Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

Using Graphs to Quantify Energetic and Structural Order in Semicrystalline Oligothiophene Thin Films†

Ellen Van,[b] Matthew Jones,[c] Eric Jankowski,[c] and Olga Wodo[a]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

In semicrystalline conjugated polymer thin films, the mobility of charges depends on the arrangement of the individual polymer chains. In particular, the ordering of the polymer backbones affects the charge transport within the film, as electron transfer generally occurs along the backbones with alternating single and double bonds. In this paper, we demonstrate that polymer ordering should be discussed not only in terms of structural but also energetic ordering of polymer chains. We couple data from molecular dynamics simulations and quantum chemical calculations to quantify both structural and energetic ordering of polymer chains. We leverage a graph-based representation of the polymer chains to quantify the transport pathways in a computationally efficient way. Next, we formulate the morphological descriptors that correlate well with hole mobility determined using kinetic Monte Carlo simulations. We show that the shortest and fastest path calculations are predictive of mobility in equilibrated morphologies. In this sense, we leverage graph-based descriptors to provide a basis for the quantitative structure property relationships.

Design, System, Application

This work deals with morphology quantification for organic electronic devices. The main finding is the identification of salient morphological features governing the hole mobility for a set of samples. Our findings provide a quantitative basis for morphology characterization, that is useful for the design of new candidate polymers and fabrication of highly efficient devices. The key element of our methodology is a highly computationally efficient graph-based representation of morphology that facilitates feature extraction. More importantly this representation can be directly used in high throughput screening of well/poorly performing morphologies. Our results constitute an important step towards establishing quantitative structure-property relationship, morphology optimization and materials design for high performing organic electronic devices.

1 Introduction

The nanomorphology of polymer thin films critically affects performance in organic optoelectronic devices such as organic solar cells[1], light emitting diodes[2], and transistors[3]. However, many aspects of the underlying physics of charge transport are still poorly understood. In general, the process of charge migration through the device heavily depends on nanomorphology: the spatial arrangement of individual polymer chains and their mutual ordering[4]. Charges migrate via discrete tunneling events - hops - between monomers of polymer chains. Hopping rates are highly sensitive to the relative position of adjacent polymer chains and the overlap of the neighboring molecular orbitals. In the semicrystalline polymer systems used to make organic electronics, charge transport is expected to be higher in the ordered regions and lower in the disordered regions, where high energy barriers inhibit charge migration. However, if energy barriers between ordered regions are too high, then charges may be trapped within the ordered phase and not contribute to bulk charge transport[5]. There are two potential solutions to the problem of trapped charges: (i) lower energy barriers (e.g. by increasing local polymer chain ordering or by selecting alternative polymer chemistries) or (ii) increase connectivity (e.g. by including long entangled polymer chains that connect disparate ordered domains).

Implementing either solution would benefit from quantitative insights into the structural features that contribute to connectivity and energetic barriers to charge transport. Such structural detail is critical to informing the design of materials and morphology in organic electronic devices but is yet to be fully quantified. In particular, studies of the electronic coupling and charge transport

[a] Department of Materials Design and Innovation, University at Buffalo, Buffalo, New York, USA Tel: +1 716 645 1377; E-mail: olgawodo@buffalo.edu
[b] Department of Mechanical and Aerospace Engineering, University at Buffalo, Buffalo, New York, USA; Currently at iiMED, E-mail: etvan@buffalo.edu
[c] Micron School of Materials Science and Engineering, Boise State University, Boise, Idaho, USA.

† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/b000000x/

of organic molecules are typically carried out *ex situ*, or given an assumed conformation[6–11]. In this work, we present a new multi-scale method for capturing both the energetic and structural aspects of the morphology governing charge transport properties. Specifically, we present a graph-based morphology representation that converts results from molecular dynamics simulations into data structures amenable to well-established algorithms[12] for quantifying the possible charge transport pathways in the graphs. We perform coarse-grained molecular dynamics simulations (CGMD) containing 250 oligomers of the frequently-studied donor polymer poly(3-hexylthiophene) (P3HT), each 15 monomers long. We then convert each morphology into a labeled, weighted, undirected graph, with nodes in the graph corresponding to polythiophene backbones and alkyl sidechains. Representing the morphology as such enables extraction of hierarchical features, here with particular focus towards the structures we hypothesize relate to charge transport. For example, we first query the graph/morphology to identify the shortest paths from any charge-carrying element to the bounding electrodes. Subsequently, we characterize these paths based on their physical lengths and carrier travel times, as well as features including the tortuosity of each route and the fraction of hops between (inter-) or along (intra-) chains en route to an electrode.

The unique feature of our approach is the flexibility to label edges between individual simulation elements (here, "beads") by Euclidean distance, energy difference, hopping rate, or any combination of these characteristics. This ability facilitates the testing of hypotheses about which features matter most to the bulk electronic properties of the material. The distances between individual beads used in our graph analysis are determined from equilibrium coarse-grained molecular dynamics simulations, from which sub-nanometer resolution of the morphological aspects of the thin films can be obtained. The electronic features such as hopping rates and energy level differences between beads are calculated from first principles after atomistic detail is projected onto the coarse-grained representations. This is the first attempt to combine data on the electronic-structure with the atomistic-level data on morphology to provide computational framework to establish quantitative structure-property relationships in organic thin films. Although morphology has been previously represented as graphs, the quantification has been focused on either meso scale[13,14] or atomistic level[15]. The work reported in this paper in an interesting step towards convergence of representation in multiscale modeling, as graphs have been mostly used to represent molecular structure[16–18].

To showcase our new method we present insights into the charge mobility of a collection of oligothiophene samples that are enabled by our graph-based approach. A kinetic Monte Carlo (KMC) model of charge transport is used to compute the hole mobilities for ordered and disordered oligothiophene morphologies[19]. The KMC simulations reveal that charge mobilities depend on morphology, varying over two orders of magnitude and in agreement with experimental measurements. Using our graph-based computational framework, we demonstrate that metrics based on the shortest paths along the network of polymer chains are sufficient for explaining how charge mobility depends on structure. We identify purely structural features, e.g. the tortuosity of the shortest path, that permits the charge transport properties of a morphology to be predicted. We find, for the morphologies studied here, the bulk mobility calculated by KMC is strongly correlated to the fastest charge pathways calculated from the graphs. The agreement of graph-based predictions with the more computationally demanding KMC calculations over two orders of magnitude of charge mobility indicates the potential for using such structural techniques to efficiently evaluate structures obtained from both simulations and experiments. Our results indicate that optimizing the morphology with respect to fastest pathways should lead to improved performance.

2 Morphology representation and quantification

Recent studies providing insights into the morphology of organic thin films have used both experimental measurements (e.g. tomography TEM[20]) and computational techniques (including molecular dynamics[15], phase field[21,22], kinetic Monte Carlo[23–26], and quantum chemical calculations[6,10,27,28]). These techniques provide structural information ranging from electronic structure on the molecular scale[6,10,27,28], up to the device scale[21,29–31]. For example, the tomography-related reconstruction typically operates on the intensity fields averaged over some nanoscale volume, the molecular dynamics simulations provide data on the position and trajectory of individual atoms, and quantum chemical calculations (QCC) like density functional theory simulations (DFT) gain insight on the energetic level. Consequently, structural descriptors vary significantly in scale and information, and translate into a diverse set of structural features that can be extracted from morphological samples. In our prior work we demonstrated that structural metrics including domain sizes, path lengths, and domain connectivity can be computed in a straightforward way from voxel-based morphology representations, providing a way to evaluate structures from either simulations or experiments[14,32]. However, it remains challenging to draw a coherent picture of which metrics at one scale impact behaviors at another scale due to the disparate length and time scales over which charges move (Figure 1).

In this paper, we work towards closing this gap by extracting structural features over several orders of magnitude, i.e. Å to 100 nm (Figure 1). In particular, we analyze the data made available by multiscale coarse-, atomistic-, and electronic structure-level studies. The morphologies we analyze were generated with CGMD[33,34], in which the locations of electronically-active chromophores in pristine polymer films are determined over tens of nanometers as a function of temperature. Chromophore positions are determined from the equilibrated ensembles of microstates using molecular dynamics simulations and electronic structure calculations, as described in section 3, using the MorphCT software package[35], following the recipe outlined in Ref. 19.

Equivalence between graph and morphology

The key idea for our morphological quantification is to represent morphologies as undirected, labeled, weighted graphs (Figure 2).

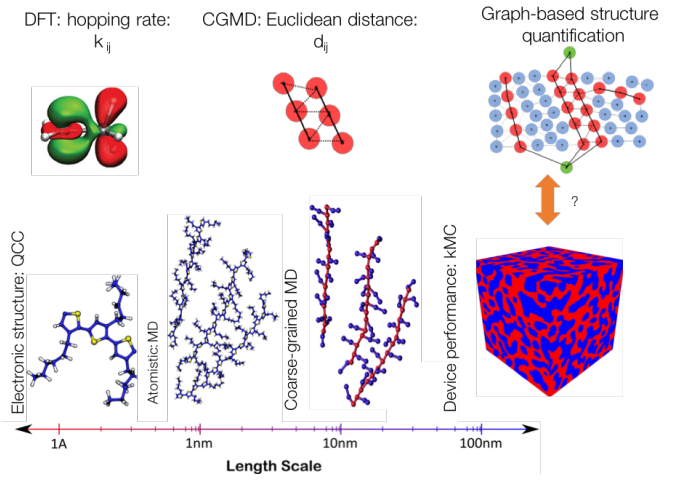

Fig. 1 This work combines hopping rates on sub-nanometer length scales with network connectivities over 100+ nanometers to infer the conditions needed for fast charge transport in organic electronic films. The bulk charge transport properties of an organic thin film depends on how the electronic orbitals of chromophores are locally oriented with respect to each other, the packing of polymer chains, and the connections between disparate nanostructured elements.

The beads from each CGMD simulation can be naturally represented as vertices with labels denoting whether the beads represent backbone or side chain simulation elements. Beads are connected via edges expressing their relation in terms of connectivity or local distances. Each edge is associated with a weight and a label that expresses the type of edge and distance between beads.

