PAPER

Check for updates

Competing single-chain folding and multi-chain aggregation pathways control solution-phase aggregate morphology of organic semiconducting polymers†

Belinda J. Boehm, [ID] [a] Christopher R. McNeill [ID] [b] and David M. Huang [ID] *[a]

Understanding the solution-phase behaviour of organic semiconducting polymers is important for systematically improving the performance of devices based on solution-processed thin films of these molecules. Conventional polymer theory predicts that polymer conformations become more compact as solvent quality decreases, but recent experiments have shown the high-performance organic-semiconducting polymer P(NDI2OD-T2) to form extended rod-like aggregates much larger than a single chain in poor solvents, with the formation of these extended aggregates correlated with enhanced electron mobility in films deposited from these solutions. We explain the unexpected formation of extended aggregates using a novel coarse-grained simulation model of P(NDI2OD-T2) that we have developed to study the effect of solvent quality on its solution-phase behaviour. In poor solvents, we find that aggregation through only a few monomers gives effectively inseparable chains, leading to the formation of extended structures of partially overlapping chains *via* non-equilibrium assembly. This behaviour requires that multi-chain aggregation occurs faster than chain folding, which we show is the case for the chain lengths and concentrations shown experimentally to form rod-like aggregates. This kinetically controlled process introduces a dependence of aggregate structure on concentration, chain length, and chain flexibility, which we show is able to reconcile experimental findings and is generalisable to the solution-phase assembly of other semiflexible polymers.

1 Introduction

Organic semiconductors (OSCs) have a number of advantages over conventional inorganic semiconductors for the fabrication of lightweight, flexible, and low-cost electronic devices. These advantages stem to a large extent from their ability to be processed from solution[1] using inexpensive printing methods.[2] However, the final thin-film microstructure is difficult to predict, particularly for semiconducting polymers, and has been found to depend on many factors, including the monomer chemical structure,[3–7] molecular weight,[8,9] solvent,[10–15] solution concentration,[16,17] and dissolution temperature[18] as well as non-equilibrium processes during or post deposition.[14,19] Device performance is closely tied to this microstructure.[7,10,20–27] Thus, the systematic design of mole-

cules and processing conditions to achieve good performance is challenging.

A better understanding of the factors that control the solution-phase morphology of semiconducting polymers can potentially help to systematically improve device performance, as this morphology has been correlated with thin-film structure, and with device performance.[9–11,18] For the high-performance semiconducting polymer poly[*N*,*N*′-bis(2-octyldodecyl) naphthalene-1,4,5,8-bis(dicarboximide)-2,6-diyl]-*alt*-5,5′-(2,2′-bithiophene) (P(NDI2OD-T2)), also known as N2200, the formation of large rod-like aggregates has been observed experimentally in poor solvents such as toluene and xylene, and has been shown to be associated with increased electron mobility in films deposited from these solutions.[10] In these poor solvents, P(NDI2OD-T2) (M_n = 31.2 kDa, polydispersity = 2.1) was shown, *via* UV-vis spectroscopy, to aggregate extensively.[10] However, counter to a conventional understanding of solution-phase aggregation of flexible polymers, which predicts the radius of gyration to decrease with decreasing solvent quality,[28] these aggregates were shown, using small-angle X-ray scattering (SAXS), to have a radius of gyration larger than that of a single chain, suggestive of the formation of extended

[a]*Department of Chemistry, School of Physical Sciences, The University of Adelaide, SA 5005, Australia. E-mail: david.huang@adelaide.edu.au; Tel: +61 8 8313 5580*
[b]*Department of Materials Science and Engineering, Monash University, Clayton, VIC 3800, Australia*
†Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d2nr04750k

multi-chain structures that are not predicted by existing theories.[10] This result contrasts with conclusions from a previous study of P(NDI2OD-T2) which, based on the lack of dependence on polymer concentration of the spectral shift in poor solvents and analytical centrifugation measurements, suggested that the aggregation behaviour in toluene is a single-chain process caused by chain collapse and folding.[12] Notably, these experiments were conducted at a much lower concentration than the SAXS measurements (up to 1 g L^{-1},[12] versus 5 g L^{-1} for the SAXS experiments[10]), as well as using significantly longer chains (118, 181, or 1105 kDa in ref. 12 versus 31 kDa in ref. 10), which may explain the reported discrepancy. Indeed, multi-chain aggregation has been observed for other semiconducting polymers such as MEH-PPV,[29] PffBT4T-2DT, D-DPP3T-EH, and PffBT4T-2OD,[18] with concentration- and molecular-weight-dependent effects observed but not fully explained.

The solution-phase conformations and dynamics of *flexible* polymers are generally well understood.[28] The conformation of a single flexible polymer chain in solution shows predictable scaling with chain length, with a scaling exponent determined by the solvent quality, *i.e.* the relative strength of the polymer–polymer, solvent–solvent, and polymer–solvent interactions. In a good solvent, in which polymer–solvent attractions dominate, the chain is in an extended conformation, whereas a more compact, collapsed structure is formed under poor solvent conditions in order to minimise unfavourable interactions with the solvent or to maximise favourable intramolecular polymer interactions. Metrics such as the radius of gyration are therefore expected to decrease as solvent quality decreases and chains become more compact. This behaviour has been observed in a number of simulation studies of sufficiently flexible single chains, with a transition from an extended coil to collapsed globule structure observed with decreasing solvent quality.[30–32]

However, semiconducting polymers typically have stiffer *semiflexible* backbones, which, coupled with the more anisotropic shape and interactions imparted by the conjugated backbone, means that the aggregate structure may deviate from that predicted for flexible chains (see ref. 33 for a more comprehensive review of the behaviour of semiflexbile polymers in dilute solutions). Indeed, for both single-[30–32] and multi-chain[34] systems, both the backbone stiffness and solvent quality have been shown to be important for predicting the types of structures formed by semiflexible polymers. For single chains (*i.e.* in very dilute solution), as chain stiffness increases, different structures, such as hairpins or toroids, take the place of disordered globules in poorer solvents, with the exact structure depending on both the solvent quality and chain flexibility.[30–32,35,36]

At higher concentrations where multiple chains are able to interact, bundles of fully overlapping chains, rather than collapsed globules, are expected for semiflexible polymers in a poor solvent due to the unfavourable bending energy. Indeed, Monte Carlo simulations have shown the equilibrium structure of multi-chain aggregates in a poor solvent to shift from a disordered globular morphology to fully overlapped twisted or folded rod-like bundles as the chain stiffness is increased.[34] However, although these fully overlapped bundles are expected to be the equilibrium structure under these conditions, due to maximising favourable polymer–polymer interactions while minimising unfavourable polymer–solvent interactions and bending energy, such structures would not lead to rod-like aggregates significantly longer than a single chain. Thus, the known equilibrium solution-phase behaviour of semiflexible polymers is not consistent with experimental observations on P(NDI2OD-T2).[10]

While a multitude of studies have examined single-chain behaviour using simulations[30–32,35–39] or theory,[40,41–44] those examining multi-chain systems, which are more relevant for the behaviour of realistic OSC systems in which chains are rarely so isolated, are less common. Although solution-phase molecular simulations of multi-chain systems of semiconducting polymers are relatively rare, owing to the need for often prohibitively large systems to explicitly account for solvent, especially for long polymer chains, studies examining OSC solubility using all-atom (AA) molecular simulation methods can be found. Some[45–47] have used mean-field solution theories such as the Flory–Huggins theory, in which simulations of short oligomers were used to estimate the Flory–Huggins parameter, which is used as a measure of solvent quality and thus the propensity for aggregation. Others[48,49] have examined the aggregation mechanism and effect of solvent and polymer properties, again using short chains or implicit solvent models. While these studies provide valuable insights into some of the many factors affecting the solution-phase morphology, the Flory–Huggins theory provides a fairly simplistic model of the effects of polymer chain length and the relative strength of the solvent–solvent, solvent–polymer, and polymer–polymer interactions on solubility; it does not capture the roles of chain stiffness and conformation and so cannot account for extended aggregates expected for P(NDI2OD-T2). Other simulation studies that have more accurately calculated solubility through free energy perturbation methods,[49] and examined aggregation mechanisms and the effect of various molecular properties on the solution-phase behaviour,[48–50] have not reached experimental chain lengths except in limited cases.[50] They have generally considered only equilibrium behaviour by using enhanced sampling methods that accelerate sampling of equilibrium polymer conformations but do not directly probe the evolution of the conformation with time, as the time scales of non-equilibrium processes such as polymer aggregation are generally not accessible to detailed AA models.[51]

In this work we have developed a systematically coarse-grained (CG) model of P(NDI2OD-T2) in order to investigate the reported formation of large extended aggregates in poor solvents,[10] to reconcile discrepancies between experimental findings on the solution-phase morphology of this polymer in such solvents,[10,12] and, more broadly, to clarify the general factors that control the solution-phase morphology of semiconducting polymers. By combining atoms with correlated

motion into a single CG site, and accounting implicitly for the solvent, the number of degrees of freedom of the system can be greatly reduced. This allows access to polymer length and time scales on the order of those studied experimentally. A similar model has previously been developed for the commonly studied semiconducting polymer P3HT[52] and used to accurately predict the experimental solution-phase conformation of this polymer, giving results consistent with a more computationally expensive AA model and experiment.

The CG model of P(NDI2OD-T2) was parameterised to reproduce the structural properties of an AA system. The methods used to parametrise and simulate the AA model are described in section 2.1, while those for the CG model are given in sections 2.2 and 2.3. The behaviour of the parameterised CG model of P(NDI2OD-T2) in simulations under conditions corresponding to varying solvent qualities is described in section 3. The types of aggregates formed under different solvent conditions are examined in section 3.1, and the aggregate structure is related to the strength of intermolecular interactions and the persistence of aggregates composed of partially overlapping chains in section 3.2. Finally, the kinetics of the competing effects of single-chain folding and multi-chain aggregation, and how these may vary with concentration, molecular weight, and chain flexibility, are considered in section 3.3.

2 Methods

All simulations were conducted using molecular dynamics (MD) with the LAMMPS software package,[53–56] and analysis and visualisation using OVITO[57] and VMD.[58] In all cases, the temperature was 300 K and, where relevant, the pressure was 1 atm.

2.1 All-atom simulations

2.1.1 Parameterisation of all-atom model. As many semiconducting polymers such as P(NDI2OD-T2) have a relatively rigid backbone and extended conjugation, their bonded parameters – particular between conjugated units – and charges are not expected to be accurately captured by general-purpose molecular-dynamics force fields.[59] To accurately model these systems, certain parameters must therefore be determined for the specific molecules of interest. Here we have based our parameterisation on the OPLS force field,[60–66] as it has been shown to accurately describe structural and thermodynamic properties of several small-molecule OSCs[67,68] and many organic liquids, which are commonly used as solvents for OSCs. We note that a previous AA model of P(NDI2OD-T2) has been parameterised with the AMBER force field,[45] but, to the best of our knowledge, no OPLS parameters exist for this polymer.

We have followed a parameterisation procedure previously used to obtain OPLS parameters for a wide variety of semiconducting polymers.[48] In all cases, van der Waals parameters were taken directly from the OPLS force field for equivalent atom types.[60–66] Atomic partial charges were obtained from quantum-chemical calculations, as described in the ESI in section S1,† with the side-chains truncated to methyl groups after the tertiary carbon (*i.e.* R–CH$_2$–CH–(CH$_3$)$_2$, where R is the monomer backbone). Note that although P(NDI2OD-T2) is typically represented as having a naphthalene diimide (NDI)–bithiophene (bTh) backbone, we separated the bTh group into two thiophene (Th) groups and modelled the monomer as Th–NDI–Th, as shown in Fig. 1, in order to increase its symmetry, allowing for a simpler and more general parameterisation. Within the polymer, the same structure will be obtained, with the only differences being in the structure of the terminal monomers (ESI Fig. S1† compares the two structures).

