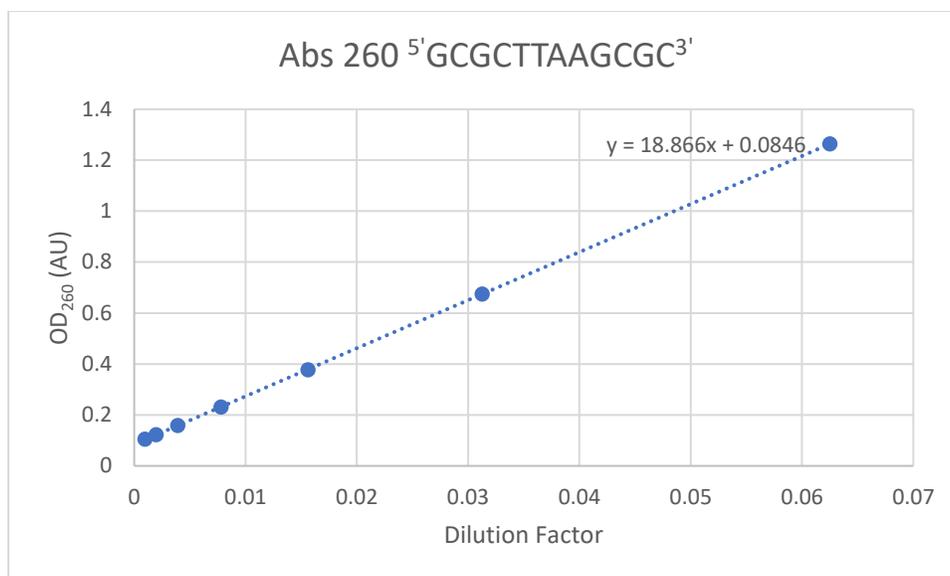


Figure A.21: Example of Spectrophotometry on rDD. In this example, a dilution factor of 1 is undiluted. The linear regression allows determination of OD<sub>260</sub> per dilution factor, giving a basis of comparison for the undiluted sample.

The synthetic yield measurement of OD<sub>260</sub>/μmol support is always calculated from a sample taken during the procedure described in Section A.6.1. Here, the oligonucleotide is typically sampled immediately after the solid support has been filtered away. One milliliter is sampled and the total volume of the sample in milliliters is recorded (at this point typically ~540 mL). In this state, the oligomer is deprotected and has been subjected to minimal purification losses, and should reflect the closest measure of the productivity of the DNA synthesizer. It is also typically noted how many grams of solid support was originally used during synthesis and what the seed loading of that solid support is. For Primer Support 5G this typically boils down to 242 μmol of loading. OD<sub>260</sub>/μmol is calculated by determining the AU of absorbance present per milliliter of the undiluted sample, then multiplying that by the total volume of the sample (~540). The resulting number is then divided by the support loading value (242 μmol).

For a typical oligonucleotide 12mer, if the DNA synthesizer is operating at peak efficiency, synthetic yield using Primer Support 5G will be approximately 90

OD<sub>260</sub>/μmol support. This value can be expected to vary from one type of solid support to the next given differences in reactive release during the deprotection step and efficiency of accessibility within the solid support structure during synthesis. For the same type of solid support, this yield value is roughly linear to the length of the oligomer being produced: a 4mer oligomer can be expected to produce ~30 OD<sub>260</sub>/μmol. To reiterate, these values are dependent on the solid support used and specialty solid supports frequently produce lower yields. If synthesis is inefficient for any reason, one can expect a decrease in this value (e.g. 60 OD<sub>260</sub>/μmol as opposed to 90 for a 12mer).

Purity of a synthesis preparation is judged by a measurement of OD<sub>260</sub>/mg. This number is obtained by dissolving a known number of milligrams of the purified and lyophilized oligonucleotide sample in 1 mL of water and then performing the spectrophotometric method described above. Upon back-calculating to the OD<sub>260</sub> value for the undiluted sample, one simply divides this value by the number of milligrams dissolved to produce this optical density. For samples with no purification except for what is described in Section A.6.2, a typical value is 20 to 21 OD<sub>260</sub>/mg. With purification including Section A.6.3.1 (the HPLC method) in addition to A.6.2, purity has been seen as high as 26 OD<sub>260</sub>/mg. With usage of X-ray diffraction methods on the LC phases to correct for contaminating mass (can be performed with data from Sections 2.0 and 3.0), these values can be used to calculate accurate empirical extinction coefficients for molarity (with the understanding that no assumption has been made about the single or double-stranded state of the oligomer during the spectrophotometric measurement.)

Beyond the specific measurements indicated above, spectrophotometry can be used to check yield at steps through the course of any purification regimen by simply sampling at every step and calculating a total OD<sub>260</sub> value for the volume of

sample on hand at that step. This method was used to determine the poor efficiency I have described when using the method from Section A.6.2 –after that step, total absorbance units for the sample have been seen to decrease by half, proving the loss.

## A.8 Specific Projects

What has been described to this point is all of the methodology to totally synthesize, purify and characterize a typical DNA oligomer with no special synthetic characteristics. I will devote this section to explaining how one produces synthetic modifications seen in papers (refs) where materials produced by me have played a role in the research, as well as other constructs which may not have seen publication.

### A.8.1 Adding a Phosphate at the 3'-Terminal

Basic oligonucleotide synthesis produces oligomers with hydroxyl groups at their 3' and 5' terminals. One special construct has appeared in work aimed at illustrating the capacity for liquid crystal phases to act as a catalytic effector to enhance chemical ligation of closely packed duplex oligomers by von Kedrowski chemistry<sup>32,33</sup>. This type of chemistry operates by producing DNA backbone ligation by a carbodiimide reaction (Figure A.22). Carbodiimide attacks a 3-terminal phosphate group and converts into urea upon facilitating the attack of a 5'-terminal hydroxyl from a neighboring oligomer and forming a phosphodiester linkage. This chemistry cannot operate without a terminal phosphate on one of the oligomers, which basic DNA synthesis does not produce.

## Carbodiimide Mediated Ligation

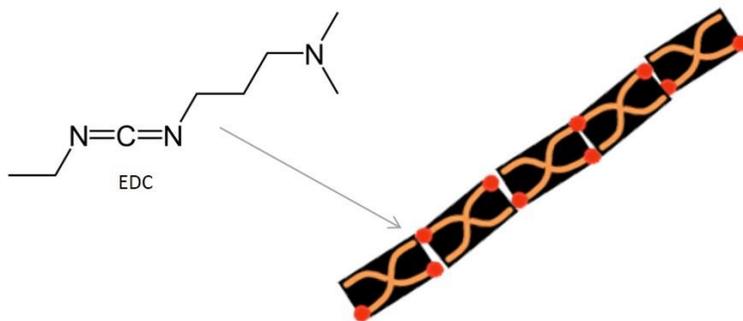
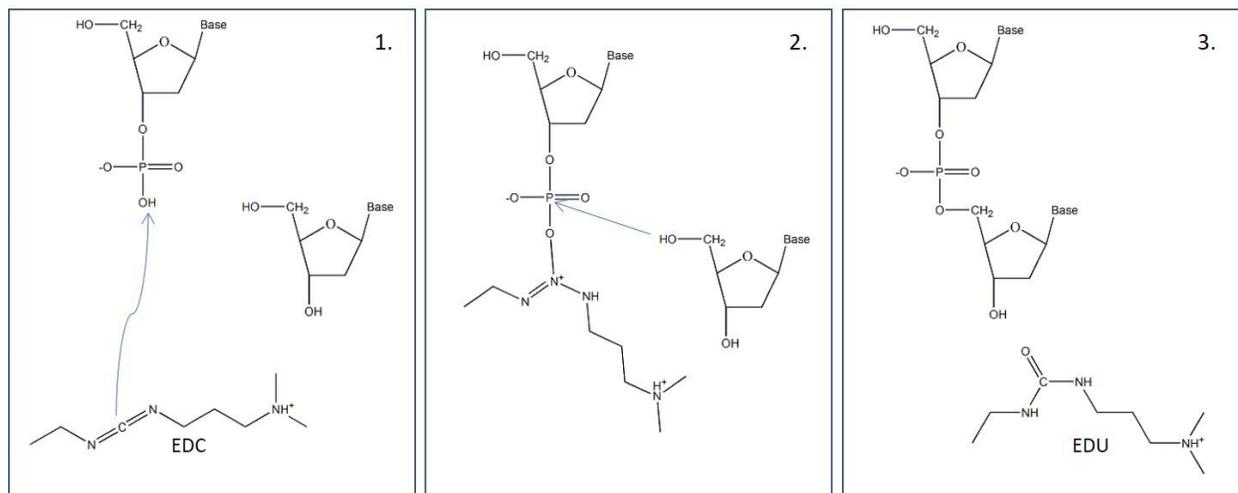


Figure A.22: Carbodiimide ligation by EDC. 1.) EDC attacks terminal phosphate group. 2.) Terminal hydroxyl attacks EDC modified phosphate. 3.) EDC group is displaced as EDU, leaving natural phosphodiester.

Oligomers decorated with a 3'-phosphate can be produced by use of a specialized solid support (Kinovate Nittophase HL Pi250 product# 01-00-36-250). This form of solid support is has no seed monomer and contains instead a decoration that allows the phosphoramidite group of the first PA added to be converted into a 3' phosphate when the completed oligomer is cleaved from the support by the treatment in Section A.6.1, as described by Kumar<sup>55</sup>.

The set-up for synthesis to generate this material is modified from the basic method described in earlier sections by inclusion of the 3'-phosphate solid support in

place of the usual solid support. To use this solid support, the synthesizer must be set to add all N-bases of the oligomer in question as opposed to the usual N-1. This must be carried out because the support omits the seed base. For the in-house oligopilot, the correct setting can be achieved by instructing the synthesizer to synthesize an extra base at the 3'-position, making the synthesizer think it's producing an N+1 oligomer (e.g. for reverse Dickerson Dodecamer 3'-phosphate, setting  $5'GCGCTTAAGCGCA^3$  where the 3'-terminal dA is actually assumed falsely to come from the solid support).

Purification of 3'phosphate modified oligomers is very similar to the unmodified version. The only significant variation is that during the process described in Section A.6.2, precipitating oligomer will tend to form an oil instead of a solid precipitate. How to deal with this is described in Section A.6.2.

It is noteworthy that some of the older forms of solid support capable of 3' chemical phosphorylation of a synthetic oligomer can be labile during the deprotection step that removes the CE protecting group. The CE-group can attack it. This can be avoided by simply using more modern support.

As a small addition, I can give guidance on how to achieve chemical phosphorylation at the 5'-terminal of an oligomer even though I've never done this myself. This is accomplished by coupling a specialized phosphoramidite as the final 5'-terminal base in a synthesis sequence (e.g. Glen Research Chemical Phosphorylation Reagent II Cat# 10-1901). For use on the oligopilot synthesizer available in-house, this sort of synthesis would have to be done in two steps. Synthesize the sequence as normal up to near completion, then omit all on-synthesizer final deprotection steps in order to leave the oligomer fully protected (leave the CE and 5'-DMT both), and keep this material in the column on the synthesizer with plans to run another synthesis on it. After completing this, select

one amidite bottle, clean it out and wash both it and the solvent lines connected to it to remove residual dN-PA, then replace the contents with the phosphorylating PA and prime the line. Perform all necessary washing and priming through the column bypass so as not to disturb the reactor column. Run a synthesis at the same scale as before performing a base addition cycle using only the bottle containing the phosphorylating PA. Include the usual deprotections to complete the synthesis and carry on afterward with deprotection and purification as usual. The method must be performed in this way because the oligopilot synthesizer has only four dN-PA bottles and addition of a fifth PA in parallel requires a new bottle that we don't have.

#### A.8.2 Synthesis of Randomer Oligonucleotides

Another common special application was the production of random sequence oligomers<sup>31</sup>, both 3'-phosphorylated (using Section A.8.1 also) and unphosphorylated. These materials are completely inhomogeneous in sequence but intended to be uniform in length. As such, during characterization, mass spec should reveal a sample of these oligomers to possess an average weight in some Poisson distribution rather than a specific discrete weight –the number of oligomer types in the distribution should go as  $4^N$  with N-length given mass spec degeneracy for oligomers that have the same number of each base.

Random sequences are formed by running the coupling step with all four bases in equimolar content simultaneously. To set this up prior to synthesis, dissolve each dN-PA to the same concentration, then add the same volume of each type to one single dN-PA reservoir bottle (there will be some left over of several bases since none require the same amount of acetonitrile to dissolve to the same

concentration.) The second part of setup is to mix the solid support on loading the reactor column: since solid support is typically decorated with the 3'-terminal base, support containing each type of base must be added to the column in an equimolar fashion. Achieving an equimolar content of the four types of basic solid support will depend on normalizing for loading variation –loading is usually close to the same, but sometimes non-identical. Exact quantities can be calculated using a weighted average, but it should be close to one quarter of each type.

To run the random sequence synthesis, set the synthesizer to draw from the single used amidite bottle for every base addition cycle –for example, if you added the mixed dN-PA to the amidite bottle, the synthesizer should be told to produce a sequence of poly-dA: 5'AAAAAAAAAAAAA3' for a random 12mer.

3'-phosphorylated random oligomers are more simple to produce than unphosphorylated oligomers because the solid support is uniform and does not require any mixing. Just perform the method in Section A.8.1 as required in parallel to the method described in this section.

Another possible way to simplify the synthesis of random sequence oligomers is to use so-called “Universal linker solid support.” Universal support is a new technology that allows a single solid support to be used for all sequences (omitting any 3'-modifications). Because usage of this support is somewhat more complicated than the basic method and requires special consideration on the in-house oligopilot synthesizer, I will detail how to use it in Section A.8.3. This support is obviously useful in production of random sequence oligomers because, again, it simplifies the method by not requiring mixing of different types of solid support.

### A.8.3 Usage of Universal Linker Solid Support

Universal linker solid support is a type of solid support that contains no pre-loaded 3'-terminal seed monomer. This material can be acquired from many solid support vendors; GE healthcare Life sciences offers it as Unylinker (Product code: 28-9964-34).

Universal linker contains the advantage that it can be used to synthesize any sequence irrespective of the 3'-terminal nucleotide, meaning that less solid support needs to be purchased in order to synthesize more sequences. On the other hand, the chemistry of this type of solid support is different than conventional support: the terminal 3'-terminal nucleoside must be added to the solid support using the equivalent of two base addition cycles instead of one. Where conventional solid support requires sufficient dN-PA to support synthesis of N-1 oligonucleotides in an oligomer of length N, universal linker requires enough dN-PA for N+1 bases, meaning that it always uses more dN-PA. For synthesis of short sequences, this means much more dN-PA is used, but usage tends to converge toward normal quantities the longer the oligomer sequence being synthesized. Less solid support needs to be purchased, but more dN-PA.

The advantages of universal linker shine through when performing methods such as that described in Section A.8.2. Only one universal linker is required to synthesize random sequence oligomers, simplifying the setup of the method.

Usage of universal linker is more complicated than basic solid support because of the more complicated attachment chemistry. The in-house Oligopilot synthesizer programs admit ease of use for basic solid support but require a special program method in order to use universal linker. Rather than starting with a more

usual method wizard (Section A.5.3), directly open the Unylinker example method in the Unicorn method editor (called RECYCLE 6mL PS5G UNY AKOPc121 Ed 007) stored in the top directory (called default/). This method contains an example sequence that can then be overwritten with the desired oligomer sequence with one extra base at the 3'-terminal (input an N+1 sequence). After adding the usual program parameters and creating the method (always save under a name different from the example method), switch the editor to text view to see the synthesizer program script, highlight the first "add\_DNA\_Base" block in the script, which is the addition routine for the most 3' base, scroll the block menu for the equivalent "add\_DNA\_Base\_Uny" and click "replace." The name of the block should then be appended with "\_Uny" in the script to signify the universal linker addition method. Save the method under the new name (important not to destroy the example method) and use as usual.

The new "add\_DNA\_Base\_Uny" program block contains the instructions necessary to double the solid support exposure to activated dN-PA at the opening of the synthesis program, in order to establish the first base at the 3'-terminal, and then to cap the exposed solid support in order to render it inactive before switching to more conventional addition cycles in the subsequent steps.

#### A.8.4 5'-Terminal Triphosphate Modification

More than decorating the terminal with a single phosphate, another possible modification is the inclusion of a full triphosphate. This modification was made following a reaction described by Lebedev et al.<sup>56</sup>.

The reactions were carried out in the reactor column on the synthesizer using two temporary jury-rigged bottles in order to carry the needed reagents into the

system. This configuration was selected to protect the reaction from the air and to make it convenient in the context of the oligopilot without requiring that chemicals be switched in and out of the normal bottles. A solvent bottle was added at the amidite-T\* line and a chemical bottle was added at the amidite-G\* line, both on valve 2. These lines are not otherwise used in normal synthesis.

The reaction is executed following the full, normal synthesis of an oligomer and then deprotecting the 5'-hydroxyl (removing the last DMT) of that oligomer while leaving the CE-protecting groups intact. The open 5'-terminal hydroxyl is then used as the target for the triphosphate addition. The triphosphate addition is carried out in two steps: first, the terminal is decorated with 2-chloro-4H-1,3,2-benzodioxaphorin-4-one (CBP), which is an activated phosphite form; second, the activated terminal is reacted with pyrophosphate to produce the final triphosphate (see Figure A.23). All of this is performed on the synthesizer using the temporary bottles.

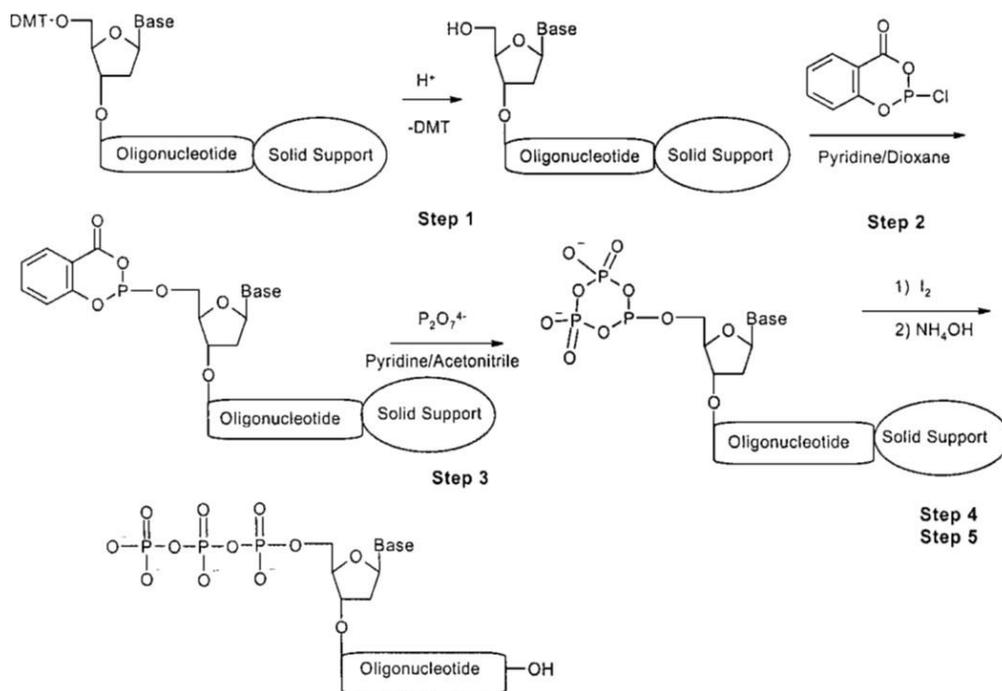


Figure A.23: Triphosphate addition at 5'-terminal of DNA. Scheme taken from Lebedev *et al.*<sup>56</sup>.

The reaction was carried out as follows. 1,4-Dioxane, as a solvent, was washed through the reactor column using the solvent bottle placed at the amidite-T\* line until the absorbance saturated. Then, 1M CBP was flowed from the chemical bottle added at the amidite-G\* line on the same valve, at 5 mL/min. As the 1 M CBP went dry, valve 2 was switched back to the amidite-T\* line in order to push the CBP to the reactor using 1,4-Dioxane (yellow CBP was clearly visible in the reactor loop). Valve 5 was then switched to position 2 in order to close the normal reactor column recycling loop and pump B was operated at 10 mL/min to recycle the CBP through the reactor column for 30 minutes at room temperature. The column was then washed out with acetonitrile to remove all unreacted CBP. 0.5 M tris(TBA) pyrophosphate was switched onto the amidite-G\* line in place of the CBP and flowed to the column in exactly the same manner, though using an acetonitrile push instead of 1,4-Dioxane. Valve 5 was again operated to close the recycle loop and pyrophosphate was recirculated through the reactor for 40 minutes to react. The column was once more washed with acetonitrile. To complete the reaction, oxidizer was pumped through the column in a mixture with acetonitrile in order to complete the reaction (acetonitrile from Pump A at 2 mL/min and oxidizer from pump B at 4 mL/min).