We distinguish two types of vertices: backbone and side chain vertices. We also add meta-vertices that correspond to the sample boundary associated with the electrodes (green and orange vertices in Figure 2). In the most general case, we can distinguish four types of edges: (i) Intra-molecular edges (backbone edges) connecting beads belonging to the same polymer chain. (ii) Inter-molecular edges connecting backbone beads belonging to adjacent polymer chains. (iii) Side-chain edges connecting side chain beads with each other and with the backbone. (iv) Secondary edges connecting all other types of beads. In this paper, however, we focus on two types: intra-molecular and inter-molecular edges, as these are directly related to hole hopping events of central importance to bulk charge transport.

Formally, we represent the morphology as a labeled, weighted, undirected graph $\mathscr{G} = (\mathscr{V}, \mathscr{E}, \mathscr{W}, \mathscr{L})$, see Figure 2. In such a graph, each bead in the CGMD morphology becomes a graph vertex, with $\mathscr{V}$ being the set of all vertices in the morphology. An associated labeling function, $\mathscr{L}$, assigns a label to each vertex in $\mathscr{V}$. Function $\mathscr{L}$ associates a color (red) to represent active bead along the backbone of the polymer, and another color (blue) to represent beads belonging to the side chains. Thus, $\mathscr{L}$ acts on each vertex in $\mathscr{V}$. Each vertex is connected to its neighboring vertices through edges. The set of all edges in the graph is denoted as $\mathscr{E}$. Finally, the weight function, $\mathscr{W} : \mathscr{E} \rightarrow R_+$, assigns a non-negative real weight to each edge in $\mathscr{E}$. In the very basic case, the weight function simply encodes the physical distance between beads. However, the function can encode any other physics-based features, such as the hopping rate of charge transfer between active beads in the morphology, which intrinsically takes into account the electronic coupling between chromophores.

Given the above definitions, the graph is constructed by the following steps:

1. Given a set of beads from CGMD simulation with (x, y, z, id), the set of vertices $\mathscr{V}$ is created. For each bead in CGMD morphology, one vertex is initiated in the graph $\mathscr{G}$. Next, each vertex has a label assigned using a mapping function defined for a given problem. Here, we define a set of three labels: backbone bead (marked red) and side chain bead (marked blue), electrode meta vertex (marked green and orange to distinguish between top and bottom electrode).

2. Next, the set $\mathscr{E}$ of edges connecting vertices in $\mathscr{V}$ is constructed. The set is determined collectively for all beads in the morphology by computing the Voronoi tessellation[36]. In particular, once the Voronoi tessellation is determined for each vertex $\sqsubseteq$, the list of nearest neighbors is retrieved. Given a vertex $\sqsubseteq$ and the list of each vertex $\sqcap$ in its nearest neighborhood, the set of edges $e(\sqsubseteq, \sqcap) \in \mathscr{E}$ is added to the graph G.

3. Each edge created in the previous step is annotated with a label $\updownarrow \in \mathscr{L}$. In this paper, we focus on two edge labels: intra-molecular and inter-molecular edges. Similar to set $\mathscr{V}$, intra-molecular edges (belonging to the same chain) are easily identified based on bead indices. Inter-chain edges (connecting adjacent beads that belongs to different chains) are determined from Voronoi tessellation.

4. Each edge in $\mathscr{E}$ is also annotated with a weight $\sqsupseteq \in \mathscr{W}$ that encodes the distance between two beads. Weights are used to calculate the shortest path. We take two approaches to define $\mathscr{W}$. In the first approach, we define a function that calculates the physical Euclidean distance between beads directly from (x, y, z) coordinates. In the second approach, we define a function encoding the hopping rates of the charges between pairs of beads, which are calculated using quantum chemical calculations (see section 3).

In summary, graph construction depends on several mapping functions that facilitate the conversion of the MD morphology to the graph. We encode local transport information in the application of the mapping functions. More importantly, this representation provides flexibility while simultaneously encoding multiscale data (CGMD and QCC) in a single graph.

Structural quantification

Once CGMD morphologies are converted into graphs, we perform calculations based on these graphs to characterize the morphology. With the graph-based morphology representation, the morphology characterization problem becomes a graph querying problem. In particular, the goal is to extract or query the graph such that various features of the underlying graph are determined. Specifically, in the context of transport properties, two types of features are of interest: the connectivity of vertices and

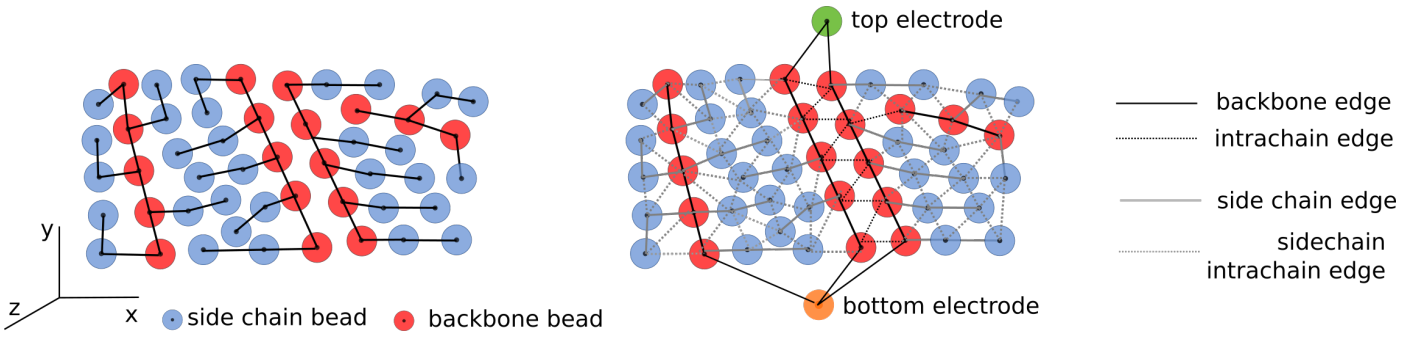

Fig. 2 The chemical topologies of the P3HT chains studied here inform the intra-molecular connections between beads (left), and we generate intermolecular edges with a Voronoi tesselation. "Boundary" vertices (denoted in green and orange) are connected to backbone beads within a cutoff distance of a periodic boundary to represent the electrodes of a hypothetical device.

the charge pathways through the morphology. To extract these features we use two general, well-studied, and efficient algorithms: the connected component algorithm, and Dijkstra's algorithm[12]. The set of features to be extracted from the graph depends on the properties of interest and the level of details encoded in a graph via labeling and weighting functions.

In the next subsections, we detail two levels of structural quantification. We first discuss a list of purely math-based descriptors. Next, we detail physics-based descriptors. We make this distinction to show the hierarchy of structural quantification but also to separate generic descriptors from problem-specific descriptors. Before providing details on descriptors, we start by introducing the graph filtering operation. This is a basic step of graph querying used in all our procedures to calculate the descriptors.

Graph filtering

Graph filtering is an important operation as it facilitates tailoring queries. Filtering is an operation during which some elements (e.g. edges) in the input graph are virtually masked while retaining only those elements satisfying specific criteria. For example, one may be interested in constructing a filtered graph, $\mathscr{G} = (\mathscr{V}, \mathscr{E})$, where $\mathscr{E}$ is a set of edges connecting vertices of the same color, e.g. red vertices corresponding to the backbone beads. This is of importance in conjugated polymeric systems, as holes hop along the backbone bead of polymers. Therefore, it is of interest to focus on the subset of vertices connected only via specific types of edges. Graph filtering is performed by defining a predicate (boolean-value function) that can operate on edges or vertices. Depending on the specific queries, the same input graph can be masked in several different ways to reflect the physics-centered aspects of the characterization. More importantly, filtering allows us to define several filter variants while operating on the same graph. In practice, it means that morphology quantification can be performed efficiently without the need to reinitialize the graph for a specific query.

Math-based descriptors

We start by introducing the connected component in graphs. Formally, a connected component is a subgraph where a path between any pair of vertices in this subgraph exists. We use the connected component algorithm[12] to identify the set of connected components in the graph representing morphology. We first construct a filtered graph consisting of vertices of a given type that are connected by specific types of edges. Next, we determine the total number of connected components in the filtered graph. Finally, we extract basic statistics such as the number of vertices constituting a given component.

Quantifying the number of connected components is of importance as it allows us to assess if the morphology is interconnected throughout the entire sample, or if it is fragmented into smaller subdomains. More importantly, we use it to identify connected components with a direct connection to the sample boundary. This is done by inspecting the neighborhood of the green meta-vertices representing the top electrode. We further compute the fraction of vertices with direct connections to the electrode. If the fraction of the vertices is high, then one could expect that carriers will be able to be extracted from the active layer. In this sense, this simple math-based descriptor can be directly used as performance indicator to differentiate good from bad morphologies.

Another math-based descriptor is based on the shortest pathway calculations. The shortest path between a source vertex $\int \in \mathscr{V}$ and a target vertex $\sqcup \in \mathscr{V}$ is the path with the smallest sum of weights of all paths between $\int$ and $\sqcup$ in $\mathscr{G}$. This is a single-source shortest path problem and so we use the Dijkstra algorithm[12,37] to determine the shortest path that a hole would need to take to reach the electrode. In particular, we are interested in determining the shortest pathways consisting of backbone vertices only with direct connection to the boundary. We translate this question into execution of this algorithm in the following way. First, we filter the graph by masking all vertices other than backbone. Next, we set the green meta vertex as source vertex. In this way, the goal is to find the shortest pathway from the meta vertex to all backbone vertices in the graph. It is achieved by executing the Dijkstra algorithm once on the entire filtered graph, which computes the shortest path for all vertices.

Once the pathways are identified, we extract their features. For example, for each path, we count the number of intra and inter-molecular edges. This information is of importance in the context of conjugated polymers, because of the nature of the hopping events. Hopping along backbone edges is considered as a fast event, while hopping events between polymer chains are expected to occur more slowly[38]. Simply by counting hop-

ping events along individual pathways, we can assess the performance of a device made with this particular morphology. To facilitate comparison between morphologies, we ensemble information for all pathways by plotting features in the form of histograms and subsequently computing basic statistics such as mean path lengths.