Most bonded parameters for bond lengths, angles, and dihedrals were taken directly from the OPLS force field for equivalent atom types,[60–66] while the equilibrium bond lengths and angles were obtained from the quantum-chemistry optimised geometry of the monomer. The exceptions were the bond stretching potential between NDI and Th groups and the NDI–Th and Th–Th dihedral potentials. These potentials were parameterised explicitly as they are important for modelling the semiflexibility of the backbone and were not expected to be accurately captured by existing OPLS parameters. They were calculated using a series of constrained quantum-chemical geometry optimisations as described in ref. 48 and ESI section S1.† All parameters used for the AA simulations in this work are tabulated in the ESI, section S13.†

2.1.2 All-atom solution-phase simulation. The non-bonded interactions for the CG model of P(NDI2OD-T2) were parameterised from AA simulations of symmetric P(NDI2OD-T2) monomers in *o*-dichlorobenzene (DCB) at a concentration of ≈55 g L^{-1} (18 monomers, 2937 solvent molecules). Note that this concentration is substantially higher than both those in
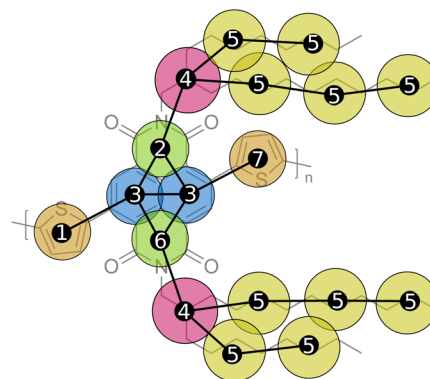
Fig. 1 Chemical structure of the P(NDI2OD-T2) monomer with the CG model overlaid. CG sites are labelled (1–7) and coloured by their site type. To preserve the backbone geometry of the AA representation, two different site types for the thiophenes (1 and 7) and imides (2 and 6) were defined, which have the same non-bonded and bond length parameters, but different bond angle parameters. Note that the terminal methyl group of one of the side-chains was not included in the coarse-graining in order to reduce the number of site types and to facilitate parameterisation.

experiments of P(NDI2OD-T2) aggregation[10,12] and used later in the CG simulations (2–10 g L^{-1}). This higher concentration was used to obtain good statistics for the configurational distributions needed for the coarse-graining procedure at a reasonable computational expense. The CG bonded parameters were parameterised based on AA simulations of P(NDI2OD-T2) trimers in DCB, so as to include the bonds between monomers, at the same concentration as the monomer simulations (8 trimers, 3229 DCB molecules). It should be noted, as shown in Fig. 1, that the backbone of each P(NDI2OD-T2) monomer consists of three molecular fragments (the NDI and two thiophene units) connected by rotatable bonds, for a total of nine backbone units per trimer, which is consistent with the chain lengths of the order of 10 backbone units used for previous coarse-grained parameterisations of semiflexible semiconducting polymers from AA models.[52,69,70] The parameters described above for P(NDI2OD-T2) were used, along with unmodified OPLS parameters[60–66] for the solvent (see ESI, section S13†). DCB was chosen as the solvent based on solution-phase UV–visible spectroscopy data of P(NDI2OD-T2),[10,12] which indicate little aggregation, guaranteeing a homogeneous system as required by the coarse-graining process.

A truncated and shifted Lennard-Jones (LJ) potential with a cutoff of 11 Å was used in the AA simulations, consistent with the parameterisation of the OPLS force field.[60,66] The simulations were carried out at constant temperature and pressure using a Nosé–Hoover thermostat[71,72] and barostat. Electrostatic interactions were calculated using the particle–particle particle–mesh (PPPM) method.[73] Hydrogen-containing bond lengths were constrained to their equilibrium lengths using the SHAKE algorithm,[74] and a timestep of 2 fs was used.[75] More details of the simulation methods are given in the ESI, section S2.†

2.2 Coarse-grained model parameterisation

The CG representation of P(NDI2OD-T2) is given in Fig. 1, in which each spherical CG site is centred at the centre-of-mass of the atoms that comprise it. Each aromatic ring was assigned to a single CG site, and the side-chain sites composed of three (site type 5) or four (site type 4) CH$_n$ groups. This mapping groups atoms whose motion is expected to be strongly correlated into the same site. We note that while the entire NDI group is relatively rigid and could theoretically be coarse-grained into a single site, doing so with a spherical CG site would not capture its significantly anisotropic shape, which is expected to impact chain packing in polymer aggregates. An alternative approach, which has previously been taken for other organic-semiconductor systems,[31,76–79] would be to coarse grain the system using anisotropic sites, which would enable the NDI group to be accurately represented by a single CG site. Site masses were taken as the sum of the masses of atoms in the AA representation that composed each CG site. Where atoms were shared between sites, such as within the NDI group, the masses were split evenly over the sites (e.g. the carbon shared between site 2 and the two sites of type 3 in the

CG representation contributed 1/3 of its mass to each of these sites; see ESI Table S15† for a list of masses).

CG simulations were carried out in implicit solvent using Langevin dynamics[80] to capture the effect of stochastic collisions with solvent molecules and frictional drag. The equations of motion of a particle i with mass m_i and position $\mathbf{r}_i$ are

$$m_i\ddot{\mathbf{r}}_i(t) = \mathbf{f}_i(t) - m_i\gamma\dot{\mathbf{r}}_i(t) + \boldsymbol{\zeta}_i(t), \qquad (1)$$

where $\mathbf{f}_i(t)$ is the force acting on particle i due to the CG potential and $m_i\gamma\dot{\mathbf{r}}_i(t)$ the frictional drag in a solvent with friction coefficient γ. $\boldsymbol{\zeta}_i(t)$ is the force due to random collisions with the solvent, which satisfies $\langle\boldsymbol{\zeta}_i(t)\rangle = 0$ and $\langle\boldsymbol{\zeta}_i(t)\boldsymbol{\zeta}_j(t')\rangle = 2\gamma k_{\mathrm{B}} T m_i \delta_{ij}\delta(t - t')$, which is an expression of the fluctuation–dissipation theorem.[81] Most simulations used a friction coefficient of $\gamma = 20$ ps^{-1}, chosen to give a monomer diffusion coefficient consistent with that of the AA model in DCB (ESI Fig. S6†). With this friction coefficient, the maximum timestep that gave stable simulations was 8 fs, which was used in all CG simulations unless otherwise stated. A number of simulations were also conducted at lower friction to speed up equilibration as well as to examine the effect of viscosity on both single-chain folding and multi-chain aggregation. These simulations used $\gamma = 2$ ps^{-1} and a timestep of 5 fs. All simulations were conducted at constant volume and temperature.

The CG model was parameterised using the iterative Boltzmann inversion (IBI) method,[82,83] which has been used previously to systematically coarse-grain OSCs.[52,84] This method aims to match the local structural distribution functions between equivalent AA and CG systems via iterative optimisation of the CG interaction potentials. We followed the procedure outlined in ref. 52 and 84, with the potential at each iteration, $U_{n+1}(x)$, updated according to

$$U_{n+1}(x) = U_n(x) + a_n \ln\left(\frac{P_n(x)}{P_{\mathrm{target}}(x)}\right), \qquad (2)$$

where $U_n(x)$ is the potential at iteration n as a function of the variable x, $0 \leq a_n \leq 1$ is a parameter that controls how much the potential changes between iterations, $P_{\mathrm{target}}(x)$ is the target AA distribution, and $P_n(x)$ is the CG distribution at iteration n. For non-bonded interactions, $P(x)$ is the radial distribution function (RDF) $g(r)$. For bonded interactions it takes the forms $P_{\mathrm{bond}}(l)/l^2$, $P_{\mathrm{angle}}(\theta)/\sin(\theta)$, $P_{\mathrm{dihed}}(\phi)$, and $P_{\mathrm{improp}}(\psi)$ for the bond length, angle, dihedral, and improper dihedral distributions, respectively, where l is the bond length, θ the bond angle, and ϕ and ψ proper and improper dihedral angles, respectively. In all cases, the resulting potentials were fit to analytical functions defined in ESI section S3.1,† giving good agreement between the CG and AA distributions, as shown in ESI section S3.3.†

The fit of the analytical CG potential functions to the Boltzmann inversion of the target distributions was used as the initial guess for all parameters, with the value of the LJ energy parameter $\varepsilon_{\mathrm{LJ}}$ for each of the non-bonded interactions constrained to be initially $0.1 \leq \varepsilon_{\mathrm{LJ}} < 1$ kcal mol^{-1} in order to prevent extensive aggregation. The constraint on $\varepsilon_{\mathrm{LJ}}$ was

removed for the iterative procedure. Bonded interactions were optimised first, by comparing the target distributions from the AA trimer simulations with distributions from an equivalent CG system. The non-bonded parameters were then optimised by comparing the target distributions from the AA monomer simulations with distributions from an equivalent an equivalent CG system. The volume of the CG system was fixed at approximately the average volume of the equilibrated AA system in each case. Further details of the coarse-graining procedure are given in the ESI, section S3.†

Optimising the non-bonded interactions independently of, and after, the bonded interactions as we have done can potentially perturb the bonded distributions so that they no longer match the corresponding AA distributions. We verified that this was not the case by comparing the AA bonded distributions to those obtained from a 100 ns simulation of the CG model with the parameterised bonded and non-bonded interactions (ESI section S3.3†). Good agreement between the AA and CG distributions was still found in all cases. The final bonded and non-bonded parameters are given in the ESI in section S14.†

2.2.1 Solvent quality. To model a range of solvent conditions, we defined two additional sets of non-bonded CG parameters to approximate solvation in a poorer solvent and a better solvent than DCB. The IBI method used in this work relies on the AA reference systems being homogeneous, which makes it challenging to parameterise models in poor solvents, in which extensive aggregation is expected. Instead of explicitly parameterising the model in other solvents, we adopted a simpler approach of scaling all the non-bonded LJ energy parameters obtained in DCB to give either 20% stronger or 20% weaker interactions. These two cases will be referred to as the poor solvent and the "good" solvent, respectively. The system with the original parameters, parameterised in DCB, will be called the intermediate solvent. Note that the original energy parameters are generally larger for interactions involving backbone sites than side-chain sites, so uniform scaling of the parameters results in a greater absolute change in the strength of backbone interactions. Note also that we did not consider the effect of solvent quality on the dihedral potentials in the model. It is important to note that some aggregation occurred in the CG simulations with all three of these solvent conditions, which means that all were, according to the conventional polymer physics definition, relatively poor solvents. We therefore use the terms "good" and intermediate to refer to the better solvents, in which, according to UV–vis spectra of P(NDI2OD-T2) in solvents such as DCB and chlorobenzene, only intermediate aggregation is observed.[10,12] The non-bonded parameters and potentials for these three cases are given in the ESI, section S14.†

To confirm that the scaled solvent parameters were reasonable representations of the behaviour of P(NDI2OD-T2) in a better and a poorer solvent than DCB, we calculated the free energy as a function of backbone centre-of-mass separation in AA systems of two P(NDI2OD-T2) monomers in DCB, 1-chloronaphthalene (a better solvent than DCB), and toluene (a poorer solvent than DCB, and one of those shown to promote extended rod-like structures experimentally[10]). This free energy was compared with the equivalent free energy calculated for CG monomers with the poor, intermediate, and "good" solvent parameters. Free energies were calculated using on-the-fly probability enhanced sampling (OPES),[85] which, similarly to metadynamics,[86] facilitates the exploration of the probability distribution of interest (here that of the centre-of-mass separation) by depositing small repulsive Gaussians (kernels) in collective-variable space over the course of the simulation in order to bias the system against exploring regions it has already visited. Further details on the method are given in the ESI, section S4.1.† These calculations were carried out using the PLUMED software package, version 2.5.4.[87,88] Comparing the final free energy curves as a function of centre-of-mass separation showed reasonably good agreement between the "good" solvent parameters and 1-chloronaphthalene, the as-parameterised DCB model and its AA equivalent, and the poor solvent parameters and toluene (ESI Fig. S15†), although the CG models do not capture some of the oscillations in the all-atom free energies, especially for DCB. This finding confirms that the poor solvent conditions used here should be a reasonable representation of the conditions that have been experimentally shown to give extended rod-like aggregates. Furthermore, the Kuhn length in single-chain simulations of our CG model, which varied from 9.2 nm and 10.0 nm as the solvent was changed from poor to "good" (see ESI Table S4†) is similar to the value of 10.2 nm obtained from fitting of small-angle neutron scattering (SANS) data for P(NDI2OD-T2) in DCB,[89] indicating that the model captures the semiflexibility of this polymer reasonably well.

