Electronic Supplementary Information

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

Belinda J. Boehm, Christopher R. McNeill, and David M. Huang

Contents

S1 Parameterisation of all-atom model of P(NDI2OD-T2)

The all-atom (AA) model of P(NDI2OD-T2) was parameterised following the method outlined in ref. 1 for conjugated polymers in the OPLS force field.[1,2] The full set of parameters, as well as the definitions of atom and bond types for each site, are given in Section S13 of this document.

To increase the symmetry and reduce the number of atom types required, a monomer of P(NDI2OD-T2) was considered as a symmetric equivalent to its more commonly represented naphthalene diimide (NDI)–bithiophene (bTh) structure, with the bTh group split between the ends of the monomer (Fig. S1). Within a polymer chain, this will give exactly the same structure, with the only difference being at the terminal monomers.

(a) Asymmetric   **(b)** Symmetric
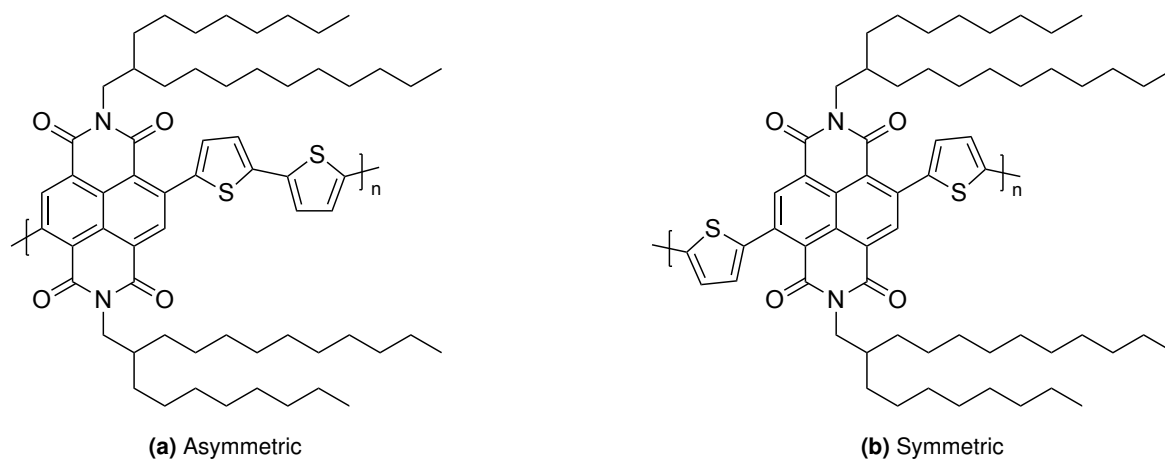
Fig. S1 Structures of the (a) more commonly used asymmetric and (b) our symmetric representation of monomers of P(NDI2OD-T2).

S1.1 Atomic point charges

Atomic point charges were calculated separately for the bTh and NDI conjugated entities of the backbone as described in ref. 1. Each unit was terminated with a thiophene ring at each end to mimic the environment of a fully conjugated backbone, the alkyl sidechains on the NDI group were truncated to methyl groups after the tertiary carbon (i.e. $-N-CH_2-CH-(CH_3)_2$), and the geometry was optimised at the B3LYP/6-31+G** level. ChelpG charges for the optimised geometries were used as atomic point charges. Due to the symmetry of the P(NDI2OD-T2) monomer, atoms in the same chemical environment were assigned the same atom type (see Fig. S27 for the atom types), and their charge averaged over equivalent atoms in the system. Any excess molecular charge, related to the introduction of the terminal thiophenes and truncation of the side-chains, was added to the carbon atoms that connect the side-chains to the backbone (i.e. the CH_2 group bonded to the N, site type 955 for the monomer unit or the central monomer of polymer, 957 for the terminal monomer of polymer; see Fig. S27 for definitions). OPLS charges[2–8] were used for the atoms in the alkyl chains.

S1.2 Bonded potentials

While most bonded parameters were obtained from a combination of the OPLS force-field parameters for equivalent atom types and the optimised geometries of a P(NDI2OD-T2) monomer, with the bond lengths and angles from the optimised geometries and the force constants from the OPLS force field, the parameters for the bond lengths and angles between the NDI and thiophene (Th) conjugated entities were obtained from quantum-chemical calculations. Specifically, the NDI–Th bond and the NDI–Th and Th–Th dihedrals were explicitly parameterised using constrained geometry optimisations.

NDI–Th bond

The NDI–Th bond was parameterised from the energy of the entire asymmetric monomer unit (B3LYP/6-31+G**) with the length of the NDI–Th bond fixed at 0.01 Å intervals from $l_0 - 0.1$ Å to $l_0 + 0.1$ Å, where l_0 is the bond length determined from the unconstrained optimisation. The potential energy was fit to a harmonic potential of the form
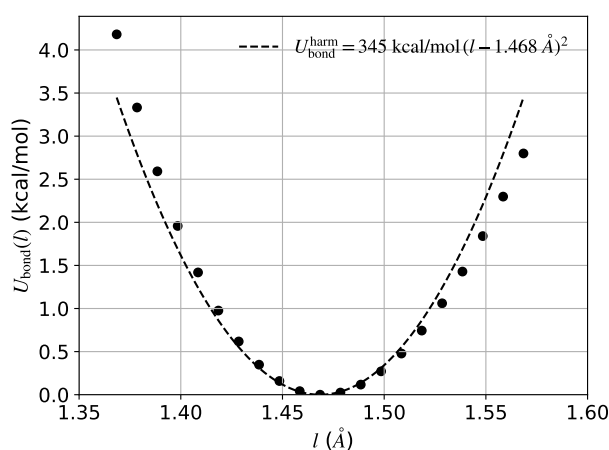
$$U_{\text{bond}}^{\text{harm}}(l) = k_{\text{b}}(l - l_0)^2, \tag{S1}$$

where k_{b} is the bond stretching coefficient (half the force constant) (Fig. S2).

Dihedral potentials

The explicitly parameterised dihedral potentials in this work used the form

$$U_{\text{dihed}}(\phi) = \sum_{n=0}^{m} k_n \cos^n(\phi), \tag{S2}$$

Fig. S2 Fit to the NDI–Th bond potential. Constrained optimisations were performed at each of the points shown, at 0.1 Å intervals from the optimised bond length of 1.468 Å. Dashed line indicates the fit to the form $U_{bond} = k_b(l - l_0)^2$, where l_0 is the value of the bond length in the unconstrained optimisation. All data with $U_{bond} < 2$ kcal/mol were used for the fit.

where ϕ is the dihedral angle and k_n are fit parameters. In LAMMPS, this potential is implemented as the "multi/harmonic" dihedral style for $m = 4$. For the Th–Th and NDI–NDI dihedrals we used $m = 4$ and 8, respectively. For $m = 4$, the dihedral form in eqn (S2) is equivalent to the OPLS form of the dihedral potential,

$$U_{dihed}^{OPLS}(\phi) = \frac{1}{2}k_1\left[1 + \cos(\phi)\right] + \frac{1}{2}k_2\left[1 - \cos(2\phi)\right] + \frac{1}{2}k_3\left[1 + \cos(3\phi)\right] + \frac{1}{2}k_4\left[1 - \cos(4\phi)\right], \qquad (S3)$$

where k_n are fit parameters, but is more flexible when $m \geq 5$. The more flexible form of eqn (S2) with $m = 8$ was required to accurately capture the height of the barrier at $\phi = 0$ for the NDI–NDI dihedral, but $m = 4$ was found to be adequate for the NDI–Th dihedral (see Fig. S3).
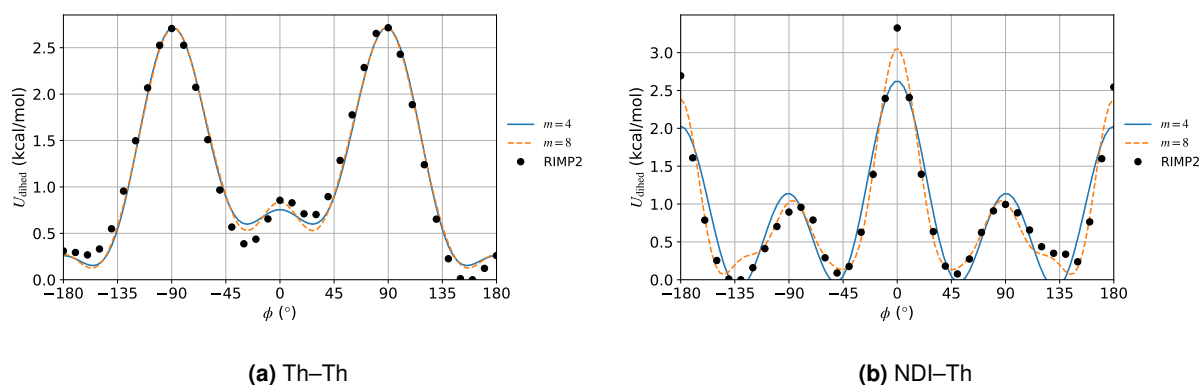
(a) Th–Th

(b) NDI–Th

Fig. S3 Comparison of the dihedral potential in eqn (S2) with $m = 4$ or $m = 8$ for fitting to the (a) Th–Th and (b) NDI–Th dihedrals. The points labelled RIMP2 are the energies obtained from quantum calculations. For the final parameters we used the fit parameters with $m = 4$ for the Th–Th potential, and $m = 8$ for the NDI–Th.

For both the NDI–Th and Th–Th dihedrals, 4 different combinations of atoms define the same dihedral due to the connectivity of the conjugated units. Rather than parameterise four interdependent dihedral potentials between each unit, we have set the parameters of three of these to zero, and parameterised the one remaining, effectively capturing the effect of all four possible dihedrals in one set of parameters. The definition of the two dihedrals, and the atoms used for their parameterisation, is shown in Fig. S4.

In order to calculate the dihedral potentials, the following procedure was followed, as outlined in ref. 1:

1. Geometry was optimised (B3LYP/6-31+G**) with the dihedral of interest fixed at 10° increments from −180 to 180°. Each optimisation started from the optimised geometry of the immediately preceding calculation, leading to structures getting trapped in geometries that were not the minimum energy for the specified dihedral angle, particularly around 0 and 180°, giving a non-symmetric potential. To obtain
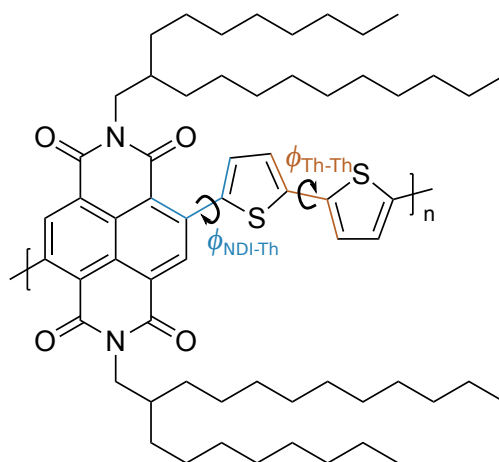
Fig. S4 The structure of a P(NDI2OD-T2) monomer unit, with the atoms used to define the NDI–Th (blue) and Th–Th (orange) dihedrals highlighted.

a symmetric potential, additional optimisations were carried out in the opposite direction from -160 to $-180°$, and from 20 to $0°$, both at $5°$ intervals, and the energies of these structures used.

2. RIMP2/cc-pVTZ single-point energy calculations on the optimised geometries with the single dihedral fixed were used to construct the MP2 dihedral potential, which was fit to eqn (S2) to give the five (or nine) coefficients k_n. The results of these calculations, and the fits to the obtained potentials are shown in Fig. S3.

3. The potential energy as a function of dihedral angle in a molecular dynamics (MD) system, including other bonded and non-bonded interactions, was obtained and compared to the previously obtained RIMP2 potential. As the energy as a function of dihedral angle in the MD simulation will include contributions from non-bonded interactions (in the OPLS force field, the 1–4 interactions, being those between atoms separated by three bonds, are set to $0.5\times$ their full value, which has been done in this work), the dihedral potential obtained from the quantum calculations cannot be used as is. The relevant dihedral potential for accurately modelling the behaviour in the MD simulation is the difference between the RIMP2 potential, and that obtained from an equivalent MD system, again calculated with the dihedral constrained at various positions over the whole $360°$ interval. For the NDI–Th dihedral, an energy minimisation at each constrained dihedral sometimes resulted in the structure becoming trapped in a local minimum (similarly to in the geometry optimisations outlined in step 1), so the MD potential was calculated using the following cooling protocol:

   (a) Dihedral of interest restrained to ϕ_0 at $10°$ intervals between 0 and $360°$ with a restraining potential of the form

   $$U_{\text{rest}} = k_{\text{rest}}\left[1 + \cos(n\phi - (\phi_0))\right], \tag{S4}$$

   with $k_{\text{rest}} = 5000$ kcal/mol and $n = 1$.

   (b) The system was cooled from 100 K to 0 K, with the temperature controlled using a Langevin thermostat ($\gamma = 10$ ps^{-1}, see eqn (1)) over 100 ps.

   (c) The simulation was run for an additional 10 ps at 0 K, then the energy minimised.

   (d) This final minimised energy at each value of the dihedral constraint was used to construct the potential energy in the same way as for Fig. S3.

   For the Th–Th dihedral, the energy with the dihedral angle restrained as outlined above was simply calculated from an energy minimisation of the structure rather than following the cooling procedure.

4. The potential obtained from the MD energy minimisation procedure was then fit to the same form as that from the quantum calculations (eqn (S2) with $m = 4$ or 8). The difference in the coefficients from the quantum fit and the MD fit then gave the final coefficients for the potential in the MD force field. A comparison between the dihedral potentials calculated from the quantum calculation and the final MD force field parameterisation is shown in Fig. S5.

(a) Th–Th
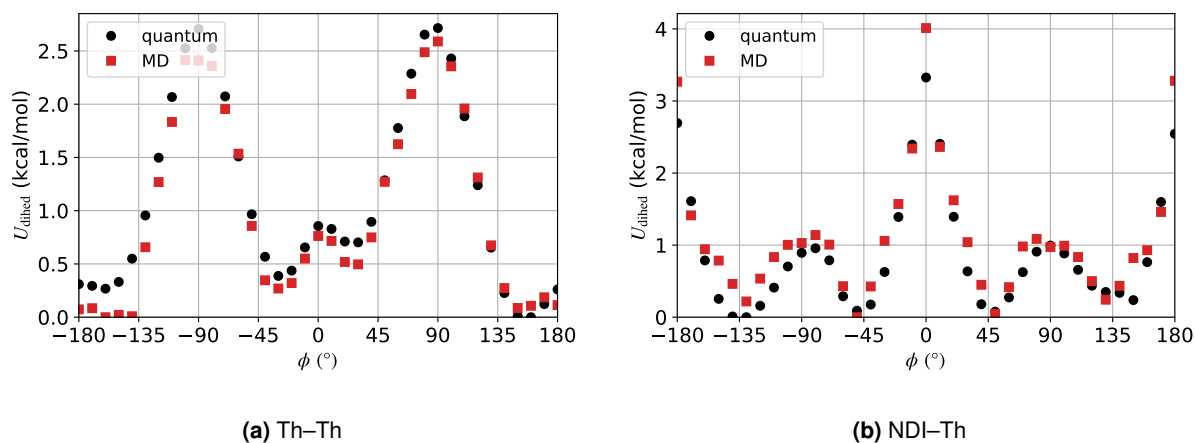
(b) NDI–Th

Fig. S5 Comparison of dihedral potentials calculated from quantum calculations (black circles) or MD simulation with the final force-field parameters (red squares) for (a) the Th–Th torsion, or (b) the NDI–Th torsion. Both show reasonable agreement between the two methods.

S2 All-atom simulation methods

To initialise the AA MD simulations of P(NDI2OD-T2) monomers in *o*-dichlorobenzene (DCB) used to parameterise the coarse-grained (CG) non-bonded parameters, the particles were packed, using Packmol,[9] with random positions and orientations into a cubic simulation box at low density ($\approx$0.5 g/cm^3, box size of (120 Å)^3) to prevent overlapping particles or interlocking rings. The box was shrunk over 40 ps to higher than the target density ($\approx$1.7 g/cm^3, box size (76 Å)^3) at constant energy, then expanded over an additional 20 ps to give approximately the expected experimental solvent density ($\approx$1.3 g/cm^3, box size (82.5 Å)^3). The energy of the system was then minimised, and velocities assigned from a Boltzmann distribution at 300 K. The simulation was then run for 200 ns at constant temperature (300 K) and pressure (1 atm), with the temperature and pressure controlled using a Nosé-Hoover thermostat (relaxation time scale = 100 fs)[10,11] and barostat (relaxation time scale = 1000 fs),[12] respectively. The equilibration time was determined from the convergence of the monomer–monomer centre-of-mass radial distribution function (RDF) (see Fig. S7), with the first 50 ns being discarded as the equilibration period.