Upon completing the reaction, the 5'-triphosphate oligomer was subjected to a DEA treatment to remove the CE-group protection from the phosphates. The DEA treatment also partly destroyed the triphosphate modification as yellow coloring was seen leaving the column. The oligomer was purified through the methods in Section A.6.1 and an HPLC round, as in Section A.6.3.1, followed by isopropanol precipitation, as in Section A.6.2. MALDI-TOF characterization (Section A.7.1) was used to locate the triphosphate decorated oligomer from the HPLC elutions.

## A.9 RNA Oligonucleotide Synthesis

RNA synthesis is almost identical to DNA synthesis except for the 2'-hydroxyl found in RNA. In order to avoid creating 5'-2' phosphodiester in the RNA, the 2'-hydroxyl must be blocked to render it inert to reaction. To this end, the RNA phosphoramidites (rN-PA) contain a tert-butyldimethylsilane (TDMS-) protecting group at its 2'-hydroxyl (Figure A.24).

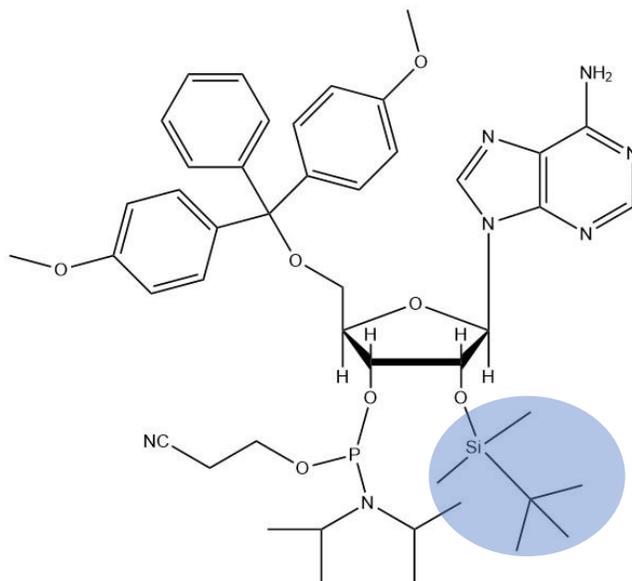


Figure A.24: rA-PA. Note the TDMS group added to protect the 2'-hydroxyl (in blue).

The side effects of adding this single protecting group are that the rN-PA are more bulky than dN-PA and that they are also more hydrophobic overall. RNA also requires a new deprotection step during the methods of Section A.6.1 in order to remove the 2'-TDMS group. The added bulkiness of the rN-PA means that 2 g of rN-PA will not synthesize as many base additions as 2 g of dN-PA, meaning that RNA are intrinsically more expensive to synthesize, added to the fact that RNA amidites are more expensive to purchase in those fractions to begin with.

Assuming RNA solid support is in hand, RNA can be synthesized on the oligopilot using the typical method wizard template RECYCLE 6mL PS5G

AKOPc121 Ed 005. In this, the sequence was set as RNA instead of DNA and the RNA sequence transcribed as per usual for DNA (the computer should allow for lower case nucleobase letters for the RNA nucleobases and accept “u” instead of “T”). A successful recycle time variable was 12 minutes to allow for the bulkiness of the molecule and the loading for the support will typically be less than for the equivalent DNA solid support. Upon creating the method, the nucleobase addition blocks will be named for the RNA base letters, but they will automatically draw from the analogous DNA amidite bottle (so, set up the synthesizer with rA in the dA bottle, rG in dG, rC in dC and rU in dT.) Synthesis should otherwise proceed indistinguishably from the DNA. RNA synthesis has been carried out successfully twice in my hands.

After synthesis, the purification methods remain unoptimized (see Figure A.20 for current efforts). The methods introduced in Section A.6.1 and A.6.2 will not work the same way for RNA as for DNA. My methods for purifying RNA are informed by a protocol acquired from an expert at GE Healthcare Lifesciences, but are currently non-identical.

Much as in Section A.6.1, perform the main deprotection overnight with 28% ammonium hydroxide at 50°C. When filtering out the solid support, I recommend using 100% ethanol for all 500 mL of washing and avoid adding water. This prevents the RNA from sedimenting on the glass frit while filtering the solid support away and will permit rotovapping of the resulting dissolved RNA. During the rotovapping step, be aware that if the water content is too high, the hydrophobicity of the TDMS-protected oligomer will cause RNA to behave as a strong surfactant that will produce an overpowering foam that prevents the rotovap from pulling liquid out efficiently. If the liquid is instead 100% ethanol, RNA remains fully miscible and the rotovap will work without a difficulty.

Unlike the situation expressed in Section A.6.1, the rotovapping must proceed to complete dryness in order to facilitate removal of the TDMS protection group. The deprotection reaction is performed in organic solution.

To begin the process of removing the TDMS, the rotovap solid is dissolved in several milliliters of Dimethylsulfoxide (DMSO). To this is added 10 mL of Triethylamine Trihydrofluoride (beware, do not touch on penalty of true suffering). The mixture is reacted for 3 hours at elevated temperature, ~60°C. Following reaction, 0.4 mL to 1 mL of 5 M Sodium perchlorate (in water) is added to the solution to help neutralize the acid. Addition of sodium perchlorate resulted in heavy precipitation and the solution emitted an odor like sewage; the precipitate can be removed by centrifuge where RNA remains in the soluble fraction. The RNA oligomer should be fully deprotected at this point and can then be carried on into additional purification.

The purification steps beyond this point are still somewhat ambiguous, but I recommend a pass through reverse phase HPLC (Section A.6.3.1) before attempting isopropanol precipitation (Section A.6.2). Immediate use of isopropanol precipitation after TDMS deprotection produced very poor purification probably because of all the unusual ions present in the mixture, though the RNA was identified by MALDI-TOF in roughly every soluble fraction obtained. As a warning, RNA is somewhat less stable than DNA because of the autoreactivity of the 2'-hydroxyl, so the appearance of failure sequences during purification should be unsurprising: the RNA can consume itself, though it is reputedly more stable in this regard at slightly alkaline pH.

## Appendix B

### Generalized nanoDNA LC Preparation Methods

Presented here is a collection of techniques used to prepare nanoDNA LC phases in (mostly) aqueous solution.

Many of the sample preparation techniques described in this section were for observation by polarized light microscopy (PLM, referring also to the microscope used for this form of microscopy). PLM is singularly useful for the examination of LC phases because of its ability to visualize microscopic structures for their birefringence. Birefringence is a phenomenon where the two perpendicular polarizations of light pass through a material with anisotropic indices of refraction, meaning that the speed of light within that material is slower along one axis of polarization than another. Birefringence is defined explicitly as the difference in index of refraction between the two polarizations.

Most molecularly pure organic substances (and many types of crystal) tend to be anisotropic where the distribution of electrons in the molecule are free to move only in particular directions within the bonding structures of that molecule. This means that light polarizations, reflecting the direction of the electric field in light, can interact with a molecule differently relative to how the molecule is oriented with respect to the impinging light. A bulk substance where these molecular

features are uniformly oriented will tend to display birefringence that can be examined to deduce how the constituent molecules in that bulk are oriented relative to the light. When the axis of motion of electrons lies along the vector describing the polarization of the E-field in light, they tend to slow the speed of light, creating a “slow axis” relative to the polarization direction of light that must occur perpendicular to the direction that light travels. This leads to a second important special axis used as a frame of reference, called the “optic axis,” which is defined as the axis within a material along which light can travel without slowing.

LC science defines a special direction within materials called the “director” which describes the orientation of LC molecules locally. For calamitic “rod-shaped” molecules, the director is defined to point along the long axis of the rod. These materials are referred to as “positively” birefringent because the slow axis of the molecule tends to align along the director of the phase so that waves of light impinging perpendicular to this axis exhibit birefringence. In this material, the optic axis lies along the director since light of any polarization can travel in the direction of the director without slowing. NanoDNA is a contrary case, called “negatively” birefringent, where the slow axis lies along the axis of DNA base-pairs, perpendicular to the long axis of the DNA double helix (and coincidentally perpendicular to the long axis of nanoDNA aggregates). The optic axis of DNA is somewhat trickier than the calamitic LC but can be considered to follow the long axis of the DNA double helix since birefringence will not be seen from light waves traveling in that direction (even though such waves would presumably end up uniformly slowed for both polarizations). Needless to say, examination of birefringence gives insight into average molecule orientation within a nanoDNA mesophase (see Figure B.1).

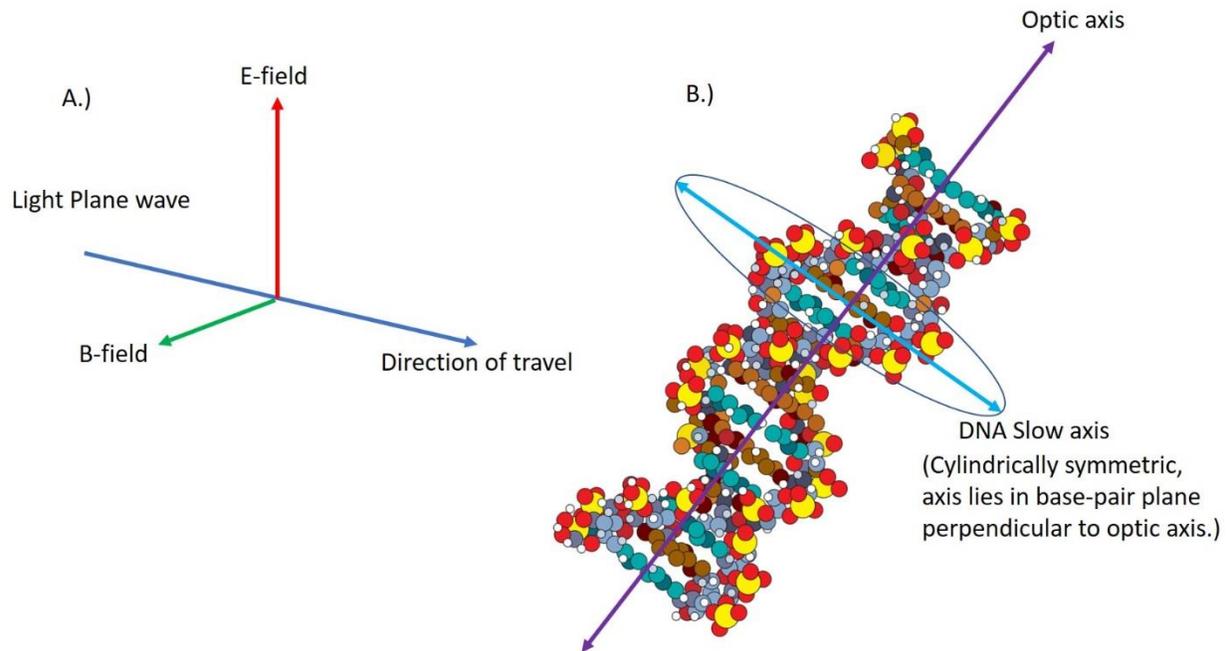


Figure B.1: Labeled optical slow axis diagram. A.) Relevant axes for impinging light. B.) Special optical axes for DNA.

To see birefringence, a PLM places two linear optical polarizers within the optics of the instrument (see Figure B.2). Between the light source and the specimen being examined is the “Polarizer” (P) which creates a uniform axis of polarization in the illuminating light. The second linear polarizer is called the “Analyzer” (A) and is placed between the microscopic specimen and the eye-piece. The axes of P and A are turned ninety degrees with respect to each other so that light passing through one cannot pass through the other. Materials lacking birefringence cannot alter light passing through them and will appear dark in the PLM since A will not pass light polarized along the direction of P. When polarized light is decomposed into its constituent vectors while passing through a birefringent specimen, one axis is slowed relative to the other, creating a beat frequency in the polarizations of the wave front interacting with itself when that light reaches A and the polarizations are forced to recombine into a single polarization. This causes

colors in the PLM image which are maximized when the slow axis of the specimen is pointed midway between P and A. If the slow axis is parallel to either P or A, the specimen is dark, forming an “extinction brush.” Specific colors occur because the distance of delay will place certain wavelengths of light into and out of phase at different delay distances, leading white light to appear red if the blue frequencies destructively interfere but the red frequencies do not. This is summed up by the Michel-Levy Chart (see Figure B.3) which allows an observed color to be associated with the thickness of a specimen and birefringence.

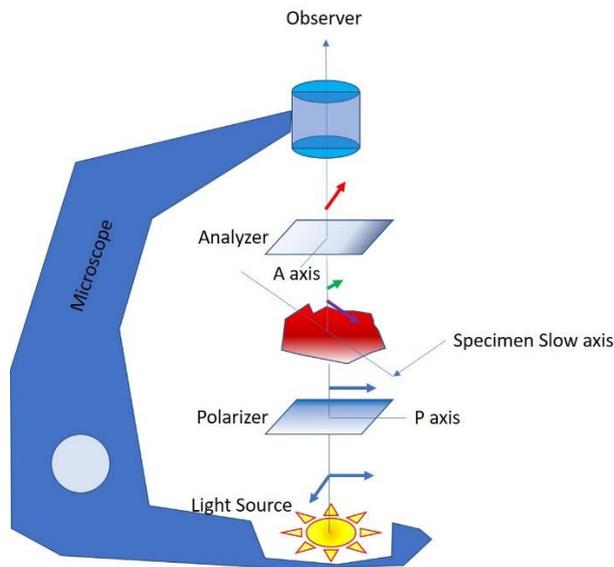


Figure B.2: Polarized light microscope diagram.

Most of the details of birefringence relative to sample thickness are used qualitatively in this work since thickness has been difficult to control in nanoDNA microscopic cells.

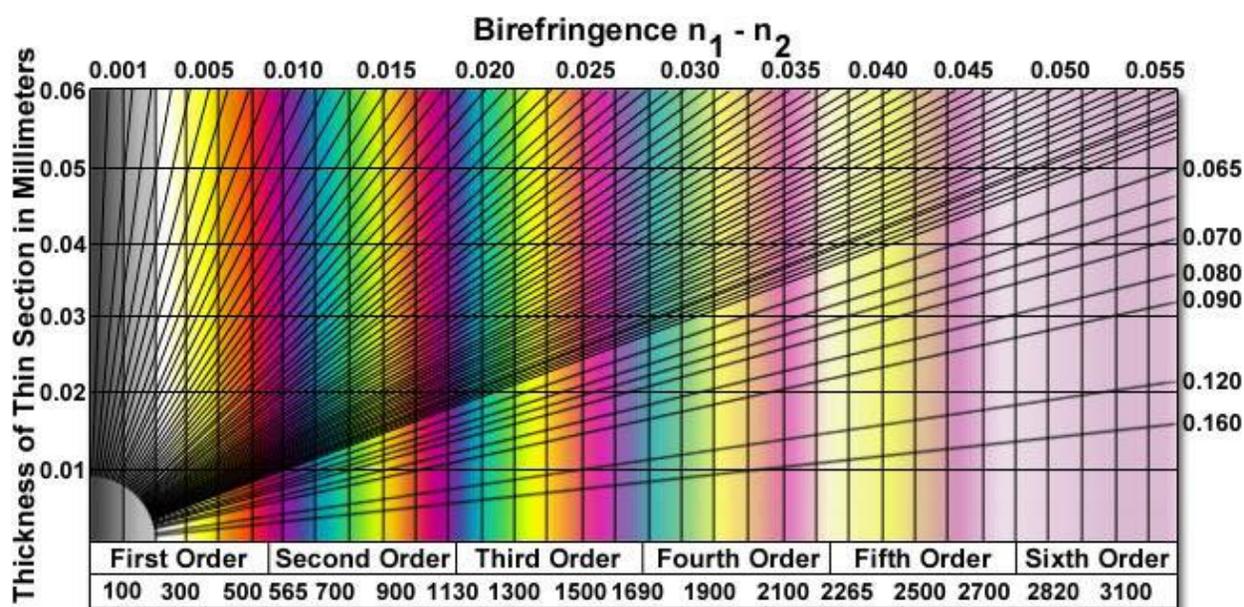


Figure B.3: Michel-Levy Chart taken from Olympus Scientific Solutions.

## B.1 nanoDNA LC Cells

Contained in this section is an overview of all of the different sample preparation formats used to examine nanoDNA LC phases. Since this LC is a lyotropic phase, it adds some complexity to preparation that is not encountered during the same type of observations made with thermotropic mesogens. The level of order seen within a phase is a direct consequence of the contents of that phase, including not only the oligonucleotide, but also the water content and any spectator ions that might be mixed in with the water. The complaint may seem mundane under other circumstances, but water presents a great challenge due to its volatility; preparation of a nanoDNA LC phase in most cell formats leads to samples that are only transiently stable. This transience can sometimes be exploited to examine multiple behaviors at once, but it becomes a deficiency during precision

measurements. Mixtures made of nanoDNA oligomer and water exhibited a penchant for slow mixing and rapid unmixing depending on what turned out to be most inconvenient behavior at the time. Sample preparation frequently demanded special consideration depending on the task at hand.

### B.1.1 Evaporative Cells

An aqueous mixture of nanoDNA is prepared to isotropic conditions (usually 100 or 200 mg/mL) and sandwiched between two pieces of glass. No seal is added around the edge and polarized light microscopy (PLM) is used to examine the fringe of the cell where the sample meets air. Water escapes by evaporation and the aqueous mixture rapidly increases concentration. Held at constant temperature, the fringe of the sample develops a concentration gradient into the interior of the cell which traverses the concentration axis of the phase diagram, highest concentration at the edge meeting the air to lowest concentration in the interior (see Figure B.4). The material textures of the nanoDNA phase diagram appear with the highest order textures at the edge and the lower order textures, ending in isotropic, in the interior.

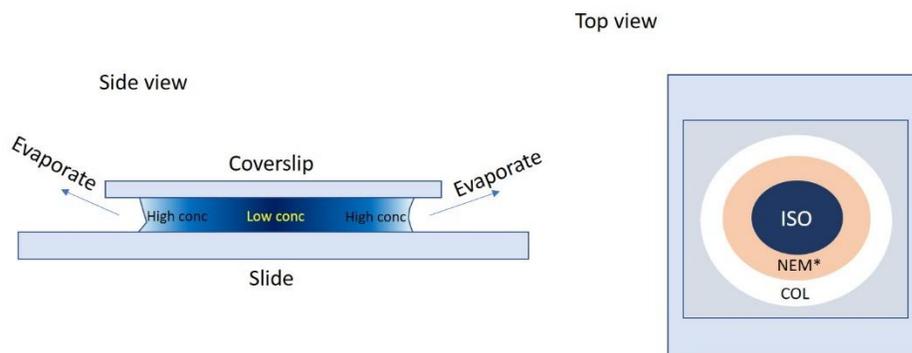


Figure B.4:  
Evaporative cell.

Evaporative cells are useful to quickly screen nanoDNA (or any lyotropic sample) for the presence of mesophases since the entire phase diagram can appear

within a single cell in stacked order on the concentration gradient. Since the loading is in the isotropic and therefore at low viscosity, the oligomer sample may be pipetted into the cell between glass plates that have a calibrated separation between them due to the inclusion of spacers.

The deficiency of evaporative cells is that they are not stable. Continuous evaporation of water means that phases are constantly changing concentrations and in ceaseless development toward higher order. Determination of local concentration is problematic given that any such measurement is not instantaneous and that the resolution size of such measurements may span an area too large to be very accurate. Evaporative cells are not stable to evaluation of the concentration axis because increasing the temperature enhances evaporation and drives the water out.

### B.1.2 Oil Sealed Cells

Evaporative cells may be temporarily stabilized by inclusion of an oil seal around the edge of the cell in order to dampen the occurrence of evaporation (see Figure B.5). Though nearly any immiscible oil can be used as the seal, some oils, such as fluorinated oils, can exhibit rapid evaporation that limits their viability. Mineral oils have fairly good long-term stability, but must be understood as imperfectly immiscible since water can cross into them in a small fraction. Mineral oils also exhibit some evaporation themselves when held at elevated temperatures.