2.2.2 Backbone flexibility. As the folding of single polymer chains has been shown to depend on backbone stiffness,[30–32] we examined two different backbone stiffnesses quantified by the Kuhn length: the as-parameterised stiffness (which we will refer to as "regular" stiffness), and a more flexible chain (referred to as "flexible"). To model the more flexible chain, the coefficients of the 1(7)–3–3 angles and 3–7–7–1 dihedral were reduced to 1% of the values of the regular stiffness backbone (see ESI section S14† for parameters and plots of these modified potentials), reducing the Kuhn length of the chain by 30–40% in the "good" solvent (ESI Fig. S16†).

2.3 Coarse-grained simulations

In order to determine the effects of the rates of single-chain folding and multi-chain aggregation, solvent quality, and backbone stiffness on the final aggregate structure, we examined the folding of a number of isolated single-chain systems of various molecular weights as a function of backbone stiffness and solvent quality, as well as multi-chain systems representative of the experimental systems studied in ref. 10. All simulations were conducted at constant volume and temperature using Langevin dynamics.

2.3.1 Single-chain simulation. Single-chain simulations were conducted for the two backbone flexibilities and the two extremes of solvent quality ("good" and poor). To examine the effect of molecular weight on folding kinetics, four different

chain lengths were studied: 10, 20, 30, and 40 monomers, corresponding to $M_n \approx$ 10, 20, 30, and 40 kDa, respectively (40 monomer chains were only simulated with the regular-flexibility backbone in the poor solvent). Simulations of a 20mer with a regular-flexibility backbone in the poor solvent were also conducted with a 10× lower friction coefficient, to determine the effect of solvent viscosity on the rate of single-chain folding. A number of independent simulations were conducted for each system type, with each initially run with non-bonded interactions modelled using purely repulsive Weeks–Chandler–Andersen (WCA) potentials to give an extended chain conformation characteristic of a good solvent, before switching to the CG LJ potentials to determine the time scale of single-chain folding. To obtain an accurate estimate of the folding time, which varied with the system parameters, a system-dependent simulation duration was used. Details of the simulation procedure are given in the ESI (section S5.1†), with a list of the systems studied and key parameters in Table S1.†

2.3.2 Multi-chain simulation. Multi-chain simulations were conducted for systems of 10, 20, 30 and 40mers of P(NDI2OD-T2) in the "good", intermediate, and poor solvents, with flexible (poor solvent only) and regular-flexibility (all three solvents) backbones. The P(NDI2OD-T2) system studied in SAXS experiments that showed extended aggregates consisted of approximately 30-monomer chains at a concentration of 5 g L^{-1}.[10] Assuming a simulation box roughly three times the polymer contour length to avoid finite-size effects, a system of 30mers at this concentration would contain too many atoms to be easily simulated on the µs time scale. Instead, we focused most of this work on the shorter 20mers. Even shorter (10mer) and longer (30mer, 40mer) chains were also considered for a few select cases. In order to achieve approximately the same behaviour as the experimental 30mer system, we set the concentration of each system so that the ratio of the polymer volume fraction, ϕ_V, to the overlap volume fraction, ϕ^*, was approximately the same for all chain lengths, and close to that in the SAXS experiments. This choice was motivated by recent work that showed the concentration of a polymer solution relative to the polymer overlap concentration to be a key predictor of OSC device performance due to its effect on the extent and type of aggregation.[17] Making the crude approximation of ideal chains gives $\phi^* \propto 1/N^{1/2}$ for polymer chain length N, and so constant ϕ_V/ϕ^* corresponds to constant $\phi_V N^{1/2}$. The concentrations that gave the same $\phi_V N^{1/2}$ as the experimental system of 30mers at 5 g L^{-1} were 4 g L^{-1} for 40mers, 6 g L^{-1} for 20mers and 8.5 g L^{-1} for 10mers. Unless otherwise stated, the results presented below are for these concentrations. A number of additional systems were studied at different $\phi_V N^{1/2}$ to elucidate the effect of concentration on multi-chain aggregation. As with the single-chain simulations, the chains were first allowed to relax to conformations consistent with a good solvent by initially using purely repulsive WCA non-bonded interactions, before switching to the CG LJ potentials. A detailed description of the simulation procedure is given in the ESI (section S5.2†), with a list of the systems studied and key parameters in Table S2.†

3 Results and discussion

3.1 Solution-phase behaviour of P(NDI2OD-T2)

We begin by analysing the solution-phase morphology of multi-chain systems of P(NDI2OD-T2) in "good", intermediate, and poor solvents. Aggregate properties and the kinetics of aggregate formation were analysed in a number of ways. For all analyses, two chains were considered to be in the same aggregate if any of their monomers had a backbone centre-of-mass separation of less than 7 Å to include chains that were part of persistent rather than transient aggregates: cut-offs significantly smaller than 7 Å missed chains that persistently aggregated, whereas cut-offs significantly larger than 7 Å counted chains that rapidly disaggregated. In all cases, aggregation occurred through interactions of the NDI groups (CG site types 2, 3, and 6) in the π-stacking direction.

3.1.1 Aggregate size (number of chains). The size of an aggregate, N_{agg}, was defined as the number of chains that it contained. Fig. 2a shows the growth over time t of the average aggregate size, $\langle N_{agg}(t)\rangle$, in each solvent. For 20mers in the poor and intermediate solvents, aggregation initially occurred
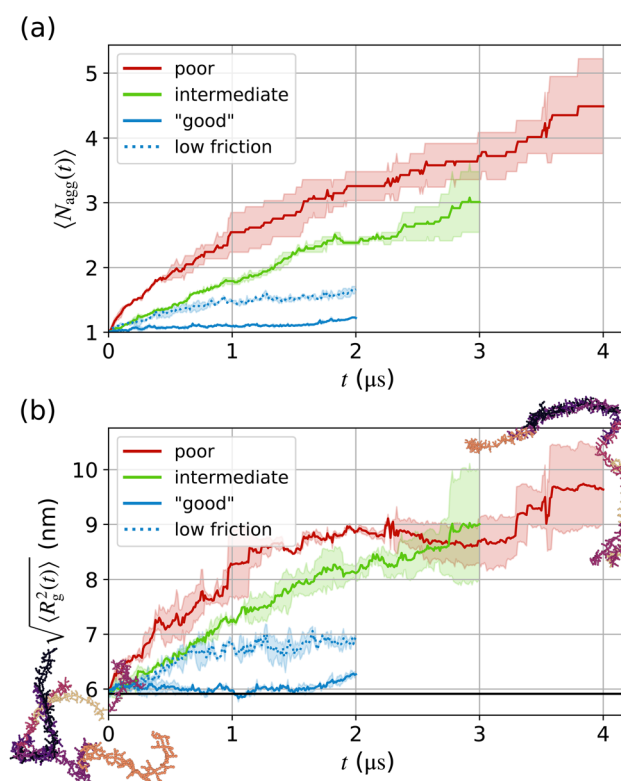
Fig. 2 (a) Average aggregate size (number of chains in aggregate) and (b) RMS aggregate radius of gyration *versus* time. Low-friction results in the "good" solvent are also shown (dotted blue line). Shaded regions indicate 95% confidence intervals based on two replicate simulations. The horizontal black line in (b) indicates the RMS R_g of a 20mer over the final 2 µs of the single-chain simulations in "good" solvent conditions. Inset images in (b) show snapshots of the same aggregate in a poor solvent at 0.7 µs and 4 µs, highlighting the more ordered, rod-like structure at later times.

rapidly, with the average aggregate size approaching three chains in the intermediate, and four chains in the poor solvent, after 3 μs. In the better ("good") solvent, aggregation occurred much slower, with the average aggregate size remaining under 2 chains, indicating the presence of many unaggregated chains. The time dependence appears to be roughly independent of chain length at the same $\phi_V N^{1/2}$ (see ESI Fig. S17†), but shows a strong dependence on concentration, which will be discussed further in section 3.3.

To get a better understanding of the long-time behaviour in the "good" solvent, which should be representative of solvents in which some aggregation is expected but the formation of rod-like aggregates is not, we conducted the same simulation with lower friction in order to speed up the dynamics of the system. Although this will not accurately capture the kinetics of aggregation in a realistic solvent, the equilibrium behaviour should be the same. Aggregation in this low-friction system occurred faster, as expected, but extensive aggregation was still not observed, with the average aggregate size remaining below 2 (Fig. 2a).

3.1.2 Aggregate conformation (radius of gyration). The experimental SAXS results showed that P(NDI2OD-T2) aggregates in extremely poor solvents may be significantly larger than in better solvents, and have a rod-like structure.[10] The conformation of aggregates in solution was characterised by their radius of gyration, R_g. Over time, as the average aggregate size increased, we observed a corresponding increase in the root-mean-squared (RMS) R_g (Fig. 2b). Separating this into the R_g of aggregates of a specific size showed that as the aggregate size (N_{agg}, number of chains) increased, the RMS R_g of the aggregate also increased beyond that of a single chain much more rapidly than would be expected if stacking in a perfect π-stacking arrangement (Fig. 3). Examining the evolution of the aggregate structure over time (see *e.g.* the structures inset in Fig. 2b) shows an aggregation mechanism in which chains initially collide with random backbone orientations, before 'zipping' up to form a more rod-like structure. Although the aggregates are still relatively small, this behaviour is indicative of the formation of extended aggregates in which chains are not fully overlapping. Due to computational constraints, it is challenging to model a system large enough to form aggregates with R_g much larger than observed here. However, the trend towards more extended structures suggests that the formation of larger, extended aggregates is expected.

3.2 Partially overlapping chains lead to extended aggregates in poor solvents

The results presented in the previous section show behaviour consistent with the formation of extended, multi-chain aggregates in poor solvents, as observed experimentally.[10] Although the increase in R_g is relatively limited due to system-size and time-scale limitations, the steady growth of the RMS R_g with aggregate size suggests that large aggregates are feasible, with values already reaching multiple times that of a single chain. Previous Monte Carlo simulations of a generic bead–spring model of a semiflexible polymer[34] indicated that the thermo-
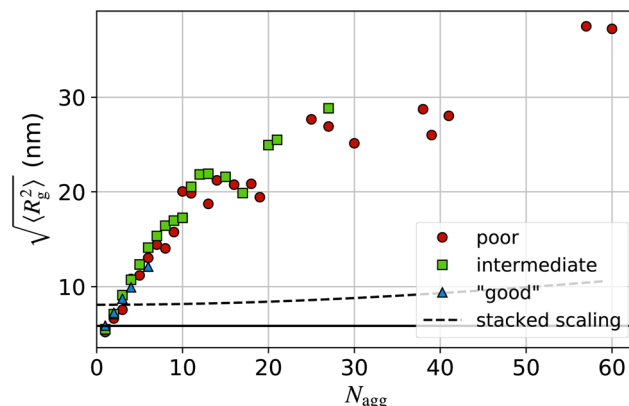
Fig. 3 RMS radius of gyration as a function of aggregate size for 20mers in varying solvent qualities. Each data point is an average over all aggregates of size N_{agg} found in simulation snapshots sampled at 10 ns intervals over the entire simulation. Note that not all possible values of N_{agg} appear in the plot as larger aggregates generally formed from the aggregation of two already large aggregates. The horizontal solid black line indicates the RMS R_g for a single 20mer in the "good" solvent conditions, calculated as described in Fig. 2. The dotted black line is an approximation of the radius of gyration of an aggregate with fully overlapping chains that form a rectangular block, calculated as $R_g^2 = (L^2/12) + (R^2/12)$ where L is the contour length of a single polymer chain (approximated as 20×1.4 nm for a chain of twenty 1.4 nm-long monomers), and R is the aggregate dimension in the π-stacking direction, in which each additional chain is assumed to add 0.4 nm.

dynamically favoured aggregate in a poor solvent is one in which all monomers between chains overlap to give a fully stacked bundle of chains. However, the formation of fully overlapping aggregates cannot explain the large values of R_g observed experimentally or in our simulations in a poor solvent. Instead, to explain the observed formation of large rod-like structures, chains must not be fully overlapping, allowing for the growth of the aggregate in a brickwork-like fashion. For non-overlapping chains to lead to significant growth of aggregates, it is necessary that these incompletely overlapped chain pairs be sufficiently stable that they are inseparable, or at least do not separate on the time scale of further aggregation, such that they become effectively trapped as additional chains are incorporated into the aggregate.

