MDPI

*Article*

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

**Madilyn E. Paul [1], Chris D. Jones [2] and Eric Jankowski [3,*]**

[1] Boise State University; madilynpaul@u.boisestate.edu
[2] Boise State University; chrisjones4@u.boisestate.edu
[3] Boise State University; ericjankowski@boisestate.edu
* Correspondence: ericjankowski@boisestate.edu; ORCID: 0000-0002-3267-1410

**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.

## 1. 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[1–6]. 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[7]. In molecular mechanics simulations, forcefields define the potential energies and therefore forces of bond, angle, torsion/dihedral, and non-bonded models of interatomic interactions[8,9]. For example, the popular Optimized Potentials for Liquid Simulations (OPLS) forcefield has mainly been built around simulating hydrocarbons and proteins [10,11]. The General Amber forcefield (GAFF) is optimized for small organic molecules[12].

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[8] 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[7]. 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.

**Citation:** Paul, M. E.; Jones, C. D.; Jankowski, E. Validating structural predictions of conjugated macromolecules in Espaloma-enabled reproducible workflows. *Int. J. Mol. Sci.* **2024**, *1*, 0. https://doi.org/

Received:
Revised:
Accepted:
Published:

**Copyright:** © 2024 by the authors. Submitted to *Int. J. Mol. Sci.* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

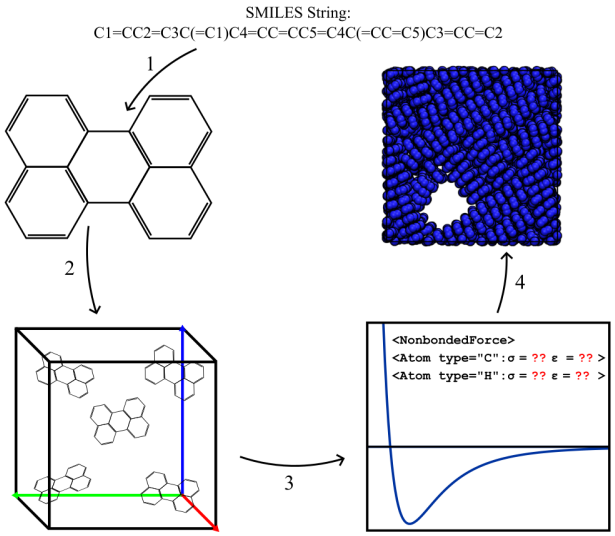


**Figure 1.** Depictions of our generalized molecular dynamics workflow. Step 1 shows the creation of an `mBuild Compound` from a SMILES string. In step 2 we create our simulation object using `flowerMD` and `PACKMOL`. We employ `Espaloma`to parameterize our molecules in step 3 and write the forcefield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

Recent efforts towards making simulations more transparent, reproducible, usable, and extensible (TRUE)[13,14] 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[15]. 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 [7]. `Espaloma` was designed for the investigation of biopolymers and has been used to accurately simulate them [16,17]. 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 1).

Here, we evaluate `Espaloma` for the modeling of highly conjugated macromolecules. We generate forcefield parameters using `Espaloma` 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 `Espaloma`'s predictive capabilities outside its core training set [7]. A schematic of the MD workflow we implemented in this work is shown in Figure 1.

## 2. Model

We consider united-atom (UA) representations of perylene and P3HT, omitting long-range electrostatics as in prior work[18,19]. 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 2. Chains of $n = 15$ repeat units are used to represent P3HT oligomers. Harmonic potentials are used to model bond-stretching (Equation 1) and angle bending (Equation 2) between pairs and triplets of bonded atoms.

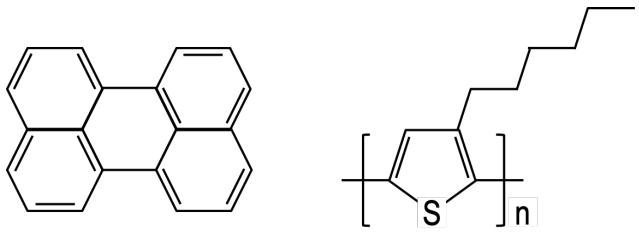


**Figure 2.** Diagram of perylene molecule (left) and poly-3-hexylthiophene (P3HT) monomer (right).

$E_{bonds}$ 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.

$$E_{bonds} = K_r(r - r_{eq})^2 \tag{1}$$

$E_{angle}$ is the potential energy of a bond angle, $K_\theta$ is the harmonic angle constant, $\theta$ is the bond angle, and $\theta_{eq}$ is the equilibrium bond angle.