The AA simulations of P(NDI2OD-T2) trimers in DCB used to parameterise the CG bonded parameters were set up similarly to the monomer simulations, with 8 trimers at the same concentration ($\approx$55 g/L, 3229 DCB molecules). The same simulation procedure as for the monomer simulations was followed, with initial (low density), intermediate (high density), and final (experimental solvent density) box sizes of (140 Å)^3 ($\approx$0.3 g/cm^3), (80 Å)^3 ($\approx$1.6 g/cm^3), and (85 Å)^3 ($\approx$1.3 g/cm^3), respectively. The system was simulated for 90 ns at constant temperature (300 K) and pressure (1 atm), again with the first 50 ns discarded as the equilibration period, based on the convergence of the bonded distributions (see Fig. S8).

S3   Coarse-graining of P(NDI2OD-T2)

All CG simulations in this work used an implicit solvent, modelled using Langevin dynamics. The friction coefficient γ (see eqn (1) of the main text) was chosen to match the mean squared displacement (MSD) (and thus the diffusion coefficient) of monomers in the AA (in DCB) and CG representations, shown in Fig. S6.
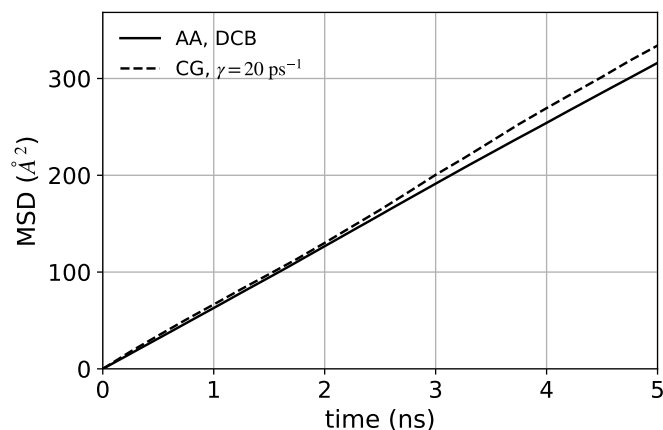
Fig. S6 MSD of P(NDI2OD-T2) monomers in the AA simulation (solid line) and CG simulation with friction coefficient set to $\gamma = 20$ ps^{-1} (dashed line) to match the diffusion coefficient in DCB.

The non-bonded interactions in the CG model of P(NDI2OD-T2) were parameterised from AA simulations of monomers in DCB. The AA simulations were deemed to be at equilibrium by 50 ns of the 200 ns simulation based on the time dependence of the monomer center-of-mass RDF (Fig. S7).
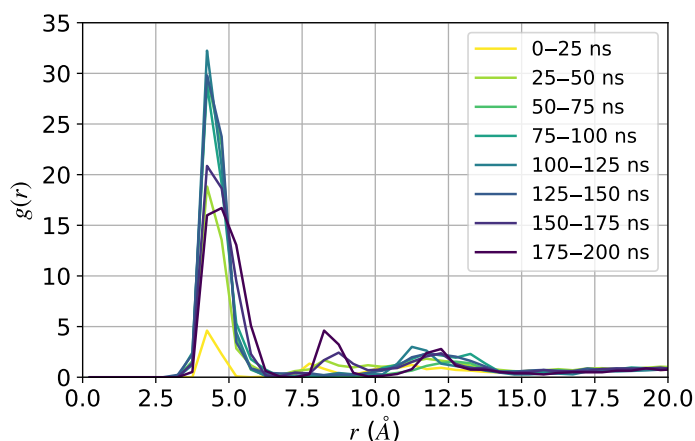
Fig. S7 Monomer center-of-mass RDF for P(NDI2OD-T2) monomers in DCB averaged over various time slices of the simulation.

The bonded interactions in the CG model of P(NDI2OD-T2) were parameterised from AA simulations of trimers, again in DCB. The equilibration period of these simulations was estimated from the time dependence of the bonded distributions. While most converged within 1 ns, the time dependence of the slowest varying distribution – the 3–1–7–3 dihedral – is shown below (Fig. S8). A 50 ns equilibration time was used based on the time dependence of this dihedral distribution.

In all cases, simulations were initialised by packing the desired monomers/oligomers into the simulation box with random positions and orientations (in the case of oligomers, each chain was a linear arrangement of monomers in which the backbone units were coplanar), and a soft-core potential of the form

$$U_{\text{soft}}(r) = A \left[1 + \cos \frac{\pi r}{r_c} \right], \, r < r_c, \tag{S5}$$

where r_c is a cutoff distance, r the distance between two particles, and A an energy pre-factor that controls the "hardness" of the potential, was applied to remove any overlaps between particles ($r_c = 6$ Å, A increased linearly from 0 to 30 kcal/mol over 5000 fs, timestep = 1 fs).
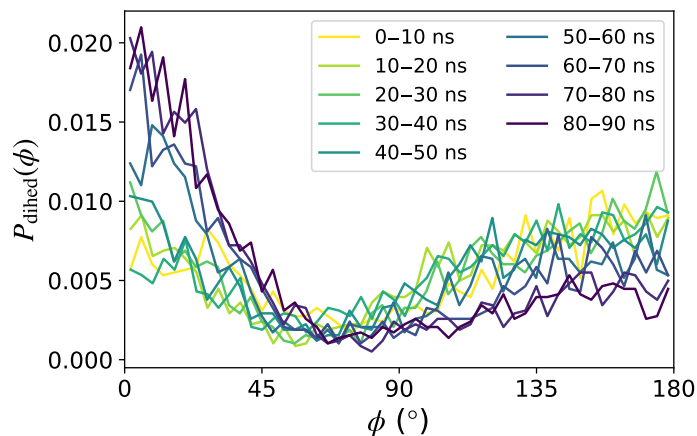
Fig. S8 Distribution of the 3–1–7–3 (site types defined in Fig. S29) dihedral angle in DCB, averaged over various time slices of the simulation. Each line is the distribution calculated over the indicated 10 ns block.

The fit of the analytical function used for each CG potential (see Section S3.1) to the Boltzmann inversion of the target distribution was used as the initial guess for all parameters, with the value of the Lennard-Jones (LJ) energy parameter ϵ_{LJ} in each of the non-bonded interactions constrained to be initially $0.1 \leq \epsilon_{LJ} < 1$ kcal/mol in order to prevent extensive aggregation. The constraint on ϵ_{LJ} was removed for the iterative procedure. Bonded interactions were optimised first, by comparing the target distributions from the AA trimer simulations (eight P(NDI2OD-T2) trimers at $\approx$55 g/L) described above with distributions from an equivalent CG system of 8 trimers in an (87 Å)^3 box (approximately the average volume in equivalent all-atom simulations). Note that not every possible bond/angle/dihedral/improper was parameterised or included in the CG model as many are fully defined by other parameters (e.g. the 3–6–3 angle was not parameterised because it is effectively fixed by the 3–6 and 3–3 bonds; see Fig. 1 for site definitions). These un-parameterised CG distributions are compared with the AA distributions for the fully parameterised model in Section S3.3. The non-bonded parameters were optimised using simulations matched to the AA system of 18 monomers at $\approx$55 g/L. These CG systems contained 18 monomers in an (84 Å)^3 box, again approximately the average volume in equivalent all-atom simulations. The discrepancy between the CG and AA RDFs at iteration n was quantified as[13]

$$b_n = \sum_{i,j} f_{ij} N_{ij}, \tag{S6}$$

where

$$f_{ij} = \frac{\int_0^{r_b} \left[g_{ij}(r) - g_{ij}^{ref}(r) \right]^2 \mathrm{d}r}{\int_0^{r_b} \left[g_{ij}^{ref}(r) \right]^2 \mathrm{d}r}, \tag{S7}$$

where $g_{ij}(r)$ is the CG RDF for site types i and j at iteration n, N_{ij} the number of ij pairs in a monomer, $g_{ij}^{ref}(r)$ the target AA RDF for each ij pair, and r_b the cutoff for the non-bonded interactions (here 15 Å). The variable b_n was found to converge after $n = 12$ iterations, with further iterations not improving the agreement between AA and CG distributions further (see Figs. S9 and S10). The parameters from iteration 12 were therefore used for further CG simulations.

S3.1 Analytical forms of bonded and non-bonded coarse-grained potentials

The following analytical functions were used for the CG potentials to facilitate simulation. See Fig. 1 for site type definitions.

Bonds: For the bonds that did not contain side-chains atoms, a harmonic potential (eqn (S8a)) was used. For those that did (site-types 4–5 and 5–5), a quartic potential was used (implemented in LAMMPS as the "class2" bond style).

$$U_{\text{bond}}^{\text{harm}}(l) = k_2(l - l_0)^2 \tag{S8a}$$

$$U_{\text{bond}}^{\text{quart}}(l) = k_2(l - l_0)^2 + k_3(l - l_0)^3 + k_4(l - l_0)^4 \tag{S8b}$$

In both cases k_i are the bond stretching coefficients, and l_0 is the equilibrium bond length between two bonded sites with bond length l.

Angles: Angles 3–2–4 and 3–6–4 were modelled with a harmonic angle potential (eqn (S9a)). The others used a quartic potential (eqn (S9b)).

$$U_{\text{angle}}^{\text{harm}}(\theta) = k_2(\theta - \theta_0)^2 \tag{S9a}$$

$$U_{\text{angle}}^{\text{quart}}(\theta) = k_2(\theta - \theta_0)^2 + k_3(\theta - \theta_0)^3 + k_4(\theta - \theta_0)^4 \tag{S9b}$$

k_i are the angle bending coefficients, and θ_0 is the equilibrium angle between three bonded sites at angle θ.

Dihedrals: All dihedrals were modelled with the potential given by eqn (S10):

$$U_{\text{dihed}}(\phi) = \sum_{n=0}^{m} k_n \cos^n(\phi), \tag{S10}$$

with $m = 4$, where k_n are fit parameters and ϕ is the dihedral angle between four bonded sites. Note this is the same as eqn (S2) used for the AA simulations.

Improper dihedrals: A cosine improper dihedral potential was used (eqn (S11), implemented in LAMMPS as the "cvff" improper style):

$$U_{\text{impro}}(\psi) = k\left[1 + d\cos(n\psi)\right], \tag{S11}$$

where k is an energy parameter, $d = 1$ or -1 and n is the multiplicity (an integer). Here we use $d = -1$ and $n = 2$.

Non-bonded interactions: Non-bonded interactions were fit to a LJ potential (eqn (S12)), with interactions cut off and shifted to zero at $r_c = 15$ Å:

$$U_{\text{LJ}} = \begin{cases} 4\epsilon_{\text{LJ}} \left[\left(\frac{\sigma_{\text{LJ}}}{r}\right)^{12} - \left(\frac{\sigma_{\text{LJ}}}{r}\right)^{6} \right], & r < r_c, \\ 0, & \text{otherwise,} \end{cases} \tag{S12}$$

where σ_{LJ} and ϵ_{LJ} are the interaction diameter and strength, respectively. A purely repulsive form of this potential was also used in order to relax the backbone conformation prior to simulation with attractive non-bonded interactions. The Weeks–Chandler–Andersen (WCA) potential was used:

$$U_{\text{WCA}} = \begin{cases} 4\epsilon_{\text{LJ}} \left[\left(\frac{\sigma_{\text{LJ}}}{r}\right)^{12} - \left(\frac{\sigma_{\text{LJ}}}{r}\right)^{6} \right] + \epsilon_{\text{LJ}}, & r < 2^{1/6}\sigma_{\text{LJ}}, \\ 0, & \text{otherwise,} \end{cases} \tag{S13}$$

where all symbols are as defined for the LJ potential. The potential is shifted to zero at the cutoff of $2^{1/6}\sigma_{\text{LJ}}$ (the position of the LJ energy minimum) to give a purely repulsive potential.

S3.2 Convergence of non-bonded parameters

Non-bonded parameters were optimised following the iterative Boltzmann inversion (IBI) protocol described in refs 14 and 15. The convergence of the RDFs was monitored using eqn (S6), as shown in Fig. S9. The parameters from iteration 12 (highlighted in orange in Fig. S9) were used as the final P(NDI2OD-T2) CG parameters. The behaviour of the RDFs, and corresponding LJ potentials, for the 5 iterations leading up to iteration 12 are shown in Fig. S10.
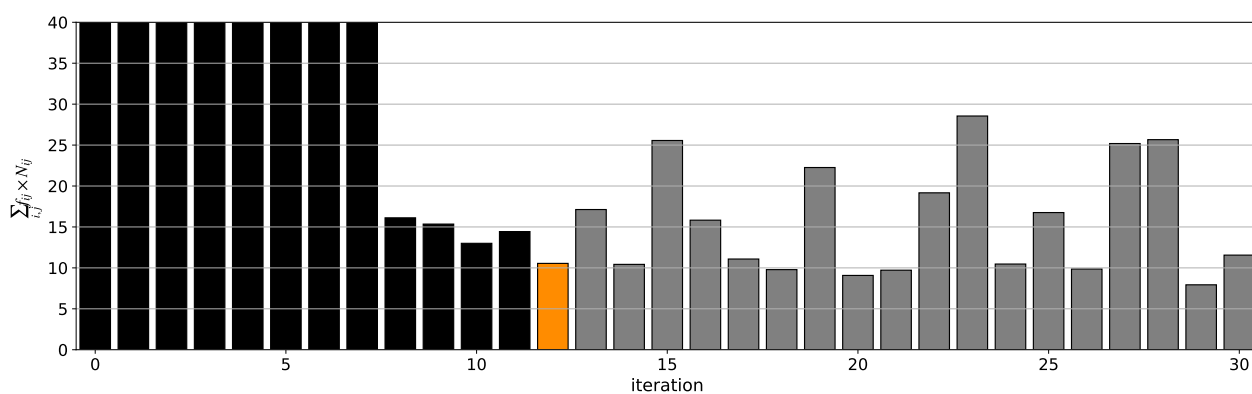
Fig. S9 Convergence of the IBI procedure for the coarse-graining of the P(NDI2OD-T2) non-bonded interactions from simulations of monomers in DCB. The orange bar indicates the iteration that was used for further simulation, as all subsequent iterations showed little to no improvement.
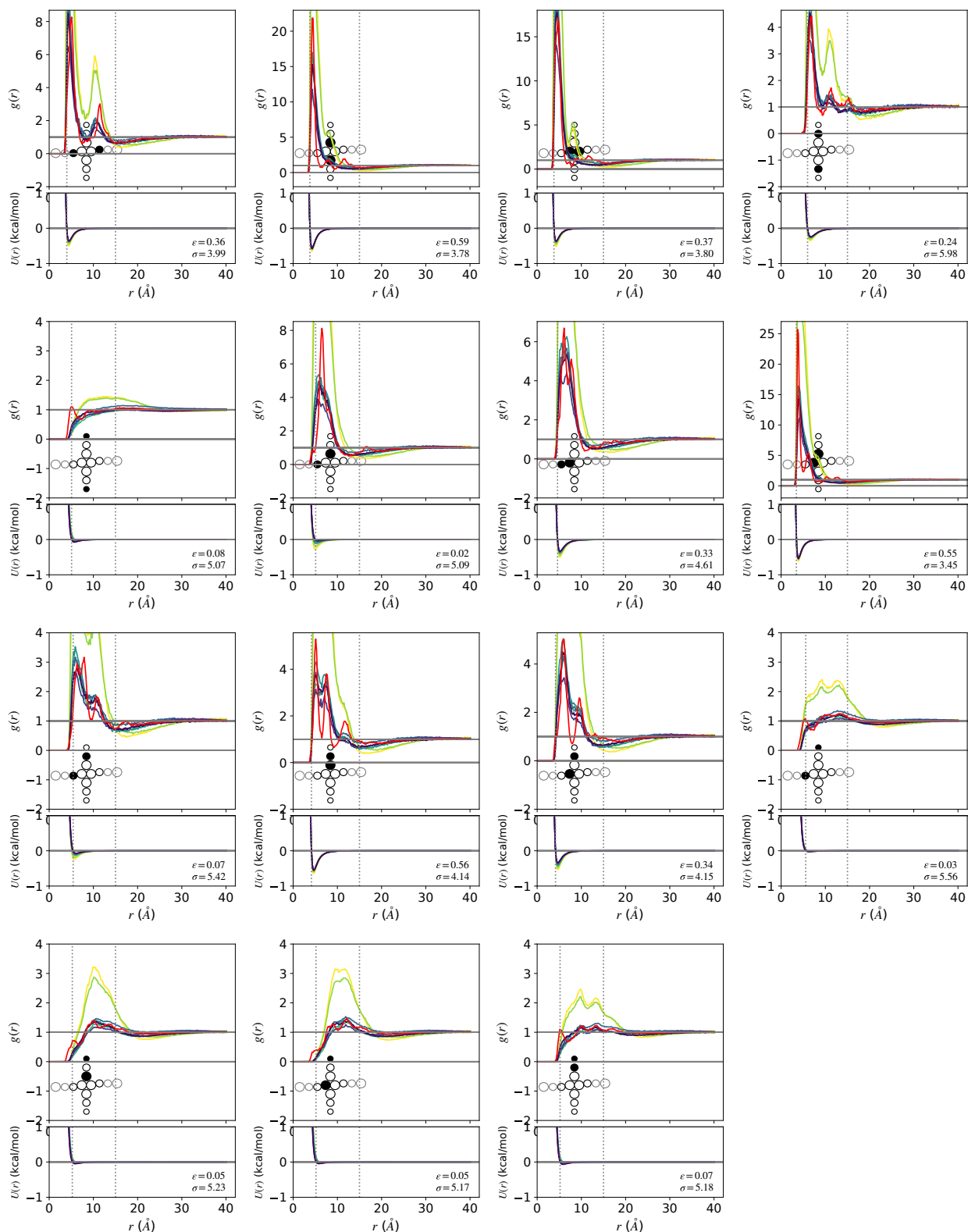
Fig. S10 Comparison of the target AA (red) and CG site–site RDFs for the final 5 iterations prior to the one that was used (coloured by iteration with yellow the earliest and purple the final iteration, number 12). The inset diagrams indicates the site types for which the distribution was calculated (for sites on different monomers). The LJ potentials that resulted in the plotted distributions are shown in the lower plots, with the parameters of the final iteration (12) listed. The plots of the LJ potentials follow the same colour mapping as the RDFs.

