

DEVELOPMENT AND APPLICATION OF COMPUTATIONAL
TOOLS FOR ORGANIC MOLECULAR STRUCTURE
PREDICTION FROM SMALL MOLECULE CRYSTAL
STRUCTURE DETERMINATION TO POLYMER PERSISTENCE
LENGTH.

by

Madilyn E. Paul

A dissertation
submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Materials Science and Engineering
Boise State University

November 2024

© 2024

Madilyn E. Paul

ALL RIGHTS RESERVED

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the dissertation submitted by

Madilyn E. Paul

Dissertation Title: Development and application of computational tools for organic molecular structure prediction from small molecule crystal structure determination to polymer persistence length.

Date of Final Oral Examination: November 21, 2024

The following individuals read and discussed the dissertation submitted by student Madilyn E. Paul, and they evaluated the student's presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

Eric Jankowski Ph.D. Chair, Supervisory Committee

Scott Phillips Ph.D. Member, Supervisory Committee

Oliviero Andreussi Ph.D. Member, Supervisory Committee

Frank Willmore Ph.D. Member, Supervisory Committee

The final reading approval of the dissertation was granted by Eric Jankowski Ph.D., Chair of the Supervisory Committee. The dissertation was approved by the Graduate College.

DEDICATION

To my family, for always believing I could do the impossible.

ACKNOWLEDGMENT

I'd like to start by thanking my advisor, Dr. Eric Jankowski. I came into Dr. Jankowski's lab in quite an unusual way, being at the end of my 3rd year in the MSE PhD program with years of research under my belt and needing to find a new research lab. Needing to find a new research group with short notice I was rather stressed when I first met with Eric, but after discussing my situation and goals with him I left the meeting feeling calm and much more capable of achieving my goals. This is the kind of environment that Eric fosters in his research group. Thank you, Eric, for always displaying your passion and enthusiasm for our research and creating such a welcoming learning experience. It has been a true privilege to work with you as well as all the excellent researchers that I've crossed paths with in the CME lab. I have gain invaluable experiences and friendships from the CME lab that I will continuously cherish.

I'd also like to thank my first research advisor, Dr. Matt King. Thank you for answering my endless questions and teaching me how to be a researcher. I can't thank the lovely MSMSE administrative staff for supporting me throughout the program enough. Each and every one of you truly goes above and beyond for the students in the MSMSE program and make the program such an enjoyable experience. I would like to give thanks to all of our funding sources. This material is based upon work supported by the National Science Foundation under Grant No. 1653954.

To my family, thank you for believing I could accomplish anything, even when I doubted it myself. To my parents, Ken and Shellie, thank you for doing everything you could to ensure I would accomplish every goal I've ever set. I wouldn't be a quarter of the person I am today without having such strong people in my corner. You both are truly my heroes. To my sister, Emily, thank you for being my role model, support system and overall my best friend. I love you all more than word can describe. To Joel, thank you for loving and supporting me through every up and down this journey has had to offer. I couldn't ask for a better partner in this life.

ABSTRACT

This dissertation utilized two areas of computational simulations for organic molecular structure prediction. First, we investigate the possible polymorphism of both barbituric acid dihydrate (BTADH) and violuric acid monohydrate (VAMH) using density functional theory. Crystalline structures that exhibit polymorphism have an increased level of uncertainty in predicting their chemical and physical properties. BTADH was previously thought to undergo a subtle phase transition from $P2_1/n$ to planar $Pnma$ at 217 K. We conclude that barbituric acid dihydrate exhibits crystalline disorder in which the molecules are crystallized in a non-merohedrally twinned $P2_1/n$ configuration below 217 K. At temperatures above 217 K, however rather than a phase transition to $Pnma$, the molecules have enough thermal energy to oscillate between the two twinned configurations. A similar planar configuration was reported for violuric acid monohydrate, which lead us to investigate the crystalline configuration. In modeling violuric acid monohydrate we redetermined the space group from the planar $Cmc2_1$ configuration to a non-planar $Pca2_1$. Both VAMH and BTADH exhibited molecule oscillations between twinned configurations that led to the incorrect determination of the planar space groups. Moving away from first principle simulations, we began to explore the realm of molecular dynamics. As apposed to density functional theory, MD simulations make approximations that significantly lower the computational cost. Due to this, we were able to investigate the bulk morphologies of

larger molecule systems, such as macromolecules and polymers. Investigating these systems has been proven difficult due to the lack of molecular information, such as bond angles and dihedral parameters. In this study, we implemented and validated the use of machine-learned forcefield parameters for macromolecules and complex polymer systems. We created an end-to-end workflow that allowed us to predict the morphologies of these complex systems beginning with a simple SMILES string of the molecules. We then used this validated workflow to predict the morphology of 13 donor-acceptor co-polymers with photovoltaic properties. Lastly, we built a function to determine the rigidity of the 13 co-polymers and compared the simulation results to published experimental results in efforts to investigate the physical properties of these polymers from the predicted morphologies.

TABLE OF CONTENTS

DEDICATION	iv
ACKNOWLEDGMENT	v
ABSTRACT	vii
LIST OF TABLES	xiii
LIST OF FIGURES	xvi
NOMENCLATURE	xxv
1 INTRODUCTION	1
1.1 Polymorphism	1
1.2 Density Functional Theory	2
1.2.1 Functionals and Basis Sets	4
1.3 Organic Photovoltaics	6
1.4 Molecular Dynamics	9
2 TRUE POLYMORPHIC PHASE TRANSITION OR DYNAMIC CRYSTAL DISORDER? AN INVESTIGATION INTO THE UNUSUAL PHASE BE- HAVIOR OF BARBITURIC ACID DIHYDRATE	14

2.1	Abstract	14
2.2	Introduction	15
2.3	Experimental Section	18
2.3.1	THz Spectroscopy	18
2.3.2	Solid-State Density Functional Theory	19
2.4	Results and Discussion	20
2.5	Conclusions	34
2.6	Acknowledgements	35
3	COMPUTATIONAL REDETERMINATION OF VIOLURIC ACID MONOHYDRATE'S CRYSTALLINE CONFORMATION	36
3.1	Summary	36
3.2	Introduction	37
3.3	Materials and Methods	39
3.3.1	Experimental THz Measurements	39
3.3.2	Solid-State Density Functional Theory	40
3.4	Results and Discussion	41
3.5	Conclusion	52
4	VALIDATING STRUCTURAL PREDICTIONS OF CONJUGATED MACROMOLECULES IN ESPALOMA-ENABLED REPRODUCIBLE WORKFLOWS	53
4.1	Abstract	53
4.2	Introduction	54
4.3	Model	56
4.4	Methods	58

4.4.1	Molecular Dynamics	59
4.4.2	Morphology Characterization	59
4.4.3	Integrating Espaloma parameterization into MoSDeF	60
4.5	Results and Discussion	62
4.5.1	Espaloma vs OPLS-UA and GAFF	62
4.5.2	Perylene	64
4.5.3	Poly-3-Hexylthiophene (P3HT)	68
4.6	Conclusions	71
5	TEMPERATURE DEPENDENT CHAIN RIGIDITY STUDY OF ORGANIC PHOTOVOLTAIC DONOR-ACCEPTOR CO-POLYMERS.	72
5.1	Summary	72
5.2	Introduction	73
5.3	Methods	78
5.4	Results and Discussion	81
5.4.1	PCPDTPT Backbone	82
5.4.2	PCPDTFBT Backbone	84
5.4.3	IDT Backbone	85
5.4.4	MD vs DFT	87
5.5	Conclusions	88
6	CONCLUSIONS	89
	REFERENCES	90
	APPENDICES	114

A	ESPALOMA FUNCTION CODE	115
B	BOND WALKER EXAMPLE CODE	120
C	PERSISTENCE LENGTH APPENDIX	123

LIST OF TABLES

2.1	Lattice parameters and calculated relative molecular orientations from BTADH experimental X-ray crystal structures reported in Ref. [1].	22
2.2	Central frequencies and linewidths of THz spectral features in the range of $20 - 95 \text{ cm}^{-1}$ for data collected at 78 K.	24
2.3	Angles between the molecular planes of two molecules within the packed unit cell, angles between the two water molecules within the unit cell and the angles between the CH_2 plane and the molecular plane of one molecule.	28
2.4	Calculated energies (per $\text{BTA}\cdot 2\text{H}_2\text{O}$) of BTADH confined to $P2_1/n$ and $Pnma$ symmetry groups at experimental lattice parameters. ΔE provided in units of kJ mol^{-1} .	29
2.5	Calculated phonon mode frequencies and intensities of full-relaxed BTADH unit cell in $P2_1/n$ symmetry using the PBE functional with the pob-TZVP and 6-311G(d,p) basis sets.	29
3.1	Heavy atom distances for intermolecular contacts observed in the $Cmc2_1$ and $P2_1$ optimized unit cells. .	44
3.2	Tabulated comparison of the unit cell parameters and electronic energies of the investigated symmetry groups.	49

4.1	Lennard-Jones diameters σ and well-depths ϵ generated by <code>espaloma</code> in the present work, and OPLS-UA and GAFF	63
4.2	A list of statepoints used in the P3HT and perylene simulations, as well as the base units for each simulation (σ, ϵ, M).	64
5.1	Persistence lengths calculated from MD predicted morphologies (l_p^{MD}) of polymers with a PCPDTPT backbone in comparison to persistence lengths calculated from SANS data (l_p^{SANS}). All l_p presented in units of Å.	83
5.2	Predicted simulation temperature at which l_p^{MD} would match l_p^{SANS} for polymers with PCPDTPT backbone. T_{SANS} calculated via linear regressions shown in Appendix C. T_{SANS} reported in units of K. . . .	83
5.3	Persistence lengths calculated from MD predicted morphologies (l_p^{MD}) of polymers with a PCPDTFBT backbone in comparison to persistence lengths calculated from SANS data (l_p^{SANS}). All l_p presented in units of Å.	84
5.4	Predicted simulation temperature at which l_p^{MD} would match l_p^{SANS} for polymers with PCPDTFBT backbone. T_{SANS} calculated via linear regressions shown in Appendix C. T_{SANS} reported in units of K. . . .	85
5.5	Persistence lengths calculated from MD predicted morphologies (l_p^{MD}) of polymers with a IDT backbone in comparison to persistence lengths calculated from SANS data (l_p^{SANS}). All l_p presented in units of Å. . .	86
5.6	Predicted simulation temperature at which l_p^{MD} would match l_p^{SANS} for polymers with IDT backbone. T_{SANS} calculated via linear regressions shown in Appendix C. T_{SANS} reported in units of K.	87

5.7 A comparison of l_p^{DFT} reported by Danielsen, et al. at Ref [2] to l_p^{MD} calculated at a temperature of 1006 K.	87
C.1 Densities at which each polymer simulation was run at. All densities reported in units of g/mL	123
C.2 Average timesteps per second (TPS) for each polymer simulation. Values averaged over each simulation with temperatures ranging from 252 K to 2013 K.	124
C.3 Caption	124
C.4 Predicted simulation temperature at which l_p^{MD} would match l_p^{SANS} . .	125

LIST OF FIGURES

1.1	Diagram of a polymorphic phase transition from Polymorph A to Polymorph B.	1
1.2	Diagram of Solvatomorphs.	3
1.3	A depiction of how density functional theory simplifies the many-body problem into a function of electron density.	4
1.4	Diagram of an OPV device. (1) Photon absorption (2) Exciton creation and migration (3) Exciton dissociation (4) Charge carrier diffusion. .	7
1.5	Depiction of a bulk heterojunction (BHJ) active layer of an organic photovoltaic device. Outlined in green is a an "island" where the electron would have no access to the electrode, forcing it to recombine with a hole, rendering it useless.	8
1.6	Plot of the Lennard-Jones potential energy.	10
1.7	Illustration of bond distance (r), bond angle (θ), and dihedral angle (ϕ)	12
2.1	A) Molecular structure of barbituric acid. Comparison of the crystalline structures of the temperature-dependent BTADH polymorphs, with the high-temperature $Pnma$ form viewed along the B) a-axis and C) c-axis, and the low-temperature $P2_1/n$ form along the D) b-axis and E) c-axis.	17

2.2	Plane orientations relating angle metrics for determining the quality of unit cell reproductions. A) BTADH molecular plane defined by ring carbonyl carbons, and that representing out-of-plane position of the ring CH ₂ group. B) Opposing molecular planes of two symmetry-related BTADH molecules in the low-temperature <i>P2₁/n</i> polymorph. C) Planes each created by a single H ₂ O molecule of the dihydrate asymmetric unit.	21
2.3	A) THz vibrational spectra of BTADH in the spectral range of 20-95 cm ⁻¹ obtained at 293 K and 78 K. B) Spectral fitting of the 78 K THz spectrum (red circles) with Lorentzian peaks (dashed lines) on exponential background (green line). The residuals of the experimental data and fitted values are indicated by blue circles.	23
2.4	A) Scan of potential energy surface along the -35.79 cm ⁻¹ imaginary phonon mode (PBE/pob-TZVP) of BTADH confined to the <i>Pnma</i> space group at 270-K lattice parameters. B) Displacement vectors (yellow arrows, magnitudes scaled ×10) for the imaginary mode shown for top two BTADH units. Displacements are oriented normal to the reflection plane (purple line) confining the molecules in <i>Pnma</i> symmetry. C) Molecular orientation of BTADH optimized in <i>P2₁/n</i> symmetry.	26

2.5 A) Mode character of the spectral feature at 33.6 cm ⁻¹ observed in the 78 K THz spectrum, represented by calculated displacement vectors (yellow arrows, magnitudes scaled ×10). B) Calculated mode displacements of the broad feature centered at 66.7 cm ⁻¹ in the 78 K THz spectrum. C) Mode character of the distinct 88.8 cm ⁻¹ vibrational mode observed in the 293 K THz spectrum (corresponding to the feature located at 92.7 cm ⁻¹ in the 78 K spectrum).	30
2.6 (A) Mode character of the spectral features at 33.6 cm ⁻¹ observed in the 78 K THz spectrum, represented by calculated displacement vectors (yellow arrows, magnitudes scaled ×10). (B) Calculated mode displacements of the broad feature centered at 66.9 cm ⁻¹ in the 78 K THz spectrum. (C) Mode character of the distinct 88.8 cm ⁻¹ vibrational mode observed in the 293 K THz spectrum (corresponding to the feature located at 92.7 cm ⁻¹ in the 78 K spectrum).	32
3.1 A) Violuric acid monohydrate (VAMH) molecular structure and B) <i>Cmc2</i> ₁ crystal configuration. Viewed along c-axis.	38
3.2 Potential energy scan along imaginary mode at 150 K.	42
3.3 A) Vibrational character of the imaginary mode clauclated at 22.33 cm ⁻¹ for VAMH constrained to <i>Cmc2</i> ₁ space group symmetry. Atomic displacement vectors are represented by yellow arrows (magnitudes scaled ×10).	43
3.4 A) <i>P</i> ₁ and <i>P</i> ₂ ₁ unit cell overlay. B) Packed <i>P</i> ₁ unit cell viewed along the c axis. C) Packed <i>P</i> ₂ ₁ unit cell viewed along the b axis.	45

3.5	Depicted planes between A) violuric acid molecules and B) water violuric acid molecular planes of VAMH confined to the $P2_1$ space group.	46
3.6	Primary electrostatic contacts observed in the VAMH crystal structures. Hydrogen bonds are indicated by dashed green lines and repulsive interactions by dashed black lines.	47
3.7	Comparison of the simulated $P2_1$ THz spectrum to the experimental THz spectrum obtained at 20 K. Lorentzian lineshapes of FWHM 5 cm ⁻¹ were convolved into calculated normal modes, and mode intensities of calculated and experimental spectra were normalized.	48
3.8	Molecular packing of the predicted VAMH $Pca2_1$ crystal structure. .	49
3.9	Comparison of the simulated $Pca2_1$ THz spectrum to the experimental THz spectrum obtained at 20 K. Lorentzian lineshapes of FWHM 5 cm ⁻¹ were convolved into calculated normal modes, and mode intensities of calculated and experimental spectra were normalized.	51
4.1	Depictions of our generalized molecular dynamics workflow. Step 1 shows the creation of an <code>mBuild Compound</code> from a SMILES string. In step 2 we create our simulation object using <code>flowerMD</code> and <code>PACKMOL</code> . We employ <code>Espaloma</code> to parameterize our molecules in step 3 and write the forcefield file. In step 4 we initialize the <code>Hoomd-Blue</code> simulation and predict the morphology of our molecules.	56
4.2	Diagram of perylene molecule (left) and poly-3-hexylthiophene (P3HT) monomer (right).	57
4.3	Blue spheres represent center-of-geometry positions used for RDFs and clustering criteria. Perylene (left), and P3HT (right).	60

4.4 Overall Espaloma-MosDeF workflow. (1) mBuild compounds are used to create the topology of an openMM molecule, (2) BondWalker uses octet rules to determine double bonds in the openMM molecule. (3) Espaloma generates forcefield parameters for the openMM molecule. (4) The ESP-UA forcefield is used to re-type the mBuild compound for use in Hoomd-Blue simulations.	61
4.5 Double- and triple-bond information that is missing from mBuild compounds is retrieved via BondWalker by iteratively checking whether the octet rules can be satisfied for all atoms after incrementing the bond character adjacent to atoms with unsatisfied octets.	62
4.6 Temperature vs density clustering order parameter (Ψ) phase diagram of perylene at 12 statepoints.	65
4.7 Snapshot of perylene taken from the most ordered morphology at a density of 0.5 g/cm^3 and temperature of 248.8 K	65
4.8 (A) Radial distribution function (RDF) of Perylene at various temperatures at a density of $0.5\text{g}/\text{cm}^3$ generated from an ESP-UA predicted morphology. (B) RDF of perylene in various phases generated from an OPLS-UA predicted morphology. OPLS-UA RDF published by Miller, et al. at Ref [3].	66

4.9 (A) Grazing incident x-ray scattering pattern of Perylene generated from an ESP-UA morphology. (B) GIXS pattern of Perylene generated from an OPLS-UA morphology. OPLS-UA GIXS pattern published by Miller, et. al. at Ref [3]. (C) Experimental XRD pattern for β -perylene, reproduced with permission from Ishii, et al. at Ref [4]. Copyright 2014 AIP Publishing LLC.	67
4.10 Temperature vs density clustering order parameter (Ψ) phase diagram of P3HT at 12 statepoints.	68
4.11 (A) Radial distribution function (RDF) of P3HT at a temperature of 304 K and a density of $0.5\text{g}/\text{cm}^3$ generated from an ESP-UA predicted morphology. (B) RDF of P3HT generated from an OPLS-UA predicted morphology. OPLS-UA RDF published by Miller, et al. in Ref [5].	69
4.12 (A) Grazing Incident X-ray Scattering Pattern of P3HT at $0.5\text{ g}/\text{cm}^3$ and 304 K generated using ESP-UA. (B) Corresponding experimental scattering pattern of P3HT. Reprinted (adapted) with permission from Ko, et al. in Ref [6]. Copyright 2012 American Chemical Society.	70
4.13 Snapshot of P3HT's most ordered morphology at a density of $0.5\text{ g}/\text{cm}^3$ and temperature of 304 K.	70
5.1 Diagram of how a more rigid polymer (high l_p) can lead to higher order in the morphology compared to a more flexible polymer (low l_p).	73
5.2 Molecular structures of the five repeating units used to build the polymers studied in this chapter. Each repeating includes a donor and acceptor region.	75
5.3 Molecular structures of the six R-groups used as side chains in this study.	77

5.4	PCPDTPT-HD geometric centers used in calculating persistence length. The pink beads represent the monomer center of mass for the backbone molecules, CPDT on the left and PT on the right. The side chains were not included in the calculation of the center of mass.	79
5.5	Diagram of a condensed polymer. Row A, B and C show the direc- tions of the bond vectors. A snapshot of a condensed PCPDTPT-HD polymer can be found at Figure C.1.	81
C.1	Snapshot of coarse grained PCPDTPT-HD at a temperature of 252 K and density of 0.05 g/mol. The red highlighted CG beads have parallel aligned bond vectors. The blue highlighted CG beads are parallel, but anti-aligned with the red.	129
C.2	Persistence length (l_p^{MD}) of PCPDTPT-eneHD vs Temperature plot with a linear regression calculated using SciPy [7] and T_{SANS} prediction.	130
C.3	Persistence length (l_p^{MD}) of PCPDTPT-eneODD vs Temperature plot with a linear regression calculated using SciPy [7] and T_{SANS} prediction.	131
C.4	Persistence length (l_p^{MD}) of PCPDTPT-HD vs Temperature plot with a linear regression calculated using SciPy [7] and T_{SANS} prediction. .	132
C.5	Persistence length (l_p^{MD}) of PCPDTPT-ODD vs Temperature plot with a linear regression calculated using SciPy [7] and T_{SANS} prediction. .	133
C.6	Persistence length (l_p^{MD}) of PCPDTPT-nC16 vs Temperature plot with a linear regression calculated using SciPy [7] and T_{SANS} prediction. .	134
C.7	Persistence length (l_p^{MD}) of PCPDTFBT-C1-BO vs Temperature plot with a linear regression calculated using SciPy [7] and T_{SANS} prediction.	135

- C.8 Persistence length (l_p^{MD}) of PCPDTFBT-C3-BO vs Temperature plot
with a linear regression calculated using SciPy [7] and T_{SANS} prediction.136
- C.9 Persistence length (l_p^{MD}) of PCPDTFBT-C4-BO vs Temperature plot
with a linear regression calculated using SciPy [7] and T_{SANS} prediction.137
- C.10 Persistence length (l_p^{MD}) of PCPDTFBT-C5-BO vs Temperature plot
with a linear regression calculated using SciPy [7] and T_{SANS} prediction.138
- C.11 Persistence length (l_p^{MD}) of PCPDTFBT-C11-BO vs Temperature plot
with a linear regression calculated using SciPy [7] and T_{SANS} prediction.139
- C.12 Persistence length (l_p^{MD}) of PIDTBT-nC16 vs Temperature plot with
a linear regression calculated using SciPy [7] and T_{SANS} prediction. . . 140
- C.13 Persistence length (l_p^{MD}) of PIDTCPDT-C11-BO vs Temperature plot
with a linear regression calculated using SciPy [7] and T_{SANS} prediction.141
- C.14 Persistence length (l_p^{MD}) of PIDTFBT-C11-BO vs Temperature plot
with a linear regression calculated using SciPy [7] and T_{SANS} prediction.142
- C.15 $\sqrt{\langle R^2 \rangle}$ vs Temperature for PCPDTFBT-C1-BO with a linear regres-
sion for the linear portion of the plot. Horizontal line plotted at $\sqrt{N_b} * l_b$.143
- C.16 $\sqrt{\langle R^2 \rangle}$ vs Temperature for PCPDTFBT-C3-BO with a linear regres-
sion for the linear portion of the plot. Horizontal line plotted at $\sqrt{N_b} * l_b$.143
- C.17 $\sqrt{\langle R^2 \rangle}$ vs Temperature for PCPDTFBT-C4-BO with a linear regres-
sion for the linear portion of the plot. Horizontal line plotted at $\sqrt{N_b} * l_b$.144
- C.18 $\sqrt{\langle R^2 \rangle}$ vs Temperature for PCPDTFBT-C5-BO with a linear regres-
sion for the linear portion of the plot. Horizontal line plotted at $\sqrt{N_b} * l_b$.144
- C.19 $\sqrt{\langle R^2 \rangle}$ vs Temperature for PCPDTFBT-C11-BO with a linear regres-
sion for the linear portion of the plot. Horizontal line plotted at $\sqrt{N_b} * l_b$.145

- C.20 $\sqrt{\langle R^2 \rangle}$ vs Temperature for PCPDTPT-ene-HD with a linear regression
for the linear portion of the plot. Horizontal line plotted at $\sqrt{N_b} * l_b$ 145
- C.21 $\sqrt{\langle R^2 \rangle}$ vs Temperature for PCPDTPT-ene-ODD with a linear regression
for the linear portion of the plot. Horizontal line plotted at $\sqrt{N_b} * l_b$ 146
- C.22 $\sqrt{\langle R^2 \rangle}$ vs Temperature for PCPDTPT-HD with a linear regression for
the linear portion of the plot. Horizontal line plotted at $\sqrt{N_b} * l_b$ 146
- C.23 $\sqrt{\langle R^2 \rangle}$ vs Temperature for PCPDTPT-nC16 with a linear regression
for the linear portion of the plot. Horizontal line plotted at $\sqrt{N_b} * l_b$ 147
- C.24 $\sqrt{\langle R^2 \rangle}$ vs Temperature for PCPDTPT-ODD with a linear regression
for the linear portion of the plot. Horizontal line plotted at $\sqrt{N_b} * l_b$ 147
- C.25 $\sqrt{\langle R^2 \rangle}$ vs Temperature for PIDTBT-nC16 with a linear regression for
the linear portion of the plot. Horizontal line plotted at $\sqrt{N_b} * l_b$ 148
- C.26 $\sqrt{\langle R^2 \rangle}$ vs Temperature for PIDTCPDT-C11-BO with a linear regression
for the linear portion of the plot. Horizontal line plotted at $\sqrt{N_b} * l_b$ 148
- C.27 $\sqrt{\langle R^2 \rangle}$ vs Temperature for PIDTBFT-C11-BO with a linear regression
for the linear portion of the plot. Horizontal line plotted at $\sqrt{N_b} * l_b$ 149

NOMENCLATURE

Abbreviations

BHJ Bulk heterojunction

BT Benzothiadiazole

BTA Barbituric acid

BTDAH Barbituric Acid Dihydrate

CG Coarse Grain

CPDT Cyclopentadithiophene

DFT Density Functional Theory

ESP-UA United atom espaloma force field

FBT Fluorobenzothiadiazole

FWHM Full width half mass

GAFF General amber force field

GAFF Generalized amber force field

GGA generalized gradient approximation

GHG Green-house gas

GIXS Grazing incident x-ray scattering

GPU Graphical processing unit

HOMO Highest occupied molecular orbital

IDT Indacenodithiophene

LDA Local density approximation

LJ Lennard-Jones

LUMO Lowest unoccupied molecular orbital

MD Molecular Dynamics

OPLS Optimized Potentials for Liquid Simulations

OPLS-UA United Atom Optimized Potentials for Liquid Simulations

OPV Organic Photovoltaic

P3HT Poly-3-hexylthiophene

PCE Power conversion efficiency

PCPDTFBT Cyclopentadithiophene-co-fluorobenzothiadiazole

PCPDTPT Cyclopentadithiophene-co-pyridal thiadiazole

PIDTBT Indacenodithiophene-co-benzothiadiazole

PIDTCPDT Indacenodithiophene-co-cyclopentadithiophene

PIDTFBT Indacenodithiophene-co-fluorobenzothiadiazole

PT Pyridalthiadiazole

PTFE Polytetrafluoroethylene

PV Photovoltaic

RDF Radial distribution factor

SANS Small angle neutron scattering

THz Terahertz

TPS Time steps per second

UA United atom

VAMH Violuric Acid Monohydrate

Symbols

ΔE Energy Difference

ϵ Depth of the Lennard-Jones potential well

$\langle R^2 \rangle$ Mean squared end-to-end distance

ϕ Dihedral angle

σ Van der Waals radius

$\sqrt{\langle R^2 \rangle}$ Root mean squared end-to-end distance

θ Bond angle

θ_0 Equilibrium bond angle

v Frequency

$d\mu/dQ$ Dipole moment derivative

$E[\Psi_i]$ Total energy of the system

k_h Harmonic bond constant

k_n Spring constant of particle n

k_θ Harmonic angle constant

l_b Average bond length

l_p Persistence length

l_p^{DFT} Persistence length calculated via density functional theory

l_p^{MD} Persistence length calculated via molecular dynamics

l_p^{SANS} Persistence length calculated via small angle neutron scattering

N_b Number of bonds in a polymer

r Bond distance

r_0 Equilibrium bond distance

E_{known} Kinetic, potential, and hartree energy

$E_v[n]$ Ground state energy

E_{XC} Exchange-correlation energy

$F[n(\mathbf{r})]$ Density functional that is independent of the external potential

$F[n]$ Electron density functional

$n(\mathbf{r})$ Electron density

r_{ij} Internuclear distance of two particles (i, j)

$U_{LJ}(r_{ij})$ Potential energy of two particles (i, j)

$v(\mathbf{r})$ External potential

Å Angstrom

Å⁻¹ Inverse angstroms

A Particle size distribution scalar

a,b,c, β Unit Cell Parameters

amu Atomic mass unit

b Exponential factor

cm⁻¹ Inverse centimeters

dt MD simulation timestep length

E Energy

g/mL Grams per mililiter

I Scattering intensity

K Kelvin

kJ Kilojoules

M Mass

N Number of particles

N_{monomers} Number of repeating units in a polymer

s Seconds

T Temperature

U Potential energy

V Volume

° Degrees

CHAPTER 1:

INTRODUCTION

1.1 Polymorphism

Polymorphism is the ability of a molecular solid, composed of the same molecular moieties, to exist in two or more crystalline conformations. These crystal conformations are called polymorphs and will transition from one to another under specific conditions, Figure 1.1. When polymorphism was first reported in 1832 by Friedrich Wöhler and Justus von Liebig, it was believed to affect only large, flexible molecules, such as proteins. However, as studies continued, it was discovered that over 50% of molecules experience polymorphism [8, 9].

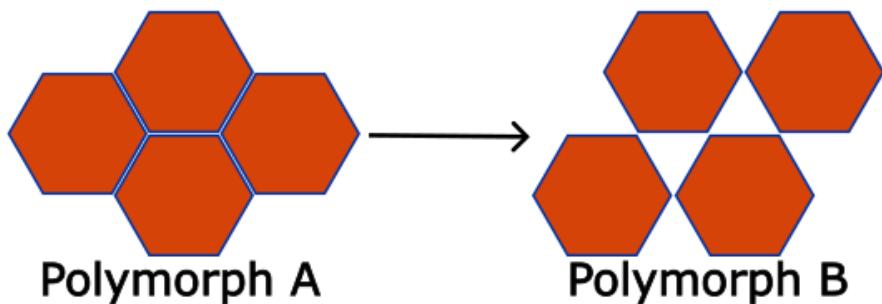


Figure 1.1: Diagram of a polymorphic phase transition from Polymorph A to Polymorph B.

Due to the abundance of polymorphs and the influence of their crystalline structure on the intermolecular interactions present in a material, polymorphs and sol-

vatomorph have become a widely growing area of study in fields of pharmaceutical development, crystal engineering, and de-novo materials design. Solvatomorphs are a specific class of polymorphs in which the solvent molecules crystallize with the solute molecules, see Figure 1.2. The system's intermolecular interactions are responsible for many of the substance's properties, and can cause differences in solubility, vibrational spectra, densities, dissolution rate, melting point, mechanical properties, etc between crystal motifs [10, 11, 12, 9]. Particularly difficult to characterize, hydrates are also a commonly observed branch of crystalline materials. Hydrates are a class of solvatomorphs in which at least one water molecule crystallizes within the unit cell with the solute molecules. These water molecules make studying polymorphism particularly difficult because of the extra degree(s) of freedom the molecules have when crystallizing and transitioning between polymorphs. In this study, we discuss our investigation of the possible polymorphism of barbituric acid dihydrate (BTADH) and violuric acid monohydrate (VAMH) using density functional theory (DFT).

1.2 Density Functional Theory

Density functional theory is a method of analysis that utilizes first principle, quantum mechanics calculations to investigate the electronic structure of materials [13, 14]. Many physical and chemical properties are governed by the material's electronic states. DFT calculations provide accurate information about your system because the calculations take into account quantum mechanical effects, however, this trade-off in accuracy comes at a high computational cost compared to other simulation methods. DFT uses the electron density surrounding a particle to calculate the energy. The particles undergo a slight change in position and the energies are calculated again. This continues until the system's energy is at a minimum.

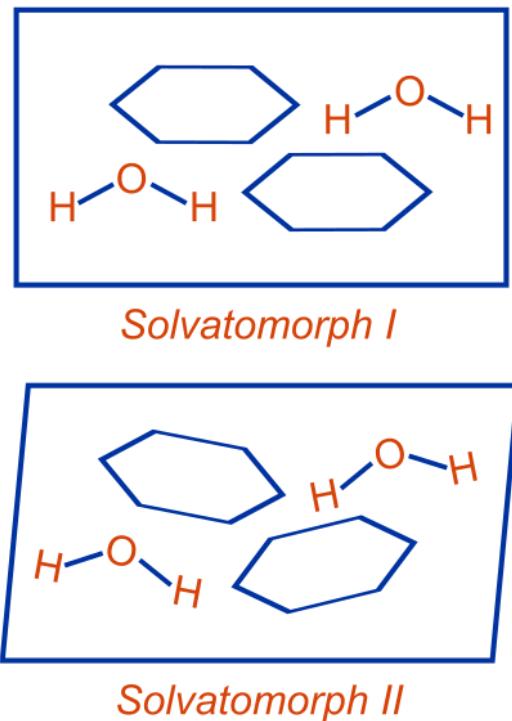


Figure 1.2: Diagram of Solvatomorphs.

DFT was developed upon the Hohenberg-Kohn theorems, the first of which was published in 1964 [15]. The first theorem states that the ground state properties, or energies, of a many body system can be calculated using only a functional ($F[n]$) of the electron density ($n(\mathbf{r})$). This theorem reduced the many body problem of N particles from $3N$ spacial coordinates down to only 3 spacial coordinates. The electron density is defined in terms of the three spacial coordinates, shown in Figure 1.3. While the first theorem describes the possibility of calculating ground state properties from electron density it does not provide any information about the functional required to do so. The second Hohenberg-Kohn theorem states that if the exact functional of a system is known, the electron density that minimizes this functional will correspond

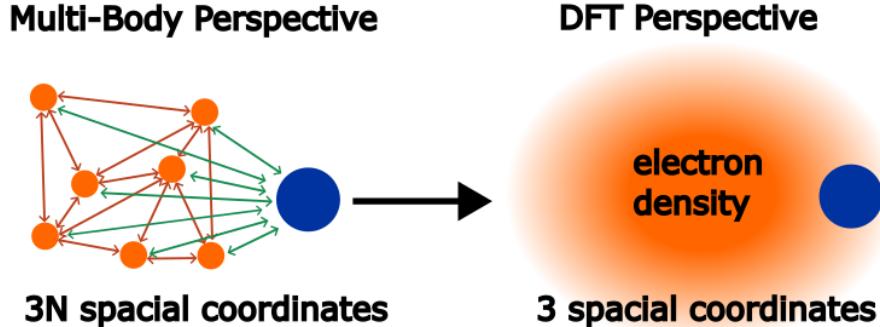


Figure 1.3: A depiction of how density functional theory simplifies the many-body problem into a function of electron density.

to the exact solution of the Schrödinger's equation [16]. Though the only system in which an exact functional has been proven is for a homogeneous electron gas, in all many-body systems the energy functional must be approximated.

1.2.1 Functionals and Basis Sets

Equation 1.1 describes DFT in a mathematical sense, where $E_v[n]$ is equal to the ground state energy of the system and $v(\mathbf{r})$ is the external potential. Equation 1.1 also details that the density functional $F[n(\mathbf{r})]$ is independent of the external potential.

$$E_v[n] \equiv \int v(\mathbf{r})n(\mathbf{r})d\mathbf{r} + F[n(\mathbf{r})] \quad (1.1)$$

We can further examine the total energy of the system using Equation 1.2.

$$E[\{\Psi_i\}] \equiv E_{\text{known}}[\{\Psi_i\}] + E_{XC}[\{\Psi_i\}] \quad (1.2)$$

The E_{known} term includes the kinetic, potential, and hartree energy, all of which can be calculated exactly for a given system. The second term of Equation 1.2 includes the exchange-correlation energy E_{XC} . This energy is unknown for many-body sys-

tems and therefore has to be approximated using density functionals. The exchange-correlation energy can only be calculated for a homogeneous, uniform electron gas, with constant electron density, which is the system that Hohenberg and Kohn used to prove their first theorem, mentioned above [15]. Different systems may require different approximations, depending on the information you are trying to learn from your simulation. This makes choosing your functional one of the most important steps in a DFT investigation. For more complex systems various types of functionals exist, including the local density approximation (LDA), generalized gradient approximation (GGA) and hybrid functionals, among others [17, 18]. LDA functionals are the most simplified models, they take into account only the electron density derived from a free electron gas. GGA functionals build on LDA functionals by also including information about the gradient observed in the electron density. This adds a level of complexity and accuracy for models that do not have uniform electron density. The functionals used in this body of work are hybrid functionals. Hybrid functionals combine the exact electron density of the system, derived by the Hartree-Fock theory, with the E_{XC} from traditional DFT functionals, such as the GGA [19, 20]. Along with choosing the functional, you must also determine a basis set to use in your DFT calculations. The basis set is set of basis functions that reduces calculations from partial derivatives to algebraic equations. Various types of basis sets exist, including plane-wave and atomic orbital. Atomic orbital basis sets place the basis equations on each atom in order to build the atomic orbitals. Chapters 2 and 3 detail how DFT was utilized to determine the crystalline configurations of two hydrate systems, violuric acid monohydrate and barbituric acid dihydrate, over a broad temperature range. Both hydrate systems had been suspected to exhibit polymorphism, but upon

further investigation it was concluded that crystalline disorder was skewing the experimental results and the accurate crystalline configurations were determined using DFT geometry optimization calculations.

1.3 Organic Photovoltaics

Photovoltaic (PV) devices are a promising solution to the rising energy crisis in America [21]. Inorganic photovoltaic cells, often composed of silicon, have a relatively high production and implementation cost[21]. Producing organic photovoltaic devices (OPVs) can be done at a potentially lower cost than silicon because the raw materials require less energy-intensive processing. Small molecule OPVs have recently achieved a photoelectric efficiency of almost 20%, which has increased impressively over recent years [22]. The Shockley-Quiesser limit suggests that there is still room for the power conversion efficiency of organic solar devices to increase by approximately 15%, leading to a PCE of 35% [23]. Introducing morphologies that have a higher charge carrier mobility have been proven to be the key to improving the cell's PCE. Organic materials are typically not thought of as semiconductors, however, it is evident throughout biology that many organic molecules can function as semiconductors. For an organic material to act as a semiconductor the molecules must exhibit significant conjugation [24]. Conjugation is the presence of overlapping pi-orbitals that create a system of delocalized electrons. These delocalized electrons are more readily excitable and are not linked to one specific atom within the molecule. As the conjugation length increases in the material, the band gap decreases, meaning that the electrons require less energy to be excited into the conduction band. Once in the conduction band the electron can contribute to conductivity [25]. In order to excite an electron from the highest occupied molecular orbital (HOMO) to the lowest

unoccupied molecular orbital (LUMO) the incident photon's energy must be equal to or greater than the excitation energy. When an electron is excited into the LUMO, where it has enough energy to move about the surface of the molecule, there is a leftover 'hole' where the electron used to be. In comparison to the negative environment surrounding it, the hole has an effective positive partial charge. The electron and hole have a coulombic attraction due from their opposing charges, resulting in a force that binds the electron and hole together into what is called an exciton [26].