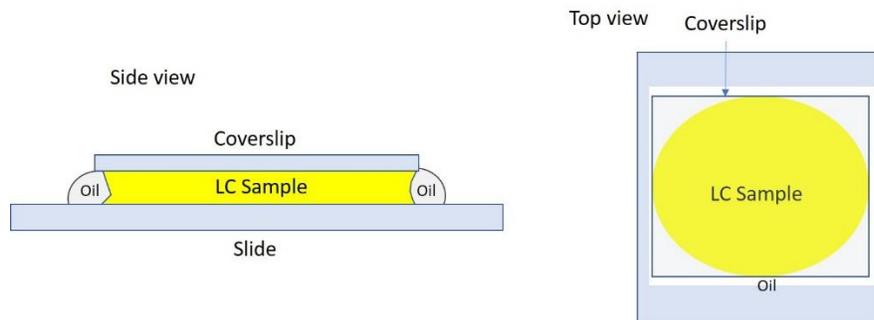


Figure B.5: Oil sealed cell.

Oil sealing an evaporative cell dampens the rate at which water leaves the cell. If the vapor pressure of the surrounding atmosphere is high, water diffusing through the oil layer will evaporate slowly enough to permit the concentration gradient in the sample to collapse by interior diffusion. This is rare in the dry climate where these experiments were carried out, so it is best understood that oil merely delays evaporation and that such a sealed cell is stable only on the scale of several days.

In addition to repurposing or stabilizing evaporative cells, an oil sealed cell can be constructed directly to examine particular oligomer concentrations. The oligomer is initially prepared to a concentration of choice and then spotted onto a glass slide in  $\sim 0.5 \mu\text{L}$  volume by a pipette where it is quickly covered by a glass coverslip and oil-sealed before the sample flows to the edge of the coverslip. This form of sealed cell is not easy to calibrate for sample thickness since the cover is added after the sample has already been spotted to the glass slide.

At room temperature, this cell type is stable for several days depending on the viscosity of the sample in use. Low concentration samples with low viscosity will tend to flow through the volume of the cell, displacing oil, until they reach edges of the coverslip, at which point they lose water rapidly through the thinness of the oil. High concentration samples do not flow as easily, but also are not easy to pipet into the cell since they may resist pipetting due to their viscosity. Further, one must always be aware that spotting of the sample onto the cell takes a finite time and that during this time, the small sample volume is exposed to the atmosphere and is constantly losing water, creating a concentration gradient at the aperture of the pipet tip in a matter seconds. High order phases may then be visible at the loading contact point in the completed cell due to a concentration gradient appearing in the pipet tip and lower order phases will tend to flow away from higher order phases as

a result of differences of viscosity. Though a sample may have been prepared at a single concentration, oil sealed cells can exhibit concentration gradient effects immediately after preparation. Further, though cells may be stable long enough to equilibrate, if the domain of the sample fragmented inside the cell prior to equilibration, the resulting daughter domains may exhibit different concentrations.

Data taken with oil-sealed cells typically reports higher order phases at lower than their actual concentrations due to evaporation. A phase prepared at low concentration loses a small amount of water in the process of loading the cell, creating a higher concentration than initially prepared, resulting in higher order phases appearing at lower apparent concentration. This effect is significantly exacerbated at high concentrations because less water is present in the sample overall and a similar evaporation rate can lead to a greater proportionate error. Reported values at high concentrations tend to be badly attenuated away from their actual values in phase diagrams made by these sorts of measurements.

A method which helps to diminish the effect of evaporation is to load through oil. Here, an oil droplet is spotted onto the glass slide, and then the oligomer sample is pipetted directly into the oil before adding the coverslip. This can reduce the period when water is evaporating out of the sample prior to sealing, but suffers from the deficiency that a low viscosity sample may force its way out from underneath the coverslip due to the excess motion of oil when dropping the coverslip into place.

High concentration samples can be challenging to work with in this sampling format because their viscosity prevents them from being easily pipetted. They sometimes need to be treated as a sticky solid and scraped directly onto the glass slide with a small spatula where their concentrations tend to change very rapidly

throughout. Under these conditions, it is not possible to put thickness spacers in a cell.

Oil sealed cells are more tolerant to temperature cycling than evaporative cells. If the sample is located a long distance from the edge of the coverslip (~0.5 cm) the sample can be very stable and will resist changing concentrations to temperatures of ~60°C. If the sample undergoes a shift in viscosity with elevated temperature, it can flow closer to the edge of the cell and may decrease in concentration stability as a result. For temperature cycling to 85°C, it is best to assume that the cell is not stable beyond one cycle, especially if the sample moves near to the edge.

### B.1.3 Double Pane Oil Sealed Cells

One feature that significantly limits the stability of an oil sealed cell is the separation of the sample from the edge of the cell. Typically, sample left in an oil sealed cell for several days will tend to evolve over time to show high order phase textures in regions that are close to the edge of the glass coverslip and lower order textures where the sample is far from the edge. Even with oil prohibiting direct contact of the surrounding atmosphere with the oligomer sample, water will escape through the oil and exit into the air and the rate at which this occurs is dependent on the distance between the edge of the cell and the location of the sample –more oil between the sample and the air means less water escaping. In oil sealed cells, as described in the previous section, there is nothing stopping the sample from flowing inside the cell to regions where it is in close proximity to the air, thus decreasing the stability of the sample.

One way to help stabilize an oil sealed cell is to introduce some form of mechanical impediment to prevent the sample from flowing too close to the edge of the cell. As it was observed that samples tend to flow toward regions of a cell that are thinner than others and that they do not usually pass the edge of the glass coverslip, it was reasoned that surface tension is one tool that might be used to accomplish this. Therefore, a second glass coverslip is introduced in order to create an edge recessed far back inside the oil layer beyond which the sample refuses to flow (see Figure B.6). This double pane coverslip design is seen to significantly increase the concentration stability of a sample.

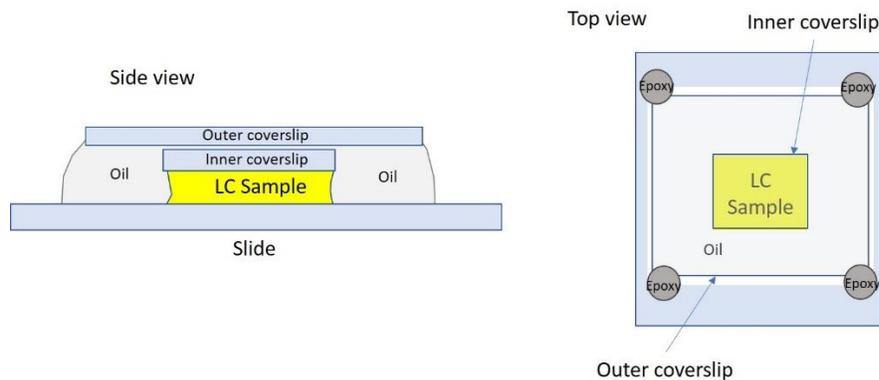


Figure B.6:  
Double pane oil  
sealed cell.

The double pane cell is prepared by cutting the inner pane, made of thin coverslip glass ( $\sim 100 \mu\text{m}$  thick), about 0.5 cm smaller on all sides than the outer pane. The sample is prepared to the desired concentration and spotted by pipet to a glass slide. The inner pane coverslip is placed over the sample and the pane is sealed by oil to help prevent evaporation while the remainder of the cell is closed. A small drop of oil is placed on top of the inner pane in order to prevent air entrapment between the panes and then the outer pane is placed over the inner pane, but centered upon it so that it overhangs the inner pane by the same amount on all sides. The outer pane is glued in place with epoxy. Oil is then added beneath the outer pane to fill the remaining open space up to the edge.

Double pane oil sealed cells exhibit significantly improved stability over oil sealed cells, making it possible to temperature cycle up to 70°C or 80°C on the short term with less evidence of concentration attenuation. This configuration of cell has been observed to hold the same texture of phase on the scale of weeks.

In one version, with a low viscosity sample, the inner pane of the cell was glued in place using spacers and measured by interferometry to calibrate thickness, and then loaded by capillary force before sealing with oil and adding the second pane. Calibrating thickness here can be challenging because the thinness of the glass can cause some interference effects on top of the air-gap of the cell.

Again, high viscosity samples did not permit use of spacers to regulate cell thickness.

#### B.1.4 Flame-Sealed Capillaries

Flat cells permit the best format for examining phase textures, but every kind of flat cell attempted or used exhibited signs of water loss in the long term or under rigorous temperature cycling. The most stable sample is the flame-sealed borosilicate melting temperature capillary (Kimble Art No. 34505-99, dimensions 1.5-1.8 x 90 mm).

The process of placing a nanoDNA LC sample inside a capillary tube proved challenging because higher order nanoDNA mesophase samples tend to have such high viscosity that they do not flow. In order to flow appreciably, samples must be heated, at which point they rapidly lose water to evaporation and become correspondingly less prone to flow. This linkage of water evaporation with sample viscosity makes it very difficult to put a nanoDNA mesophase into a capillary and still know exactly the concentration after closure. Additionally, closing these

capillaries proved problematic because any sample left smeared at one end or the other, even in very small amounts, burns during flame sealing and prevents a clean, stable closure.

One way to avoid all of these problems is to construct the mesophase inside the capillary. Oligomer is added to the capillary as a granular powder and then water is added to it afterward. Water and granular powder can both be centrifuged into a capillary cleanly and with minimum effort independently of one another, after which the capillary is sealed. The water-oligomer mixture can then be equilibrated with temperature cycling. By monitoring how much of each component is added in weight, the concentration can be determined.

This method can be executed in different centrifuges, but the most effective machine will depend on the preparation of the solid oligomer. If the oligomer is lyophilized into a low density sponge (from high proportion of water and low proportion of DNA), a strong centrifuge will be needed to load it into the capillary, but if the oligomer is very dense and granular (from low proportion of water and high DNA in lyophilization) it can be tapped in by hand with little or no centrifuging. This latter technique makes it possible to load capillaries less robust than the preferred borosilicate variety listed above, but it has not been mastered. The high force centrifuge (a Sorvall R70 with an AH-629 swinging bucket rotor) usually needed to spin down low-density oligomer powder necessitates the use of robust cylindrical capillaries. Capillaries of different cross section or thinner glass are pulverized by this treatment for lack of sufficient tensile strength and strong centrifuges that place force in a shearing direction across the capillary rather than along the long axis, as would be encountered in a fixed angle rotor, will tend to crush anything made of glass. Holders of the swinging bucket rotor typically used were fitted with a Delrin holder floating on a layer of dense fluorinated oil (3M

Fluorinert FC-72 or FC-77). 3,500 rpms was often sufficient to spin down nanoDNA oligomer, but 20,000 rpm was possible for stubborn samples.

Preparation of capillary samples proceeded as follows. A selected capillary was tested to see that it slid freely in and out of the holder inside the bucket capsule of the centrifuge rotor and the top end of the capillary shortened by several centimeters to prevent it from hitting the top of the bucket during centrifugation inside the rotor. The closed end of the capillary was checked to see that it remained intact. The capillary was weighed to mark its empty weight to the 1/100<sup>th</sup> of a milligram. The capillary was then loaded by pressing its open end gently down into a glass crucible containing powdered nanoDNA in order to force small amounts of the DNA powder into the opening of the capillary. A glass crucible works better than plastic during loading in order to avoid the capillary shaving the plastic chunks into the sample and contaminating it with solid particulates. DNA should not be pressed with too much force into the opening of the capillary in order to avoid jamming the powder and preventing it from falling easily to the bottom of the capillary. Loading pressure should be light and the capillary should frequently be pulled back and tapped to bring nanoDNA to the bottom, with only a very small amount loaded at a time. Typical samples load a total of 1-2 mg of nanoDNA powder. Low density nanoDNA powder may stick along the sides of the capillary, necessitating a centrifuging step in order to bring DNA dust to the bottom of the capillary and clean the sides of remaining solid -- 3,500 rpm is typically enough centrifugal force to bring the powder to the bottom of the capillary. The capillary is then weighed again in order to mark its total mass plus that of the newly added DNA powder. The quantity of water necessary to render the desired concentration is calculated as an estimate and that amount is added into the mouth of the capillary. 1 to 2  $\mu$ L of water can be added quite easily by micropipet while larger quantities

should be added by hypodermic needle in order to insert the water more deeply. The capillary is immediately centrifuged again to bring the water down onto the nanoDNA powder. Powder and water usually do not mix well at this point. The actual quantity of water added never quite matches the amount desired, so the capillary must be weighed again to determine if an acceptable amount of water is inside. If the water is sufficiently close to the amount desired, this is taken to be a stopping point, though more water can be added using the same methods. For closure, the capillary is taken to an oxygen-propane torch and sealed by pull-closure: the flame is directed at a point midway down the capillary and the top and bottom of the capillary are pulled away from each other after the glass softens, causing the capillary to split in half with both ends at the flamed point sealing up. The end of the capillary containing the sample should be sealed now from both ends. All sections of the capillary are weighed a final time. At this point, the sample is usually sealed but unmixed.

There was some initial concern that this method of creating the sample inside the capillary would yield false weights if water wets the borosilicate glass and the full water weight is spread across the walls and fails to reach the bottom of the capillary before closure. This was tested by spinning water into the capillary, then inverting the capillary and spinning it back out again: the mass difference of the capillary before and after was less than the error of the scale.

Once water and nanoDNA powder are united at the bottom of the capillary, they must be mixed. Mixing is accomplished by heating the sample to 60°C or 70°C for a short time. Typically, the capillary is positioned such that the portion containing the sample is at the coolest point on the heating element in order to force water vapor to remain with the nanoDNA and not migrate out. If the nanoDNA is

not too spongy as a powder, the capillary can then be left at room temperature overnight and is found to be fully equilibrated after.

If the nanoDNA powder is very spongy, as the DNA dissolves into the added water it may release air bubbles that open a space that prevents the water from connecting farther distant DNA. Usually, the temperature ramp to 70°C will cause vapor to cross such an impediment and wet the still-dry DNA, but there will remain a large concentration differential. This can now be collapsed by returning the capillary to the centrifuge and spinning 3,500 to 7,000 rpms. At this point, the water should reach the bottom. Another temperature ramp to 70°C (or into the isotropic) will induce more complete uniformity to the concentration at this point.

Often, after this process, the sample will be very much more uniform in concentration, but may exhibit large bubbles incorporated with the sample layer inside the capillary. If removal of these bubbles is desired, heat the centrifuge bucket until it's ~90°C in nearly boiling water, then remove it from the water, add the capillary to it and centrifuge for 5 min at 3,500 to 7,000 rpm. The heated centrifuge bucket will put the sample into an isotropic state and the centrifugal force will eject the bubbles.

Equilibration of a capillary sample can be judged by examining the behavior of the mesogen when crossing a phase transition during a temperature ramp, particularly while decreasing temperature and crossing the line from Iso to another phase. The sample is well equilibrated if the ordered phase begins to appear at points everywhere in the capillary simultaneously. A concentration gradient is apparent if the ordered phase begins to appear at different points in the capillary at significantly different temperatures.

Phases inside a capillary are significantly more difficult to examine for their texture than phases in a flat cell. The curved surface of a cylindrical capillary will act as a lens that distorts light passing through it, making focus difficult. This can be partly remedied by building a refraction index matching device (see Figure B.7). Such a device is built by stacking and epoxy gluing a series of cut glass slides to form a trough wide and tall enough to accommodate the capillary lying on its side. The trough is filled with oil, covered by a glass slide and then adjusted until all bubbles have exited from beneath the covering slide so that microscopic examination is only through oil, glass and sample. Heavy mineral oil matches index with glass fairly well, but a carefully selected oil would be ideal. With indices matched, it should be more possible to focus on fine features inside the capillary while performing microscopy. This solution is imperfect because the refraction index of glass does not match that of water, meaning that light passing through the interior of the capillary gets bent at the sample-glass interface, creating a small distortion that cannot be addressed.

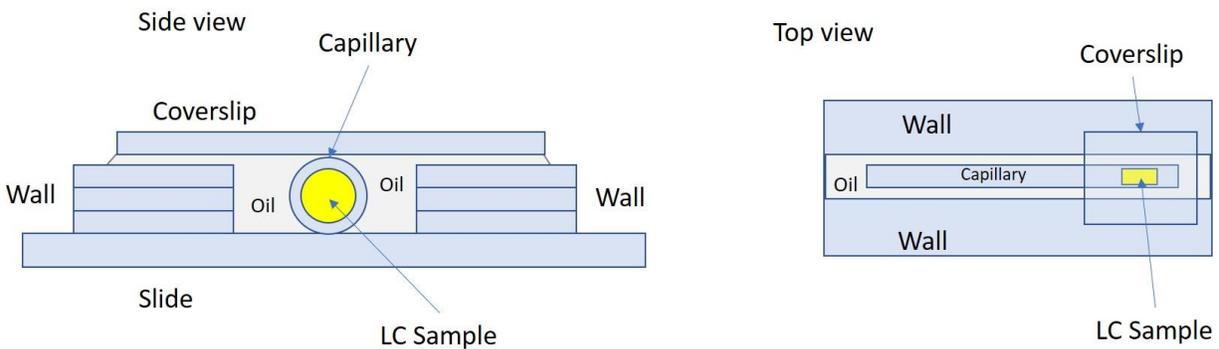


Figure B.7: Index matching device for observing LC sample in a round profile capillary.

### B.1.5 Oil Sealed Free Surface Cells

Because higher order phases of nanoDNA (and similar mesophases) do not appreciably flow due to their very high viscosity, it was noted that experimental

preparation of high order phases presents the unique challenge where such a sample cannot be easily transplanted from a vessel where the water-nanoDNA mixture was made to an examination substrate. Viscosity can be reduced by heating a sample, but this presents the contradictory challenge where water rapidly evaporates out of the sample, quickly altering the concentration and potentially the phase behavior. For high concentration phases, this is doubly challenging because water is not present at a very large quantity, meaning that little evaporation more significantly alters the concentration. In the case of dNTP liquid crystals (encountered in Section 4.0), the chemical itself is of uncertain stability at elevated temperatures, making it prudent to avoid large temperature swings during preparation.

While precision at concentration preparation becomes very hit or miss, one method for examining the morphology of high order phases is to adopt roughly the opposite strategy from what is used in the evaporative cell (seen in Section B.1.1). Here, the sample is obtained as a dried solid on a glass substrate, sealed by a droplet of oil so that water evaporates only slowly and then a droplet of water is added onto the solid through the oil layer and permitted to diffuse into the solid sample (see Figure B.8). At the contact line between the water and solid oligomer (or dNTP) the mixture explores the various mesophases available to it.

Despite the lack of precision in the concentration, these samples can be established to favor very high concentration phases. This is accomplished by limiting the quantity of water added relative to the solid oligomer sample mass present on the substrate.

Preparation methods are slightly different for nanoDNA versus dNTP. NanoDNA is typically stored as a lyophilized solid while dNTP is acquired often

pre-mixed in an aqueous phase at a predetermined concentration. Differences between the preparative methods will be noted where relevant.

The first step of preparation is to acquire a cut section of glass intended for use as the microscopy substrate. This glass should be cleaned and carefully weighed. Of note, the method of cleaning can affect how sample will later dry onto the surface of this substrate: cleaning with acetone will result in a different dried sample texture from cleaning with water. Whether this fundamentally alters the sample is unknown. For the sake of consistency and safety, clean with water and dry the substrate by heating it in an oven or on a hot stage. After the substrate is clean, determine and record the weight.

Sample can now be added to the surface of the substrate. For nanoDNA oligomer, add the solid sample directly to a spot on the surface of the substrate and note the weight. Best results will be at  $\leq 1$  mg of sample. Add sufficient water as a droplet over the nanoDNA solid to cause that solid sample to dissolve into isotropic phase; a 20  $\mu\text{L}$  droplet of water should suffice. For dNTP, pipet the aqueous sample to a spot on the surface of the substrate, usually adding a known volume at a known concentration. As dNTP concentrations are usually  $\sim 100$  mM, it may be necessary to add several tens of  $\mu\text{L}$  to the slide in order to attain a sufficiently measurable bulk of solid later; a 10-20  $\mu\text{L}$  drop is sufficient for microscopy purposes.