S3.3   Comparison of coarse-grained and all-atom distributions

Following the parameterisation of both the bonded and non-bonded interactions, all the bonded distributions were recalculated with the final non-bonded interactions to determine whether the parameterisation of the non-bonded interactions affected the previously obtained bonded distributions. Distributions between additional bonded sites, which were not explicitly parameterised as their behaviour should be captured by the combination of other parameters, were also examined to determine whether they behaved as expected. The simulations used to generate these were set up identically to the CG trimer simulations used to parameterise the bonded interactions (18 trimers in (87 Å)^3 box). The same soft potential as for the previous simulations was used to remove overlaps, and the simulations run for a total of 100 ns (timestep = 5 fs), with the bonded distributions calculated from the final 50 ns. In all cases, good agreement was found between the AA (solid grey line) and CG (dotted black line) distributions, even for the potentials that were not explicitly parameterised.
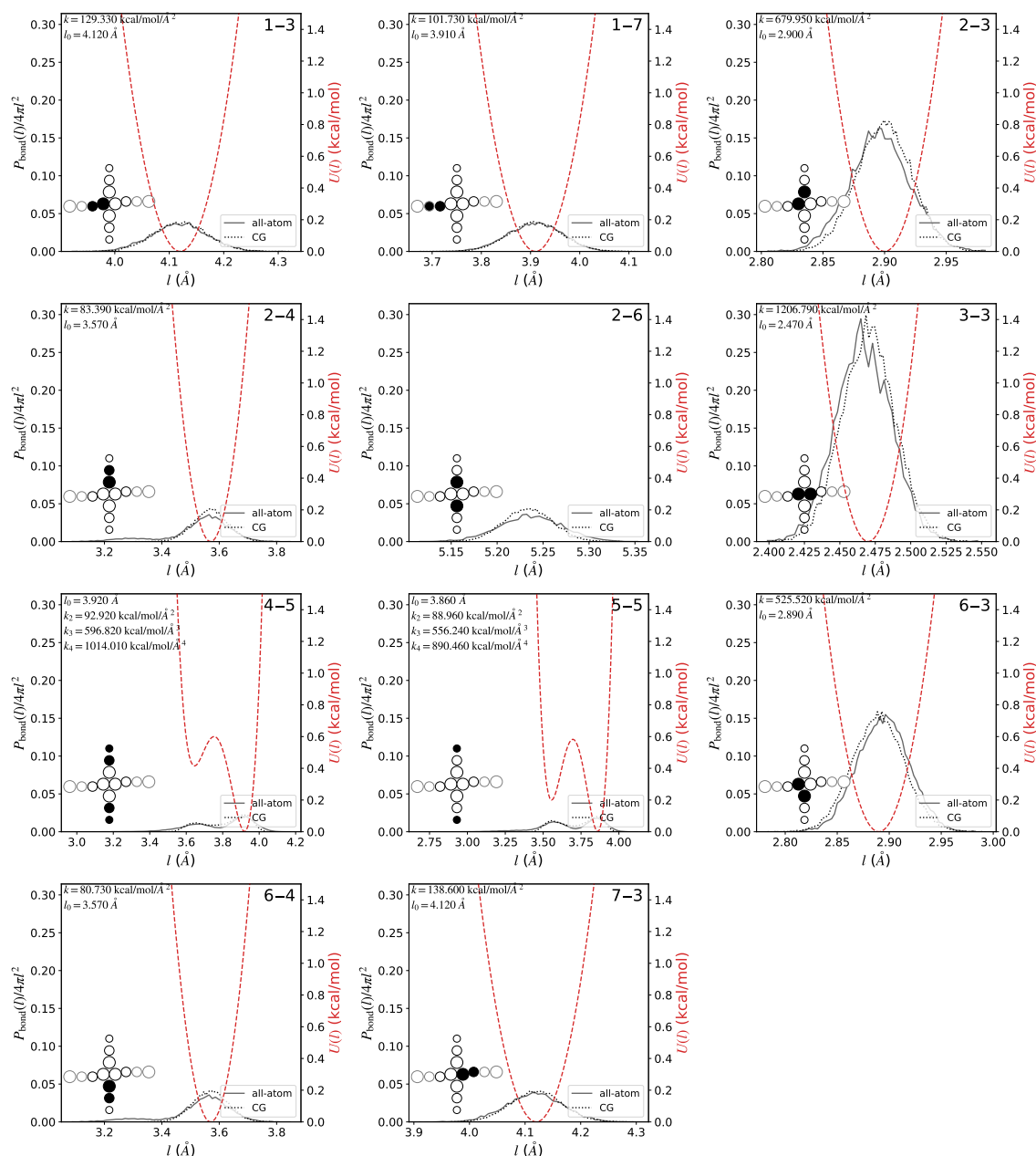
Fig. S11 Final bond distributions calculated from simulations with the full set of optimised parameters (including non-bonded). Where the bond length was explicitly parameterised, the final potential is shown in red, and its parameters given in the top left corner of each plot. An inset diagram shows a representation of P(NDI2OD-T2) with the pair of atoms defining the bond highlighted.
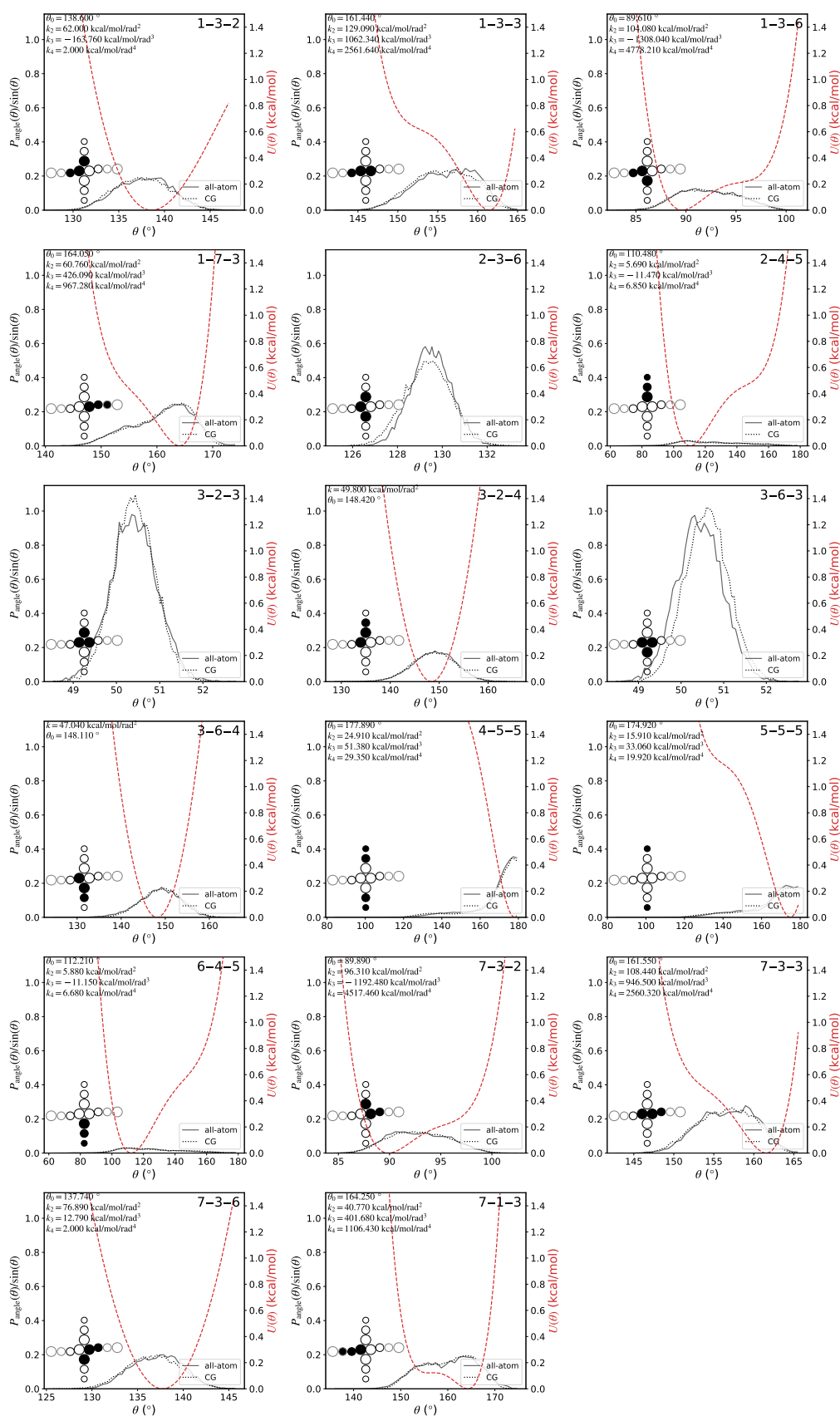
Fig. S12 Final angle distributions calculated from simulations with the full set of optimised parameters (including non-bonded). Where the angle was explicitly parameterised, the final potential is shown in red, and its parameters given in the top left corner of each plot. An inset diagram shows a representation of P(NDI2OD-T2) with the trio of atoms defining the angle highlighted.
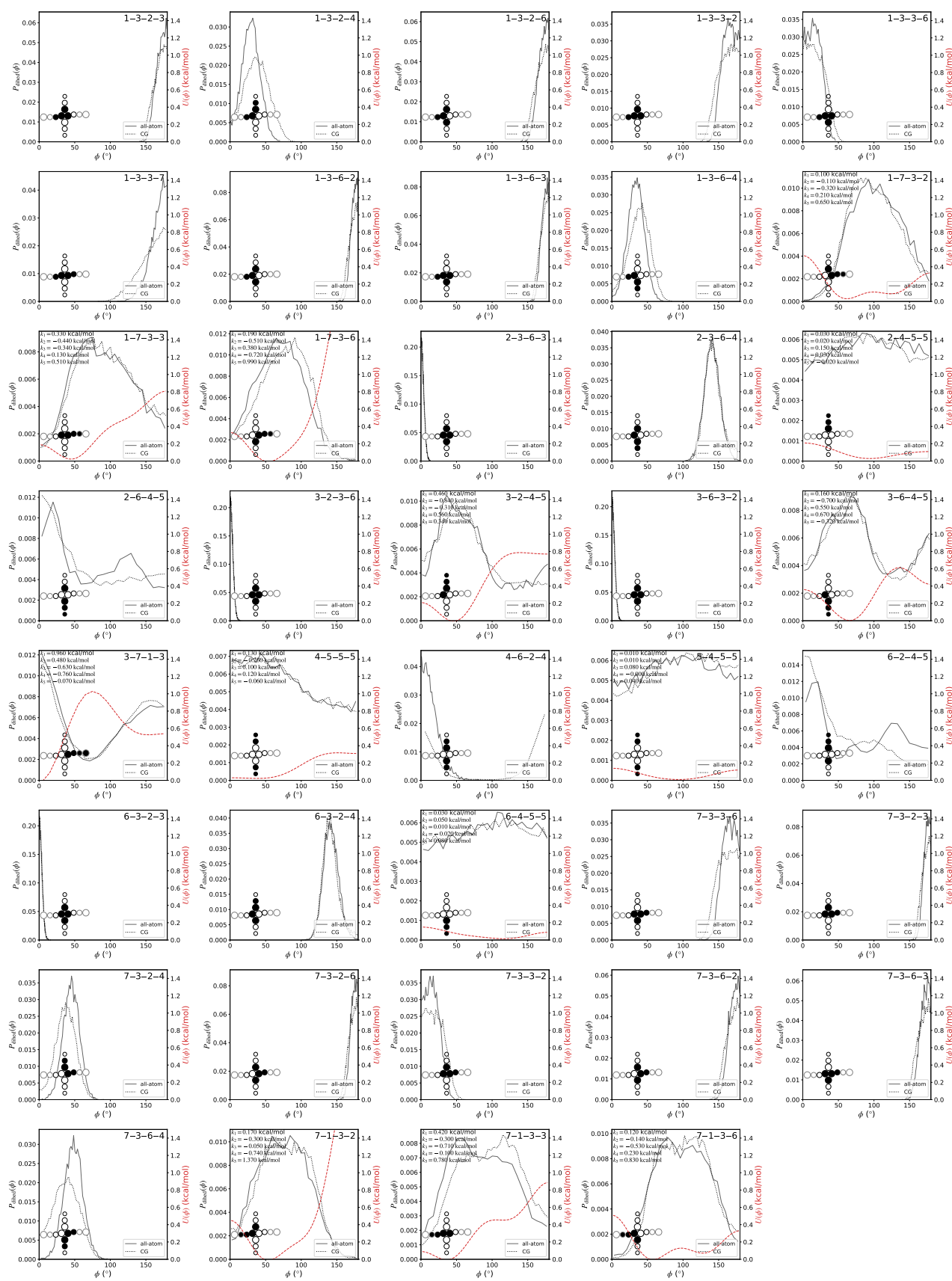
Fig. S13 Final dihedral distributions calculated from simulations with the full set of optimised parameters (including non-bonded). Where the dihedral was explicitly parameterised, the final potential is shown in red, and its parameters given in the top left corner of each plot. An inset diagram shows a representation of P(NDI2OD-T2) with the four atoms defining the dihedral highlighted.
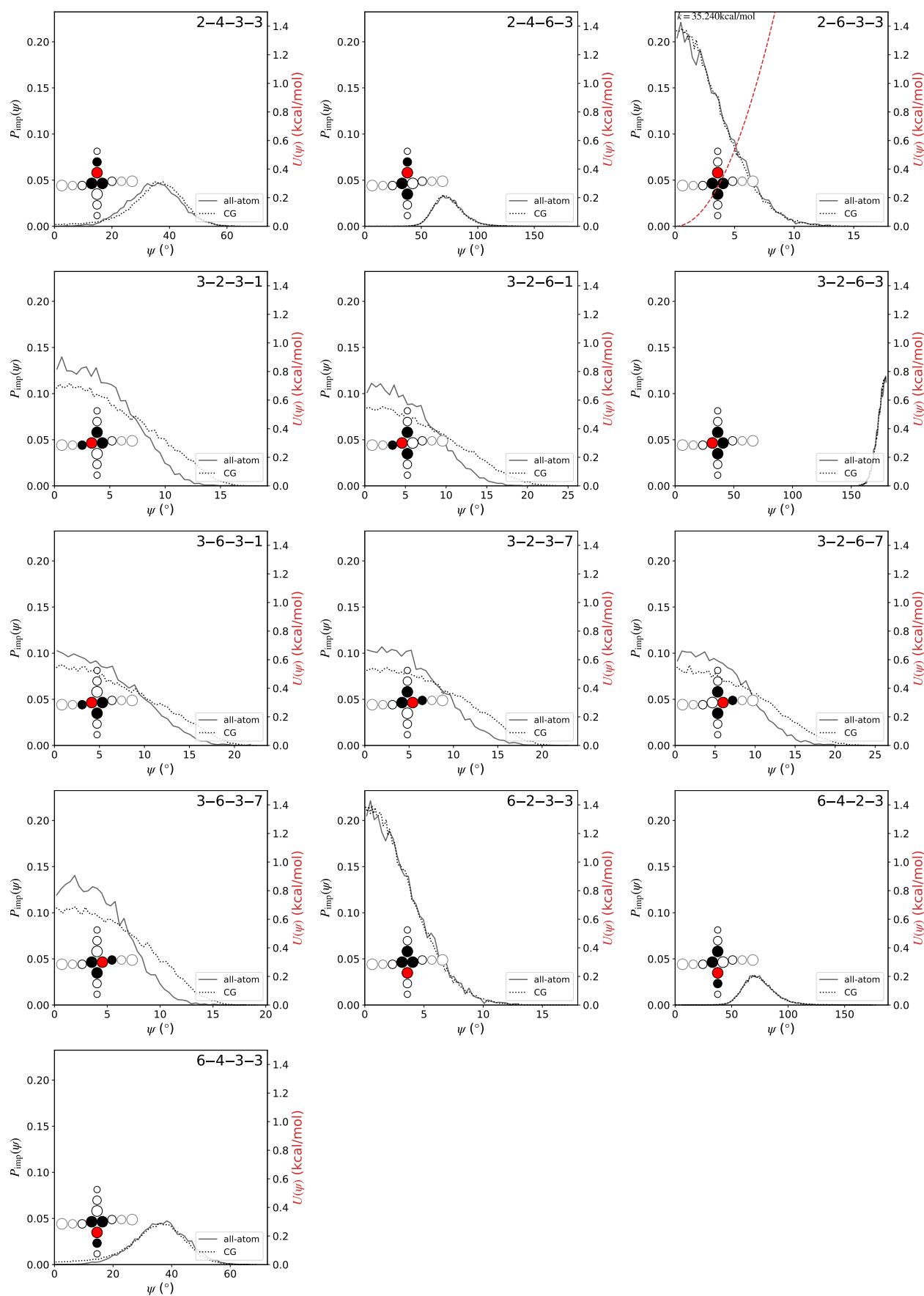
Fig. S14 Final improper distributions calculated from simulations with the full set of optimised parameters (including non-bonded). Where the improper was explicitly parameterised, the final potential is shown in red, and its parameters given in the top left corner of the plot. An inset diagram shows a representation of P(NDI2OD-T2) with the four sites defining the improper highlighted (central site highlighted in red).