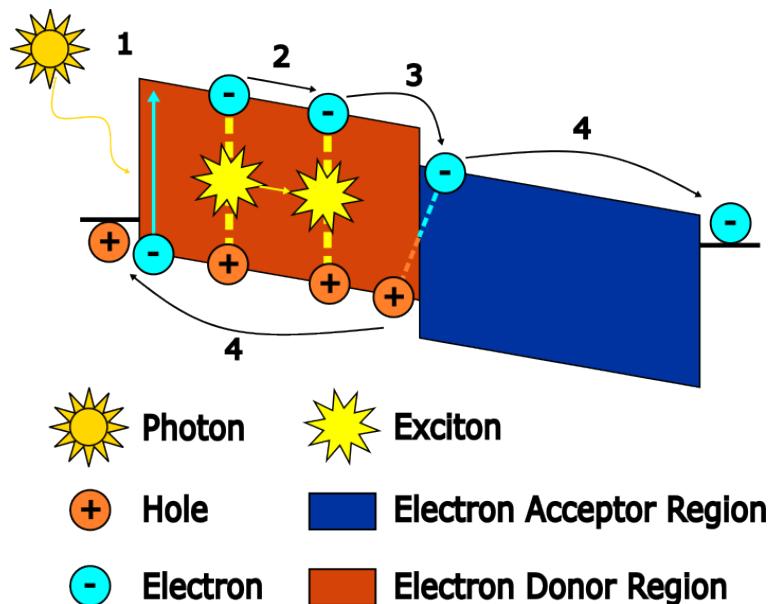


Figure 1.4: Diagram of an OPV device. (1) Photon absorption (2) Exciton creation and migration (3) Exciton dissociation (4) Charge carrier diffusion.

The photoelectric effect, described above, is what makes OPV devices functional. In an OPV device, when an incident photon hits the surface it is passed through into the active layer, where an exciton is created (Figure 1.4, step 1). The exciton diffuses through the active layer until it reaches a grain boundary between an acceptor region and donor region, where exciton dissociation occurs (Figure 1.4, steps 2 & 3).

Once the coulombic connection between the electron and hole is broken the electron diffuses through the acceptor region to be harvested at the cathode and the hole through the donor region to be harvested at the anode (Figure 1.4, step 4) [27].

The active layer motif depicted in Figure 1.5 is called a bulk heterojunction active layer [28]. These donor/acceptor regions can come from two different organic semiconductor molecules or a co-polymer with differing acceptor monomers and donor monomers [29]. The donor/acceptor regions need to exhibit a significant difference in energy levels in order to cause exciton dissociation when the exciton crosses the grain boundary. Once the excitons are dissociated the charge carriers have to pass through the active layer to their respective anodes [30]. If there is not a clear path for the electron or hole to travel to electrodes these charges are typically lost to exciton recombination. This is why the morphology of the active layer is so important to the power conversion efficiency.

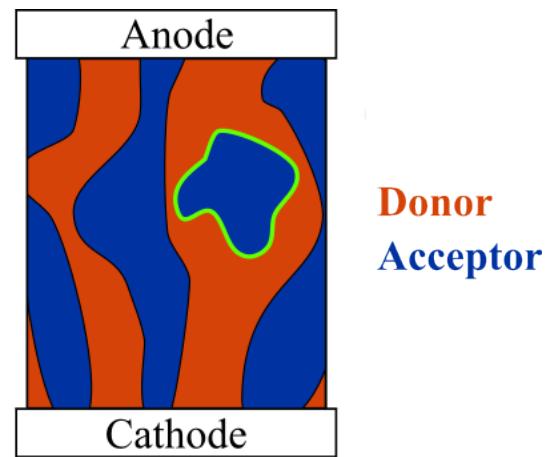


Figure 1.5: Depiction of a bulk heterojunction (BHJ) active layer of an organic photovoltaic device. Outlined in green is a an "island" where the electron would have no access to the electrode, forcing it to recombine with a hole, rendering it useless.

Understanding the influence of structural conformation on OPV device performance brings us to the importance of density functional theory calculations and molecular dynamics simulations. DFT has many applications, one of which is geom-

etry optimizations of small molecule systems. From these geometry optimizations we can determine the most energetically favorable conformation of our system. Molecular dynamics (MD) simulations (as discussed below) are an efficient way to determine the morphology of the active layer and the properties associated with it [31]. MD simulations are used for investigating the structures of OPV active layers because they follow the self-assembly of the copolymers in a bulk heterojunction. MD simulations give us information about the thermodynamics of the self-assembly as well as the kinetics of how these molecules interact to form the equilibrium structures. MD simulations also have the benefit of simulating large molecule systems at a lower computational cost than DFT. The information gained from DFT calculations in conjunction with MD simulations can give great insight into how our copolymers will self-assemble during in the active layer during device production [32].

1.4 Molecular Dynamics

Molecular dynamics is a computational simulation method used to predict equilibrium morphologies of many particle systems. Newton's equations of motion are solved numerically to generate trajectories of how molecular systems evolve over time [33]. The basic algorithm behind an MD simulation is: (1) calculate potentials and forces from particle positions. (2) Update particle velocities using the forces calculated in (1). (3) Update particle positions from the velocities calculated in (2). (4) Continue (1)-(3) with updated particle positions until equilibrium is achieved. Once equilibrium is achieved we calculate the decorrelation time (time steps between independent samples) and calculate system properties by averaging independent microstates.

The forces of attraction and repulsion enacted on one particle by another are described by the particles' pair potential. The pair potential is essentially the po-

tential energy of two interacting particles. A common pair potential used in MD simulations is the Lennard-Jones (LJ) potential [34, 35]. The LJ-potential, depicted in Figure 1.6 and explained mathematically in Equation 1.3, describes the potential energy $U_{LJ}(r_{i,j})$ of two particles, i and j , as a function of internuclear distance, $r_{i,j}$.

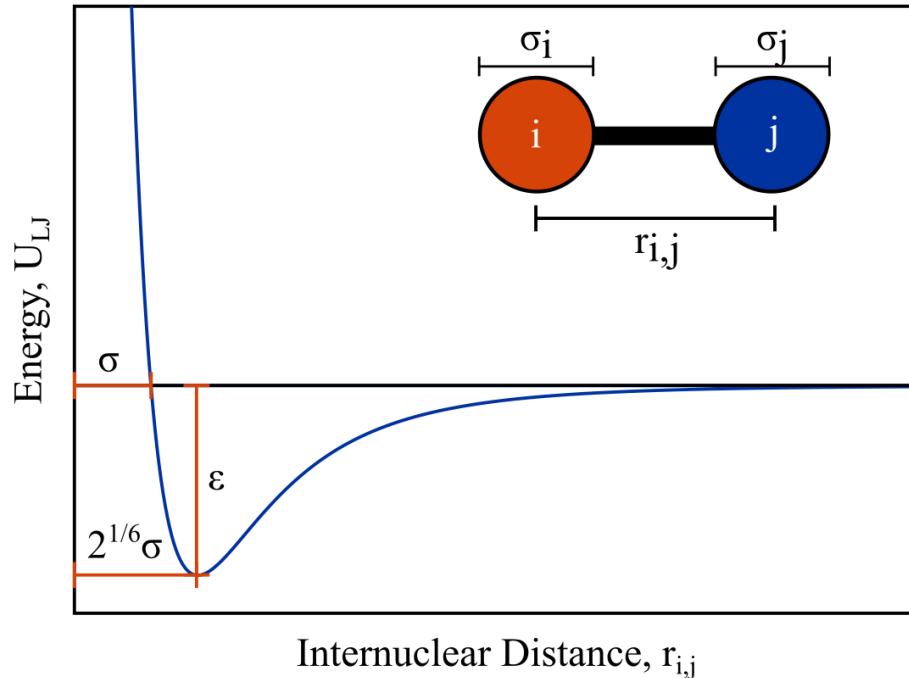


Figure 1.6: Plot of the Lennard-Jones potential energy.

$$U_{LJ}(r_{ij}) = 4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right] \quad (1.3)$$

Pair interactions are governed by σ and ϵ . σ is the internuclear distance at which the potential energy between two particles is zero and epsilon is the depth of the well in the potential energy plot. ϵ is a measure of how attractive the particle is and σ can be thought of as the diameter of the particle. As shown in Figure 1.6, the potential energy reaches a minimum when the particles are at a distance of $2^{1/6}\sigma$ apart. At

distances closer than $2^{1/6}\sigma$, the particles repel one another and at distances greater than $2^{1/6}\sigma$ the forces felt are negative, which accelerates the particles toward one another. The greater the distance gets from $2^{1/6}\sigma$, the lesser the effect the attractive force has on the other particle. In this study, we used the Lorentzian combination rule for sigma and a geometric combining rule for epsilon, shown in Equation 1.4.

$$\sigma_{i,j} = \frac{\sigma_i + \sigma_j}{2}, \epsilon_{i,j} = \sqrt{\epsilon_i * \epsilon_j} \quad (1.4)$$

Other properties, such as bond, angle and dihedral, also dictate how particles interact and how a system evolves over time. In MD simulations bonds and bond angles are described as harmonic oscillators, using Equation 1.5, where $U(r)$ is the potential energy of the bond, r is the internuclear distance between the bonded particles, r_0 is the equilibrium internuclear distance and k_h is the bond's harmonic spring constant. A diagram depicting the bonds, angles, and dihedral angles is shown in Figure 1.7.

$$U(r) = \frac{1}{2}k_h(r - r_0) \quad (1.5)$$

Equation 1.6 described the potential energy $U(\theta)$ associated with the bond angle, θ , where θ_0 is the equilibrium bond angle and k_θ is the harmonic angle constant.

$$U(\theta) = \frac{1}{2}k_\theta(\theta - \theta_0) \quad (1.6)$$

Dihedrals are described using Equation 1.7, where k_n is the spring constant of the corresponding particle, ϕ is the angle between atoms planes within the dihedral, and $U(\phi)$ is the potential energy of the dihedral.

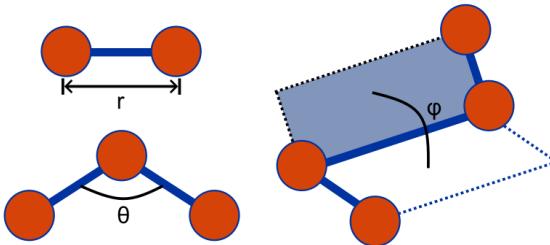


Figure 1.7: Illustration of bond distance (r), bond angle (θ), and dihedral angle (ϕ)

$$U(\phi) = \frac{1}{2}k_1(1+\cos(\phi)) + \frac{1}{2}k_2(1+\cos(2\phi)) + \frac{1}{2}k_3(1+\cos(3\phi)) + \frac{1}{2}k_4(1+\cos(4\phi)) \quad (1.7)$$

In our simulation framework we try to follow the TRUE model, ensuring that our simulations are transparent, reproducible, usable and extensible. In MD simulations a forcefield is used to efficiently store and keep track of all the variables used in the pair, angle, bond, harmonic potentials mentioned above [36, 37]. Each unique pair interaction requires its own unique set of parameters to model the pair of particles using the equations listed above. The process of creating and applying a forcefield is a common barrier to the transparency and reproducibility of our TRUE simulations [38]. It is frequent in initializing a simulation of a new molecule to have missing forcefield parameters. In these cases the parameters must be either added to an existing forcefield file or a new forcefield file must be produced. In this study, we have developed a workflow that implements an open source software to create a forcefield file from a molecule graph [39]. We have validated this workflow for complex, conjugated polymers by comparing simulation results of two molecules to previously published results (Chapter 4). After validating this workflow it was utilized to simulate 13

different complex polymers (Chapter 5).

In comparison to MD simulations, DFT calculations come with a much higher computational cost. Due to this, larger systems that can be equilibrated using MD cannot be simulated with DFT. On the other hand, however, MD simulations often don't provide as much information about a system as DFT calculations do. This is due to the neglect of quantum effects in MD simulations. Due to these differences, in this dissertation we utilize DFT to predict the crystalline conformation of small molecules like BTADH and VAMH, but employ MD simulations to determine how macromolecules and polymers interact to self-assemble into bulk morphologies. In both cases, however, molecular simulations are used over experiment due to experimental constraints, such as difficulty in structure determination and experimental costs and time. The results from the molecular simulations are then used to direct experimental research toward materials of interest for the given application.

CHAPTER 2:

TRUE POLYMORPHIC PHASE TRANSITION OR DYNAMIC CRYSTAL DISORDER? AN INVESTIGATION INTO THE UNUSUAL PHASE BEHAVIOR OF BARBITURIC ACID DIHYDRATE

The following chapter contains published work written by me with the guidance of Dr. Matthew King. This work is published at Ref. [14].

2.1 Abstract

Crystalline barbituric acid dihydrate (BTADH) was previously thought to undergo a subtle temperature-dependent phase transition from a high-temperature orthorhombic *Pnma* space group to a non-merohedrally twinned monoclinic *P2₁/n* phase below a transition temperature of 217 K. Questions remain on the true nature of the unusual transformation and the structure of the high-temperature crystal phase due to difficulties in resolving the *Pnma* structure from X-ray diffraction data and because of the subtlety of the structural transition. In this study, terahertz (THz) spectroscopy and solid-state density functional theory (DFT) were utilized to explore this sus-

pected phase transition and uncover the principle physical mechanisms contributing to this anomalous transition. Our findings suggest that at temperatures above the previously reported phase transition temperature of 217 K, the crystal does not exist in the *Pnma* space group configuration. However, two equivalent favorable energetic states with *P2₁/n* symmetry related by the twinning plane lie on either side of the proposed *Pnma* structure. It is these same local potential energy minima that guide the crystal system to a non-merohedrally twinned configuration identified in the low-temperature crystal structures. At temperatures above 217 K, the system possesses the thermal energy necessary to readily transition between these degenerate states separated by the low-lying *Pnma* structural barrier. The evidence acquired by THz spectroscopy and DFT indicate the absence of a true temperature-dependent phase change, and rather a system that exists in a tenuous thermally disordered state above the previously alleged transition temperature.

2.2 Introduction

The crystalline conformation of compounds can have a large influence on material properties, such as solubility, dissolution rate, melting point, and mechanical properties, and these can vary significantly between polymorphs of a compound [10, 11, 12]. As polymorphism has a profound influence on such physicochemical properties, the study of the growth and transformation of polymorphs and solvatomorphs have become a quite impactful area of interest for pharmaceutical development, crystal engineering, and de novo materials design[40, 41, 42, 43, 44, 45, 46]. However, detailed experimental and computational analyses of complex target systems can often be difficult. In ongoing efforts to improve computational methodologies to describe the behavioral intricacies of crystal systems, it is useful to accompany such approaches

with experimental spectroscopic analyses in the study of polymorphism, phase transitions, and related phenomena. In this study, solid-state density functional theory (DFT) and terahertz (THz) spectroscopy were used to explore the underlying physical mechanisms leading to a subtle temperature-dependent phase transition exhibited by the barbituric acid dihydrate (BTADH) crystal system. THz spectroscopy has been shown to be a powerful tool for the investigation of molecular crystals and polymorphism since THz radiation can be used to probe phonon vibrations sensitive to the specific intermolecular interactions present in a particular molecular packing configuration [47, 48, 49, 50, 51]. In combination with DFT calculations, valuable information can be obtained regarding structure, energetics, and vibrational properties [52, 53]. Crystalline BTADH is reported to undergo a transition from a high-temperature orthorhombic *Pnma* phase to a non-merohedrally twinned monoclinic *P2₁/n* phase below a transition temperature of 217 K (Figure 2.1). The BTADH system has been extensively studied by X-ray diffraction, with most attempts at resolving the high-temperature phase resulting in the investigators reluctantly coming to the same conclusions. The most thorough study was performed by Nichol and Clegg in 2005, whom measured the BTADH system at 14 temperatures ranging from 100–270 K, with meticulous focus on temperatures surrounding the suspected transition temperature reported at 216–217 K [1]. Most unusual is the subtlety of the phase transformation undergoing no changes in hydrogen bonding motifs upon the *Pnma* to *P2₁/n* transition, only an out-of-plane rotation of the barbituric acid molecules and modest displacements of the two H₂O molecules and the sp³-hybridized carbon of barbituric acid. The authors ultimately determined a high-temperature *Pnma* phase, which was in line with two preceding studies also proposing the *Pnma* space group at room tem-

perature [54, 55]. These studies reported quite large atomic displacements along the b-axis, perpendicular to the mirror plane containing all atoms, besides the mirrored sp₃ hydrogens. Furthermore, Al-Karaghouli et al. studied the BTADH system by neutron diffraction and proposed that the crystal may take a Pn21a configuration or that the H₂O oxygen atoms may not lie within the mirror plane [55]. However, neither proposition gave satisfactory results, leading both studies to settle on the *Pnma* space group assignment. All previous reports do agree that the BTADH crystal system at temperatures below 217 K adopts the *P2₁/n* space group and exhibits non-merohedral twinning relating equivalent *P2₁/n* unit cells by a two-fold rotation about the b-axis.

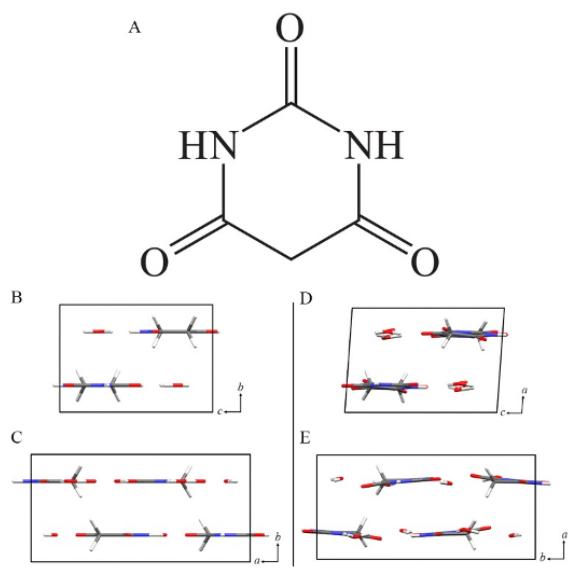


Figure 2.1: A) Molecular structure of barbituric acid. Comparison of the crystalline structures of the temperature-dependent BTADH polymorphs, with the high-temperature *Pnma* form viewed along the B) a-axis and C) c-axis, and the low-temperature *P2₁/n* form along the D) b-axis and E) c-axis.

2.3 Experimental Section

2.3.1 THz Spectroscopy

Anhydrous barbituric acid (BTA) was obtained from Sigma-Aldrich (99% purity). BTADH was crystallized from a saturated aqueous solution by slow evaporation at room temperature over a period of one week under darkness. Samples for THz measurements were mixed with polytetrafluoroethylene (PTFE) powder at concentrations of approximately 4.5% by mass. The samples were pulverized using a stainless steel amalgamator (Dentsply Rinn 3110-3A) to minimize particle size and scattering of THz radiation. Approximately 0.50 g of the sample mixtures were pressed into pellets under a measured pressure of 2000 psi using a hydraulic press (ICL EZ-Press 12) equipped with a 13 mm stainless steel die. PTFE pellets for use as reference blanks were prepared in the same manner. THz spectra were recorded using a custom time-domain pulsed THz spectrometer based on an amplified Ti:sapphire laser system (Coherent, Inc., Santa Clara, CA), described in detail elsewhere [56]. Data were acquired at 293 and 78 K with samples under vacuum. Samples and blanks were scanned 32 times and data were averaged for each individual data set. Each 32-ps scan window consisted of 3200 data points representing the THz waveform, producing an effective instrument resolution was approximately 1.0 cm^{-1} . The ratio of the power spectra obtained from the Fourier-transformed data sets of the sample and blank produced the THz absorption spectrum. Each THz spectrum presented in this work is the average of four individual THz spectra.

2.3.2 Solid-State Density Functional Theory

All calculations were performed using the Crystal14 software package [57, 58]. Solid-state DFT calculations utilized the PBE,[59] M06,[60] and B3LYP [61, 62] functionals with atom-centered cc-pVDZ,[63] 6-31G(d,p),[64, 65] 6-311G(d,p),[66] and pob-TZVP [67] basis sets. Initial lattice dimensions and atomic coordinates were acquired from X-ray crystallographic data from Cambridge Crystallographic Database[1, 68]. The sampling rate as a function of k points used in defining the real space density matrix and Monkhorst grid in reciprocal space was determined for total energy convergence criteria of $\Delta E < 10^{-8}$ hartree for geometry optimizations and $\Delta E < 10^{-11}$ hartree for normal-mode calculations [69, 70]. In order to meet these convergence criteria, shrinking factors of 6 where used for the BTADH crystal calculations. The total number of grid points varied between calculations depending on initial crystal geometry. The radial and angular distributions of points were defined by a pruned (75,974) integration grid for each calculation. A pair-wise (D2) damped empirical energy correction for London-type dispersion forces as proposed by Grimme et al [71, 72]. was included in all DFT calculations. The global scaling factors, s6, used for the dispersion correction term were 0.75, 0.25, and 1.05 for PBE, M06, and B3LYP, respectively [60, 71, 72, 73]. Additional dispersion parameters, including C6 coefficients, van der Waals radii, cut-off distance, and dampening function steepness, were taken from Grimme et al [71, 72]. Modifications to van der Waals radii were made in accordance with Civalleri et al. [74]. for the application of D2 dispersion corrections in molecular crystals. The truncation tolerances for the Coulomb and HF exchange integral series are defined by the program keyword TOLINTEG, and were set to values of 10^{-8} , 10^{-8} , 10^{-8} , 10^{-8} , and 10^{-16} hartree. Frequencies of normal modes were calculated within

the harmonic approximation by numerical differentiation of the analytical gradient of the potential energy with respect to atomic position[75]. The IR intensities for normal modes were calculated from the dipole moment derivatives ($d\mu/dQ$) determined by the Berry phase approach of calculating Born charges as polarization differences between equilibrium and distorted geometries [76].

2.4 Results and Discussion

An initial screening of density functional and basis set combinations was performed to determine those best suited to reproduce the experimental X-ray crystal structures of BTADH by solid-state DFT. The PBE, M06, and B3LYP density functionals were employed, representing GGA, meta-GGA, and hybrid classes of functionals, respectively. Each density functional was combined with two different commonly used double- ζ basis sets – 6-31G(d,p) and cc-pVDZ – and two triple- ζ basis sets – 6-311G(d,p) and pob-TZVP. Geometry optimizations were performed in both fixed lattice dimensions of the 100-K structure and full-cell optimizations allowing lattice parameters to freely optimize within the constraints of space group symmetry. The 100-K X-ray crystal structure in $P2_1/n$ was used as the starting point for all calculations. $Pnma$ unit cells of the high-temperature crystal phase were not used for this initial evaluation as the crystal structures in $Pnma$ symmetry confine the BTA and H₂O molecular orientations to planar configurations, thus no valuable information into functional/basis set performance can be obtained regarding intermolecular interactions from such calculations. The quality of structural reproduction was based on the notable structural variations presented through the suspected temperature-dependent phase transition using the series of BTADH crystal structures over the temperature range of 100–270 K reported in Ref. 18. For the geometry optimizations in fixed lattice dimensions,

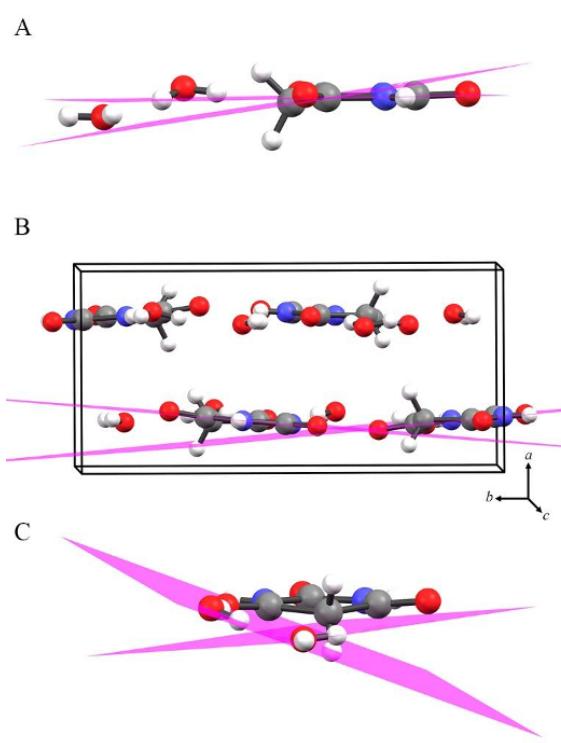


Figure 2.2: Plane orientations relating angle metrics for determining the quality of unit cell reproductions. A) BTADH molecular plane defined by ring carbonyl carbons, and that representing out-of-plane position of the ring CH₂ group. B) Opposing molecular planes of two symmetry-related BTADH molecules in the low-temperature *P2₁/n* polymorph. C) Planes each created by a single H₂O molecule of the dihydrate asymmetric unit.

the angle between molecular planes between BTA molecules within a unit cell, the deviation of the CH₂ group from the BTA molecular plane, and the angle of planes formed by each of the two water molecules in the dihydrate asymmetric unit were used as metrics for structural reproduction (Figure 2.2). For full-geometry optimizations, lattice parameters *a*, *b*, *c*, and β were evaluated in addition to the aforementioned conformational metrics. These data for the experimental X-ray structures are provided in Table 2.1.

It was determined from the extensive series of DFT structural calculations that the

Table 2.1: Lattice parameters and calculated relative molecular orientations from BTADH experimental X-ray crystal structures reported in Ref. [1].

Experimental X-ray										
T (K)	Group	Space	a (Å)	b (Å)	c (Å)	β (°)	V (Å ³)	CH ₂ -Mol (°) ^a	Mol-Mol (°) ^b	H ₂ O-H ₂ O (°) ^c
270	<i>Pnma</i>		6.2144	12.7512	8.8841	90	703.99	—	—	—
230	<i>Pnma</i>		6.1739	12.7594	8.8831	90	699.77	—	—	—
220	<i>Pnma</i>		6.1665	12.7630	8.8814	90	699.00	—	—	—
219	<i>Pnma</i>		6.1624	12.7570	8.8780	90	697.90	—	—	—
218	<i>Pnma</i>		6.1626	12.7570	8.8760	90	697.80	—	—	—
217	<i>Pnma</i>		6.1770	12.7850	8.8980	90	702.70	—	—	—
216	<i>P2₁/n</i>		6.1567	12.7330	8.8650	91.180	694.80	1.85	2.98	28.87
215	<i>P2₁/n</i>		6.1580	12.7514	8.8763	91.263	696.83	2.47	2.97	8.85
210	<i>P2₁/n</i>		6.1538	12.7470	8.8770	91.627	696.00	2.85	3.6	20.31
200	<i>P2₁/n</i>		6.1313	12.7030	8.8456	92.187	688.50	3.72	5.11	26.94
190	<i>P2₁/n</i>		6.1377	12.7306	8.8641	92.528	691.94	4.81	5.9	16.72
170	<i>P2₁/n</i>		6.1270	12.7253	8.8633	93.068	690.06	5.8	7.18	20.06
150	<i>P2₁/n</i>		6.1130	12.7149	8.8564	93.437	687.14	6.7	8.22	25.09
100	<i>P2₁/n</i>		6.0970	12.7152	8.8587	94.051	685.05	8.14	9.96	27.36
DFT Full-Cell Geometry Optimizations in P2₁/n										
		a (Å)	b (Å)	c (Å)	β (°)	V (Å ³)	CH ₂ -Mol (°) ^a	Mol-Mol (°) ^b	H ₂ O-H ₂ O (°) ^c	
PBE/6-311G(d,p)		5.9263	12.6285	8.7625	94.455	653.81	8.76	10.53	24.04	
PBE/pob-TZVP		6.0755	12.4857	8.6603	94.438	654.97	11.39	15.53	40.25	

^aCH₂-Molecule planes (Fig. 2A); ^bMolecule-molecule planes (Fig. 2B); ^cWater-water planes (Fig. 2C).

PBE density function in conjunction with the tested triple- ζ basis sets, 6-311G(d,p) or pob-TZVP, were best suited for the computational treatment of the BTADH crystal system. The results of the computational screening are provided as **Supporting Information**. All subsequent calculations in this study utilized PBE/6-311G(d,p) and PBE/pob-TZVP. Results of the full unit cell geometry optimization using these functional/basis set combinations are also provided in **Table 2.1**. THz spectra of BTADH in the range of 20–95 cm⁻¹ were obtained at 293 K and 78 K (Figure 2.3A). Notable differences are present between the spectra taken at the two temperatures. At room-temperature, there is an absence of definitive spectral features below the one resolved peak at 88.8 cm⁻¹. Upon cooling to 78 K, six additional features emerge in the THz spectrum. Most conspicuous is the relatively narrow peak appearing at 33.6 cm⁻¹ for which no evidence of any corresponding absorption intensity is present at 293 K. Large changes between THz spectra collected at different temperatures are usually indicative of a phase transformation leading to the activation/deactivation

of phonon modes corresponding to the distinctive crystalline phases. In this case, however, there is not clear evidence of a phase transition between 293 and 78 K as many similarities remain between the two spectra, although many modes seem to be washed out at room temperature. There is an indication of underlying modes across the 293 K spectrum as broad absorptions resulting in a mostly featureless background that become resolved at low temperature. There is also a clear shift in the 88.8 cm^{-1} feature at 293 K to higher frequency upon cooling, which is typically observed temperature-dependent behavior for phonon modes.

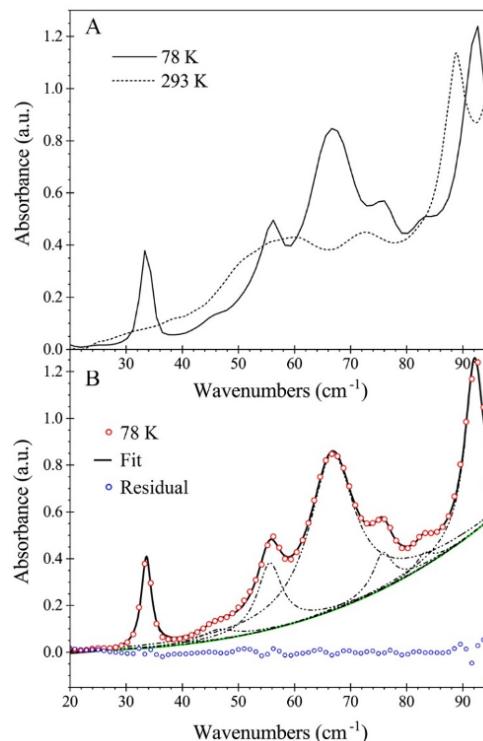


Figure 2.3: A) THz vibrational spectra of BTADH in the spectral range of $20\text{-}95\text{ cm}^{-1}$ obtained at 293 K and 78 K. B) Spectral fitting of the 78 K THz spectrum (red circles) with Lorentzian peaks (dashed lines) on exponential background (green line). The residuals of the experimental data and fitted values are indicated by blue circles.

Another peculiar observance in the BTADH THz spectra collected at 78 K are the extremely broad symmetric absorptions between 40 and 80 cm⁻¹, and the discrepancy between linewidths of all present spectral features. A multiple peak fitting procedure was performed to determine the FWHM of the absorptions. Lorentzian line shapes for the seven distinguishable features were added to an exponential function representing the rising background, which is a product of Mie and Rayleigh scattering of THz radiation from the granular polycrystalline sample [77]. The Lorentzian peaks and background were simultaneously fit to the experimental spectrum (Figure 2.3B). Peak centers and FWHMs are provided in Table 2.2. The exponential function of form $I = Avb$, where I is the scattering intensity, A is the scalar accounting for particle size distribution, and b is the exponential factor influencing the frequency-dependence of scattering, yielded optimal values of $A = 1.7 \times 10^{-7}$ and $b = 3.34$. This value of b is in accordance with previously reported values using the same exponential form to correct for scattering effects observed in THz spectra [50, 77].

Table 2.2: Central frequencies and linewidths of THz spectral features in the range of 20 – 95 cm⁻¹ for data collected at 78 K.

Peak Center (cm ⁻¹)	FWHM (cm ⁻¹)
33.6	2.2
46.3	7.3
55.6	5.4
66.7	9.7
75.7	4.3
83.1	3.6
92.1	4.0

The spectral fitting routine resulted in a high-quality fit to the THz spectrum. Large deviations in FWHMs were determined for the observed peaks, ranging from 2.2 cm⁻¹ for the mode centered at 33.6 cm⁻¹ to a FWHM of 9.7 cm⁻¹ for the 66.7

cm^{-1} feature. Peak symmetries and the quality of fit using single Lorentzian functions suggests that all seven resolvable features, including the broad absorption at 66.7 cm^{-1} , are due to single vibrational mode and not the product of contributions from multiple underlying modes. Such large discrepancies in peak widths are not typically observed in THz spectra of organic materials, although phonon mode softening along phase transition coordinates and a corresponding reduction in vibrational lifetimes can result in homogeneous broadening of peaks [78, 79]. Changes in peak width are typically observed as the system nears a phase transition temperature [80]. However, this would not be expected for BTADH in the THz spectrum collected at 78 K, as this temperature is well below the suspected phase transition at 217 K. Phonon vibrational frequencies were calculated by solid-state DFT to provide information on mode characters to aid in explanation of the unusual mode behavior exhibited by BTADH. Both triple- ζ basis sets, pob-TZVP and 6-311G(d,p), that best performed for the initial structural evaluations were applied with the PBE density functional for normal mode calculations. The high-temperature crystal phase in *Pnma* symmetry produced three imaginary vibrational frequencies in calculations utilizing either basis set. Examination of the atom displacement eigenvectors for each imaginary mode revealed that the most atomic motions are normal to the reflection plane imposed by the *Pnma* space group for all atoms confined by this symmetry operator (Figure 2.4). Scanning of the potential energy surface along the imaginary modes produces a symmetric double-well potential with a low lying energy barrier (Figure 2.4A). Displacement of BTADH to either side of this double-well potential places the crystal into corresponding *P2₁/n* structures related by a two-fold rotation about the b-axis, which is equivalent to the low-temperature phase.

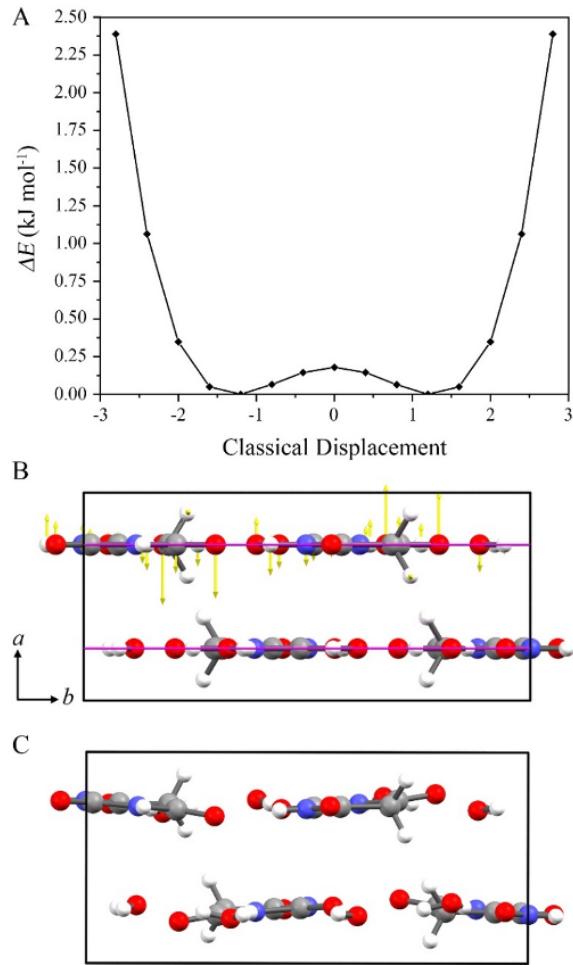


Figure 2.4: A) Scan of potential energy surface along the -35.79 cm^{-1} imaginary phonon mode (PBE/pob-TZVP) of BTADH confined to the $Pnma$ space group at 270-K lattice parameters. B) Displacement vectors (yellow arrows, magnitudes scaled $\times 10$) for the imaginary mode shown for top two BTADH units. Displacements are oriented normal to the reflection plane (purple line) confining the molecules in $Pnma$ symmetry. C) Molecular orientation of BTADH optimized in $P2_1/n$ symmetry.

Starting from the planar 270-K *Pnma* crystal structure, symmetry constraints were removed and the geometry was re-optimized within fixed experimental X-ray lattice dimensions using PBE/pob-TZVP. The unit cell symmetry was relaxed to *P2₁/n*, analogous to the suspected low-temperature phase, thus removing the reflection planes that confine the atoms in the *Pnma* space group. From the optimized geometry, the measured molecular plane angles, as described in Figure 2.2 and Table 2.1, were 7.93°, 11.79°, and 36.82° corresponding to CH₂–molecule, molecule–molecule, and H₂O–H₂O planes, respectively. These angles are remarkably similar to those observed for the X-ray crystal structures collected at low temperature. The relaxed-symmetry structural optimizations were repeated for each of the crystal dimensions measured at the 14 different temperatures in the range of 100–270 K using both the pob-TZVP and 6-311G(d,p) basis sets (Table 2.3). Analysis of the optimized geometries revealed little variation in measured angles over the temperature range. There was a general trend of angles increasing with decreasing temperature, but with no significant variations occurring between adjacent temperature steps. An additional series of calculations were performed using the lattice dimensions of the suspected *Pnma* phase, i.e. at temperatures above 217 K. The crystal structures were restricted to either *P2₁/n* or *Pnma* symmetry to evaluate differences in unit cell energies between the two crystal settings. In both basis sets, the *P2₁/n* symmetry produced the lower energy unit cell for all temperature lattice parameters (Table 2.4). The difference in energy between the *P2₁/n* or *Pnma* space groups were small in each case, ranging from -0.985 to -1.379 kJ mol⁻¹ in the pob-TZVP basis, and -0.804 to -1.234 kJ mol⁻¹ in 6-311G(d,p). These results show a slight energetic preference to *P2₁/n* symmetry in unit cell dimensions at all temperatures above 216 K.

Table 2.3: Angles between the molecular planes of two molecules within the packed unit cell, angles between the two water molecules within the unit cell and the angles between the CH₂ plane and the molecular plane of one molecule.