To characterise whether polymer chains in aggregates were overlapping or not, and whether they were likely to be trapped in those structures, we have defined three order parameters: N_{pair}, N_{total}, and N_{trap}, as illustrated in Fig. 4. For all three quantities, monomers were considered to overlap if their centre-of-mass separation was less than 7 Å. N_{pair} defines the number of overlapping monomers between a given pair of polymer chains. A value $<N$ (or $N_{pair}/N < 1$), where N is the polymer chain length, indicates that two chains only overlap partially. N_{total} extends this parameter to include the number of overlapping monomers between a chain and any other chain. Therefore, a value of N (or $N_{total}/N = 1$) indicates either a pair of fully overlapping chains, or a chain that is fully covered by multiple other chains in a partially overlapping fashion.
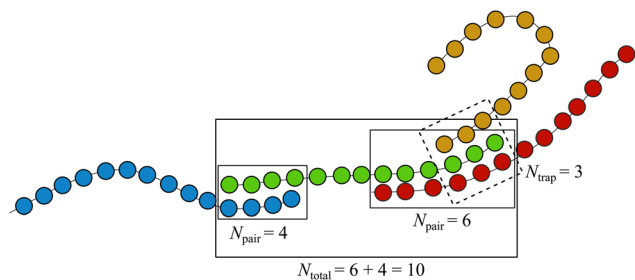
Fig. 4 Definitions of order parameters quantifying chain overlap (for the green chain). N_{pair} is the number of overlaps between a single pair of chains. N_{total} is the overlaps between a chain and any other chain. N_{trap} is the number of monomers on the specified chain that have a monomer from a different chain on each face.

Finally we considered monomers to be trapped in an aggregated structure if they had a monomer on each face. N_{trap} was thus defined as the number of monomers in aggregates that overlap with two other monomers on separate chains.

3.2.1 Chain overlap fraction. As incompletely overlapping chains are required to give a substantial increase in R_g as aggregates grow, we first examined the number of overlaps between pairs of chains, N_{pair}. Only chains with overlaps were counted so this variable has a minimum value of 1. Fig. 5a shows the evolution of $\langle N_{pair} \rangle$ over time for 20mers in the three different solvent conditions studied. By approximately 1 µs, the average overlap fraction $\langle N_{pair} \rangle/N$ of 20mers in the poor solvent has converged to 0.4 (8 overlaps) and does not appear to increase further over the rest of the simulation. This is well below the expected 100% overlap predicted previously as the equilibrium structure.[34] In the "good" solvent, however, $\langle N_{pair} \rangle/N$ is still increasing, albeit very slowly. In better solvents in which the effective interchain interactions are still attractive, we expect that chains are able to separate rapidly enough that less thermodynamically favourable structures, being those held together by only a few monomers, do not become kinetically trapped by the aggregation of more chains around them. Over time, this behaviour, where thermodynamically less favourable partially overlapping chains can separate, will give aggregates that tend toward the equilibrium state, for which the average overlap fraction $\langle N_{pair} \rangle/N$ will depend on the competition between the attractive interchain energy and chain entropy. While Fig. 5a suggests that this process may be occurring, especially in the "good" solvent, the time scale of this process appears to be too long to observe full rearrangement within the simulation duration. We have compared the behaviour in the "good" solvent with an equivalent system with lower friction to elucidate the equilibrium behaviour. This low-friction system showed greater overlaps between aggregated chains, indicating that when able, the system appears to converge towards an equilibrium state with more overlapped chains than in the kinetically trapped poor-solvent system, despite the solvent quality being better.

Comparing the behaviour for different chain lengths at the same $\phi_V N^{1/2}$ (ESI Fig. S18†), a slightly lower average overlap
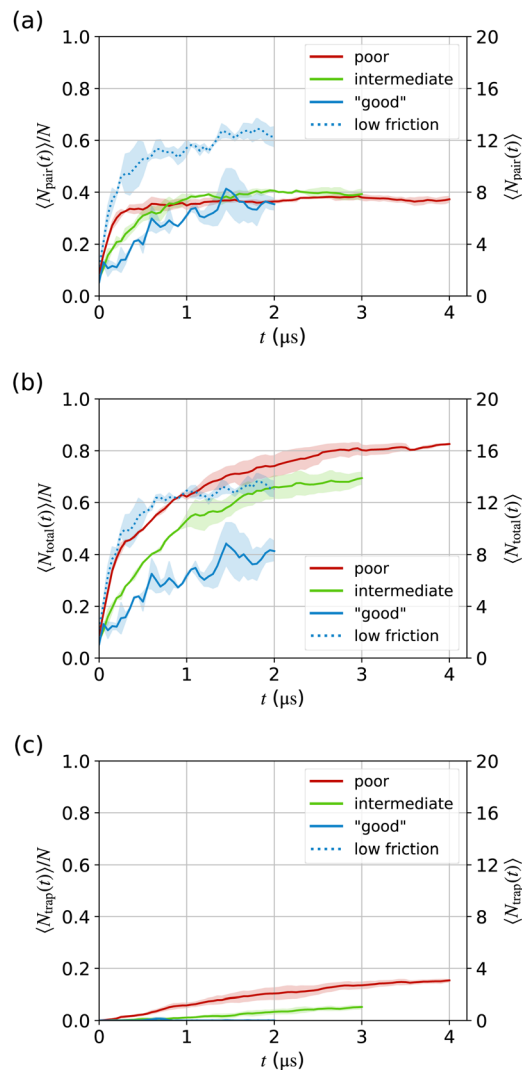
Fig. 5 Average (a) chain overlap fraction $\langle N_{pair} \rangle/N$, (b) total overlap fraction $\langle N_{total} \rangle/N$, and (c) trapped monomer fraction $\langle N_{trap} \rangle/N$ *versus* time in multi-chain 20mer systems in different solvents. The dotted blue line indicates the simulations in the "good" solvent with low friction. Shaded regions indicate 95% confidence intervals calculated for the two replicate simulations. Only chains that overlapped with at least one other chain were counted in calculating these quantities so in (a) and (b) the value can never be zero. Accordingly, neither (a) nor (b) alone gives insight into the degree of aggregation. In (c), a value of zero indicates that all monomers in the system that overlap with any other chains overlap only on one face. For the two "good" solvent curves, this value is always zero.

fraction $\langle N_{pair} \rangle/N$ was observed with increasing chain length, which may be attributed to faster folding of the longer single chains. This behaviour will be discussed in more detail in section 3.3. It should also be noted that we have assumed ideal chains in using $\phi_V N^{1/2}$ to scale the polymer concentration, from which the behaviour of our CG model in poor solvents is likely to deviate.

3.2.2 Stability of partially overlapping aggregates. The differences in the evolution of $\langle N_{pair} \rangle$ over time in the "good"

and poor solvents, with the structure able to rearrange towards more fully overlapping in the "good" solvent but trapped in partially overlapping structures in poorer solvents, suggests that aggregation through fewer monomers is sufficient to hold two chains together as solvent quality decreases. If partially overlapping chains are effectively inseparable, at least on the time scale of becoming trapped by further aggregation, a build-up of extended aggregates with increasing R_g will occur.

The strength of the attraction between two monomers in the different solvents was estimated from the free energy as a function of intermolecular separation calculated from OPES simulations (Fig. 6). Although this free energy was calculated as a function of distance only (*i.e.* not considering the orientation of the particles, which is important for distinguishing different aggregate geometries), the minimum at ≈4 Å is due almost exclusively to π-stacked structures as it is the only configuration that allows such close packing. The free energy preference for aggregation of a pair of monomers was approximately 3.8 kcal mol^{-1} (6.4 $k_B T$) in the poor solvent, 2.2 kcal mol^{-1} (3.7 $k_B T$) in the intermediate solvent, and 0.9 kcal mol^{-1} (1.5 $k_B T$) in the "good" solvent ($k_B T$ at T = 300 K). In the poor solvent, this attraction is sufficiently strong that even chains held together by a single monomer are unlikely to separate often, allowing for the build-up of large aggregates. Additionally, the convergence of $\langle N_{pair}\rangle/N$ to a value far less than 1 in the poor and intermediate solvents (Fig. 5) indicates that with an average N_{pair} of just 8 in the poor solvent, the average number of overlaps neither decreases nor increases with time. This finding again indicates that chains that are much less than fully overlapping are stable for long periods of time in the poorer solvents.

3.2.3 Trapping of aggregates. Although it appears that aggregates in which chains overlap by only a few monomers are stable enough that the chains become effectively inseparable, further aggregation, by which new chains create stacked structures in which parts of a central chain are sandwiched

between two other chains, may result in trapping of the non-equilibrium structure, making these partially overlapping structures even more long-lived. We have quantified this behaviour *via* N_{trap}, the number of monomers that have a monomer on each face (Fig. 4). This variable increased over time in the poor and intermediate solvents, but did not go above zero in the "good" solvent over the simulation duration (Fig. 5c), highlighting again that the chains in the "good" solvent should be able to rearrange towards the fully overlapping structure, whereas those in the poorer solvents will eventually become trapped in partially overlapping structures.

3.3 Single-chain folding is slower than multi-chain aggregation for sufficiently stiff backbones

The results of the previous section show that multi-chain P(NDI2OD-T2) aggregates in which pairs of chain do not fully overlap are sufficiently stable in the poor solvent that they do not separate on the time scale of further aggregation. A further condition that must be satisfied for the build-up of extended rod-like aggregates is that the chains must aggregate before they are able to fold into more compact conformations. Thus, we turn our attention to the relative rates of single-chain folding and multi-chain aggregation.