S4   OPES simulations

S4.1   Reweighting procedure

Free-energy calculations for the AA and CG monomer systems were used to estimate the relative strength of the interactions between monomers in different solvents. These calculations used the on-the-fly probability enhanced sampling (OPES) procedure.[16] As OPES biases the behaviour of the monomers, the free energy cannot simply be calculated from the probability distribution in these simulations, but must be reweighted to properly account for the effect of this bias. The sampled distribution $P_{\text{bias}}(s)$ is related to the unbiased equilibrium distribution $P(s)$, where s is the collective variable(s) that is (are) being biased (here the center-of-mass separation) through the addition of a bias $V(s)$, where

$$V(s) = k_{\text{B}} T \ln \left(\frac{P_{\text{bias}}(s)}{P(s)} \right). \tag{S14}$$

The bias introduced by the harmonic wall that restrains the center-of-mass separation of the monomers to be less than 20 Å was also accounted for in this manner. The weights to return to the unbiased distribution are then simply $w = \exp \left(\frac{V(s)}{k_{\text{B}} T} \right)$. This can be used to calculate an estimate of unbiased distribution, which in PLUMED[17,18] is done using kernel density estimation, as

$$\langle \hat{P}(s,t) \rangle = \frac{\sum_{t'=0}^{t} w(t') K(s - s(t'), \sigma_{\text{K}})}{\sum_{t'=0}^{t} w(t')}, \tag{S15}$$

where $w(t')$ is the weight at time t' ($w(t') = \exp \left(V(s,t')/k_{\text{B}} T \right)$), and $K(s - s(t'), \sigma_{\text{K}})$ are kernels (here Gaussians), centered at $s(t)$ with bandwidth σ_{K} (here 0.25 Å) and summed over all times between t' and t. This procedure is implemented in PLUMED as `reweight_bias`, combined with the `histogram` functionality to calculate weighted probability densities.[17,18]

S4.2   All-atom free-energy calculations

For all three solvents, a pair of P(NDI2OD-T2) monomers in the π-stacked configuration, with centre-of-mass (of NDI groups) separation of 3 Å, was placed in the centre of a $(100 \text{ Å})^3$ simulation box, and 1000 solvent molecules packed with random position and orientation around it[9] to give a low-density system in which solvent molecules did not overlap or interlock. As with the AA simulations described previously, the system was shrunk over 40 ps to a higher density (box size $(55 \text{ Å})^3$), then expanded over a further 20 ps to give a $(60.8 \text{ Å})^3$ box. The system was then simulated at constant temperature (300 K) and pressure (1 atm) for 5 ns to properly solvate the monomers and to reach the appropriate density. As OPES generally converges faster if it starts near the expected free energy minimum, we placed a harmonic restraining potential of the form

$$U_{\text{w}}(r) = \begin{cases} k_{\text{w}} (r - r_{\text{w}})^2, & r > r_{\text{w}}, \\ 0, & \text{otherwise}, \end{cases} \tag{S16}$$

on the monomer–monomer centre-of-mass separation r, which acts to keep the separation below r_{w}. Here we set r_{w} to 10 Å and k_{w} to 100 kcal/mol/Å² to ensure the monomers stayed close to the expected free energy minimum of the directly π-stacked structure. After 5 ns, this harmonic wall was removed and the OPES simulation begun from the final configuration of this setup procedure. Three collective variables were biased in this procedure: the centre-of-mass separation (where the centre-of-mass was calculated as the centre-of-mass of the NDI group), the angle between the vectors normal to the planes of each NDI group (angle 1), and the angle between this vector of one monomer and the vector connecting the centres-of-mass of the two monomers (angle 2). For this work, we were interested solely in the free-energy as a function of the monomer separation rather than both separation and orientation, but we note that biasing the angles, which may be expected to be slowly varying collective variables, aids in achieving faster convergence. The OPES simulation was conducted with a kernel deposition frequency of 5000 timesteps, where kernels were initially of width 0.2 Å for the distance constraint, and 0.06 and 0.12 radians, respectively, for angles 1 and 2. These initial kernel widths were chosen to be approximately the standard deviation of the unbiased variable at equilibrium. A bias factor, which is used to determine the shape of the target probability distribution,[16] of 10 and an estimated free energy barrier of 30 kcal/mol were used. Additionally, to prevent the monomers becoming too separated and sampling large regions of unimportant (large separation) space, we placed a harmonic wall (eqn (S16)) at a centre-of-mass separation of 20 Å, again with force constant 100 kcal/mol/Å². OPES simulations were then run for approx 1.6 µs in 1-chloronaphthalene (CN), 1.7 µs in DCB, and 1.2 µs in toluene (TOL).

S4.3   Coarse-grained free-energy calculations

Similar OPES simulations were conducted at constant volume and temperature (300 K) for the three sets of CG parameters in order to validate the choice of parameters for the "good" and poor solvents, and confirm

the agreement between the AA and CG models for the system parameterised in DCB. These simulations were initialised by placing two monomers in the centre of a $(60 \text{ Å})^3$ simulation box (approximately the size of the simulation box for the AA free energy calculations) again in the directly π-stacked configuration, with centre-of-mass separation of 5 Å. A soft-core potential (identical to that used previously) was applied to remove particle overlaps, before switching to the LJ parameters corresponding to the "good", intermediate (DCB), or poor solvents. The energy was minimised, then OPES run for 2 μs with a timestep of 8 fs. As with the AA calculations, the centre-of-mass separation was defined in terms of the centre-of-mass of the NDI groups (site types 2, 3, and 6 in the CG model), and a repulsive harmonic wall (eqn (S16), $k_w = 100$ kcal/mol/Å^2) used to keep the monomer centres-of-mass within 20 Å of each other. All of the OPES and reweighting parameters were the same as for the AA equivalent, although in this case, only the separation (not the angles) was biased.

S4.4 Comparison of AA and CG OPES results

To validate the choice of scaled CG non-bonded energy parameters, where the interactions corresponding to P(NDI2OD-T2) in a better and poorer solvent were obtained by uniform scaling of the parameters from DCB, we compared the free energy as a function of center-of-mass separation of the three sets of CG parameters used in this work with the AA equivalents in TOL (a poor solvent for P(NDI2OD-T2), and one in which rod-like aggregates are observed experimentally), *o*-dichlorobenzene (DCB) (the solvent in which intermediate solvent interactions were parameterised), and CN (a good solvent for P(NDI2OD-T2)). In all three cases, the scaled parameters give reasonable agreement between the CG and AA free energies. Importantly, the poor solvent parameters are a good representation of the TOL environment in which the formation of extended rod-like aggregates is expected.
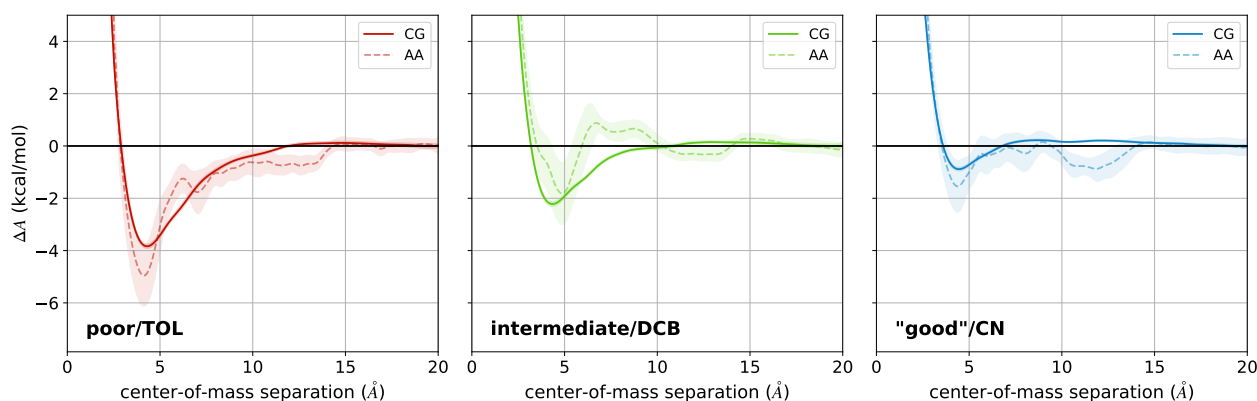
Fig. S15 Free energy as a function of center-of-mass separation for the interaction between two monomers of P(NDI2OD-T2) in either the AA (dashed lines) or the CG (solid lines) representation. Error bars (shaded) indicate two standard errors, calculated by block averaging (they are approximately the thickness of the line for the CG curves). The poor solvent parameters were obtained by scaling the parameters obtained from AA simulations in DCB to be 20% stronger, while the "good" solvent parameters are 20% weaker than the DCB parameterisation.

S5 Coarse-grained simulation methods

S5.1 Single-chain simulations

All systems were set up following the same procedure. First, the single (linear) chain was placed in the centre of the simulation box (side length of 420, 840, 1260, and 1680 Å for the 10, 20, 30, and 40 monomer chains, respectively; approximately three times the contour length of the polymer chain). A soft-core potential was applied to remove any overlaps between CG sites (following the same procedure as applied previously). The resulting system was simulated with a purely repulsive WCA potential (LJ potential truncated and shifted at its minimum, eqn (S13)) equivalent of the CG potential in order to allow the chain to relax away from the fully extended conformation to the equilibrium conformation given purely repulsive non-bonded interactions. The temperature was controlled using a Langevin thermostat. Simulation times with the WCA potentials for 10mers were 100 ns with the regular (matching DCB) Langevin friction coefficient, 500 ns for 20 and 30mers with the 10× lower friction coefficient, and 750 ns for the 40mers, with the same lower friction coefficient. This time was sufficient that the radius of gyration had plateaued. The non-bonded parameters were then switched back to the LJ parameters and the Langevin friction coefficient returned to the value parameterised for DCB. For each system, a number of independent simulations were carried out, with the simulation duration chosen for each system type in order to obtain an accurate estimate of the single-chain folding time, as summarised in Table S1.

Table S1 Number of independent simulations (n_{sim}) and duration of each simulation (t_{sim}) for single-chain simulations for systems with different solvent quality, chain backbone flexibility, chain length (N), and Langevin dynamics friction coefficient (γ).

solvent	flexibility	N	γ (ps^{-1})	n_{sim}	t_{sim} (µs)
poor	regular	10	20	80	25
poor	regular	20	20	80	20
poor	regular	30	20	80	15
poor	regular	40	20	100	13.5
poor	regular	20	2	20	5
poor	flexible	10	20	20	10
poor	flexible	20	20	20	10
poor	flexible	30	20	20	10
"good"	regular	10	20	20	10
"good"	regular	20	20	20	10
"good"	regular	30	20	20	10
"good"	flexible	10	20	20	10
"good"	flexible	20	20	20	10
"good"	flexible	30	20	20	10

S5.2 Multi-chain simulations

A variety of systems with different chain lengths N, concentration, and flexibility were used to study the effect of these properties on the aggregation behaviour of P(NDI2OD-T2). Table S2 summarises the conditions of each simulation, including the number of chains and box size used to achieve the target concentration, as well as the calculated aggregation time, τ_c, for each system. To examine the average behaviour of most systems, a number of systems were simulated starting from multiple independent starting configurations.

For all systems, the polymers were initially placed in the simulation box at random positions and orientations with a fully extended backbone. As the random packing was likely to lead to overlaps between chains, a soft-core potential (eqn (S5)) was applied to remove overlapping CG sites (same procedure and parameters as described previously). As with the single-chain simulations, the chain conformations were allowed to relax by initially simulating with purely repulsive non-bonded interactions (WCA equivalent of the CG LJ potentials). Simulation times with the WCA potentials for 10mers were 100 ns with the regular (matching DCB) Langevin friction coefficient, 500 ns for 20 and 30mers with the 10× lower friction coefficient, and 750 ns for 40mers also with the 10× lower friction coefficient, as described for the single-chain simulations. The non-bonded interactions were switched to the LJ potentials, the friction coefficient returned to the desired value (generally that parameterised to match DCB, but 10× lower in some cases), and the systems then simulated for 1–4.5 µs using Langevin dynamics to simulate an implicit solvent environment at a temperature of 300 K. This duration was sufficient that extensive aggregation was observed in the poorer solvents. In the poor solvent, the aggregate sizes were approaching the size of the box after approximately 3–4 µs, so further increasing the simulation time was unlikely to yield physically meaningful results without also increasing the system size.

Table S2 List of the multi-chain systems studied in this work. N is the length (number of monomers) of the polymer chain, γ is the Langevin dynamics friction coefficient, and t_{sim} is the duration of each of n_{sim} independent simulations started from different initial configurations. $\phi_{\text{V}} N^{1/2}$ for the different polymer concentrations C and number of chains, N_{c}, packed into a box with side length L_{box} are also included. The aggregation time τ_{c} for each system, measured as the time for the concentration of single chains to drop to 25% of the original concentration, is given in the final column. Where this value is "N/A", the single-chain concentration never dropped to below 25% of the original concentration.

N	solvent	C (g/L)	flexibility	γ (ps^{-1})	N_{c}	L_{box} (nm)	$\phi_{\text{V}} N^{1/2}$	n_{sim}	t_{sim} (μs)	τ_{c} (μs)
10	poor	8.5	regular	20	112	60	0.014	2	1.5	0.42
20	poor	6	regular	20	94	80	0.014	2	4	0.69
20	poor	6	flexible	20	94	80	0.014	2	3	N/A
20	poor	6	regular	2	94	80	0.014	2	1	0.06
20	intermediate	6	regular	20	94	80	0.014	2	3	1.34
20	"good"	6	regular	20	94	80	0.014	2	2	N/A
20	"good"	6	regular	2	94	80	0.014	2	2	N/A
20	"good"	6	flexible	20	94	80	0.014	2	1	N/A
30	poor	5	regular	20	101	100	0.014	3	1	0.79
40	poor	4	regular	20	104	120	0.014	3	3	1.65
10	poor	10	regular	20	39	40	0.016	3	2	0.32
10	poor	10	flexible	20	39	40	0.016	1	5	0.86
20	poor	12	regular	20	187	80	0.027	2	1	0.18
20	poor	2	flexible	20	31	80	0.0046	2	3	N/A

S6   Properties of the CG P(NDI2OD-T2) polymer chain

We quantified the flexibility of the polymer chains with regular and flexible backbones using the Kuhn length, b, calculated from simulations of single chains of length 10, 20, or 30 monomers under "good" solvent conditions. The flexible chains, where the bonded parameters that define the flexibility (see Section S14) were set to 1% of the original parameters, have a Kuhn length approximately 30–40% shorter than the regular backbone (Fig. S16). The Kuhn length b was calculated as

$$b = \frac{\langle \bar{R}^2 \rangle}{R_{\max}}, \tag{S17}$$

where $\bar{R}$ is the end-to-end distance, measured as the distance between the centers-of-mass of the first and last monomers, and $R_{\max}$ is the contour length, approximated as $(1.4 \text{ nm}) \times N$, where 1.4 nm is the approximate distance between equivalent thiophene groups of adjacent monomers and N the number of monomers in the chain.
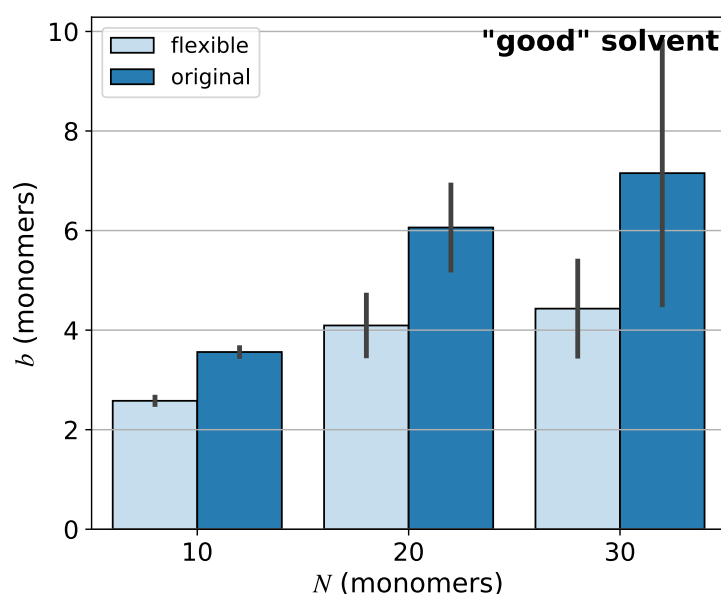
Fig. S16 Kuhn length b of P(NDI2OD-T2) with the flexible and regular-flexibility backbone in the "good" solvent. Error bars are the standard deviation of the mean Kuhn length calculated over 1–10 µs of the 20 independent simulations for each chain length and flexibility.