T (K)	CH ₂ -Molecule ^a (°)			Molecule-Molecule ^b (°)			H ₂ O-H ₂ O ^c (°)		
	pob-TZVP	6-311G(d,p)	exp.	pob-TZVP	6-311G(d,p)	exp.	pob-TZVP	6-311G(d,p)	exp.
270	7.93	8.40	—	11.79	11.27	—	36.82	35.86	—
230	7.48	7.80	—	11.19	10.55	—	34.67	33.19	—
220	7.39	7.67	—	11.07	10.41	—	34.27	32.70	—
219	7.38	7.66	—	11.12	10.43	—	34.22	32.75	—
218	7.39	7.67	—	11.14	10.45	—	34.29	32.82	—
217	7.39	7.57	—	10.88	10.19	—	34.09	32.26	—
216	8.35	8.70	1.85	11.77	11.20	2.98	35.07	33.59	28.87
215	8.34	8.63	2.47	11.57	10.95	2.97	34.65	32.86	8.85
210	8.62	8.85	2.85	11.71	11.06	3.60	34.67	32.66	20.31
200	9.12	9.39	3.72	12.25	11.70	5.11	35.20	33.53	26.94
190	9.35	9.50	4.81	12.08	11.46	5.90	34.87	32.66	16.72
170	9.65	9.79	5.80	12.17	11.56	7.18	34.53	32.07	20.06
150	9.83	9.94	6.70	12.20	11.63	8.22	34.16	31.62	25.09
100	10.05	10.02	8.14	12.06	11.44	9.96	33.22	30.19	27.36

^{a,b,c}Refer to Fig. 2 for plane definitions: ^aFig. 2A; ^bFig. 2B; ^cFig. 2C.

THz vibrational frequencies were calculated for the low-temperature *P*2₁/*n* phase using the fully relaxed unit cell structures (lattice dimensions provided in Table 2.1) in PBE/pob-TZVP and PBE/6-311G(d,p) (Figure 2.5). The computed mode frequencies were mostly overestimated in comparison to the THz spectrum obtained at 78 K (Table 2.5). This is typical in such calculations as the lattice dimensions used for normal mode calculations are contracted from the actual experimental dimensions due to neglect of thermal expansion in the DFT-optimized unit cells and contributions from basis set superposition error [81, 82]. The result of constricting the unit cell dimensions is the general blue-shifting of vibrational frequencies as extensively shown in previous reports. In addition, vibrational anharmonicity of phonon modes contributes to overestimation of normal modes calculated in a harmonic limit [47, 83, 84]. Frequencies were scaled by 0.80 in the spectral overlays provided in Figure 2.5 to aid in visual comparison. Lorentzian lineshapes were convolved into calculated modes using the FWHM of the 33.6 cm⁻¹ spectral feature to emphasize the large differences in linewidths observed in the THz spectrum. From the spectral

T(K) ^a	$\Delta E/\text{pob-TZVP}$	$\Delta E/6-311G(d,p)$
270	-1.379	-1.234
230	-1.116	-0.893
220	-1.050	-0.844
219	-1.050	-0.860
218	-1.050	-0.866
217	-0.985	-0.804

^aLattice parameters for unit cells at variable temperatures are listed in Table 2.1.

Table 2.4: Calculated energies (per BTADH·2H₂O) of BTADH confined to P2₁/n and *Pnma* symmetry groups at experimental lattice parameters. ΔE provided in units of kJ mol⁻¹.

simulations, the vibrational modes contributing to the observed THz absorptions can be assigned. A notable discrepancy between the phonon frequencies calculated by the two basis sets is the underestimation of the mode corresponding to the 33.6 cm⁻¹ peak when using the 6-311G(d,p) basis set.

Table 2.5: Calculated phonon mode frequencies and intensities of full-relaxed BTADH unit cell in *P2₁/n* symmetry using the PBE functional with the pob-TZVP and 6-311G(d,p) basis sets.

pob-TZVP	Int ^a	6-311G(d,p)	Int
43.46 ^b	5.14	30.14	3.16
54.47	0.01	44.05	0.13
61.43	0.04	58.89	0.00
67.17	1.86	65.13	0.69
78.70	0.40	72.49	0.13
85.07	10.35	80.62	9.81
103.23	0.00	82.94	35.50
107.29	15.68	96.21	2.57
121.83	28.40	110.36	27.21
122.75	4.41	115.01	1.97

^aIntensity (km mol⁻¹); ^bFrequency (cm⁻¹)

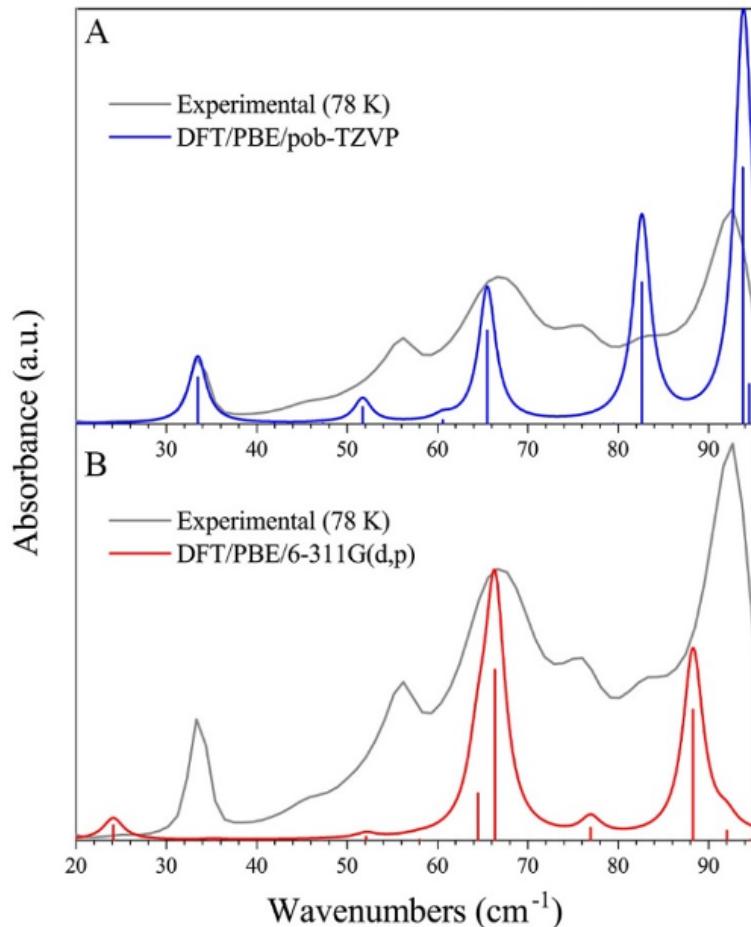


Figure 2.5: A) Mode character of the spectral feature at 33.6 cm^{-1} observed in the 78 K THz spectrum, represented by calculated displacement vectors (yellow arrows, magnitudes scaled $\times 10$). B) Calculated mode displacements of the broad feature centered at 66.7 cm^{-1} in the 78 K THz spectrum. C) Mode character of the distinct 88.8 cm^{-1} vibrational mode observed in the 293 K THz spectrum (corresponding to the feature located at 92.7 cm^{-1} in the 78 K spectrum).

Of primary interest to explain the unexpected behavior of BTADH are the mode characters of the THz absorptions at 33.6, 66.7, and 92.1 cm⁻¹. The respective assignments of these modes were made by inspection of the spectral overlays, and are the 43.46, 85.07, and 121.83 cm⁻¹ modes for PBE/pob-TZVP and the 30.14, 82.94, and 110.36 cm⁻¹ modes for PBE/6-311G(d,p). Comparison of the vibrational motions for the assigned modes between the two basis sets confirms that these are indeed analogous modes. To reiterate the significance of these specific phonon vibrations, the 33.6 cm⁻¹ mode appears in the THz spectrum at low temperature, but is ostensibly absent at room-temperature. The absorption at 66.7 cm⁻¹ is unusually broad and is also not directly evident at room temperature. The 92.1 cm⁻¹ mode, which is located at 88.8 cm⁻¹ at room temperature, is present in the THz spectra at both 78 and 293 K and exhibits seemingly typical temperature-dependent phonon behavior. The calculated eigenvectors for each of the three normal modes are shown in Figure 2.6. It can be seen that the 33.6 and 66.7 cm⁻¹ modes are along the direction of the suspected phase change relating the *P*2₁/*n* and *Pnma* symmetry cells perpendicular to the ⟨100⟩ lattice plane (*P*2₁/*n* bc plane). The difference in the modes principally lies in the relative molecular translational and rotational contributions to each. The 33.6 cm⁻¹ mode has primarily translational character with neighboring BTA molecules moving in opposing phase. The 66.7 cm⁻¹ exhibits the adjacent BTA molecules rotating in-phase along the c-axis about their respective molecular centers of mass. In combination with the associated motions of the two H₂O molecules, this vibration moves along a minima of the potential energy surface represented by the double-well potential shown in Figure 2.4A relating two equivalent energetic minima. Finally, the vibrational mode corresponding the peaks at 88.8 and 92.7 cm⁻¹ in the 293 K and 78 K

spectra, respectively, consists of combined internal deformation of the BTA molecule, notably in the CH_2 groups, mixed with rotational character about the a -axis. The H_2O molecular motions are both translational and rotational in character.

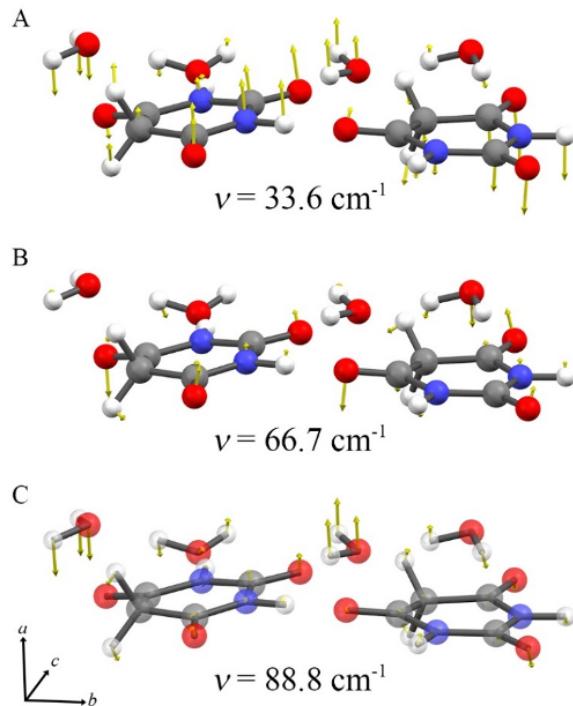


Figure 2.6: (A) Mode character of the spectral features at 33.6 cm^{-1} observed in the 78 K THz spectrum, represented by calculated displacement vectors (yellow arrows, magnitudes scaled $\times 10$). (B) Calculated mode displacements of the broad feature centered at 66.9 cm^{-1} in the 78 K THz spectrum. (C) Mode character of the distinct 88.8 cm^{-1} vibrational mode observed in the 293 K THz spectrum (corresponding to the feature located at 92.7 cm^{-1} in the 78 K spectrum).

The unusual THz spectra obtained for BTADH can be explained by examining the vibrational mode characters. In the room-temperature spectrum, phonon modes having displacements primarily along the a -axis and are of rotational and translational character are not distinctly present, although a very broad absorption profile across the collection bandwidth is observed. The one distinct spectral feature at 293

K is due to a vibrational mode containing to a large extent intramolecular deformations, making it nearly independent of the barrier transition coordinate. We propose that subtle crystal disorder exists at higher temperatures about the low-energy barrier lying on the $\langle 100 \rangle$ lattice plane, or equivalently, the mirror plane confining the crystal system in imposed *Pnma* symmetry. From our combined experimental and computational findings, the crystal system can more accurately be described in a monoclinic *P2₁/n* setting having a β angle 90° . Two degenerate energetic states exist in *P2₁/n* related by a two-fold rotation about the b-axis, and the two states are separated by a low-energy barrier corresponding to the originally reported *Pnma* structure. Given the quite low energy barrier as reported in Table 2.4, the crystal structure would be able to transition through the closely separated energy states, assuming there is sufficient thermal energy within the system. In this case, the resulting crystal structure would have the appearance of a time-averaged structure in a planar configuration that could be represented in *Pnma* symmetry. Upon cooling below 217 K, the BTADH crystal structure settles into a *P2₁/n* configuration. The degenerate energy states available for BTADH to access at higher temperatures provide the origins of the non-merohedral twinning phenomenon observed by X-ray diffraction in the system at reduced temperatures. The diminished thermal fluctuations in the crystal systems locks BTADH into either of the *P2₁/n* configurations resulting in the possible twinning of neighboring unit cells joined about a two-fold b-axis rotation. Evidence of crystal twinning is also presented in the THz spectrum obtained at 78 K. Vibrational modes with motions in the direction of the transition coordinate connecting the two equivalent *P2₁/n* structures, i.e. those which would pass through a near planar *Pnma* structure with large atomic displacements, exhibit

significant mode softening as these modes approach the transition state along their vibrational coordinates. Other similar vibrational modes that would not cross this planar structural barrier but do, however, have primary displacements perpendicular to the inversion plane appear in the THz spectrum collected at 78 K without significant broadening of spectral features. These same vibrational modes are not observed at 293 K as the modes are effectively washed out due to the subtle crystal disorder above 217 K.

2.5 Conclusions

Solid-state DFT and THz spectroscopy data suggest that at temperatures above the previously reported phase transition temperature of 217 K, the crystal does not exist in the *Pnma* space group configuration. However, two equivalent energetic states with *P2₁/n* symmetry lie on either side of the proposed *Pnma* structure. It is these same minima that guide the crystal system to a non-merohedrally twinned configuration at decreased temperatures. At temperatures above 217 K, the system contains the thermal energy necessary to readily transition between these degenerate states separated by the nearly energetically equivalent *Pnma* structural barrier. The spatial and energetic proximity of the *P2₁/n* and *Pnma* structures result in the subtle disorder of the BTADH system at higher temperatures. It is this disorder which initially led investigators through the comprehensive examination of the complexity in resolving what was thought to be a high-temperature phase, for which the *Pnma* space group was settled despite remarkable efforts to define the system otherwise. The evidence presented in the current study point to an absence of a true temperature-dependent phase change, and rather to a system that exists in a delicately disordered state above the once-thought transition temperature.

2.6 Acknowledgements

We would like to acknowledge high-performance computing support of the R2 compute cluster (DOI: 10.18122/B2S41H) provided by Boise State University's Research Computing Department. The Authors would like to thank Boise State University for continued support.

CHAPTER 3:

COMPUTATIONAL REDETERMINATION OF VIOLURIC ACID MONOHYDRATE'S CRYSTALLINE CONFORMATION

The following chapter contains yet to be published work written by me with the guidance of Dr. Matthew King.

3.1 Summary

An investigation into the crystalline structure of violuric acid monohydrate (VAMH) using density functional theory (DFT) and terahertz (THz) spectroscopy resulted in the redetermination of the previously reported orthorhombic *Cmc*2₁ crystal structure. Calculated imaginary modes across a broad range of unit cell volumes called into question the validity of the reported planar *Cmc*2₁ space group ($a = 6.0754 \text{ \AA}$, $b = 14.3430 \text{ \AA}$, $c = 7.5288 \text{ \AA}$, $Z = 4$), determined by single-crystal X-ray diffraction. A scan of the potential energy surface along the imaginary mode displacement coordinate of the *Cmc*2₁ structure resulted in a symmetric double-welled potential, and led to the conclusion that the apparent planar *Cmc*2₁ configuration was a result of the averaging of two energetically equivalent non-planar configurations. A further geometry optimization with fully reduced symmetry resulted in a perceived monoclinic

$P2_1$ configuration ($a = 6.31442 \text{ \AA}$, $b = 7.31395 \text{ \AA}$, $c = 7.74019 \text{ \AA}$, $\beta = 113.4318^\circ$, $Z = 2$) with the molecules shifting slightly out of their reported planar configuration. Comparison of the experimental and simulated THz spectra ($10 - 130 \text{ cm}^{-1}$) of the $P2_1$ unit cell, however, showed poor alignment with many experimental vibrational modes unaccounted for in the calculations. A geometry optimization of an expanded $2 \times 2 \times 2$ supercell with P_1 symmetry resulted in the perceived orthorhombic $Pca2_1$ space group ($a = 7.32092 \text{ \AA}$, $b = 6.31631 \text{ \AA}$, $c = 28.37480 \text{ \AA}$, $Z = 8$). No imaginary modes were observed in the normal-mode calculation for the $Pca2_1$ unit cell along with nearly harmonic potentials for the low-frequency phonon modes. Good alignment of the experimental and theoretical THz spectra resulted in the conclusion that crystalline VAMH resides in an orthorhombic $Pca2_1$ configuration.

3.2 Introduction

The structural conformation of crystalline materials largely influences the physical properties displayed by the material, such as the mechanical properties, melting point, dissolution rate, and optical properties.[10, 11, 43] Comprehension of these physical properties has a necessary impact on the fields of crystal engineering and de novo materials design[85, 86, 87]. Investigation of elaborate molecular crystal systems using combined computational analysis and experimental spectroscopic measurements provides the ability to accurately describe the physical phenomena leading to their observed behaviors [88, 50, 48, 89, 90]. In this study the crystalline configuration of violuric acid monohydrate (VAMH) was elucidated via solid-state density functional theory (DFT) and terahertz (THz) spectroscopy. The crystal structure of VAMH was previously reported to display a peculiar orthorhombic $Cmc2_1$ configuration at 150 K as determined by single-crystal X-ray diffraction (Figure 3.1).[91] Although

the molecular structure of violuric acid consists of a planar substituted six-member ring, its 1:1 co-crystallization with water makes it unlikely for all VAMH crystalline molecular orientations to lie on parallel reflections planes given the enhanced conformational flexibility of the hydrogen bonding network with the inclusion of water [92].

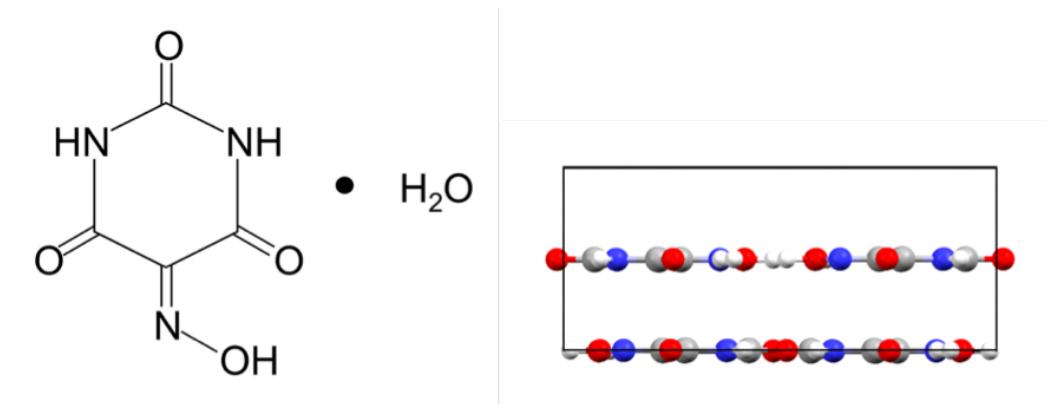


Figure 3.1: A) Violuric acid monohydrate (VAMH) molecular structure and B) *Cmc2₁* crystal configuration. Viewed along c-axis.

In a previous study we explored a proposed polymorphic phase transition of barbituric acid dihydrate, a structurally similar compound to VAMH. We concluded through experimental and computational analyses that the reported temperature-dependent phase transition was actually a result of thermal-induced crystal dynamic disorder and a relationship between a high-temperature disordered state and a low-temperature static state, both of which maintained the same crystal symmetry. The proposed phase transition occurred around 217 K, with the molecules shifting from a planar (*Pnma*) to non-planar configuration (*P2₁/n*) as the temperature of the system was decreased [1, 93, 94]. Our data was consistent with a dynamic crystal disorder

resulting from the oscillation of two non-merohedrally twinned $P2_1/n$ configurations. This oscillation was observed as a planar configuration by X-ray crystallography due to the average of the two energetically equivalent and closely spaced $P2_1/n$ states. Due to the structural similarity between BTADH and VAMH, the accuracy of the VAMH planar $Cmc2_1$ crystal structure was called into question [14]. Investigation into the crystal conformation of VAMH at 150 K concluded with a redetermination of the space group from planar $Cmc2_1$ to non-planar $Pca2_1$. Initial normal-mode calculations of the $Cmc2_1$ unit cell resulted in an imaginary mode. Relaxed symmetry geometry optimizations with the initial $Cmc2_1$ unit cell parameters resulted in the perceived Z=2 unit cell with $P2_1$ symmetry. Upon further analysis via THz spectrometry it was discovered that the simulated $P2_1$ and experimental spectra showed poor agreement, with the number of calculated low-frequency phonon modes falling short of those observed experimentally. A final geometry optimization of a 2x2x2 supercell with P_1 symmetry yielded a $Pca2_1$ space group (Z=8). The simulated $Pca2_1$ and experimental THz spectra displayed good alignment, concluding that the redetermined crystal structure of VAMH is of $Pca2_1$ symmetry.

3.3 Materials and Methods

3.3.1 Experimental THz Measurements

Samples were prepared for terahertz spectroscopy measurements by initially grinding a 2.5% w/w mixture of violuric acid with polytetrafluoroethylene (PTFE) in order to homogenize the samples and reduce particle size in order to minimize scattering effects. The sample-PTFE mixture was then pressed into a free-standing 13mm diameter pellet using a hydraulic pellet press (Specac Ltd., UK), yielding an 888.34 mg

pellet with a thickness of 3.03mm. A second pellet using only PTFE was used as a standard reference, with the same dimensions as the sample-containing pellet. The pellets were mounted in a closed-cycle liquid helium cryostat (Crycocoool Industries, USA), which permitted variable temperature acquisition between 20 and 300 K. For the terahertz time-domain spectroscopy measurements, a commercial THz-TDS system (Toptica Photonics AG, DE) was used, which consists of a pair of fiber-coupled emitter and receiver modules. Four off-axis parabolic mirrors were used to collimate and focus the terahertz radiation, and the entire spectrometer was continuously purged with nitrogen gas to minimize absorption from atmospheric water vapor. At each temperature, 20,000 terahertz time-domain waveforms were averaged for both the sample and blank, which was repeated four times. The time-domain data were Fourier transformed to obtain terahertz power spectra, and each sample-blank pair was used to produce absorption spectra, with four absorption spectra per temperature averaged to yield the final spectra presented.

3.3.2 Solid-State Density Functional Theory

The Crystal17 software package was utilized to perform all DFT calculations, using the PBE-D3 functional and pob-TZVP [95] basis set combination [58, 57]. Default values were used for the three-body (D3) empirical energy correction for London-type dispersion forces [58, 96, 97, 98]. Atom-only geometry optimizations were run on VAMH confined to the *Cmc2₁*space group. Initial lattice dimensions and atomic coordinates for this optimization were obtained from the Cambridge Crystallographic Database [91, 68]. Space group perceptions were performed using the Avogadro software package with atom position tolerances of 0.01 Å or less [99]. Constant volume geometry optimizations were performed with energy convergence criteria of $\Delta E \pm 10^{-5}$

6 hartree for initial optimizations and $\Delta E \geq 10^{-8}$ hartree for geometry optimizations used for frequency calculations. Normal mode calculations were run for the optimized structures using the total energy convergence criteria of $\Delta E \geq 10^{-11}$ hartree. The optimal sampling rate for the k points used to define the real space density matrix and the reciprocal space Monkhorst grid were determined for the total energy convergence of $\Delta E \geq 10^{-11}$ hartree for the $Cmc2_1$ optimization [100, 70]. Shrinking factors of 6 were used for all calculations in order to meet the above listed energy convergence criteria. The radial and angular distributions of points were defined by a pruned (75,974) integration grid for each calculation. The truncation tolerances for the Coulomb and HF exchange integral series were set to 10^{-8} , 10^{-8} , 10^{-8} , 10^{-8} , and 10^{-16} hartree. The IR intensities for the normal mode calculations were determined by the Berry phase approach of calculating polarization differences between equilibrium and distorted geometries utilizing the dipole moment derivatives ($d\mu/dQ$) [75, 101, 102].

3.4 Results and Discussion

Initial geometry optimizations of VAMH were performed with the molecules confined to $Cmc2_1$ symmetry group with fixed lattice parameters corresponding to the experimental X-ray crystal structure determined at 150 K ($a = 6.0754$ Å, $b = 14.3430$ Å, $c = 7.5288$ Å, $Z = 4$). The atomic coordinates from the $Cmc2_1$ cell optimizations were used in a normal mode calculation for VAMH, resulting in a single imaginary mode at -22.33 cm $^{-1}$. The potential energy surface following the imaginary mode displacements were scanned, resulting in a double-well potential, with a small energy barrier of 0.058 kJ mol $^{-1}$ between the two energetically equivalent minima (Figure 3.2). Graphical representations of the displacement vectors for the imaginary mode were constructed to view the molecular displacements associated with the imaginary mode

(Figure 3.3). As seen, the atomic displacements were observed to be equal, but opposite across the vertical mirror plane (parallel to the b-c plane) within the molecule, suggesting an in-phase rocking motion between two mirrored configurations.

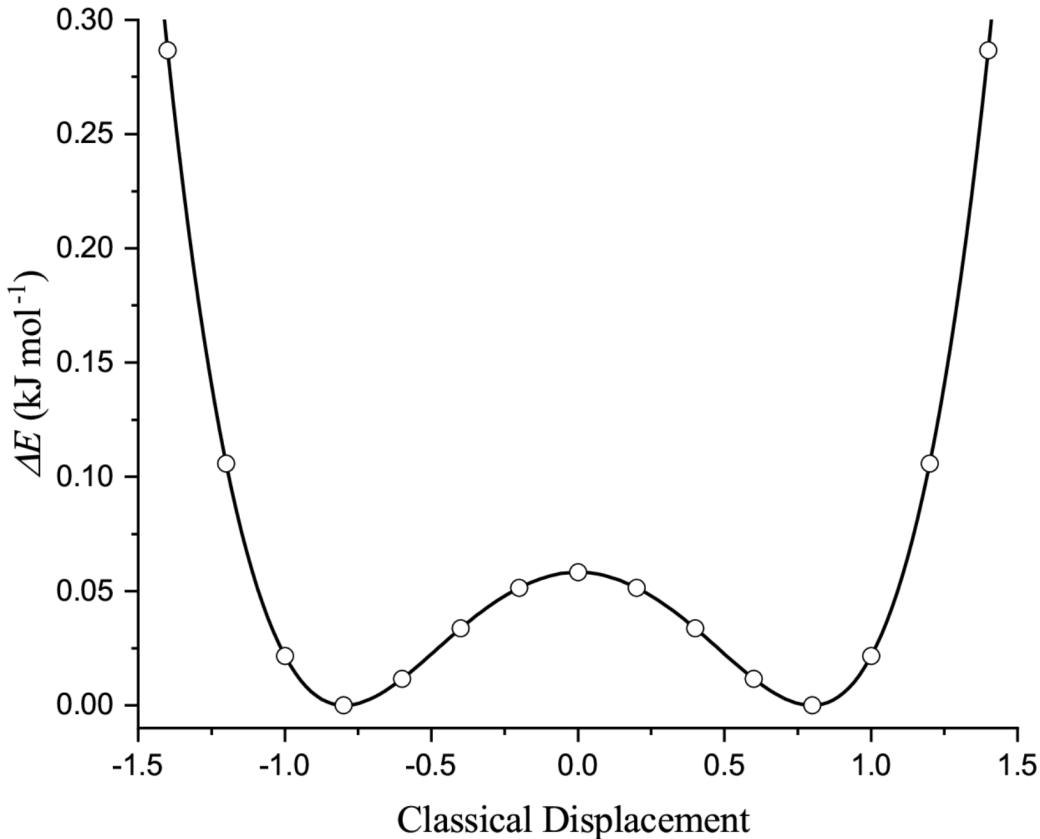


Figure 3.2: Potential energy scan along imaginary mode at 150 K.

To further investigate the energetically favorable crystalline packing of VAMH, all symmetry constraints were removed, relaxing the $Cmc2_1$ space group to P_1 . Constant-volume geometry optimizations were performed at various volumes to investigate the effect of unit cell size (i.e. thermal expansion) on the optimization of the molecules. The lattice parameters and calculated electronic energies of the optimized structures

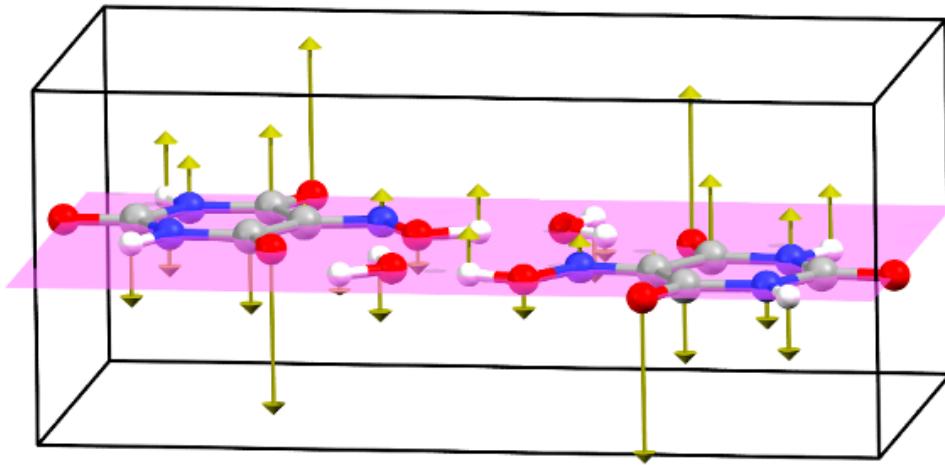


Figure 3.3: A) Vibrational character of the imaginary mode calculated at 22.33 cm⁻¹ for VAMH constrained to *Cmc2*₁ space group symmetry. Atomic displacement vectors are represented by yellow arrows (magnitudes scaled $\times 10$).

are tabulated in the Supporting Information. The molecules in each of the resulting optimized structures were shifted slightly out of plane, as seen in Figure 3.4 and Figure 3.5. The optimized *P*₁ structure corresponding to the 150-K experimental unit cell volume was analyzed, and a higher-symmetry *P*₂₁ molecular configuration was perceived with parameters $a = 6.31442 \text{ \AA}$, $b = 7.31395 \text{ \AA}$, $c = 7.74019 \text{ \AA}$, $\beta = 113.4318^\circ$, $Z = 2$. The *P*₂₁ space group contained only two asymmetric units, which is a reduction of half from the original *Cmc2*₁ space group. Subsequent constant volume geometry optimizations with *P*₂₁ symmetry were conducted resulting in the *b*-cell length being reduced by approximately half from the *Cmc2*₁ to *P*₂₁ configuration, along with the β -angle increasing from 90° to approximately 113°. An overlay of the *P*₁ (green) and *P*₂₁ (black) unit cells within the crystal structure is shown in

Figure 3.4A.

Correlating the static molecular orientations observed in the P_1 and $P2_1$ optimized structures with the imaginary mode character obtained from the $Cmc2_1$ normal mode calculations revealed that the out-of-plane molecular shifts correspond to the direction of the imaginary mode displacement vectors shown in Figure 3.5. This break in planar symmetry from $Cmc2_1$ to $P2_1$ can be explained by the changes in intermolecular contacts between the two configurations (Figure 3.6, Table 3.1). Relaxation of the crystal symmetry permits the formation of more stable hydrogen bonding and an increased distance of repulsive O…O contacts between adjacent VA molecules (labeled as contact 5). Considerably more favorable direct hydrogen bonding involving the water molecule was obtained in $P2_1$, most markedly for contact 2 which was reduced by -0.501 Å in comparison to the $Cmc2_1$ structure. Contacts 1 and 4 were modestly reduced by -0.033 and -0.028 Å, respectively. The contact 3 distance remained relatively unchanged between the two crystal structures. Also contributing to the addition stabilization of the $P2_1$ optimized structure is the increase in the short-contact O…O distance between carbonyl oxygens of neighboring VA molecules by 0.048 Å, thereby reducing the repulsive strain forces present within the crystal structure.

Table 3.1: Heavy atom distances for intermolecular contacts observed in the $Cmc2_1$ and $P2_1$ optimized unit cells.

Contact	Distance (Å)		
	$Cmc2_1$	$P2_1$	Difference
1	2.507	2.474	-0.033
2	2.969	2.468	-0.501
3	2.758	2.761	0.003
4	2.676	2.648	-0.028
5	2.745	2.793	0.048

ht!

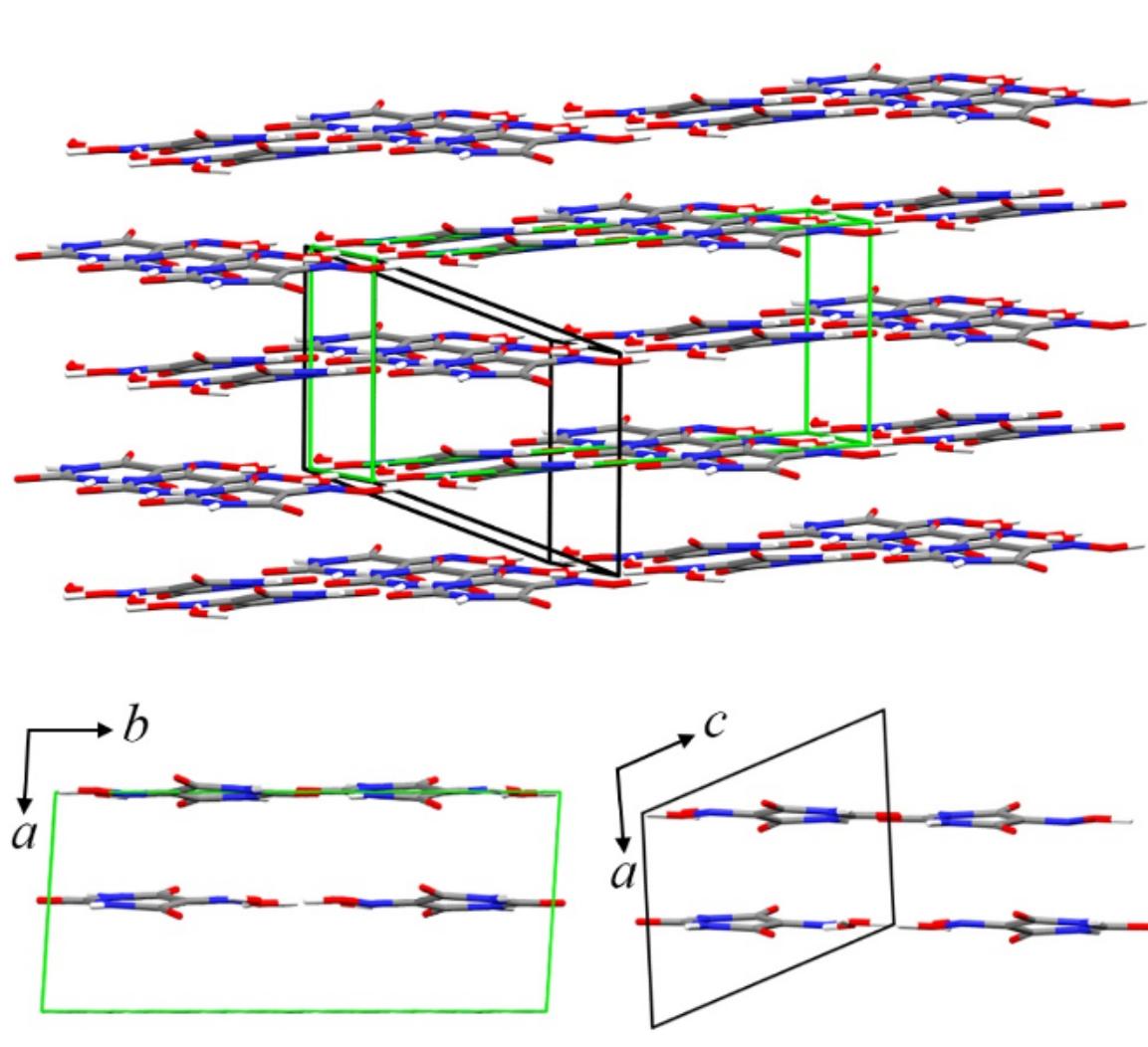


Figure 3.4: A) P_1 and P_{21} unit cell overlay. B) Packed P_1 unit cell viewed along the c axis. C) Packed P_{21} unit cell viewed along the b axis.

ht!

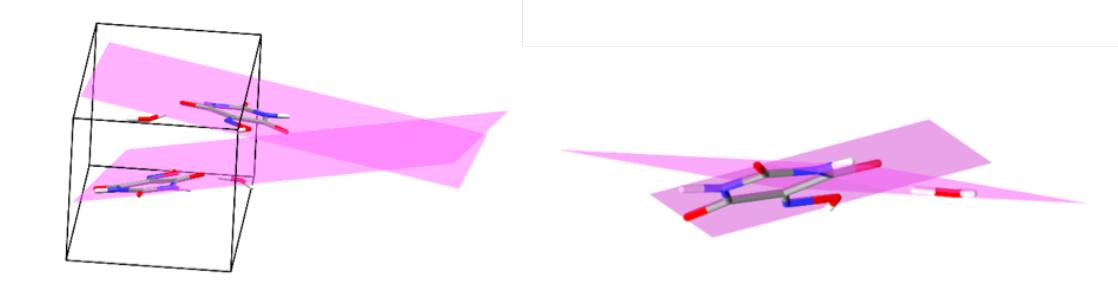


Figure 3.5: Depicted planes between A) violuric acid molecules and B) water violuric acid molecular planes of VAMH confined to the $P2_1$ space group.

In effort to validate the proposed $P2_1$ symmetry configuration of VAHM, the phonon vibrational frequencies were calculated using the optimized crystal coordinates. As expected, there were no imaginary modes calculated in the normal mode analysis of the $P2_1$ crystal structure, suggesting that a minimum energy configuration was achieved. The calculated low-frequency vibrational modes were subsequently compared to experimental THz spectrum of VAMH obtained in the spectral range of 10-130 cm⁻¹ (Figure 3.7). It was observed that the simulated $P2_1$ THz spectra did not align well with the experimental spectrum, with a notably lower calculated density of states. Only 5 IR-active phonon modes having appreciable absorption intensities were calculated in the probed spectral range while 9 modes could clearly be resolved in the experimental THz spectrum, along with several evident underlying modes contributing to observed peak asymmetry. The additional peaks in the experimental spectrum could not be accounted for with the predicted $P2_1$ crystal structure, however the calculated modes could be tentatively correlated to observed

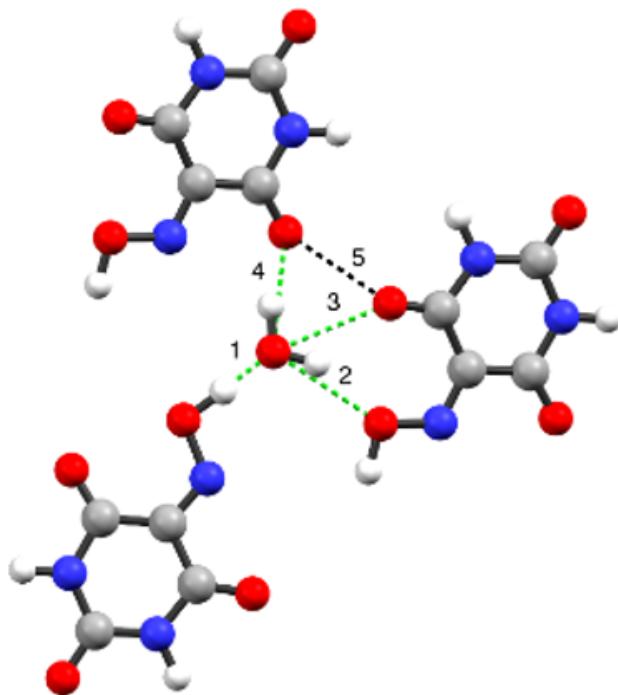


Figure 3.6: Primary electrostatic contacts observed in the VAMH crystal structures. Hydrogen bonds are indicated by dashed green lines and repulsive interactions by dashed black lines.

spectral features. This indicated that the basic VAMH molecular orientations predicted in the $P2_1$ configuration were perhaps correct, but the $Z=2$ unit cell was too reduced and thereby limited the number of phonon modes present in the calculations.