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

Cosine series are used to model proper dihedrals (Equation 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, $n$, $\phi$ is the dihedral angle, and $\gamma$ is the phase angle.

$$E_{dihedral} = \sum_{n=1}^{4} \frac{K_n}{2}[1 + \cos(n\phi - \gamma)] \tag{3}$$

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

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)[20] truncated at $r_{cut} = 2.5\sigma$, where $\sigma$ is the length unit corresponding to the largest simulation element being represented. Here, this corresponds to sulfur (S) simulation elements for P3HT ($\sigma = 3.565$Å), and carbon (C1) elements for perylene ($\sigma = 3.380$Å). We apply the Lorentzian combination rule for sigma and a geometric combination rule for epsilon, shown in Equation 5.

$$\sigma_{i,j} = \frac{\sigma_i + \sigma_j}{2}, \epsilon_{i,j} = \sqrt{\epsilon_i * \epsilon_j} \tag{5}$$

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.

## 3. Methods

In this section we detail the molecular dynamics simulations performed with `HOOMD-Blue`[21] to predict equilibrium morphologies, describe the new tools we develop for integrating `Espaloma` into workflows utilize the Molecular Simulation Design Framework (MoSDeF)[22], and detail morphology characterization that underpins model validation.

### *3.1. 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`[23] 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

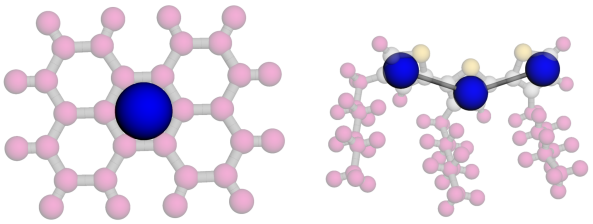

**Figure 3.** Blue spheres represent center-of-geometry positions used for RDFs and clustering criteria. Perylene (left), and P3HT (right).

throughout this work. We initialize our system using `PACKMOL`[24] within `flowerMD`[25] 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 $5 \times 10^6$ 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[26,27] 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.

### *3.2. Morphology Characterization*

Equilibrium morphologies are quantified using radial distribution functions (RDF) and simulated grazing-incident x-ray scattering (GIXS) implemented in `freud`[28]. RDFs are calculated between perylene centers of mass (Figure 3a), and for P3HT, the sulfur of the thiophene rings (Figure 3b). Both the RDF and GIXS analyses are averaged over the last 20 independent snapshots of each simulation.

To compare with prior work, we calculate an order parameter $\Psi$ as used by Miller et al. [19]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 6Å and the best-fit planes through each moiety are within 10 degrees of being parallel. The code for calculating RDFs, GIXS, $\Psi$ and phase diagrams is available at github.com/madilynpaul/Espaloma-Validation.

### *3.3. Integrating Espaloma parameterization into MoSDeF*

Our simulation workflows are built around MosDeF tools[29], in particular `mBuild`[30] 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). This workflow fits into the overall MD workflow (Section 4.2, Figure 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 5). **(3)** Pass this "`BondWalked`" molecule to `Espaloma`for forcefield parameterization, writing these parameters to a forcefield `xml` file. **(4)** Use the atom labels given by `Espaloma`to 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/madilynpaul/Espaloma-Validation.

## 4. 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.

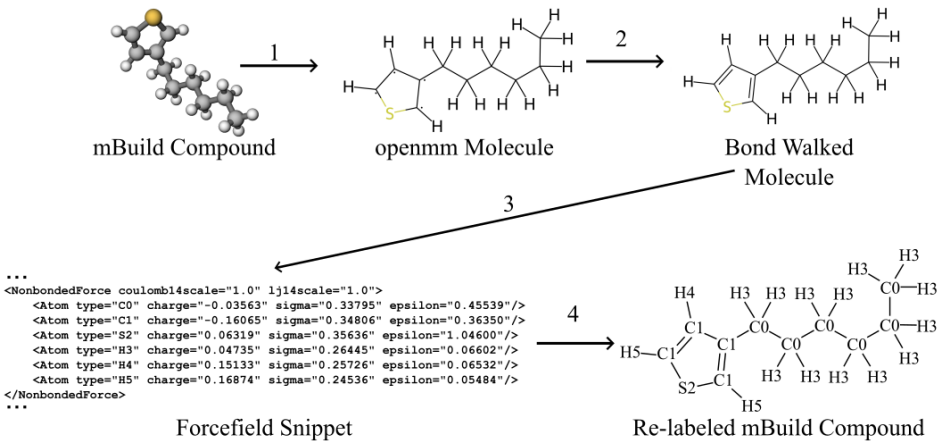


**Figure 4.** Overall `Espaloma-MosDeF` workflow. (1) `mBuild` compounds are used to create the topology of an `openMM` molecule, (2) `BondWalker` uses octet rules to determine double bonds in the `openMM` molecule. (3) `Espaloma` generates forcefield parameters for the `openMM` molecule. (4) The `ESP-UA`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