S7   Chain length and concentration dependence of aggregation properties

The effect of chain length and concentration at constant $\phi_V N^{1/2}$ on the evolution of aggregate size and structure over time is shown in Figs. S17 and S18. The rate of aggregate growth appears to be roughly independent of chain length, particularly at early time (Fig. S17a), indicating that the choice of constant $\phi_V N^{1/2}$ is a reasonable way to compare systems with different chain lengths. The number of overlaps between aggregated chains appears to decrease slightly for the 20 and 30mers (Fig. S18a).

(a) Average aggregate size

(b) Radius of gyration
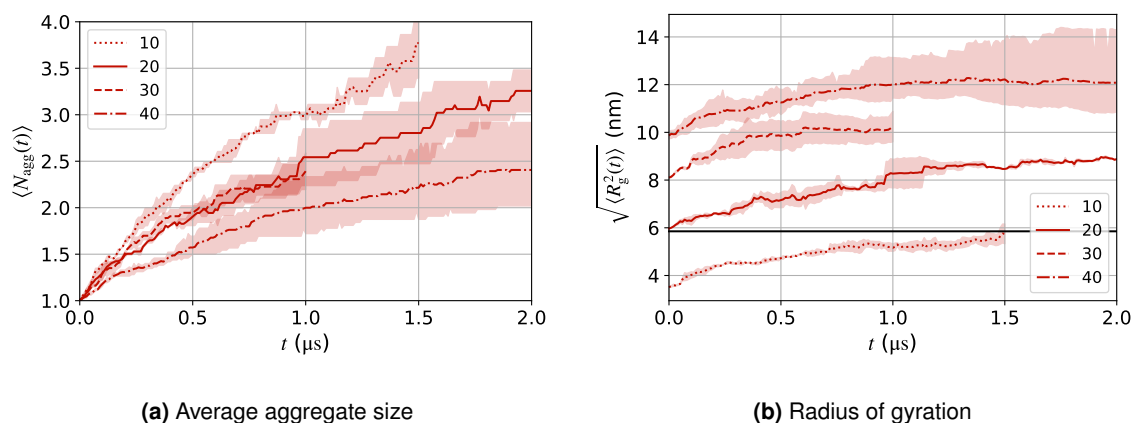
Fig. S17 (a) Average aggregate size and (b) root-mean-squared (RMS) radius of gyration versus time under poor solvent conditions at constant $\phi_V N^{1/2}$ for various values of the polymer chain length N (specified in the legend) and polymer volume fraction ϕ_V. Only the first 2 µs are shown. The horizontal black line in (b) indicates the R_g of a single 20mer in the "good" solvent for reference.

(a) $\langle N_{pair} \rangle / N$
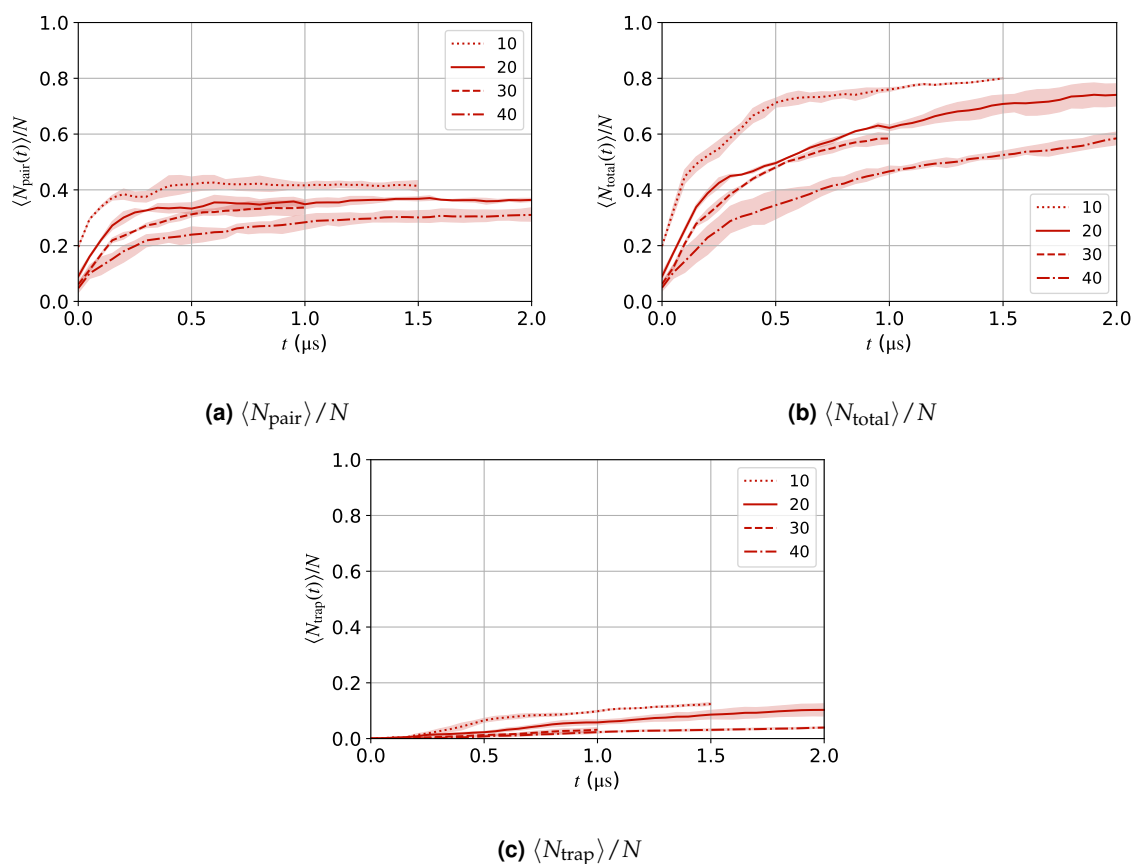
(b) $\langle N_{total} \rangle / N$

(c) $\langle N_{trap} \rangle / N$

Fig. S18 (a) $\langle N_{pair} \rangle / N$, (b) $\langle N_{total} \rangle / N$, and (c) $\langle N_{trap} \rangle / N$ versus time in the poor solvent at constant $\phi_V N^{1/2}$ for various values of the polymer chain lengths N (specified in the legend) and polymer volume fraction ϕ_V. Only the first 2 µs are shown.

S8 Single-chain folding kinetics

S8.1 Intermediate- and late-time 2D histograms of single-chain behaviour

The distribution of single-chain conformations at intermediate (4.5–5.5 μs) and late (9–10 μs) times (early time given in the main text) in the poor solvent are shown in Fig. S19. These distributions highlight that the flexible chains fold more readily than the chains with regular flexibility. A variety of structures were observed, corresponding to extended coils (high R_g, high κ^2), hairpins/racquets (lower R_g, high κ^2), and toroids (lower R_g, lower κ^2). At late time, toroid structures appear to be generally favoured for the flexible backbones, especially the shorter chains, while both hairpins and toroids were common for the regular-flexibility backbones and the longer flexible chains.

(a) 4.5–5.5 μs   **(b)** 9–10 μs
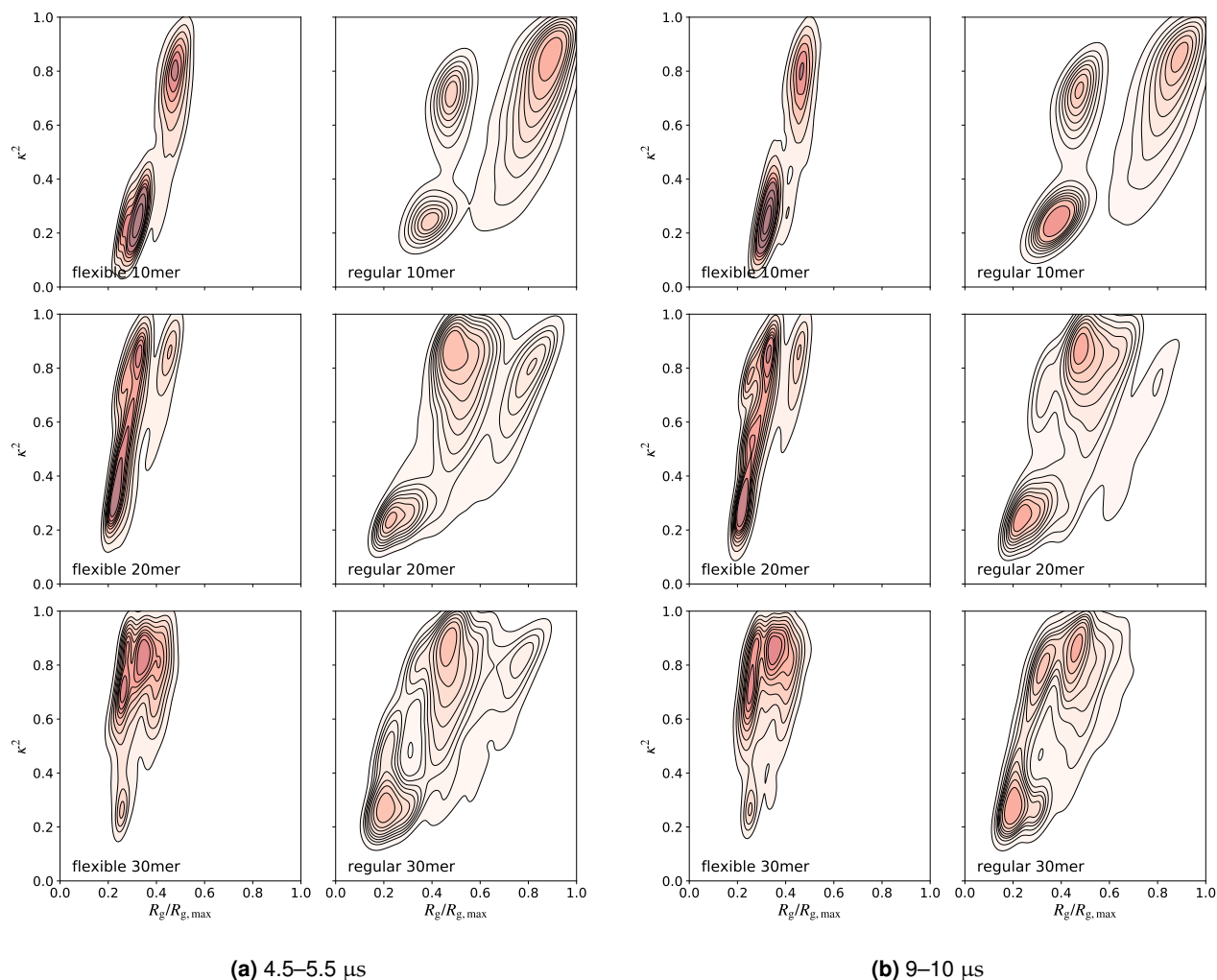
Fig. S19 2D histograms of the radius of gyration, R_g, and shape anisotropy, κ^2, for 20 (flexible chains) or 80 (regular chains) independent simulations of various chain lengths in poor solvent conditions. Early time distributions (0.5–1 μs) can be found in the main text (Fig. 7). The R_g is normalised by $R_{g,max}$, the R_g of a fully extended rod,[19] with $R_g^2 = L^2/12$ and $L = 1.4$ nm per monomer. The colour scale is the same for all plots, with darker colours corresponding to higher probability.

S8.2 2D histograms of single-chain 40mers

The distribution of single-chain conformations at early (0.5–1µs), intermediate (4.5–5.5 µs) and late (9–10 µs) times for single 40mer chains in the poor solvent are shown in Fig. S20.

(a) 0.5–1 µs (b) 4.5–5.5 µs (c) 9–10 µs
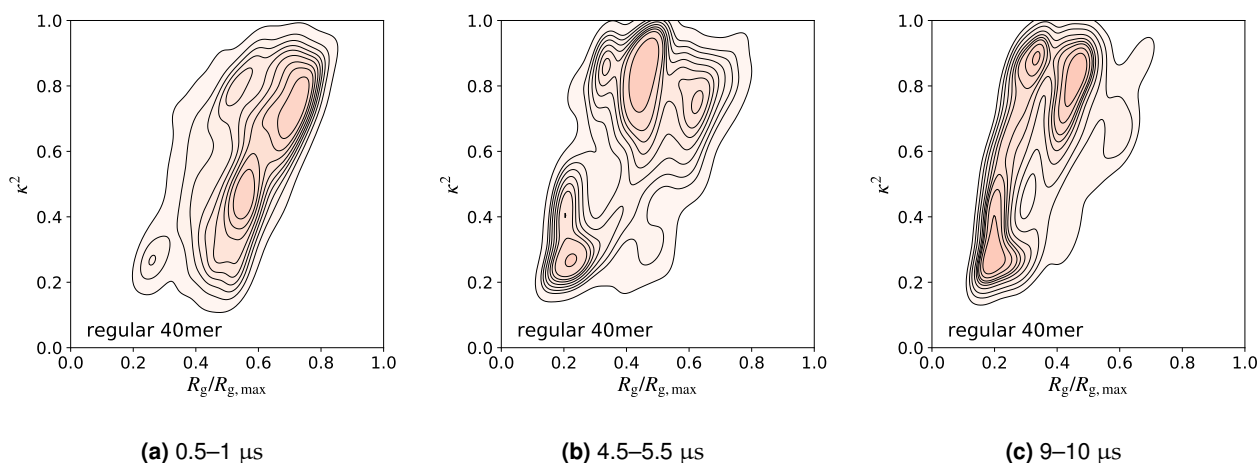
Fig. S20 2D histograms of the radius of gyration, R_g, and shape anisotropy, κ^2, for 100 independent simulations of various chain lengths in poor solvent conditions. The R_g is normalised by $R_{g,max}$, the R_g of a fully extended rod,[19] with $R_g^2 = L^2/12$ and $L = 1.4$ nm per monomer. The colour scale is the same for all plots, with darker colours corresponding to higher probability.

S8.3 Scaling of folding rate with chain length

The rate of single-chain folding for semiflexible polymers has been previously shown to scale with $N^{1/3}$ for chains of length N.[20] We have fit the folding rate (calculated as $1/\tau_s$, where τ_s was calculated from the first-contact time described in the main text) to an equation of the form $1/\tau_s = aN^{1/3} + b$, where a and b are fit parameters. For the regular backbones, this scaling accurately describes the behaviour of the regular-flexibility backbones calculated from the first-contact time (Fig. S21). The fit parameter b shifts the x-intercept of the plot to higher N (here $N \approx 7.5$), which is consistent with single chains not folding ($\tau_s \to \infty$ and so $1/\tau_s \to 0$) for chains with length on the order of a single Kuhn length (approx 5–7 monomers for the model here).
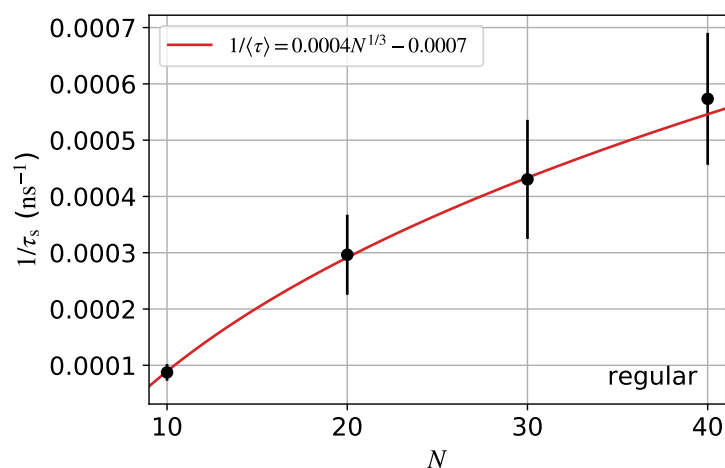
Fig. S21 Scaling of folding rate $1/\tau_s$ with $N^{1/3}$ for the single-chain simulations with the regular-flexibility backbone in the poor solvent. Error bars are two standard errors of the first-contact time τ_s calculated over the 80–100 independent simulations. τ_s was calculated as the first contact time defined in the main text. The red line is the fit to the data.

S9   Scaling of kinetics of multi-chain aggregation vs single-chain collapse

A derivation is given below of the scaling with chain length of the critical concentration at which the rates of single-chain folding and multi-chain aggregation are equal, both with and without hydrodynamic interactions (HI).