2.1 Physics-based descriptor

Math-based descriptors are further post-processed to construct application-specific descriptors. In this paper, we seek to explain the large mobility difference between several samples by leveraging connected components and shortest pathway calculations. We define three levels of physics-based quantification. We first calculate the tortuosity of paths from each active bead. Next, we compare the frequency of intra- and inter-molecular hops along the pathways. Finally, we look into the fastest rather than the shortest pathways.

The tortuosity is defined as the the ratio of the path length through the media (taking into account various structural obstacles) to the length of corresponding straight path. The straight path is considered as the ideal reference path. By this definition, a tortuosity of $\tau = 1$ describes a path through the morphology that is identical to the straight path. The larger the tortuosity, the more tortuous the pathway is.

Each shortest pathway is also inspected to find the number of hops along and between polymer chains. Each pathway is retrieved along with the corresponding edge labels. Next, simple counting is applied to determine the number of intra-molecular (backbone) edges and inter-molecular hops constituting individual pathways. The ratio of inter-molecular hops to the total number of hops in the given pathway is determined. In general, this descriptor allows us to assess the structural order. If the polymer chains are nicely aligned then the shortest paths should be built from intra-molecular edges. For highly disordered morphologies, no clear pattern should be observed.

All descriptors defined so far rely on the Euclidean distance as a edge weight. The next descriptor weights the edges based on the electronic structure of polymer backbone. In particular, we find the fastest rather than shortest path from each vertex to the electrode. Dijkstra's algorithm is executed on the same graph with one difference: edges in the graph are weighted by hopping time calculated from the local hopping rate – k_{ij} (see subsection 3 for details), instead of the Euclidean distance between beads i and j. The corresponding total travel time is computed for each i pathway, by summing the travel time between adjacent bead (j) and $(j+1)$ belonging to the pathway:

$$t_i = \sum_{j}^{n-1} d_{(j)(j+1)}/k_{(j)(j+1)} \tag{1}$$

where $d_{(j)(j+1)}$ is the Euclidean distance and $k_{(j)(j+1)}$ is the hopping rate between between bead (j) and $(j+1)$, respectively. Similar to previous descriptors, we calculate this descriptor for each red vertex (active bead) in the graph and determine the histogram for the ensemble of paths.

Finally, we also quantify structural ordering using the orientational order parameter[39]:

$$S_2 = \left\langle \frac{3}{2}\cos^2\theta - \frac{1}{2} \right\rangle \tag{2}$$

where θ is the angle between the unit end-to-end vector of a backbone with respect to the average end-to-end backbone vector and the angle brackets denotes the average over all backbones in a snapshot. S_2 is often used to distinguish isotropic ($S_2 \approx 0$) and nematic phases ($S_2 \approx 1$) in liquid crystals, and is thus a useful descriptor for the alignment of P3HT oligomers studied here.

The set of graph-based descriptors can be expanded depending on type of questions to be asked and data available. More details on other descriptors can be found in our other papers[14,20,40].

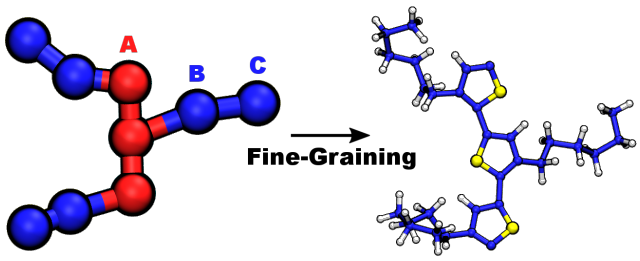

Fig. 3 The P3HT model used in this investigation, depicted using both the coarse-grained and atomistic representations of three repeat units. The blue coarse-grained beads represent three alkyl carbons and their associated hydrogens and the red coarse-grained beads represent thiophene rings.

3 Multi-scale data generation

MD simulations

The morphologies explored in this investigation were generated previously using coarse-grained molecular dynamics simulations performed by Marsh et al.[33,34]. Only the most important components of the model will be discussed here. For more details, including a discussion of coarse-grained forcefield, selected statepoints, equilibration criteria, and the simulation volume parameters, the interested reader is directed to Marsh et al. and the accompanying citations therein[34].

Systems contained 250 chains of 100% regioregular P3HT, each consisting of 15 monomers. Simulating oligomers as opposed to longer polymers is advantageous as the systems equilibrate in significantly less computation time, while still providing excellent structural agreement with experimental morphologies consisting of long chains[33]. Additionally, these chains ensure that we can study periodic, experimentally relevant simulation volumes ($\sim$ (10 nm)^3) and densities ($\sim$ 1.1 g/cm^3) without including unphysical self-interactions across the periodic boundary condition. Such interactions can lead to chain entanglements, which prevent chains from crystallizing out of the amorphous state[41]. Each monomer was modeled using 3 coarse-grained beads - one corresponding to the thiophene ring, and two corresponding to the flexible hexyl sidechain, as can be seen from the coarse-grained depiction of the model in Figure 3. All beads in the system had unit diameter σ, the fundamental distance unit, where

$\sigma = 3 \times 10^{-10}$ m. Bead masses were calculated in terms of the fundamental mass unit $M = 1.4 \times 10^{-25}$ kg. The interactions between beads were mediated by a simplified version of the atomistically-derived forcefield proposed by Huang et al.[42–44], which has obtained good agreement with both structural and electronic data in the past[33,34,41]. Pairs, triplets, and quadruplets of bonded beads were constrained by harmonic bonds, angles and dihedrals respectively, while non-bonded pair interactions were governed by a Lennard-Jones interaction. Energies are given in terms of the energy unit $\varepsilon = 2.7 \times 10^{-21}$ J. The complete set of forcefield parameters can be found in the SI Section S1.

The primary simulations were performed using the GPU-accelerated molecular dynamics simulation suite HOOMD-Blue[45]. Initially, the chains were randomly arranged in a large, cubic, periodically-bound simulation volume, which was then equilibrated isobarically and isothermally using the NPT ensemble at a pressure of $P = 0.1\varepsilon/\sigma$ (98.7 atm) and a temperature $T = 2.5\varepsilon/k_B$ (216 °C). This had the effect of shrinking down the simulation volume to obtain a system with experimentally relevant density $\rho = 1.1$ gcm^{-3} at $T = 25°C$. The morphology was then cooled by several successive phases of simulations at constant pressure, $P^* = 0.1$, to $T^* = 1.5$. At each phase, T^* was decremented by 0.25 dimensionless units, and then permitted to equilibrate, in order that 52 ± 12 statistically independent trajectory frames could be obtained. Equilibration times were determined by measuring the time taken for the potential energy (a proxy for molecular structure) to statistically decorrelate, the full details of which can be found in the SI Section S2. Although the equilibration times were strongly temperature dependent, on average $2.9 \times 10^8 \pm 1.6 \times 10^8$ time steps were required for the structure to reach equilibrium. Each time step corresponded to 0.0004τ, where τ is the fundamental time unit $\tau = \sqrt{M\sigma^2/\varepsilon} = 2.16 \times 10^{-12}$ s, resulting in an average equilibration time of ~250 ns. The example morphologies are shown in Figure 4.

Fine-graining

In order to obtain the electronic properties of the chains in the equilibrated morphologies, quantum chemical calculations must be run. These calculations require atomic resolution, necessitating a "fine-graining" process that can return the coarse-grained simulations to a realistic atomistic morphology, where the centers-of-mass of the coarse-grained beads are the same as the centers-of-mass of the atomistic moieties/functional groups that they represent. By fine-graining in this way, we can leverage the computational efficiencies, large simulation volumes, and long equilibration times afforded by coarse-grained molecular dynamics simulation, while still obtaining useful electronic structure information. There are several viable fine-graining methodologies available that have obtained good structural agreement with experiment - particularly in the fields of organic and biomolecular simulation[46–48]. In this investigation, we use the complete coarse-grained to charge-transport methodology outlined by Jones and Jankowski[19]. Firstly, the atoms corresponding to each atomistic moiety (e.g. a thiophene ring for coarse-grained beads of type A; see Figure 3) are projected around the center-of-mass of the corresponding coarse-grained bead, with arbitrary rotation. The resultant molecular morphologies are not physical, but a series of canonical NVT molecular dynamics simulations can be performed to relax the system, with the centers-of-mass constrained to ensure that the equilibrated structure obtained from the coarse-grained simulations is not modified. Generally, this series begins with a simulation where non-bonded pair interactions are disabled and the chains are permitted to relax according to the intra-molecular forcefield. This allows molecular moieties to obtain a physical rotation within the molecule, but results in significant atomic overlap between molecules. As such, the simulations is followed by a relaxation using a "soft" dissipative particle dynamics pair potentials where atoms are permitted to intersect each other but overlap is gradually reduced throughout the morphology. Finally, several short molecular dynamics simulations are performed sequentially in which the atomistic forcefield is gradually introduced to eliminate any atomic overlaps and obtain a final relaxed conformation. For this investigation, a sequence of 8 molecular dynamics simulations was found to produce realistic chain conformations, the specifications of which are listed in SI Section S3.

Quantum Chemical Calculations

In order to perform kinetic Monte Carlo simulations that predict charge mobilities using the fine-grained atomistic representations we require carrier hopping rates between "chromophores." A chromophore is defined as an electronically-active moiety over which a charge carrier can be assumed to be delocalized. DFT simulations have shown that, in P3HT, carriers are delocalized over a region consisting of around 7 monomers[49,50]. Such long delocalization lengths are difficult to reconcile in an oligomeric system consisting of 15-mers. Instead, we make the assumption that each chromophore consists of a single P3HT monomer, providing parity with the graph-based model described above, where each monomer is represented by a single graph vertex. We find that the hopping rate distribution for monomeric chromophores is strongly bimodal, with hops within the same chain occurring several orders of magnitude faster than hops between chains. Intra-molecular hops are therefore comparatively instantaneous, suggesting that carrier delocalization lengths across several monomers drop out as a natural feature of our model. In addition, utilizing monomeric chromophores removes the requirement to recalibrate the chromophore length in the model when considering the charge transport of other systems. Our model is therefore more generally applicable and useful, as it can be used to simulate different polymers without prior knowledge of the molecule's delocalization length.