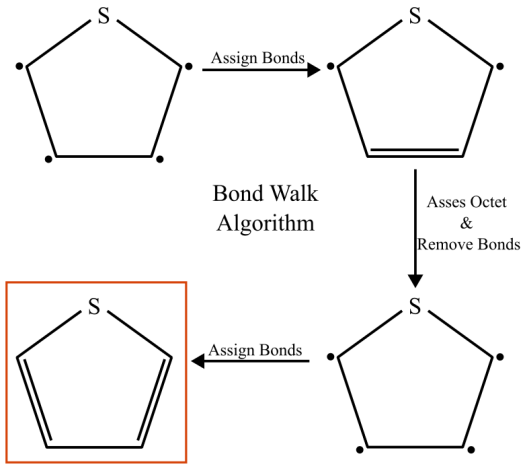


**Figure 5.** Double- and triple-bond information that is missing from `mBuild` compounds is retrieved via `BondWalker` by iteratively checking whether the octet rules can be satisfied for all atoms after incrementing the bond character adjacent to atoms with unsatisfied octets.

### 4.1. Espaloma vs OPLS-UA and GAFF

The nonbonded pair interaction for perylene and P3HT are presented in Table 1, where they are compared against `OPLS-UA` and `GAFF` parameters. We first consider the LJ $\sigma$

**Table 1.** Lennard-Jones diameters $\sigma$ and well-depths $\epsilon$ generated by `espaloma` in the present work, and `OPLS-UA` and `GAFF`.

| | $\sigma$ (Å) | | | $\epsilon$ (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 |

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 $\epsilon$ values and note that `ESP-UA` is very similar to `GAFF` once again, with `OPLS-UA` as the outlier.

Normalizing the $\epsilon_{C1}$ and $\epsilon_{C0}$ values by $\epsilon_S$ provides a clearer picture of the range of attractions in each model than simple comparison the absolute values of each $\epsilon$. 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 $\epsilon_{C1}/\epsilon_S$ and $\epsilon_{C0}/\epsilon_S$ ratios and C0 $\sigma$ 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 2. These statepoints are chosen to replicate structures sampled in prior work from Miller et al.[19]. 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[31]. We expect to observe various phases (ordered, liquid, vapor, disordered) over this range of statepoints.

### 4.2. Perylene

The ordering ($\Psi$) of perylene modeled with `ESP-UA` as a function of temperature and density is summarized in Figure 6. Consistent with Ref. [19], the most ordered structure (T = 250 K, $\rho$ = 0.5 g/cm$^3$) exhibits significant $\pi$-stacking, which is visible in Figure 7. At lower temperatures, we observe a transition from highly ordered to less ordered as the density

**Table 2.** A list of statepoints used in the P3HT and perylene simulations, as well as the base units for each simulation ($\sigma, \epsilon, M$).

| | **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 |
| $\sigma$ (Å) | 3.56 | 3.40 |
| $\epsilon$ (kJ/mol) | 1.046 | 0.360 |
| $N_{monomers}$ | 15 | 1 |