With an aqueous droplet of sample on the slide, place the slide in a rough vacuum and dry it 45 minutes to 1 hour. The isotropic droplet will flatten into a disc shape on the glass surface and may turn cloudy. A cloudy texture is often found to correspond to birefringence when examining the sample under polarized light microscopy, so this can constitute an extreme concentration phase on the phase diagram and may warrant examination on its own.

The composite of dried sample and glass substrate should be weighed. This gives an estimate of any residual water in the system and can be useful for concentration calculations of the material.

An oil drop is added over the solid sample to seal it. Water is then added by micropipette through the oil layer onto the surface of the sample beneath. As mentioned, the amount of water added is typically sufficient such that if the oligomer and dNTP fully mix, it will produce a certain desired, very high concentration. The sample is then allowed to age, typically at a controlled temperature on a microscope hot stage. Birefringence will evolve at the contact line between the water and the solid sample where water will gradually diffuse into the solid and fringes of the solid will dissolve out into the isotropic water, creating a concentration gradient at the interface.

## Fabrication of oil sealed free surface cell

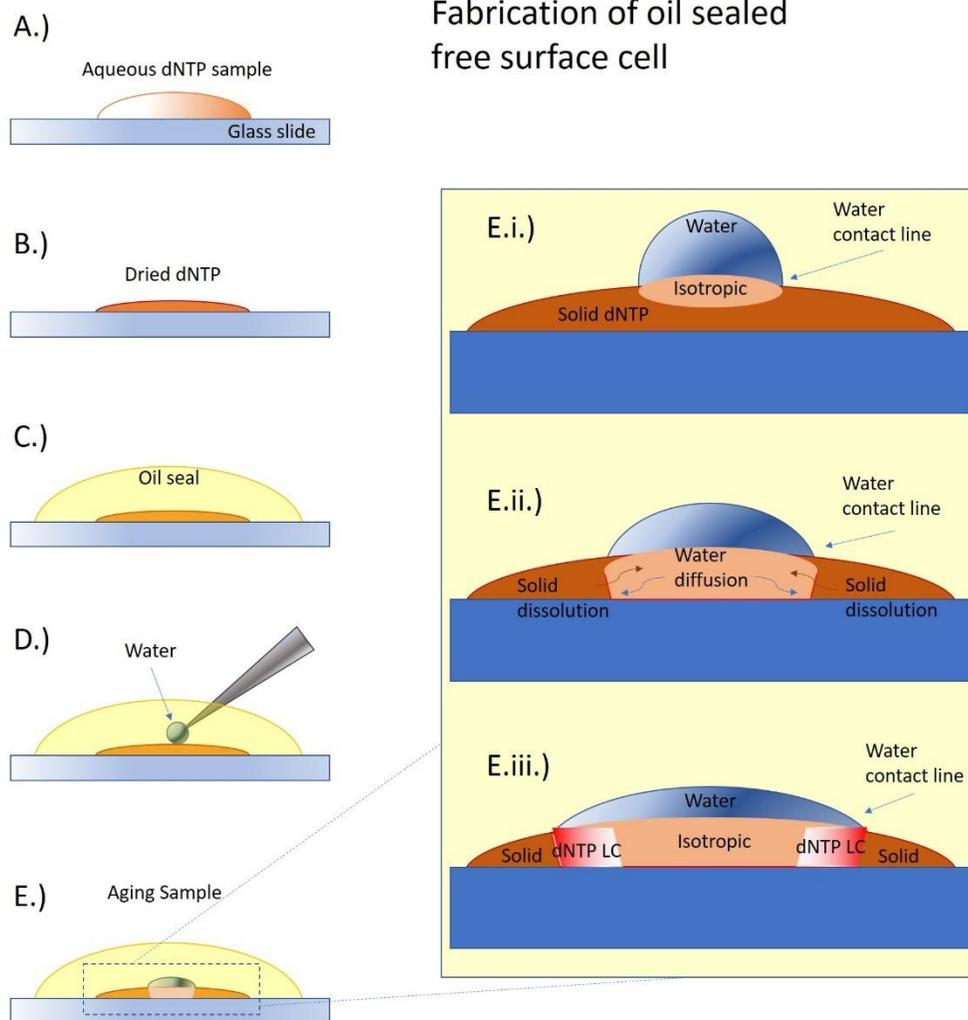


Figure B.8: Preparation and development of oil sealed free surface cell, particularly as seen during experimentation with dNTP. A.) Aqueous sample is pipetted onto a glass substrate. B.) The sample is allowed to dry on the substrate under vacuum. C.) An oil seal is added over the dry sample to remove it from direct contact with the atmosphere. D.) A small quantity of water is pipetted through the oil seal onto the surface of the solid sample. E.) The water and solid samples mix. E.i.) an isotropic region develops at the contact interface, typically melting straight through the solid sample at the earliest stage of mixing. E.ii.) water and solid sample mobilize across the contact line creating concentration gradients. E.iii.) birefringent domains appear as the concentration gradients steepen to higher concentrations due to the overall paucity of water relative to the amount of available solid.

This approach has several complications that should be noted. First, the oil-sealed free surface sample is not stable in true long term particularly under

conditions where humidity is high; water from humidity can invade the sample through the large oil interface. The oil seal constitutes only a “good” barrier to water entering and leaving, but very imperfectly under conditions where small amounts of water can be continuously forced through by a large chemical potential. Limiting the size of this interface in the long term can increase stability. Second, this configuration of sample cell places a curved air-oil interface between the microscope objective and the sample, creating optical distortions microscopically which can make photography challenging. Both of these complications, to stability and to microscopic imaging clarity, can be addressed by simply dropping a glass coverslip over the sample after birefringent phase gradients have been established, reducing the oil interface size with inclusion of a glass wall covering most of the access and flattening the interface so that light does not refract unduly when passing through it. Be warned that dropping a glass slide onto this form of sample will tend to displace anything with low viscosity beneath it; if significant quantities of low viscosity ISO or NEM phases are present, they will be rearranged and the oil flows may carry them out from underneath the falling glass. High viscosity phases are observed to remain undisturbed.

It has not been possible to calibrate cell thickness in this sample format and birefringence measurements can only reveal the sign of birefringence.

## B.2 Concentration Calculations

As mentioned, nanoDNA mesophase behavior is pivotally dependent on the content ratio between oligomer and water and any spectator that might be occupying a significant volume in the sample. Presented here are calculations used

to determine nanoDNA concentrations in various different circumstances encountered during experimentation.

Almost all concentration calculations are based upon mass measurements made using a Toledo-Mettler precision scale.

### B.2.1 Legacy Concentration Calculations

These simpler calculations were used as bench calculations for preparing flat cells and capillaries due to their portability; for the complete treatment refer to the next section.

Concentration ( $c$ ) was always reported in milligrams of nanoDNA per milliliter of total volume.

$$c = \frac{m_{dna}}{V_t} \quad (\text{B.1})$$

The density of nanoDNA was taken to be  $\sim 1800$  mg/mL (as in Nakata<sup>25</sup>). Total volume ( $V_t$ ) of a sample is composed of water ( $V_w$ ) and nanoDNA ( $V_{dna}$ ) where the density of water is taken to be 1000 mg/mL (or 1 g/mL) as per the standard of water density. Concentration is simply mass of DNA in milligram ( $m_{dna}$ ) per total volume. During preparation of a sample, a target concentration is arbitrarily decided and nanoDNA mass is weighed into a preparatory vessel, so that  $m_{dna}$  and  $c$  are both known and the water volume is calculated in order to produce the desired concentration.

$$V_t = V_w + V_{dna} \quad (\text{B.2})$$

$$V_t - V_{dna} = V_w \quad (\text{B.3})$$

The value for density of DNA can be subbed in for volume of DNA from the density relation.

$$\rho_{dna} = \frac{m_{dna}}{V_{dna}} \rightarrow V_{dna} = \frac{m_{dna}}{\rho_{dna}} \quad (\text{B.4})$$

These can be substituted into the volume relation. Total volume is also multiplied by 1 to insert  $m_{dna}$  with the other term so that DNA mass can be pulled out.

$$V_t \frac{m_{dna}}{m_{dna}} - \frac{m_{dna}}{\rho_{dna}} = V_w \quad (\text{B.5})$$

$$m_{dna} \left( \frac{V_t}{m_{dna}} - \frac{1}{\rho_{dna}} \right) = V_w \quad (\text{B.6})$$

Total volume can then be extracted by substituting in the desired concentration, and by making the approximation that the weighed mass ( $m_{meas}$ ) is the same as the DNA mass ( $m_{dna}$ ).

$$m_{meas} \left( \frac{1}{c} - \frac{1}{\rho_{dna}} \right) = V_w \quad (\text{B.7})$$

This calculation is readily used and portable for determining the amount of water which needs to be added to a measured mass of DNA in order to produce a desired concentration. This calculation is sufficient for design of flat-celled samples given the evaporation of water from such samples, making concentration mostly approximate to begin with.

While the equation B.7 is useful in planning and constructing samples, the situation encountered during production of a capillary cell affords a retroactive measurement of how much water mass ended up actually added to the sample. The associated concentration in such a cell can be calculated by inverting equation B.7.

$$c = \frac{1}{\frac{V_w}{m_{meas}} + \frac{1}{\rho_{dna}}} \quad (\text{B.8})$$

The density value for DNA used here turns out to be too high, creating a systematic error that swings toward overly high concentrations. Further, it was determined later that DNA oligomer mass is not ever reportable as purely DNA, containing a variety of potential contaminants, including potential excess water mass, counter ions and spectator ions introduced as salt. For purposes of DNA LC phase production, the formation of phases tends to be dependent on at least some of these contaminants, most critically counter ion content, necessitating better precision at calculating concentration.

### B.2.2 Higher Precision Concentration Calculations

This method parallels the previous method but moves to accommodate volumes that can be attributed to potential sources of error. There are two direct measurements that are made of samples in the process of preparation, the mass of the solid containing the DNA and the volume of the water added to it. The total volume is the sum of the DNA volume plus all the potential contaminant volumes.

$$V_t = V_{dna} + V_w + V_{xw} + V_{Na} + V_{con} \quad (\text{B.9})$$

The water deliberately added to the sample ( $V_w$ ) must be corrected for any amount of water not deliberately added ( $V_{xw}$ ). Moreover, DNA is not present in a pH neutral form without an amount of Sodium ( $V_{Na}$ ) and very likely also an amount of spectator ions, probably from salt contamination ( $V_{con}$ ). Concentration is DNA per total volume (eqn B.1).

Measurement of the solid mass cannot be taken directly as the mass for DNA; sample mass containing DNA also contains most of the spectators, counter ions and excess water noted above. Added water is assumed to be deionized and highly pure, so all contaminants come in with the measured mass. The total measured mass

( $m_{meas}$ ) is expanded as a sum of DNA mass ( $m_{dna}$ ), excess water mass ( $m_{xw}$ ), counter ion mass ( $m_{Na}$ ) and spectator content ( $m_{con}$ ).

$$m_{meas} = m_{dna} + m_{xw} + m_{Na} + m_{con} \quad (B.10)$$

The constituent masses in this relation can be expressed in proportions of various useful masses. Sodium counter ion must be present in a solid DNA sample at a stoichiometric quantity to the DNA given the requirement that the counter ion must neutralize the DNA charge at neutral pH; counter ion is therefore present in a direct proportion to DNA. Excess water and spectator content can be taken as proportions of measured mass which must be obtained from additional measurements.

$$m_{Na} = k_1 m_{dna} \quad (B.11)$$

$$m_{xw} = k_2 m_{meas} \quad (B.12)$$

$$m_{con} = k_3 m_{meas} \quad (B.13)$$

These three relations can be used to reduce the masses (eqn B.10) to only total measured mass of the solid and the DNA mass.

$$m_{meas} = m_{dna} + k_2 m_{meas} + k_1 m_{dna} + k_3 m_{meas} \quad (B.14)$$

$$m_{meas}(1 - k_2 - k_3) = m_{dna}(1 + k_1) \quad (B.15)$$

This gives a relation to determine mass of DNA as a proportion of the measured mass.

$$m_{dna} = m_{meas} \left( \frac{1 - k_2 - k_3}{1 + k_1} \right) \quad (B.16)$$

As densities for water, DNA (eqn B.4) and sodium ion are all knowable, these too can be used to help determine the overall concentration. Further, as the identities of the spectators cannot be well known, these are left unrepresented.

$$V_{xw} = \frac{m_{xw}}{\rho_{xw}} = \frac{k_2 m_{meas}}{\rho_{xw}} \quad (\text{B.17})$$

$$V_{Na} = \frac{m_{Na}}{\rho_{Na}} = \frac{k_1 m_{dna}}{\rho_{Na}} \quad (\text{B.18})$$

These can be substituted into the total volume relation (equation B.9).

Because it is a measured quantity,  $V_w$  is unaltered.

$$V_t = \frac{m_{dna}}{\rho_{dna}} + V_w + \frac{k_2 m_{meas}}{\rho_w} + \frac{k_1 m_{dna}}{\rho_{Na}} + V_{con} \quad (\text{B.19})$$

Concentration can be brought in as before, as a ratio of DNA mass per total volume.

$$V_t = m_{dna} \left( \frac{1}{\rho_{dna}} + \frac{V_w}{m_{dna}} + \frac{k_2 m_{meas}}{\rho_w m_{dna}} + \frac{k_1}{\rho_{Na}} + \frac{V_{con}}{m_{dna}} \right) \quad (\text{B.20})$$

$$\frac{1}{c} = \left( \frac{1}{\rho_{dna}} + \frac{V_w}{m_{dna}} + \frac{k_2 m_{meas}}{\rho_w m_{dna}} + \frac{k_1}{\rho_{Na}} + \frac{V_{con}}{m_{dna}} \right) \quad (\text{B.21})$$

$$c = \left( \frac{1}{\rho_{dna}} + \frac{k_1}{\rho_{Na}} + \frac{1}{m_{dna}} \left( \frac{k_2 m_{meas}}{\rho_w} + V_w + V_{con} \right) \right)^{-1} \quad (\text{B.22})$$

DNA mass is eliminated by substituting its proportion of measured mass (eqn B.16).

$$c = \left( \frac{1}{\rho_{dna}} + \frac{k_1}{\rho_{Na}} + \frac{1}{m_{meas}} \left( \frac{1 + k_1}{1 - k_2 - k_3} \right) \left( \frac{k_2 m_{meas}}{\rho_w} + V_w + V_{con} \right) \right)^{-1} \quad (\text{B.23})$$

$$c = \left( \frac{1}{\rho_{dna}} + \frac{k_1}{\rho_{Na}} + \left( \frac{1 + k_1}{1 - k_2 - k_3} \right) \left( \frac{k_2}{\rho_w} + \frac{V_w}{m_{meas}} + \frac{V_{con}}{m_{meas}} \right) \right)^{-1} \quad (\text{B.24})$$

The form of this final equation can be cleaned up by extracting several of the quantities as independent constants.

$$G_1 = \frac{1 + k_1}{1 - k_2 - k_3} \quad (\text{B.25})$$

$$G_2 = \frac{V_w}{m_{meas}} \quad (\text{B.26})$$

$$G_{con} = \frac{V_{con}}{m_{meas}} \quad (\text{B.27})$$

$$c = \left( \frac{1}{\rho_{dna}} + \frac{k_1}{\rho_{Na}} + G_1 \left( \frac{k_2}{\rho_w} + G_2 + G_{con} \right) \right)^{-1} \quad (\text{B.28})$$

The concentration equation depends on accurate measures of the densities of each species. The density of water is well understood as 1000 mg/mL. With an ionic radius of 1.16 Å and a molecular weight of 22.9898 g/mol, sodium density is calculated to be  $\rho_{Na} = 6091.8$  mg/mL. Replacing the value of 1800 mg/mL, DNA density is corrected to 1687 mg/mL based on the average Van Der Waals radius of oligonucleotide crystal structure<sup>44</sup>. In most nanoDNA calculations, the spectator contaminant is taken to be close to negligible so that constant  $G_{con}$  is omitted. Constant  $k_2$  depends on the dryness of the solid nanoDNA sample at the time it is measured by scale. For most nanoDNA samples prepped by lyophilization, contaminant water content can also be taken as negligible, allowing  $k_2$  to be brought to zero. Constant  $k_1$  must be calculated for the specific oligomer species being used. For example, Drew Dickerson Dodecamer, 5'CGCGAATTCGCG3'; molecular weight with phosphates fully deprotonated, 3636.4 g/mol with eleven phosphates, requires eleven counter ions.  $11 \times 22.9898 = 252.89$ ;  $m_{Na} = k_1 m_{dna}$  would imply  $k_1 = 252.89/3636.4 = 0.0695$ . Constant  $G_2$  can be calculated directly from the measurements of water volume and solid mass added during sample preparation.

With adjustments noted in the previous paragraph, the concentration calculation simplifies.

$$c = \left( \frac{1}{\rho_{dna}} + \frac{k_1}{\rho_{Na}} + (1 + k_1)G_2 \right)^{-1} \quad (\text{B.29})$$

Constant  $G_2$  can be used to modify concentrations calculated as in Section B.2.1. In this way  $G_2$  can be acquired from benchtop calculations.

$$G_2 = \frac{1}{c_{old}} - \frac{1}{\rho_{old}} \quad (\text{B.30})$$

Here,  $\rho_{old}$  is the DNA density used in Section B.2.1 and  $c_{old}$  is the concentration calculated during preparation.

Depending on what system is being examined, accuracy in calculation of concentration can be adjusted by including whatever additional terms are necessary to cope with any discerned contaminants.

## Appendix C

### X-ray Diffraction Theory and Methods

Contained in this section is a detailed overview of all theory and methods used to interpret X-ray diffraction data obtained during the research presented in this thesis. The derivation presented here is motivated and informed by Drenth<sup>57</sup> and Guinier<sup>58</sup>, but does not necessarily duplicate those works since the purpose here is specialized most directly toward the 2D mesophase structures encountered with nanoDNA.

#### C.1 Useful Basic Theory of X-ray Diffraction in Liquid Crystals

X-ray diffraction patterns from crystals or semi-crystals is the result of radiation scattering from a series of scatterers positioned in space. For a simple version of this model, a scatterer is defined to be a point-like object that preserves the phase of the radiation field that it scatters and where all such objects scatter with the same intensity.

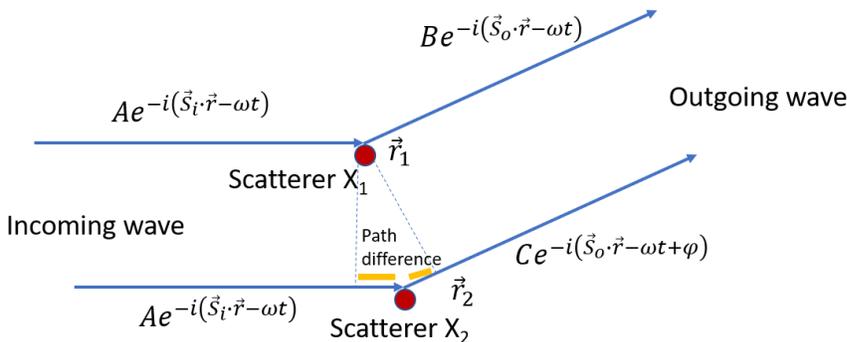


Figure C.1: Incoming wave scattered from two identical scattering bodies as sampled along a predetermined outgoing path.

As illustrated in Figure C.1, two scatterers,  $X_1$  at position  $\vec{r}_1$  and  $X_2$  at position  $\vec{r}_2$  scatter incident radiation from a wave vector  $\vec{S}_i$ . The wavelength of the radiation is related to the wave vector as  $|\vec{S}| = 2\pi/\lambda$ . If in the far-field limit one examines radiation scattered along a particular outgoing wave vector  $\vec{S}_o$ , one would discover that the wave component resulting from scattering by  $X_2$  has a phase shift from that scattered by  $X_1$  due to the length difference between the two paths. Along the direction denoted by the outgoing wave vector, scattering intensity cannot be observed if the two waves are out of phase, requiring a phase shift  $\varphi = 2\pi$ . Preservation of phase by the scatterers establishes the boundary condition that the phasing of the scattered wave match that of the incident wave.