In this work, the energy splitting in dimer method is used to calculate the electronic transfer integral from one chromophore (denoted as subscript i) to another (subscript j)[51]. The method estimates electronic orbital overlap between the chromophore pair by determining the frontier molecular orbitals of the chromophore when it is isolation, and calculating the splitting as the second chromophore is placed with the correct relative position and orientation as determined by the molecular morphology. In

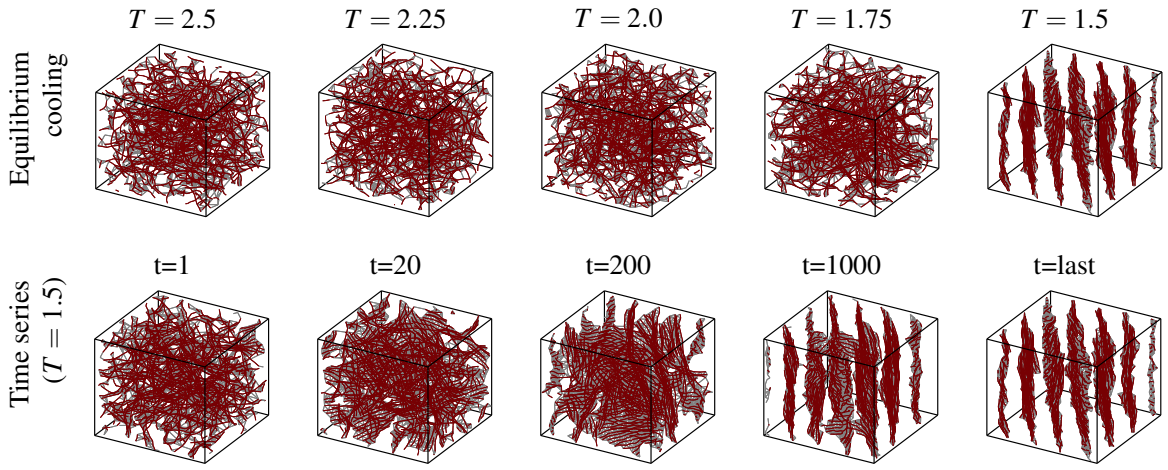

Fig. 4 Representative snapshots of backbone structure from CGMD simulations as a function of equilibration temperature (top panel: $T = 1.5$ to $T = 2.5$) and simulation time (bottom panel: $t = 1$ to 1500 million time steps).

donor materials, the important molecular orbital is the highest occupied molecular orbital (HOMO) in the isolated case, which splits to form a new HOMO and HOMO-1 when the chromophore pair is considered, whereas for acceptors, the lowest unoccupied molecular orbital (LUMO) splits to form a LUMO and LUMO+1. The magnitude of the splitting between the donor P3HT chromophores identified in this investigation is related to the electronic transfer integral, T_{ij}:

$$|T_{ij}| = \frac{1}{2}\sqrt{(E_{\text{HOMO}} - E_{\text{HOMO-1}})^2 - (\Delta E_{ij})^2}, \qquad (3)$$

where ΔE_{ij} describes the difference in HOMO energies between the two chromophores i and j in isolation, and $(E_{\text{HOMO}} - E_{\text{HOMO-1}})$ represents the splitting of the HOMO level observed when the pair is considered. Transfer integrals T_{ij} are calculated between all chromophore pairs that share a Voronoi cell edge within the morphology, again providing parity with the graph-based model. Note that, using this method, the energy levels of the chromophore pair are calculated with the pair *ex situ*, so only interactions between the molecular orbitals of the two chromophores in question are considered, disregarding any higher-order interactions with other nearby chromophores. This therefore represents a "best-case scenario" for the transfer integral - the interactions with other nearby chromophores would likely reduce the overlap and splitting of the considered pair, decreasing the electronic coupling between them.

The orbital energies for single- and pairs-of-chromophores are determined from semi-empirical ZINDO/S calculations using the ORCA quantum chemical simulation suite[52]. While calculations at the ZINDO level tend not to be as rigorous as more computationally expensive levels of DFT, they can quickly obtain molecular orbital energy spectra and are often used to determine excitonic[53] and electronic transfer integrals[54] for a variety of systems, including siloles[55], copolymers[56], and polythiophenes such as P3HT[19,41,53]. Furthermore, such calculations are expected to provide good agreement with *ab initio* quantum chemical methodologies for systems similar to those studied here[57,58]. Generally the energy calculations required to obtain the molecular orbital energy spectrum for a single P3HT monomer takes around a second on a desktop computer. This is beneficial, as our simulations of 250 chains contained 3750 chromophores consisting of 27 atoms, and more than 40,000 chromophore pairs - an untenably large dataset to analyze using conventional DFT methods. ZINDO/S is therefore an excellent candidate for our uses, with comparable computational cost, accuracy, and failure rates when calculating single-point energies to cutting-edge quantum chemical methodologies such as the recently proposed IPEA tight-binding method[59].

Kinetic Monte Carlo Simulations

The network of transfer integrals calculated through Quantum Chemical methods is then submitted to the kinetic Monte Carlo (KMC) algorithm to predict the motion of charges throughout the system - quantified by the carrier mobility. In these simulations, charge carriers are considered as semi-classical particles that are completely delocalized over a single chromophore, i, and can hop a distance r_{ij} to a destination chromophore, j, at a rate based on Marcus theory[60]:

$$k_{ij} = \frac{|T_{ij}|^2}{\hbar}\sqrt{\frac{\pi}{\lambda_{ij} k_B T_{\text{KMC}}}}\exp\left(\frac{r_{ij}}{\alpha}\right)\exp\left[-\frac{(\Delta E_{ij} + \lambda_{ij})^2}{4\lambda_{ij} k_B T_{\text{KMC}}}\right], \qquad (4)$$

where k_B is Boltzmann's constant, T_{KMC} is the KMC simulation temperature, and λ_{ij} is the reorganization energy, which is the energy required to completely polarize and depolarize a chromophore in response to a charge carrier hopping from one site to another. For P3HT with a chromophore length of a single monomer, $\lambda_{ij} = 306$ meV, which matches that predicted by DFT and the Su-Schrieffer-Heeger model[61]. In all KMC simulations, $T_{\text{KMC}} = 290K$ to simulate a device at approximately room temperature. This temperature is not to be confused with the MD simulation temperature, which was used to generate the different state-points in this work. During the MD simulations, T corresponds to the final temperature that the system of chains was cooled to

(from the polymer melt phase) and equilibrated at in order to stimulate crystallization[34]. This process corresponds to the device manufacturing stage, where thermal and solvent annealing is used to 'lock in' a particular conformation and structure. During the KMC simulations, the solvent in the thin film is assumed to have completely evaporated, leaving behind a morphological structure that is no longer evolving. We are therefore justified in 'freezing' the motion of molecules in the system, effectively considering the charge transport of a fully encapsulated device and ignoring any degradation effects.

Note that eq. 4 includes an additional exponential term $\exp\left(\frac{r_{ij}}{\alpha}\right)$ not present in conventional Marcus theory. This term is based on Mott's variable-range hopping theory[62] (VRH) and is sometimes used in polymer hopping theory and computation[63,64]. Here, we include a VRH term with a single tunable parameter, $\alpha = 2$ Å (determined from preliminary work), to control the degree to which long-range hops are penalized in our simulation. This is important for systems with variable or large chromophore sizes, where long center-of-mass chromophore separations are not significantly offset by reduced transfer integrals, resulting in larger carrier distances traveled in the same amount of time, higher mobilities, and a preference for disordered morphological structures. In this work, the P3HT monomer sizes and deviation in r_{ij} are sufficiently small that the VRH term results in a reduced absolute value of mobility across all the considered systems, with no appreciable change in the qualitative trends considered in this investigation.

The simulations proceed by injecting a single charge carrier onto a random chromophore in the system and calculating the rates to all permitted neighbours (i.e. those with non-zero transfer integrals that are within 1 nm of the initial site). Simulation wait times for each permitted hop to occur, t_{ij}, are calculated using the KMC algorithm:

$$t_{ij} = -\frac{\ln(x)}{k_{ij}}, \tag{5}$$

where $0 \leq x < 1$ is a uniformly distributed random number, which adds variance to the hopping rate k_{ij}. The hopping event with the shortest wait time is selected as the most probable behavior for the carrier. The carrier is moved to the destination site, the total simulation time is incremented by t_{ij}, and the algorithm repeats for the permitted neighbours of the new chromophore. In these simulations, only a single carrier is run at any given time, so no Coulombic interactions need to be considered. The charge density in the morphology is therefore low, and the simulation more closely approximates time-of-flight mobility experiments than field-effect transistor measurements. Additionally, periodic boundaries are used, allowing charges to wrap back into the simulation volume instead of leaving the morphology. The entire system available for carriers to move in is therefore a periodic supercell, in which the $(10\text{ nm})^3$ structure is repeated for several hundred nanometers in each direction. Finally, no external electric field is applied across the simulated volume, resulting in zero-field charge transport measurements. Using this periodic supercell, MorphCT therefore calculates the device-scale carrier mobilities over hundreds of nm, while maintaining strict molecular resolution and explicitly calculating in-cell molecular conformations and energetic disorder[19,35]. This differs significantly from previous kinetic Monte Carlo codebases that operate on the mesoscale, where transport properties tend to be averaged and molecular orientations are abstracted out in favor of exploring device domain composition for systems generated using Cahn-Hilliard[29–31] or Ising-model[23,65] techniques. Each carrier is permitted to hop in this way until a specific simulation time, t, is reached, at which point its displacement from its origin is calculated. The mean squared displacements, $< x^2 >$, of 10,000 carriers are obtained for 15 different values of t logarithmically spaced between 1 ps and 1 ns. For purely diffusive transport there is a linear dependence of $< x^2 >$ on t, the gradient of which is the three-dimensional diffusion coefficient, D. The zero-field carrier mobility can therefore be determined as:

$$\mu_0 = \frac{qD}{6k_BT}, \tag{6}$$

where q is the elementary charge.