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

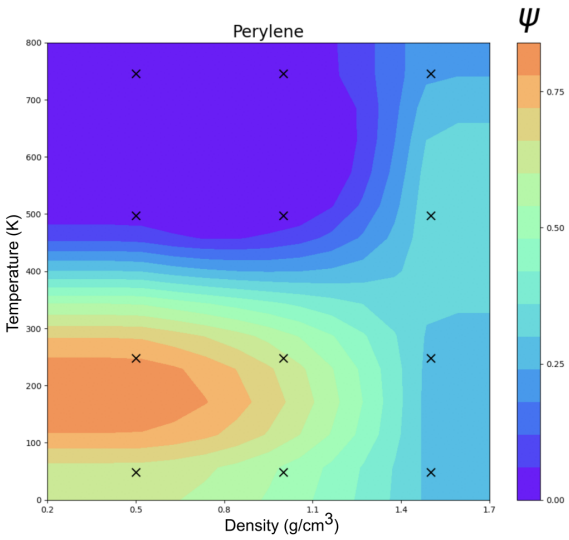


**Figure 6.** Temperature vs density clustering order parameter (Ψ) phase diagram of perylene at 12 statepoints.

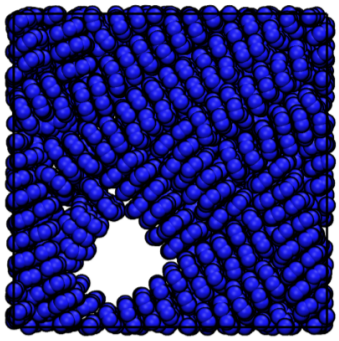

**Figure 7.** Snapshot of perylene taken from the most ordered morphology at a density of 0.5 g/cm$^3$ and temperature of 248.8 K.

order-disorder transition temperature observed here around 470 K is roughly 0.79 that of the 600 K in Ref. [18] and Botoshansky et al[32]. This is explained by the difference in base unit between `Espaloma` and `OPLS-UA` perylene forcefields: Both models have only C1 atomtypes but $\epsilon_{OPLS-UA} = 0.4602$ kJ/mol and $\epsilon_{ESP-UA} = 0.3635$ 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/cm$^3$ 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.

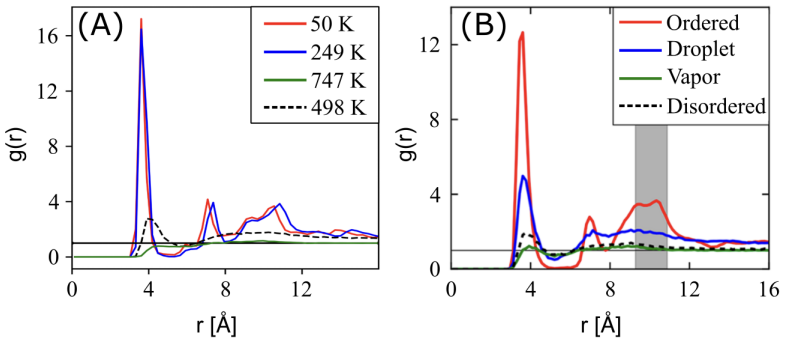


**Figure 8.** (A) Radial distribution function (RDF) of Perylene at various temperatures at a density of 0.5g/cm$^3$ generated from an ESP-UA predicted morphology. (B) RDF of perylene in various phases generated from an OPLS-UA predicted morphology. OPLS-UA RDF published by Miller, et al. at Ref [18].

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 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 $0.8A^{-1}$, while with ESP-UA here $q_y = 1.9A^{-1}$ in better agreement with Ishii and Miyasaka[33]. To briefly summarize, modeling perylene with ESP-UA parameterization