In light of the information provided by the experimental THz spectra, a $2 \times 2 \times 2$ supercell was generated from the original $Cmc2_1$ crystal structure, and a geometry optimization was performed with reduced P_1 symmetry under constant volume constraints. The resulting structure was then analyzed for symmetry elements present within the P_1 supercell. Using an atomic-position tolerance of 0.05 Å, a new unit

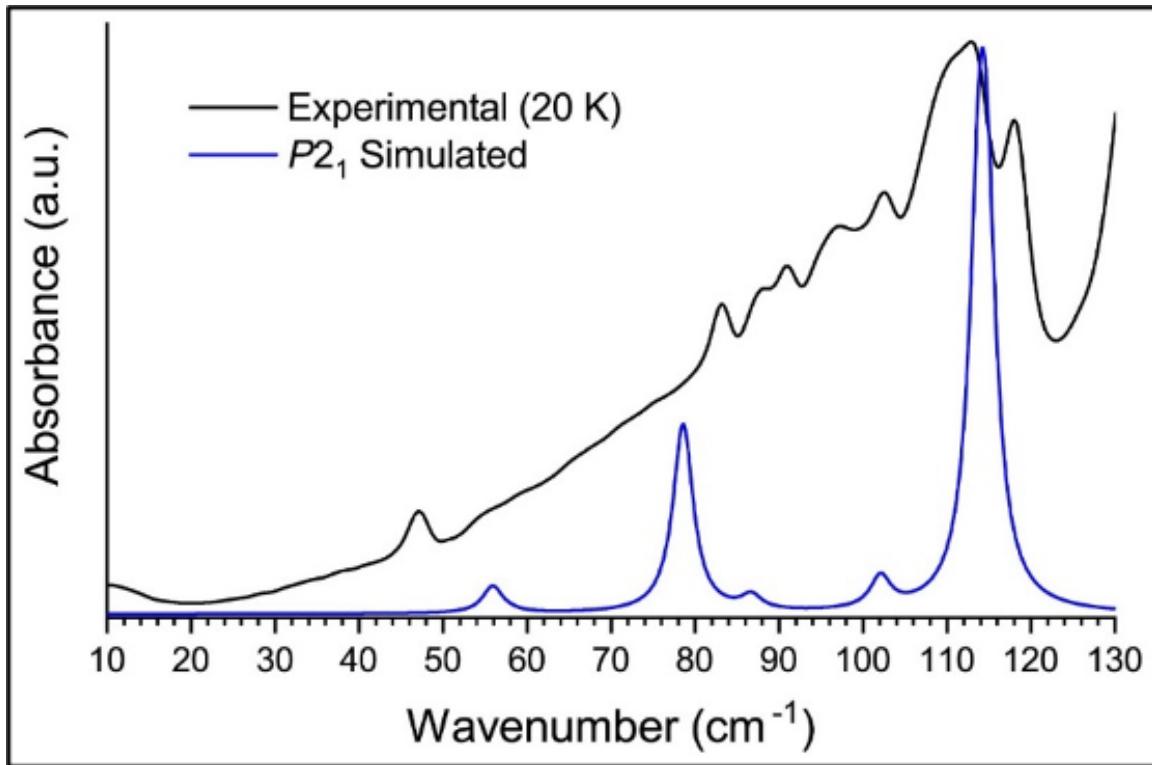


Figure 3.7: Comparison of the simulated $P2_1$ THz spectrum to the experimental THz spectrum obtained at 20 K. Lorentzian lineshapes of FWHM 5 cm⁻¹ were convolved into calculated normal modes, and mode intensities of calculated and experimental spectra were normalized.

cell of $Pca2_1$ symmetry was perceived to be contained within the optimized P_1 supercell structure (Figure 3.8). The VAMH molecular orientations of the larger $Pca2_1$ ($Z=8$, $Z'=2$) resembled those observed in the previously predicted $P2_1$ ($Z=2$, $Z'=1$) crystal structure. However, the VAMH molecular pairings along the elongated c axis were now related by a screw axis as opposed to restricted translationally invariant in the reduced $P2_1$ packing configuration. The predicted lattice demensions of the $Pca2_1$ and the $P2_1$ crystal structures provided in Table 3.2.

Comparison of calculated unit cell energies shows the energetic favorability of the $Pca2_1$ unit cell configuration, with relative energy 1.052 kJ mol⁻¹ molecule⁻¹ lower in

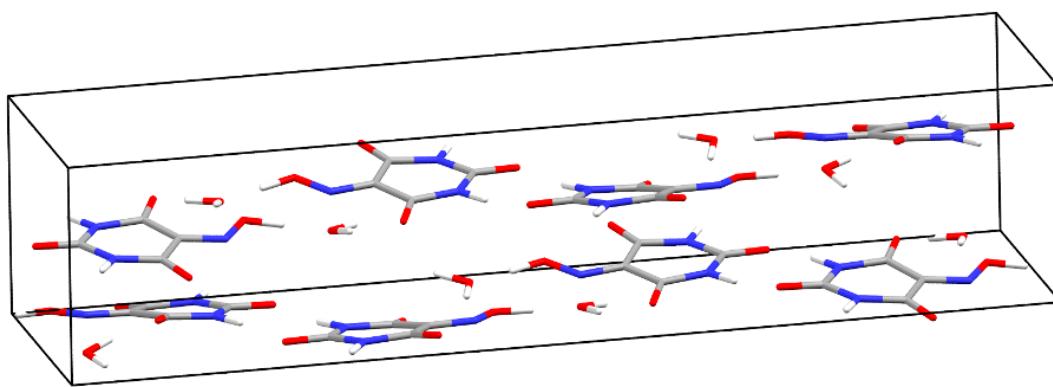


Figure 3.8: Molecular packing of the predicted VAMH *Pca*2₁ crystal structure.

energy than the *P*2₁ cell and 1.647 kJ mol-1 molecule-1 lower than the constrained *Cmc*2₁ structure (Table 3.2). The energy difference separating each crystal form is small and presents a low transition barrier between equivalent configurations related by the *Cmc*2₁ reflection planes. The phase coexistence is likely the cause of the apparent crystal disorder and the original conclusion of VAMH residing in the *Cmc*2₁ space group, as this would represent the average of two mirrored equivalent energetic minima.

Table 3.2: Tabulated comparison of the unit cell parameters and electronic energies of the investigated symmetry groups.

	a (Å)	b (Å)	c (Å)	beta (°)	Relative Energies (kJ mol-1 molecule-1) ^a
^b Cmc2₁	6.0754	14.3430	7.5288	90	1.647
P2₁	6.31442	7.31395	7.74019	113.4318	1.052
Pca2₁	7.32092	6.31631	28.37480	90	0.000

^aRelative energies obtained from constant-volume optimizations; ^bRef [91].

Normal modes of vibration were calculated for the predicted *Pca*2₁ crystal structure and the low-frequency vibrational modes were compared with the experimental

THz spectrum (Figure 3.9). A good quality fit of the simulated spectrum to the experimental was obtained. The number of calculated modes is sufficient to assign many of the experimental peaks, with additional low-intensity calculated modes that likely contribute to the high and uneven baseline of the THz spectrum, as well as to the peak asymmetry observed in several of the experimental features. The calculated vibrational frequencies were generally underestimated upon inspection of the spectral comparison. A reasonable explanation for this result is the comparison of the simulated spectrum calculated from an effective 150 K unit cell volume *s* the structure was optimized under constant-volume conditions from the original *Cmc2*₁ crystal structure measured at this temperature [78]. The simulated spectrum is overlayed with the experimental spectrum obtained at 20 K for enhanced resolution of experimental spectral features.

Through a systematic computational analysis of the VAMH crystal structure in conjunction with experimental validation by low-frequency THz spectroscopy, a re-determination of the VAMH crystal structure was achieved. The previously reported *Cmc2*₁ crystal structure exhibiting a planar arrangement of violuric acid and water molecules was shown to be inaccurate and likely the result of structural averaging during experimental X-ray measurements. Similar behavior was previously reported for the structurally similar barbituric acid dihydrate, for which a planar configuration of *Pnma* symmetry was reported for the suspected low-temperature polymorph. However, it was conclusively demonstrated that the crystal structure maintained *P2*₁/n symmetry, that which was reported as the high-temperature polymorph. In the cases of both VAMH and barbituric acid dihydrate, the planar crystalline configuration represents an averaged structure at the transition point connecting two symmetry-

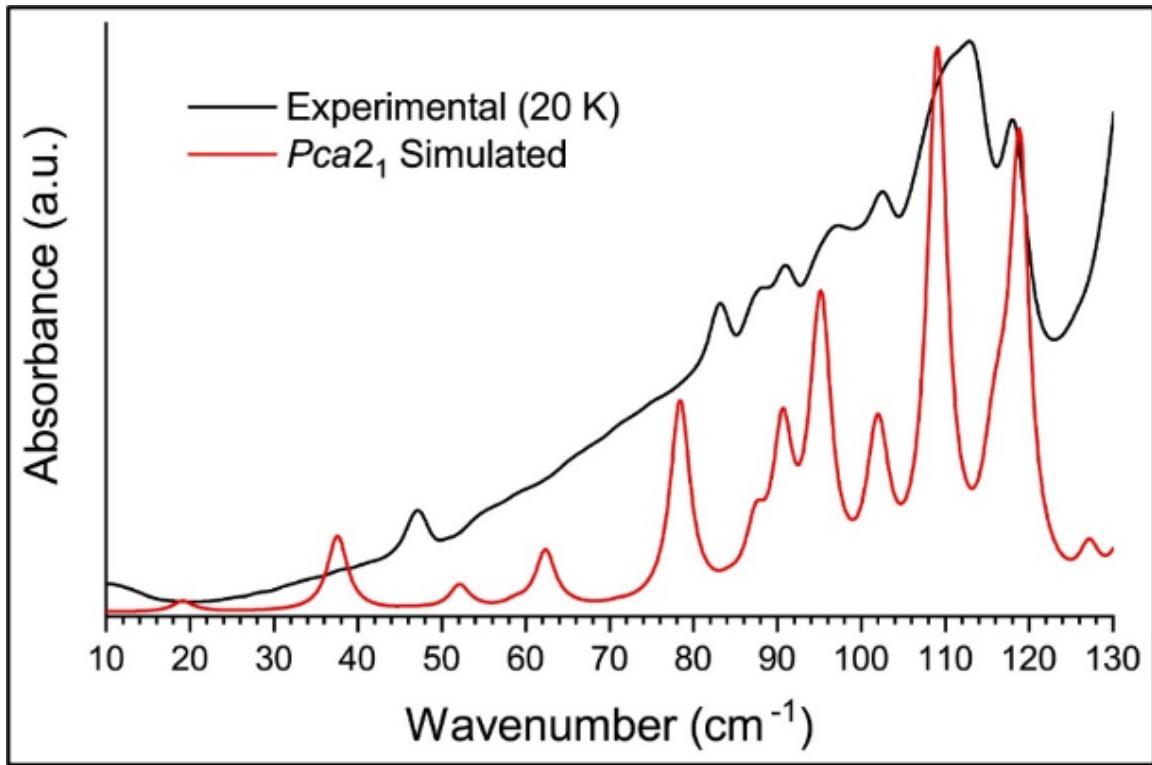


Figure 3.9: Comparison of the simulated $Pca2_1$ THz spectrum to the experimental THz spectrum obtained at 20 K. Lorentzian lineshapes of FWHM 5 cm^{-1} were convolved into calculated normal modes, and mode intensities of calculated and experimental spectra were normalized.

equivalent mirrored unit cell structures. At the two minima, distinct crystal symmetries exist that do not conform to the reflection planes imposed by the $Cmc2_1$ and $Pnma$ space groups, respectively. In the present study, a more energetically favorable configuration of $Pca2_1$ was identified for VAMH. The validity of the predicted structure was verified using THz spectroscopy. Normal mode calculations of the optimized $Pca2_1$ crystal structure produced phonon modes that were in good agreement with the experimental THz spectra.

3.5 Conclusion

Investigation into the crystal structure of VAMH demonstrated the theoretical and experimental data did not support the proposed $Cmc2_1$ symmetry configuration. Imaginary modes were observed across a broad range of defined unit cell volumes while VAMH was confined to the planar $Cmc2_1$ configuration. Initial geometry optimizations with all symmetry operations removed resulted in a perceived $Z=2$ unit cell of space group $P2_1$. The shift in molecular orientation was supported by the plotting of the displacement vectors along the observed imaginary mode. This information supported that the actual crystalline arrangement of VAMH deviated from the planar configuration, and that the planar structure may be the averaged structure connecting two energetically equivalent non-planer structures. Explained by the double-well potential and imaginary modes revealed with VAMH confined to the $Cmc2_1$ space group. However, when the simulated THz spectrum of the $P2_1$ unit cell was compared to the experimentally obtained spectrum, there was poor agreement between the two. Further evaluation of the original crystal structure yielded an orthorhombic $Pca2_1$ space group configuration following the generation and structural optimization of a 2x2x2 supercell ($Z=32$) in P_1 symmetry. The normal-mode calculation for the $Pca2_1$ unit cell resulted in no imaginary modes and nearly harmonic potentials for the low-frequency phonon modes. When comparing the simulated $Pca2_1$ and experimental THz spectra, there was good alignment between theory and experiment. This study provides a conclusive redetermination of the violuric acid monohydrate crystal structure in $Pca2_1$ symmetry.

CHAPTER 4:

VALIDATING STRUCTURAL PREDICTIONS OF CONJUGATED MACROMOLECULES IN ESPALOMA-ENABLED REPRODUCIBLE WORKFLOWS

The following chapter contains yet to be published work written by me with the guidance of Dr. Eric Jankowski. The following work has been submitted to the International Journal of Molecular Sciences and is expected to be published in January 2025.

4.1 Abstract

We incorporate `Espaloma` forcefield parameterization into `MoSDeF` tools for performing molecular dynamics simulations of organic molecules with `Hoomd-Blue`. We compare equilibrium morphologies predicted for perylene and poly-3-hexylthiophene (P3HT) with the `ESP-UA` forcefield in the present work against prior work using the `OPLS-UA` forcefield. We find that after resolving chemical ambiguities in molecular topologies, `ESP-UA` is similar to `GAFF`. We observe clustering/melting phase behavior to be similar between `ESP-UA` and `OPLS-UA`, but that the base energy unit of `OPLS-UA` better connects to experimentally-measured transition temperatures. Short-range ordering

measured by radial distribution functions is essentially identical between the two forcefields, and long-range ordering measured by grazing incidence x-ray scattering is qualitatively similar, with **ESP-UA** matching experiments better than **OPLS-UA**. We conclude that **Espaloma** offers promise in the automated screening of molecules from more complex chemical spaces.

4.2 Introduction

Molecular simulations offer promise for the high-throughput screening of thermodynamically stable structures from families of molecules. Such screening studies can identify chemistries and conditions where self-assembly of desired structures are optimized, and can provide insight into identifying chemistry-structure-property relationships that are otherwise inaccessible with experimental methods[103, 104, 105, 106, 107, 108]. One challenge of screening studies, particularly of polymers, is that the parameters that define the interactions between and within model molecules may not have been defined for a given “off-the-shelf” forcefield[39]. In molecular mechanics simulations, forcefields define the potential energies and therefore forces of bond, angle, torsion/dihedral, and non-bonded models of interatomic interactions[109, 110]. For example, the popular Optimized Potentials for Liquid Simulations (OPLS) forcefield has mainly been built around simulating hydrocarbons and proteins [111, 36]. The General Amber forcefield (GAFF) is optimized for small organic molecules[112].

An example of missing parameters arises when attempting to simulate the popular organic semiconductor poly(3-hexylthiophene) (P3HT). When using OPLS, we find the forcefield is missing parameter definitions for hydrogen-carbon-carbon-sulfur (H-C-C-S) dihedrals. In such cases, the simulator typically has had two options: (1) they might reuse a similar parameterization from a forcefield based on their chemical

intuition, or (2) they can perform quantum chemical calculations to identify new forcefield parameters[109] that best fit the calculated potential energies. The former option has the downside of missing important physics and giving inaccurate results, while the latter option requires additional software fluency and calculation time[39]. Both options suffer from an additional shortcoming: the forcefields that result are *different* from the forcefields from which they were derived, which complicates data provenance and reproducibility.

Recent efforts towards making simulations more transparent, reproducible, usable, and extensible (TRUE)[113, 114] have identified programmatic workflows as a best practice. In the present context, this means that simulation scripts that fully specify forcefields contribute to more reproducible simulations, even if the forcefields are custom derivatives of prior work[115]. Quantum chemical calculations could therefore be included into the programmatic specification of custom forcefields and in principle be shared as reproducible workflows.

However, re-running quantum chemical calculations itself introduces software dependency issues and extended runtimes that are redundant. A new tool that ameliorates this problem is **Espaloma**, which is an open-source package that uses graph neural networks to perceive chemical environments in a molecular graph and predict molecular mechanics forcefield parameters [39]. **Espaloma** was designed for the investigation of biopolymers and has been used to accurately simulate them [116, 117]. Of central importance to the present work is **Espaloma**'s promise to programmatically “fill in” missing parameters for organic macromolecules and polymers, enabling more TRUE screening studies without extra quantum chemical calculations (Step 3 of Figure 4.1).

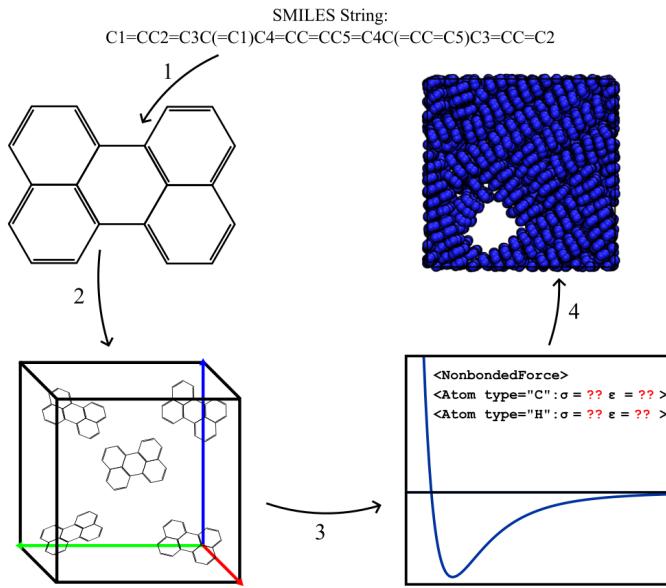


Figure 4.1: Depictions of our generalized molecular dynamics workflow. Step 1 shows the creation of an `mBuild Compound` from a SMILES string. In step 2 we create our simulation object using `flowerMD` and `PACKMOL`. We employ `Espalomato` parameterize our molecules in step 3 and write the forcefield file. In step 4 we initialize the HOOMD-Blue simulation and predict the morphology of our molecules.

Here, we evaluate `Espalomato` for the modeling of highly conjugated macromolecules. We generate forcefield parameters using `Espalomato` and compare them to the “off-the-shelf” OPLS-UA and GAFF forcefields where possible. We compare phase diagrams and morphologies from prior work to validate and quantify `Espalomato`’s predictive capabilities outside its core training set [39]. A schematic of the MD workflow we implemented in this work is shown in Figure 4.1.

4.3 Model

We consider united-atom (UA) representations of perylene and P3HT, omitting long-range electrostatics as in prior work[3, 5]. Spherical simulation elements are used to

represent each “heavy” atom and its bonded hydrogens. Here, the heavy atoms are carbon and sulfur, topologically connected as in Figure 4.2. Chains of $n = 15$ repeat

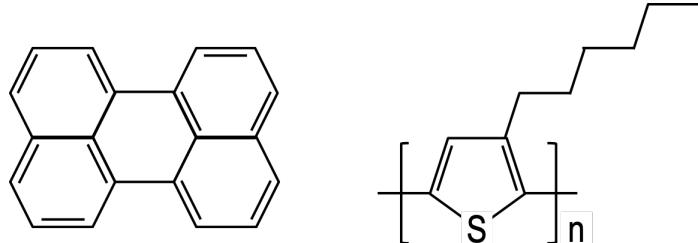


Figure 4.2: Diagram of perylene molecule (left) and poly-3-hexylthiophene (P3HT) monomer (right).

units are used to represent P3HT oligomers. Harmonic potentials are used to model bond-stretching (Equation 4.1) and angle bending (Equation 4.2) between pairs and triplets of bonded atoms. E_{bonds} is the potential energy of a bond, K_r is the bond’s harmonic spring constant, r is the internuclear distance, and r_{eq} is the equilibrium internuclear distance.

$$E_{bonds} = K_r(r - r_{eq})^2 \quad (4.1)$$

E_{angle} is the potential energy of a bond angle, K_θ is the harmonic angle constant, θ is the bond angle, and θ_{eq} is the equilibrium bond angle.

$$E_{angles} = K_\theta(\theta - \theta_{eq})^2 \quad (4.2)$$

Cosine series are used to model proper dihedrals (Equation 4.3) across quadruplets of bonded atoms, where $E_{dihedral}$ is the potential energy of the dihedral, K_n is the force constant of the corresponding particle, n , ϕ is the dihedral angle, and γ is the phase angle.

$$E_{dihedral} = \sum_{n=1}^4 \frac{K_n}{2} [1 + \cos(n\phi - \gamma)] \quad (4.3)$$

$$E_{LJ}(r_{i,j}) = 4\epsilon_{i,j} \left[\left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - \left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^6 \right] \quad (4.4)$$

Simulation elements that are not bonded or separated by more than three bonds within a molecule interact only via the 12-6 Lennard-Jones potential (Equation 4.4) [118] truncated at $r_{cut} = 2.5\sigma$, where σ is the length unit corresponding to the largest simulation element being represented. Here, this corresponds to sulfur (S) simulation elements for P3HT ($\sigma = 3.565\text{\AA}$), and carbon (C1) elements for perylene ($\sigma = 3.380\text{\AA}$). We apply the Lorentzian combination rule for sigma and a geometric combination rule for epsilon, shown in Equation 4.5.

$$\sigma_{i,j} = \frac{\sigma_i + \sigma_j}{2}, \epsilon_{i,j} = \sqrt{\epsilon_i * \epsilon_j} \quad (4.5)$$

The parameterization of the bond, angle, dihedral, and nonbonded interaction potentials are generated via `Espaloma` and we describe details and challenges with this in the Methods.

4.4 Methods

In this section we detail the molecular dynamics simulations performed with `Hoomd-Blue`[119] to predict equilibrium morphologies, describe the new tools we develop for integrating `Espaloma` into workflows utilize the Molecular Simulation Design Framework (MoSDeF)[120], and detail morphology characterization that underpins model validation.

4.4.1 Molecular Dynamics

Simulations are performed on the Fry high performance computing cluster at Boise State University using HOOMD-Blue on NVIDIA P100 and V100 GPUs. We use `signac`[121] to manage simulation workspaces and job submission. Simulation scripts are available online at github.com/madilynpaul/Espaloma-Validation. Equilibrium morphologies of perylene and P3HT are predicted in the canonical ensemble (constant number of particles N , volume V , and temperature T). Periodic boundary conditions of cubic volumes are used throughout this work. We initialize our system using PACKMOL[122] within `flowerMD`[123] to create initial low-density volumes that are randomized at high T (1653 K for P3HT and 696 K for perylene) and a shrinking simulation runs for 5×10^6 time steps to the target state point density. Newton's equations of motion are integrated using a two-step MTK velocity-verlet implementation of Nosé-Hoover chains[124, 125] with a step size of $dt = 0.0003$ for P3HT and $dt = 0.0001$ for perylene. Simulation volumes are then instantaneously quenched to the statepoint temperature, where the potential energy trajectory is analyzed to determine the onset of equilibrium. Once the potential energy has stabilized, the simulations continue until decorrelation time measurements provide at least 50 statistically independent snapshots have been generated.

4.4.2 Morphology Characterization

Equilibrium morphologies are quantified using radial distribution functions (RDF) and simulated grazing-incident x-ray scattering (GIXS) implemented in `freud`[126]. RDFs are calculated between perylene centers of mass (Figure 4.3a), and for P3HT, the sulfur of the thiophene rings (Figure 4.3b). Both the RDF and GIXS analyses are averaged over the last 20 independent snapshots of each simulation.

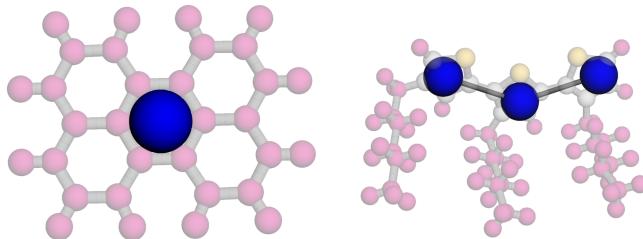


Figure 4.3: Blue spheres represent center-of-geometry positions used for RDFs and clustering criteria. Perylene (left), and P3HT (right).

To compare with prior work, we calculate an order parameter Ψ as used by Miller et al. [5] that measures the fraction of perylene molecules—or monomers for P3HT—belonging to clusters of at least size 6. Two perylene molecules—or two thiophene rings for P3HT—are considered clustered if their centers are within 6 Å and the best-fit planes through each moiety are within 10 degrees of being parallel. The code for calculating RDFs, GIXS, Ψ and phase diagrams is available at github.com/madilypaul/Espaloma-Validation.

4.4.3 Integrating Espaloma parameterization into MoSDef

Our simulation workflows are built around MosDef tools[127], in particular `mBuild` [128] whose `mbuild Compound` objects are incompatible with the `openMM Molecule` objects expected by `Espaloma`. To incorporate `Espaloma` into MoSDef workflows we develop a helper function termed “`BondWalker`” that can convert `mbuild Compounds` into `openMM Molecules` (Figure 4.4). This workflow fits into the overall MD workflow (Section 4.2, Figure 4.1) within Step 4. The general procedure is: **(1)** Create an `openMM Molecule` topology from connectivity information of the `mBuild Compound` to be parameterized. **(2)** Rebuild the missing double and triple bonds in the `openMM Molecule` using the `BondWalker` function (See Figure 4.5). **(3)** Pass the “`BondWalked`” molecule to `Espaloma` for forcefield parameterization, writing these parameters to

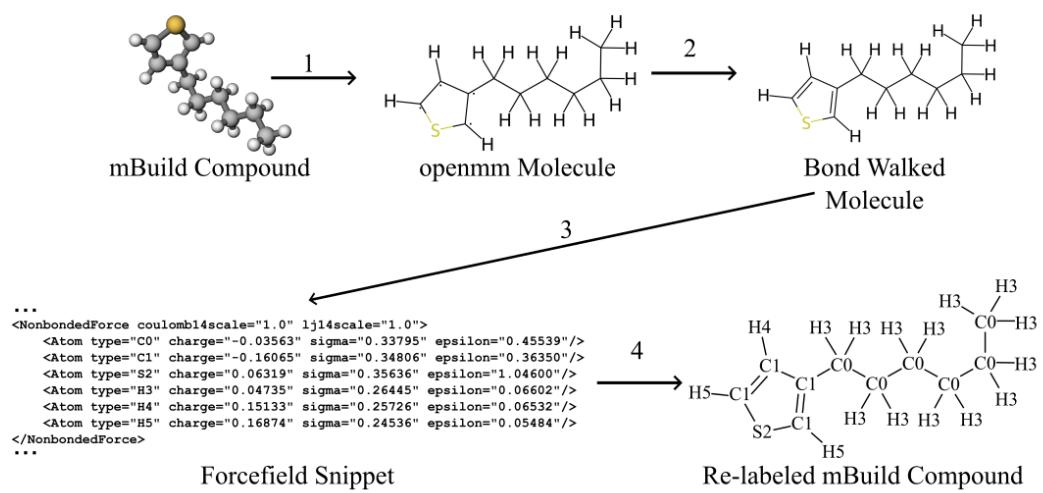


Figure 4.4: Overall Espaloma-MosDef workflow. (1) mBuild compounds are used to create the topology of an openMM molecule, (2) BondWalker uses octet rules to determine double bonds in the openMM molecule. (3) Espaloma generates forcefield parameters for the openMM molecule. (4) The ESP-UA force-field is used to re-type the mBuild compound for use in HOOMD-Blue simulations.

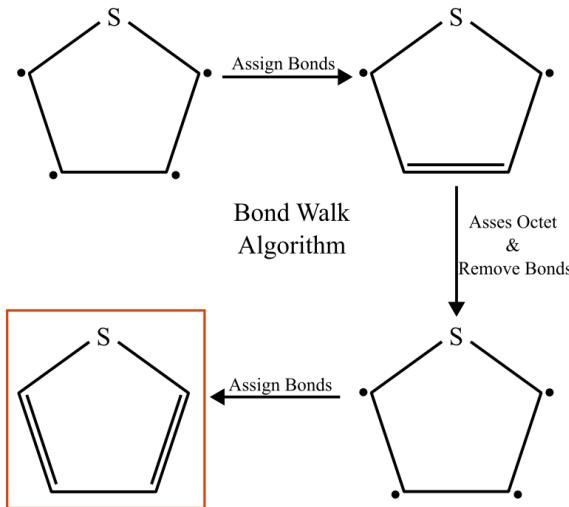


Figure 4.5: Double- and triple-bond information that is missing from `mBuild` compounds is retrieved via `BondWalker` by iteratively checking whether the octet rules can be satisfied for all atoms after incrementing the bond character adjacent to atoms with unsatisfied octets.

a forcefield `xml` file. (4) Use the atom labels given by `Espalomato` rename the atoms in our `mBuild Compound`. This last step ensures that the correct parameters will be applied to the corresponding atoms. A full tutorial of `Espaloma` forcefield and typed `mBuild` Compound generation can be found at github.com/madilynlpaul/Espaloma-Validation.

4.5 Results and Discussion

Here we report on the differences between the ESP-UA, GAFF, and OPLS-UA parameters followed by a comparison between ESP-UA generated morphologies and OPLS-UA morphologies generated in prior work.

4.5.1 Espaloma vs OPLS-UA and GAFF

The nonbonded pair interaction for perylene and P3HT are presented in Table 4.1, where they are compared against OPLS-UA and GAFF parameters. We first consider

Table 4.1: Lennard-Jones diameters σ and well-depths ϵ generated by espaloma in the present work, and OPLS-UA and GAFF.

	σ (Å)			ϵ (kJ/mol)		
	C1	C0	S	C1	C0	S
Espaloma	3.481	3.380	3.564	0.3635	0.4554	1.046
OPLS-UA	3.436	3.905	3.436	0.4602	0.7113	1.339
GAFF	3.3997	3.3997	3.564	0.3598	0.4577	1.046

the LJ σ values, and note that OPLS-UA is somewhat of an outlier, with much larger C0 diameters, while ESP-UA and GAFF are similar to each other. The ESP-UA C1's are slightly larger than both OPLS-UA and GAFF, while the ESP-UA C0's are slightly smaller than GAFF. GAFF and ESP-UA agree on the diameter of S atoms in P3HT. We next consider the ϵ values and note that ESP-UA is very similar to GAFF once again, with OPLS-UA as the outlier.

Normalizing the ϵ_{C1} and ϵ_{C0} values by ϵ_S provides a clearer picture of the range of attractions in each model than simple comparison the absolute values of each ϵ . For ESP-UA, $\epsilon_{C1}/\epsilon_S = 0.3475$, is nearly identical to that of OPLS-UA: $\epsilon_{C1}/\epsilon_S = 0.3437$. For C0, ESP-UA $\epsilon_{C0}/\epsilon_S = 0.4353$, is about 20% weaker than that of OPLS-UA: $\epsilon_{C0}/\epsilon_S = 0.5312$. To briefly summarize, the C1 parameterizations are close-but-not identical between ESP-UA and OPLS-UA, so we expect perylene simulations to be similar between the two models. However, because the ϵ_{C1}/ϵ_S and ϵ_{C0}/ϵ_S ratios and C0 σ values vary significantly between ESP-UA and OPLS-UA, it is unclear whether P3HT morphologies and phase behavior with ESP-UA will match those generated previously with OPLS-UA.

We perform MD simulations of each molecule to compare morphologies generated with the present ESP-UA parameterization against those previously generated with OPLS-UA. A summary of simulation state points and units for each molecule is provided in Table 4.2. These statepoints are chosen to replicate structures sampled in prior

Table 4.2: A list of statepoints used in the P3HT and perylene simulations, as well as the base units for each simulation (σ, ϵ, M).

	P3HT	Perylene
Temperature Range (K)	[60.4,304.4,427.7,608.9]	[49.8,248.8,497.7,746.5]
Density Range (g/cm ³)	[0.25,0.5,1.0]	[0.5,1.0,1.5]
N	100	250
dt	0.0003	0.0001
M (amu)	32.06	12.011
σ (Å)	3.56	3.40
ϵ (kJ/mol)	1.046	0.360
N _{monomers}	15	1

work from Miller et al.[5]. Each temperature range encompasses the solid-liquid transition temperatures of approximately 500 K and 550 K for P3HT and perylene, respectively. The density ranges are chosen with respect to previous MD studies as well as experimental thin-film studies for P3HT[129]. We expect to observe various phases (ordered, liquid, vapor, disordered) over this range of statepoints.

4.5.2 Perylene

The ordering (Ψ) of perylene modeled with ESP-UA as a function of temperature and density is summarized in Figure 4.6. Consistent with Ref. [5], the most ordered structure ($T = 250$ K, $\rho = 0.5$ g/cm³) exhibits significant π -stacking, which is visible in Figure 4.7. At lower temperatures, we observe a transition from highly ordered to less ordered as the density increases, also in agreement with prior work. This is due to the steric hindrance of having a high number of molecules in a restricted volume is observed in prior work. However, the order-disorder transition temperature observed here around 470 K is roughly 0.79 that of the 600 K in Ref. [3] and Botoshansky et al[130]. This is explained by the difference in base unit between Espaloma and OPLS-UA perylene forcefields: Both models have only C1 atomtypes but $\epsilon_{OPLS-UA} =$

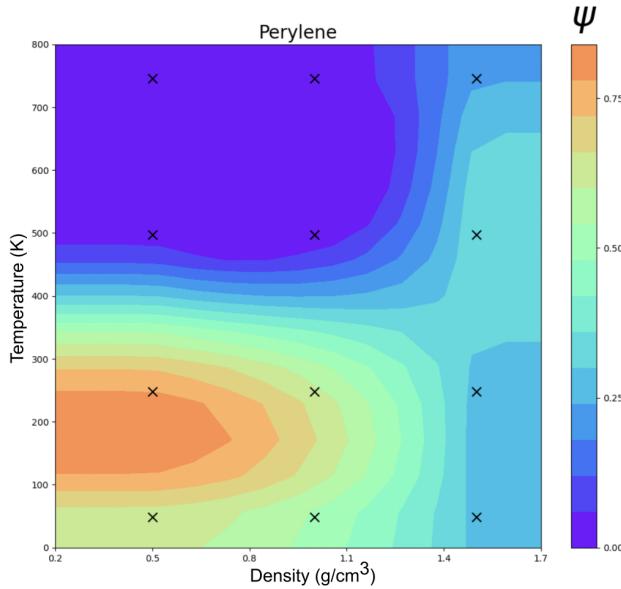


Figure 4.6: Temperature vs density clustering order parameter (Ψ) phase diagram of perylene at 12 statepoints.

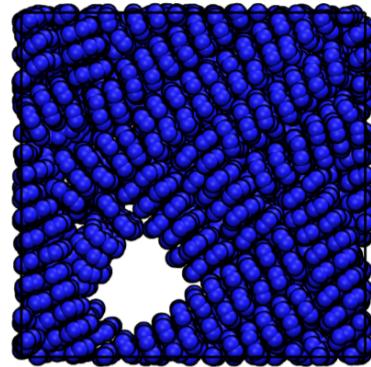


Figure 4.7: Snapshot of perylene taken from the most ordered morphology at a density of $0.5 \text{ g}/\text{cm}^3$ and temperature of 248.8 K .

0.4602 kJ/mol and $\epsilon_{ESP-UA} = 0.3635 \text{ kJ/mol}$. The ratio between these two energy units $\frac{\epsilon_{ESP-UA}}{\epsilon_{OPLS-UA}} = 0.79$ matches the transition temperature discrepancy. That is, in dimensionless units, there is no significant difference in the temperature-density phase diagrams between the two forcefields. Given that the energy unit of OPLS-UA gives an

order-disorder transition temperature in agreement with experiments, this suggests that the absolute ESP-UA energy units may benefit from empirical rescaling.

A comparison to the RDFs reported in Miller, et al. at density of 1.7 g/cm³ and the present work shows good agreement in short-range ordering as a function of temperature. First peak locations for the ordered systems in the ESP-UA and OPLS-UA RDF's align at 3.8 Å. First peak locations in the disordered systems produced by ESP-UA and OPLS-UA both appear at approximately 4 Å. Differences in peak intensity can be explained by the difference in density between the two systems. As expected, the RDFs calculated at higher temperatures show less intense peaks. We observe high ordering at 50 K and 249 K, in correspondence to the RDF reported by Miller, et al. We observe consistent RDF peak locations between the ESP-UA simulation results at temperature of 498 K and the OPLS-UA simulation results reported in the droplet phase by Miller, et al.

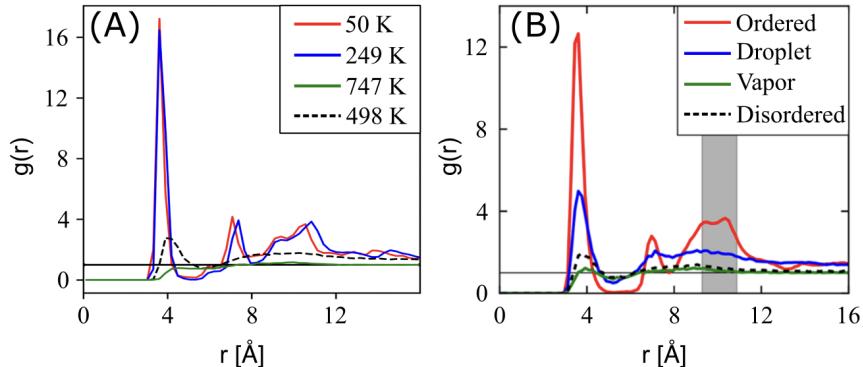


Figure 4.8: (A) Radial distribution function (RDF) of Perylene at various temperatures at a density of 0.5g/cm³ generated from an ESP-UA predicted morphology. (B) RDF of perylene in various phases generated from an OPLS-UA predicted morphology. OPLS-UA RDF published by Miller, et al. at Ref [3].