4 Algorithmic and performance details

We first provide technical details of the analysis. The key element required to translate the CGMD results into a graph is the construction of the neighborhood. We leverage Voronoi diagrams to determine the neighborhood of each bead. In particular, we use the voro++[36] package, due to its computational efficiency and flexibility. Once the graph is constructed, the graph quantification is implemented in C++ using Boost library[66].

The mobility calculation pipeline developed here leverages several freely available software packages, with runtimes varying over a few orders of magnitude. The initial MD simulations require ∼100 wall-clock hours on a single GPU to generate CGMD morphologies at one thermodynamic state point and relax the fine-grained atomistic representations from these morphologies using HOOMD-Blue[45]. Note that these simulations are a prerequisite for obtaining the inputs to both MorphCT and the graph-based calculations. In the MorphCT pipeline[19,35], the ZINDO/S calculations for every chromophore and neighbor pair within each morphology are parallelized over 14 CPU cores, and require ∼10 wall-clock hours using ORCA[52]. The KMC simulations require ∼12 hours to obtain sufficiently small statistical errors for carrier trajectories and calculate mobilities for one morphology, when parallelized over 14 CPU cores. Conversely, the graph-based morphology quantifications developed here take around 3 minutes on a desktop for morphologies with 500,000 atoms. The disparity in computation time between the graph-based morphology quantifications and QCC/KMC demonstrates the performance benefits such methods represent if they can provide similar insight into structure-property relationships.

5 Results

In this section we describe graph-based measurements of ordered and disordered P3HT morphologies with a focus on structural and energetic disorder. We seek to identify structural features that can explain the two order of magnitude difference in mobility cal-

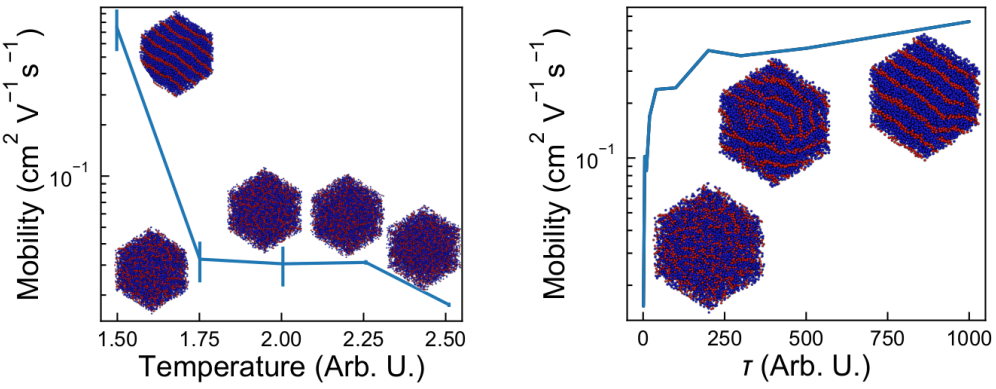

Fig. 5 The predicted zero-field hole mobility for each investigated P3HT morphology as functions of a) temperature, and b) structural evolution time for the $T = 1.5$ system. The corresponding coarse-grained morphologies are shown for select state points (a: $T = 1.5$, 1.75, 2.0, 2.25, 2.5, b: $\tau = 0$, 200, 1000). Error bars on the mobility are calculated based on 5 structurally decorrelated replicate simulations performed at each temperature.

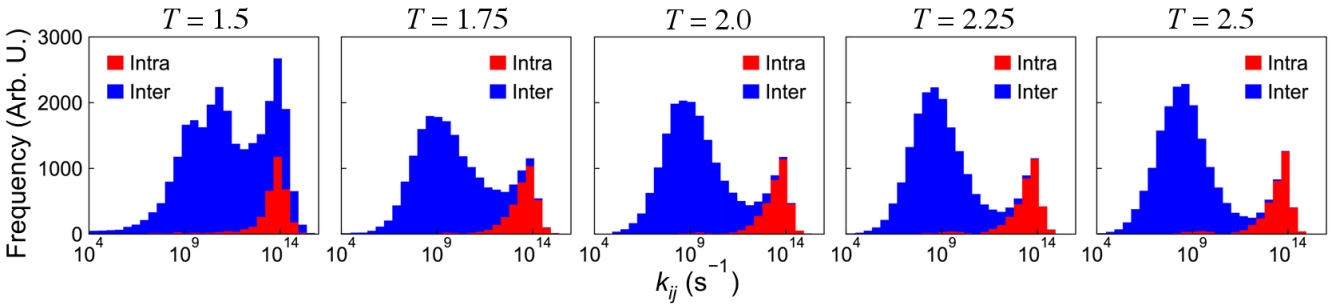

Fig. 6 The corresponding distributions of Marcus hopping rates between chromophores as a function of temperature. The histogram bars are stacked (blue frequencies are not "hiding" behind red), with red bar corresponding to the rates of intra-molecular carrier hops and blue bars describing inter-molecular hops.

culated from by KMC, which depends only on how the polymer chains are arranged (Figure 5). Towards this aim we answer three questions related to the charge transport: Is the structure interconnected? What are the characteristic features of the comprising paths? Which descriptor traces the mobility quantitatively?

5.1 Is the structure interconnected?

An interconnected network, one in which there is a finite path connecting each vertex to every other vertex, is associated with favorable charge transport. Interconnectedness is not equivalent to structural order, however. For example, it is possible for highly ordered lamellae of P3HT to be disconnected due to the insulating effect of alkyl side chains separating the backbones along which charges prefer to migrate. On the other hand, a disordered melt of P3HT chains may be highly connected, though unfavorable orientations between neighboring chromophores may hinder transport between otherwise "connected" vertices. Consequently, measuring the connectivity of the semiconductor network should provide better, or at least additional, insight into its charge transport properties than the a structural order parameter alone.

Quantifying connectivity is straightforward with graph-based representations of morphology. In particular, we query the graph to determine the set of connected component both in terms of its cardinality as well as the size of individual components. As detailed in section 2, a connected component is a subgraph where a path between any pair of vertices in this subgraph exists. This formal definition maps well to the practical question we pose in the context of the charge transport. The goal is to check how many regions within the morphology exist. In the ideal case, one connected component exists that is well connected to boundary associated with the target electrode (here green meta vertex). If one connected component exists, it effectively means that any charge has means to reach any other bead/vertex in the morphology, given enough time. On the other hand, if multiple connected components within morphology exist, particularly those disconnected from the boundaries, then the probability of trapping charges (and therefore recombination with an opposing charge) increases.

We calculate three descriptors related to interconnectedness. First, we look into the total number of connected components, $\mathscr{N}_{CC}$. Each connected component consists of one type of vertex, i.e. backbone beads (BB) or side chain beads (SCh). Next, we look at the connected components that have any connection to the boundary. This means that we exclude the islands that are isolated from the boundaries. We call these components useful as they can contribute to the charge transport, as opposite to islands. Specifically, we consider a domain as useful if it is connected to the top boundary of the morphology, $\mathscr{N}_{CC}^{useful}$. The selection of the top or bottom boundary for this measure is equivalent and arbitrary, and we justify the selection of a single boundary to represent the electrode of relevance for one carrier type (here, holes) for systems where the size of the morphology is representative of devices, which is the case here. Finally, we calculate the fraction of useful domain, f_{CC}^{useful}, as the fraction of beads (equivalently, vertices) that are connected to the top boundary of the sample.

In Table 1, we present three descriptors detailed above for five morphologies, each equilibrated at a different temperature. Interestingly, all analyzed morphologies show a high degree of connectivity. Regardless of the annealing temperature, the connectivity is almost ideal. Although the number of connected components varies (from 1 to 5), the overall fraction of connected domains is almost 100%. This observation means that connectivity metrics cannot be used in isolation to explain the differences in charge mobility for these morphologies.

Similar observation were made for time series sequence of $T = 1.5$ morphologies. Similarly, the fraction of connected domains is nearly 100% irrespective of structural order. Even in the early stages of the annealing, when carrier mobility is low, the connectivity is already high. We measured connectivity for all three possible choices of boundaries (along each axis) and verified that the connectivity results were independent of which boundaries were chosen. In summary, all of the morphologies we analyze here are nearly identical in terms of their complete connectivity.

5.2 What are the characteristic features of paths?

In the next level of morphology quantification, we focus on various features of the shortest pathways. As explained in section 2, for each red vertex in the graph we compute the shortest pathway towards green meta vertex using the Dijkstra algorithm (Euclidean distance as edge weight). As an outcome, for each backbone bead, the shortest pathway is identified. Given all shortest pathways (3750 for each morphology), we perform a three-level pathway quantification for a series of morphologies. The results for 10 morphologies (five equilibration temperatures, and five increasingly ordered snapshots from $T = 1.5$) are included in Figures 7, 8 and 9. In particular, the histogram of the pathway length is plotted in Figure 7. The pathway tortuosity is depicted in Figure 8. Finally, the ratio of intermolecular hops to all hopes along each pathway is shown in Figure 9.

As shown in Figure 7, histograms of pathway lengths are quite similar across the ten morphologies studied here, with some variation in the fraction of paths in the $9-14$nm range. The more disordered a morphology is, the longer the path charges must take to make it to a boundary. On average, it takes 6.97nm for sample $T = 2.5$ to reach the boundary of the domain, while it take 4.35nm for sample $T = 1.5$. In the increasingly ordered morphologies on the second row of Figure 7, we observe also a diminishing fraction of paths in the $9-14$nm range as ordering increases. Here, the average path length decreases from 5.85 to 4.35nm, from first to last snapshot, respectively. These observations suggest that relatively long pathways are correlated with lower charge mobility, but in isolation are insufficient to explain the large differences in mobility calculated with KMC.