Single-chain folding, τ_s

We consider a system of polymer chains of degree of polymerisation N and chain concentration C (monomer concentration $c = NC$) transferred from a good to a poor solvent at time $t = 0$. From the power-law scaling of the single-chain folding time τ_s with N measured in our and other[20] simulations of semiflexible polymers, and assuming that folding is diffusion-controlled,[21] the time scale for single-chain folding in a given solvent is expected to obey

$$\tau_s = \frac{L_0^2}{D_0} N^\alpha, \tag{S18}$$

where D_0 is the monomer diffusion coefficient, L_0 is some length scale, and α is a scaling exponent. Based on previous work,[20] and the simulations here of single-chain folding (Fig. S21), $\alpha \approx -\frac{1}{3}$. The scaling with D_0 in eqn (S18) is also consistent with our simulations, which indicate that τ_s is proportional to the Langevin friction coefficient γ (see Table 1 of the main paper), which is in turn proportional to $1/D_0$.

Multi-chain aggregation, τ_c

We assume that multi-chain aggregation is diffusion limited and occurs via binary collisions between spherical aggregates, with the radius and diffusion coefficient of an aggregate of i chains R_i and D_i, respectively, and ignore the effect of chain collapse on R_i. The kinetics of aggregation is governed by a series of kinetic equations for the concentration C_i of aggregates of size i:[22]

$$\frac{dC_i}{dt} = \frac{1}{2} \sum_{j=1}^{i-1} K_{j,i-j} C_j C_{i-j} - \sum_{j=1}^{\infty} K_{i,j} C_i C_j, \quad i \geq 2, \tag{S19}$$

$$\frac{dC_1}{dt} = -\sum_{j=1}^{\infty} K_{1,j} C_1 C_j, \tag{S20}$$

where, assuming all collisions lead to aggregation,[23]

$$K_{i,j} = 4\pi (D_i + D_j)(R_i + R_j) \left[1 + \frac{(R_i + R_j)}{\sqrt{\pi (D_i + D_j)t}} \right]. \tag{S21}$$

Note that the derivation of eqn (S21) assumes that particles in the system are initially separated by more than $(R_i + R_j)$, since aggregation occurs when particles reach a separation of $(R_i + R_j)$. Above the overlap concentration, all particles will on average be separated from their nearest neighbours by less than this distance, so aggregation will be instantaneous (the aggregation time will be zero), and so the crossover concentration becomes ill-defined.

Eqns (S19)–(S21) are typically solved ignoring the time-dependent term in eqn (S21), for which analytical solutions exist in some cases (e.g. constant $K_{i,j} = K$). But this term can only be ignored if

$$\frac{(R_i + R_j)}{\sqrt{\pi (D_i + D_j)t}} \ll 1, \tag{S22}$$

i.e.

$$t \gg \frac{(R_i + R_j)^2}{\pi (D_i + D_j)}. \tag{S23}$$

Comparing an estimate of this time scale using the RMS radius of gyration and diffusion coefficient from the simulations of single chains in the "good" solvent with the aggregation time from multi-chain simulations in the poor solvent, this condition does not appear to hold in our multi-chain simulations, even for aggregate size $i = j = 1$. In fact, the time-independent part of eqn (S21) appears negligible for large N (see Table S3; here, $R_i = R_j = R_1$, $D_i = D_j = D_1$, and so $\frac{(R_i+R_j)^2}{\pi(D_i+D_j)} = \frac{2R_1^2}{\pi D_1}$, and we take $R_1 \approx R_g(t=0)$ from simulations of single chains).

Table S3 Properties of single-chain diffusion and aggregation, compared with the multi-chain aggregation time τ_c, all measured in the poor solvent with viscosity corresponding to implicit DCB. D_1 and R_1 are the diffusion coefficient and size of a single polymer chain of length N. All data is for chains with the regular backbone flexibility and the multi-chain simulations were at the concentrations that gave constant $\phi_V N^{1/2}$.

N	D_1 ($\text{Å}^2/\text{ns}$)	R_1^2 (Å^2)	$2R_1^2/\pi D_1$ (μs)	τ_c (μs)
10	1.23	1200	0.6	0.42
20	0.85	3400	2.5	0.69
30	0.47	8000	11	0.60

In all cases, aggregation occurs on a shorter time scale (and much shorter for $N \geq 20$) than the time scale for which $K_{i,j}$ is constant. Furthermore, we are interested in the conditions for which $\tau_s \approx \tau_c$. As τ_s decreases with increasing N, we expect the time-dependent part of $K_{i,j}$ to be come dominant for larger N.

Unfortunately, eqns (S19)–(S21) cannot be solved analytically in this case. Thus, for simplicity, we only consider the reaction of single chains to form aggregates of size two, which should accurately approximate the kinetics at short enough times that the population of larger aggregates is small. In this case, eqns (S19) and (S20) become (with $K \equiv K_{1,1}$)

$$\frac{dC_2}{dt} = KC_1^2, \tag{S24}$$

$$\frac{dC_1}{dt} = -KC_1^2, \tag{S25}$$

and eqn (S21) becomes (with $R \equiv R_1$ and $D \equiv D_1$)

$$K = K_{1,1} = 16\pi DR \left(1 + \frac{\sqrt{2}R}{\sqrt{\pi Dt}}\right)$$

$$\equiv A + Bt^{-1/2}, \tag{S26}$$

where

$$A = 16\pi DR \tag{S27}$$

and

$$B = 16\sqrt{2\pi}D^{1/2}R^2. \tag{S28}$$

Combining eqns (S25) and (S26),

$$\frac{dC_1}{dt} = -(A + Bt^{-1/2})C_1, \tag{S29}$$

which can be solved with initial condition $C_1(0) = C$ to give

$$\frac{1}{C_1(t)} = \frac{1}{C} + At + 2Bt^{1/2}. \tag{S30}$$

Defining the characteristic aggregation time τ_c as the time for the concentration of single chains to fall to a fraction x of the initial concentration, i.e.

$$C_1(\tau_c) = x\frac{c}{N}, \tag{S31}$$

then inserting eqn (S31) into eqn (S30) and rearranging gives

$$0 = A\tau_c + 2B\tau_c^{1/2} + \left(\frac{x-1}{x}\right)\frac{N}{c}, \tag{S32}$$

which has the general solution

$$\tau_c = \frac{2B^2 - AG \pm 2\sqrt{B^4 - AB^2G}}{A^2}, \tag{S33}$$

where

$$G \equiv \left(\frac{x-1}{x} \right) \frac{N}{c}. \tag{S34}$$

We will restrict ourselves to two cases:

1. $A\tau_c \gg 2B\tau_c^{1/2}$
2. $A\tau_c \ll 2B\tau_c^{1/2}$

noting that case 2 appears to correspond to the conditions in this work, in which the time-dependent part of eqn (S21) dominates.

Case 1: $A\tau_c \gg 2B\tau_c^{1/2}$

Under these conditions, eqn (S32) becomes

$$\tau_c = -\frac{G}{A} = \frac{N}{Ac} \left(\frac{1-x}{x} \right). \tag{S35}$$

Inserting the definition of A from eqn (S27) gives

$$\tau_c = \frac{1}{16\pi DR} \left(\frac{1-x}{x} \right) \frac{N}{c}. \tag{S36}$$

If we assume

$$D = D_0 N^\beta \tag{S37}$$

and

$$R = R_0 N^\nu, \tag{S38}$$

$$\tag{S39}$$

then

$$\tau_c = \frac{1}{16\pi D_0 R_0} \left(\frac{1-x}{x} \right) \frac{N^{(1-\beta-\nu)}}{c} \propto \frac{N^{(1-\beta-\nu)}}{c}. \tag{S40}$$

Defining the critical monomer concentration, $c^\dagger$, as c where $\tau_c = \tau_s$, combining eqns (S40) and (S18) gives

$$c^\dagger = \frac{1}{16\pi R_0 L_0^2} \left(\frac{1-x}{x} \right) N^{(1-\alpha-\beta-\nu)} \propto N^{(1-\alpha-\beta-\nu)}. \tag{S41}$$

Case 2: $A\tau_c \ll 2B\tau_c^{1/2}$

Under the conditions corresponding to case 2, eqn (S32) gives

$$\tau_c = \frac{N^2}{4B^2 c^2} \left(\frac{1-x}{x} \right)^2. \tag{S42}$$

Inserting the definition of B from eqn (S28):

$$\tau_c = \frac{1}{2^{11}\pi DR^4} \left(\frac{1-x}{x} \right)^2 \frac{N^2}{c^2}. \tag{S43}$$

Using eqns (S37) and (S38)

$$\tau_c = \frac{1}{2^{11}\pi D_0 R_0^4} \left(\frac{1-x}{x} \right)^2 \frac{N^{(2-\beta-4\nu)}}{c^2} \propto \frac{N^{(2-\beta-4\nu)}}{c^2}. \tag{S44}$$

Again defining the critical monomer concentration, $c^\dagger$, where $\tau_c = \tau_s$, combining eqns (S44) and (S18) gives

$$c^\dagger = \frac{1}{(2^{11}\pi)^{1/2} R_0^2 L_0} \left(\frac{1-x}{x} \right) N^{(1-\alpha/2-\beta/2-2\nu)} \propto N^{(1-\alpha/2-\beta/2-2\nu)}. \tag{S45}$$

The scaling of $c^\dagger$ with N is therefore quite different in cases 2 (eqn (S45)) and (eqn (S41)).

Comparison of scaling

As described above, from both the literature[20] and the simulations in this work, $\alpha \approx -1/3$. Given no HI in the Langevin dynamics simulations, $\beta = -1$.[24] Assuming the polymer conformation (initially) corresponds to a good solvent, $\nu \approx 0.6$. With these exponents, from eqns (S41) in case 1 and (S45) in case 2:

$$\textbf{Case 1:} \quad c^{\dagger} \propto N^{(1-\alpha-\beta-\nu)} = N^{(1+1/3+1-0.6)} = N^{1.73} \tag{S46}$$

$$\textbf{Case 2:} \quad c^{\dagger} \propto N^{(1-\alpha/2-\beta/2-2\nu)} = N^{(1+1/6+1/2-2\times 0.6)} = N^{0.47} \tag{S47}$$

Note that case 2 corresponds to the time-scale regime of the simulation in which the time-dependent part of the aggregation rate coefficient dominates.

Calculation of $c^{\dagger}$

To determine $c^{\dagger}$ from Langevin dynamics simulations for case 2 (which is the case for the simulations conducted in this work), the fraction x of single unaggregated chains in solution used to define the aggregation time was chosen to be $x = 0.25$. This value is kept fixed for all the analysis. Note that, from eqns (S41) and eqn (S45), the scaling of $c^{\dagger}$ with chain length N is not predicted to depend on the choice of x, although the value of $c^{\dagger}$ at a given value of N is predicted to be proportional to $(x^{-1}-1)$. From eqn (S44),

$$\tau_{\text{c}} = \frac{f(N)}{c^2}, \tag{S48}$$

where $f(N)$ is a function of N that depends on the simulation system properties and conditions but not on c. From τ_{c} and c measured in the simulation, $f(N)$ for each chain length N simulated can be determined. Then, setting $\tau_{\text{c}} = \tau_{\text{s}}$, where τ_{s} is the single-chain collapse time measured in the simulation , $c^{\dagger}$ for each N simulated can be determined as

$$c^{\dagger}(N) = \left(\frac{f(N)}{\tau_{\text{s}}}\right)^{1/2}. \tag{S49}$$

Correcting the scaling for hydrodynamic interactions

The simulations in this work used Langevin dynamics, in which HI are ignored. The scaling relationships in eqns (S46) and (S47) therefore need to be corrected to account for the effect of HI. From ref. 24, the Kirkwood formula[25] for the translational diffusion coefficient of a polymer of degree of polymerisation N is

$$D = \frac{D_0}{N} + \frac{k_{\text{B}}T}{6\pi\eta}\left\langle \frac{1}{R_{\text{H}}}\right\rangle \tag{S50}$$

where D_0 is the monomer diffusion coefficient and R_{H} the hydrodynamic radius, with

$$\left\langle \frac{1}{R_{\text{H}}}\right\rangle = \frac{1}{N^2}\sum_{i\neq j}\left\langle \frac{1}{r_{ij}}\right\rangle, \tag{S51}$$

where r_{ij} is the distance between monomers i and j.

From Zimm theory (see e.g. refs. 19 and 26)

$$\frac{k_{\text{B}}T}{6\pi\eta}\left\langle \frac{1}{R_{\text{H}}}\right\rangle = \frac{D_0^*}{N^{\nu}}, \tag{S52}$$

where ν is the scaling exponent for polymer size, and D_0^* is the segmental diffusion coefficient, which is expected to be similar to the monomer diffusion coefficient D_0. Assuming $D_0^* \approx D_0$, and combining eqns (S52) and (S50) gives

$$D \approx D_0\left(\frac{1}{N} + \frac{1}{N^{\nu}}\right). \tag{S53}$$

Note that eqn (S53) is incorrect for $N = 1$, where $D = D_0$, but should be reasonably accurate for $N \gg 1$.

Since the scaling of the τ_{c} and $c^{\dagger}$ involves D in the denominator (see eqns (S40) and (S44)), which contributes a scaling of $c^{\dagger}$ with $N^{-\beta}$ for case 1 and $N^{-\beta/2}$ for case 2, correcting the $c^{\dagger}$ for $\beta = -1$ from Langevin dynamics simulations involves multiplying $c^{\dagger}$ calculated from the simulations by

$$\textbf{Case 1:} \quad \frac{1}{N}\left(\frac{1}{N} + \frac{1}{N^{\nu}}\right)^{-1} = \frac{N^{\nu}}{N + N^{\nu}} \tag{S54}$$

$$\textbf{Case 2:} \quad \frac{1}{N^{1/2}}\left(\frac{1}{N} + \frac{1}{N^{\nu}}\right)^{-1/2} = \left(\frac{N^{\nu}}{N + N^{\nu}}\right)^{1/2} \tag{S55}$$

Thus, if $c^\dagger_{\text{noHI}}(N)$ is the critical concentration from a Langevin dynamics simulation for chain length N, where "noHI" denotes the absence of hydrodynamic interactions, the critical concentration in the presence of hydrodynamic interactions can be estimated by multiplying this value by eqn (S55) (for case 2) to give

$$c^\dagger_{\text{HI}}(N) \approx c^\dagger_{\text{noHI}}(N) \left(\frac{N^\nu}{N + N^\nu} \right)^{1/2}. \tag{S56}$$

For $N \gg 1$ and $\nu < 1$, eqn (S56) simplifies to

$$c^\dagger_{\text{HI}}(N) \approx c^\dagger_{\text{noHI}}(N) N^{\frac{1}{2}(\nu - 1)}. \tag{S57}$$

So given $c^\dagger_{\text{noHI}}(N)$ from fitting the data for $c^\dagger$ versus N without hydrodynamic interactions, and assuming a value of ν (e.g. 0.6 for a good solvent), $c^\dagger$ with hydrodynamic interactions can be predicted using eqn (S57) for $N \gg 1$.

S10   Concentration effects

S10.1   Effect of concentration on aggregate size

The effect of concentration on the average aggregate size as a function of time was examined by comparing three sets of simulations with the same chain length and properties, but different concentrations: regular-flexibility 10mers at 8.5 and 10 g/L, regular-flexibility 20mers at 6 and 12 g/L, and flexible 20mers at 2 and 6 g/L. More rapid aggregate growth, and the formation of larger aggregates, was observed at higher concentration. Comparing the 6 and 12 g/L simulations, τ_c is approx. 4× smaller at the higher concentration (see Table S2), consistent with the predicted $1/c^2$ scaling from the theory (eqn (S44)).

(a) Regular 10mers

(b) Regular 20mers

(c) Flexible 20mers
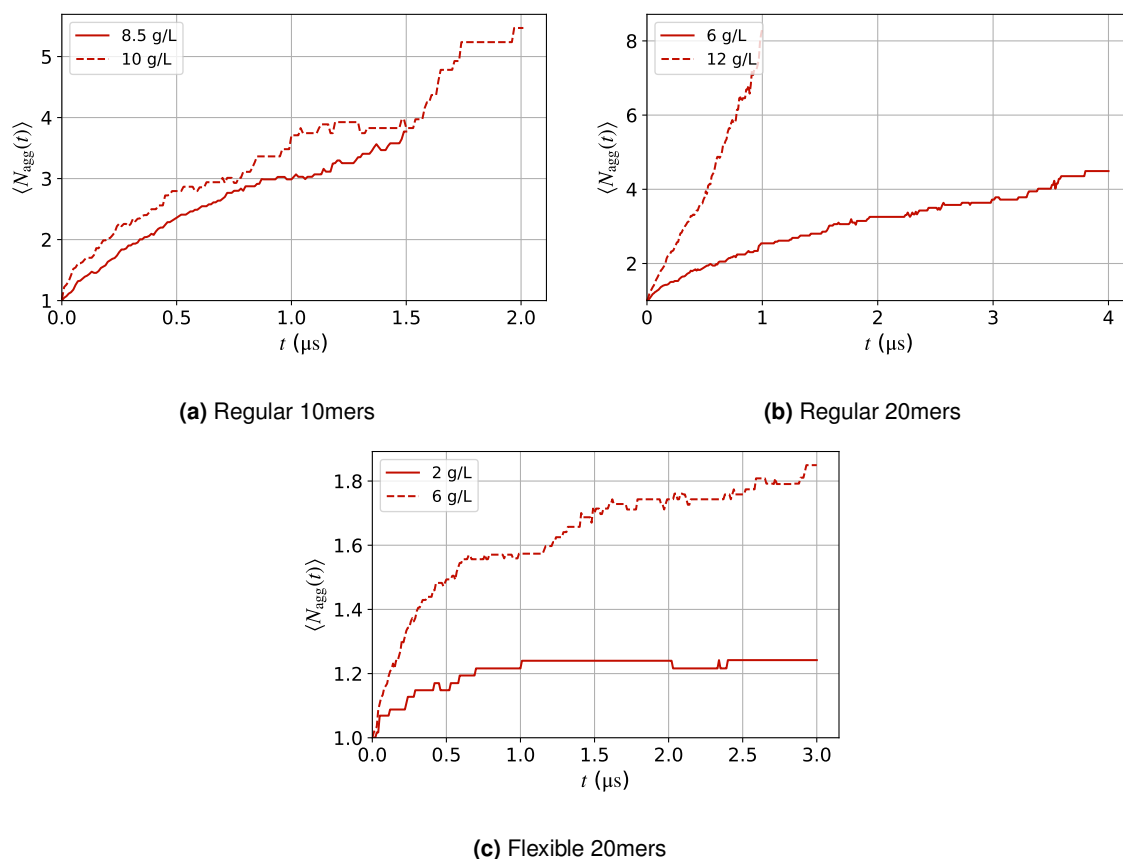
Fig. S22 Average aggregate size (number of chains) versus time at different concentrations for different chain lengths and flexibilities. Higher concentrations (dashed lines) give more rapid aggregate growth and overall larger aggregates.