**3.3.1 Single-chain folding: expected structure and kinetics.** Single CG P(NDI2OD-T2) chains were studied in the poor solvent for flexible and regular-flexibility backbones of length 10, 20, and 30 monomers. 40mers were considered only for the regular-flexibility chains. The single-chain conformation was characterised by the radius of gyration R_g and shape anisotropy κ^2, defined as

$$\kappa^2 = \frac{3}{2}\frac{\lambda_x^4 + \lambda_y^4 + \lambda_z^4}{(\lambda_x^2 + \lambda_y^2 + \lambda_z^2)^2} - \frac{1}{2}, \quad (3)$$

where λ_i are the eigenvalues of the gyration tensor. The shape anisotropy is 0 for a spherical aggregate, and 1 for a linear aggregate, allowing rod-like aggregates to be distinguished from more disordered structures that are likely to be closer to spherical. Based on these characteristics, three broad classes of structure were observed: the extended chain (large radius of gyration, shape anisotropy >0.5), hairpin (smaller radius of gyration, shape anisotropy >0.5), and toroid (small radius of gyration, shape anisotropy <0.5). Over 80 independent 10–25 μs simulations, most chains (whether regular or flexible) displayed a transition to a collapsed conformation within 10 μs. The exception to this behaviour was the 10mers, which were too short to consistently give collapsed structures within the 25 μs time period, assuming the collapsed structure even has significant probability at equilibrium for such short chains. Collapsed structures were either hairpins or toroids, with the 2D distributions as a function of R_g and κ^2, calculated at early (0.5–1 μs, corresponding to the time required to achieve an average aggregate size $\langle N_{agg}(t)\rangle \approx 2$ in the multi-chain simulations), intermediate (4.5–5.5 μs), and late times (9–10 μs) given in Fig. 7 (early time) and in the ESI, Fig. S19† (intermediate and late time). Fig. 7 highlights that, while some
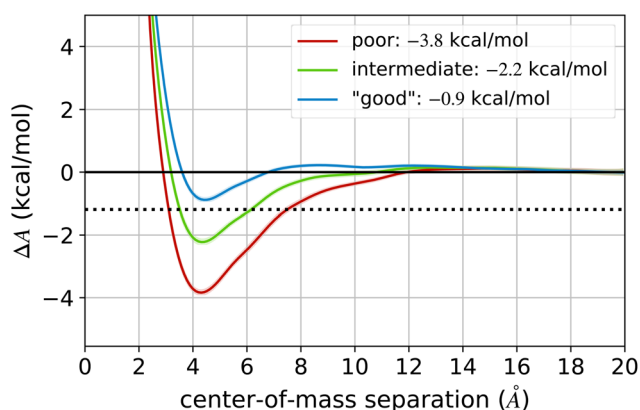
Fig. 6 Free energy *versus* NDI centre-of-mass separation of two CG P(NDI2OD-T2) monomers in various solvents. The minimum value of ΔA in each system is given in the legend. The black dotted line indicates $-2k_B T$ at T = 300 K. A comparison with the same free energy in the AA system can be found in the ESI (Fig. S15†).
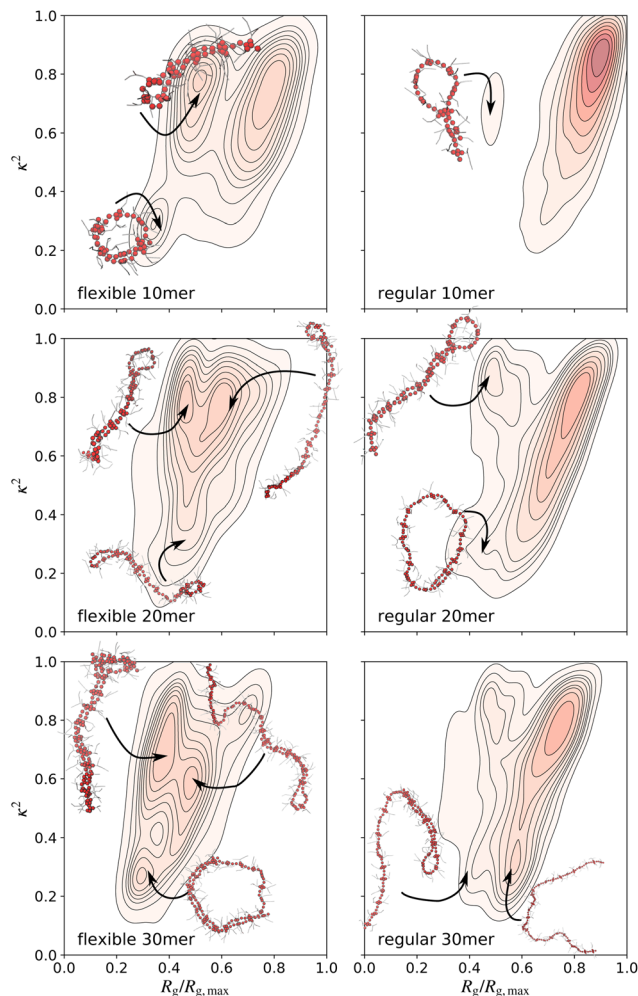
Fig. 7 2D histogram of the radius of gyration R_g and shape anisotropy κ^2 calculated over 80 independent single-chain simulations of various chain lengths in poor solvent conditions. The distributions were calculated over the period 0.5–1 μs, corresponding to the time by which the average aggregate size $\langle N_{agg}(t)\rangle$ in the poor solvent for both flexible and regular backbones was approximately 2 in the multi-chain simulations. The R_g is normalised by $R_{g,max}$, the R_g of a fully extended rod, with $R_{g,max}^2 = L^2/12$ and $L = 1.4$ nm per monomer. Representative conformations are shown near their corresponding peak in the distribution. The colour scale is the same in all cases with darker regions corresponding to higher probability.

chains may have folded by 1 μs, most chains, especially for the regular-flexibility backbones, remained extended on the time scale of initial multi-scale aggregation. The more flexible backbones fold faster, and a greater proportion of these are expected to be folded prior to multi-chain aggregation occurring. Equivalent 2D histograms for the 40mers are given in the ESI in Fig. S20.†

An approximate time scale for single-chain folding, τ_s, was determined as the average time for two monomers in the chain to come into contact (defined as a center-of-mass separation of 7 Å, averaged over 80–100 (or 20 for flexible chains) independent single-chain simulations), consistent with the

definition used in previous work for the folding of semiflexible polymers.[38] We will refer to this as the "first-contact time". This definition of the folding time assumes that, once in contact, the chain does not unfold, which is consistent with the observed behaviour. It also assumes that all chains fold (form at least one contact) within the simulation duration; however, while all the longer chains, and flexible chains of any length, folded, only ≈75% of the 10mers with the regular-flexibility backbone folded within the simulation duration. For those that did not fold, the first-contact time was set to the simulation duration (*e.g.* 25 μs for 10mers) for the purposes of calculating the folding time. The calculated folding time in this case is therefore a lower bound on the actual folding time.

The folding time was found to be 0.5–1 μs for flexible chains and 1.7–11.4 μs for regular chains (Table 1). The slower chain collapse for stiffer chains is consistent with previous reports on the collapse dynamics of single semiflexible chain as a function of flexibility.[35] In terms of the dependence of the kinetics of chain collapse on molecular weight (chain length), scaling of the folding rate with $N^{1/3}$ has been previously reported,[38] and the behaviour of the single chains in this work is consistent with this scaling (ESI Fig. S21†).

3.3.2 Kinetics of multi-chain aggregation. The aggregation kinetics were approximated based on the depletion of single chains in solution, from which the aggregation time τ_c was approximated as the time for the concentration of unaggregated chains in solution to fall to 25% of the original concentration. This proportion corresponded to an average aggregate size of ≈2–2.5 in the simulations in the poor solvent, and so this definition should be representative of the characteristic time scale of multi-chain aggregation. The values of τ_c for the multi-chain systems studied are given in Table 1 alongside the time scales of the corresponding single-chain folding. For 20mers of regular backbone flexibility in the poor solvent, the aggregation time scale is almost five times shorter than that for single-chain folding.

3.3.3 Controlling relative rates of single-chain folding and multi-chain aggregation. As the multi-chain behaviour described above is kinetically controlled, it is expected to

Table 1 Single-chain folding time, τ_s, multi-chain aggregation time, τ_c, and estimated critical concentration $c^\dagger$ at which $\tau_s = \tau_c$ from simulations in the poor solvent for varying chain flexibility, chain length N, and solvent friction coefficient γ. In cases where τ_c is not given, the single-chain concentration in the multi-chain systems remained higher than 25% of the initial concentration over the simulation duration

Flexibility	N	γ (ps^{-1})	τ_s (μs)	τ_c (μs)	$c^\dagger$ (g L^{-1})
Regular	10	20	11.4 ± 2.0	0.42 ± 0.06	1.6 ± 0.2
Regular	20	20	3.37 ± 0.81	0.69 ± 0.09	2.7 ± 0.4
Regular	30	20	2.32 ± 0.57	0.79 ± 0.13	2.9 ± 0.4
Regular	40	20	1.74 ± 0.36	1.65 ± 0.67	3.9 ± 0.9
Regular	20	2	0.37 ± 0.18	0.056 ± 0.003	2.4 ± 0.9
Flexible	10	20	0.98 ± 0.33	—	—
Flexible	20	20	0.56 ± 0.20	—	—
Flexible	30	20	0.45 ± 0.23	—	—

depend on the concentration of the system. Assuming that multi-chain aggregation is a diffusion-limited bimolecular process that occurs *via* binary collisions (all of which lead to aggregation) between aggregates and is dominated by the aggregation of single chains to give an aggregate of two chains, as shown in ESI section S9,† for the conditions studied here under which aggregation occurs on times scales significantly shorter than R^2/D, where R is the typical size and D the typical diffusion coefficient of the aggregating species, the aggregation time scale can be approximated as

$$\tau_c \approx f(N)/c^2, \tag{4}$$

where $c = CN$ is the monomer concentration (or, equivalently, the mass concentration) for chains of length N and concentration C, and $f(N)$ is a function of N that depends on system properties besides c. We define a critical monomer (or mass) concentration $c^\dagger$ above which multi-chain aggregation is expected to be faster than single-chain collapse, by setting $\tau_c = \tau_s$. Combined with eqn (4), this gives

$$c^\dagger(N) \approx \left(\frac{f(N)}{\tau_s}\right)^{1/2}, \tag{5}$$

where $f(N)$ can be determined from τ_c and c measured in the multi-chain simulations. The value of this concentration for regular-flexibility backbones of different chain lengths, N, in the poor solvent is shown in Fig. 8.

If $D \sim N^\beta$ and $R \sim N^\nu$, where β and ν are scaling exponents, the scaling of $f(N) \sim N^{(2-\beta-4\nu)}$ is predicted; then, assuming that $\tau_s \propto N^\alpha$, eqn (5) predicts the scaling (see ESI section S9†)

$$c^\dagger(N) \sim N^{(1-\alpha/2-\beta/2-2\nu)}. \tag{6}$$

We showed previously that $\alpha \approx -1/3$. Given that the polymer conformation was initially that in a good solvent, $\nu \approx 0.6$ according to the Flory theory.[28] In the absence of hydrodynamic interactions, which are neglected in the Langevin dynamics simulations that we have used, the polymer diffusion coefficient is expected to scale with N^{-1},[90] so $\beta = -1$. Thus, under conditions corresponding to this work, eqn (6) predicts a scaling of $c^\dagger \sim N^{0.47}$. The best-fit scaling observed in the simulations was $c^\dagger \sim N^{0.58}$, which is reasonably close to the predicted scaling given the significant approximations made in this simple theory. Note that, accounting for statistical uncertainty, the theoretical $N^{0.47}$ scaling is also consistent with the simulation values in Fig. 8.

Arguably the most significant assumption in the theory is that aggregation only occurs between two single chains to give an aggregate of size two. This is not the case for the simulations studied in this work, in which many aggregates form containing more than two chains (see Fig. 3). In addition, the definition of τ_c as the time taken for the concentration of single chains to fall to 25% of the original concentration, although consistent with the notion that the aggregation time scale should correspond to when most chains are in aggregates, is somewhat arbitrary. However, according to the theory in ESI section S9,† the specific choice of this proportion (x) of
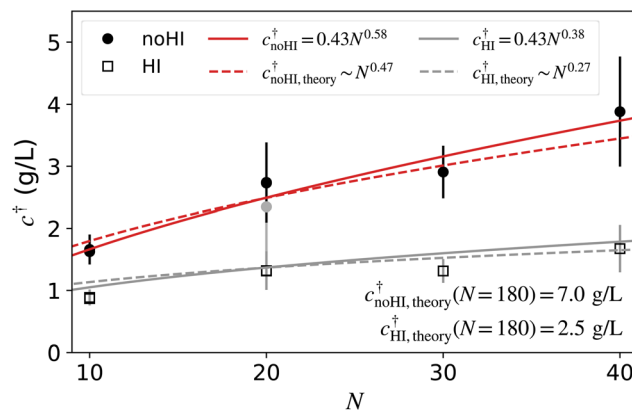
Fig. 8 Critical concentration above which multi-chain aggregation is expected to be faster than single-chain folding as a function of polymer chain length in the poor solvent. For 10 and 20mers, there are multiple indistinguishable overlapping points present for the same chain length, calculated using τ_c from multi-chain simulations at different concentrations. The solid circles are values calculated from the Langevin dynamics simulations (*i.e.* with no hydrodynamic interactions (HI)). The red curves are power-law fits to these data, either with the fit parameters unconstrained (solid line) or with the power-law exponent constrained to that expected from theory (dashed line). The grey circle for N = 20 corresponds to the low-friction system and was not included in the fit. Unfilled squares are values corrected for HI using eqn (7). The grey curves are unconstrained (solid line) and constrained (dashed line) power-law fits to these corrected data analogous to the red lines for the uncorrected data. The value of $c^\dagger$ at N = 180, corresponding to the work in ref. 12, calculated from the theoretical scaling both with ($c^\dagger_{HI,theory}$) and without ($c^\dagger_{noHI,theory}$) the hydrodynamic correction, is indicated on the graph. Error bars indicate two standard errors.

single chains used to define τ_c is not expected to affect the scaling of $c^\dagger$ with N, although $c^\dagger$ at a given N is predicted to scale with ($x^{-1} - 1$).