To help broaden our understanding of path characteristics, we next turn to path tortuosity. With the definition introduced in section 2, straight pathways have tortuosity close to unity. The higher the tortuosity is, the more curvy the pathway. Histograms of tortuosity for all 10 samples are depicted in Figure 8. Clearly, the highly ordered $T = 1.5$ sample has a distribution of tortuosi-

Table 1 First order descriptors capturing the connectivity of the network for five morphologies: f_{CC}^{useful} is the bead fraction connected to top boundary of the sample, $\mathscr{N}_{CC}^{useful}$ is the number of connected components directly connected to top surface of the samples, $\mathscr{N}_{CC}$ is the total number of connected components, $\mathscr{E}_{inter}$ is the total number of inter-chain edges connecting backbone beads from directly adjacent polymer chains, $\mathscr{E}_{intra}$ is the total number of intra-chain edges connecting backbone beads, BB denotes backbone beads, and SCh denotes side chains. Physics-based descriptors tortuosity and S_2 are included for reference, as well as the the hole mobility calculated from KMC simulations, μ_0 (given in cm^2/Vs). Standard deviations of the last digits are given in parentheses, except for $\mathscr{N}_{CC}$. *E.g.*, 1.28(1) means 1.28 ± 0.01 and 6793(111) means 6793 ± 111.

	T=1.5	1.75	2.0	2.25	2.5
f_{CC}^{useful}	1.00(0)	1.00(0)	1.00(0)	1.00(0)	1.00(0)
$\mathscr{N}_{CC}$ [BB+SCh]	5 [4+1]	5 [4+1]	3 [2+1]	2 [1+1]	2 [1+1]
$\mathscr{E}_{inter}$	10,946(37)	6,793(111)	5,968(98)	5,436(66)	5,059(50)
$\mathscr{E}_{intra}$	3,342(8)	3,401(10)	3,408(8)	3,417(6)	3,428(10)
Tortuosity	1.10(0)	1.28(1)	1.29(1)	1.30 (1)	1.33(2)
S_2	0.933(9)	0.37(1)	0.410(8)	0.458(6)	0.511(7)
μ_0	0.8(2)	0.032(8)	0.030(8)	0.0306(9)	0.017(5)

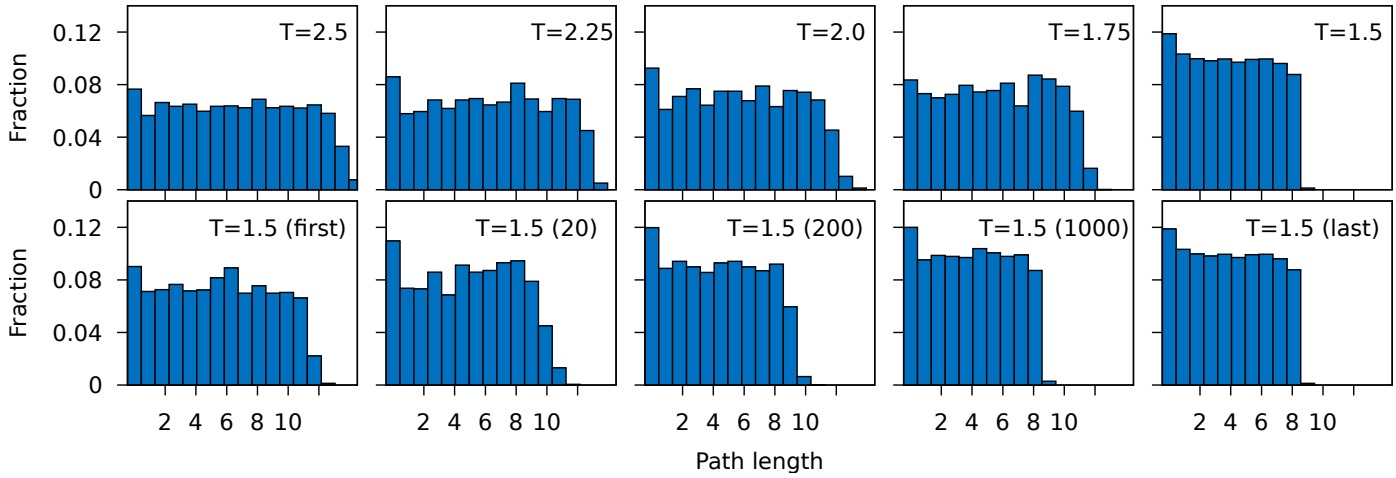

Fig. 7 Shortest pathway length distributions for the 10 morphologies analyzed. The top row corresponds to annealing temperature ($T = 2.5$ to 1.5), the bottom row corresponds to the increasingly ordered morphologies during the time evolution of the $T = 1.5$ simulation. Highly ordered morphologies and those with the highest charge mobilities have few paths over 8nm in length.

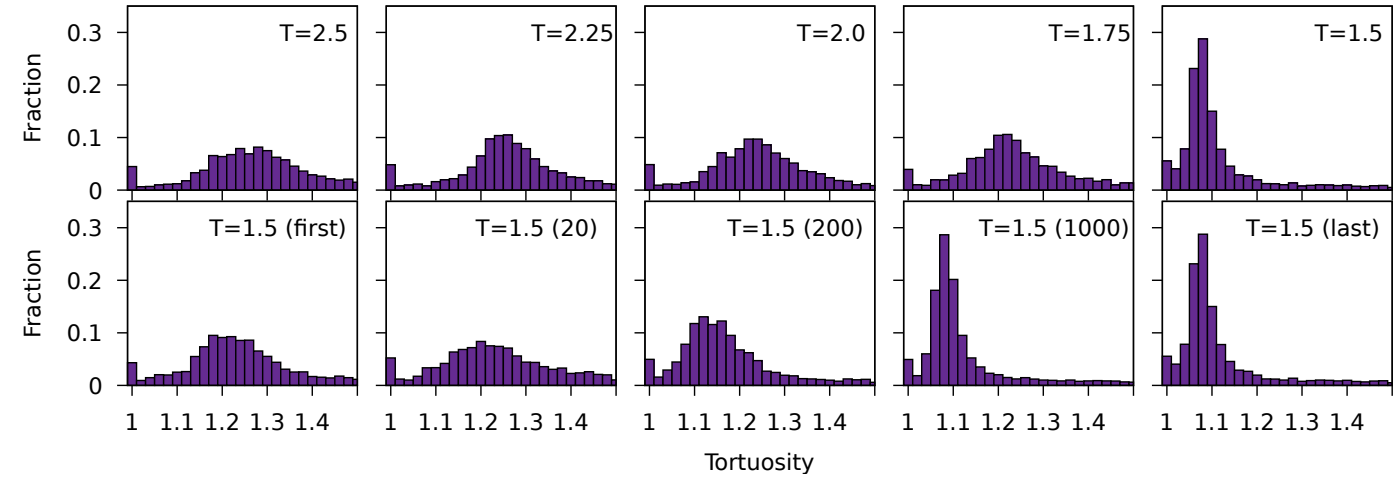

Fig. 8 Tortuosity distributions for the 10 morphologies analyzed: The top row corresponds to annealing temperature ($T = 2.5$ to 1.5), the bottom row corresponds to the temporal evolution for $T = 1.5$. The most ordered, highest-mobility morphologies have a significant fraction of low-tortuosity pathways.

ties that differs significantly from the disordered cases. All other samples have similarly tortuous pathways, with mean tortuosity of 1.30. The bottom row of Figure 8 depicts the temporal evolution of tortuosity for sample $T = 1.5$ from the first time step until the morphology is considered to be equilibrated. Initially the tortuosity is relatively high with many curvy pathways. Over time, the pathways become more straight and the distribution shifts left towards the lower tortuosities. We note that low mobility morphologies are characterized by higher tortuosity (see Table 1). High mobility morphologies are characterized with tortuosity closer to one (straight paths), and smaller variance. That the morphologies with more straight pathways have the highest charge mobilities suggest that these pathways have a significant impact on charge transport. Comparing to the histograms of path length (Figure 7), here we notice our first clear distinction in our descriptors between low and high mobility samples. These two features provide a qualitative means to explain and potentially classify well and poorly performing devices.

Due to the correlation between short, straight pathways and high mobility, we further analyze the fraction of inter-molecular hops along these pathways, which are straightforward to query from the graph representation. That is, "Are these fast charge transport pathways mostly along single chains, or made from multiple chains?" Figure 9 shows the fraction of inter-molecular hops along each path. When inspecting the histogram for all 5 temperatures, we notice much clearer distinction between high mobility and lower mobility morphologies through the lens of inter-molecular hops. In particular, a full 58% of the pathways in the high mobility ($T = 1.5$) sample have no hops along a single chain! This finding is surprising because of the expectation that charge transport is facilitated along the conjugated backbone of a polymer chain relative to hopping between chains. We observe that the more disordered (higher temperature) morphologies are comprised of pathways ranging from purely intra-molecular hops to purely inter-molecular hops with a mean around 25% of inter-molecular hops. In summary, these results suggest that the pathways that contribute the most to high charge mobilities are short (low-tortuosity) pathways, and that these short pathways have a high proportion of hops *between* chains.

To further probe the link between path tortuosity and mobility, we investigate and compare our graph metrics with the distributions of Marcus hopping rates calculated with quantum chemical methods (Figure 6). The number of both inter- and intra-molecular hops with a non-zero hopping rate is much higher for $T = 1.5$ than any other sample presented in Figure 6. This insight from QCC in terms of individual hopping rates is also mirrored in the total number of inter-molecular edges in the graph. In Table 1, we list the total number of edges for the five samples with varying temperature, counting only pairs of neighbors between backbone beads. The total number of intra-molecular edges $\mathscr{E}_{intra}$ does not vary much between samples because it depends on the (constant) total number of bonds between backbone beads, and how many of these bonds are severed due to the selection of the boundary vertices. The total number of inter-molecular edges $\mathscr{E}_{inter}$ differs significantly between samples. In particular, sample $T = 1.5$ shows approximately double the quantity of intermolecular edges, despite having the same number of chromophores. Consequently, this descriptor can also be used to quickly screen high and low mobility devices.

In summary, we identified three purely graph-based descriptors that can be used for quick screening of the morphologies. These key descriptors include: tortuosity, intermolecular hop ratio, and the total number of intermolecular edges in the graph. All these descriptors can be quickly calculated using a graph-based morphology representation and as evidenced by the present work on P3HT, offer unique screening capabilities. Using these three purely structural descriptors, our quantification method revealed that the $T = 1.5$ sample is the least torturous with pathways consisting mostly of inter-molecular hops. All together these structural features distinguish this high mobility morphology from the disordered, low-mobility samples. However, these qualitative links between structure and mobility do not yet explain the two order of magnitude difference in mobility measured with KMC. In the next subsection we detail the results for graph-based analysis that accounts for both structural and energy disorder.