S10.2   Effect of flexibility on critical concentration

The critical concentration $c^\dagger$ was calculated from simulations of various chain lengths, at different concentrations, and for polymers with different flexibilities. While the results for the regular-flexibility backbones are presented in the main text in Fig. 8, the effect of chain flexibility was not presented. An additional data point from the more flexible chains is shown in Fig. S23, as well as a breakdown of the regular-flexibility points by the concentration of the multi-chain simulations.

The majority of simulations in this work were conducted at concentrations higher than $c^\dagger$ ($\phi_V N^{1/2}$ given by dashed black line in Fig. S23a), though the flexible chain simulations approached, or dropped below, this critical concentration. Accordingly, although all simulations in poor solvents showed some degree of multi-chain aggregation, there was substantially less in the case of flexible chains. The scaling of this critical concentration with chain length N is discussed in the main text, and can explain the conflicting behaviour in the two published experimental studies.[27,28] We also examined one system at a concentration expected to be well below the critical concentration for the flexible 20mers in poor solvent (2 g/L). Although this system initially still showed a small amount of aggregation, due to random placement of chains in the box resulting in chains that were close enough to aggregate before they folded, there was significantly less multi-chain aggregation observed in this case, and almost none after the initial aggregation (Fig. S22c).

(a) Volume fraction relative to overlap concentration   **(b)** Concentration
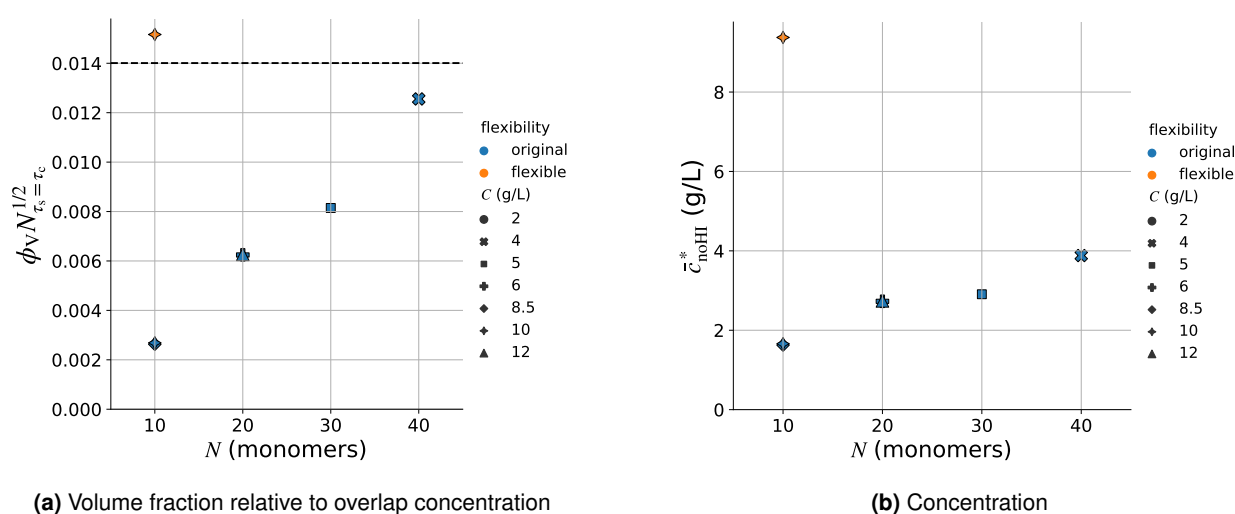
Fig. S23 Scaling with chain length N of the critical concentration at which $\tau_{\mathrm{s}} = \tau_{\mathrm{c}}$, presented as either (a) $\phi_{\mathrm{V}} N^{1/2}$ or (b) concentration (g/L). The points corresponding to different concentrations were calculated using the aggregation time from multi-chain simulations at the labelled concentration. The horizontal dashed line in (a) indicates the value of $\phi_{\mathrm{V}} N^{1/2}$ used for the majority of simulations in this work. Systems at concentrations below the points on these plots are expected to favour single-chain folding, while higher concentrations should lead to more prominent multi-chain aggregation.

S11   Effect of viscosity on multi-chain aggregation

As the experimental viscosity of DCB and TOL differ significantly,[29] it is important to consider the effect of this parameter on the single- and multi-chain behaviour of the CG model. The effect of this change in viscosity on the multi-chain aggregation properties is shown in Figs. S24 and S25. τ_c in the low-viscosity solvent is approximately $10\times$ smaller than in implicit DCB (0.06 vs 0.7 µs, calculated as the time for the single-chain concentration to fall below 25% its original concentration). This scaling is consistent with the $10\times$ smaller friction coefficient for the low-viscosity solvent.

(a) Average aggregate size

(b) Radius of gyration
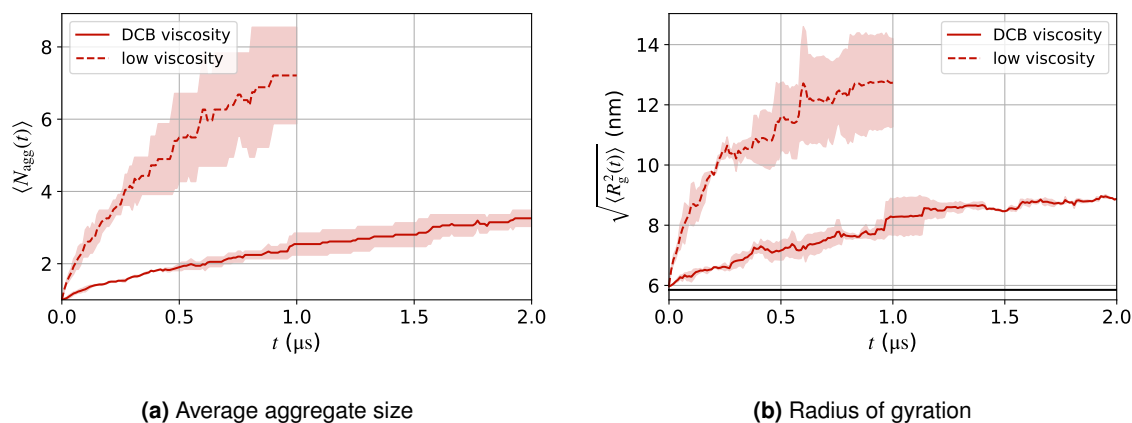
Fig. S24 (a) Average aggregate size and (b) RMS radius of gyration as a function of time for different solvent viscosities under poor-solvent conditions. The horizontal black line in (b) indicates R_g of a single 20mer in the "good" solvent for reference.

(a) $\langle N_{\text{pair}} \rangle / N$
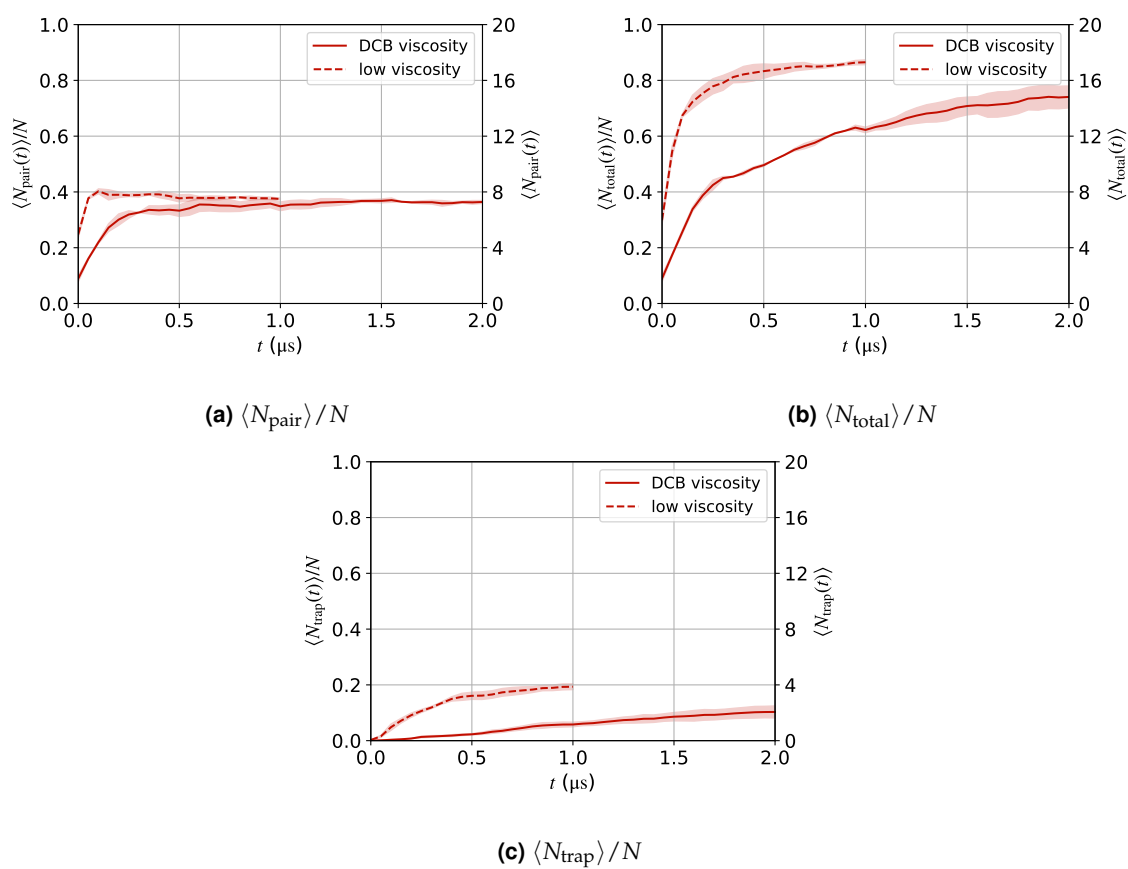
(b) $\langle N_{\text{total}} \rangle / N$

(c) $\langle N_{\text{trap}} \rangle / N$

Fig. S25 (a) $\langle N_{\text{pair}} \rangle / N$, (b) $\langle N_{\text{total}} \rangle / N$, and (c) $\langle N_{\text{trap}} \rangle / N$ as a function of time for 20mers in different solvent viscosities under poor solvent conditions.

S12   Backbone flexibility

S12.1   Effect of overlaps and backbone flexibility on Kuhn length and bending rigidity

The effect of backbone aggregation on chain flexibility is discussed briefly in the main text. Fully overlapping chains were shown to increase the overall rigidity of the aggregate backbone, with a 3–5-fold increase in the Kuhn length observed on going from a single free chain to an aggregated pair, even on transitioning from a "good" to a poor solvent. The Kuhn lengths and bending rigidities for the single-chain systems not listed in the main text, and for a fully overlapping aggregate of flexible chains in poor solvent, are given in Table S4.

Table S4 Kuhn length, b (Å or number of monomers), and bending rigidity, $\kappa_b = bk_B T/2$ (assuming a worm-like chain [30]), for 30mers with different backbone flexibilities and solvent qualities. Where b is given as a number of monomers, each monomer was assumed to be 1.4 nm long. Systems labelled "aggregated" are from a single 1 μs simulation of two fully overlapping 30mers, with listed values averaged over the final 100 ns for each chain in the aggregate pair. The single-chain simulations (labelled "single") use data from the same 10–15 μs single-chain simulations described previously, using the values averaged over the 0.9–1 μs period in each of the 20 (flexible chains in both solvent, and regular chains in "good" solvent) or 80 (regular chains in poor solvent) independent simulations.

flexibility	solvent	system	b (Å)	b (monomers)	κ_b (kcal/mol.Å)
flexible	poor	single	24.04	1.72	7.16
flexible	poor	aggregated	287.28	20.52	85.63
regular	poor	single	92.18	6.58	27.48
regular	poor	aggregated	279.91	20.00	83.44
flexible	"good"	single chain	57.04	4.07	17.00
regular	"good"	single chain	100.17	7.15	29.86

S12.2   Effect of backbone flexibility on chain overlaps

The effect of backbone flexibility on some of the properties defining the aggregate structure of P(NDI2OD-T2) in both "good" and poor solvents is shown below. The overall aggregate size (in terms of number of monomers and R_g) is given in the main text Fig. 9, and chain overlap fractions (N_{pair}/N, N_{total}/N, and N_{trap}/N) are shown in Fig. S26. For the poor solvent, a similar degree of overlap was observed for aggregated chains with the flexible and regular backbones. This indicates that, although less aggregation was observed overall for flexible chains (Fig. 9), where aggregation did occur, it was in a similar manner to the regular backbone, with partially overlapping chains. Overall, more chain collapse was observed prior to aggregation for the flexible chains, resulting in less aggregation and an overall lower R_g (due to collapse of either single chains or partially overlapping aggregates), though the initial aggregation process followed a similar path for both backbone flexibilities, giving partially overlapping chains with $N_{pair}/N \approx 0.4$ in the poor solvent. The difference was more prominent in the "good" solvent, in which aggregation did not initially occur irreversibly as it did in the poor solvent. The more compact chains formed with the flexible backbones resulted in both less aggregation and fewer overlaps in the "good" solvent compared with the same system with the stiffer backbone. These chains were able to collapse more prior to extensive aggregation, with R_g^2 of the single flexible chains being approximately 60% of the regular ones in the "good" solvent.

(a) $\langle N_{\mathrm{pair}} \rangle / N$
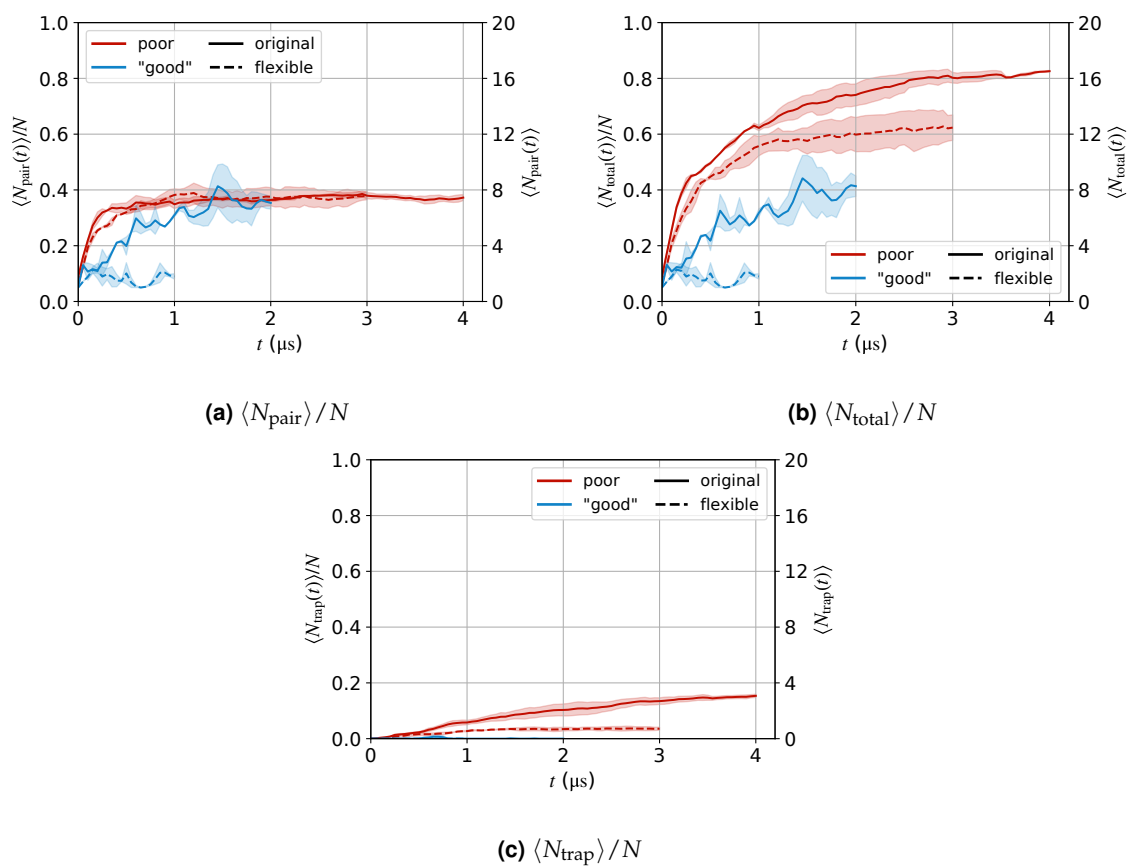
(b) $\langle N_{\mathrm{total}} \rangle / N$

(c) $\langle N_{\mathrm{trap}} \rangle / N$

Fig. S26 (a) $\langle N_{\mathrm{pair}} \rangle / N$, (b) $\langle N_{\mathrm{total}} \rangle / N$, and (c) $\langle N_{\mathrm{trap}} \rangle / N$ as a function of time for different backbone flexibilities of 20mers at 6 g/L for "good" and poor solvent conditions.