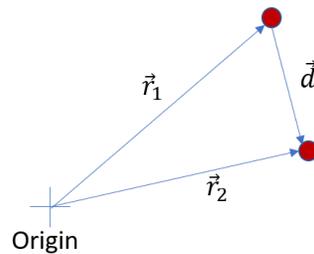
$$\text{At } X_1: \vec{S}_i \cdot \vec{r}_1 = \vec{S}_o \cdot \vec{r}_1 \quad (\text{C.1})$$

$$\text{At } X_2: \vec{S}_i \cdot \vec{r}_2 = \vec{S}_o \cdot \vec{r}_2 \quad (\text{C.2})$$

Here, the time dependence has been canceled across the equal sign in order to show only the phasing boundary condition at the two scattering centers. In this, no common phase is yet required during scattering, except that scattering can only be seen along the direction of the outgoing wave vector if the phases in the two equations are matched, which they may not be.

Position  $\vec{r}_2$  may be regarded as a translation  $\vec{d}$  from position  $\vec{r}_1$ .

$$\vec{r}_2 = \vec{r}_1 + \vec{d} \quad (\text{C.3})$$



This translation can be included in the phasing argument for scatterer  $X_2$ .

$$\vec{S}_i \cdot (\vec{r}_1 + \vec{d}) = \vec{S}_o \cdot (\vec{r}_1 + \vec{d}) \quad (\text{C.4})$$

The displaced phase can then be rearranged in order to separate the terms.

$$\vec{S}_i \cdot \vec{r}_1 + \vec{S}_i \cdot \vec{d} = \vec{S}_o \cdot \vec{r}_1 + \vec{S}_o \cdot \vec{d} \quad (\text{C.5})$$

$$\vec{S}_i \cdot \vec{r}_1 = \vec{S}_o \cdot \vec{r}_1 + (\vec{S}_o - \vec{S}_i) \cdot \vec{d} \quad (\text{C.6})$$

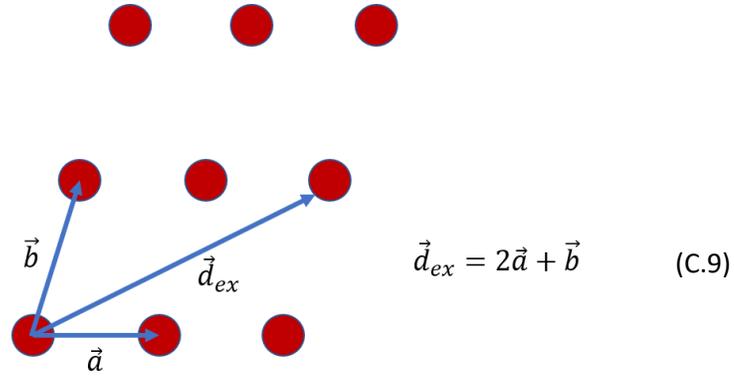
Clearly, scattering from  $X_2$  will be in phase with that from  $X_1$  if the term accumulated in the displacement is some integer multiple of  $2\pi$ .

$$(\vec{S}_o - \vec{S}_i) \cdot \vec{d} = 2\pi n \quad (\text{C.7})$$

The difference between the incoming and outgoing wave vectors can be renamed as the scattering vector  $\vec{q} = \vec{S}_o - \vec{S}_i$ .

$$\vec{q} \cdot \vec{d} = 2\pi n \quad (\text{C.8})$$

This can be expanded specifically into a 2D version of the Laue equations and the Bravais lattice which is particularly useful for analysis of semi-crystals as found with nanoDNA LC phases. Since the scattered intensity from one or two individual scatters is assumed to be very small, scattering intensity can be increased by placing additional scatterers into a repeated array where scattering would be coherent from any scattering centers in the array if they matched the displacement condition above. Placement in a lattice simply allows the scattering condition to be repeated between individual scatterers in the lattice array and scattering intensity seen along any particular outgoing scattering vector would be due to the number of scatterers in the array. The displacement vector  $\vec{d}$  can be written as a sum of the indexing vectors of the lattice, in this case a 2D lattice with characteristic vectors  $\vec{a}$  and  $\vec{b}$ .



The characteristic vectors  $\vec{a}$  and  $\vec{b}$  define the shape and dimensions of the basis unit cell used to construct a crystal where the displacement vector  $\vec{d}_{ex}$  is an example of a possible displacement within the crystal written in terms of multiples of these characteristic vectors. Any displacement to indistinguishable scatterers can be written in terms of integers  $n$  and  $m$ .

$$\vec{d}_{nm} = n\vec{a} + m\vec{b} \quad (\text{C.10})$$

These displacements can be substituted into the scattering vector phase relation, where  $h, n, k$  and  $m$  are all integers and any products of these are also integers.

$$\vec{q} \cdot (n\vec{a} + m\vec{b}) = n\vec{q} \cdot \vec{a} + m\vec{q} \cdot \vec{b} = n2\pi h + m2\pi k = 2\pi(hn + km) \quad (\text{C.11})$$

In this,  $n$  and  $m$  are indexes of lattice spacing and  $h$  and  $k$  are multiples of the  $2\pi$  phase that must be obeyed in order to produce visible scattering at any particular scattering vector. The individual relations with the lattice index vectors produce the Laue equations for the 2D system.

Laue Equations:

$$\vec{q} \cdot \vec{a} = 2\pi h \quad (\text{C.12})$$

$$\vec{q} \cdot \vec{b} = 2\pi k \quad (\text{C.13})$$

For the crystalline system, diffraction for a particular scattering vector  $\vec{q}$  is not observed unless the Laue equations are satisfied.

Further, the scattering vectors can also be assembled into another lattice with a similar vector construction as encountered in the real lattice. Here  $(h, k)$  are equivalent to Miller Index notation.

$$\vec{Q}_{hk} = h\vec{A} + k\vec{B} \quad (\text{C.14})$$

The vectors  $\vec{A}$  and  $\vec{B}$  are selected according to several rules intended to enforce the dot product on lattices where the basis vectors are not necessarily orthogonal. For a 2D lattice as introduced above, the necessary rules are as follows.

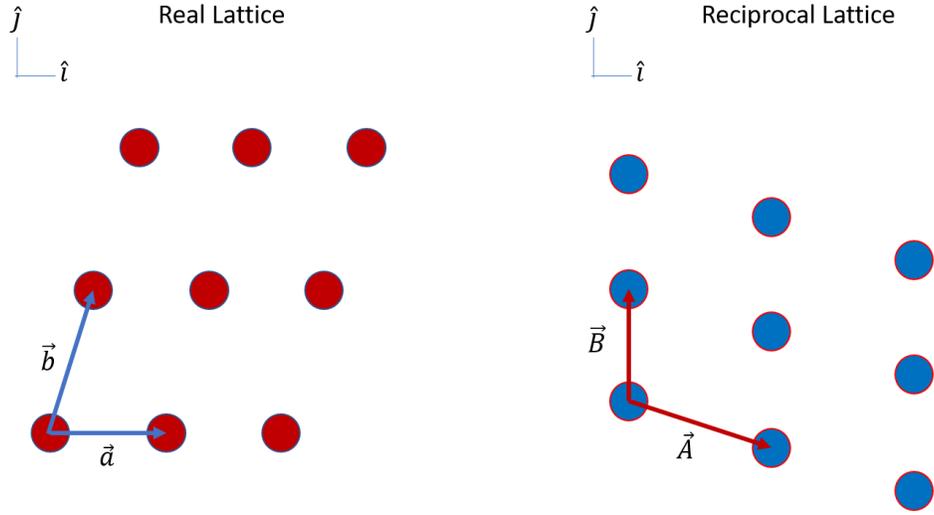
$$\vec{A} \cdot \vec{a} = 2\pi, \quad \vec{B} \cdot \vec{b} = 2\pi, \quad \vec{A} \cdot \vec{b} = 0, \quad \vec{B} \cdot \vec{a} = 0 \quad (\text{C.15})$$

From these rules, using cartesian basis vectors  $\vec{a} = a_1\hat{i} + a_2\hat{j}$  and  $\vec{b} = b_1\hat{i} + b_2\hat{j}$ , the  $\vec{A}$  and  $\vec{B}$  basis vectors are found by the following relations.

$$\vec{A} = \frac{2\pi}{|\vec{a} \times \vec{b}|} (b_2\hat{i} - b_1\hat{j}) \quad (\text{C.16})$$

$$\vec{B} = \frac{2\pi}{|\vec{a} \times \vec{b}|} (-a_2\hat{i} + a_1\hat{j}) \quad (\text{C.17})$$

Multiples of  $\vec{A}$  and  $\vec{B}$  basis vectors can be used to define any scattering vector that might be observed during a scattering experiment. For appropriately oriented real 2D lattices, these vectors span a 2D reciprocal lattice in the plane of the detector. This lattice is called “reciprocal” because it has units of inverse length and because the basis vectors may be thought of as scaled inverses of the real space lattice vectors. For the real 2D lattice depicted above, the reciprocal lattice would take on the following structure.



As the reciprocal lattice is the only observable during an X-ray diffraction experiment, vectors  $\vec{A}$  and  $\vec{B}$  are known while vectors  $\vec{a}$  and  $\vec{b}$  must be calculated. From the relations above, the real space lattice vectors can be obtained by similar equations as seen for  $\vec{A}$  and  $\vec{B}$ .

$$\vec{a} = \frac{2\pi}{|\vec{A} \times \vec{B}|} (B_2 \hat{i} - B_1 \hat{j}) \quad (\text{C.18})$$

$$\vec{b} = \frac{2\pi}{|\vec{A} \times \vec{B}|} (-A_2 \hat{i} + A_1 \hat{j}) \quad (\text{C.19})$$

Where scattering reflections in the 2D reciprocal lattice are indexed by  $h$  for  $\vec{A}$  and  $k$  for  $\vec{B}$ , the Laue equations are reproduced by substituting  $\vec{Q}_{hk} = h\vec{A} + k\vec{B}$  for the scattering vector into the phase relation.

$$(\text{C.20}) \quad \vec{Q}_{hk} \cdot \vec{d}_{nm} = (h\vec{A} + k\vec{B}) \cdot (n\vec{a} + m\vec{b}) = h\vec{A} \cdot \vec{a}n + h\vec{A} \cdot \vec{b}m + k\vec{B} \cdot \vec{a}n + k\vec{B} \cdot \vec{b}m$$

$$\vec{Q}_{hk} \cdot \vec{d}_{nm} = h2\pi n + 0 + 0 + k2\pi m \quad (\text{C.21})$$

$$\vec{Q}_{hk} \cdot \vec{d}_{nm} = 2\pi(hn + km) \quad (\text{C.22})$$

A useful expansion to this machinery is to develop a means of examining reflection omissions in the reciprocal lattice –the nature of these omissions will be discussed shortly. The points of the reciprocal lattice are positions in the far field

where the scattered wave produces intensity; intensity  $I$ , of course, is the square of the electric field  $E$  of the wave.

$$I = E^*E \quad (\text{C.23})$$

The vectoral quantities of wave polarization are omitted for simplicity and the wave is considered complex. Reconsider the situation of only scatterer  $X_1$  at position  $\vec{r}_1$  in Figure C.1 above. At the location of the scattering event, the incident wave is converted into the outgoing wave by a transformation.

$$E(\vec{r}_1) = B e^{i\vec{s}_o \cdot \vec{r}_1} = T A e^{i\vec{s}_i \cdot \vec{r}_1} \quad (\text{C.24})$$

The transformation mediating this single interaction may be solved at the location  $\vec{r}_1$ .

$$T = \frac{B}{A} e^{i\vec{s}_o \cdot \vec{r}_1 - i\vec{s}_i \cdot \vec{r}_1} = \frac{B}{A} e^{i(\vec{s}_o - \vec{s}_i) \cdot \vec{r}_1} = \frac{B}{A} e^{i\vec{q} \cdot \vec{r}_1} \quad (\text{C.25})$$

The ratio of  $B/A$  is the scattering strength  $f$  of scatterer  $X_1$ . The position where this scattering event occurs,  $\vec{r}_1$ , is a mysterious location in microscopic space and the initial phase of the incident wave is not known. On the other hand, it is possible to mark relative phases between neighboring scatterers in an array. The scattered wave, considered to be the E-field  $E$ , from all such scatterers builds up as a sum in the usual superposition where each scatterer is given a strength  $f$ . As before, scatterer  $X_2$  can be described as shifted from the location of  $X_1$  by the displacement of  $\vec{d}$  and scattered waves from  $X_1$  and  $X_2$  are identical except for this phase and possibly the strengths of each scatterer.

$$E = E_{X_1} + E_{X_2} \quad (\text{C.26})$$

$$E = T(X_1)A e^{i(\vec{s}_i \cdot \vec{r}_1 + \delta)} + T(X_2)A e^{i(\vec{s}_i \cdot \vec{r}_1 + \delta)} \quad (\text{C.27})$$

$$E = f_{X_1} e^{i\vec{q} \cdot \vec{r}_1} A e^{i(\vec{s}_i \cdot \vec{r}_1 + \delta)} + f_{X_2} e^{i\vec{q} \cdot (\vec{r}_1 + \vec{d})} A e^{i(\vec{s}_i \cdot \vec{r}_1 + \delta)} \quad (\text{C.28})$$

$$E = \left( f_{X_1} e^{i\vec{q}\cdot\vec{r}_1} + f_{X_2} e^{i\vec{q}\cdot(\vec{r}_1+\vec{d})} \right) A e^{i(\vec{S}_i\cdot\vec{r}_1+\delta)} \quad (\text{C.29})$$

Each term of the sum sets a field produced by each scatterer  $X_1$  and  $X_2$  relative to the position  $\vec{r}_1$  using the phase arguments previously developed and some unknown initial phase of the incident wave  $\delta$ . The E-field found at the detector would be obtained by propagating the phase over the distance between the sample and detector, but this additional phase is irrelevant to the current purposes since it is evenly accumulated for every scattering contribution and would not impact intensities specific to the reciprocal lattice: it could be contained within  $\delta$ . The sum in the coefficient here can be collected and expanded to include as many scatterers as one pleases.

The construct denoting the E-field is readily taken through the arguments already developed for the crystal array in this section, where each crystal unit cell could be expanded to contain more than just a single scatterer.  $E$  would be a sum of the waves scattered simultaneously by every unit cell in a crystal domain for a total intensity of  $V$ , where each cell contains  $v$  scatterers. Incorporating the previous work in this section,  $\vec{q} \rightarrow \vec{Q}_{hk}$ ,  $\vec{r}_1 \rightarrow \vec{d}_{nm}$  and  $\vec{d}_l$  is now some positioning vector in the basis of  $\vec{a}$  and  $\vec{b}$  (must be less than 1) describing the locations of the  $v$  objects inside each identical unit cell. With all contributions collected as a sum,

$$E = V \left( \sum_l^v f_l e^{i\vec{Q}_{hk}\cdot\vec{d}_{nm}+i\vec{Q}_{hk}\cdot\vec{d}_l} \right) A e^{i(\vec{S}_i\cdot\vec{d}_{nm}+\delta)} \quad (\text{C.30})$$

$$E = V e^{i\vec{Q}_{hk}\cdot\vec{d}_{nm}} \sum_l^v f_l e^{i\vec{Q}_{hk}\cdot\vec{d}_l} A e^{i(\vec{S}_i\cdot\vec{d}_{nm}+\delta)} \quad (\text{C.31})$$

The phase factor carried out of the sum is always automatically 1 at any position in the reciprocal lattice where the Laue conditions are met.

$$E = V \left( \sum_l^v f_l e^{i\vec{Q}_{hk}\cdot\vec{d}_l} \right) A e^{i(\vec{S}_i\cdot\vec{d}_{nm}+\delta)} \quad (\text{C.32})$$

In the far field, the intensity of a reciprocal lattice point is determined by convention from the E-field.

$$I = E^*E = V^2 \sum_g^v f_g e^{-i\vec{Q}_{hk}\cdot\vec{d}_g} \sum_l^v f_l e^{i\vec{Q}_{hk}\cdot\vec{d}_l} A^2 e^{i(\vec{S}_i\cdot\vec{d}_{nm}+\delta)-i(\vec{S}_i\cdot\vec{d}_{nm}+\delta)} \quad (\text{C.33})$$

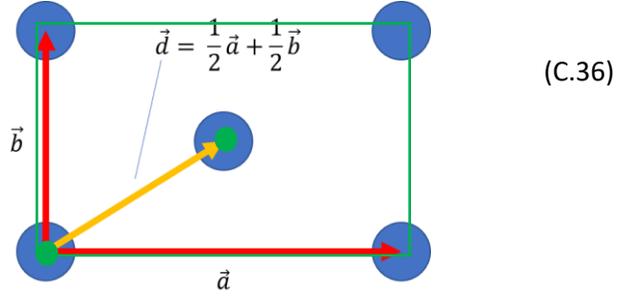
A simplification is to avoid the absolute magnitude which *unquestionably* contains some corrections from the naïve value given here as  $(AV)^2$  by turning this to a proportionality.

$$I \propto \sum_g^v f_g e^{-i\vec{Q}_{hk}\cdot\vec{d}_g} \sum_l^v f_l e^{i\vec{Q}_{hk}\cdot\vec{d}_l} \quad (\text{C.34})$$

The transformation used to acquire this intensity is called the structure factor.

$$T(h, k) = \sum_l^v f_l e^{i\vec{Q}_{hk}\cdot\vec{d}_l} \quad (\text{C.35})$$

Structure factor is useful for analyzing scattering reflection omissions present in the reciprocal lattice, a feature that is important to a few nanoDNA X-ray diffraction behaviors. When the unit cell of a lattice contains multiple identical scatterers, a few arrangements are of particular importance. Foremost for nanoDNA is the face centered rectangular lattice, which will produce omissions in the reciprocal lattice depending on how  $\vec{A}$  and  $\vec{B}$  are defined. As mentioned with the structure factor, the displacement vector  $\vec{d}$  is assigned fractional displacements of the real space unit cell as defined by  $\vec{a}$  and  $\vec{b}$ .



In this unit cell, marked with a green box, only the scattering centers marked with green spots are considered to be in the cell; the other three scatters are in three neighboring cells. The scatterer at the corner of the cell is at the origin of measurement for that cell and has a displacement of zero while the scatterer in the middle of the cell is located at  $\vec{d} = \frac{1}{2}\vec{a} + \frac{1}{2}\vec{b}$ . Both scatterers are of identical strength. With use of these displacements, the structure factor can be written for these two scatterers.

$$T(h, k) = e^{i\vec{Q}_{hk} \cdot 0} + e^{i\vec{Q}_{hk} \cdot (\frac{1}{2}\vec{a} + \frac{1}{2}\vec{b})} \quad (\text{C.37})$$

The rules of the basis vectors established above are applied to simplify the structure factor.

$$T(h, k) = 1 + e^{i(h\vec{A} + k\vec{B}) \cdot (\frac{1}{2}\vec{a} + \frac{1}{2}\vec{b})} = 1 + e^{i(\frac{h}{2}\vec{A} \cdot \vec{a} + \frac{k}{2}\vec{B} \cdot \vec{b})} = 1 + e^{i(\frac{h}{2}2\pi + \frac{k}{2}2\pi)} \quad (\text{C.38})$$

$$T(h, k) = 1 + e^{i\pi(h+k)} \quad (\text{C.39})$$

With the structure factor, the intensity can then be written.