Long-range order is measured via simulated GIXS at temperature of 250 K and

density of 0.5 g/cm^3 . Bragg reflections are observed along both the x and y axes, indicating significant long-range order and close packed columns, which are observed in Figure 4.7. One major difference between the ESP-UA generated GIXS pattern here and the OPLS-UA pattern in prior work is the q_y location of the 001 reflection: From OPLS-UA the 001 reflection occurs around 0.8\AA^{-1} , while with ESP-UA here $q_y = 1.9\text{\AA}^{-1}$ in better agreement with Ishii and Miyasaka[131]. To briefly summarize,

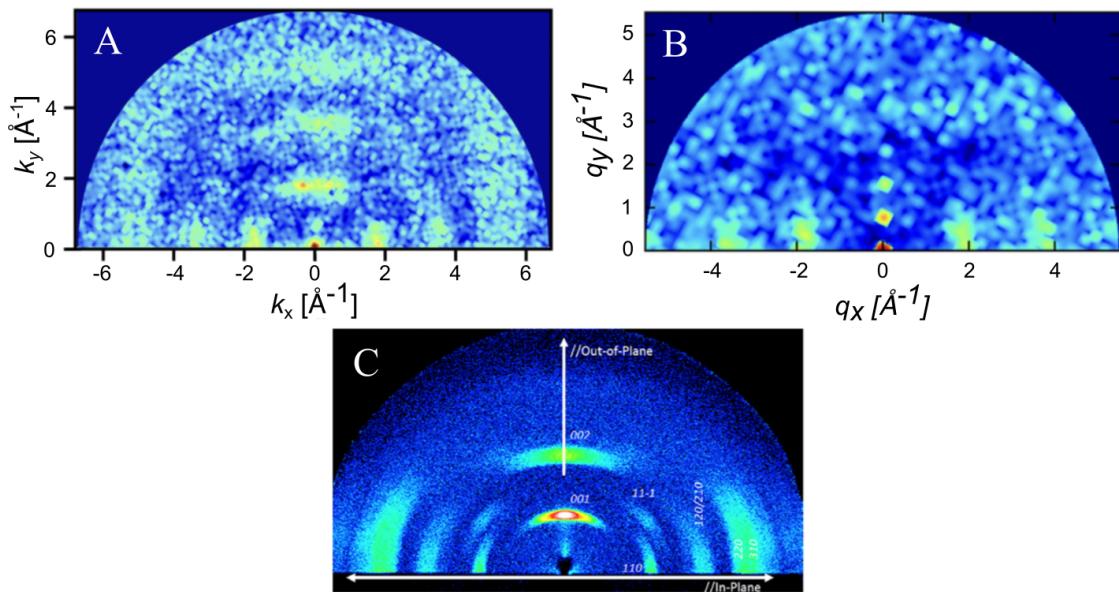


Figure 4.9: (A) Grazing incident x-ray scattering pattern of Perylene generated from an ESP-UA morphology. (B) GIXS pattern of Perylene generated from an OPLS-UA morphology. OPLS-UA GIXS pattern published by Miller, et. al. at Ref [3]. (C) Experimental XRD pattern for β -perylene, reproduced with permission from Ishii, et al. at Ref [4]. Copyright 2014 AIP Publishing LLC.

modeling perylene with ESP-UA parameterization results in phase behavior and short-range ordering in agreement with prior OPLS-UA work, however long-range ordering as measured by GIXS matches experiments better, while the base energy unit of

OPLS-UA gives better temperature correspondence.

4.5.3 Poly-3-Hexylthiophene (P3HT)

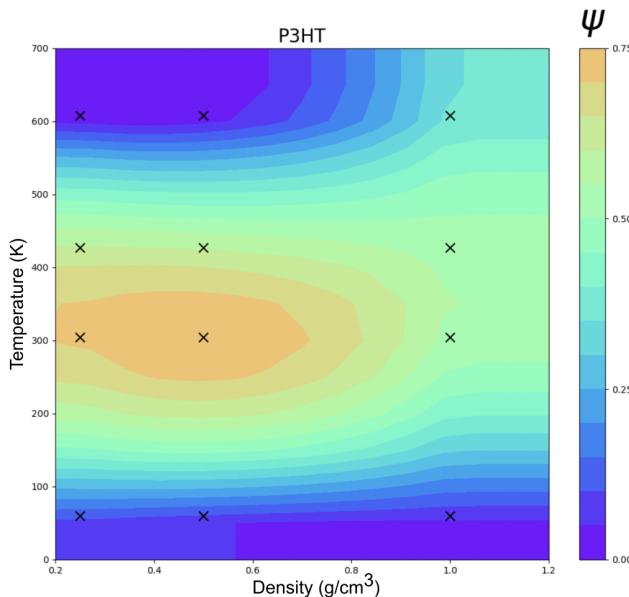


Figure 4.10: Temperature vs density clustering order parameter (Ψ) phase diagram of P3HT at 12 statepoints.

Figure 4.10 shows the clustering order parameter (Ψ) phase diagram of P3HT as a function of temperature and density. As observed with perylene, the transition temperatures are too low relative to experiments, and this is again explained by the differences in ϵ units between ESP-UA and OPLS-UA. ESP-UA gives an ϵ of 0.4554 kJ/mol for the aliphatic side chain carbons of P3HT, while OPLS-UA uses an ϵ of 0.7113 kJ/mol. This creates more ordered side chains in the OPLS-UA simulations in comparison to the ESP-UA simulations. This observation is supported by the work in Reference [132], stating that lower the ϵ values for side chains of P3HT monomers results in lamellar structures at lower temperatures than if the ϵ was higher [132]. Figure 4.11 shows the radial distribution function of ESP-UA P3HT (left) and OPLS-UA P3HT

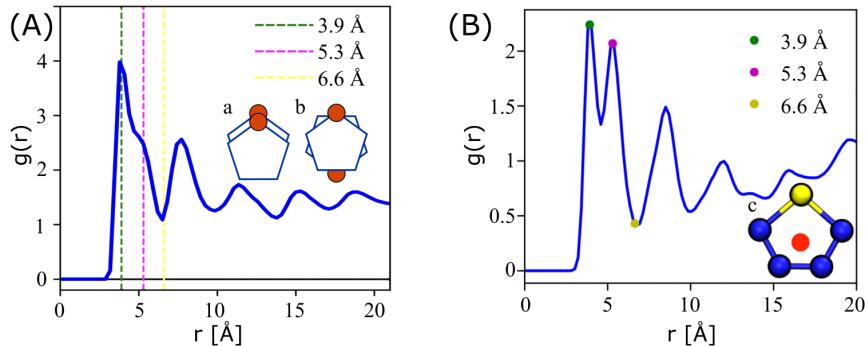


Figure 4.11: (A) Radial distribution function (RDF) of P3HT at a temperature of 304 K and a density of 0.5g/cm³ generated from an ESP-UA predicted morphology. (B) RDF of P3HT generated from an OPLS-UA predicted morphology. OPLS-UA RDF published by Miller, et al. in Ref [5].

(right). The ESP-UA RDF is generated at a temperature of 304 K and a density of 0.5g/cm³. Three vertical dashed lines in the ESP-UA RDF correspond to the local maxima and minima highlighted in the OPLS-UA RDF. The first peak in each corresponds to the aligned π -stacking of the thiophene rings, shown in Figure 4.11a. The second peak aligns with the anti-aligned π -stacking of the thiophene rings, shown in Figure 4.11b. The OPLS-UA RDF is calculated using the geometric center of thiophene ring, shown in Figure 4.11c. The ESP-UA RDF is calculated using only the S-S interactions, excluding those in the same chain. The sulfurs were chosen for the ESP-UA RDF because they are central to the ring, hold the most mass and have the largest radius.

The GIXS scattering pattern of the most ordered structure (0.5 g/cm³, 304 K) displays significant correlation to the experimental scattering pattern of P3HT (Figure 4.12). Peaks are observed at approximately 1.65 Å⁻¹ corresponding to the (010) plane, as well as peaks along the (100) plane spaced approximately 0.3 Å⁻¹ apart. This correlation between experiment and simulation confirms that the ESP-UA is able

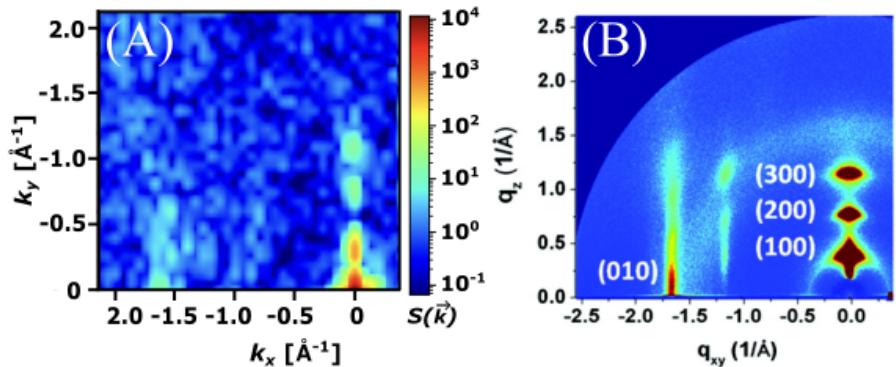


Figure 4.12: (A) Grazing Incident X-ray Scattering Pattern of P3HT at 0.5 g/cm^3 and 304 K generated using ESP-UA. (B) Corresponding experimental scattering pattern of P3HT. Reprinted (adapted) with permission from Ko, et al. in Ref [6]. Copyright 2012 American Chemical Society.

to responsibly parameterize and represent the P3HT polymer in simulations. The lamellar structure represented by the scattering pattern in Figure 4.12 is shown in Figure 4.13.

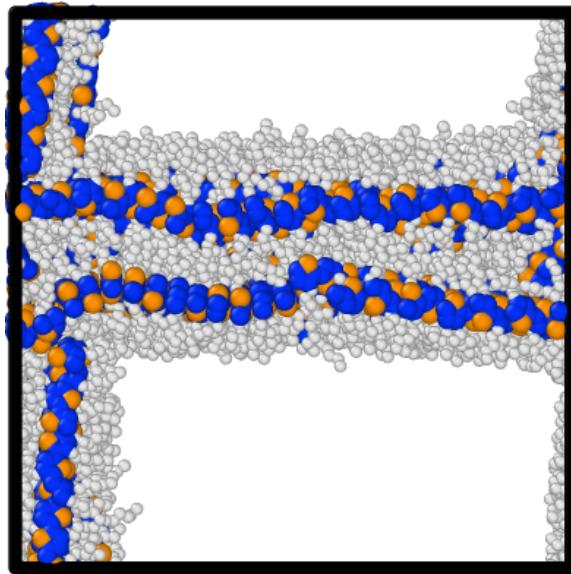


Figure 4.13: Snapshot of P3HT's most ordered morphology at a density of 0.5 g/cm^3 and temperature of 304 K .

4.6 Conclusions

We successfully incorporate **Espaloma** forcefield generation into MoSDeF workflows for investigating organic molecule phase behavior via MD simulations. **Espaloma** generates reasonable forcefield parameters for macromolecules with high aromaticity as well as for thiophene-based conjugated polymers as measured by GIXS, RDF, and phase behavior. The GIXS scattering patterns for both perylene and P3HT showed long range order that was consistent with published experimental work, and in the case of perylene has better agreement than **OPLS-UA** in matching the length scale of the 100 reflection. One caveat is that the absolute energy units generated by **ESP-UA** may be too low for the molecules of interest. Here we find that multiplying **ESP-UA** ϵ values by 1.26 to be in closer agreement with **OPLS-UA** gives better agreement for P3HT and Perylene’s experimental phase transitions between ordered and disordered phases. Whether this rescaling rule is universal or is dependent upon the specific chemistries remains the subject of future work. Nevertheless, the present observations inform confidence in using **Espaloma** to quickly generate forcefields for molecules that are missing information in “off-the-shelf” forcefields like **OPLS-UA** and **GAFF**, subject to the energy-rescaling caveat above. We conclude that **Espaloma** holds promise as a component in high-throughput screenings of organic molecules for phase behavior in systems from polymer thermoplastics to organic photovoltaics to macromolecular drug packing and more.

CHAPTER 5:

TEMPERATURE DEPENDENT CHAIN RIGIDITY STUDY OF ORGANIC PHOTOVOLTAIC DONOR-ACCEPTOR CO-POLYMERS.

The following chapter contains yet to be published work written by me with the guidance of Dr. Eric Jankowski.

5.1 Summary

Donor-acceptor (DA) co-polymers are currently a interesting area of study because of their promise in improving the field of optoelectronics because of their tunable band gaps. Applying these polymers in photovoltaic devices has proven to increase the power conversion efficiency of the device. In this study, we implemented molecular dynamics simulations to investigate the rigidity of complex organic polymers. Utilizing the previously validated workflow (Chapter 4) we built and parameterized 13 conjugated, organic polymers and predicted the bulk morphologies over a large temperature range. From the morphologies we employed a new function to calculated the persistence length l_p^{MD} of each polymer throughout their trajectories. We compare the results to experimental persistence lengths calculated via small angle

neutron scattering, l_p^{SANS} , and to theoretical results calculated via density functional theory, l_p^{DFT} . We found that for most of the polymers there was good agreements between the l_p^{MD} and l_p^{SANS} . From this information we are able to draw conclusions about how differences in side chains and backbone chemistry effects the rigidity of DA co-polymers.

5.2 Introduction

The morphology that OPV polymers self-assemble into has a significant impact on the power conversion efficiency (PCE) of the solar devices that are made from these materials. Currently, there is some ambiguity as to why certain morphologies lead to an increase in the PCE as well as what materials might self-assemble into these more efficient morphologies. Drawing connections between the physical and/or chemical properties of polymers to higher efficiency morphologies will enable the discovery of new materials that could increase the PCE of OPV devices. Donor-acceptor (DA) polymers are of a great interest in OPV device design due to their tunable band-gaps and high charge carrier mobilities [133]. DA polymers function as both the electron donor and electron acceptor by alternating electron rich and electron de-

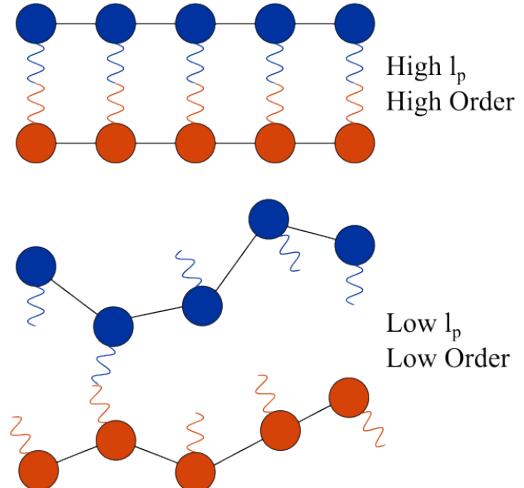


Figure 5.1: Diagram of how a more rigid polymer (high l_p) can lead to higher order in the morphology compared to a more flexible polymer (low l_p).

ficient environments. DA polymers are so exciting because we can in theory tune their optoelectronic properties by altering the donors and acceptors as well as their sequencing, all while determining the thermodynamic stability using MD simulations. The persistence length (l_p) of a polymer is used to describe the polymer's rigidness, a smaller l_p correlates to a more flexible polymer. The l_p is one measure of both thermodynamic stability as well as a prediction of how well the polymer might order in a bulk morphology. The l_p is a characteristic of polymer chains in a given environment that gives a lot of insight into how the polymer(s) interact and self-assemble into given morphologies. In some cases, the higher the persistence length of a polymer, the higher the observed order of the polymer bulk-morphology will be [134]. Previous studies have found that the straighter a polymer with optoelectronic properties is, the greater the probability of π -stacking of the backbones, leading to a greater charge carrier mobility in the bulk morphology [135]. Danielsen et al. has drawn a connection between the rigidity of DA copolymers and their molecular structure and side chain size via small angle neutron scattering (SANS). With the abundance of possible donors and acceptors and infinite possible combinations it is impossible to experimentally synthesize and test many DA polymers. Due to this we turn toward molecular dynamics simulations. In this study, we present a computational workflow of modeling and predicting the persistence length of DA polymers. We validate this workflow with 13 OPV polymers by comparing the persistence lengths calculated via MD (l_p^{MD}) to those presented by Danielsen, et al. using DFT (l_p^{DFT}) and SANS (l_p^{SANS}).

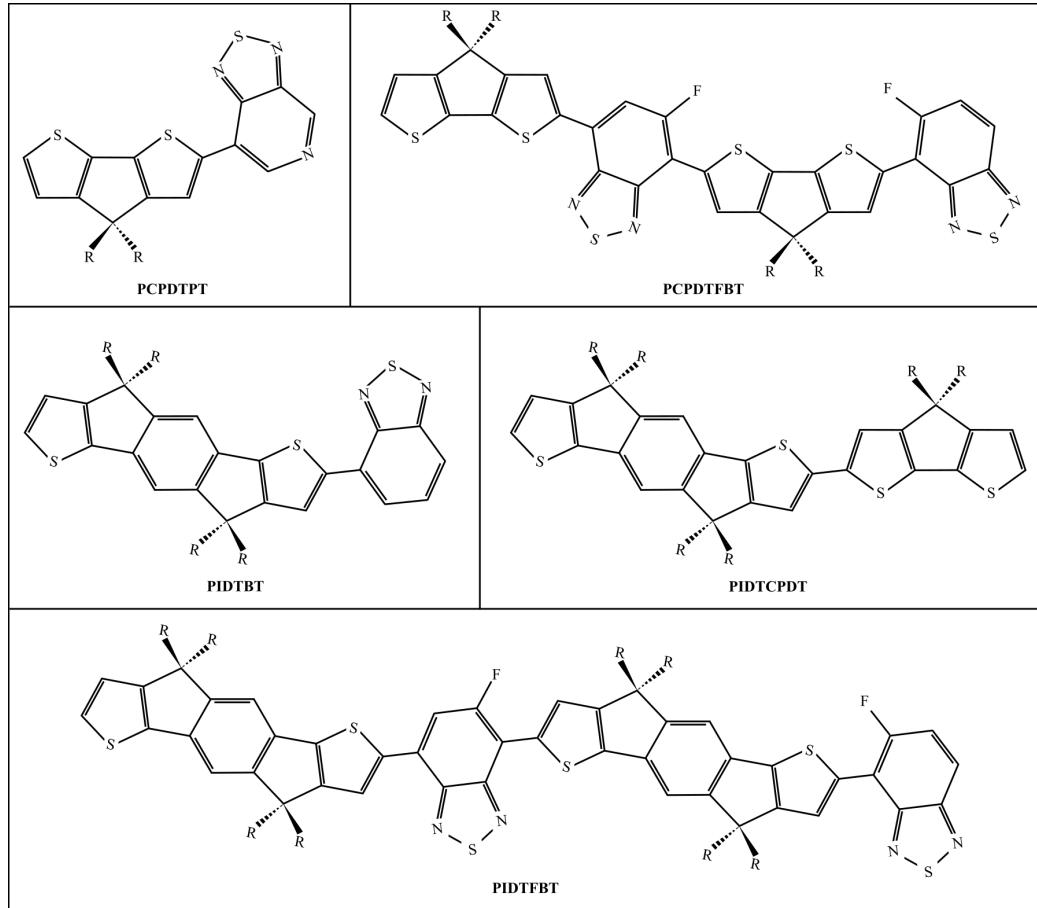


Figure 5.2: Molecular structures of the five repeating units used to build the polymers studied in this chapter. Each repeating includes a donor and acceptor region.

The monomer fragments used to create the monomers in this study are cyclopentadithiophene (CPDT), pyridalthiadiazole (PT), benzothiadiazole (BT), fluorobenzothiadiazole (FBT), and indacenodithiophene (IDT). Each repeating unit was built to exhibit both donor and acceptor regions. Figure 5.2 depicts how the repeating units, or monomers were built from the monomer fragments listed above. The polymers with PIDTBT, PCPDTPPT and PIDTCPDT backbones have an A-B-A-B polymerization pattern. PCPDTFBT and PIDTFBT polymers also consider the F functional group location, altering between a meta and an ortho position. The polymerization pattern for PIDTFBT and PCPDTFBT is A-B_m-A-B_o, where B_m is meta-FBT and B_o is ortho-FBT. In this study, we investigate the effect of backbone chemistry as well as side chain complexity on the rigidity of the polymer, the side-chains are shown in Figure 5.3. Calculating the persistence length is not only a way to gain more of an understanding of polymer physics, but is also another means to validate the workflow developed in Chapter 4.

We are using an implicit solvent model in this study, so we utilize temperature to mimic solvent effects, to justify this we must dive into polymer theory. We first need to define an ideal chain. An ideal chain is a model in polymer theory where monomer units along the polymer have no interactions. In a real chain, interactions would be observed between monomers as well as between the polymer and solvent. The theta temperature (Θ_T) is the temperature at which a real polymer behaves as an ideal chain [136, 137]. Similarly, the theta solvent (Θ_S) is the solvent in which the polymer behaves as an ideal chain. Because we are modeling our solvent implicitly, we use the temperature to tune the ideality of our polymer chains, therefore mimicking how the solvent would effect chain behavior in experiment. We calculate the temperature

at which we would expect the l_p^{MD} and l_p^{SANS} to match and compare this to the calculated Θ_T in order to draw conclusions on the solvent effects C_6D_5Cl has on each polymer in the experimental data.

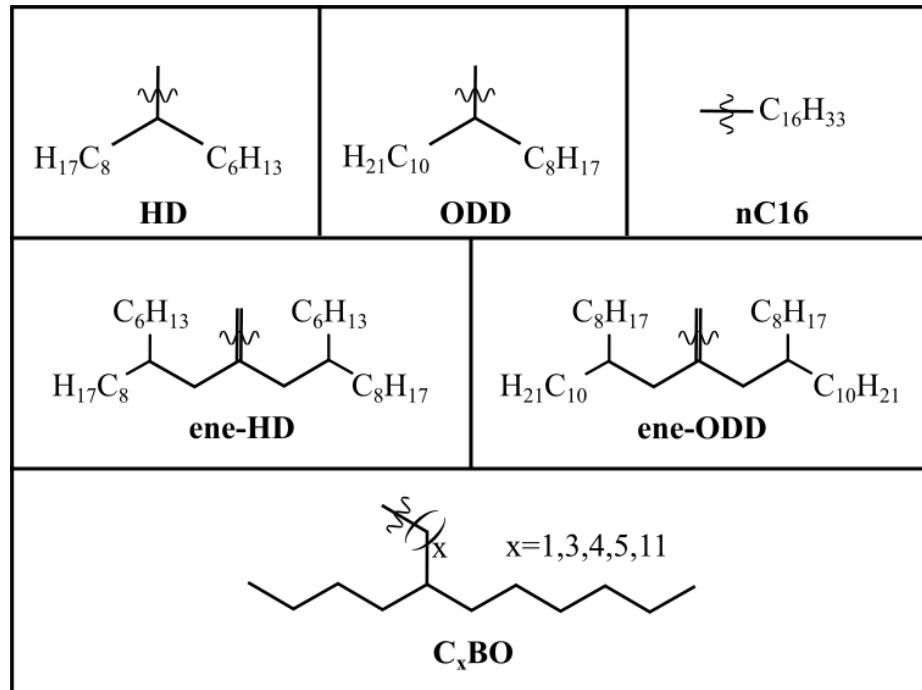


Figure 5.3: Molecular structures of the six R-groups used as side chains in this study.

5.3 Methods

In this study, we conducted molecular dynamics simulations using `Hoomd-Blue` to investigate the persistence length of 13 various polymers. Simulations were performed on the Fry high performance computing cluster at Boise State using NVIDIA P100 and V100 GPUs. Each simulation was conducted in the canonical ensemble (NVT), in which the number of particles, volume, and temperature are all held constant throughout the simulation. Particle positions and velocities were updated using a two-step velocity-verlet integration of Newton's laws of classical mechanics. The `FlowerMD` software package was utilized to construct our simulation object and initialize the simulations [123]. All simulations were initialized using the `Pack` functionality, in which the polymers are randomly placed in a cubic simulation box with periodic boundary conditions enforced. The simulation box size is calculated from the desired density, which is listed in Table C.1 for each polymer in Appendix C. Thirteen conjugated polymers with similar backbone structures and varying side chain lengths were simulated in this study. An `espaloma` forcefield was generated for each polymer using the workflow outlined in Chapter 4. All simulations were run using an uncharged, united-atom model by removing the partial charges of each particle and adding the masses of the hydrogens into the neighboring, bonded atom. For each simulation we initialized a polymer of 250 monomer units in a large simulation space. Simulations were conducted at temperatures of 252, 503, 1006, 1258, 1510, 1761, 2013 K to capture how the persistence lengths evolve with temperature. The high temperatures were used to investigate how the persistence length might vary at high temperatures, but we focus on temperatures 252 K - 1006 K for analysis because temperatures above this are unrealistic for experimental settings. Simulations were run for 5×10^7

timesteps at a timestep size, $dt = 0.001\tau$. $\tau = \sqrt{M * L^2 / E}$, where M, L, and E are the reference mass, length, and energy of the simulation, respectively. From this information, the each timestep was 1.97 fs, resulting in a 98 ns simulation.

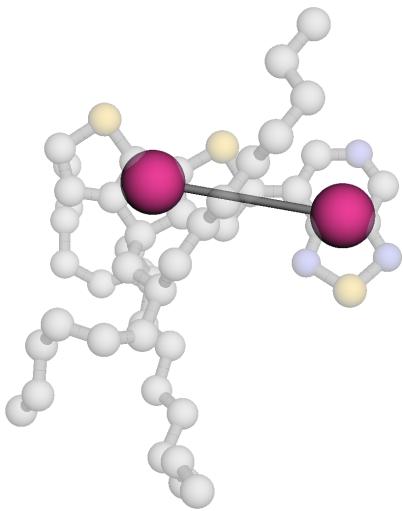


Figure 5.4: PCPDTPT-HD geometric centers used in calculating persistence length. The pink beads represent the monomer center of mass for the backbone molecules, CPDT on the left and PT on the right. The side chains were not included in the calculation of the center of mass.

at the center of mass of the input indices provided by the user. The geometric centers were calculated by adding a bead to each fully-rotatable unit to ensure that no sensitivity was lost in the persistence length calculation. The persistence length is determined to be the polymer length between the first monomer in the polymer chain

Once the simulations reached equilibrium and had at least 50 independent samples for each simulation the geometric center of masses were calculated using the GRiTS software [138]. This software takes an input of either the SMILES string for the portion of the molecule that you are replacing with a bead or the indices that should be replaced by a bead. In this study we used particle indices to manually determine where the beads should be placed rather than using SMILES string matching. In Figure 5.4 we show an example of how a PCPDTPT-HD monomer's center of mass was calculated and where the sphere was placed. The beads are placed

and the first fully decorrelated monomer unit. A depiction of this is shown in Figure 5.1. For this study, the persistence length was calculated by determining how each geometric center bead in the polymer correlates to the first bead at index 0. The correlation function ($C(n)$) is shown in Equation 5.1. This equation explains how the angle correlation of the n^{th} bead is equal to the averaged dot products of the first bond vector (\hat{a}_0) and the n^{th} bond vector (\hat{a}_{0+n}).

$$C(n) = \langle \cos \theta_{0,0+n} \rangle = \langle \hat{a}_0 \cdot \hat{a}_{0+n} \rangle \quad (5.1)$$

The exponential relationship of the $C(n)$ and n is shown in Equation 5.2. The exponential coefficient of the autocorrelation function can be used to determine the persistence length, l_p from the average bond length, l_b .

$$C(n) \approx \exp\left(\frac{-nl_b}{l_p}\right) \approx \langle \hat{a}_0 \cdot \hat{a}_{0+n} \rangle \quad (5.2)$$

Using these relationships, a python function was written to calculate l_p of a polymer from the predicted morphology. In this function the autocorrelation of the bond vectors in the polymer backbone was calculated for 50 independent samples. The autocorrelation of each decorrelated snapshot was fit using the `curve_fit` function in SciPy's `Optimize module` [7]. Using the average bond length of the backbone bonds and the exponential coefficient gained from fitting the autocorrelation function to an exponential decay, the l_p was calculated using the relationship in Equation 5.2. The l_p 's for each snapshot were then averaged together and the standard deviations were calculated using NumPy's `Statistics module` [139]. The full script for the persistence length function can be found at github.com/madilynpaul/persistence_length

as well as in Appendix C. By plotting the root mean square end-to-end distance, $\sqrt{\langle R^2 \rangle}$, versus the temperature of the simulation we were able to calculate the approximate theta temperature of each polymer. This was accomplished by determining the length at which the polymer will behave as an ideal chain, Equation 5.3.

$$\langle R^2 \rangle = N_b l_b^2 \quad (5.3)$$

We then calculated the $\sqrt{\langle R^2 \rangle}$ by taking the normalization of the end-to-end vector at each temperature and, using a linear regression, calculated the temperature at which the root mean square end-to-end distance was equivalent to $\sqrt{N_b} * l_b^2$, where N_b is the number of bonds in the polymer and l_b is the average bond length. The $\sqrt{\langle R^2 \rangle}$ vs T plots can be found in Appendix C.

5.4 Results and Discussion

We have successfully simulated 13 polymers with proven OPV properties. Each polymer consisted of 250 repeating units and was simulated in a large volume with no other molecules present. For most polymers we see an increase in persistence length

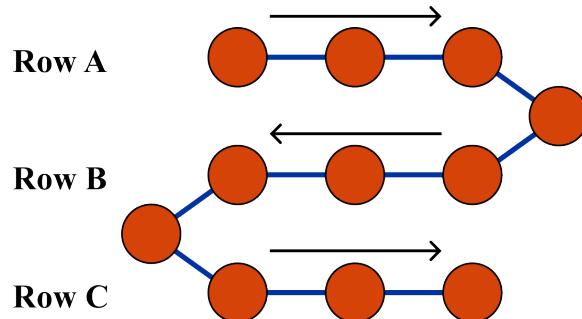


Figure 5.5: Diagram of a condensed polymer. Row A, B and C show the directions of the bond vectors. A snapshot of a condensed PCPDTPT-HD polymer can be found at Figure C.1.

as the temperature increases, until we reach a plateau in persistence length around 1000 K. This trend is expected, at lower temperatures we observe the polymers condensing and folding in on themselves (Figure 5.5), decreasing the persistence length. At temperatures above 1000 K the polymer has so much thermal energy it can easily overcome the steric hindrances leading to a more flexible polymer.

5.4.1 PCPDTPT Backbone

The l_p^{MD} 's for polymers with a cyclopentadithiophene-co-pyridalthiadiazole (PCPDTPT) backbone (Figure 5.2) are shown in Table 5.1. The lower l_p^{MD} 's observed at lower temperatures is expected and can be explained using the diagram in Figure 5.5. When the polymer condenses on itself the l_p decreases and non-bonded interactions between rows of the polymer (Rows A, B, and C) become more prevalent. We observe good alignment between the l_p^{MD} 's calculated over this temperature range and the l_p^{SANS} 's calculated at room temperature. While we observe reasonable l_p values, we do not observe the same trend in l_p with increasing side chain size/complexity as the l_p^{SANS} . The strong non-bonded interactions between side chains observed at low temperatures lead us to believe that having more than one polymer in the simulation is necessary for properly observing the side chain effects on backbone rigidity.

By fitting the l_p^{MD} with a linear regression we can predict what simulation temperature would give similar results to the SANS data. These linear regressions can be found in Appendix C. Due to solvent effects we expect the l_p^{SANS} values to align with different temperatures for each polymer. The simulation temperature at which we can predict similar l_p^{MD} values to the l_p^{SANS} values were predicted using a linear regression and can be found in Table 5.2. The backbone rigidity and temperature effects are not the only properties that can influence the persistence length. We can

Polymer	l_p			
	252 K	503 K	1006 K	SANS
PCPDTPT-ene-HD	39.8 ± 1.4	48.4 ± 1.2	110.7 ± 28	76.6
PCPDTPT-ene-ODD	21.9 ± 0.1	38.5 ± 1.1	97.8 ± 26	83.4
PCPDTPT-HD	36.8 ± 0.3	55.3 ± 9.3	100.4 ± 22	47.3
PCPDTPT-nC16	40.2 ± 0.7	60.3 ± 1.4	96.0 ± 22	61
PCPDTPT-ODD	56.4 ± 1.1	63.6 ± 2.6	99.6 ± 34	54.9

Table 5.1: Persistence lengths calculated from MD predicted morphologies (l_p^{MD}) of polymers with a PCPDTPT backbone in comparison to persistence lengths calculated from SANS data (l_p^{SANS}). All l_p presented in units of Å.

draw a connection between the solvent efficiency and the l_p . If the T_{SANS} values are greater than the Θ_T we can conclude that the experimental solvent (C_6D_5Cl) is a good solvent for the given polymer. C_6D_5Cl appears to be a good solvent for PCPDTPT-ene-HD, PCPDTPT-ene-ODD and PCPDTPT-n-C16, while it is a less than ideal solvent for PCPDTPT-HD and PCPDTPT-ODD. This comparison can be justified in the same manor that we explain the temperature effects on a polymer. If the polymer is solvated in a poor solvent it will condense into itself, and if it is a good solvent for the polymer we can expect that the polymer will extend into the solvent.

Polymer	T_{SANS} (K)	Θ_T (K)	T_{SANS}/Θ_T
PCPDTPT-ene-HD	692	502.5	1.38
PCPDTPT-ene-ODD	885	456.2	1.94
PCPDTPT-HD	389	507.0	0.767
PCPDTPT-nC16	526	488.3	1.08
PCPDTPT-ODD	278	450.5	0.616

Table 5.2: Predicted simulation temperature at which l_p^{MD} would match l_p^{SANS} for polymers with PCPDTPT backbone. T_{SANS} calculated via linear regressions shown in Appendix C. T_{SANS} reported in units of K.

5.4.2 PCPDTFBT Backbone

We observe similar trends in l_p^{MD} with temperature as seen in the polymers with PCPDPTP backbones as we do with cyclopentadithiophene-co- fluorobenzothiadiazole (PCPDTFBT) backbone polymers. In PCPDTFBT-C1-BO, however, we observe a higher l_p^{MD} at 503 K than at 1006 K. When calculating the l_p by fitting the auto-correlation to an exponential function we don't take into account the bond vectors that may be aligned with the initial bond vector, but that shouldn't be contributing to the persistence length. This is a correction that would need to be accounted for in future calculations. The uncertainty at 1006 K is large enough to incorporate the l_p^{MD} at 503 K, statistically stating that the values are essentially equal.

Polymer	252 K	503 K	1006 K	l_p^{SANS}
PCPDTFBT-C1-BO	37.8 ± 0.5	115.3 ± 1.9	99.7 ± 21	67
PCPDTFBT-C3-BO	47.5 ± 0.4	57.6 ± 2.8	98.9 ± 22	78.4
PCPDTFBT-C4-BO	46.3 ± 1.1	65.8 ± 3.1	105.1 ± 20	86.4
PCPDTFBT-C5-BO	35.4 ± 0.3	68.7 ± 1.9	104.3 ± 28	114
PCPDTFBT-C11-BO	34.6 ± 0.2	55.9 ± 2.6	104.0 ± 26	291

Table 5.3: Persistence lengths calculated from MD predicted morphologies (l_p^{MD}) of polymers with a PCPDTFBT backbone in comparison to persistence lengths calculated from SANS data (l_p^{SANS}). All l_p presented in units of Å.

The l_p^{MD} values match well with the l_p^{SANS} values for all polymers with the PCPDTFBT backbone besides that with the C11-BO side chain. The C11-BO side chain is much larger and bulkier than the other side chains, which could be contributing to the improper modeling. It could also be observed that PCPDTFBT-C11-BO may exhibit strong solvent interactions, in which solvent molecules would need to be present within the simulation to properly model the interactions. The fact that the

polymers fold in on themselves and exhibit side chain interactions at low temperatures leads us to believe that a simulation with multiple polymers could lead to increased persistence lengths due to side chain interactions.

Polymer	T_{SANS} (K)	Θ_T (K)	T_{SANS}/Θ_T
PCPDTFBT-C1-BO	326	406.5	0.801
PCPDTFBT-C3-BO	736	515.7	1.43
PCPDTFBT-C4-BO	767	568.5	1.35
PCPDTFBT-C5-BO	1091	491.9	2.22
PCPDTFBT-C11-BO	3034.7	528.1	5.75

Table 5.4: Predicted simulation temperature at which l_p^{MD} would match l_p^{SANS} for polymers with PCPDTFBT backbone. T_{SANS} calculated via linear regressions shown in Appendix C. T_{SANS} reported in units of K.

The predicted simulation temperature at which the l_p^{MD} and l_p^{SANS} will match, T_{SANS} can be found in Table 5.4, along with a comparison of the Θ_T . The T_{SANS} increases with increasing side chain size. C_6D_5Cl appears to be a good solvent for all PCPDTFBT polymers besides PCPDTFBT-C1-BO, as $T_{SANS} < \Theta_T$. The values of T_{SANS} are reasonable for all PCPDTFBT backbone polymers besides PCPDTFBT-C11-BO, with a value of 3034.7 K. In our MD model we observe a plateau in the persistence length at temperatures above 1200 K. The experimental values for persistence length of PCPDTFBT-C11-BO are larger than the greatest persistence lengths we observe. Investigations into our model as well as the model being used to fit the SANS data to calculate the persistence length are necessary in order to gain insight into the origin of these differences.

5.4.3 IDT Backbone

The final DA co-polymers we investigated were ones with indacenodithiophene (IDT) as the donor molecule and three different acceptor molecules, CPDT, FBT, and

BT. We successfully simulated these rather large co-polymers with little issues. The timesteps per second (TPS) for these simulations were significantly lower than the other polymer simulations due to the increased complexity and size of the polymer backbones. We have predicted reasonable values for l_p^{MD} , we observe an increase

Polymer	l_p			
	252 K	503 K	1006 K	SANS
PIDTBT-nC16	65.4 ± 0.2	57.8 ± 1.6	200.6 ± 26	1310
PIDTCPDT-C11-BO	36.0 ± 0.2	76.4 ± 8.7	182.2 ± 49	236
PIDTFBT-C11-BO	79.4 ± 3.9	90.6 ± 23	178.4 ± 32.7	254

Table 5.5: Persistence lengths calculated from MD predicted morphologies (l_p^{MD}) of polymers with a IDT backbone in comparison to persistence lengths calculated from SANS data (l_p^{SANS}). All l_p presented in units of Å.

in l_p^{MD} with the increase in temperature, with only one outlier of PIDTBT-nC16 at 503 K, Table 5.5. We also observe good overlap between the l_p^{MD} and l_p^{SANS} for the PIDTFBT-C11-BO and PIDTCPDT-C11-BO polymers. Molecular dynamics predicts a significantly lower l_p^{MD} for PIDTBT-nC16 than the presented l_p^{SANS} . However, the l_p^{SANS} for all three of these polymers is reported as longer than the contour length (L_c) of the polymer in experiment. This makes the direct comparison of l_p^{MD} and l_p^{SANS} ineffective. Both the experimental and DFT values presented for these three co-polymers came with some caveats, leading us to believe that the method of analysis used in calculating l_p^{DFT} and l_p^{SANS} breaks down with larger, more complex polymers requiring additional steps to predict the l_p^{DFT} . For all three IDT polymers the l_p^{SANS} reported is longer than the contour length of the polymer, due to the $l_p^{SANS} \gtrsim L_c$ fitting the l_p^{SANS} into the linear regression of l_p^{MD} results in very high, inaccurate T_{sans} values. Again, due to the plateau observed in l_p^{MD} at temperatures above 1200 K, we cannot confidently say that simulating these polymers at the T_{SANS}

temperature will result in $l_p^{MD} \approx l_p^{SANS}$. Further investigation into both theoretical and experiential results are necessary to draw conclusions.