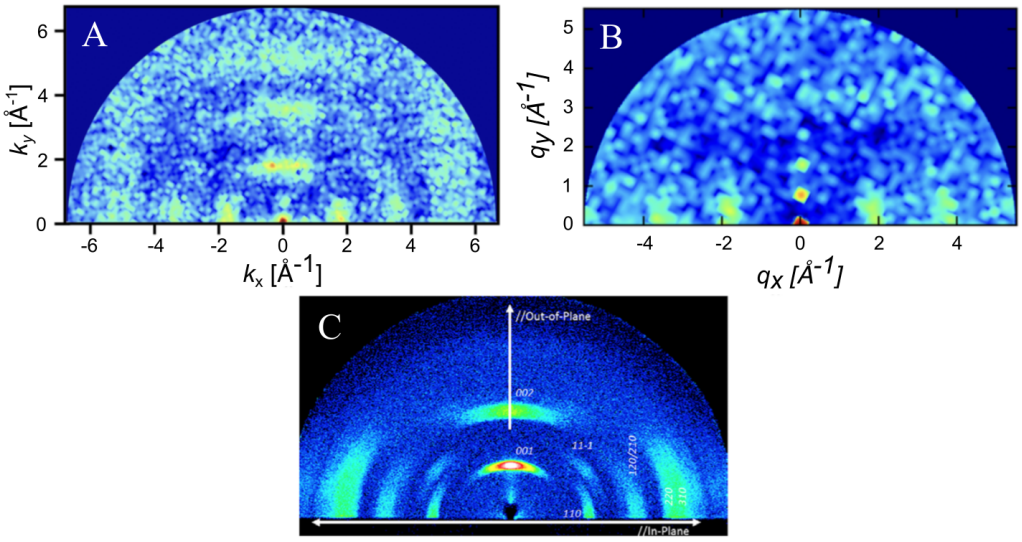


**Figure 9.** (A) Grazing incident x-ray scattering pattern of Perylene generated from an ESP-UA morphology. (B) GIXS pattern of Perylene generated from an OPLS-UA morphology. OPLS-UA GIXS pattern published by Miller, et. al. at Ref [18]. (C) Experimental XRD pattern for $\beta$-perylene, reproduced with permission from Ishii, et al. at Ref [34]. Copyright 2014 AIP Publishing LLC.

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.

### 4.3. Poly-3-Hexylthiophene (P3HT)

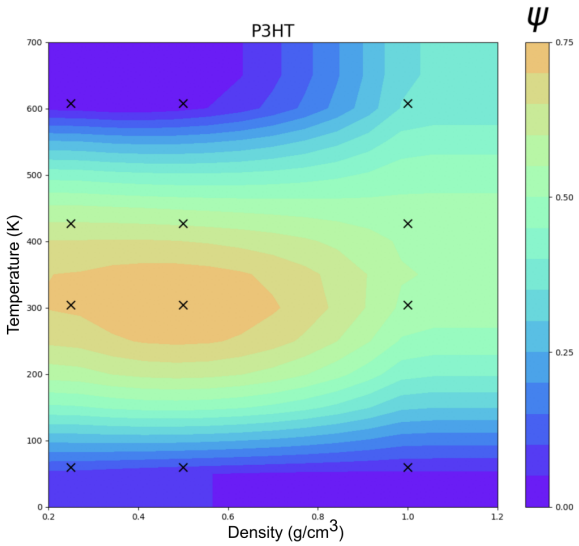


**Figure 10.** Temperature vs density clustering order parameter (Ψ) phase diagram of P3HT at 12 statepoints.

Figure 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 $\epsilon$ units between `ESP-UA` and `OPLS-UA`. `ESP-UA` gives an $\epsilon$ of 0.4554 kJ/mol for the aliphatic side chain carbons of P3HT, while `OPLS-UA` uses an $\epsilon$ 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 [35], stating that lower the $\epsilon$ values for side chains of P3HT monomers results in lamellar structures at lower temperatures than if the $\epsilon$ was higher [35]. Figure 11 shows the radial distribution function of `ESP-UA` P3HT

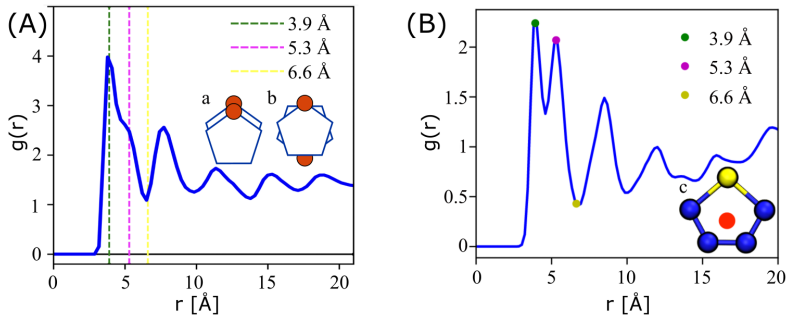


**Figure 11.** (A) Radial distribution function (RDF) of P3HT at a temperature of 304 K and a density of 0.5g/cm$^3$ generated from an `ESP-UA` predicted morphology. (B) RDF of P3HT generated from an `OPLS-UA` predicted morphology. `OPLS-UA` RDF published by Miller, et al. in Ref [19].

(left) and `OPLS-UA` P3HT (right). The `ESP-UA` RDF is generated at a temperature of 304 K and a density of 0.5g/cm$^3$. 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 $\pi$-stacking of the thiophene rings, shown in Figure 11a. The second peak aligns with the anti-aligned $\pi$-stacking of the thiophene rings, shown in Figure 11b. The `OPLS-UA` RDF is calculated using the geometric center of thiophene ring, shown in Figure 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