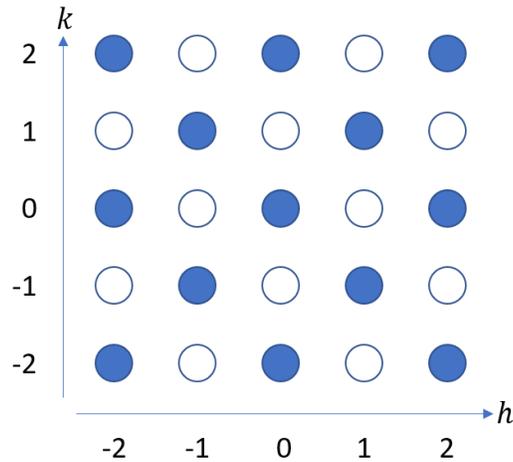
$$I \propto T^*T = (1 + e^{-i\pi(h+k)})(1 + e^{i\pi(h+k)}) \quad (\text{C.40})$$

$$I \propto 1 + e^{i\pi(h+k)} + e^{-i\pi(h+k)} + e^{i\pi(h+k) - i\pi(h+k)} = 1 + 1 + 2 \left( \frac{e^{i\pi(h+k)} + e^{-i\pi(h+k)}}{2} \right)$$

$$I \propto 2 + 2\cos[\pi(h+k)] \quad (\text{C.41})$$

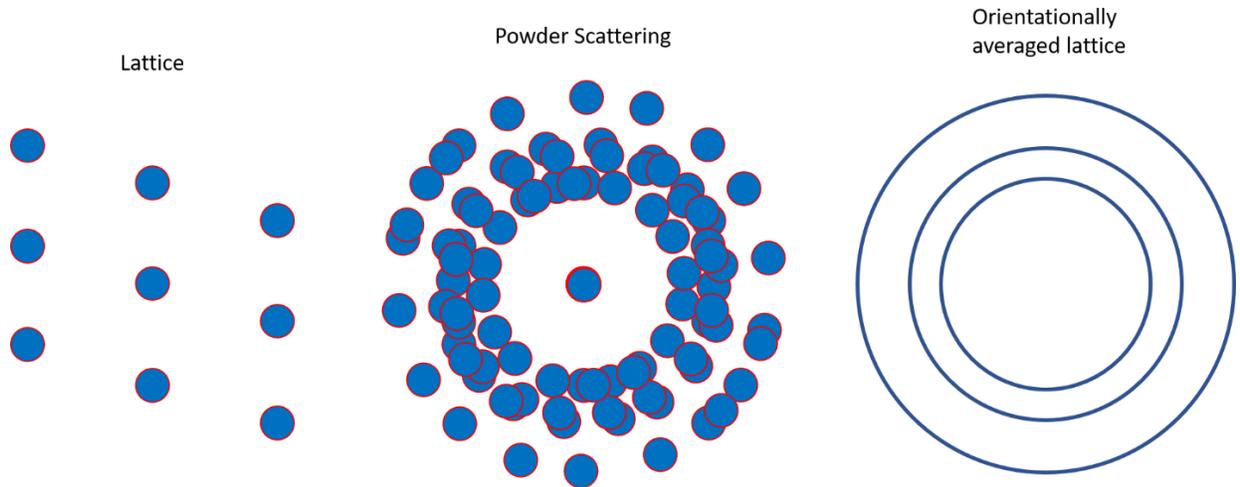
This expression of the intensity gives either 4 or 0 depending on the integer arguments of  $h$  and  $k$  in the cosine. If  $h+k$  is even, the intensity is 4. If  $h+k$  is odd,

there is no intensity at the detector. These are omissions in the reciprocal lattice where the second scatterer in the unit cell has provided a wave contribution of equal intensity to the first, but out phase by  $\pi$ . The scattering condition for this reciprocal lattice is therefore that reflections only occur if the sum of  $h$  and  $k$  is even.



The filled spots indicate locations where the scattering condition is met while empty spots represent lattice points where the scattering is out of phase and therefore canceled out. This is only one possible example of how this technique can be used and it also exemplifies the potential for degeneracy in assignment of lattice spacing vectors since the face-centered cell introduced above could also be a rhombic cell.

The machinery here are sufficient to reconstruct a real space 2D lattice given an observable reciprocal lattice. In practice, matters tend to be somewhat more complicated because the observation of such a lattice depends on the homogeneous orientation of lattice-bearing domains within the sample. If a sample contains many small-sized domains where the identical lattice is oriented differently in every domain, the reciprocal lattice as expressed on a 2D detector becomes circularly symmetrized.



When this “powder averaging” occurs, features of the lattice become degenerate and different reciprocal lattice sets may fit the same powder average. In these situations, lattices can be identified computationally by comparing the ratios of different ring radii in reciprocal space. Two dimensional nanoDNA lattices have been amenable to this technique and identification of relevant LC lattices will be tackled by case where pertinent.

For the situation where the system has been symmetrized in this way, it is useful to consider reducing the dimensionality of the scattering to 1D, where the real space displacement is always perpendicular to the scattering vector, to produce a simplified version of the Laue relations.

$$\vec{q} \cdot \vec{d} = 2\pi \quad \rightarrow \quad q = \frac{2\pi}{d} \quad (\text{C.42})$$

This adjustment produces a real-space perpendicular displacement between arrays of scatterers without necessarily placing those scatterers on the relevant lattice in 2D or even 3D space, which can provide some insight into the nature of the array when not all features of the lattice have been deduced. The above formula tells the distance between parallel arrays of electron density but gives no

information about how the elements of these arrays are arranged along the extent of a given array.

This relationship is close to what would be encountered in a 1D crystal lattice, which is also relevant to DNA in the first approximation given the regular 1D spacing of base-pairs along the axis of a single DNA duplex. Base-stacking in nanoDNA gives a 1D reflection at wide angle that is very similar to the 1D lattice formed by chromonic dyes Sunset Yellow and DSCG (data not included in this thesis). It must be noted however that the actual 2D scattering from a single DNA duplex is significantly more complicated as a result of the double helical arrangement of the scatterers<sup>59</sup>.

Most of the details in this section are presented under the approximation that scatterers are simple point-like objects that preserve the phase of radiation that they scatter. This may not agree with what occurs in reality. Real scatterers in a crystal lattice may be highly compound objects constructed of many atoms, while the atoms themselves are a variety of different types of extended object with non-uniform electron density. The actual scattering from such a more complicated scatterer will not preserve phase at every angle the same way simply because the internal combination and arrangement of atoms in the scatterer is subject to the physics detailed above, giving angles where scattering phases cancel out internally without including the external physics of the lattice. This can also give rise to lattice points where omissions might otherwise be expected, but intensity is observed none-the-less due to the structure factor or even the form factors of individual scatterers.

For the sake of completeness, it must be noted that the structure factor can, in principle, be used to access the electron density of an arrangement of scatterers.

As defined above, the structure factor is assembled from an array of dimensionless point scatterers.

$$T(h, k) = \sum_{j=0}^N f_j e^{i\vec{Q}(h,k) \cdot \vec{d}_j} \quad (\text{C.43})$$

This might be taken over into a continuum where the point-like scatterers are instead a distributed density  $\rho(\vec{r})$  within the space of the unit cell. This density is essentially “distributed density of scattering strength,” which can be constructed from the electron density. In this case, the formulation is again in 2D, keeping with the rest of this section.

$$T(h, k) = \int d^2r \rho(\vec{r}) e^{i\vec{Q}(h,k) \cdot \vec{r}} \quad (\text{C.44})$$

The formerly discrete variables of  $q$  space can also be promoted to a continuum and scattering strength density is found by a Fourier transform with the space of integration being the area of the unit cell and the available size of reciprocal space. In this particular case, the  $2\pi$  normalization factor needed for the usual transform is locked up in the definition of the dot product from the orthogonal basis vectors defining the reciprocal and real lattices.

$$\begin{aligned} \int dhdk T(h, k) e^{-i\vec{Q}(h,k) \cdot \vec{r}'} &= \int dhdk \int d^2r \rho(\vec{r}) e^{i\vec{Q}(h,k) \cdot (\vec{r} - \vec{r}')} \\ \int dhdk T(h, k) e^{-i\vec{Q}(h,k) \cdot \vec{r}'} &= \int d^2r \rho(\vec{r}) \int dhdk e^{i\vec{Q}(h,k) \cdot (\vec{r} - \vec{r}')} \\ \int dhdk T(h, k) e^{-i\vec{Q}(h,k) \cdot \vec{r}'} &= \int d^2r \rho(\vec{r}) \delta^2(\vec{r} - \vec{r}') \\ \rho(\vec{r}') &= \int dhdk T(h, k) e^{-i\vec{Q}(h,k) \cdot \vec{r}'} \quad (\text{C.45}) \end{aligned}$$

This offers a method by which the object distribution can essentially be determined from the structure factor when noting how the structure factor is

obtained from the observed scattering intensities  $I(h, k)$  at the detector (which are essentially discretized by a series of  $\delta$ -functions in  $(h, k)$  space).

$$|T(h, k)| = \sqrt{I(h, k)} \quad (\text{C.46})$$

The only stumbling block here is that  $T(h, k)$  is complex, giving rise to the so-called phase problem.

$$T(h, k) = \sqrt{I(h, k)}e^{i\delta} \quad (\text{C.47})$$

The scattering strength density in the unit cell is calculated from the following form.

$$\rho(\vec{r}') = \int dhdk \sqrt{I(h, k)}e^{-i(\vec{Q}(h,k)\cdot\vec{r}'-\delta)} \quad (\text{C.48})$$

With a proper normalization of  $I$  obtained experimentally and some solution for the phase  $\delta$ , one can calculate the electron density per position in the unit cell of a crystal lattice by these means. Most of this last is not used in this thesis given the highly degenerate nature of the circularly averaged data that will be presented and is added purely for the edification of the author and his abiding curiosity for the methods used to determine molecular structures.

## C.2 Analysis of X-ray Scattering Peak Shapes

Additional useful information about the diffraction is contained in the scattering peak shape.

In an ideal system, with crystals of infinite size, all X-ray scattering off of the sample would be perfectly identical across the entire sample, directing scattering for a particular crystalline feature equally well for every represented scatterer onto the same scattering vector. Under these conditions, lattice points in reciprocal space

would appear to be  $\delta$ -functions, with all scattered intensity directed exactly to a point. Intensity falls off quickly going away from these points where the scattering phases from different locations in the crystal rapidly becomes disordered.

In reality, experimental systems are not identical throughout a given sample; scatterers may be subtly different from one another, or lattice spacing may be imperfectly reproduced across a domain. Even the size of a crystal domain influences the sharpness of the peak where small domains tend to produce broader peaks than large domains, as expressed by the Scherrer Equation. From these effects and others, scattering along a particular scattering vector is softened so that intensity is spread out around some average vector. As Scherrer Equation related domain size contributions become limited when crystal domains become much larger than 100-200 nanometers<sup>58</sup>, peak broadening can reflect direct variation of crystal ordering. If the instrument is kept the same from one sample to the next so that peak broadening due to instrumental effects is regularized throughout an experiment, peak broadening observed during a particular experiment or between similar samples can be ascribed to dynamical variation of the samples, potentially giving some clues about relative crystal ordering between observations.

X-ray diffraction is regularly used to report on the uniformity of ordering in an LC phase. For example, in a conventional calamitic nematic liquid crystal thought to occupy a Maier-Suape distribution, X-ray scattering from pair-wise side-side correlations in the mesophase results in a Gaussian scattering vector peak shape where the half-width of the peak directly reflects the order parameter of the phase<sup>60</sup>. While this specific model can be useful elsewhere, reflections from nanoDNA mesophases, both nematic and columnar, typically do not fit well to Gaussian peaks, suggesting that other models are needed to interpret the scattering peak shape. Regardless of the specific peak shape used, the general trend overall is

that non-instrumental peak broadening is associated with decreased order in such a mesophase where the domain size is generally considered sufficiently large as to not noticeably broaden the peak shape.

The method used to judge nanoDNA mesophase order by peak shape in this thesis was originated by R. Hosemann in his treatment of pseudocrystals<sup>61,62</sup>. In a real physical crystal where scatterer positioning in a unit cell varies slowly between neighboring cells, accumulated small positioning errors gradually propagate until the sum of such errors effectively negates positional correlation between widely separated cells in a crystal domain. The size scale over which these positioning errors accumulate to exceed the size of the unit cell is often called the structural coherence length (derived A. Guinier Chapter 9, pgs 309-330<sup>58</sup>). Peak shape associated with this form of broadening is closely Lorentzian where the full width at half maximum (FWHM) is approximately the direct reciprocal of the structural coherence length. Since structural coherence would be a characteristic feature of the lattice, this form of broadening tends to soften peaks more strongly going to large  $q$ -values, causing WAXS peaks to broaden much more than SAXS peaks, approximately going as diffraction peak order squared,  $n^2$ .

In this thesis, data sets are analyzed upon data reduction from 2D detector data to a 1D circular average and peak shape is analyzed by fitting with a Lorentzian as a function of reciprocal space position.

$$I(q) = \left(\frac{2A}{\pi}\right) \frac{\Delta q}{4(q - q_0)^2 + (\Delta q)^2} \quad (\text{C.49})$$

This gives intensity of the scattering  $I(q)$  where  $A$  is a free parameter to accommodate the peak height,  $q$  is the domain of a 1D reciprocal space,  $q_0$  is the center of the peak and  $\Delta q$  is the FWHM of the peak. Structural coherence length  $d_c$  is taken from the FWHM.

$$d_c = \frac{1}{\Delta q} \quad (\text{C.50})$$

This relation is justified to work for lowest order diffraction peak harmonics.

A full derivation of this relation exceeds the scope of need here and may be obtained from a competent text on X-ray powder diffraction<sup>58</sup>.

### C.3 Experimental Methods for nanoDNA X-ray Diffraction

All X-ray diffraction studies reported in this thesis were carried out on beamline 7.3.3 at the Advanced Light Source at Lawrence Berkeley Labs in Berkeley California<sup>63</sup>. Beamline 7.3.3 uses 10 keV X-rays, with a corresponding wavelength of 1.24 Å and a flux of 10<sup>12</sup> photons/sec. The beam dimensions were approximately 300 x 700 μm. All data reported in this thesis were shot in Wide Angle Scattering geometry (WAXS) with a sample-to-detector path length of about 30 cm and detected by a two-dimensional Pilatus 2M detector with a CCD pixel size of 0.172 mm. Thermal control was accomplished during diffraction experiments by use of a modified Instec microscope hot-stage with windows made of either Mica or Kapton tape and with cooling supplemented by a circulating chiller filled with an antifreeze solution. 2D to 1D Data reduction was carried out in Igor Pro version 6 with the Nika SAS 2D package<sup>64</sup>. Every X-ray “shot” was typically actually a pair of shots where the detector is displaced fractionally in the vertical direction between shots in order to generate a tiled data set that hides the panel seams of the CCD boards in the detector in the vertical direction (but not in the horizontal direction).

NanoDNA samples were prepared in cylindrical borosilicate capillaries by the methods described in Section B.1.4 to concentrations calculated as described in Sections B.2.1 and B.2.2. Since X-ray diffraction experiments typically involved

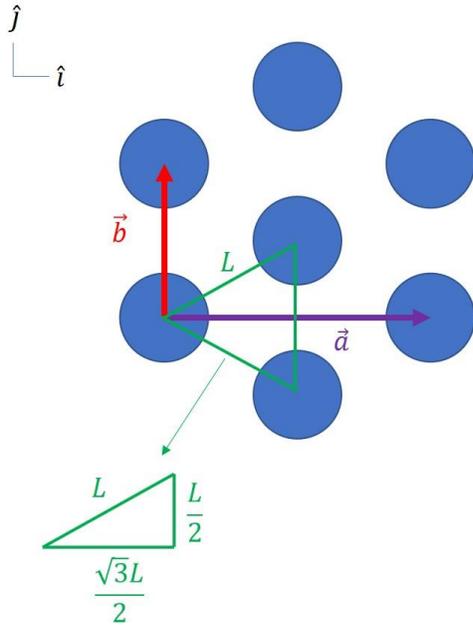
many X-ray shots at a given sample, exposure times were restricted to periods as short as practical to acquire the desired data, typically less than 2 seconds, in order to limit the rate of radiation exposure damage.

As implied in Section C.1, nanoDNA LC phase scattering data is frequently from samples that are unaligned polydomains, giving rise to powder averaged scattering that appears as a series of concentric 2D rings. Data reduction carried out as described above takes a 2D ring pattern, centers the image at the unscattered beam using a calibration standard of Silver Behenate and performs a circular average over all angles at each radius in order to report  $\langle q(r) \rangle$  vs. scattering intensity. The resulting data is then analyzed case-by-case in an effort to extract real-space lattice information using the theory presented in Sections C.1 and C.2.

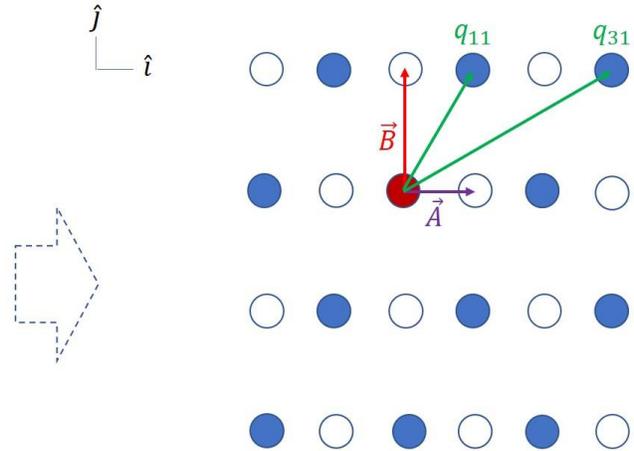
#### C.4.1 Analysis of Two Dimensional Hexagonal Lattices

To exemplify a model that will be very specifically used concerning the higher order phases of rDD and in the examination of Blunt End 4mer  $5'GTAC3'$ , recall the theory of the face centered rectangular lattice put forward in Section C.2.1. This particular lattice model creates a cell where the 2D crystal dimensions are perpendicular to one another, easing certain calculations, which can be associated with a hexagonal 2D lattice by the following construction. A similar calculation can be made by simply noting that the reciprocal lattice of a hexagonal lattice is also a hexagonal lattice, allowing the ratio of scattering peak radii to be obtained directly—the rectangular face centered cell offered here is simply introduced out of convenience to familiarize the reader with its potential in order to expand its use elsewhere in this thesis. In this depiction of the hexagonal reciprocal lattice, filled circles are observed reflections while empty circles are omissions.

Real Hexagonal Lattice:



Reciprocal Lattice:



The real hexagonal lattice built of equilateral triangle facets with length  $L$  can be instead reduced to a rectangle with perpendicular sides  $\vec{a}$  and  $\vec{b}$ . The lattice unit cell is then a face-centered rectangle. A 30-60-90 right-triangle gives the lengths of the rectangular basis vectors in terms of  $L$ .

$$\vec{a} = \sqrt{3}L\hat{i} \quad \vec{b} = L\hat{j} \quad (\text{C.51})$$

Because these vectors are perpendicular, the conversion from real space to reciprocal space is especially easy using the methods of Section C.1.

$$\vec{A} = \frac{2\pi}{\sqrt{3}L}\hat{i} \quad \vec{B} = \frac{2\pi}{L}\hat{j} \quad (\text{C.52})$$

In the reciprocal lattice, which appears directly in the diffraction pattern, first order and second order reflections are six-fold degenerate in that all such reflections lie the same radius from beam center. When the reciprocal lattice is powder-averaged, this leads to only two observable peaks. As in the figure above, the first order peak will be called  $q_{11}$  even though this peak is equal to  $q_{20}$ ,

$q_{1-1}$ ,  $q_{-1-1}$ ,  $q_{-20}$  and  $q_{-11}$  in magnitude. Similarly, the second order peak will be called  $q_{31}$  among its six variants. Lengths for  $q_{11}$  and  $q_{31}$  can be obtained in terms of the basis vectors by constructing them as if they were 2D vectors.

$$q_{11} = 2 \left( \frac{2\pi}{\sqrt{3}L} \right) = \frac{4\pi}{\sqrt{3}L} \quad (\text{C.53})$$

$$q_{31} = \sqrt{\left( 3 \frac{2\pi}{\sqrt{3}L} \right)^2 + \left( \frac{2\pi}{L} \right)^2} = \frac{4\pi}{L} \quad (\text{C.54})$$

This provides a method for identifying potential  $q_{11}$  and  $q_{31}$  reflections from their ratio.

$$\frac{q_{11}}{q_{31}} = \frac{4\pi L}{4\pi\sqrt{3}L} = \frac{1}{\sqrt{3}} \quad (\text{C.55})$$

The standard in this thesis for assignment of hexagonal lattices from powder averaged data is to look for this ratio between the first and second order peaks of the 1D diffraction. After these peaks have been assigned, the real space lattice spacing parameter  $L$  can be identified from  $q_{11}$ .

$$L = \frac{4\pi}{\sqrt{3}q_{11}} \quad (\text{C.56})$$

Also frequently used is the area of the face-centered rectangular unit cell for a hexagonal lattice ( $A_{hex}$ ) obtained by multiplying the lengths of  $\vec{a}$  and  $\vec{b}$ .