5.3 Which descriptor traces the mobility quantitatively?

Two notable takeaways from Table 1 are: (1) that S_2 is inversely correlated with mobility for the disordered ($T > 1.5$) morphologies, and (2) that tortuosity tracks mobility. Upon closer inspection we find that a linear fit of $\log(tortuosity)$ vs. mobility has an R-squared value of 0.987, which is consistent with the hypothesis that activated processes on the shortest networks determine mobility. The observation that S_2 is a poor predictor of mobility is somewhat surprising from the qualitative intuition that more ordered structures should have lower hopping barriers and higher mobilities. The poor correlation of S_2 underscores the utility of physics-based descriptors like tortuosity that include detailed network structure.

Figure 10 depicts overlaid histograms of the total travel time along the fastest pathway. The total travel time defined in Equation 1 was used as the morphology descriptor (hopping rate as edge weight). This is in contrast to the results presented in the previous subsection, where we used Euclidean distance as edge weight to find the shortest pathways. Since the total travel time relies on the local hopping rate calculations and the overall structural arrangement of the polymer chain, we consider this descriptor as one coupling energetic and structural features of the morphologies.

The histogram of fastest pathways reveal that indeed high mobility samples ($T = 1.5$) consists of significantly faster pathways. As shown in Figure 10, the average travel time for this sample is two orders of magnitude shorter than the average time for the $T = 2.5$ sample. We see similar trends when we look at the temporal evolution of sample $T = 1.5$ (Figure 10). In comparing both the temperature-series data and time-series data we find strong correlation between the mean travel time measured through the shortest paths and the mobility calculated from first-principles methods (SI Section S4). This is an interesting result, given that the calculation through Dijkstra's algorithm considers the entirety of the path to obtain the shortest route, whereas the

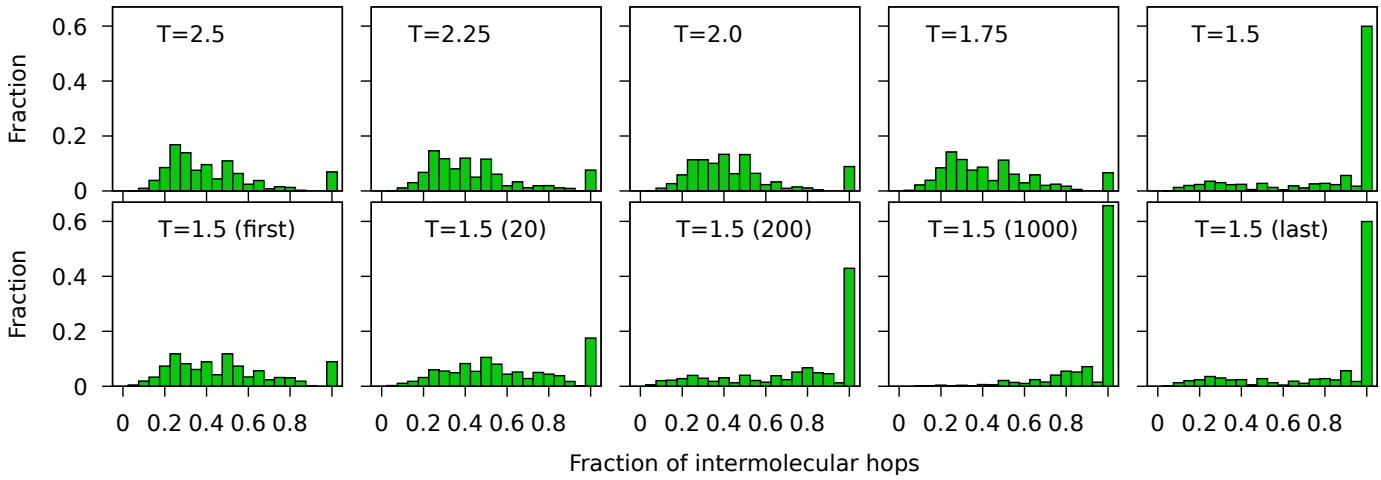

Fig. 9 Fraction of inter-molecular hop distributions along the shortest pathways for the 10 morphologies analyzed: The top row corresponds to annealing temperature ($T = 2.5$ to 1.5), the bottom row corresponds to temporal evolution for $T = 1.5$. The most ordered, highest-mobility morphologies have a high fraction of paths that are only between, never along, the polymer backbones.

KMC only identifies a preferential route for the carrier on a hop-by-hop basis by considering only the immediate neighbors to the charged chromophore. Given that in the KMC we average the carrier motion over 10,000 carriers that have started at randomly selected chromophores in the system, this correlation indicates that all routes eventually guide the carriers onto the fastest routes through the crystal. We therefore suggest that, within an ordered crystal like our $T = 1.5$ morphology, the hopping landscape is relatively 'flat', in that long, circuitous transport routes contribute negligibly to the overall carrier mobility. This contrasts with transport behavior over larger scales than explored here, where long, low-resistance pathways in the form of "tie chains" between multiple ordered domains are widely believed to be critical to efficient charge transport through the amorphous phase[4,5,41,67]. It is important to understand the intra- and inter-chain composition of pathways for all significant phases within the morphology, however, charge transport within crystals is often ignored in both experimental and theoretical investigations due to the comparatively high resultant mobility. In this work, we therefore present a computational tool that can quantitatively investigate the important transport features. Due to computational constraints we showcase it for short polymer chains. We hope to extend our methodology to systems consisting of longer chains with larger molecular weights and provide further insight into these mechanisms.

In summary, the total travel time as a descriptor traces the mobility prediction most closely among all descriptors we defined so far. It is able to capture the differences between morphologies both qualitatively and quantitatively. The exact matching between graph-based descriptors and KMC prediction of morphology is still to be found. However, at this stage, the approach can be used for quick screening of morphologies. It should be kept in mind that the total travel time relies on the local hopping rate from QCC that can be computationally expensive. The next step would be to establish the relationship between the structural features of the local arrangement of chromophores and hopping rate such that this step can be avoided.

6 Conclusion

In this work, we present a multi-scale framework to extract salient features of morphologies governing hole mobility. The key idea behind the framework is to represent morphology as a weighted, labeled, undirected graph. This representation allows us to include information from electronic structure calculations in measurements of hierarchical structural features from CGMD morphologies by leveraging well-studied algorithms to query the graph. We have calculated the hole mobility for a series of morphologies containing the donor polymer P3HT, simulated at different state points, and discovered that, out of the descriptors studied, the "fastest pathways" descriptor traces the hole mobility the best, with good quantitative agreement. However, these calculations depend on computationally expensive QCC. A viable alternative that circumvents this computational cost is the log(tortuosity), which also correlates well with mobility. Therefore, tortuosity calculations performed directly on data from simulations, computed tomography, or other structural data can be used for screening purposes at significantly reduced computational cost.

While the present work does not definitively answer which molecular design rules will optimize OPV performance, we derive the following insights from the present analysis. First, the observed significance of shortest paths informs the design choice of selecting chemistries that are thermodynamically favored to pi-stack with the bounding interfaces of the device: This would facilitate the assembly of active layer clusters oriented with short pathways optimized for transport. Second, to maximize the self-assembly in a chemistry like P3HT, the film should be processed at the highest temperature possible below the order-disorder temperature. Third, we learn that is not necessarily the case that chemistries that robustly self-assembly ordered structures are optimal for OPVs, but rather we should identify chemistries that robustly self-assemble interconnected, short pathways. This sug-

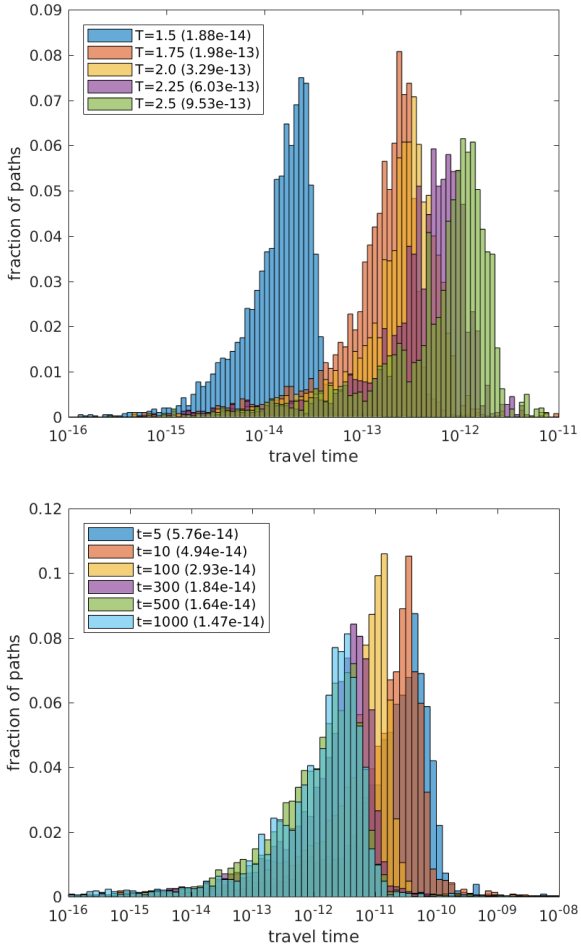

Fig. 10 Histogram of shortest path travel times for five equilibration temperatures (top), and six snapshots from the T=1.5 equilibration. Note the two orders of magnitude difference in average travel time between T=1.5 and T=2.5. The travel times of charges along the shortest paths from a vertex to the nearest boundary vertex correlate well with the mobility calculation from kinetic Monte Carlo calculations (Table 1).

gests that smaller planar molecules that are thermodynamically favored to assemble pi-stacks, but have shorter relaxation times than polymers, may offer some experimental avenues to high mobility if processed properly. Fourth, by being able to independently assess network connectivity and structure, we may begin thinking about correlating chemical features directly with connectivity without the proxy of morphological order, which would provide more specific experimental guidance than "Choose chemistries that assemble well". Finally, this work helps to design more efficient computational studies of OPV materials: The observation of diminishing mobility returns as a function of time in Fig. 5 informs more efficient screening strategies that relax the requirement that each MD trajectory be run to equilibration. That is, if a high-throughput screening study is searching a set of chemistries for those that assemble high mobility morphologies, the present work can be used to justify running simulations until the mobilities plateau, which may save significant time compared to running until the morphologies equilibrate.