Another simplifying assumption of the theory used to derive the scaling of $c^\dagger$ with N is that the process of single-chain folding does not affect the multi-chain aggregation rate. As chains fold, they gradually become more compact, reducing the diffusion-limited collision rate between chains. Thus, there is a complex interdependence between the single-chain folding and multi-chain aggregation that cannot be fully captured by the simple theory used here. This means that the estimated critical concentrations are a lower bound: accounting for chain collapse during aggregation will increase the concentration at which aggregation dominates single-chain collapse. Nevertheless, especially for the shorter chains, for which the size difference between a fully extended and collapsed chain is less significant, the calculated concentrations should be a reasonable approximation to the actual concentrations at which folding occurs as fast as interchain aggregation.

Although these simulations used Langevin dynamics, in which hydrodynamic interactions are neglected, an approximate correction to account for the effect of hydrodynamics on the polymer diffusion coefficient can be applied based on the Kirkwood formula for the translational diffusion coefficient of

a macromolecule[90,91] (see ESI section S9, eqn (S50)–(S57)†). This amounts to

$$c_{HI}^{\dagger} = c_{noHI}^{\dagger} \left(\frac{N^{\nu}}{N + N^{\nu}} \right)^{1/2} \qquad (7)$$

$$\approx c_{noHI}^{\dagger} N^{\frac{1}{2}(\nu-1)}, \text{ for } N \gg 1, \nu < 1. \qquad (8)$$

The values of $c^{\dagger}$ obtained from eqn (7) are included in Fig. 8, along with the theoretical scaling of $c_{HI}^{\dagger} \sim N^{0.27}$ from eqn (8) with $\nu = 0.6$, corresponding to the initial good-solvent conformation of the polymer chain. Note that the rate of single-chain folding was not adjusted for hydrodynamic interactions as it is expected to depend on the rate of monomer diffusion rather than that of the whole polymer.[92] The monomer diffusion coefficient was parameterised in the CG Langevin dynamics simulations to match that in the explicit-solvent AA simulations, which include all hydrodynamic interactions for the specified system Hamiltonian, as all particles were explicitly modelled using purely deterministic dynamics. Accordingly, the CG Langevin dynamics simulations effectively account for the hydrodynamics in the AA model at the monomer level.

The predicted values of $c^{\dagger}$ (including hydrodynamic interactions) under conditions corresponding to those used in the work of ref. 10 (30mers, $c \approx 5 \text{ g L}^{-1}$) and ref. 12 (180mers, $c < 1 \text{ g L}^{-1}$) can help explain the contrasting observations in these studies. For 30mers, such as those used in ref. 10 for which extended rod-like aggregates were observed, the critical concentration is predicted to be approximately 1 g L^{-1}, well below the concentrations used in the experiments. At concentrations of 5 g L^{-1} (roughly corresponding to the concentrations used in the simulations conducted in this work) multi-chain aggregation is therefore expected to dominate single-chain collapse, giving rise to the observed rod-like aggregates. The effect of concentration on the behaviour of a number of different systems that are otherwise identical is given in the ESI in Fig. S22,† highlighting that more rapid aggregation, and the formation of larger aggregates, is indeed observed at higher concentrations, with the concentration dependence of the aggregation rate with eqn (4). Extrapolating the observed chain length dependence (Fig. 8) to longer chains (*e.g.* 180mers, consistent with ref. 12), single-chain folding is expected to be the dominant pathway at concentrations up to approximately 2.4 g L^{-1}. These concentrations are above those used in the experiments of up to 1 g L^{-1}.[12] The predicted folding behaviour is therefore consistent with the experimental observations for these longer chains at lower concentrations. This kinetic effect, by which the relative rates of single chain folding and multi-chain aggregation are important for predicting the aggregate structure, reconciles the apparent discrepancy between the experimental studies, and highlights the importance of both concentration and chain length for achieving the desired thin-film morphology.

It is important to note here that $c^{\dagger}$ scales very differently with N compared with the overlap volume fraction ϕ^{*}, which

has been used previously[17] to predict aggregation properties. The work of ref. 17, which considered a single polymer (DPP-DTT, which is significantly different chemically to P(NDI2OD-T2)) at concentrations close to the overlap concentration, suggested that the optimal concentration for achieving high-performing organic field-effect transistor (OFET) devices is the polymer overlap concentration. If this is the case, the optimal concentration is expected to decrease with N, and $c^{\dagger}$ should be roughly constant for constant $\phi_V N^{1/2}$ (approximating $\phi^{*} \sim \phi_V N^{1/2}$). Fig. S23a† shows that this is not the case for the simulations in this work, with the value of $\phi_V N^{1/2}$ at which $\tau_s = \tau_c$ scaling roughly linearly with N.

Our work suggests that there is a lower concentration than the overlap concentration, determined by the relative rates of single-chain collapse and multi-chain aggregation, which might more accurately predict the transition to extended aggregates correlated with good device performance. It should be noted, however, that the theory presented for the scaling of this critical concentration (ESI section S9†) breaks down at the overlap concentration as aggregation will be instantaneous ($\tau_c = 0$) when the chains on average overlap, resulting in a critical concentration that is ill-defined. The values of $c^{\dagger}$ calculated from the short-chain simulations are well below this overlap concentration ($\approx 25 \text{ g L}^{-1}$ for 30mers, assuming the size is the radius of gyration in a good solvent, and higher for shorter chains), though are approaching the estimated overlap concentration for 180mers ($\approx 6 \text{ g L}^{-1}$, calculated using the scaling of R_g with N obtained from the shorter chains in good solvent). This calculated overlap concentration is, however, a lower bound on the value, which will be higher in poor solvents where single chains are more collapsed, so the estimated values of $c^{\dagger}$ for 180mers are still expected to be reasonable.

3.3.4 Effect of solvent viscosity on relative rates of single-chain folding and multi-chain aggregation. All the previous analysis was conducted using the same solvent viscosity (friction coefficient chosen to match diffusion of CG and AA monomers in DCB) in order to facilitate comparison between different solvent qualities. However, the viscosity of toluene (0.560 mPa s at 25 °C) is approximately half that of DCB (1.324 mPa s at 25 °C).[93] It is therefore important that the effect of viscosity on the competition between single-chain folding and multi-chain aggregation be considered, as it should affect the rates of both processes. Based on the theory presented in the ESI (section S9†), the rates of both single-chain folding[92] and multi-chain aggregation are expected to scale linearly with viscosity, as they both depend on the diffusion coefficient of either the monomer or polymer, which from the Stokes–Einstein equation are inversely proportional to solvent viscosity.

To determine the effect of viscosity in the simulations, the single-chain folding and multi-chain aggregation time scales were calculated for a system with Langevin friction coefficient 1/10th the value used for all other simulations in implicit DCB. The measured time constants for single-chain folding (τ_s) and multi-chain aggregation (τ_c) are given alongside the DCB-viscosity results in Table 1. Both the single- and multi-

chain aggregation time scales were found to scale approximately linearly with γ, indicating that while viscosity will change τ_c and τ_s, it will do in such a way that it is not expected to change the calculated value of $c^\dagger$. Indeed, the low-viscosity system is included as one of the 20mer points in Fig. 8 and shows roughly the same scaling of $c^\dagger$ with N as the higher viscosity points.

3.3.5 Effect of aggregation on backbone stiffness. P(NDI2OD-T2) consists of a fused-ring NDI system connected through a bTh group. Flexibility of the backbone therefore comes largely from the rotatable Th–Th and Th–NDI bonds. As aggregation occurs in a manner in which both the NDI and Th groups π stack, aggregation has the effect of reducing the flexibility of the chain. The average Kuhn length b of each chain in a pair of aggregated (fully overlapping, $N_{pair}/N = 1$) 30mers from a simulation in the poor solvent (20.0 monomers, assuming a monomer length of 1.4 nm) was found to be three times that of a single 30mer in the same solvent (6.58 monomers), corresponding to a substantial increase in bending rigidity. This increased backbone stiffness means that folding of sections of the polymer where aggregation has occurred becomes highly unlikely. Details of measurements of the Kuhn length for these and other systems are given in the ESI, section S12.1.†

To determine whether this regime, where the chains are so covered as to prevent further folding, is relevant for the aggregation observed here, the fraction of monomers in aggregates that interacted with other monomers in any other chain was calculated. This variable, N_{total}/N, defined in Fig. 4, gives the total number of monomer–monomer interactions between one chain and any other chain. In the poor and intermediate solvents, this quantity was in excess of 80% of the full chain length (about 16 monomers for 20mers; Fig. 5b) after 4 μs of simulation, indicating that chains that are in aggregates are almost fully covered by other chains. Although the small regions where chains are not overlapped may still be able to fold, the aggregates will be substantially stiffer than the single chains, and effectively stuck in an extended state, from which the further build up of extended rod-like structures can occur.

3.3.6 Effect of backbone flexibility on multi-chain aggregation. To better understand the effect of the single-chain folding kinetics on multi-chain aggregation properties, we examined the behavior of the same P(NDI2OD-T2) polymer with the artificially flexible backbone. Single chains of this flexible polymer exclusively collapsed into more compact structures within the 10 μs single-chain simulations, with relaxation times on the order of 1 μs, rather than remaining extended (Fig. 7 and S19†). At the same concentration as the regular-flexibility 20mers (6 g L^{-1}), the flexible 20mers did not meet the metric discussed above for the calculation of τ_c (single-chain concentration fallen to 25% of the original concentration) within the simulation time of 3 μs. The average aggregate size also remained below 2 chains over this entire period. Given the value of τ_s of ≈1 μs for these chains, the critical concentration $c^\dagger$ must be >6 g L^{-1}. For the concentrations simulated, single-chain collapse therefore dominates multi-chain

aggregation for the flexible chains, in stark contrast to the stiffer regular-flexibility chains.