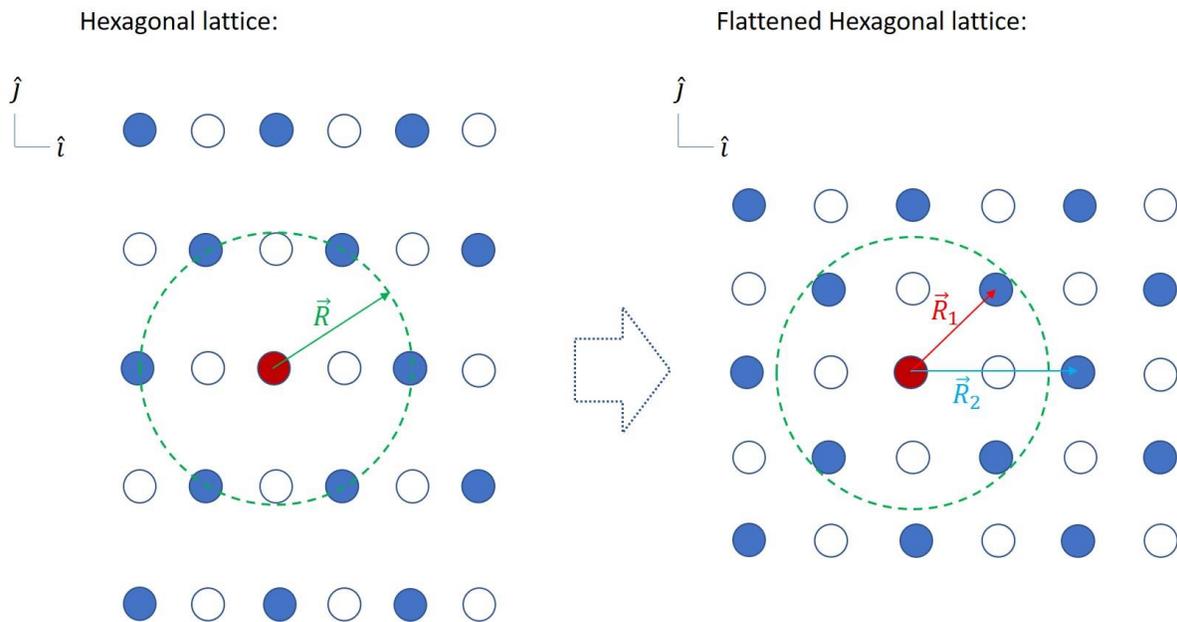
$$A_{hex} = \sqrt{3}L^2 = \frac{16\pi^2}{\sqrt{3}(q_{11})^2} \quad (\text{C.57})$$

#### C.4.2 Analysis of Flattened Hexagonal Lattices

As hexagonal lattices are highly important to DNA mesophases, deformations of this sort of lattice have also proven important. The first kind of distortion used in analysis was the idea of the flattened hexagonal lattice (FHL). An FHL is obtained

from a hexagonal lattice in a face-centered rectangular definition by altering the lengths of the basis vectors without changing the angle between them. A common required calculation was to compare an apparent hexagonal lattice to an FHL in order to predict the hexagonal  $q_{11}$  reflection, called here  $q_{hex}$ , on the postulate that the FHL and hexagonal lattices both have the same area in their unit cells.

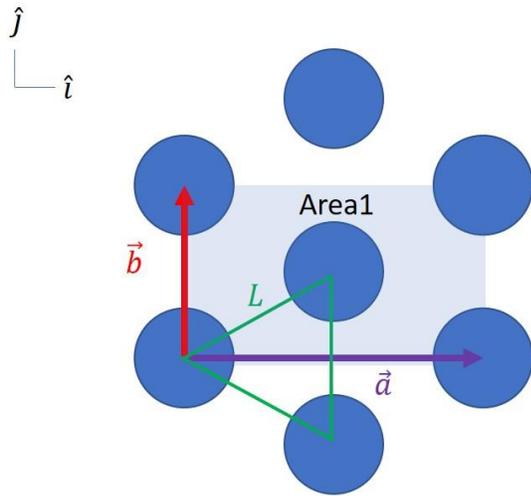
For a 2D hexagonal lattice as seen by powder average in reciprocal space, all six points of the hexagon would occur at the same radius and would overlap to hide their multiplicity. For a squished hexagon, the resulting transformation in reciprocal space would be to move all of the reflections away from their common radius, some closer to the beamstop and some farther, in order to deal with the conservation of area. For this particular transformation, the hexagonal  $q_{11}$  reflection should split into two reflections.



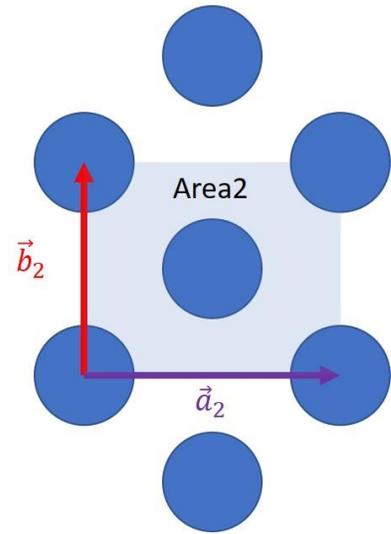
As usual, for these variations of the face-centered rectangular cell, in reciprocal space, the filled circles imply observed reflections while the empty circles are expected omissions.

We could then construct a model to determine the relationship between radius  $\vec{R}$  and radii  $\vec{R}_1$  and  $\vec{R}_2$  such that the areas of both *real space* lattices are equal. Again, using the methods of Section C.1, it is helpful to return to the rectangular face-centered lattice of Section C.4.1 since the area of a rectangle is easily calculated.

Real Hexagonal Lattice:



Flattened Hexagonal Lattice:



For this form of the real space hexagonal lattice, relevant to the reciprocal version depicted above, Area1 ( $A_{hex}$ ) and Area2 ( $A_{FHL}$ ) are calculated as:

$$A_{hex} = \sqrt{3}L^2$$

$$A_{FHL} = |\vec{a}_2||\vec{b}_2| \quad (C.58)$$

The areas are then set equal  $A_{hex} = A_{FHL}$ . Such area conservation would be expected for superstructures that are incompressible in the dimension perpendicular to the 2D lattice. This results in a relation between the hexagonal lattice spacing parameter  $L$  and the basis vectors of the flattened lattice.

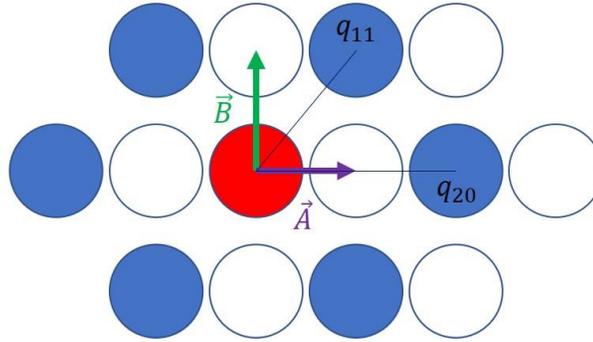
$$\sqrt{3}L^2 = |\vec{a}_2||\vec{b}_2| \quad (C.59)$$

From Section C.4.1, the first order hexagonal reflections are related to the lattice spacing parameter in a known way. Taking the absolute value of the radius of the reflection  $|\vec{R}| = q_{hex}$ .

$$L = \frac{4\pi}{\sqrt{3} q_{hex}} \quad (C.60)$$

The first order reflections of the flattened hexagonal lattice may be taken from the geometry of the appropriate face-centered rectangular lattice.

Flattened Hexagonal Reciprocal Lattice:



From this, both reflections can be written in terms of the reciprocal basis vectors.

$$q_{20} = 2|\vec{A}| \quad (C.61)$$

$$q_{11} = \sqrt{|\vec{A}|^2 + |\vec{B}|^2} \quad (C.62)$$

And the real lattice basis vectors are easily obtained from the reciprocal lattice basis given that they are perpendicular.

$$|\vec{A}| = \frac{2\pi}{|\vec{a}|} \quad (C.63)$$

$$|\vec{B}| = \frac{2\pi}{|\vec{b}|} \quad (C.64)$$

Which produces the observed  $q$ -values in terms of the real lattice basis vectors.

$$q_{20} = \frac{4\pi}{|\vec{a}|} \quad (\text{C.65})$$

$$q_{11} = \sqrt{\frac{4\pi^2}{|\vec{a}|^2} + \frac{4\pi^2}{|\vec{b}|^2}} \quad (\text{C.66})$$

Basis vector  $\vec{b}$  can be decoupled from  $\vec{a}$  in the  $q_{11}$  equation by substitution of the other equation for  $q_{20}$ .

$$q_{11} = \sqrt{\frac{(q_{20})^2}{4} + \frac{4\pi^2}{|\vec{b}|^2}} \quad (\text{C.67})$$

And this equation can be rearranged to place  $|\vec{b}|$  purely in terms of the observables.

$$|\vec{b}| = \frac{2\pi}{\sqrt{(q_{11})^2 - \left(\frac{q_{20}}{2}\right)^2}} \quad (\text{C.68})$$

With  $q_{hex}$ ,  $q_{11}$  and  $q_{20}$  in hand, these can be substituted into  $\sqrt{3}L^2 = |\vec{a}_2||\vec{b}_2|$ .

$$\sqrt{3} \left( \frac{4\pi}{\sqrt{3} q_{hex}} \right)^2 = \frac{4\pi}{q_{20}} \frac{2\pi}{\sqrt{(q_{11})^2 - \left(\frac{q_{20}}{2}\right)^2}} \quad (\text{C.69})$$

Which simplifies to an easier form.

$$q_{hex} = \left( \frac{2q_{20}}{\sqrt{3}} \sqrt{(q_{11})^2 - \left(\frac{q_{20}}{2}\right)^2} \right)^{\frac{1}{2}} \quad (\text{C.70})$$

This enables prediction of a hexagonal first order  $q_{11}$  reflection from the properly assigned first order reflections of a hypothetical FHL. Higher order reflections can be predicted by construction of the reciprocal lattice basis vectors from the two fundamental reflections.

$$\vec{A} = \frac{q_{20}}{2} \hat{i} \quad (\text{C.71})$$

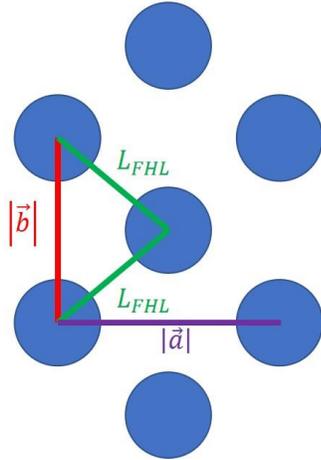
$$\vec{B} = \sqrt{(q_{11})^2 - \left(\frac{q_{20}}{2}\right)^2} \hat{j} \quad (\text{C.72})$$

Including Miller index notation, a formula useful for predicting higher order reflections can be written from this.

$$q_{hk} = \sqrt{(h^2 - k^2) \frac{(q_{20})^2}{4} + (kq_{11})^2} \quad (\text{C.73})$$

The dimensional parameters of the real space FHL are summarized together below. While the scale here is arbitrary, the unit cell of the flattened face-centered lattice is as shown.

Flattened Hexagonal Lattice:



$$|\vec{a}| = \frac{4\pi}{q_{20}} \quad (\text{C.74})$$

$$L_{FHL} = 2\pi \sqrt{\frac{1}{(q_{20})^2} + \frac{1}{4(q_{11})^2 - (q_{20})^2}} \quad (\text{C.75})$$

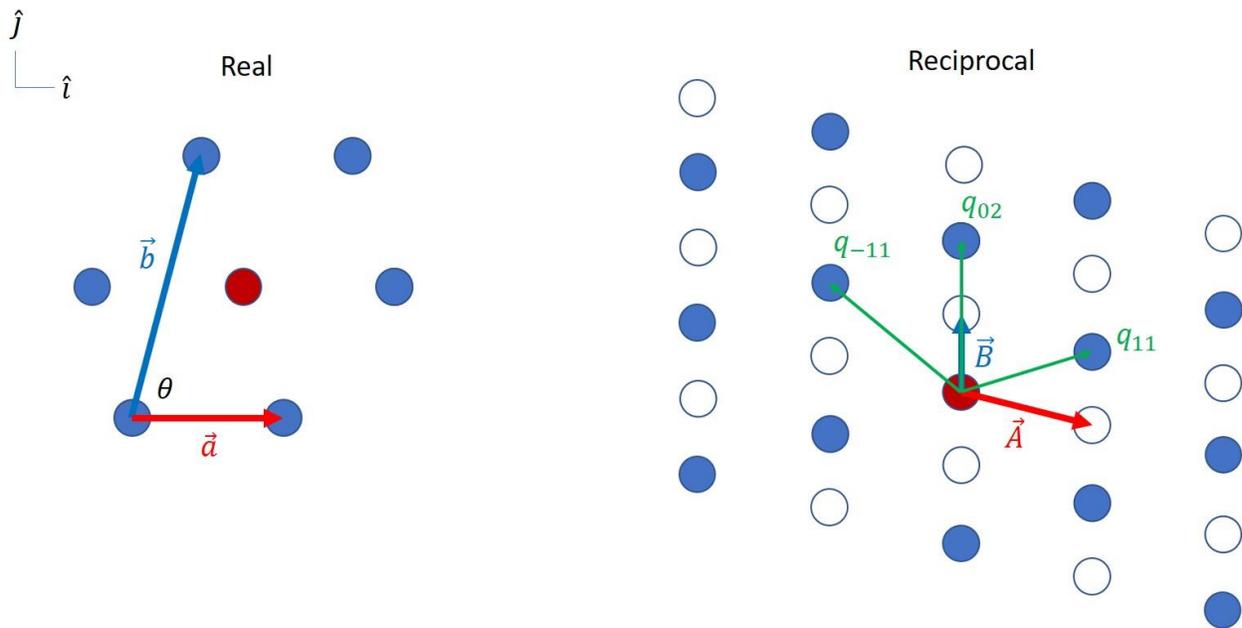
$$|\vec{b}| = \frac{2\pi}{\sqrt{(q_{11})^2 - \left(\frac{q_{20}}{2}\right)^2}} \quad (\text{C.76})$$

### C.4.3 Sheared Hexagonal Lattice

Sheared hexagonal lattice (SHL) is the other type of hexagonal lattice deformation that has been examined during data analysis encountered in this thesis. The shearing deformation differs from the flattening deformation in that it

would be expected to split the hexagonal  $q_{11}$  reflection (called here  $q_{hex}$ ) into three peaks instead of just two as seen in Section C.4.2. As in the previous section, a practical analysis is to compare the area of the SHL to that of a normal hexagonal lattice in order to predict  $q_{hex}$  based upon the three fundamental reflections generated by the SHL.

Sheared Hexagonal lattice:



For conserved concentration with no compression along the z-axis, area of a hexagonal lattice must equal the area of the equivalent SHL. Parameters of the SHL are as shown above where the face centered rectangular cell has become a parallelogram. In the reciprocal lattice, the empty circles are systematic omissions while the filled circles are observed reflections.

Acquisition of  $q_{hex}$  and  $A_{hex}$  are detailed in Section C.4.1, but in summary,

$$A_{hex} = \frac{16\pi^2}{\sqrt{3}(q_{hex})^2}$$

On the other side of the area equivalence, the SHL unit cell is simply a parallelogram where the area can be found using a cross product of the lattice parameters  $\vec{a}$  and  $\vec{b}$ .

$$A_{SHL} = |\vec{a} \times \vec{b}| = |\vec{a}||\vec{b}|\sin\theta \quad (C.77)$$

From this,  $A_{hex} = A_{SHL}$ , which gives,

$$\frac{16\pi^2}{\sqrt{3}(q_{hex})^2} = |\vec{a}||\vec{b}|\sin\theta \quad (C.78)$$

The lattice parameters of the sheared hexagonal lattice can be constructed:

$$\vec{a} = a\hat{i} \quad (C.79)$$

$$\vec{b} = b\cos\theta\hat{i} + b\sin\theta\hat{j} \quad (C.80)$$

Using the methods of Section C.1, these lattice parameter vectors can be used to construct the reciprocal lattice basis vectors.

$$\vec{A} = \frac{2\pi}{a}\hat{i} - \frac{2\pi}{a\tan\theta}\hat{j} \quad (C.81)$$

$$\vec{B} = \frac{2\pi}{b\sin\theta}\hat{j} \quad (C.82)$$

The reflections  $q_{11}$ ,  $q_{-11}$  and  $q_{02}$  are then constructed as vectors from the reciprocal lattice basis.

$$\vec{q}_{11} = \vec{A} + \vec{B} \quad \vec{q}_{02} = 2\vec{B} \quad \vec{q}_{-11} = -\vec{A} + \vec{B} \quad (C.83)$$

Vector length is obtained with a dot product, which can be shortened simply to the square:

$$(q_{02})^2 = 4\vec{B} \cdot \vec{B} = 4\left(\frac{2\pi}{b\sin\theta}\right)\left(\frac{2\pi}{b\sin\theta}\right) = \frac{16\pi^2}{b^2\sin^2\theta} \quad (C.84)$$

$$(q_{11})^2 = \vec{A} \cdot \vec{A} + 2\vec{A} \cdot \vec{B} + \vec{B} \cdot \vec{B} = \frac{4\pi^2}{a^2} + \frac{4\pi^2}{a^2\tan^2\theta} - 2\left(\frac{4\pi^2}{ab\sin\theta\tan\theta}\right) + \frac{4\pi^2}{b^2\sin^2\theta}$$

$$(q_{-11})^2 = \vec{A} \cdot \vec{A} - 2\vec{A} \cdot \vec{B} + \vec{B} \cdot \vec{B} = \frac{4\pi^2}{a^2} + \frac{4\pi^2}{a^2 \tan^2 \theta} + 2 \left( \frac{4\pi^2}{ab \sin \theta \tan \theta} \right) + \frac{4\pi^2}{b^2 \sin^2 \theta}$$

These equations are enough to construct the pieces needed to find the area. Most immediately,  $\sin \theta$  comes from  $q_{02}$  and parameter  $b$ .

$$\sin^2 \theta = \frac{16\pi^2}{b^2(q_{02})^2} \quad \sin \theta = \frac{4\pi}{bq_{02}} \quad (\text{C.85})$$

This equation helps remove the tangents with some simple trig:

$$\frac{1}{\tan \theta} = \sqrt{\frac{1}{\sin^2 \theta} - 1} = \sqrt{\frac{b^2(q_{02})^2}{16\pi^2} - 1} \quad (\text{C.86})$$

Relations involving  $q_{11}$  and  $q_{-11}$  are then constructed to remove  $\theta$  by back-substitution.

$$(q_{11})^2 = \left( \frac{1}{a^2} + \frac{1}{b^2} \right) \frac{b^2(q_{02})^2}{4} - \frac{2\pi q_{02}}{a} \sqrt{\frac{b^2(q_{02})^2}{16\pi^2} - 1} \quad (\text{C.87})$$

$$(q_{-11})^2 = \left( \frac{1}{a^2} + \frac{1}{b^2} \right) \frac{b^2(q_{02})^2}{4} + \frac{2\pi q_{02}}{a} \sqrt{\frac{b^2(q_{02})^2}{16\pi^2} - 1} \quad (\text{C.88})$$

Neither of these equations, which are simplified from above, allows the parameters  $a$  and  $b$  to be directly separated from each other. Separation can be achieved by using these equations together to create combinations. The first combination is to simply add the squares.

$$(q_{11})^2 + (q_{-11})^2 = \left( \frac{1}{a^2} + \frac{1}{b^2} \right) \frac{b^2(q_{02})^2}{2} \quad (\text{C.89})$$

Which can then be simplified to solve  $a$  in terms of  $b$ .

$$a = \frac{bq_{02}}{\sqrt{2((q_{11})^2 + (q_{-11})^2) - (q_{02})^2}} \quad (\text{C.90})$$

A second combination is obtained by subtracting the squares.

$$(q_{-11})^2 - (q_{11})^2 = \frac{4\pi q_{02}}{a} \sqrt{\frac{b^2(q_{02})^2}{16\pi^2} - 1} \quad (\text{C.91})$$

With back-substitution of the solution for  $a$  in terms of  $b$ , this equation can be applied to solve for  $b$  exclusively in terms of observables after some significant algebraic manipulation.

$$b = 4\pi \left( \frac{2((q_{11})^2 + (q_{-11})^2) - (q_{02})^2}{2(q_{02})^2((q_{11})^2 + (q_{-11})^2) - (q_{02})^4 - ((q_{-11})^2 - (q_{11})^2)^2} \right)^{\frac{1}{2}} \quad (\text{C.92})$$

Because these expressions became quite cumbersome and need to be solved with a circular permutation of  $q_{11}$ ,  $q_{-11}$  and  $q_{02}$  (1D WAXS data can only be narrowed to three possible assignment combinations of these peaks), calculation of  $b$  was automated with the software package Mathematica.  $a$  and  $\sin\theta$  were then calculated from  $b$  using their interdependence. These numbers allow determination of  $q_{hex}$  of the hexagonal lattice possessing the same area as the SHL encountered here.

$$q_{hex} = \frac{4\pi}{\sqrt{\sqrt{3}absin\theta}} \quad (\text{C.93})$$

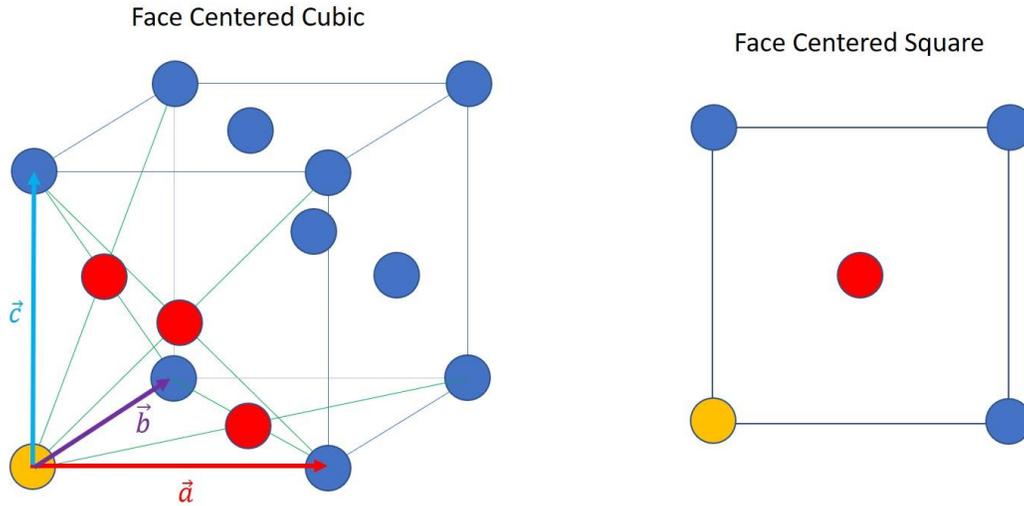
In addition, as suggested above, area can be directly acquired for comparison.

$$A_{SHL} = absin\theta \quad (\text{C.94})$$

#### C.4.4 Face Centered Cubic Lattice

One particular three-dimensional crystal lattice is useful to work regarding nanoDNA 4mer  ${}^5\text{GCCG}^3$ , the Face Centered Cubic (FCC) crystal lattice. What is presented here is a quick overview of this well-known crystal system such that it

can used elsewhere in this thesis. FCC is a 3D version of the face centered square lattice which is a specialized form of the face centered rectangular unit cell used to analyze the 2D hexagonal lattice in previous sections but where the edges of the unit cell are all of equal length.