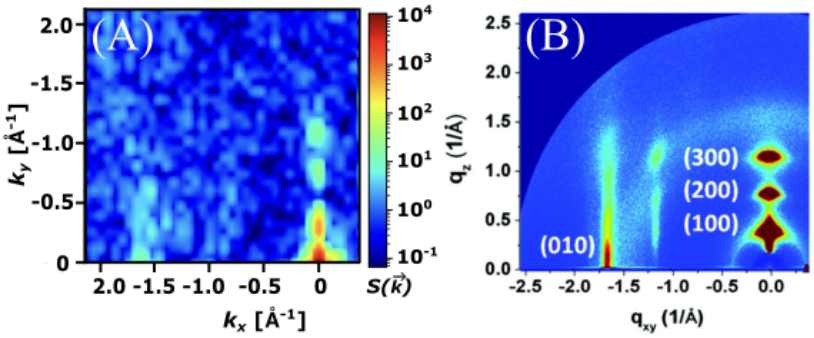


**Figure 12.** (A) Grazing Incident X-ray Scattering Pattern of P3HT at 0.5 g/$cm^3$ and 304 K generated using `ESP-UA`. (B) Corresponding experimental scattering pattern of P3HT. Reprinted (adapted) with permission from Ko, et al. in Ref [36]. Copyright 2012 American Chemical Society.

of the most ordered structure (0.5 g/cm$^3$, 304 K) displays significant correlation to the experimental scattering pattern of P3HT (Figure 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 to responsibly parameterize and represent the P3HT polymer in simulations. The lamellar structure represented by the scattering pattern in Figure 12 is shown in Figure 13.

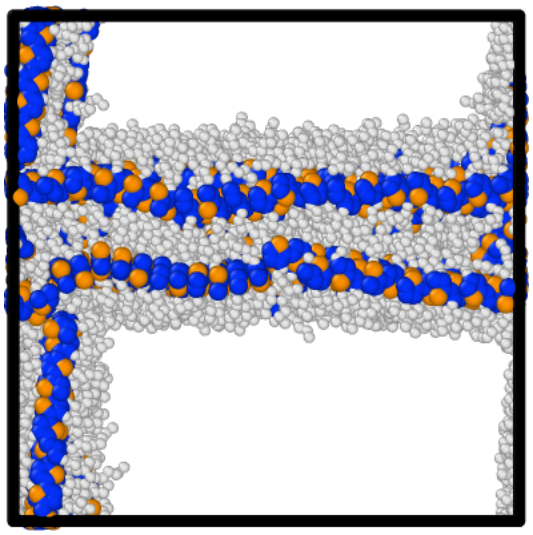

**Figure 13.** Snapshot of P3HT's most ordered morphology at a density of 0.5 g/cm$^3$ and temperature of 304 K.

## 5. 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` $\epsilon$ 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.

**Author Contributions:** Funding acquisition, Eric Jankowski; Investigation, Madilyn Paul; Methodology, Madilyn Paul; Software, Madilyn Paul and Chris Jones; Supervision, Eric Jankowski; Validation, Madilyn Paul; Visualization, Madilyn Paul; Writing – original draft, Madilyn Paul and Eric Jankowski; Writing – review editing, Madilyn Paul, Chris Jones and Eric Jankowski.

All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by NSF grant number 2411302.

**Data Availability Statement:** Example data and submission scripts for generating all data is available at github.com/madilynpaul/Espaloma-Validation.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Daglar, H.; Keskin, S. Recent advances, opportunities, and challenges in high-throughput computational screening of MOFs for gas separations. *Coordination Chemistry Reviews* **2020**, *422*, 213470. https://doi.org/https://doi.org/10.1016/j.ccr.2020.213470.
2. Liu, X.; Shi, D.; Zhou, S.; Liu, H.; Liu, H.; Yao, X. Molecular dynamics simulations and novel drug discovery. *Expert opinion on drug discovery* **2018**, *13*, 23–37.
3. Lin, X.; Li, X.; Lin, X. A review on applications of computational methods in drug screening and design. *Molecules* **2020**, *25*, 1375.
4. Afzal, M.A.F.; Browning, A.R.; Goldberg, A.; Halls, M.D.; Gavartin, J.L.; Morisato, T.; Hughes, T.F.; Giesen, D.J.; Goose, J.E. High-throughput molecular dynamics simulations and validation of thermophysical properties of polymers for various applications. *ACS Applied Polymer Materials* **2020**, *3*, 620–630.
5. Meier, M.A.; Hoogenboom, R.; Schubert, U.S. Combinatorial methods, automated synthesis and high-throughput screening in polymer research: The evolution continues. *Macromolecular rapid communications* **2004**, *25*, 21–33.
6. Quach, C.D.; Gilmer, J.B.; Pert, D.; Mason-Hogans, A.; Iacovella, C.R.; Cummings, P.T.; McCabe, C. High-throughput screening of tribological properties of monolayer films using molecular dynamics and machine learning. *The Journal of Chemical Physics* **2022**, *156*.
7. Wang, Y.; Fass, J.; Kaminow, B.; Herr, J.E.; Rufa, D.; Zhang, I.; Pulido, I.; Henry, M.; Chodera, J.D. End-to-End Differentiable Molecular Mechanics Force Field Construction. *Chemical Science* **2022**, *13*, 12016–12033, [2010.01196 [physics]]. https://doi.org/10.1039/D2SC02739A.
8. Hopfinger, A.; Pearlstein, R. Molecular mechanics force-field parameterization procedures. *Journal of computational chemistry* **1984**, *5*, 486–499.
9. Vanommeslaeghe, K.; Guvench, O.; et al. Molecular mechanics. *Current pharmaceutical design* **2014**, *20*, 3281–3292.