Polymer	T_{sans} (K)	Θ_T (K)	T_{SANS}/Θ_T
PIDTB-nC16	6785	411.7	16.5
PIDTCPDT-C11-BO	1290	511.1	2.52
PIDTFBT-C11-BO	1591	405.0	3.93

Table 5.6: Predicted simulation temperature at which l_p^{MD} would match l_p^{SANS} for polymers with IDT backbone. T_{SANS} calculated via linear regressions shown in Appendix C. T_{SANS} reported in units of K.

5.4.4 MD vs DFT

Polymer	l_p^{DFT}	l_p^{MD}	l_p^{SANS}
PCPDTFBT-C1-BO	71.1	99.7 ± 21	67
PCPDTFBT-C3-BO	71.1	98.9 ± 22	78.4
PCPDTFBT-C4-BO	71.1	105.1 ± 20	86.4
PCPDTFBT-C5-BO	71.1	104.3 ± 28	114
PCPDTFBT-C11-BO	71.1	104.0 ± 26	291
PCPDTPT-ene-HD	48.0	110.7 ± 28	76.6
PCPDTPT-ene-ODD	48.0	97.8 ± 26	83.4
PCPDTPT-HD	44.9	100.4 ± 22	47.3
PCPDTPT-nC16	44.9	96.0 ± 22	61
PCPDTPT-ODD	44.9	99.6 ± 34	54.9
PIDTB-nC16	($>1 \mu\text{m}$)	200.6 ± 26	1310
PIDTCPDT-C11-BO	224	182.2 ± 49	236
PIDTFBT-C11-BO	($>1 \mu\text{m}$)	178.4 ± 33	254

Table 5.7: A comparison of l_p^{DFT} reported by Danielsen, et al. at Ref [2] to l_p^{MD} calculated at a temperature of 1006 K.

The results reported by Danielsen, et al. for l_p^{DFT} are presented in Table 5.7 as well at the l_p^{MD} and the l_p^{SANS} [2]. Dihedral potentials were calculated using optimized scans of the central dihedral angle (ϕ) of each monomer unit [140]. To minimize computational cost monomer units were utilized rather than polymers and side chains

were replaced by methyl groups. A numerical average of backbone conformations was estimated from a set of dihedral angles, from this the l_p^{DFT} was approximated. In the DFT approach accuracy is lost in modeling all side chains as methyl groups and only modeling one monomer unit. The DFT approach also does not model temperature or solvents, whereas MD simulations are able to model long polymers, complex side chains as well as temperature and can be easily modified to model explicit solvents.

5.5 Conclusions

The molecular dynamics approach presented here has a significantly lower computational cost, allowing for the simulation of long oligomers with full side chains. Both these factors add to our confidence in the accuracy of the persistence length calculations. We observe that molecular simulations can predict side-chain influences on persistence length. The calculated l_p^{MD} show good agreement to the l_p^{SANS} for the majority of the polymers in this study in terms of magnitudes and trends. From this information we can conclude that the MD workflow presented in Chapter 4 is accurately predicting bulk morphologies of DA-polymers. From this work we conclude that l_p is a relatively simple measure of the thermodynamic stability of a polymer as well as a prediction of the bulk morphology and we present a MD workflow that accurately predicts and calculates the l_p of DA co-polymers.

CHAPTER 6:

CONCLUSIONS

Over the course of this dissertation we have tied together two very different, but theoretically similar projects. Beginning with attempting to model the polymorphic phase transition of barbituric acid dihydrate, we discovered that what was reported to be a very subtle phase transition at 217 K was actually the presence of crystalline disorder. We aimed to model this phase transition via density functional theory (DFT) due to the present of hydrate molecules in the crystal structure, but when two equivalent energy well were observed in the potential energy scan of the structure at temperatures above 217 K it became clear that this phase transition might not exist. This paired with the widening of the peaks in the terahertz spectra, lead us to conclude that there was twinning within the crystalline structure. Further investigation into this revealed that above 217 K the molecules had enough energy to oscillate between the two energetically favorable configurations. When this structure was analyzed using x-ray crystallography the oscillating molecule's positions were averaged out into a perfectly planar configuration, leading the investigators to report that the molecules were confined to a *Pnma* space group. Upon finding this crystal disorder we became skeptical of the planar configuration reported for violuric acid monohydrate because of the similarities in structure between the two compounds. When investigating VAMH a double-welled potential and imaginary mode lead us to prematurely conclude that

VAMH had a space group of $P2_1$ rather than $Cmc2_1$. When evaluating the terahertz spectra of VAMH generated in the $P2_1$ space group to the experimental there was not good peak alignment. This led us to increase the size of the unit cell and remove all symmetry operations, allowing the molecules to freely optimize. The molecules were then perceived to belong to the orthorhombic $Pca2_1$ space group. In this space group the double-welled potential disappeared and the experimental spectra aligned well with the theoretical spectra. The second half of this dissertation focused on developing tools for structure determination via molecular dynamics. With a focus on organic macromolecules and polymers we developed a workflow that implemented Espaloma generated forcefields into the molecular dynamics framework. It was necessary to validate the accuracy of the Espaloma forcefields for conjugated, organic molecule before using the workflow to predict morphologies of these kinds of materials. To validate the workflow we predicted the morphologies of two widely studied organic molecules, perylene and poly-3-hexylthiophene (P3HT). We found that Espaloma produces similar forcefield parameters to the generalized amber forcefield (GAFF). The predicted morphologies of both materials with the espaloma forcefield showed great overlap with those predicted using the OPLS-UA forcefield. Once the workflow was validated for this use we utilized it to predict the morphologies of 13 conjugated donor-acceptor co-polymers over a range of temperatures. A function was written to calculate the persistence length from the predicted morphologies of the 13 polymers. This function was validated against persistence lengths calculated via DFT and small angle neutron scattering patterns.

REFERENCES

- [1] Gary S. Nichol and William Clegg. A variable-temperature study of a phase transition in barbituric acid dihydrate. *Acta Crystallographica: Section B (Wiley-Blackwell)*, 61(4):464 – 472, 2005.
- [2] Scott P. O. Danielsen, Colin R. Bridges, and Rachel A. Segalman. Chain stiffness of donor–acceptor conjugated polymers in solution. *Macromolecules*, 55(2):437–449, 2022-01-25.
- [3] Evan D. Miller, Matthew L. Jones, and Eric Jankowski. Enhanced computational sampling of perylene and perylothiophene packing with rigid-body models. *ACS Omega*, 2(1):353–362, 2017-01-31.
- [4] Ayumi Ishii, Ajay Kumar Jena, and Tsutomu Miyasaka. Fully crystalline perovskite-perylene hybrid photovoltaic cell capable of 1.2 v output with a minimized voltage loss. *APL Materials*, 2(9):091102, 2014.
- [5] Evan Miller, Matthew Jones, Michael Henry, Paul Chery, Kyle Miller, and Eric Jankowski. Optimization and validation of efficient models for predicting polythiophene self-assembly. *Polymers*, 10(12):1305, 2018-11-26.
- [6] Sangwon Ko, Eric T. Hoke, Laxman Pandey, Sanghyun Hong, Rajib Mondal, Chad Risko, Yuanping Yi, Rodrigo Noriega, Michael D. McGehee, Jean-

- Luc Brédas, Alberto Salleo, and Zhenan Bao. Controlled conjugated backbone twisting for an increased open-circuit voltage while having a high short-circuit current in poly(hexylthiophene) derivatives. *Journal of the American Chemical Society*, 134(11):5222–5232, 2012.
- [7] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.
- [8] Aurora J. Cruz-Cabeza, Susan M. Reutzel-Edens, and Joel Bernstein. Facts and fictions about polymorphism. *Chemical Society Reviews*, 44(23):8619–8635, 2015.
- [9] Alexandre Tkatchenko Johannes Hoja, Anthony M. Reilly. First-principles modeling of molecular crystals: structures and stabilities, temperature and pressure. *WIREs computational molecular science*, 2016.
- [10] James Swarbrick Harry G. Brittain. *Polymorphism in Pharmaceutical Solids: Second edition*, volume 192. CRC Press, 2 edition, 2016.

- [11] Joel Bernstein. Polymorphism a perspective. *Crystal Growth & Design*, 11(3):632–650, 2011-03-02.
- [12] Denis Gentili, Massimo Gazzano, Manuela Melucci, Derek Jones, and Massimiliano Cavallini. Polymorphism as an additional functionality of materials for technological applications at surfaces and interfaces. *Chemical Society Reviews*, 48(9):2502–2517, 2019.
- [13] Kyle A. Baseden and Jesse W. Tye. Introduction to density functional theory: Calculations by hand on the helium atom. *Journal of Chemical Education*, 91(12):2116–2123, 2014-12-09.
- [14] Madilyn E. Paul, Thiago H. da Silva, and Matthew D. King. True polymorphic phase transition or dynamic crystal disorder? an investigation into the unusual phase behavior of barbituric acid dihydrate. *Crystal Growth & Design*, 19(8):4745–4753, 2019-08-07.
- [15] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Physical Review*, 136(3):B864–B871, 1964-11-09.
- [16] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140(4):A1133–A1138, 1965-11-15.
- [17] Teffanie Goh. Exchange-correlation functionals: GGA vs. meta-GGA. *N/A*, 2020.
- [18] Philip J. Hasnip, Keith Refson, Matt I. J. Probert, Jonathan R. Yates, Stewart J. Clark, and Chris J. Pickard. Density functional theory in the solid state.

- Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 372(2011):20130270, 2014-03-13.
- [19] A. Arbuznikov. Hybrid exchange correlation functionals and potentials: Concept elaboration. *Journal of Structural Chemistry*, 48:S1–S31, 12 2007.
 - [20] N L Heda, Hukmi Chand Suthar, Kishor Kumar, Ushma Ahuja, and B L Ahuja. Assessment of hybrid density functional in revealing electronic response and half-metallic character of tm_2O_3 ($\text{tm} = \text{ti, v and cr}$) and their compton spectroscopy. *Bulletin of Materials Science*, 46(4):1 – 13, 2023.
 - [21] Katherine A. Mazzio and Christine K. Luscombe. The future of organic photovoltaics. *Chemical Society Reviews*, 44(1):78–90, 2015.
 - [22] Weiwei Sun, Haiyang Chen, Ben Zhang, Qinrong Cheng, Heyi Yang, Ziyuan Chen, Guang Zeng, Junyuan Ding, Weijie Chen, and Yaowen Li. host-guest active layer enabling annealing-free , non-halogenated green solvent processing for high-performance organic solar cells †. *Chinese Journal of Chemistry*, 40(24):2963–2972, 2022-12-15.
 - [23] Owen D. Miller, Eli Yablonovitch, and Sarah R. Kurtz. Strong internal and external luminescence as solar cells approach the shockley–queisser limit. *IEEE Journal of Photovoltaics*, 2(3):303–311, 2012-07.
 - [24] Ruvini S. Kularatne, Harsha D. Magurudeniya, Prakash Sista, Michael C. Biewer, and Mihaela C. Stefan. Donor–acceptor semiconducting polymers for

- organic solar cells. *Journal of Polymer Science Part A: Polymer Chemistry*, 51(4):743–768, 2013-02-15.
- [25] William D. Callister and David G. Rethwisch. *Materials science and engineering: an introduction*. John Wiley & Sons, 8th ed edition, 2010. OCLC: ocn401168960.
- [26] Renata Lopes Reis and Demétrio A. da Silva Filho. The effect of positional disorder and the beer-lambert law in organic photovoltaics: A kinetic monte carlo simulation analysis. *Journal of Molecular Modeling*, 28(10):330, 2022-10.
- [27] Wei Zhang, Jefferson Maul, Diana Vulpe, Peyman Z. Moghadam, David Fairen-Jimenez, Daniel M. Mittleman, J. Axel Zeitler, Alessandro Erba, and Michael T. Ruggiero. Probing the mechanochemistry of metal–organic frameworks with low-frequency vibrational spectroscopy. *The Journal of Physical Chemistry C*, 122:27442–27450, 12 2018.
- [28] Hugo Gaspar, Flávio Figueira, Luiz Pereira, Adélio Mendes, Júlio Viana, and Gabriel Bernardo. Recent developments in the optimization of the bulk heterojunction morphology of polymer: Fullerene solar cells. *Materials*, 11(12):2560, 2018-12-16.
- [29] Jorge Piris, Tienke E. Dykstra, Artem A. Bakulin, Paul H.M. van Loosdrecht, Walter Knulst, M. Tuan Trinh, Juleon M. Schins, and Laurens D.A. Siebbeles. Photogeneration and ultrafast dynamics of excitons and charges in p3ht/PCBM blends. *The Journal of Physical Chemistry C*, 113(32):14500–14506, 2009-08-13.
- [30] Chase L. Radford and Timothy L. Kelly. Controlling solid-state structure and

- film morphology in non-fullerene organic photovoltaic devices. *Canadian Journal of Chemistry*, 99(12):921–932, 2021-12.
- [31] Tao Chen, Rui Shi, Ruohua Gui, Haixia Hu, Wenqing Zhang, Kangning Zhang, Bin Cui, Hang Yin, Kun Gao, and Jianqiang Liu. Fluorination of terminal groups promoting electron transfer in small molecular acceptors of bulk heterojunction films. *Molecules*, 27(24):9037, 2022-12-18.
- [32] Frederik C. Krebs. All solution roll-to-roll processed polymer solar cells free from indium-tin-oxide and vacuum coating steps. *Organic Electronics*, 10(5):761–768, 2009-08.
- [33] Daan Frenkel and Berend Smit. *Understanding molecular simulation: from algorithms to applications*. Number 1 in Computational science series. Academic Press, 2nd ed edition, 2002.
- [34] Katharina Vollmayr-Lee. Introduction to molecular dynamics simulations. *American Journal of Physics*, 88(5):401–422, 2020-05.
- [35] Johannes Lenhard, Simon Stephan, and Hans Hasse. On the history of the lennard-jones potential. *Annalen der Physik*, 536(6):2400115, 2024.
- [36] Mohammad M. Ghahremanpour, Julian Tirado-Rives, and William L. Jorgensen. Refinement of the optimized potentials for liquid simulations force field for thermodynamics and dynamics of liquid alkanes. *The Journal of Physical Chemistry B*, 126(31):5896–5907, 2022-08-11.
- [37] Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and

- David A. Case. Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174, 2004.
- [38] Eric Jankowski, Neale Ellyson, Jenny W. Fothergill, Michael M. Henry, Mitchell H. Leibowitz, Evan D. Miller, Mone’t Alberts, Samantha Chesser, Jaime D. Guevara, Chris D. Jones, Mia Klopfenstein, Kendra K. Noneman, Rachel Singleton, Ramon A. Uriarte-Mendoza, Stephen Thomas, Carla E. Estridge, and Matthew L. Jones. Perspective on coarse-graining, cognitive load, and materials simulation. *Computational Materials Science*, 171:109129, 2020-01.
- [39] Yuanqing Wang, Josh Fass, Benjamin Kaminow, John E. Herr, Dominic Rufa, Ivy Zhang, Iván Pulido, Mike Henry, and John D. Chodera. End-to-end differentiable molecular mechanics force field construction. *Chemical Science*, 13(41):12016–12033, 2022.
- [40] Sergi Galindo, Adrián Tamayo, Francesca Leonardi, and Marta Mas-Torrent. Control of polymorphism and morphology in solution sheared organic field-effect transistors. *Advanced Functional Materials*, 27(25):n/a – N.PAG, 2017.
- [41] Lucia Maini, Federico Gallino, Massimo Zambianchi, Margherita Durso, Massimo Gazzano, Katia Rubini, Denis Gentili, Ilse Manet, Michele Muccini, Stefano Toffanin, Massimiliano Cavallini, and Manuela Melucci. Chemical design enables the control of conformational polymorphism in functional 2,3-thieno(bis)imide-ended materials. *Chemical Communications*, 51(11):2033 – 2035, 2015.

- [42] Rupa Hiremath, Joseph A. Basile, Stephen W. Varney, and Jennifer A. Swift. Controlling molecular crystal polymorphism with self-assembled monolayer templates. *Journal of the American Chemical Society*, 127(51):18321 – 18327, 2005.
- [43] Mariacristina Cocca, Maria Laura Di Lorenzo, Mario Malinconico, and Vincenzo Frezza. Influence of crystal polymorphism on mechanical and barrier properties of poly(l-lactic acid). *European Polymer Journal*, 47(5):1073 – 1080, 2011.
- [44] Alfred Y. Lee, Deniz Erdemir, and Allan S. Myerson. Crystal polymorphism in chemical process development. *Annual Review of Chemical and Biomolecular Engineering*, 2(1):259–280, 2011-07-15.
- [45] Brian Moulton and Michael J. Zaworotko. From molecules to crystal engineering: Supramolecular isomerism and polymorphism in network solids. *Chemical Reviews*, 101(6):1629, 2001.
- [46] S. Baskar Raj, P. T. Muthiah, U. Rychlewska, and B. Warzajtis. Pseudo-polymorphism and crystal engineering: hydrogen-bonded supramolecular networks in trimethoprim m-chlorobenzoate and trimethoprim m-chlorobenzoate dihydrate. *CrystEngComm*, 5(9):48–53, 2003. Publisher: The Royal Society of Chemistry.
- [47] Michael T. Ruggiero and J. Axel Zeitler. Resolving the origins of crystalline anharmonicity using terahertz time-domain spectroscopy and ab initio simulations. *Journal of Physical Chemistry B*, 120(45):11733 – 11739, 2016.

- [48] Michael T. Ruggiero, Johanna Kölbel, Qi Li, and J. Axel Zeitler. Predicting the structures and associated phase transition mechanisms in disordered crystals via a combination of experimental and theoretical methods. *Faraday Discuss.*, 211(0):425–439, 2018. Publisher: The Royal Society of Chemistry.
- [49] Matthew D. King, William D. Buchanan, and Timothy M. Korter. Understanding the terahertz spectra of crystalline pharmaceuticals: Terahertz spectroscopy and solid-state density functional theory study of (s)-(+)-ibuprofen and (RS)-ibuprofen. *Journal of Pharmaceutical Sciences*, 100(3):1116 – 1129, 2011.
- [50] Matthew D. King, William D. Buchanan, and Timothy M. Korter. Identification and quantification of polymorphism in the pharmaceutical compound diclofenac acid by terahertz spectroscopy and solid-state density functional theory. *Analytical Chemistry*, 83(10):3786 – 3792, 2011.
- [51] Sean P. Delaney, Duohai Pan, Michael Galella, Shawn X. Yin, and Timothy M. Korter. Understanding the origins of conformational disorder in the crystalline polymorphs of irbesartan. *Crystal Growth & Design*, 12(10):5017–5024, 2012-10-03.
- [52] Michael T. Ruggiero, J. Axel Zeitler, and Timothy M. Korter. Concomitant polymorphism and the martensitic-like transformation of an organic crystal. *Phys. Chem. Chem. Phys.*, 19(42):28502–28506, 2017. Publisher: The Royal Society of Chemistry.
- [53] Damian G. Allis, Darya A. Prokhorova, and Timothy M. Korter. Solid-state modeling of the terahertz spectrum of the high explosive HMX. *The Journal of Physical Chemistry A*, 110(5):1951–1959, 2006-02-01.

- [54] G. A. Jeffrey, S. Ghose, and J. O. Warwicker. The crystal structure of barbituric acid dihydrate. *Acta Crystallographica*, 14(8):881–887, aug 1961.
- [55] A. R. Al-Karaghiali, B. Abdul-Wahab, E. Ajaj, and S. Al-Asaff. A neutron diffraction study of barbituric acid dihydrate. *Acta Crystallographica Section B*, 33(6):1655–1660, jun 1977.
- [56] Neilson R. Rexrode, Jordan Orien, and Matthew D. King. Effects of solvent stabilization on pharmaceutical crystallization: Investigating conformational polymorphism of probucol using combined solid-state density functional theory, molecular dynamics, and terahertz spectroscopy. *The Journal of Physical Chemistry A*, 123(32):6937–6947, 2019-08-15.
- [57] Roberto Dovesi, Roberto Orlando, Alessandro Erba, Claudio M. Zicovich-Wilson, Bartolomeo Civalleri, Silvia Casassa, Lorenzo Maschio, Matteo Ferrabone, Marco De La Pierre, Philippe D’Arco, Yves Noël, Mauro Causà, Michel Rérat, and Bernard Kirtman. Crystal14 : A program for the ab initio investigation of crystalline solids. *International Journal of Quantum Chemistry*, 114:1287–1317, 10 2014.
- [58] R. Dovesi, V. R. Saunders, C. Roetti, R. Orlando, C. M. Zicovich-Wilson, F. Pascale, B.;Doll K. Civalleri, N. M. Harrison, I. J. Bush, P. D’Arco, M. Llunell, M. Causà, Y. Noël, L. Maschio, A. Erba, M. Rerat, and S. Casassa. Crystal17 user’s manual, 2017.
- [59] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, Oct 1996.

- [60] Yan Zhao and Donald G. Truhlar. The m06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four m06-class functionals and 12 other functionals. *Theoretical Chemistry Accounts: Theory, Computation, & Modeling*, 120(1):215 – 241, 2008.
- [61] Chengteh Lee, Weitao Yang, and Robert G. Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37:785–789, Jan 1988.
- [62] Axel D. Becke. Density-functional thermochemistry. III. the role of exact exchange. *The Journal of Chemical Physics*, 98(7):5648–5652, 1993-04-01.
- [63] Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *Journal of Chemical Physics*, 90(2):1007, 1989.
- [64] W. J. Hehre, R. Ditchfield, and J. A. Pople. Self—consistent molecular orbital methods. XII. further extensions of gaussian—type basis sets for use in molecular orbital studies of organic molecules. *The Journal of Chemical Physics*, 56(5):2257–2261, 1972-03-01.
- [65] P. C. Hariharan and John A. Pople. The influence of polarization functions on molecular orbital hydrogenation energies. *Theoretica chimica acta*, 28:213–222, 1973.
- [66] R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople. Self-consistent molecular

- orbital methods. XX. a basis set for correlated wave functions. *The Journal of Chemical Physics*, 72(1):650–654, 1980-01-01.
- [67] Michael F. Peintinger, Daniel Vilela Oliveira, and Thomas Bredow. Consistent gaussian basis sets of triple-zeta valence with polarization quality for solid-state calculations. *Journal of Computational Chemistry*, 34(6):451 – 459, 2013.
- [68] Colin R. Groom, Ian J. Bruno, Matthew P. Lightfoot, and Suzanna C. Ward. The cambridge structural database. *Acta Crystallographica Section B*, 72(2):171–179, 2016-04.
- [69] G. Gilat. Analysis of methods for calculating spectral properties in solids. *Journal of Computational Physics*, 10(3):432–465, 1972.
- [70] Hendrik J. Monkhorst and James D. Pack. Special points for brillouin-zone integrations. *Phys. Rev. B*, 13:5188–5192, Jun 1976.
- [71] Stefan Grimme. Accurate description of van der waals complexes by density functional theory including empirical corrections. *Journal of Computational Chemistry*, 25(12):1463–1473, 2004-09.
- [72] Stefan Grimme. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *Journal of Computational Chemistry*, 27(15):1787–1799, 2006-11-30.
- [73] Amir Karton, David Gruzman, and Jan M. L. Martin. Benchmark thermochemistry of the $\text{cnh}2n+2$ alkane isomers ($n = 2\text{-}8$) and performance of dft and composite ab initio methods for dispersion-driven isomeric equilibria. *The Journal of Physical Chemistry A*, 113(29):8434–8447, July 2009.

- [74] Bartolomeo Civalleri, Claudio M. Zicovich-Wilson, Loredana Valenzano, and Piero Ugliengo. B3lyp augmented with an empirical dispersion term (b3lyp-d*) as applied to molecular crystals. *CrystEngComm*, 10(4):405–410, 2008. Publisher: The Royal Society of Chemistry.
- [75] F. Pascale, C. M. Zicovich-Wilson, F. López Gejo, B. Civalleri, R. Orlando, and R. Dovesi. The calculation of the vibrational frequencies of crystalline compounds and its implementation in the CRYSTAL code. *Journal of Computational Chemistry*, 25(6):888–897, 2004-04-30.
- [76] C. M. Zicovich-Wilson, F. Pascale, C. Roetti, V. R. Saunders, R. Orlando, and R. Dovesi. Calculation of the vibration frequencies of alpha-quartz: The effect of hamiltonian and basis set. *Journal of Computational Chemistry*, 25(15):1873–1881, 2004-11-30.
- [77] Y. C. Shen, P. F. Taday, and M. Pepper. Elimination of scattering effects in spectral measurement of granulated materials using terahertz pulsed spectroscopy. *Applied Physics Letters*, 92(5):051103, 2008-02-04.
- [78] Aurélien M. A. Leguy, Alejandro R. Goñi, Jarvist M. Frost, Jonathan Skelton, Federico Brivio, Xabier Rodríguez-Martínez, Oliver J. Weber, Anuradha Palipurath, M. Isabel Alonso, Mariano Campoy-Quiles, Mark T. Weller, Jenny Nelson, Aron Walsh, and Piers R. F. Barnes. Dynamic disorder, phonon lifetimes, and the assignment of modes to the vibrational spectra of methylammonium lead halide perovskites. *Phys. Chem. Chem. Phys.*, 18(39):27051–27066, 2016. Publisher: The Royal Society of Chemistry.

- [79] Margarita Correa, Arun Kumar, Ram S. Katiyar, and J. M. Scott. Study of optical phonon lifetimes near the diffuse phase transition in relaxor ferroelectric thin films. *Journal of Physics Condensed Matter*, 20(12):125211, feb 2008.
- [80] Chan La-o vorakiat, Huanxin Xia, Jeannette Kadro, Teddy Salim, Daming Zhao, Towfiq Ahmed, Yeng Ming Lam, Jian-Xin Zhu, Rudolph A. Marcus, Maria-Elisabeth Michel-Beyerle, and Elbert E. M. Chia. Phonon mode transformation across the orthohombic–tetragonal phase transition in a lead iodide perovskite $\text{CH}_3\text{NH}_3\text{PbI}_3$: A terahertz time-domain spectroscopy approach. *The Journal of Physical Chemistry Letters*, 7(1):1–6, 2016-01-07.
- [81] Matthew D. King, William D. Buchanan, and Timothy M. Korter. Application of london-type dispersion corrections to the solid-state density functional theory simulation of the terahertz spectra of crystalline pharmaceuticals. *Phys. Chem. Chem. Phys.*, 13(10):4250–4259, 2011. Publisher: The Royal Society of Chemistry.
- [82] Matthew D. King and Timothy M. Korter. Modified corrections for london forces in solid-state density functional theory calculations of structure and lattice dynamics of molecular crystals. *The Journal of Physical Chemistry A*, 116(25):6927–6934, 2012-06-28.
- [83] Elyse M. Kleist, Corinna L. Koch Dandolo, Jean-Paul Guillet, Patrick Mounaix, and Timothy M. Korter. Terahertz spectroscopy and quantum mechanical simulations of crystalline copper-containing historical pigments. *The Journal of Physical Chemistry A*, 123(6):1225–1232, 2019-02-14.

- [84] Matthew D. King, William D. Buchanan, and Timothy M. Korter. Investigating the anharmonicity of lattice vibrations in water-containing molecular crystals through the terahertz spectroscopy of *l*-serine monohydrate. *The Journal of Physical Chemistry A*, 114(35):9570–9578, 2010-09-09.
- [85] Linda Lange and Gabriele Sadowski. Polymorphs, hydrates, cocrystals, and cocrystal hydrates: Thermodynamic modeling of theophylline systems. *Crystal Growth & Design*, 16:4439–4449, 8 2016.
- [86] D Singhal. Drug polymorphism and dosage form design: a practical perspective. *Advanced Drug Delivery Reviews*, 56:335–347, 2 2004.
- [87] Rakesh Radhakrishnan and K. G. Sreejalekshmi. Computational design, synthesis, and structure property evaluation of 1,3-thiazole-based color-tunable multi-heterocyclic small organic fluorophores as multifunctional molecular materials. *The Journal of Organic Chemistry*, 83:3453–3466, 4 2018.
- [88] Thiago H. da Silva, Neilson R. Rexrode, and Matthew D. King. Time-domain terahertz spectroscopy and solid-state density functional theory analysis of p-nitrophenol polymorphs. *Journal of Infrared, Millimeter, and Terahertz Waves*, 41:1337–1354, 11 2020.
- [89] Martin Hutereau, Peter A. Banks, Ben Slater, J. Axel Zeitler, Andrew D. Bond, and Michael T. Ruggiero. Resolving anharmonic lattice dynamics in molecular crystals with x-ray diffraction and terahertz spectroscopy. *Physical Review Letters*, 125:103001, 8 2020.

- [90] Hiromichi Hoshina, Takuro Kanemura, and Michael T. Ruggiero. Exploring the dynamics of bound water in nylon polymers with terahertz spectroscopy. *The Journal of Physical Chemistry B*, 124:422–429, 1 2020.
- [91] Gary S. Nichol and William Clegg. Violuric acid monohydrate: a definitive redetermination at 150 k. *Acta Crystallographica Section E Structure Reports Online*, 61(11):o3788–o3790, 2005-11-15.
- [92] Liang Pu, Yueming Sun, and Zhibing Zhang. Hydrogen bonding in the side-on rings of the hydrates with one acetic acid molecule. *Journal of Molecular Liquids*, 154:124–130, 7 2010.
- [93] G. A. Jeffrey, S. Ghose, and J. O. Warwicker. The crystal structure of barbituric acid dihydrate. *Acta Crystallographica*, 14:881–887, 8 1961.
- [94] A. R. Al-Karaghoubi, B. Abdul-Wahab, E. Ajaj, and S. Al-Asaff. A neutron diffraction study of barbituric acid dihydrate. *Acta Crystallographica Section B Structural Crystallography and Crystal Chemistry*, 33:1655–1660, 6 1977.
- [95] Michael F. Peintinger, Daniel Vilela Oliveira, and Thomas Bredow. Consistent gaussian basis sets of triple-zeta valence with polarization quality for solid-state calculations. *Journal of Computational Chemistry*, 34:451–459, 3 2013.
- [96] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu. *The Journal of Chemical Physics*, 132:154104, 4 2010.

- [97] Stefan Grimme, Stephan Ehrlich, and Lars Goerigk. Effect of the damping function in dispersion corrected density functional theory. *Journal of Computational Chemistry*, 32:1456–1465, 5 2011.
- [98] Stefan Grimme, Andreas Hansen, Jan Gerit Brandenburg, and Christoph Bannwarth. Dispersion-corrected mean-field electronic structure methods. *Chemical Reviews*, 116:5105–5154, 5 2016.
- [99] Marcus D. Hanwell, Donald E. Curtis, David C. Lonie, Tim Vandermeersch, Eva Zurek, and Geoffrey R. Hutchison. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics*, 4(1):17, 2012-08-13.
- [100] G Gilat. Analysis of methods for calculating spectral properties in solids. *Journal of Computational Physics*, 10:432–465, 12 1972.
- [101] Konstantin N. Kudin, Roberto Car, and Raffaele Resta. Berry phase approach to longitudinal dipole moments of infinite chains in electronic-structure methods with local basis sets. *The Journal of Chemical Physics*, 126:234101, 6 2007.
- [102] Lian-Ao Wu, Jing Sun, and Ji-Yu Zhong. A new approach to calculating the berry phase. *Physics Letters A*, 183:257–262, 12 1993.
- [103] Hilal Daglar and Seda Keskin. Recent advances, opportunities, and challenges in high-throughput computational screening of mofs for gas separations. *Coordination Chemistry Reviews*, 422:213470, 2020.
- [104] Xuewei Liu, Danfeng Shi, Shuangyan Zhou, Hongli Liu, Huanxiang Liu, and

- Xiaojun Yao. Molecular dynamics simulations and novel drug discovery. *Expert opinion on drug discovery*, 13(1):23–37, 2018.
- [105] Xiaoqian Lin, Xiu Li, and Xubo Lin. A review on applications of computational methods in drug screening and design. *Molecules*, 25(6):1375, 2020.
- [106] Mohammad Atif Faiz Afzal, Andrea R Browning, Alexander Goldberg, Mathew D Halls, Jacob L Gavartin, Tsuguo Morisato, Thomas F Hughes, David J Giesen, and Joseph E Goose. High-throughput molecular dynamics simulations and validation of thermophysical properties of polymers for various applications. *ACS Applied Polymer Materials*, 3(2):620–630, 2020.
- [107] Michael AR Meier, Richard Hoogenboom, and Ulrich S Schubert. Combinatorial methods, automated synthesis and high-throughput screening in polymer research: The evolution continues. *Macromolecular rapid communications*, 25(1):21–33, 2004.
- [108] Co D Quach, Justin B Gilmer, Daniel Pert, Akanke Mason-Hogans, Christopher R Iacovella, Peter T Cummings, and Clare McCabe. High-throughput screening of tribological properties of monolayer films using molecular dynamics and machine learning. *The Journal of Chemical Physics*, 156(15), 2022.
- [109] AJ Hopfinger and RA Pearlstein. Molecular mechanics force-field parameterization procedures. *Journal of computational chemistry*, 5(5):486–499, 1984.
- [110] Kenno Vanommeslaeghe, Olgun Guvench, et al. Molecular mechanics. *Current pharmaceutical design*, 20(20):3281–3292, 2014.

- [111] William L. Jorgensen and Julian Tirado-Rives. The opls [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, 110(6):1657–1666, 1988.
- [112] Junmei Wang, Romain M Wolf, James W Caldwell, Peter A Kollman, and David A Case. Development and testing of a general amber force field. *Journal of computational chemistry*, 25(9):1157–1174, 2004.
- [113] Matthew W Thompson, Justin B Gilmer, Ray A Matsumoto, Co D Quach, Parashara Shamaprasad, Alexander H Yang, Christopher R Iacobella, Clare McCabe, and Peter T Cummings. Towards molecular simulations that are transparent, reproducible, usable by others, and extensible (true). *Molecular physics*, 118(9-10):e1742938, 2020.
- [114] Eric Jankowski, Neale Ellyson, Jenny W Fothergill, Michael M Henry, Mitchell H Leibowitz, Evan D Miller, Samantha Chesser, Jaime D Guevara, Chris D Jones, Mia Klopfenstein, et al. Perspective on coarse-graining, cognitive load, and materials simulation. *Computational Materials Science*, 171:109129, 2020.
- [115] Christoph Klein, Andrew Z. Summers, Matthew W. Thompson, Justin B. Gilmer, Clare McCabe, Peter T. Cummings, Janos Sallai, and Christopher R. Iacobella. Formalizing atom-typing and the dissemination of force fields with foyer. *Computational Materials Science*, 167:215–227, Sep 2019.
- [116] Connor M. Davel, Timotej Bernat, and Michael R. Shirts Jeffery R. Wagner.

- Parameterization of general organic polymers within the open force field framework. *ChemRxiv*, 2023.
- [117] Kenichiro Takaba, Anika J. Friedman, Chapin E. Cavender, Pavan Kumar Behara, Iván Pulido, Michael M. Henry, Hugo MacDermott-Opeskin, Christopher R. Iacobella, Arnav M. Nagle, Alexander Matthew Payne, Michael R. Shirts, David L. Mobley, John D. Chodera, and Yuanqing Wang. Machine-learned molecular mechanics force fields from large-scale quantum chemical data. *Chemical Science*, 15(32):12861–12878, 2024.
- [118] John Lennard-Jones. The electronic structure of some diatomic molecules. *Transactions of the Faraday Society*, 25(0):668–686, 1929.
- [119] Joshua A. Anderson, Jens Glaser, and Sharon C. Glotzer. HOOMD-blue: A python package for high-performance molecular dynamics and hard particle monte carlo simulations. *Computational Materials Science*, 173:109363, 2020.
- [120] Peter T Cummings, Christopher R Iacobella, Akos Ledeczi, Eric Jankowski, Arthi Jayaraman, Jeremy C Palmer, Edward J Maginn, Sharon C Glotzer, Joshua A Anderson, J Ilja Siepmann, et al. Open-source molecular modeling software in chemical engineering focusing on the molecular simulation design framework. 2021.
- [121] Carl S. Adorf, Paul M. Dodd, Vyas Ramasubramani, and Sharon C. Glotzer. Simple data and workflow management with the signac framework. *Computational Materials Science*, 146:220–229, 2018.
- [122] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez. P <span

- style="font-variant:small-caps;">ACKMOL
- : A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry*, 30(13):2157–2164, 2009-10.
- [123] Marjan Albooyeh, Chris Jones, Rainier Barrett, and Eric Jankowski. Flowermd: Flexible library of organic workflows and extensible recipes for molecular dynamics. *Journal of Open Source Software*, 8(92):5989, 2023.
- [124] Glenn J. Martyna, Douglas J. Tobias, and Michael L. Klein. Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics*, 101(5):4177–4189, 1994.
- [125] J. Cao and G. J. Martyna. Adiabatic path integral molecular dynamics methods. II. algorithms. *The Journal of Chemical Physics*, 104(5):2028–2035, 1996.
- [126] Vyas Ramasubramani, Bradley D. Dice, Eric S. Harper, Matthew P. Spellings, Joshua A. Anderson, and Sharon C. Glotzer. freud: A software suite for high throughput analysis of particle simulation data. *Computer Physics Communications*, 254:107275, 2020.
- [127] Peter T. Cummings, Clare McCabe, Christopher R. Iacovella, Akos Ledeczi, Eric Jankowski, Arthi Jayaraman, Jeremy C. Palmer, Edward J. Maginn, Sharon C. Glotzer, Joshua A. Anderson, J. Ilja Siepmann, Jeffrey Potoff, Ray A. Matsumoto, Justin B. Gilmer, Ryan S. DeFever, Ramanish Singh, and Brad Crawford. Open-source molecular modeling software in chemical engineering focusing on the molecular simulation design framework. *AICHE Journal*, 67(3), 2021.
- [128] Christoph Klein, János Sallai, Trevor J. Jones, Christopher R. Iacovella, Clare

- McCabe, Peter T. Cummings, Randall Q Snurr, Claire S. Adjiman, and David A. Kofke. *A Hierarchical, Component Based Approach to Screening Properties of Soft Matter*, page 79–92. Molecular Modeling and Simulation. Springer Singapore, 2016.
- [129] Gregory M Newbloom, Katie M Weigandt, and Danilo C Pozzo. Structure and property development of poly (3-hexylthiophene) organogels probed with combined rheology, conductivity and small angle neutron scattering. *Soft Matter*, 8(34):8854–8864, 2012.
- [130] Mark Botoshansky, Frank H Herbstein, and Moshe Kapon. Towards a complete description of a polymorphic crystal: The example of perylene: Redetermination of the structures of the ($z= 2$ and 4) polymorphs. *Helvetica chimica acta*, 86(4):1113–1128, 2003.
- [131] Ayumi Ishii and Tsutomu Miyasaka. A high voltage organic–inorganic hybrid photovoltaic cell sensitized with metal–ligand interfacial complexes. *Chem. Commun.*, 48:9900–9902, 2012.
- [132] Hilary S. Marsh, Eric Jankowski, and Arthi Jayaraman. Controlling the morphology of model conjugated thiophene oligomers through alkyl side chain length, placement, and interactions. *Macromolecules*, 47(8):2736–2747, 2014-04-22.
- [133] Chandra Shekar Sarap, Yashpal Singh, John Michael Lane, and Neeraj Rai. Electronic properties and optical spectra of donor–acceptor conjugated organic polymers. *Scientific Reports*, 13(1):21587, 2023-12-07.