Conflicts of interest

There are no conflicts to declare.

Acknowledgments

This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053575[68]. This material is based upon work supported by the National Science Foundation under Grant No. 1653954.

Notes and references

1 C. J. Brabec, N. S. Sariciftci, J. C. Hummelen *et al.*, *Advanced functional materials*, 2001, **11**, 15–26.
2 Y.-S. Tyan, *Journal of Photonics for Energy*, 2011, **1**, 011009–011009.
3 C. D. Dimitrakopoulos and D. J. Mascaro, *IBM Journal of research and development*, 2001, **45**, 11–27.
4 R. Noriega, J. Rivnay, K. Vandewal, F. P. Koch, N. Stingelin, P. Smith, M. F. Toney and A. Salleo, *Nature materials*, 2013, **12**, 1038–1044.
5 S. Schott, E. Gann, L. Thomsen, S. H. Jung, J. K. Lee, C. R. McNeill and H. Sirringhaus, *Advanced Materials*, 2015, **27**, 7356–7364.
6 J.-L. Brédas, D. Beljonne, V. Coropceanu and J. Cornil, *Chemical Reviews*, 2004, **104**, 4971–5004.
7 V. Coropceanu, J. Cornil, D. A. da Silva Filho, Y. Olivier, R. Silbey and J.-L. Brédas, *Chemical Reviews*, 2007, **107**, 926–952.
8 J. Kirkpatrick, V. Marcon, J. Nelson, K. Kremer and D. Andrienko, *Physical Review Letters*, 2007, **98**, 227402.
9 A. Troisi, *Advanced Materials*, 2007, **19**, 2000–2004.
10 Y.-k. Lan and C.-i. Huang, *The Journal of Physical Chemistry B*, 2008, **112**, 14857–14862.
11 Y.-K. Lan and C.-I. Huang, *The Journal of Physical Chemistry B*, 2009, **113**, 14555–14564.
12 T. H. Cormen, *Introduction to algorithms*, MIT press, 2009.

13 C. Du, Y. Ji, J. Xue, T. Hou, J. Tang, S.-T. Lee and Y. Li, *Scientific Reports*, 2015, **5**, 16854.

14 O. Wodo, S. Tirthapura, S. Chaudhary and B. Ganapathysubramanian, *Organic Electronics*, 2012, **13**, 1105–1113.

15 C.-K. Lee, O. Wodo, B. Ganapathysubramanian and C.-W. Pao, *ACS applied materials & interfaces*, 2014, **6**, 20612–20624.

16 B. G. Sumpter and D. W. Noid, *Macromolecular theory and simulations*, 1994, **3**, 363–378.

17 V. A. Blatov, A. P. Shevchenko and D. M. Proserpio, *Crystal Growth & Design*, 2014, **14**, 3576–3586.

18 M. Witman, S. Ling, P. Boyd, S. Barthel, M. Haranczyk, B. Slater and B. Smit, *ACS Central Science*, 2018, **4**, 235–245.

19 M. L. Jones and E. Jankowski, *Molecular Simulation*, 2017, **43**, 756–773.

20 O. Wodo, J. Roehling, A. Moule and B. Ganapathsubramanian, *Energy and Environmental Science*, 2013, **6**, 3060–3070.

21 O. Wodo and B. Ganapathysubramanian, *Computational Materials Science*, 2012, **55**, 113–126.

22 J. J. Michels and E. Moons, *Macromolecules*, 2013, **46**, 8693–8701.

23 C. Groves, R. G. E. Kimber and A. B. Walker, *The Journal of Chemical Physics*, 2010, **133**, 144110.

24 R. G. E. Kimber, A. B. Walker, G. E. Schröder-Turk and D. J. Cleaver, *Phys. Chem. Chem. Phys.*, 2010, **12**, 844–851.

25 C. Groves, *Energy & Environmental Science*, 2013, **6**, 1546.

26 M. L. Jones, B. Chakrabarti and C. Groves, *The Journal of Physical Chemistry C*, 2014, **118**, 85–91.

27 J.-L. Brédas, J. P. Calbert, D. da Silva Filho and J. Cornil, *Proceedings of the National Academy of Sciences*, 2002, **99**, 5804–5809.

28 W.-Q. Deng and W. A. Goddard, *The Journal of Physical Chemistry B*, 2004, **108**, 8614–8621.

29 B. P. Lyons, N. Clarke and C. Groves, *The Journal of Physical Chemistry C*, 2011, **115**, 22572–22577.

30 B. P. Lyons, N. Clarke and C. Groves, *Energy & Environmental Science*, 2012, **5**, 7657.

31 M. L. Jones, R. Dyer, N. Clarke and C. Groves, *Phys. Chem. Chem. Phys.*, 2014, **16**, 20310–20320.

32 O. Wodo, J. D. Roehling, A. J. Moulé and B. Ganapathysubramanian, *Energy & Environmental Science*, 2013, **6**, 3060–3070.

33 E. Jankowski, H. S. Marsh and A. Jayaraman, *Macromolecules*, 2013, **46**, 5775–5785.

34 H. S. Marsh, E. Jankowski and A. Jayaraman, *Macromolecules*, 2014, **47**, 2736–2747.

35 M. L. Jones, *MorphCT v2.2, doi:10.5281/zenodo.1228321*, 2017.

36 C. Rycroft, *Voro++: A three-dimensional Voronoi cell library in C++, permalink: https://escholarship.org/uc/item/8sf4t5x8*, 2009.

37 E. W. Dijkstra, *Numerische mathematik*, 1959, **1**, 269–271.

38 P. Carbone and A. Troisi, *The Journal of Physical Chemistry Letters*, 2014, **5**, 2637–2641.

39 M. R. Wilson, *J. Mol. Liq.*, 1996, **68**, 23–31.

40 O. Wodo, S. Tirthapura, S. Chaudhary and B. Ganapathysubramanian, *Journal of Applied Physics*, 2012, **12**, 06431.

41 M. L. Jones, D. M. Huang, B. Chakrabarti and C. Groves, *Journal of Physical Chemistry C*, 2016, **120**, 4240–4250.

42 D. M. Huang, R. Faller, K. Do and A. J. MouleÌĄ, *Journal of Chemical Theory and Computation*, 2010, **6**, 526–537.

43 D. M. Huang, A. J. Moule and R. Faller, *Fluid Phase Equilibria*, 2011, **302**, 21–25.

44 K. N. Schwarz, T. W. Kee and D. M. Huang, *Nanoscale*, 2013, **5**, 2017–2027.

45 J. A. Anderson, C. D. Lorenz and A. Travesset, *Journal of Computational Physics*, 2008, **227**, 5342–5359.

46 B. Hess, S. León, N. van der Vegt and K. Kremer, *Soft Matter*, 2006, **2**, 409–414.

47 C. Peter and K. Kremer, *Soft Matter*, 2009, **5**, 4357.

48 O. Bezkorovaynaya, A. Lukyanov, K. Kremer and C. Peter, *Journal of Computational Chemistry*, 2012, **33**, 937–949.

49 D. P. McMahon, D. L. Cheung, L. Goris, J. DacunÌČa, A. Salleo and A. Troisi, *The Journal of Physical Chemistry C*, 2011, **115**, 19386–19393.

50 D. H. K. Murthy, M. Gao, M. J. W. Vermeulen, L. D. A. Siebbeles and T. J. Savenije, *The Journal of Physical Chemistry C*, 2012, **116**, 9214–9220.

51 M. Pope and C. E. Swenberg, *Electronic Processes in Organic Crystals and Polymers*, Oxford University Press, New York, 2nd edn, 1999.

52 F. Neese, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2012, **2**, 73–78.

53 M. E. Köse, *Journal of Physical Chemistry C*, 2011, **115**, 13076–13082.

54 J. Kirkpatrick, *International Journal of Quantum Chemistry*, 2008, **108**, 51–56.

55 C. Risko, G. P. Kushto, Z. H. Kafati and J. L. Brédas, *Journal of Chemical Physics*, 2004, **121**, 9031–9038.

56 L. Yang, J. K. Feng, Y. Liao and A. M. Ren, *Polymer*, 2005, **46**, 9955–9964.

57 M. Rubio, M. Merchán, E. Ortí and B. O. Roos, *The Journal of Chemical Physics*, 1995, **102**, 3580–3586.

58 M. Rubio, M. Merchán, E. Ortí and B. O. Roos, *Chemical Physics Letters*, 1996, **248**, 321–328.

59 V. Ásgeirsson, C. A. Bauer and S. Grimme, *Chemical Science*, 2017, **8**, 4879–4895.

60 R. A. Marcus, *Annual Review of Physical Chemistry*, 1964, **15**, 155–196.

61 E. Johansson and S. Larsson, *Synthetic Metals*, 2004, **144**, 183–191.

62 N. Mott, *Journal of Non-Crystalline Solids*, 1968, **1**, 1–17.

63 D. N. Tsigankov and A. L. Efros, *Physical Review Letters*, 2002, **88**, 176602.

64 R. P. Fornari and A. Troisi, *Physical Chemistry Chemical Physics*, 2014, **16**, 9997.

65 P. K. Watkins, A. B. Walker and G. L. B. Verschoor, *Nano Letters*, 2005, **5**, 1814–1818.

66 J. G. Siek, L.-Q. Lee and A. Lumsdaine, *The Boost Graph Library: User Guide and Reference Manual, Portable Documents*, Pearson Education, 2001.

67 M. Brinkmann and P. Rannou, *Macromolecules*, 2009, 1125–1130.

68 J. Towns, T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. D. Peterson, R. Roskies, J. R. Scott and N. Wilkens-Diehr, *Comput. Sci. Eng.*, 2014, **16**, 62–74.