10. Jorgensen, W.L.; Tirado-Rives, J. 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* **1988**, *110*, 1657–1666. https://doi.org/10.1021/ja00214a001.
11. Ghahremanpour, M.M.; Tirado-Rives, J.; Jorgensen, W.L. Refinement of the Optimized Potentials for Liquid Simulations Force Field for Thermodynamics and Dynamics of Liquid Alkanes. *The Journal of Physical Chemistry B* **2022-08-11**, *126*, 5896–5907. https://doi.org/10.1021/acs.jpcb.2c03686.
12. Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and testing of a general amber force field. *Journal of computational chemistry* **2004**, *25*, 1157–1174.
13. Thompson, M.W.; Gilmer, J.B.; Matsumoto, R.A.; Quach, C.D.; Shamaprasad, P.; Yang, A.H.; Iacovella, C.R.; McCabe, C.; Cummings, P.T. Towards molecular simulations that are transparent, reproducible, usable by others, and extensible (TRUE). *Molecular physics* **2020**, *118*, e1742938.
14. Jankowski, E.; Ellyson, N.; Fothergill, J.W.; Henry, M.M.; Leibowitz, M.H.; Miller, E.D.; Chesser, S.; Guevara, J.D.; Jones, C.D.; Klopfenstein, M.; et al. Perspective on coarse-graining, cognitive load, and materials simulation. *Computational Materials Science* **2020**, *171*, 109129.
15. Klein, C.; Summers, A.Z.; Thompson, M.W.; Gilmer, J.B.; McCabe, C.; Cummings, P.T.; Sallai, J.; Iacovella, C.R. Formalizing atom-typing and the dissemination of force fields with foyer. *Computational Materials Science* **2019**, *167*, 215–227. https://doi.org/10.1016/j.commatsci.2019.05.026.
16. Davel, C.M.; Bernat, T.; Jeffery R. Wagner, M.R.S. Parameterization of general organic polymers within the Open Force Field framework. *ChemRxiv* **2023**.
17. Takaba, K.; Friedman, A.J.; Cavender, C.E.; Behara, P.K.; Pulido, I.; Henry, M.M.; MacDermott-Opeskin, H.; Iacovella, C.R.; Nagle, A.M.; Payne, A.M.; et al. Machine-learned molecular mechanics force fields from large-scale quantum chemical data. *Chemical Science* **2024**, *15*, 12861–12878. https://doi.org/10.1039/D4SC00690A.
18. Miller, E.D.; Jones, M.L.; Jankowski, E. Enhanced Computational Sampling of Perylene and Perylothiophene Packing with Rigid-Body Models. *ACS Omega* **2017-01-31**, *2*, 353–362. https://doi.org/10.1021/acsomega.6b00371.
19. Miller, E.; Jones, M.; Henry, M.; Chery, P.; Miller, K.; Jankowski, E. Optimization and Validation of Efficient Models for Predicting Polythiophene Self-Assembly. *Polymers* **2018-11-26**, *10*, 1305. https://doi.org/10.3390/polym10121305.
20. Lennard-Jones, J. The electronic structure of some diatomic molecules. *Transactions of the Faraday Society*, *25*, 668–686. https://doi.org/10.1039/TF9292500668.
21. Anderson, J.A.; Glaser, J.; Glotzer, S.C. HOOMD-blue: A Python package for high-performance molecular dynamics and hard particle Monte Carlo simulations. *Computational Materials Science* **2020**, *173*, 109363. https://doi.org/10.1016/j.commatsci.2019.109363.
22. Cummings, P.T.; Iacovella, C.R.; Ledeczi, A.; Jankowski, E.; Jayaraman, A.; Palmer, J.C.; Maginn, E.J.; Glotzer, S.C.; Anderson, J.A.; Ilja Siepmann, J.; et al. Open-source molecular modeling software in chemical engineering focusing on the Molecular Simulation Design Framework **2021**.
23. Adorf, C.S.; Dodd, P.M.; Ramasubramani, V.; Glotzer, S.C. Simple data and workflow management with the signac framework. *Computational Materials Science* **2018**, *146*, 220–229. https://doi.org/10.1016/j.commatsci.2018.01.035.
24. Martínez, L.; Andrade, R.; Birgin, E.G.; Martínez, J.M. P ACKMOL : A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry* **2009-10**, *30*, 2157–2164. https://doi.org/10.1002/jcc.21224.
25. Albooyeh, M.; Jones, C.; Barrett, R.; Jankowski, E. FlowerMD: Flexible Library of Organic Workflows and Extensible Recipes for Molecular Dynamics. *Journal of Open Source Software* **2023**, *8*, 5989. https://doi.org/10.21105/joss.05989.
26. Martyna, G.J.; Tobias, D.J.; Klein, M.L. Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics*, *101*, 4177–4189. https://doi.org/10.1063/1.467468.
27. Cao, J.; Martyna, G.J. Adiabatic path integral molecular dynamics methods. II. Algorithms. *The Journal of Chemical Physics*, *104*, 2028–2035. https://doi.org/10.1063/1.470959.
28. Ramasubramani, V.; Dice, B.D.; Harper, E.S.; Spellings, M.P.; Anderson, J.A.; Glotzer, S.C. freud: A Software Suite for High Throughput Analysis of Particle Simulation Data. *Computer Physics Communications* **2020**, *254*, 107275. https://doi.org/https://doi.org/10.1016/j.cpc.2020.107275.