- [134] Jaime Martin, Emily C. Davidson, Cristina Greco, Wenmin Xu, James H. Bannock, Amaia Agirre, John De Mello, Rachel A. Segalman, Natalie Stingelin, and Kostas Ch. Daoulas. Temperature-dependence of persistence length affects phenomenological descriptions of aligning interactions in nematic semiconducting polymers. *Chemistry of Materials*, 30(3):748–761, 2018.
- [135] Scott Himmelberger and Alberto Salleo. Engineering semiconducting polymers for efficient charge transport. *MRS Communications*, 5(3):383–395, 2015.
- [136] M. Doi. *Introduction to Polymer Physics*. Introduction to Polymer Physics. Clarendon Press, 1996.
- [137] Michael Rubinstein and Ralph H. Cobly. *Polymer physics*. Oxford University Press, Oxford, 2003.
- [138] Jenny W. Fothergill. *OPEN-SOURCE WORKFLOWS FOR REPRODUCIBLE MOLECULAR SIMULATION*. PhD thesis, Boise State University, 2022.
- [139] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.
- [140] Wenlin Zhang, Enrique D. Gomez, and Scott T. Milner. Predicting chain di-

mensions of semiflexible polymers from dihedral potentials. *Macromolecules*, 47(18):6453–6461, 2014.

APPENDIX A:
ESPALOMA FUNCTION CODE

Espaloma_Function.py	page 1/5
1 import mbuild as mb	
2 import numpy	
3 from mbuild.lib.recipes.polymer import Polymer	
4 import espaloma as esp	
5 from writers import foyer_xml_writer	
6 from writers.foyer_xml_writer import mbuild_to_foyer_xml	
7 from functions import bond_walk	
8 from functions.bond_walk import MadAtom, MadBond, BondWalker	
9 import os	
10 import parmed as pmd	
11 import networkx as nx	
12 if not os.path.exists("espaloma_model.pt"):	
13 os.system("wget http://data.wangyq.net/espaloma_model.pt")	
14	
15 def build_chain(monomer, length, min_energy):	
16 chain = Polymer()	
17 chain.add_monomer(compound=monomer,	
18 indices=monomer.bond_indices,	
19 separation=0.154,	
20 replace=True)	
21 chain.build(n=length)	
22 if min_energy == True:	
23 chain.energy_minimize()	
24 chain.energy_minimize()	
25 return chain	

Espaloma_Function.py	page 2/5
26	
27 def espaloma(MONOMER,XML_FILEPATH,TYPED_FILEPATH,DIMER=False):	
28 '''if you plan to parameterize a polymer made up of your monomer set dimer = True,	
29 if you are simulating only the monomer set dimer = False	
30 '''	
31 if DIMER == True:	
32 monomer = build_chain(monomer=MONOMER,length=1,min_energy=True)	
33 dimer = build_chain(monomer=MONOMER,length=2,min_energy=True)	
34 mb_mol = mb.Compound()	
35 mb_mol.add([monomer,dimer])	
36 monomer.translate([1,1,1])	
37 dimer.translate([-1,-1,-1])	
38 else:	
39 mb_mol = MONOMER	
40	
41 '''Create an OpenMM Molecule object from mBuild Compound	
42 and run BondWalker Function to restore double bonds.'''	
43 rdkit_mol = mb_mol.to_rdkit()	
44 from openff.toolkit.topology import Molecule	
45 comp = Molecule.from_rdkit(rdkit_mol,allow_undefined_stereo=True,hydrogens_are_explicit=True)	
46 bonds = [b for b in comp.bonds]	
47 for i in range(len(bonds)):	
48 bonds[i].bond_order = 1	
49 b = BondWalker(comp)	
50 molecule = b.fill_in_bonds()	
51	
52 '''Run OpenMM Molecule through espaloma code'''	
53 molecule_graph = esp.Graph(molecule)	
54 espaloma_model = esp.get_model("latest")	
55 espaloma_model(molecule_graph.heterograph)	
56 openmm_system = esp.graphs.deploy.openmm_system_from_graph(molecule_graph,charge_method="nn")	
57	
58 '''Store the results for each in something more accessible'''	
59 pair_forces = openmm_system.getForces()[0]	
60 angle_forces = openmm_system.getForces()[2]	
61 bond_forces = openmm_system.getForces()[3]	
62 torsion_forces = openmm_system.getForces()[1]	

Espaloma_Function.py page 3/5

```

63     """generate atom_type type_map"""
64     topology = molecule.to_topology()
65     openmm_topology = topology.to_openmm()
66     structure = pmd.openmm.load_topology(topology=openmm_topology, system=openmm_system)
67     structure.bonds.sort(key=lambda x: x.atom1.idx)
68
69     for i in range(len(molecule.atoms)):
70         if molecule.atoms[i].atomic_number == 6:
71             molecule.atoms[i].name = 'C'
72         if molecule.atoms[i].atomic_number == 1:
73             molecule.atoms[i].name = 'H'
74         if molecule.atoms[i].atomic_number == 7:
75             molecule.atoms[i].name = 'N'
76         if molecule.atoms[i].atomic_number == 16:
77             molecule.atoms[i].name = 'S'
78         if molecule.atoms[i].atomic_number == 8:
79             molecule.atoms[i].name = 'O'
80         if molecule.atoms[i].atomic_number == 9:
81             molecule.atoms[i].name = 'F'
82
83     Gopenmm = nx.Graph()
84     Gparmed = nx.Graph()
85     for i in range(bond_forces.getNumBonds()):
86         Gopenmm.add_edge(bond_forces.getBondParameters(index=i)[0],bond_forces.getBondParameters(index=i)[1])
87
88     for b in structure.bonds:
89         Gparmed.add_edge(b.atom1.idx,b.atom2.idx)
90
91     particle_types = []
92     type_map = dict()
93     tree_openmm = nx.bfs_tree(Gopenmm,0)
94     tree_parmed = nx.bfs_tree(Gparmed,0)
95     if nx.is_isomorphic(Gopenmm,Gparmed):
96         for i in range(pair_forces.getNumParticles()):
97             pair_parms = pair_forces.getParticleParameters(index=i)
98             sigma = pair_parms[1]/pair_parms[1].unit
99             epsilon = pair_parms[2]/pair_parms[2].unit
100            if (sigma, epsilon) not in particle_types:
101                particle_types.append((sigma, epsilon))
102            type_map[molecule.atoms[i].molecule_atom_index] = "".join([molecule.atoms[i].name ,
103                                         str(particle_types.index((sigma, epsilon)))])
104

```

<pre> 105 106 '''generate force dictionaries''' 107 108 bond_types = [] 109 bond_dict = dict() 110 111 for i in range(bond_forces.getNumBonds()): 112 bond_parms = bond_forces.getBondParameters(index=i) 113 l0 = bond_parms[2]/bond_parms[2].unit 114 k = bond_parms[3]/bond_parms[3].unit 115 bond_dict[type_map[bond_parms[0]],type_map[bond_parms[1]]] = {'k':k,'l0':l0} 116 117 angle_types = [] 118 angle_dict = dict() 119 120 for i in range(angle_forces.getNumAngles()): 121 angle_parms = angle_forces.getAngleParameters(index=i) 122 k = angle_parms[4]/angle_parms[4].unit 123 t0 = angle_parms[3]/angle_parms[3].unit 124 angle_dict[type_map[angle_parms[0]],type_map[angle_parms[1]],type_map[angle_parms[2]]] = {'k':k,'t0':t0} 125 126 dihedral_types = [] 127 dihedral_dict = {} 128 129 for i in range(torsion_forces.getNumTorsions()): 130 if i%6==0: 131 periodicity=[] 132 phase = [] 133 k = [] 134 135 dihedral_parms = torsion_forces.getTorsionParameters(index=i) 136 periodicity.append(dihedral_parms[4]) 137 phase.append(dihedral_parms[5]/dihedral_parms[5].unit) 138 k.append(dihedral_parms[6]/dihedral_parms[6].unit) 139 dt = (type_map[dihedral_parms[0]],type_map[dihedral_parms[1]],type_map[dihedral_parms[2]], 140 type_map[dihedral_parms[3]]) 141 if periodicity[-1]==6: 142 dihedral_dict[dt] = {'periodicity':periodicity,'k':k,'phase':phase} 143 144 nonbonded_types = [] 145 nonbonded_dict = {} 146 147 for i in range(pair_forces.getNumParticles()): 148 nonbonded_parms = pair_forces.getParticleParameters(index=i) 149 charge = nonbonded_parms[0]/nonbonded_parms[0].unit 150 sigma = nonbonded_parms[1]/nonbonded_parms[1].unit 151 epsilon = nonbonded_parms[2]/nonbonded_parms[2].unit 152 nonbonded_types.append((charge,sigma,epsilon)) 153 nonbonded_dict[(type_map[i])]={'charge':charge,'sigma':sigma,'epsilon':epsilon} </pre>	page 4/5
---	----------

	Espaloma_Function.py	page 5/5
149	'''rename the atoms in the mBuild Compound to match espaloma forcefield	
150	and remove extra dimer from molecule object:'''	
151	for index in type_map:	
152	mb_mol[index].name = type_map[index]	
153		
154	if DIMER == True and mb_mol.children[1].n_particles > mb_mol.children[0].n_particles:	
155	dim = mb_mol.children[1]	
156	mb_mol.remove(objs_to_remove=dim)	
157		
158	'''write the forcefield parameters into an XML file:'''	
159	mbuild_to_foyer_xml(
160	file_name=XML_FILEPATH,	
161	compound=mb_mol,	
162	bond_params=bond_dict,	
163	angle_params=angle_dict,	
164	dihedral_params=dihedral_dict,	
165	dihedral_type="periodic",	
166	non_bonded_params=nonbonded_dict,	
167	combining_rule="geometric",	
168	name="",	
169	version="",	
170	coulomb14scale=1.0,	
171	lj14scale=1.0)	
172		
173	'''Create an mol2 file with the correct atom types corresponding to the espaloma forcefield'''	
174	mb_mol.save(TYPED_FILEPATH,overwrite=True)	
175		
176	if os.path.exists(TYPED_FILEPATH):	
177	print('Typed mol2 file created.')	
178	else:	
179	print('Failed to create typed mol2')	
180	if os.path.exists(XML_FILEPATH):	
181	print('xml file created.')	
182	else:	
183	print('Failed to create xml file')	
184		
185		

APPENDIX B:
BOND WALKER EXAMPLE CODE

<pre> 1 import mbuild as mb 2 import random 3 import numpy 4 import random 5 import warnings 6 warnings.filterwarnings("ignore") 7 8 octet = {6 : 4, 7 : 3, 1 : 1, 8 : 2, 16 : 2, 9:1} 9 atom_type = {6:"C",7:"N",8:"O",1:"H",16:"S",9:"F"} 10 11 class MadAtom(): 12 13 def check_happy(self): 14 if self.desired_order == self.current_order: 15 self.happy = True 16 return 17 self.happy= False 18 19 def add_order(self,o): 20 self.current_order += o 21 self.check_happy() 22 for b in self.bonds: 23 b.check_done() 24 25 def status(self): 26 return self.desired_order - self.current_order 27 28 def __str__(self): 29 return "{}{}: {}-{}, {}".format(atom_type[self.atomic_number],self.index,self.desired_order,self.current_order,self.happy) 30 31 def __init__(self, i, an, do, co): 32 self.index = i 33 self.atomic_number = an 34 self.desired_order = do 35 self.current_order = co 36 self.happy = False 37 self.check_happy() 38 self.bonds = [] </pre>	page 1/3
---	----------

<pre> 39 40 class MadBond(): 41 def __init__(self,a1,a2,bo,i): 42 self.atom1 = a1 43 self.atom2 = a2 44 self.order = bo 45 self.happy = False 46 self.index = i 47 48 def __repr__(self): 49 return "Bond {} {}-{}{} O={}, {}".format(self.index,atom_type[self.atom1.atomic_number], 50 self.atom1.index,atom_type[self.atom2.atomic_number], 51 self.atom2.index,self.order,self.happy) 52 53 def check_done(self): 54 if self.atom1.atomic_number == 1: 55 self.happy = True 56 return 57 if self.atom2.atomic_number == 1: 58 self.happy = True 59 return 60 if self.atom1.happy and self.atom2.happy: 61 self.happy = True 62 return 63 self.happy = False 64 65 def add_order(self,o): 66 self.order += o 67 self.atom1.add_order(o) 68 self.atom2.add_order(o) 69 self.check_done() 70 </pre>	page 2/3
--	----------

<pre> 71 class BondWalker(): 72 73 def __init__(self, mol): 74 self.mol = mol 75 self.atom_list = [] 76 for atom in mol.atoms: 77 self.atom_list.append(MadAtom(atom.molecule_atom_index, 78 atom.atomic_number, 79 octet[atom.atomic_number], 80 len([b for b in mol.bonds if atom.molecule_atom_index in [b.atom1_index,b.atom2_index]]))) 81 #assumes we have no double bonds at the start 82 self.bond_list = [MadBond(self.atom_list[b.atom1_index], 83 self.atom_list[b.atom2_index], 84 b.bond_order,i) 85 for i,b in enumerate(mol.bonds)] # .bond_order inside each of these, and can use atom1_index atom2_index 86 for b in self.bond_list: 87 b.atom1.bonds.append(b) 88 b.atom2.bonds.append(b) 89 b.check_done() 90 91 def done(self): 92 for i in self.atom_list: 93 if i.happy == False: 94 return False 95 for i in self.bond_list: 96 if i.happy == False: 97 return False 98 return True 99 100 def bond_increasable(self, b): 101 if not b.atom1.happy and not b.atom2.happy: 102 if b.atom1.status() > 0: 103 if b.atom2.status() >0: 104 b.add_order(1) #this triggers each atom to increment, and *their* bonds to check happiness 105 if self.done(): 106 return True 107 for a in self.atom_list: 108 if a.happy == False: 109 for bx in a.bonds: 110 if self.bond_increasable(bx): 111 return True 112 #no bonds around this unhappy atom "a" are increasable 113 b.add_order(-1) 114 return False 115 116 def fill_in_bonds(self): 117 for a in self.atom_list: #since some of the unhappy bonds might end up being OK, let's iterate atoms 118 if a.happy == False: 119 for b in a.bonds: #try incrementing every bond around every unhappy atom 120 if self.bond_increasable(b): 121 print("Done!") 122 for b1,b2 in zip(self.bond_list, self.mol.bonds): 123 b2.bond_order=b1.order 124 return self.mol 125 print("Sorry- I couldn't solve this") 126 </pre>	<p>bondwalker.py</p> <p>page 3/3</p>
---	--------------------------------------

APPENDIX C:

PERSISTENCE LENGTH APPENDIX

Polymer	Density
PCPDTFBT-C1-BO	0.05
PCPDTFBT-C3-BO	0.05
PCPDTFBT-C4-BO	0.01
PCPDTFBT-C5-BO	0.05
PCPDTFBT-C11-BO	0.005
PCPDTPT-ene-HD	0.005
PCPDTPT-ene-ODD	0.05
PCPDTPT-HD	0.05
PCPDTPT-nC16	0.05
PCPDTPT-ODD	0.05
PIDTB-nC16	0.05
PIDTCPDT-C11-BO	0.05
PIDTFBT-C11-BO	0.005

Table C.1: Densities at which each polymer simulation was run at. All densities reported in units of g/mL

Polymer	TPS
PCPDTFBT-C11-BO	1027.1953
PCPDTFBT-C1-BO	1085.1254
PCPDTFBT-C3-BO	1134.8646
PCPDTFBT-C4-BO	1021.8457
PCPDTFBT-C5-BO	959.0328
PCPDPTPT-ene-HD	1987.0726
PCPDPTPT-ene-ODD	1717.0476
PCPDPTPT-HD	1443.4256
PCPDPTPT-nC16	1481.9449
PCPDPTPT-ODD	1425.6319
PIDTBT-nC16	1025.8267
PIDTCPDT-C11-BO	646.3932
PIDTFBT-C11-BO	421.9683

Table C.2: Average timesteps per second (TPS) for each polymer simulation. Values averaged over each simulation with temperatures ranging from 252 K to 2013 K.

Polymer	l_p			
	252 K	503 K	1006 K	SANS
PCPDTFBT-C1-BO	37.8 ± 0.1	115.3 ± 0.5	99.7 ± 5.9	67
PCPDTFBT-C3-BO	47.5 ± 0.1	57.6 ± 0.8	98.9 ± 6.1	78.4
PCPDTFBT-C4-BO	46.3 ± 0.3	65.8 ± 0.9	105.1 ± 5.6	86.4
PCPDTFBT-C5-BO	35.4 ± 0.1	68.7 ± 0.5	104.3 ± 7.8	114
PCPDTFBT-C11-BO	34.6 ± 0.1	55.9 ± 0.7	104.0 ± 7.4	291
PCPDPTPT-ene-HD	39.8 ± 0.4	48.4 ± 0.3	110.7 ± 7.9	76.6
PCPDPTPT-ene-ODD	21.9 ± 0.1	38.5 ± 0.3	97.8 ± 7.3	83.4
PCPDPTPT-HD	36.8 ± 0.1	55.3 ± 2.6	100.4 ± 6.3	47.3
PCPDPTPT-nC16	40.2 ± 0.2	60.3 ± 0.4	96.0 ± 6.3	61
PCPDPTPT-ODD	56.4 ± 0.3	63.6 ± 0.7	99.6 ± 9.6	54.9
PIDTBT-nC16	65.4 ± 0.2	57.8 ± 0.4	200.6 ± 25.5	1310
PIDTCPDT-C11-BO	36.0 ± 0.1	76.4 ± 2.4	182.2 ± 13.7	236
PIDTFBT-C11-BO	79.4 ± 1.1	90.6 ± 6.5	178.4 ± 9.2	254

Table C.3: Caption

Polymer	T_{SANS}
PCPDTFBT-C1-BO	326
PCPDTFBT-C3-BO	736
PCPDTFBT-C4-BO	767
PCPDTFBT-C5-BO	1091
PCPDTFBT-C11-BO	3035
PCPDTPT-ene-HD	692
PCPDTPT-ene-ODD	885
PCPDTPT-HD	389
PCPDTPT-nC16	526
PCPDTPT-ODD	278
PIDTB-T-nC16	6785
PIDTCPDT-C11-BO	1290
PIDTFBT-C11-BO	1591

Table C.4: Predicted simulation temperature at which l_p^{MD} would match l_p^{SANS} .

	persistence_length.py	page 1/3
1	import numpy as np	
2	import MDAnalysis as mda	
3	from MDAnalysis.analysis import polymer	
4	import matplotlib.pyplot as plt	
5	import scipy	
6	from cmeutils.sampling import autocorr1D	
7		
8	def persistence_length(filepath,start,stop,interval):	
9	"""	
10	filepath needs to be a format in which you can	
11	create an mdanalysis universe from, we mostly use gsd files	
12	"""	
13	u = mda.Universe(topology=filepath)	
14		
15	"""rewrite atom indices""" bond_indices = [] particle_index = [0] for i in range(len(u.bonds)): a = u.bonds[i].atoms.indices bond_indices.append(list(a)) if particle_index[-1] in bond_indices[i]: atom1 = bond_indices[i][0] atom2 = bond_indices[i][1] if atom1 not in particle_index: particle_index.append(atom1) if atom2 not in particle_index: particle_index.append(atom2)	
21		
22		
23		
24		
25		
26		
27		

	persistence_length.py	page 2/3
28		
29	"""create bonds list"""	
30	l_ps = []	
31	bond_len = []	
32	for t in u.trajectory[start:stop:interval]:	
33	bonds = []	
34	unit_bonds = []	
35	angles = []	
36		
37	for i in range(len(particle_index)):	
38	b = t.positions[particle_index[i]]-t.positions[particle_index[i-1]]	
39	bonds.append(b)	
40	l2 = t.dimensions[0]/2	
41	for i,b in enumerate(bonds):	
42	for j,x in enumerate(b):	
43	if x>l2:	
44	bonds[i][j] = x-l2*2	
45	if x<-l2:	
46	bonds[i][j] = x+l2*2	
47	a = b/np.linalg.norm(b)	
48	unit_bonds.append(a)	
49		
50	avg_traj_lb = np.average([np.linalg.norm(b) for b in bonds])	
51	bond_len.append(avg_traj_lb)	

	persistence_length.py	page 3/3
52		
53	for i in range(len(unit_bonds)-1):	
54	b1 = unit_bonds[0]	
55	b2 = unit_bonds[0+i]	
56	dot_product = np.dot(b1,b2)	
57	angles.append(dot_product)	
58		
59	auto = autocorr1D(angles)	
60	x = [i for i in range(0,2*len(auto),2)]	
61		
62	'''set negative results to 0'''	
63	for r in range(len(auto)):	
64	if auto[r] < 0:	
65	auto[r] = 0	
66		
67	def expfunc(x, a):	
68	return np.exp(-x/a)	
69		
70	exp_coeff = scipy.optimize.curve_fit(expfunc,x,auto)[0][0]	
71		
72	l_p = exp_coeff * avg_traj_lb	
73	l_ps.append(l_p)	
74		
75	'''average the data from trajectories together'''	
76	av_lb = np.average(bond_len)	
77	av_lp = np.average(l_ps)	
78	std_dev = np.std(l_ps)	
79		
80	return av_lb,av_lp, std_dev	

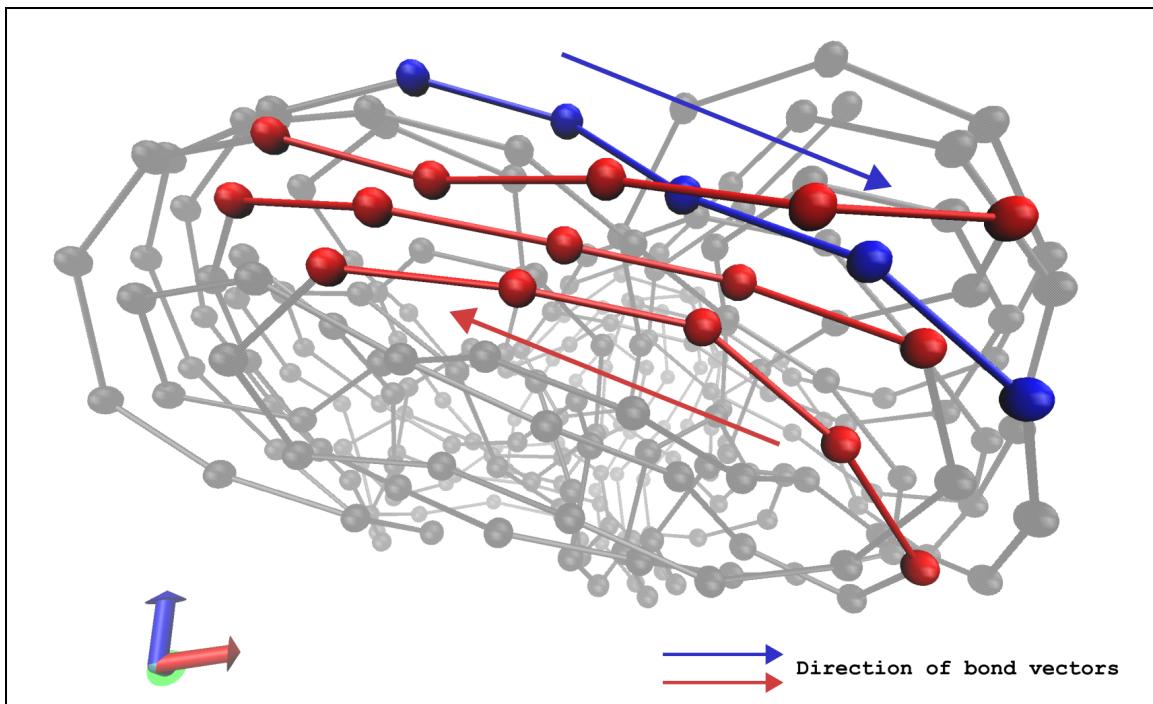


Figure C.1: Snapshot of coarse grained PCPDTPT-HD at a temperature of 252 K and density of 0.05 g/mol. The red highlighted CG beads have parallel aligned bond vectors. The blue highlighted CG beads are parallel, but anti-aligned with the red.

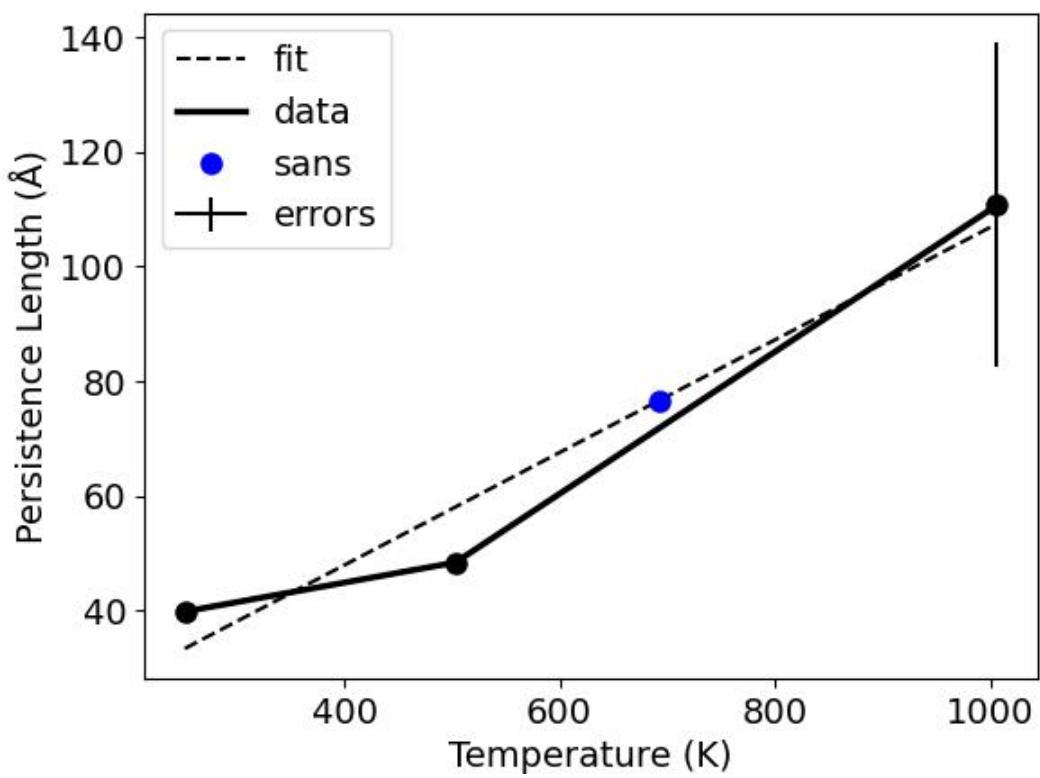


Figure C.2: Persistence length (l_p^{MD}) of PCPDTPT-eneHD vs Temperature plot with a linear regression calculated using SciPy [7] and T_{SANS} prediction.

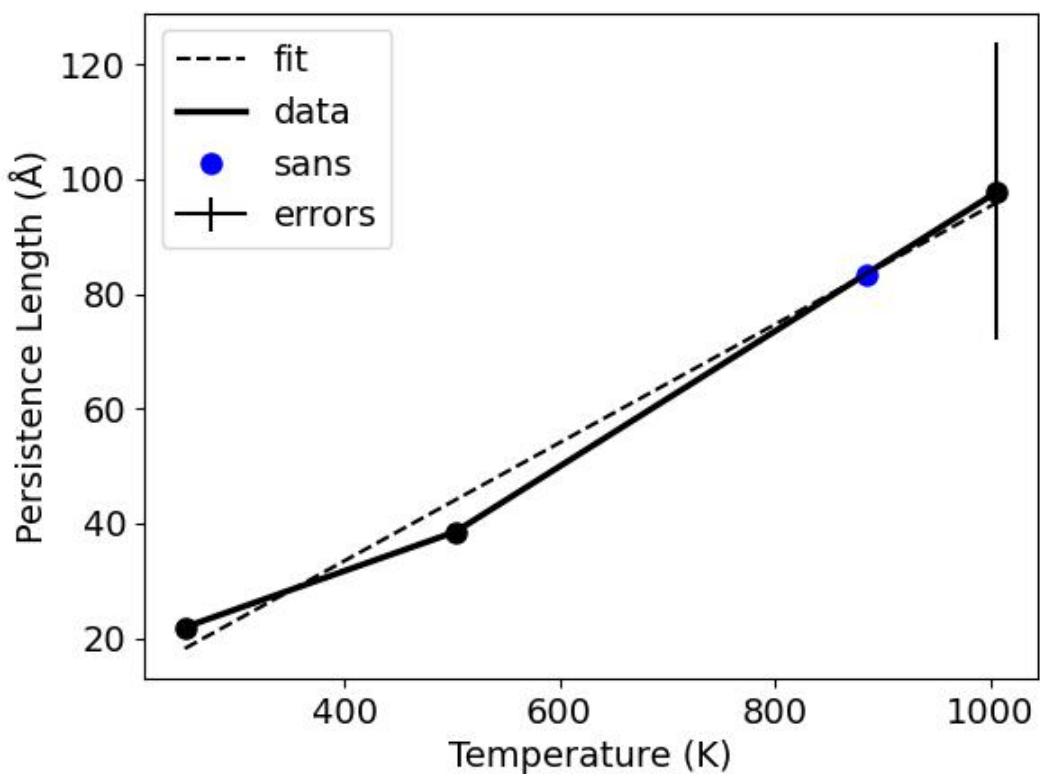


Figure C.3: Persistence length (l_p^{MD}) of PCPDPT-eneODD vs Temperature plot with a linear regression calculated using SciPy [7] and T_{SANS} prediction.

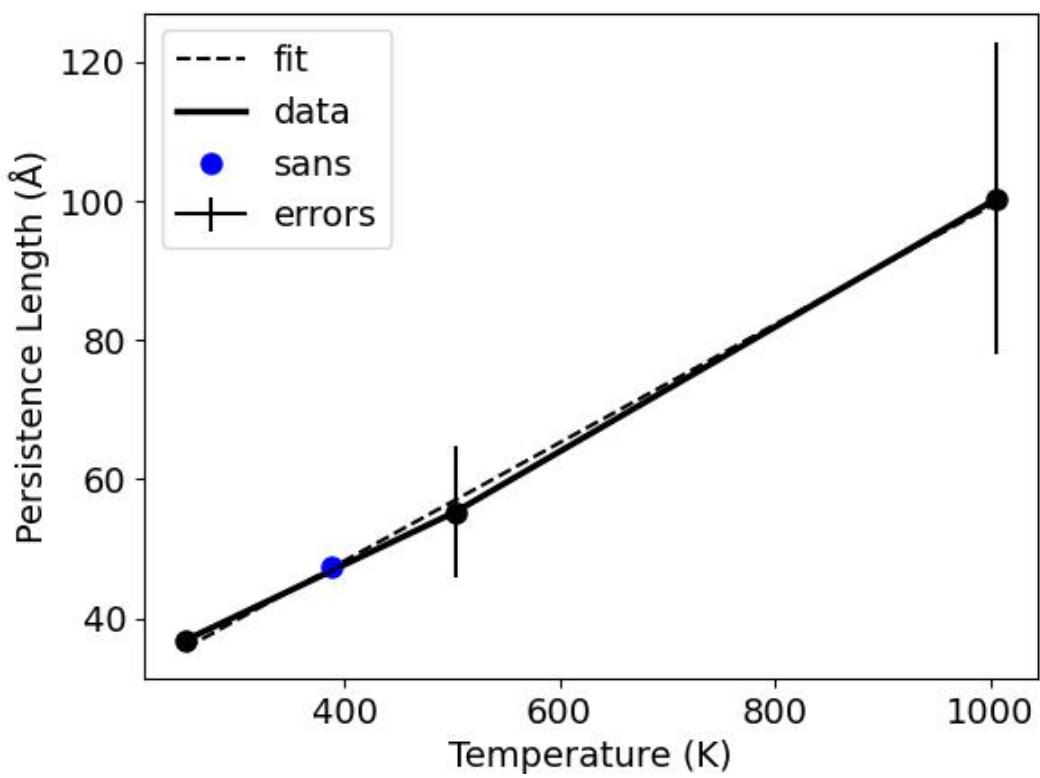


Figure C.4: Persistence length (l_p^{MD}) of PCPDTPT-HD vs Temperature plot with a linear regression calculated using SciPy [7] and T_{SANS} prediction.

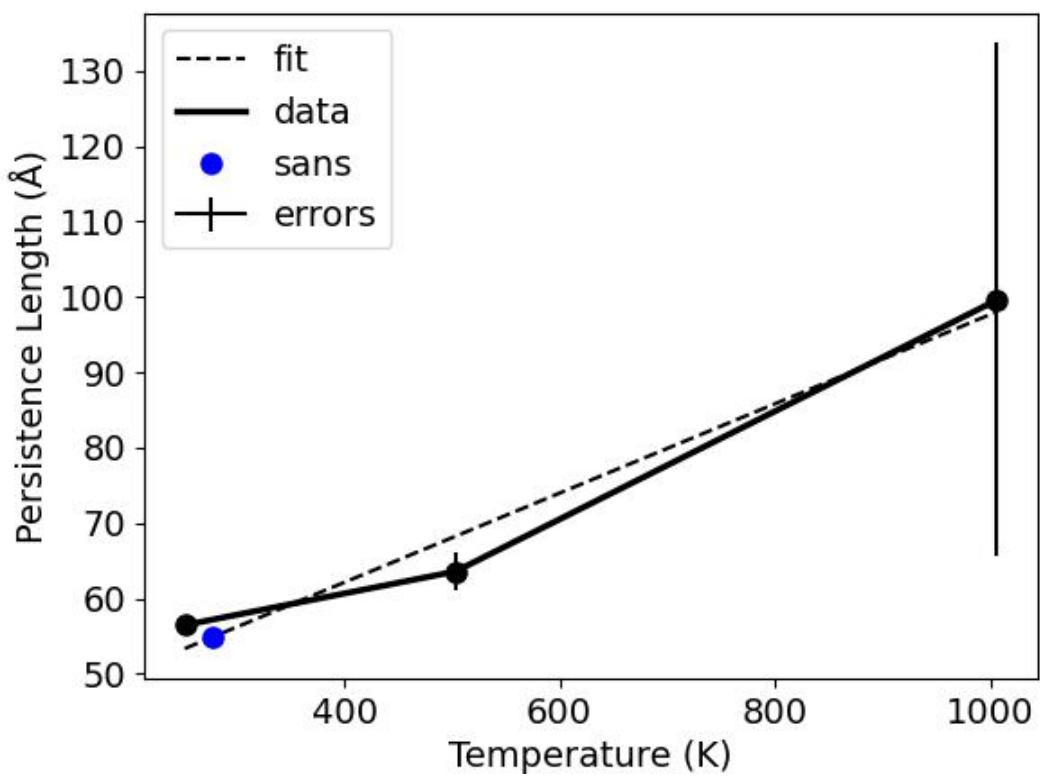


Figure C.5: Persistence length (l_p^{MD}) of PCPDTPT-ODD vs Temperature plot with a linear regression calculated using SciPy [7] and T_{SANS} prediction.

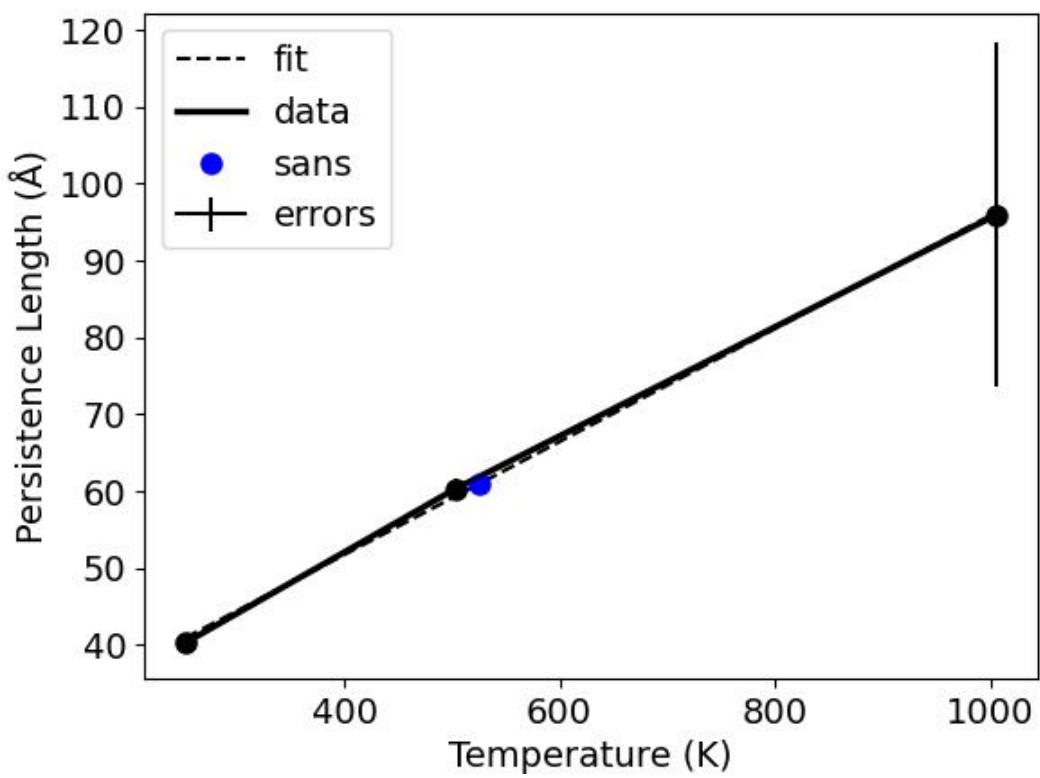


Figure C.6: Persistence length (l_p^{MD}) of PCPDTPT-nC16 vs Temperature plot with a linear regression calculated using SciPy [7] and T_{SANS} prediction.