Shown here, the FCC has a unit cell containing four identical bodies, one marking the corner of the cubic crystal cell (in orange) and three others in the adjoining faces (in red). This unit would then have translation symmetry to any unit elsewhere in the crystal as located by multiples of the unit cell basis vectors  $\vec{a}$ ,  $\vec{b}$  and  $\vec{c}$ , which all have equal length and are located along the three perpendicular edges of the cube.

Systematic omissions in the reciprocal lattice are most easily calculated using a version of the structure factor containing these four scatterers. Within the unit cell, the four scatterers are located by fractional displacements of the crystal basis vectors.

$$\vec{d}_1 = 0, \quad \vec{d}_2 = \frac{\vec{a}}{2} + \frac{\vec{b}}{2}, \quad \vec{d}_3 = \frac{\vec{b}}{2} + \frac{\vec{c}}{2}, \quad \vec{d}_4 = \frac{\vec{c}}{2} + \frac{\vec{a}}{2} \quad (\text{C.95})$$

Where the scattering intensities of each scatterer are set to unity, this gives the following structure factor:

$$T = 1 + e^{i(h\vec{A}+k\vec{B}+l\vec{C})\cdot\left(\frac{\vec{a}}{2}+\frac{\vec{b}}{2}\right)} + e^{i(h\vec{A}+k\vec{B}+l\vec{C})\cdot\left(\frac{\vec{b}}{2}+\frac{\vec{c}}{2}\right)} + e^{i(h\vec{A}+k\vec{B}+l\vec{C})\cdot\left(\frac{\vec{c}}{2}+\frac{\vec{a}}{2}\right)} \quad (\text{C.96})$$

Or,

$$T(h, k, l) = 1 + e^{i\pi(h+k)} + e^{i\pi(h+l)} + e^{i\pi(k+l)} \quad (\text{C.97})$$

This structure factor is then converted to the form which would appear in expressions of intensity.

$$T^*T = (1 + e^{-i\pi(h+k)} + e^{-i\pi(h+l)} + e^{-i\pi(k+l)})(1 + e^{i\pi(h+k)} + e^{i\pi(h+l)} + e^{i\pi(k+l)})$$

With some work, the following is obtained.

$$\begin{aligned} I(h, k, l) &\propto T^*T \\ &= 4 + 2\cos\pi(h+k) + 2\cos\pi(h+l) + 2\cos\pi(k+l) + 2\cos\pi(k-l) \\ &\quad + 2\cos\pi(h-l) + 2\cos\pi(h-k) \end{aligned} \quad (\text{C.98})$$

This can then be used to produce expected reflections for the prototypical FCC lattice by substituting in the values of the Miller indices  $h$ ,  $k$  and  $l$ . Table C.1 contains expected reflections for Miller indices in one octant for  $h$ ,  $k$  and  $l$  assigned as 0 to 2. Repeated results are grouped.

Table C.1

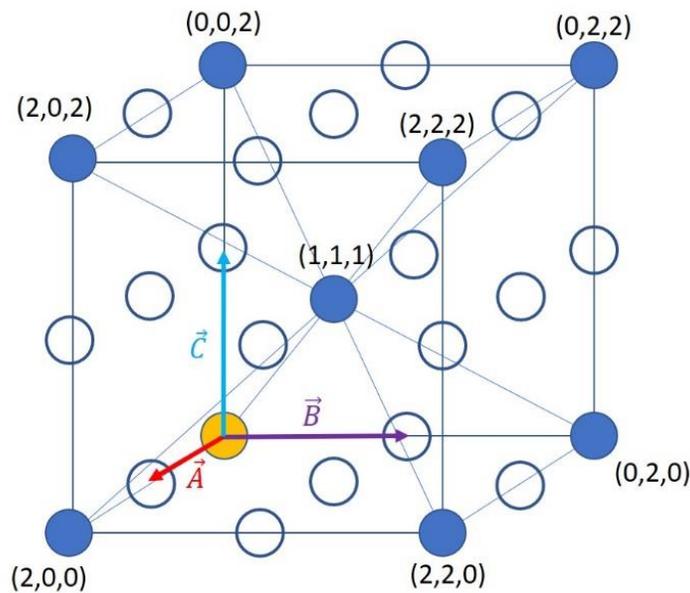
Index	Reflection expected?	Duplicates
(0,0,0)	Origin	
(1,0,0)	No	(0,1,0) (0,0,1)
(2,0,0)	Yes	(0,2,0) (0,0,2)
(1,1,0)	No	(1,0,1) (0,1,1)
(2,1,0)	No	(0,2,1) (1,0,2) (1,2,0) (0,1,2) (2,0,1)
(2,2,0)	Yes	(2,0,2) (0,2,2)
(2,1,1)	No	(1,2,1) (1,1,2)
(2,2,1)	No	(2,1,2) (1,2,2)
(2,2,2)	Yes	
(1,1,1)	Yes	

This table covers all the unique indexing of the reciprocal space of an FCC telling which indices should be present or absent. Other octants in reciprocal space would duplicate this octant by switching the index signs (+/-) as needed.

Despite 1D powder averaging, the basis vectors of the FCC reciprocal space are all equal  $\vec{A} = \vec{B} = \vec{C}$ , which means that all possible  $q$  values of a cubic lattice can be assigned based on deciding only a single fundamental basis vector.  $q$  values can then be predicted by the following formula and checked as present or absent in the data based on Table C.1.

$$q(h, k, l) = |\vec{A}| \sqrt{h^2 + k^2 + l^2} \quad (\text{C.99})$$

If necessary, Table C.1 can be expanded by noting that the expected reflections produce a body centered cubic lattice in reciprocal space. Here, filled circles are expected reflections while empty circles are omitted reflections.



The real space dimensions of the FCC are easily determined from the fundamental reflection of the reciprocal lattice since all basis vectors are purely orthogonal.

$$|\vec{a}| = \frac{2\pi}{|\vec{A}|} \quad (\text{C.100})$$

## REFERENCES

1. Levine, J. S. *Prentice Hall: Biology*. (Prentice Hall, 2007).
2. Line, M. A. The enigma of the origin of life and its timing. *Microbiology* **148**, 21–27 (2002).
3. Goldenfeld, N. & Woese, C. Life is Physics: Evolution as a Collective Phenomenon Far From Equilibrium. *Annu. Rev. Condens. Matter Phys.* **2**, 375–399 (2011).
4. Joyce, G. F. The antiquity of RNA-based evolution. *Nature* **418**, 214–221 (2002).
5. Biello, D. & Biello, D. The Origin of Oxygen in Earth's Atmosphere. *Scientific American*  
Available at: <https://www.scientificamerican.com/article/origin-of-oxygen-in-atmosphere/>.  
(Accessed: 19th March 2018)
6. D, E. W. & E. *Biochemistry and Molecular Biology*. (Oxford University Press, 2005).
7. Baker, T. A. *DNA Replication*. (W H Freeman & Co, 1992).
8. Gilbert, W. Origin of life: The RNA world. *Nature* **319**, 618 (1986).
9. Robertson, M. P. & Joyce, G. F. The Origins of the RNA World. *Cold Spring Harb. Perspect. Biol.* **4**, a003608–a003608 (2012).
10. Lakdawala, S. S., Fodor, E. & Subbarao, K. Moving On Out: Transport and Packaging of Influenza Viral RNA into Virions. *Annu. Rev. Virol.* **3**, 411–427 (2016).
11. Petrov, A. S. *et al.* History of the ribosome and the origin of translation. *Proc. Natl. Acad. Sci.* **112**, 15396–15401 (2015).
12. Kovacs, N. A., Petrov, A. S., Lanier, K. A. & Williams, L. D. Frozen in Time: The History of Proteins. *Mol. Biol. Evol.* **34**, 1252–1260 (2017).
13. Leslie E., O. Prebiotic Chemistry and the Origin of the RNA World. *Crit. Rev. Biochem. Mol. Biol.* **39**, 99–123 (2004).

14. Yang, Y.-W., Zhang, S., McCullum, E. O. & Chaput, J. C. Experimental Evidence That GNA and TNA Were Not Sequential Polymers in the Prebiotic Evolution of RNA. *J. Mol. Evol.* **65**, 289–295 (2007).
15. Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D. & Sutherland, J. D. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat. Chem.* **7**, 301–307 (2015).
16. WESTHEIMER, F. H. Why Nature Chose Phosphates. **235**, 6 (1987).
17. Famiano, M. A., Boyd, R. N., Kajino, T. & Onaka, T. Selection of Amino Acid Chirality via Neutrino Interactions with  $^{14}\text{N}$  in Crossed Electric and Magnetic Fields. *Astrobiology* **18**, 190–206 (2017).
18. Rossi, M., Zanchetta, G., Klussmann, S., Clark, N. A. & Bellini, T. Propagation of Chirality in Mixtures of Natural and Enantiomeric DNA Oligomers. *Phys. Rev. Lett.* **110**, (2013).
19. Kvenvolden, K. *et al.* Evidence for Extraterrestrial Amino-acids and Hydrocarbons in the Murchison Meteorite. *Nature* **228**, 923–926 (1970).
20. Gennes, P. G. de & Prost, J. *The Physics of Liquid Crystals*. (Clarendon Press, 1995).
21. Franklin, R. E. & Gosling, R. G. Molecular Configuration in Sodium Thymonucleate. *Nature* **171**, 740–741 (1953).
22. Watson, J. D. Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. 1953. *JAMA J. Am. Med. Assoc.* **269**, 1966–1967 (1993).
23. Livolant, F., Levelut, A. M., Doucet, J. & Benoit, J. P. The highly concentrated liquid-crystalline phase of DNA is columnar hexagonal. *Nature* **339**, 724–726 (1989).
24. Yevdokimov, Y. M., Salyanov, V. I., Semenov, S. V. & Skuridin, S. G. *DNA Liquid-Crystalline Dispersions and Nanoconstructions*. (CRC Press, 2011).

25. Nakata, M. *et al.* End-to-End Stacking and Liquid Crystal Condensation of 6- to 20-Base Pair DNA Duplexes. *Science* **318**, 1276–1279 (2007).
26. Zanchetta, G., Bellini, T., Nakata, M. & Clark, N. A. Physical Polymerization and Liquid Crystallization of RNA Oligomers. *J. Am. Chem. Soc.* **130**, 12864–12865 (2008).
27. Zanchetta, G., Nakata, M., Buscaglia, M., Clark, N. A. & Bellini, T. Liquid crystal ordering of DNA and RNA oligomers with partially overlapping sequences. *J. Phys. Condens. Matter* **20**, 494214 (2008).
28. Kuriabova, T., Betterton, M. D. & Glaser, M. A. Linear aggregation and liquid-crystalline order: comparison of Monte Carlo simulation and analytic theory. *J. Mater. Chem.* **20**, 10366 (2010).
29. Asakura, S. & Oosawa, F. On Interaction between Two Bodies Immersed in a Solution of Macromolecules. *J. Chem. Phys.* **22**, 1255–1256 (1954).
30. Zanchetta, G., Nakata, M., Buscaglia, M., Bellini, T. & Clark, N. A. Phase separation and liquid crystallization of complementary sequences in mixtures of nanoDNA oligomers. *Proc. Natl. Acad. Sci.* **105**, 1111–1117 (2008).
31. Bellini, T. *et al.* Liquid crystal self-assembly of random- sequence DNA oligomers. *Comput. Biol.* **6**
32. Taran, O., Thoennesen, O., Achilles, K. & von Kiedrowski, G. Synthesis of information-carrying polymers of mixed sequences from double stranded short deoxynucleotides. *J. Syst. Chem.* **1**, 9 (2010).
33. Fraccia, T. P. *et al.* Abiotic ligation of DNA oligomers templated by their liquid crystal ordering. *Nat. Commun.* **6**, (2015).

34. Fraccia, T. P. *et al.* Liquid Crystal Ordering and Isotropic Gelation in Solutions of Four-Base-Long DNA Oligomers. *ACS Nano* **10**, 8508–8516 (2016).
35. Wing, R. *et al.* Crystal structure analysis of a complete turn of B-DNA. *Nature* **287**, 755–758 (1980).
36. Yang, C., Kim, E. & Pak, Y. Free energy landscape and transition pathways from Watson–Crick to Hoogsteen base pairing in free duplex DNA. *Nucleic Acids Res.* **43**, 7769–7778 (2015).
37. Bonazzi, S. *et al.* Four-stranded aggregates of oligodeoxyguanylates forming lyotropic liquid crystals: a study by circular dichroism, optical microscopy, and x-ray diffraction. *J. Am. Chem. Soc.* **113**, 5809–5816 (1991).
38. Davis, J. T. G-Quartets 40 Years Later: From 5'-GMP to Molecular Biology and Supramolecular Chemistry. *Angew. Chem. Int. Ed.* **43**, 668–698 (2004).
39. Mezzina, E. *et al.* The Self-Assembly of a Lipophilic Guanosine Nucleoside into Polymeric Columnar Aggregates: The Nucleoside Structure Contains Sufficient Information To Drive the Process towards a Strikingly Regular Polymer. *Chemistry* **7**, 388–395 (2001).
40. Pinnavaia, T. J. *et al.* Alkali metal ion specificity in the solution ordering of a nucleotide, 5'-guanosine monophosphate. *J. Am. Chem. Soc.* **100**, 3625–3627 (1978).
41. Lydon, J. Chromonic review. *J. Mater. Chem.* **20**, 10071 (2010).
42. Lydon, J. Chromonic mesophases. *Curr. Opin. Colloid Interface Sci.* **8**, 480–490 (2004).
43. Lydon, J. Chromonic liquid crystalline phases. *Liq. Cryst.* **38**, 1663–1681 (2011).
44. Nadassy, K. Standard atomic volumes in double-stranded DNA and packing in protein-DNA interfaces. *Nucleic Acids Res.* **29**, 3362–3376 (2001).

45. Park, H.-S. *et al.* Self-Assembly of Lyotropic Chromonic Liquid Crystal Sunset Yellow and Effects of Ionic Additives. *J. Phys. Chem. B* **112**, 16307–16319 (2008).
46. Klump, H. H., Völker, J., Maeder, D. L., Niermann, T. & Sobolewski, C. H. M. Conformational changes in nucleic acids/chromatin structure. *Thermochim. Acta* **193**, 391–415 (1991).
47. Vesnaver, G. & Breslauer, K. J. The contribution of DNA single-stranded order to the thermodynamics of duplex formation. *Proc. Natl. Acad. Sci.* **88**, 3569–3573 (1991).
48. Kornyshev, A. A., Lee, D. J., Leikin, S. & Wynveen, A. Structure and interactions of biological helices. *Rev. Mod. Phys.* **79**, 943–996 (2007).
49. Beaucage, S. L. & Caruthers, M. H. Deoxynucleoside phosphoramidites—A new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Lett.* **22**, 1859–1862 (1981).
50. Caruthers, M. H. The Chemical Synthesis of DNA/RNA: Our Gift to Science. *J. Biol. Chem.* **288**, 1420–1427 (2013).
51. Dahl, B. H., Nielsen, J. & Dahl, O. Mechanistic studies on the phosphoramidite coupling reaction in oligonucleotide synthesis. I. Evidence for nucleophilic catalysis by tetrazole and rate variations with the phosphorus substituents. *Nucleic Acids Res.* **15**
52. Sauer, S. The essence of DNA sample preparation for MALDI mass spectrometry. *J. Biochem. Biophys. Methods* **70**, 311–318 (2007).
53. Shahgholi, M., Garcia, B. A., Chiu, N. H. L., Heaney, P. J. & Tang, K. Sugar additives for MALDI matrices improve signal allowing the smallest nucleotide change (A:T) in a DNA sequence to be resolved. *Nucleic Acids Res.* **29**, e91–e91 (2001).
54. Absorbance. *Wikipedia* (2017).

55. Kumar, A. A New Solid Phase Method for the Synthesis of Oligonucleotides with Terminal - 3'-Phosphate. *Nucleosides Nucleotides* **12**, 441–447 (1993).
56. Lebedev, A. V., Koukhareva, I. I., Beck, T. & Vaghefi, M. M. PREPARATION OF OLIGODEOXYNUCLEOTIDE 5'-TRIPHOSPHATES USING SOLID SUPPORT APPROACH. *Nucleosides Nucleotides Nucleic Acids* **20**, 1403–1409 (2001).
57. Drenth, J. *Principles of Protein X-ray Crystallography*. (Springer Science & Business Media, 2002).
58. Guinier, A. *X-Ray Diffraction: In Crystals, Imperfect Crystals, and Amorphous Bodies*. (Courier Corporation, 2013).
59. Kittel, C. X-Ray Diffraction from Helices: Structure of DNA. *Am. J. Phys.* **36**, 610–616 (1968).
60. Luckhurst, G. R. & Zannoni, C. Why is the Maier–Saupe theory of nematic liquid crystals so successful? *Nature* **267**, 412–414 (1977).
61. Hosemann, R. Die parakristalline Feinstruktur natürlicher und synthetischer Eiweisse. Visuelles Näherungsverfahren zur Bestimmung der Schwankungstensoren von Gitterzellen. *Acta Crystallogr.* **4**, 520–530 (1951).
62. Lindenmeyer, P. H. & Hosemann, R. Application of the Theory of Paracrystals to the Crystal Structure Analysis of Polyacrylonitrile. *J. Appl. Phys.* **34**, 42–45 (1963).
63. Hexemer, A. *et al.* A SAXS/WAXS/GISAXS Beamline with Multilayer Monochromator. *J. Phys. Conf. Ser.* **247**, 012007 (2010).
64. Ilavsky, J. Nika: software for two-dimensional data reduction. *J. Appl. Crystallogr.* **45**, 324–328 (2012).