Comparing the aggregate size (number of monomers) and radius of gyration of the flexible and regular-flexibility backbones shows a slower rate of aggregate growth, and generally more compact structures for the flexible chains than the stiffer regular chains (Fig. 9), as expected from the relative rates of folding and aggregation. This behaviour can be attributed to a more rapid collapse into hairpin/toroid structures, which has the twofold effect of reducing the collision rate due to the more compact structures, and giving more compact structures when collisions do occur, as chains may already be partially collapsed. Examining a system with an even lower concentration (2 g L^{-1}) showed the same behaviour, with very little multi-chain aggregation observed over the simulated time period (Fig. 9a). While there was still a brief initial aggregation period, during which chains that were initially positioned close to each other were able to aggregate prior to folding, little aggregation was observed after this point with the
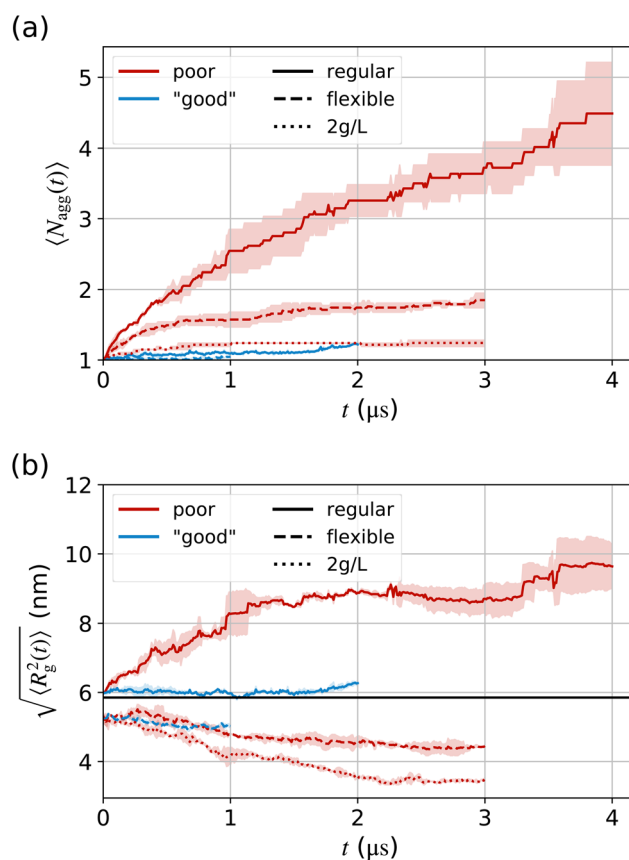
Fig. 9 Comparison of the multi-chain aggregation kinetics of flexible and regular-flexibility 20mers in "good" and poor solvent. (a) Average aggregate size (number of chains in aggregate) and (b) RMS radius of gyration *versus* time. The horizontal black line in (b) indicates the RMS radius of gyration for single regular-flexibility chains in the "good" solvent, calculated as described in Fig. 2. Results for flexible chains in the poor solvent at two concentrations that are expected to be lower than $c^\dagger$ (2 and 6 g L^{-1}) are also presented (dotted and dashed red lines).
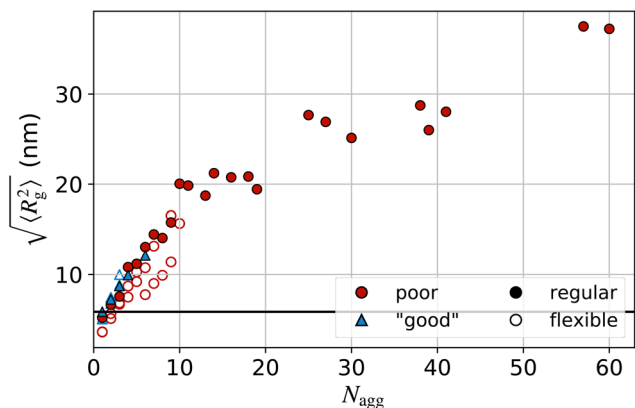
Fig. 10 RMS radius of gyration as a function of aggregate size for regular-flexibility (filled symbols) and flexible (unfilled symbols) backbones in a "good" (blue triangles) or poor (red circles) solvent at a concentration of approximately 6 g L^{-1}. The horizontal black line indicates the value of R_g for single 20mers of the regular backbone in the "good" solvent.

average aggregate size remaining well below 2 over the entire simulation period. Although multi-chain aggregation is not completely prevented at this lower concentration, it is greatly suppressed and could be expected to lead to different final aggregate properties, as observed experimentally.[10,12]

The lower radius of gyration of the flexible chains in the poor solvent observed in Fig. 9b could be attributed to both less aggregation than for the regular-flexibility backbone and more compact aggregates even when consisting of many chains. From the behaviour in Fig. 9a, the flexible-chain aggregates were generally smaller (contained fewer polymer chains) than those with the regular backbone flexibility, indicating that less aggregation does occur as previously discussed. From examination of the R_g of aggregates of various sizes (Fig. 10) it can also be seen that when larger aggregates did form with the flexible backbone, they were generally more compact (lower R_g) than their regular-flexibility counterparts. Overall, the more rapid single-chain collapse of the flexible polymer appears to lead to a stronger preference for intrachain aggregation compared with the regular-flexibility chain. This has the combined effect of reducing the number of aggregates, due to a lower probability of collisions between the more compact aggregates, and giving slightly more compact aggregates where aggregation does occur. Similar behaviour could likely be obtained in a more dilute system of stiffer chains, which, although they take longer to fold, could be expected to collapse prior to extensive multi-chain aggregation at low enough concentration.

4 Conclusions

The solution-phase morphology and dynamics of the organic semiconducting polymer P(NDI2OD-T2) was studied using coarse-grained molecular dynamics simulations in order to understand the reported formation of extended rod-like aggregates in poor solvents. We found that sufficiently strong inter-

molecular attractions (equivalent to poor solvent quality), for which interaction through only a few monomers resulted in effectively inseparable chains, led to the build up of extended aggregates of partially overlapping chains with radii of gyration exceeding that of a single chain. Over time, a trend towards more linear, rod-like aggregates was also observed, consistent with experimental results.[10]

We proposed that this behaviour, which is not predicted by existing theories of polymer solubility, in which decreasing solvent quality is conventionally associated with the formation of compact aggregates, is a kinetic effect associated with the relative rates of multi-chain aggregation and single-chain folding. The formation of extended aggregates is expected under conditions in which aggregation occurs faster than folding, assuming interchain attraction is strong enough to hold chains together in an only partially overlapping chain configuration. Firstly, we showed that under conditions that correspond to P(NDI2OD-T2) in the poor solvent toluene, at concentrations representative of experiments in which rod-like aggregates were observed, aggregated chains overlapping by only around 40% of their full chain length were stable over the duration of the simulations. Under conditions corresponding to a better solvent, this overlap fraction increased over the entire simulation duration, and was expected to reach close to the full chain length. This finding is consistent with the difference between the experimentally observed behaviour in good–intermediate and poor solvents, with rod-like aggregates observed in the poor solvents, and structures in the better solvents showing sizes consistent with single chains, despite some aggregation occurring, suggesting almost fully overlapping chains.

For semiflexible polymers, a class that describes many organic semiconductors, the folding of a single polymer chain is expected to depend on the chain stiffness. By comparing coarse-grained simulations of P(NDI2OD-T2) with a backbone parameterised to match the flexibility of the all-atom model with those of a much more flexible equivalent, we found more rapid folding of the flexible chain. In both cases, the folding rate also displayed a chain-length dependence, increasing with increasing chain length as has previously been reported.[38] By comparing the approximate time scales characterising single-chain folding and multi-chain aggregation in the poor solvent, we determined approximate concentrations at which each of these processes are expected to dominate. A theory relating this critical concentration to the chain length was developed, and the simulations were found to agree well with the predictions. The critical concentration depended both on backbone flexibility, with a more flexible backbone expected to result in predominantly single-chain folding at higher concentrations than a more rigid one, and chain length, with longer chains transitioning from single-chain folding to multi-chain aggregation at higher concentrations due to their more rapid folding. In comparing the simulated solution-phase behaviour of flexible and regular P(NDI2OD-T2) chains, this proposed dependence was observed, with the more flexible chains giving more compact structures and less multi-chain aggregation. This

finding rationalises apparent discrepancies between experimental measurements of the P(NDI2OD-T2):toluene system[10,12] and emphasises the importance of both concentration and chain length on predicting solution-phase behaviour.

Overall, multi-chain aggregation, resulting in the formation of extended rod-like aggregates, is expected to occur under conditions in which (1) partially overlapping chains are inseparable over sufficiently long time scales that they do not rearrange to the energetically favourable fully-overlapped chains before becoming trapped, and (2) single-chain folding occurs slowly enough that it is not expected to occur before multi-chain aggregation prevents further folding. The relative rates of the single- and multi-chain pathways that control the second of these conditions depend on the polymer concentration, chain length, and backbone flexibility. Although we have assumed these processes to be independent, they are likely to show a complex interdependence, with the progress along the single-chain folding pathway affecting the aggregation rate. A more complex model that accounts for these processes more completely, as well as explicitly including the effects of hydrodynamics, will further improve understanding of the solution-phase behaviour of semiflexible polymers. Finally, we have studied this behaviour using a coarse-grained model systematically parameterised to accurately represent P(NDI2OD-T2). However, these results are not expected to be specific to this molecule, with the reported dependence of the solution-phase morphology on solvent quality, backbone flexibility, concentration, chain length, and solvent viscosity expected to be applicable more generally to any semiflexible polymer.

Author contributions

BJB: Investigation, methodology, formal analysis, visualization, writing – original draft. CRM: Conceptualization, writing – review and editing. DMH: Conceptualization, methodology, formal analysis, supervision, writing – review and editing.

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

References

1 A. C. Arias, J. D. MacKenzie, I. McCulloch, J. Rivnay and A. Salleo, *Chem. Rev.*, 2010, **110**, 3–24.

2 F. C. Krebs, T. Tromholt and M. Jørgensen, *Nanoscale*, 2010, **2**, 873–886.

3 X. Liu, B. He, A. Garzón-Ruiz, A. Navarro, T. L. Chen, M. A. Kolaczkowski, S. Feng, L. Zhang, C. A. Anderson, J. Chen and Y. Liu, *Adv. Funct. Mater.*, 2018, **28**, 1801874.

4 J. Liu, L.-K. Ma, Z. Li, H. Hu, F. K. Sheong, G. Zhang, H. Ade and H. Yan, *J. Mater. Chem. A*, 2018, **6**, 23270–23277.

5 M. Alkan and I. Yavuz, *Phys. Chem. Chem. Phys.*, 2018, **20**, 15970–15979.

6 B. Kang, R. Kim, S. B. Lee, S.-K. Kwon, Y.-H. Kim and K. Cho, *J. Am. Chem. Soc.*, 2016, **138**, 3679–3686.

7 G. Yang, Z. Li, K. Jiang, J. Zhang, J. Chen, G. Zhang, F. Huang, W. Ma and H. Yan, *Sci. China: Chem.*, 2017, **60**, 545–551.

8 M. M. Nahid, R. Matsidik, A. Welford, E. Gann, L. Thomsen, M. Sommer and C. R. McNeill, *Adv. Funct. Mater.*, 2017, **27**, 1604744.

9 J. A. Bartelt, J. D. Douglas, W. R. Mateker, A. E. Labban, C. J. Tassone, M. F. Toney, J. M. J. Fréchet, P. M. Beaujuge and M. D. McGehee, *Adv. Energy Mater.*, 2014, **4**, 1301733.

10 M. M. Nahid, A. Welford, E. Gann, L. Thomsen, K. P. Sharma and C. R. McNeill, *Adv. Electron. Mater.*, 2018, **4**, 1700559.

11 A. Luzio, L. Criante, V. D'Innocenzo and M. Caironi, *Sci. Rep.*, 2013, **3**, 3425.

12 R. Steyrleuthner, M. Schubert, I. Howard, B. Klaumünzer, K. Schilling, Z. Chen, P. Saalfrank, F. Laquai, A. Facchetti and D. Neher, *J. Am. Chem. Soc.*, 2012, **134**, 18303–18317.

13 Y.-Q. Zheng, Z.-F. Yao, T. Lei, J.-H. Dou, C.-Y. Yang, L. Zou, X. Meng, W. Ma, J.-Y. Wang and J. Pei, *Adv. Mater.*, 2017, **29**, 1701072.

14 H. Hu, K. Zhao, N. Fernandes, P. Boufflet, J. H. Bannock, L. Yu, J. C. de Mello, N. Stingelin, M. Heeney, E. P. Giannelis and A. Amassian, *J. Mater. Chem. C*, 2015, **3**, 7394–7404.

15 Y. Huang, H. Cheng and C. C. Han, *Macromolecules*, 2010, **43**, 10031–10037.

16 P. Cheng, C. Yan, Y. Li, W. Ma and X. Zhan, *Energy Environ. Sci.*, 2015, **8**, 2357–2364.

17 R. Venkatesh, Y. Zheng, C. Vierson, A. Liu, C. Silva, M. Grover and E. Reichmanis, *ACS Mater. Lett.*, 2021, 1321–1327.