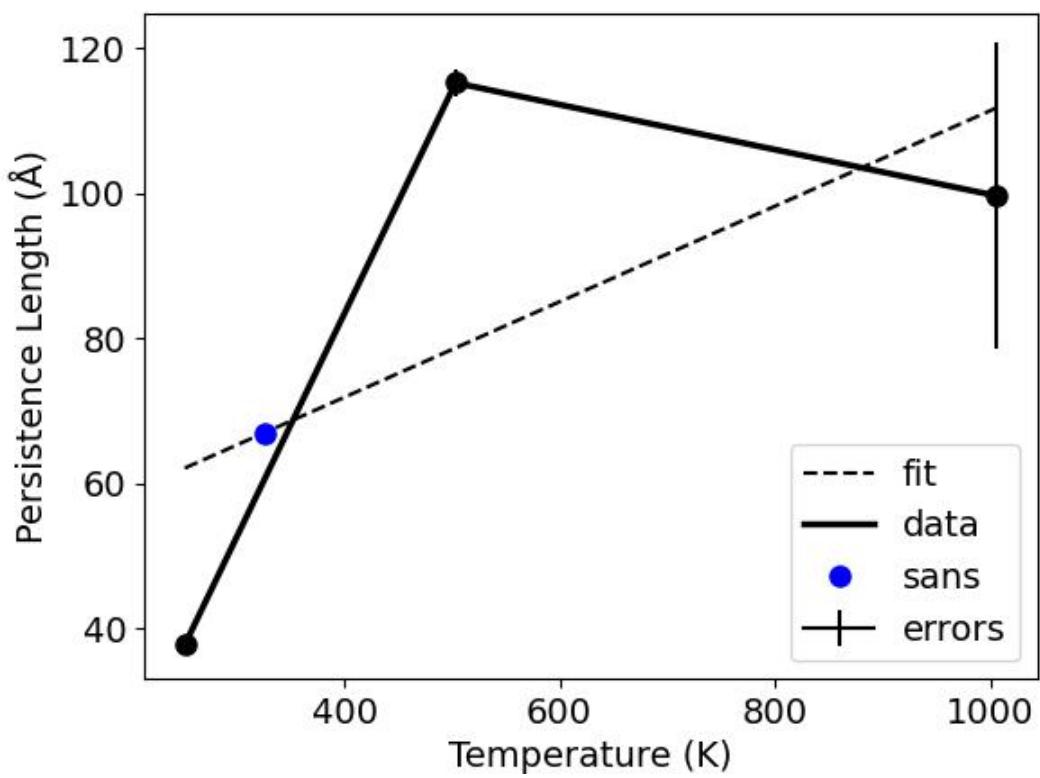


Figure C.7: Persistence length (l_p^{MD}) of PCPDTFBT-C1-BO vs Temperature plot with a linear regression calculated using SciPy [7] and T_{SANS} prediction.

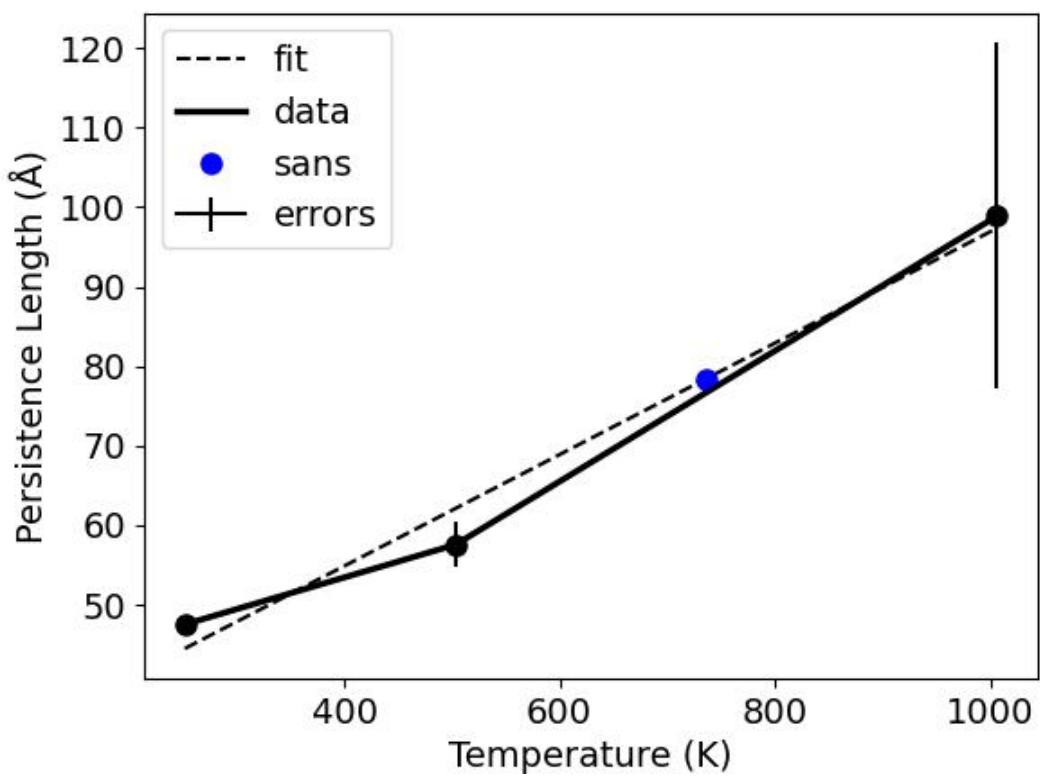


Figure C.8: Persistence length (l_p^{MD}) of PCPDTFBT-C3-BO vs Temperature plot with a linear regression calculated using SciPy [7] and T_{SANS} prediction.

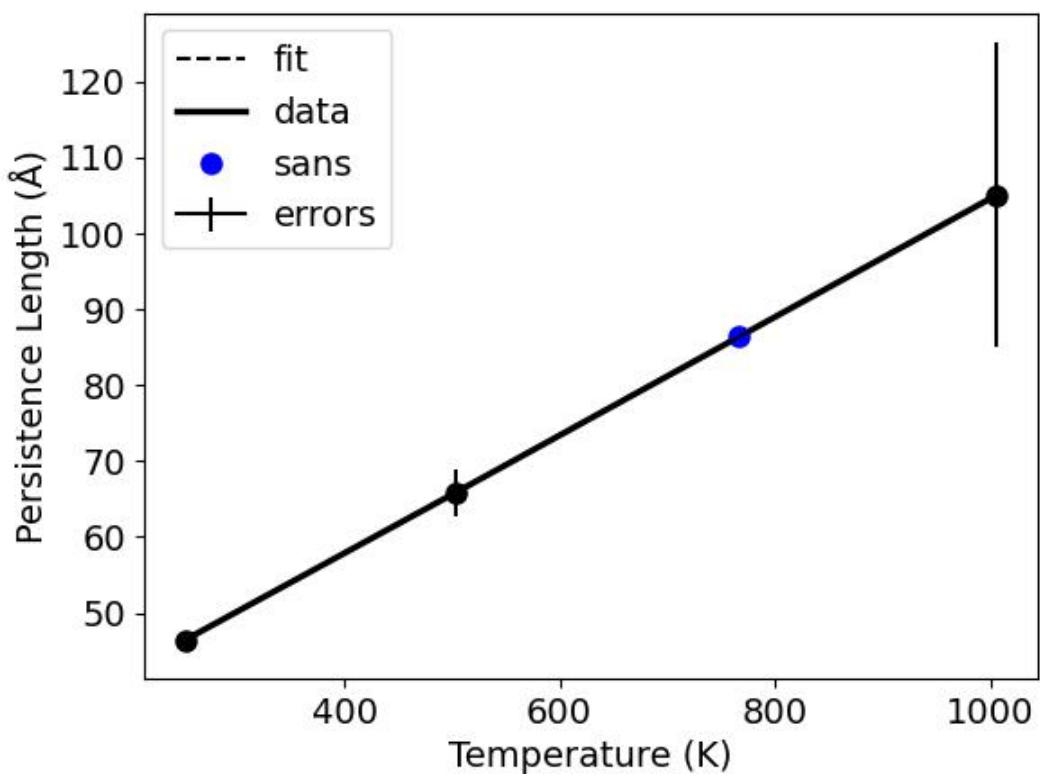


Figure C.9: Persistence length (l_p^{MD}) of PCPDTFBT-C4-BO vs Temperature plot with a linear regression calculated using SciPy [7] and T_{SANS} prediction.

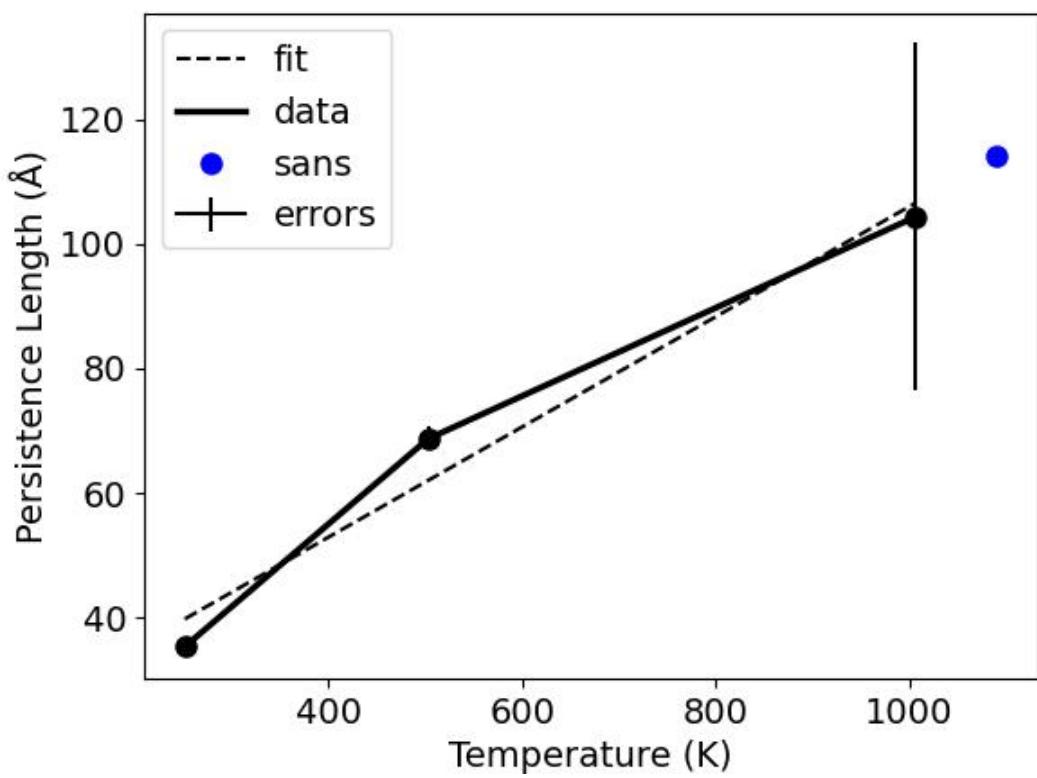


Figure C.10: Persistence length (l_p^{MD}) of PCPDTFBT-C5-BO vs Temperature plot with a linear regression calculated using SciPy [7] and T_{SANS} prediction.

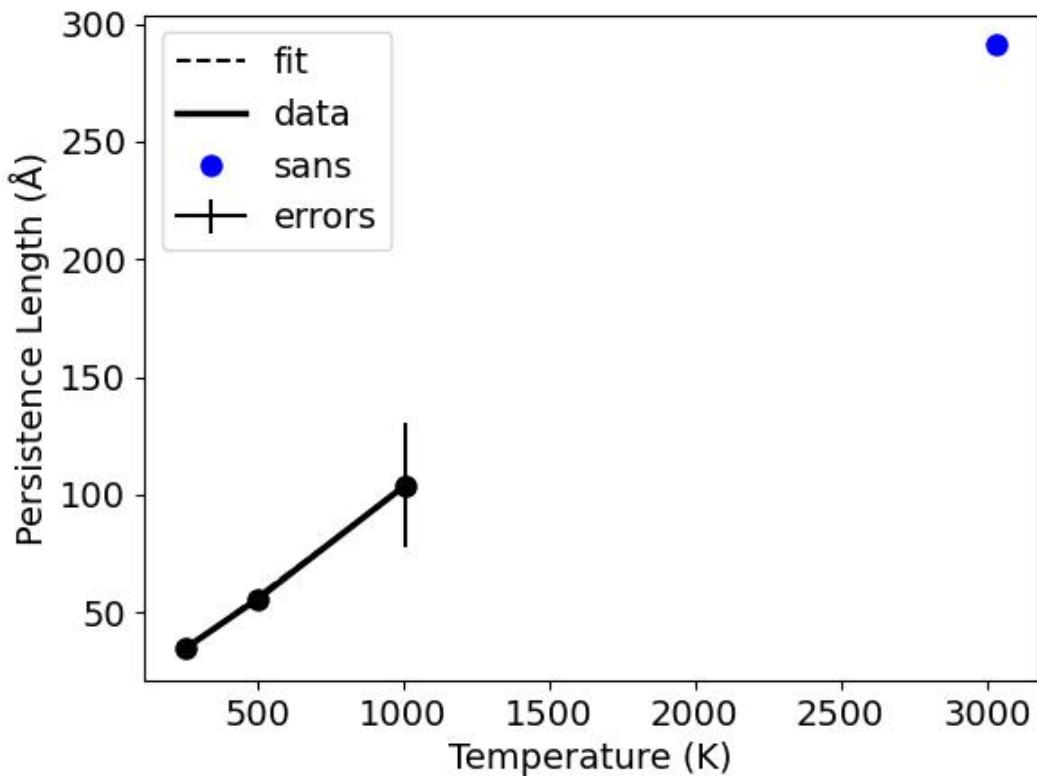


Figure C.11: Persistence length (l_p^{MD}) of PCPDTFBT-C11-BO vs Temperature plot with a linear regression calculated using SciPy [7] and T_{SANS} prediction.

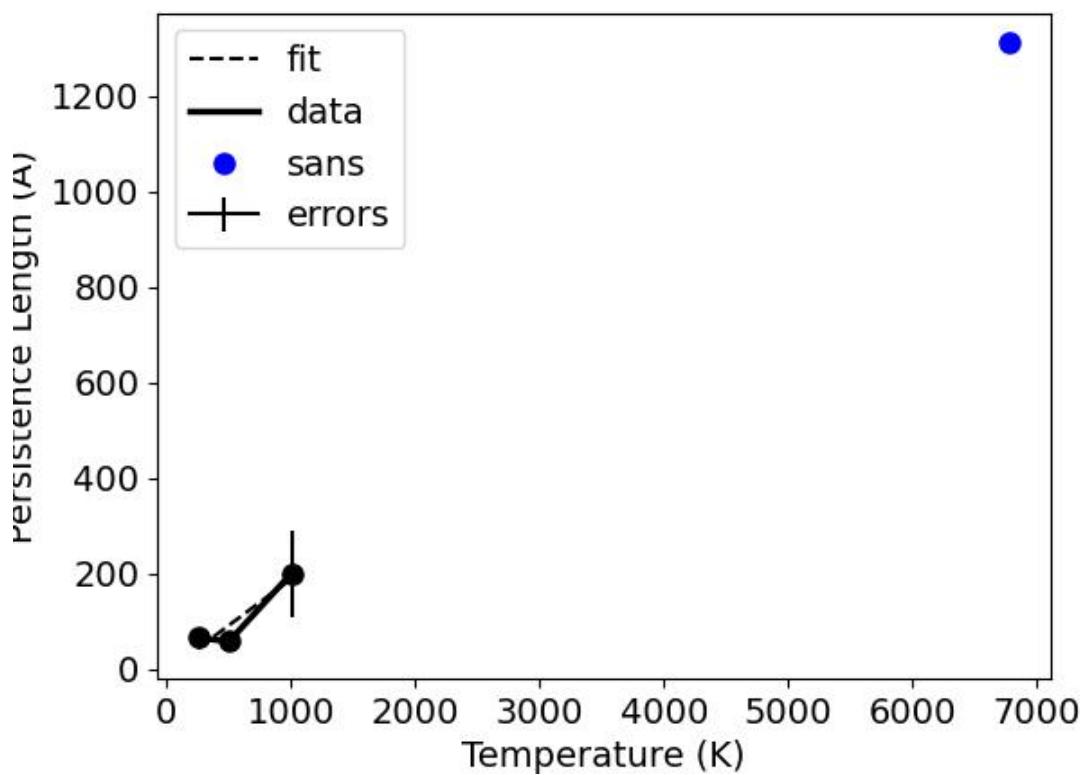


Figure C.12: Persistence length (l_p^{MD}) of PIDTBT-nC16 vs Temperature plot with a linear regression calculated using SciPy [7] and T_{SANS} prediction.

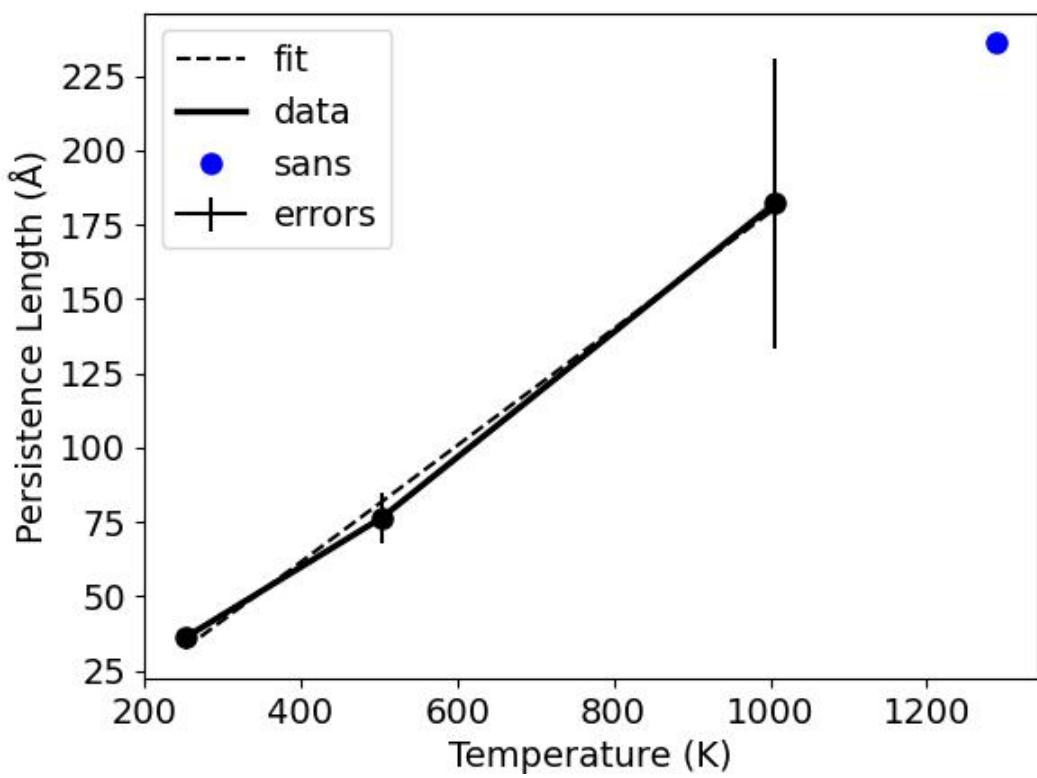


Figure C.13: Persistence length (l_p^{MD}) of PIDTCPDT-C11-BO vs Temperature plot with a linear regression calculated using SciPy [7] and T_{SANS} prediction.

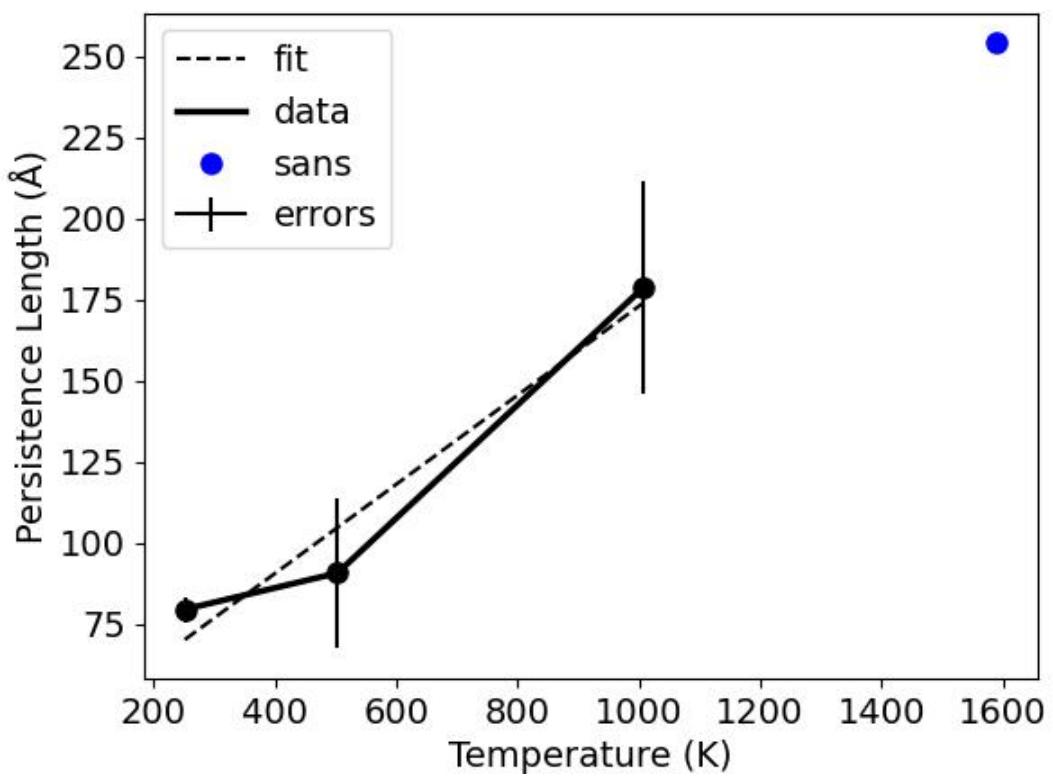


Figure C.14: Persistence length (l_p^{MD}) of PIDTFBT-C11-BO vs Temperature plot with a linear regression calculated using SciPy [7] and T_{SANS} prediction.

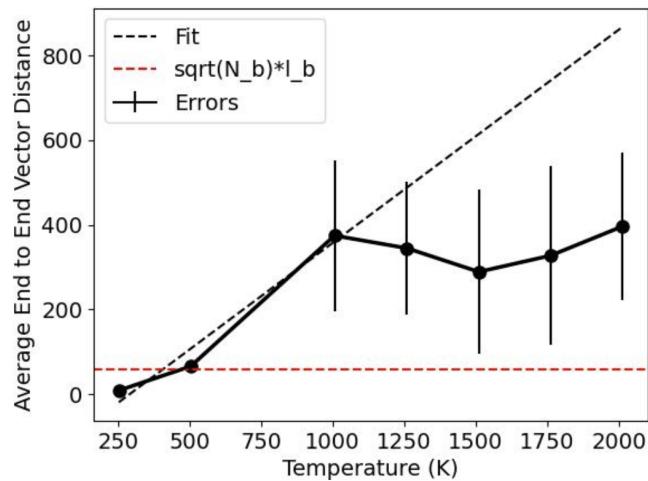


Figure C.15: $\sqrt{\langle R^2 \rangle}$ vs Temperature for PCPDTFBT-C1-BO with a linear regression for the linear portion of the plot. Horizontal line plotted at $\sqrt{N_b} * l_b$.

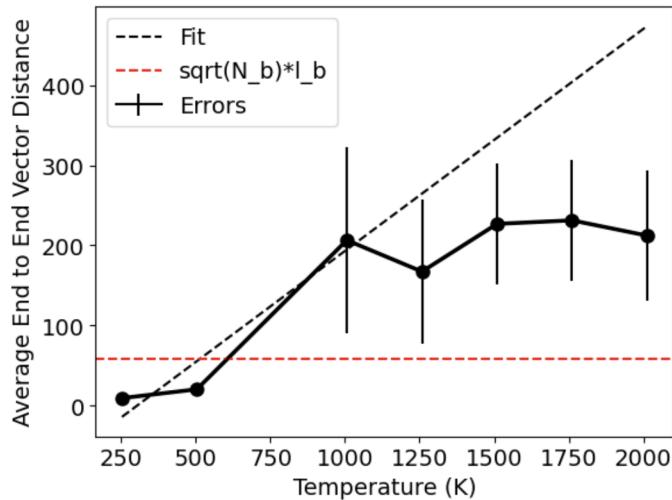


Figure C.16: $\sqrt{\langle R^2 \rangle}$ vs Temperature for PCPDTFBT-C3-BO with a linear regression for the linear portion of the plot. Horizontal line plotted at $\sqrt{N_b} * l_b$.

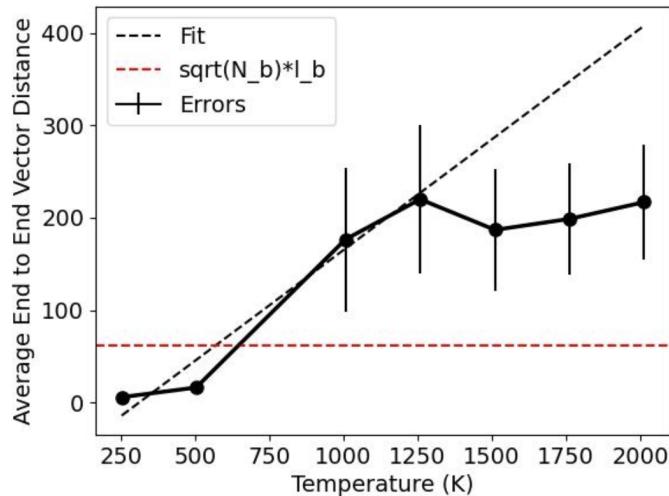


Figure C.17: $\sqrt{\langle R^2 \rangle}$ vs Temperature for PCPDTFBT-C4-BO with a linear regression for the linear portion of the plot. Horizontal line plotted at $\sqrt{N_b} * l_b$.

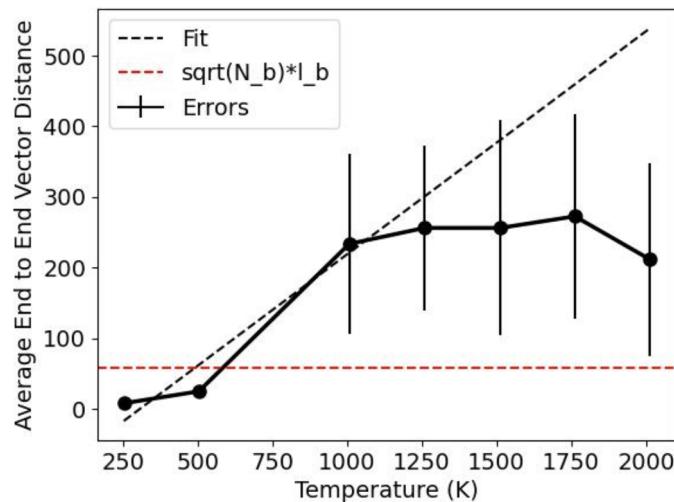


Figure C.18: $\sqrt{\langle R^2 \rangle}$ vs Temperature for PCPDTFBT-C5-BO with a linear regression for the linear portion of the plot. Horizontal line plotted at $\sqrt{N_b} * l_b$.

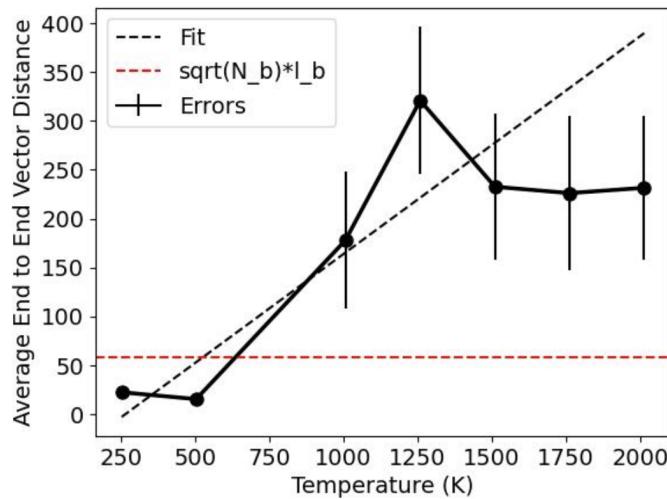


Figure C.19: $\sqrt{\langle R^2 \rangle}$ vs Temperature for PCPDTFBT-C11-BO with a linear regression for the linear portion of the plot. Horizontal line plotted at $\sqrt{N_b} * l_b$.

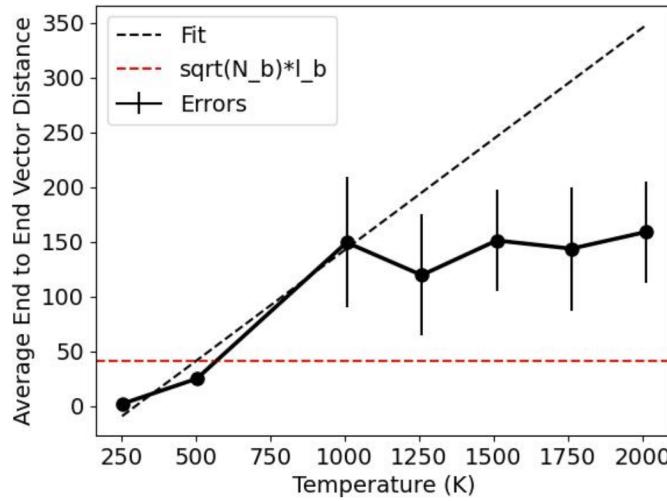


Figure C.20: $\sqrt{\langle R^2 \rangle}$ vs Temperature for PCPDTP-ene-HD with a linear regression for the linear portion of the plot. Horizontal line plotted at $\sqrt{N_b} * l_b$.

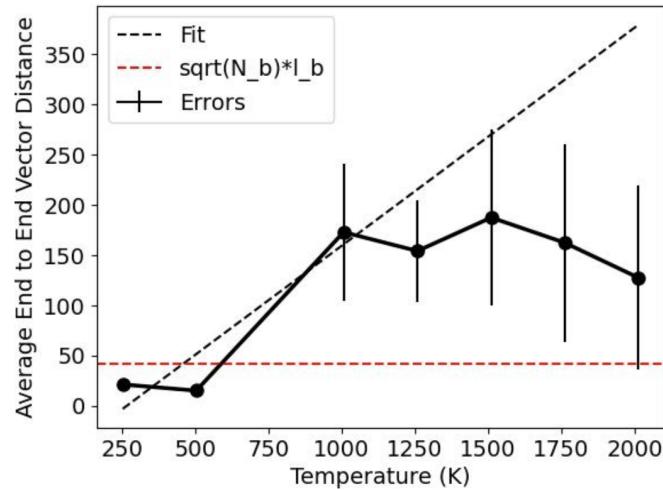


Figure C.21: $\sqrt{\langle R^2 \rangle}$ vs Temperature for PCPDTPT-ene-ODD with a linear regression for the linear portion of the plot. Horizontal line plotted at $\sqrt{N_b} * l_b$.

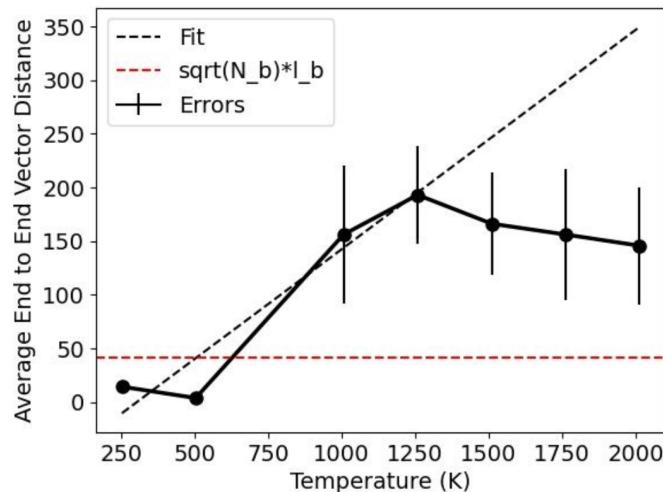


Figure C.22: $\sqrt{\langle R^2 \rangle}$ vs Temperature for PCPDTPT-HD with a linear regression for the linear portion of the plot. Horizontal line plotted at $\sqrt{N_b} * l_b$.

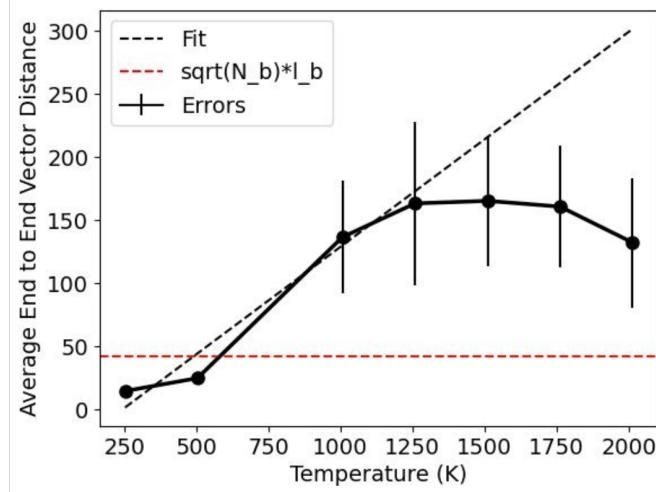


Figure C.23: $\sqrt{\langle R^2 \rangle}$ vs Temperature for PCPDTPT-nC16 with a linear regression for the linear portion of the plot. Horizontal line plotted at $\sqrt{N_b} * l_b$.

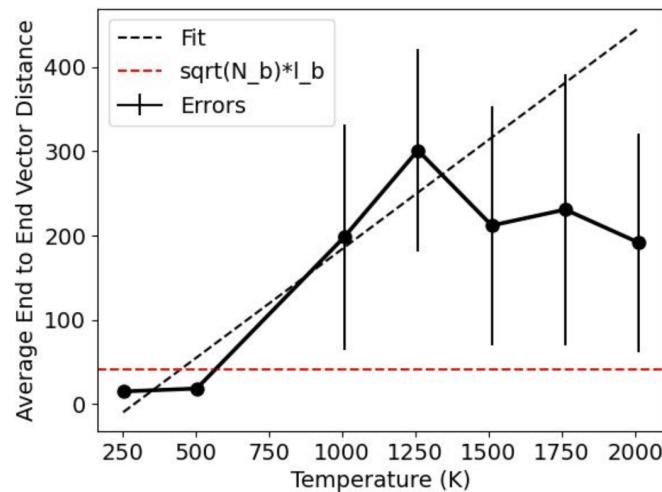


Figure C.24: $\sqrt{\langle R^2 \rangle}$ vs Temperature for PCPDTPT-ODD with a linear regression for the linear portion of the plot. Horizontal line plotted at $\sqrt{N_b} * l_b$.

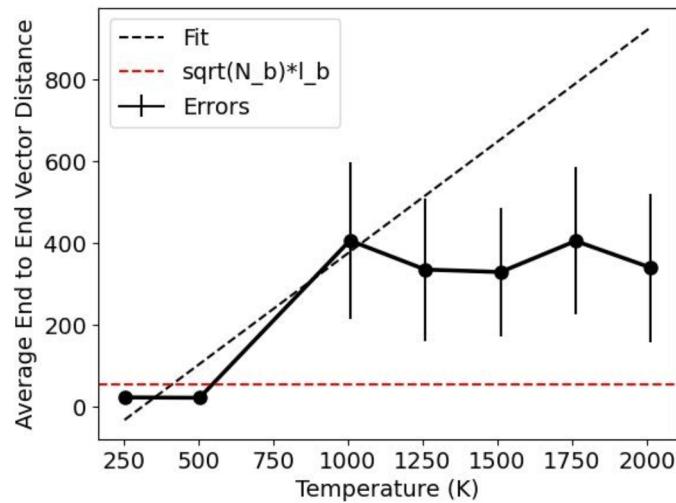


Figure C.25: $\sqrt{\langle R^2 \rangle}$ vs Temperature for PIDTBT-nC16 with a linear regression for the linear portion of the plot. Horizontal line plotted at $\sqrt{N_b} * l_b$.

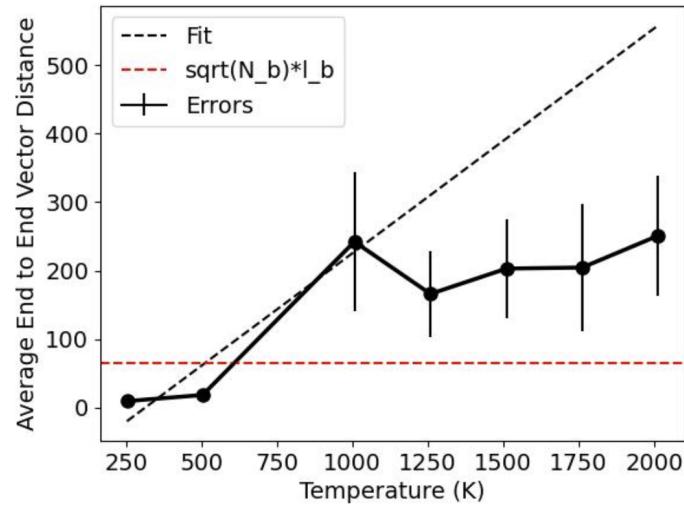


Figure C.26: $\sqrt{\langle R^2 \rangle}$ vs Temperature for PIDTCPDT-C11-BO with a linear regression for the linear portion of the plot. Horizontal line plotted at $\sqrt{N_b} * l_b$.

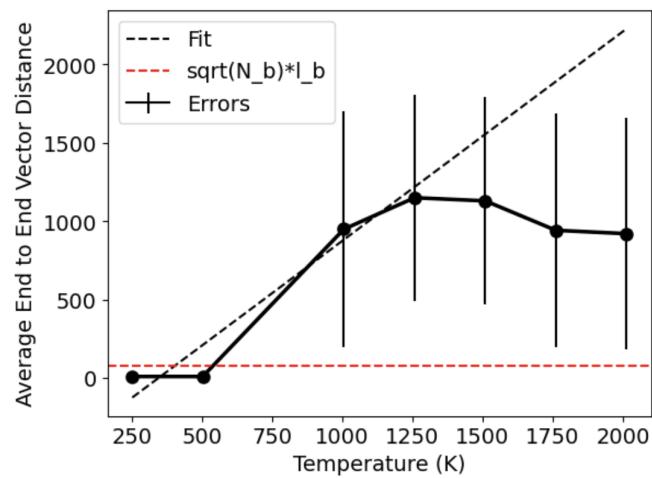


Figure C.27: $\sqrt{\langle R^2 \rangle}$ vs Temperature for PIDTBFT-C11-BO with a linear regression for the linear portion of the plot. Horizontal line plotted at $\sqrt{N_b} * l_b$.