29. Cummings, P.T.; MCabe, C.; Iacovella, C.R.; Ledeczi, A.; Jankowski, E.; Jayaraman, A.; Palmer, J.C.; Maginn, E.J.; Glotzer, S.C.; Anderson, J.A.; et al. Open-source molecular modeling software in chemical engineering focusing on the Molecular Simulation Design Framework. *AIChE Journal* **2021**, *67*. https://doi.org/10.1002/aic.17206.
30. Klein, C.; Sallai, J.; Jones, T.J.; Iacovella, C.R.; McCabe, C.; Cummings, P.T.; Snurr, R.Q.; Adjiman, C.S.; Kofke, D.A., A Hierarchical, Component Based Approach to Screening Properties of Soft Matter. In *Foundations of Molecular Modeling and Simulation: Select Papers from FOMMS 2015*; Molecular Modeling and Simulation, Springer Singapore, 2016; p. 79–92. https://doi.org/10.1007/978-981-10-1128-3_5.
31. Newbloom, G.M.; Weigandt, K.M.; Pozzo, D.C. Structure and property development of poly (3-hexylthiophene) organogels probed with combined rheology, conductivity and small angle neutron scattering. *Soft Matter* **2012**, *8*, 8854–8864.
32. Botoshansky, M.; Herbstein, F.H.; Kapon, M. 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* **2003**, *86*, 1113–1128.
33. Ishii, A.; Miyasaka, T. A high voltage organic–inorganic hybrid photovoltaic cell sensitized with metal–ligand interfacial complexes. *Chem. Commun.* **2012**, *48*, 9900–9902. https://doi.org/10.1039/C2CC34829B.
34. Ishii, A.; Jena, A.K.; Miyasaka, T. Fully crystalline perovskite-perylene hybrid photovoltaic cell capable of 1.2 V output with a minimized voltage loss. *APL Materials*, *2*, 091102. https://doi.org/10.1063/1.4895039.
35. Marsh, H.S.; Jankowski, E.; Jayaraman, A. Controlling the Morphology of Model Conjugated Thiophene Oligomers through Alkyl Side Chain Length, Placement, and Interactions. *Macromolecules* **2014-04-22**, *47*, 2736–2747. https://doi.org/10.1021/ma5000267.
36. Ko, S.; Hoke, E.T.; Pandey, L.; Hong, S.; Mondal, R.; Risko, C.; Yi, Y.; Noriega, R.; McGehee, M.D.; Brédas, J.L.; et al. 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* **2012**, *134*, 5222–5232. https://doi.org/10.1021/ja210954r.