18 M. Li, H. Bin, X. Jiao, M. M. Wienk, H. Yan and R. A. J. Janssen, *Angew. Chem., Int. Ed.*, 2020, **59**, 846–852.

19 S. Liu, W. M. Wang, A. L. Briseno, S. C. B. Mannsfeld and Z. Bao, *Adv. Mater.*, 2009, **21**, 1217–1232.

20 H. Sirringhaus, P. J. Brown, R. H. Friend, M. M. Nielsen, K. Bechgaard, B. M. W. Langeveld-Voss, A. J. H. Spiering, R. a. J. Janssen, E. W. Meijer, P. Herwig and D. M. de Leeuw, *Nature*, 1999, **401**, 685–688.

21 A. Salleo, *Mater. Today*, 2007, **10**, 38–45.

22 L. H. Jimison, A. Salleo, M. L. Chabinyc, D. P. Bernstein and M. F. Toney, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2008, **78**, 125319.

23 M. L. Jones, D. M. Huang, B. Chakrabarti and C. Groves, *J. Phys. Chem. C*, 2016, **120**, 4240–4250.

24 K. Park, E.-Y. Shin, X. Jiao, C. R. McNeill, Y.-H. Kim, S.-K. Kwon and Y.-Y. Noh, *ACS Appl. Mater. Interfaces*, 2019, **11**, 35185–35192.

25 J. Rivnay, M. F. Toney, Y. Zheng, I. V. Kauvar, Z. Chen, V. Wagner, A. Facchetti and A. Salleo, *Adv. Mater.*, 2010, **22**, 4359–4363.

26 A. Wadsworth, H. Chen, K. J. Thorley, C. Cendra, M. Nikolka, H. Bristow, M. Moser, A. Salleo, T. D. Anthopoulos, H. Sirringhaus and I. McCulloch, *J. Am. Chem. Soc.*, 2020, **142**, 652–664.

27 R. Noriega, J. Rivnay, K. Vandewal, F. P. V. Koch, N. Stingelin, P. Smith, M. F. Toney and A. Salleo, *Nat. Mater.*, 2013, **12**, 1038–1044.

28 M. Rubinstein and R. H. Colby, *Polymer Physics*, Oxford University Press, 2003.

29 D. Wang, Y. Yuan, Y. Mardiyati, C. Bubeck and K. Koynov, *Macromolecules*, 2013, **46**, 6217–6224.

30 M. Marenz and W. Janke, *Phys. Rev. Lett.*, 2016, **116**, 128301.

31 A. E. Cohen, N. E. Jackson and J. J. de Pablo, *Macromolecules*, 2021, **54**, 3780–3789.

32 J. Wu, C. Cheng, G. Liu, P. Zhang and T. Chen, *J. Chem. Phys.*, 2018, **148**, 184901.

33 J. Zierenberg, M. Marenz and W. Janke, *Polymers*, 2016, **8**, 333.

34 J. Zierenberg and W. Janke, *EPL*, 2015, **109**, 28002.

35 A. Lappala and E. M. Terentjev, *Macromolecules*, 2013, **46**, 7125–7131.

36 Y. D. Gordievskaya and E. Y. Kramarenko, *Soft Matter*, 2019, **15**, 6073–6085.

37 H. Noguchi and K. Yoshikawa, *J. Chem. Phys.*, 2000, **113**, 854–862.

38 A. Montesi, M. Pasquali and F. C. MacKintosh, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2004, **69**, 021916.

39 M. Kong, I. Saha Dalal, G. Li and R. G. Larson, *Macromolecules*, 2014, **47**, 1494–1502.

40 W. Huang, M. Huang, Q. Lei and R. G. Larson, *Polymers*, 2016, **8**, 264.

41 Y. A. Kuznetsov, E. G. Timoshenko and K. A. Dawson, *J. Chem. Phys.*, 1996, **105**, 7116–7134.

42 N. Yoshinaga, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2008, **77**, 061805.

43 T. X. Hoang, A. Giacometti, R. Podgornik, N. T. T. Nguyen, J. R. Banavar and A. Maritan, *J. Chem. Phys.*, 2014, **140**, 064902.

44 T. Sakaue and K. Yoshikawa, *J. Chem. Phys.*, 2002, **117**, 6323–6330.

45 C. Caddeo, D. Fazzi, M. Caironi and A. Mattoni, *J. Phys. Chem. B*, 2014, **118**, 12556–12565.

46 C. Caddeo and A. Mattoni, *Macromolecules*, 2013, **46**, 8003–8008.

47 T. Wang and J.-L. Brédas, *J. Am. Chem. Soc.*, 2021, **143**, 1822–1835.

48 N. E. Jackson, K. L. Kohlstedt, B. M. Savoie, M. Olvera de la Cruz, G. C. Schatz, L. X. Chen and M. A. Ratner, *J. Am. Chem. Soc.*, 2015, **137**, 6254–6262.

49 D. R. Reid, N. E. Jackson, A. J. Bourque, C. R. Snyder, R. L. Jones and J. J. de Pablo, *J. Phys. Chem. Lett.*, 2018, **9**, 4802–4807.

50 S. M. Ryno and C. Risko, *Phys. Chem. Chem. Phys.*, 2019, **21**, 7802–7813.

51 E. J. García and H. Hasse, *Eur. Phys. J.: Spec. Top.*, 2019, **227**, 1547–1558.

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

53 S. J. Plimpton, *J. Comput. Phys.*, 1995, **117**, 1–19.

54 A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott and S. J. Plimpton, *Comput. Phys. Commun.*, 2022, **271**, 108171.

55 W. M. Brown, P. Wang, S. J. Plimpton and A. N. Tharrington, *Comput. Phys. Commun.*, 2011, **182**, 898–911.

56 W. M. Brown, A. Kohlmeyer, S. J. Plimpton and A. N. Tharrington, *Comput. Phys. Commun.*, 2012, **183**, 449–459.

57 A. Stukowski, *Modell. Simul. Mater. Sci. Eng.*, 2009, **18**, 015012.

58 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.

59 K. H. DuBay, M. L. Hall, T. F. Hughes, C. Wu, D. R. Reichman and R. A. Friesner, *J. Chem. Theory Comput.*, 2012, **8**, 4556–4569.

60 W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225–11236.

61 W. L. Jorgensen and N. A. McDonald, *J. Mol. Struct.: THEOCHEM*, 1998, **424**, 145–155.

62 N. A. McDonald and W. L. Jorgensen, *J. Phys. Chem. B*, 1998, **102**, 8049–8059.

63 M. L. P. Price, D. Ostrovsky and W. L. Jorgensen, *J. Comput. Chem.*, 2001, **22**, 1340–1352.

64 R. C. Rizzo and W. L. Jorgensen, *J. Am. Chem. Soc.*, 1999, **121**, 4827–4836.

65 E. K. Watkins and W. L. Jorgensen, *J. Phys. Chem. A*, 2001, **105**, 4118–4125.

66 G. A. Kaminski, R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen, *J. Phys. Chem. B*, 2001, **105**, 6474–6487.

67 V. Marcon and G. Raos, *J. Am. Chem. Soc.*, 2006, **128**, 1408–1409.

68 A. Pizzirusso, M. Savini, L. Muccioli and C. Zannoni, *J. Mater. Chem.*, 2010, **21**, 125–133.

69 C.-K. Lee, C.-W. Pao and C.-W. Chu, *Energy Environ. Sci.*, 2011, **4**, 4124–4132.

70 C. K. Lee, C. C. Hua and S. A. Chen, *Macromolecules*, 2011, **44**, 320–324.

71 W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695–1697.

72 W. Shinoda, M. Shiga and M. Mikami, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2004, **69**, 134103.

73 R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles*, CRC Press, Boca Raton, 1988.

74 J.-P. Ryckaert, G. Ciccotti and J. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.

75 G. J. Martyna, D. J. Tobias and M. L. Klein, *J. Chem. Phys.*, 1994, **101**, 4177–4189.

76 H. T. L. Nguyen and D. M. Huang, *J. Chem. Phys.*, 2022, **156**, 184118.

77 A. S. Bowen, N. E. Jackson, D. R. Reid and J. J. de Pablo, *J. Chem. Theory Comput.*, 2018, **14**, 6495–6504.

78 A. Khot and B. M. Savoie, *Macromolecules*, 2021, **54**, 4889–4901.

79 D. M. Friday and N. E. Jackson, *Macromolecules*, 2022, **55**, 1866–1877.

80 T. Schneider and E. Stoll, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1978, **17**, 1302–1322.

81 N. Van Kampen, *Stochastic Processes in Physics and Chemistry*, Elsevier, Amsterdam, 3rd edn, 2007, pp. 219–243.

82 D. Reith, M. Pütz and F. Müller-Plathe, *J. Comput. Chem.*, 2003, **24**, 1624–1636.

83 R. Faller and D. Reith, *Macromolecules*, 2003, **36**, 5406–5414.

84 D. M. Huang, R. Faller, K. Do and A. J. Moulé, *J. Chem. Theory Comput.*, 2010, **6**, 526–537.

85 M. Invernizzi and M. Parrinello, *J. Phys. Chem. Lett.*, 2020, **11**, 2731–2736.

86 A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 12562–12566.

87 G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, *Comput. Phys. Commun.*, 2014, **185**, 604–613.

88 M. Bonomi, G. Bussi, C. Camilloni, G. A. Tribello, P. Banáš, A. Barducci, M. Bernetti, P. G. Bolhuis, S. Bottaro, D. Branduardi, R. Capelli, P. Carloni, M. Ceriotti, A. Cesari, H. Chen, W. Chen, F. Colizzi, S. De, M. De La Pierre, D. Donadio, V. Drobot, B. Ensing, A. L. Ferguson, M. Filizola, J. S. Fraser, H. Fu, P. Gasparotto, F. L. Gervasio, F. Giberti, A. Gil-Ley, T. Giorgino, G. T. Heller, G. M. Hocky, M. Iannuzzi, M. Invernizzi, K. E. Jelfs, A. Jussupow, E. Kirilin, A. Laio, V. Limongelli, K. Lindorff-Larsen, T. Löhr, F. Marinelli, L. Martin-Samos, M. Masetti, R. Meyer, A. Michaelides, C. Molteni, T. Morishita, M. Nava, C. Paissoni, E. Papaleo, M. Parrinello, J. Pfaendtner, P. Piaggi, G. Piccini, A. Pietropaolo, F. Pietrucci, S. Pipolo, D. Provasi, D. Quigley, P. Raiteri, S. Raniolo, J. Rydzewski, M. Salvalaglio, G. C. Sosso, V. Spiwok, J. Šponer, D. W. H. Swenson, P. Tiwary, O. Valsson, M. Vendruscolo, G. A. Voth, A. White and The PLUMED consortium, *Nat. Methods*, 2019, **16**, 670–673.

89 M. Xiao, R. L. Carey, H. Chen, X. Jiao, V. Lemaur, S. Schott, M. Nikolka, C. Jellett, A. Sadhanala, S. Rogers, S. P. Senanayak, A. Onwubiko, S. Han, Z. Zhang, M. Abdi-Jalebi, Y. Zhang, T. H. Thomas, N. Mahmoudi, L. Lai, E. Selezneva, X. Ren, M. Nguyen, Q. Wang, I. Jacobs, W. Yue, C. R. McNeill, G. Liu, D. Beljonne, I. McCulloch and H. Sirringhaus, *Sci. Adv.*, 2021, **7**, eabe5280.

90 B. Liu and B. Dünweg, *J. Chem. Phys.*, 2003, **118**, 8061–8072.

91 J. G. Kirkwood and J. Riseman, *J. Chem. Phys.*, 1948, **16**, 565–573.

92 M. Karplus and D. L. Weaver, *Protein Sci.*, 1994, **3**, 650–668.

93 W. M. Haynes, *CRC Handbook of Chemistry and Physics*, CRC Press LLC, Oakville, United Kingdom, 95th edn, 2014.