S13   All-atom parameters

The all-atom parameters for the symmetric P(NDI2OD-T2) model are reproduced below. The atom and bond types are defined in Fig. S27. Where parameters were not parameterised explicitly (for the side-chains for example), they were taken directly from the OPLS force field for equivalent atom types.[2–8]
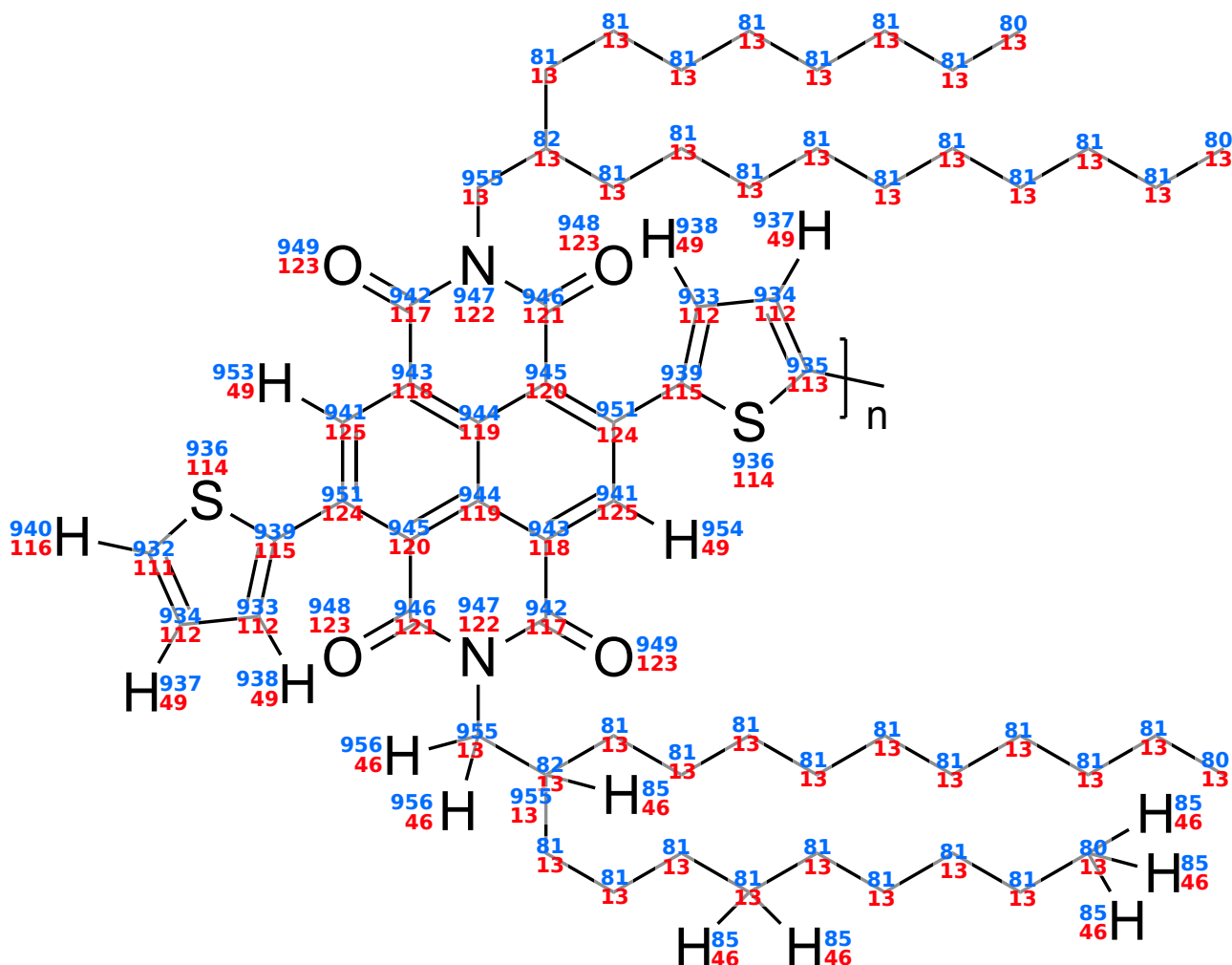
Fig. S27 Definition of the different atom (blue) and bond (red) types for this parameterisation of P(NDI2OD-T2). Atom types with values 932 and above differ from those in the OPLS force field, with their partial charges obtained from quantum-chemical calculations, but their van der Waals parameters came from the OPLS force field for equivalent atom types. Bond types with values 110 and above also differ from those in the OPLS force field, with the bond lengths and angles obtained from the optimised geometry, and force constants from the OPLS force field for equivalent atom types. The energy parameters for the 113–113 bond length, and the 120–124–113–112, and 112–113–113–112 dihedrals were explicitly parameterised (as described in Section S1) as these potentials were not expected to be accurately represented by the OPLS force field. For the sake of representing both a terminal and central monomer of a polymer chain in a single image, the left terminus is shown with the site types corresponding to a terminal monomer, and the right to a central monomer (connecting to another monomer). For clarity, only a few side-chain hydrogens are shown.

S13.1 Non-bonded parameters

Table S5 Masses, partial charges, and LJ parameters for the AA symmetric P(NDI2OD-T2) monomer model. The non-bonded potential between two particles of type i and j is defined as

$$U_{\mathrm{NB}}(r_{ij}) = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}},$$

where σ_{ii} is the homonuclear LJ diameter, ϵ_{ii} the homonuclear LJ interaction strength, and q_i the charge. Geometric mixing rules were used to define the heteronuclear parameters from the homonuclear ones: $\sigma_{ij} = \sqrt{\sigma_i \sigma_j}$ and $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$. The excess charge of the system, resulting from the truncation of the side-chains in the model parameterisation and the use of OPLS charges for the final side-chain atoms, was added to atom type 955 (or 957 for teminal monomers of polymer chain) to enforce charge neutrality. The resulting charge of this atom type was approximately -0.08 *e*, but the charges used in the force field depended on whether the monomer was an isolated monomer (955^m), central polymer unit (955^{pc}), or terminal polymer unit (957) in order to maintain charge neutrality. Charges and LJ parameters for atom types 80, 81, 82, and 85 came directly from the OPLS force field. Note that 1–2 and 1–3 interactions (between particles separated by one and two bonds, respectively) were set to zero, and the 1–4 LJ interactions (between particles separated by three bonds) to 0.5 times their full values, consistent with the OPLS force field.

atom type (i)	mass (g/mol)	q_i (e)	ϵ_{ii} (kcal/mol)	σ_{ii} (Å)
932	12.01	0.126452	0.07	3.55
933	12.01	-0.222396	0.07	3.55
934	12.01	-0.271944	0.07	3.55
935	12.01	0.1209635	0.07	3.55
936	32.06	-0.1260135	0.25	3.55
937	1.008	0.1944945	0.03	2.42
938	1.008	0.1890015	0.03	2.42
939	12.01	0.110991	0.07	3.55
940	1.008	0.182183	0.03	2.42
941	12.01	-0.099982	0.07	3.55
942	12.01	0.6655695	0.105	3.75
943	12.01	-0.205615	0.07	3.55
944	12.01	0.083396	0.07	3.55
945	12.01	-0.1794755	0.07	3.55
946	12.01	0.5946645	0.105	3.75
947	14.01	-0.344225	0.17	3.25
948	16.00	-0.45413	0.21	2.96
949	16.00	-0.5356155	0.21	2.96
950	12.01	0.149833	0.07	3.55
951	12.01	0.132339	0.07	3.55
952	1.008	0.128893	0.03	2.42
953	1.008	0.043813	0.03	2.42
954	1.008	0.128152	0.03	2.42
955^m	12.01	-0.0667025	0.066	3.5
955^{pc}	12.01	0.120969	0.066	3.5
956	1.008	0.07051275	0.03	2.42
957	12.01	0.040045	0.066	3.5
80	12.01	-0.18	0.066	3.5
81	12.01	-0.12	0.066	3.5
82	12.01	-0.06	0.066	3.5
85	1.008	0.06	0.03	2.5

^{m} charge for monomer
^{pc} charge for central monomer of polymer chain

S13.2   Bonded parameters

Bonds, angles, dihedrals, and impropers for the AA model of P(NDI2OD-T2) are listed in terms of their bond types (red numbers in Fig. S27).

Table S6 Bond length parameters. Bond stretching coefficients, k_2, were taken directly from the OPLS force field with the exception of the identical 115–124 and 111–124 bonds, which were parmeterised explicitly. Most equilibrium bond lengths (l_0) were obtained from the optimised geometry of P(NDI2OD-T2). Bonds 13–13 and 13–46 contain only side-chain atoms so both k_2 and l_0 were taken from the OPLS force field. The harmonic bond length potential is defined in eqn (S8a).

bond	bond style	k_2 (kcal/mol/Å^2)	l_0 (Å)
49–112	harmonic	367	1.084
111–112	harmonic	546	1.371
111–114	harmonic	250	1.733
112–112	harmonic	469	1.419
112–113	harmonic	546	1.383
112–115	harmonic	546	1.379
113–113	harmonic	512	1.449
113–114	harmonic	250	1.751
114–115	harmonic	250	1.757
115–124*	harmonic	345	1.468
116–111	harmonic	367	1.081
117–118	harmonic	400	1.486
117–122	harmonic	490	1.400
117–123	harmonic	570	1.225
118–119	harmonic	469	1.417
118–125	harmonic	469	1.382
119–119	harmonic	469	1.424
119–120	harmonic	469	1.420
120–121	harmonic	400	1.492
120–124	harmonic	469	1.400
121–122	harmonic	490	1.408
121–123	harmonic	570	1.224
122–13	harmonic	337	1.482
124–49	harmonic	367	1.084
124–125	harmonic	469	1.415
125–49	harmonic	367	1.084
111–124*	harmonic	345	1.468
13–13	harmonic	268	1.529
13–46	harmonic	340	1.090

* Bonds for which force constant was explicitly parameterised

Table S7 Bond angle parameters. Angle bending coefficients, k_2, were taken directly from the OPLS force field. Most equilibrium bond angles (θ_0) were obtained from the optimsed geometry of P(NDI2OD-T2). Angles 13–13–13, 46–13–13, and 46–13–46 contain only side-chain atoms so both k_2 and θ_0 were taken from the OPLS force field. The harmonic bond angle potential is defined in eqn (S9a).

angle	angle style	k_2 (kcal/mol/rad^2)	θ_0 (°)
111–112–112	harmonic	70	112.8
112–112–49	harmonic	35	123.6
112–112–113	harmonic	70	113.5
112–113–113	harmonic	70	128.9
112–113–114	harmonic	74	110.2
113–114–111	harmonic	74	91.8
113–114–115	harmonic	74	92.3
114–111–112	harmonic	70	111.7
114–113–113	harmonic	74	120.8
114–115–112	harmonic	74	110.0
115–112–49	harmonic	35	122.8
115–112–112	harmonic	70	113.9
112–115–124	harmonic	70	129.6
115–124–125	harmonic	70	116.2
115–124–120	harmonic	70	125.6
124–125–49	harmonic	35	120.2
124–125–118	harmonic	63	121.7
124–120–119	harmonic	63	119.9
114–115–124	harmonic	70	120.1
125–124–120	harmonic	63	119.0
125–118–119	harmonic	63	120.4
125–118–117	harmonic	85	119.2
118–119–119	harmonic	63	118.4
118–119–120	harmonic	63	121.0
118–117–123	harmonic	80	122.0
118–117–122	harmonic	35	116.8
118–125–49	harmonic	35	118.1
117–118–119	harmonic	85	120.5
117–122–121	harmonic	70	124.7
117–122–13	harmonic	50	117.6
122–121–120	harmonic	70	117.5
122–121–123	harmonic	80	120.0
122–117–123	harmonic	80	121.2
122–13–46	harmonic	35	106.2
122–13–13	harmonic	80	114.9
121–122–13	harmonic	50	117.4
121–120–119	harmonic	85	118.9
121–120–124	harmonic	85	121.1
123–121–120	harmonic	80	122.5
120–119–119	harmonic	63	120.6
120–124–49	harmonic	35	119.0
116–111–112	harmonic	35	132.1
116–111–114	harmonic	35	125.0
113–112–49	harmonic	35	123.4
13–13–13	harmonic	38.35	112.7
46–13–13	harmonic	37.5	110.7
46–13–46	harmonic	33	107.8

Table S8 Dihedral parameters. Unless otherwise stated, parameters were taken directly from the OPLS force field for equivalent atom types. The "multi/harmonic" dihedral styles are of the form in eqn (S2) with $m = 4$ or 8 for the multi/harmonic4 and multi/harmonic8 potentials, respectively. The "opls" style dihedral has the form given in eqn (S3). X is a wild-card atom. For the three dihedrals that were explicitly parameterised (112–113–113-112, 112–115–124–120, and 112–111–124–120), the other dihedrals involving the same two central atoms had all energy parameters set to zero, such that the dihedral was completely controlled by the one set of parameters. All energy parameters k_i have units of kcal/mol.

dihedral	dihedral style	k_0	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8
X–114–111–X	opls	0	2.8	0	0					
X–111–112–X	opls	0	7.25	0	0					
X–112–112–X[a]	opls	0	7.25	0	0					
X–112–113–X	opls	0	7.25	0	0					
X–113–114–X	opls	0	2.8	0	0					
112–113–113–114[b]	opls	0	0	0	0					
112–113–113–112[b]	multi/harmonic4	2.542	0.042	−7.139	−0.351	3.121				
114–113–113–114[b]	opls	0	0	0	0					
X–114–115–X	opls	0	2.8	0	0					
X–115–112–X	opls	0	7.25	0	0					
112–115–124–120[b]	multi/harmonic8	0.802	0.321	−2.186	−1.377	−24.648	0.893	48.435	0.984	−30.962
112–115–124–125[b]	opls	0	0	0	0					
114–115–124–125[b]	opls	0	0	0	0					
114–115–124–120[b]	opls	0	0	0	0					
112–111–124–120[b]	multi/harmonic8	0.802	0.321	−2.186	−1.377	−24.648	0.893	48.435	0.984	−30.962
112–111–124–125[b]	opls	0	0	0	0					
114–111–124–125[b]	opls	0	0	0	0					
114–111–124–120[b]	opls	0	0	0	0					
X–124–125–X	opls	0	7.25	0	0					
X–125–118–X	opls	0	7.25	0	0					
X–118–117–122	opls	0	1.1	0	0					
X–118–117–123	opls	0	2.1	0	0					
118–117–122–121[c]	opls	0	7.25	0	0					
118–117–122–13[c]	opls	0	7.25	0	0					
123–117–122–X[c]	opls	0	2.1	0	0					

dihedral	dihedral style	k_0	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8
120–121–122–117[c]	opls	0	7.25	0	0					
120–121–122–13[c]	opls	0	7.25	0	0					
123–121–122–X[c]	opls	0	2.1	0	0					
X–122–13–X	opls	0	0	0	0					
X–120–121–123	opls	0	2.1	0	0					
X–120–121–122	opls	0	1.1	0	0					
X–119–120–X	opls	0	7.25	0	0					
X–118–119–X	opls	0	7.25	0	0					
X–119–119–X	opls	0	7.25	0	0					
X–120–124–X	opls	0	7.25	0	0					
13–13–13–13	opls	1.3	−0.05	0.2	0					
46–13–13–13	opls	0	0	0.3	0					
46–13–13–46	opls	0	0	0.3	0					

[a] ref 1
[b] parameterised in this work
[c] ref 31

Table S9 Improper torsion parameters. All parameters are from the OPLS force-field for equivalent atom types. The cosine potential defined in eqn (S11) (implemented in LAMMPS as "cvff") was used for all improper torsions.

improper	improper style	k (kcal/mol)	d	n
X-X-117-123	cvff	10.5	−1	2
X-X-118-X	cvff	1.1	−1	2
X-X-119-X	cvff	1.1	−1	2
X-X-120-X	cvff	1.1	−1	2
X-X-121-123	cvff	10.5	−1	2
X-X-122-X	cvff	1.1	−1	2
X-X-124-X	cvff	1.1	−1	2
X-X-125-X	cvff	1.1	−1	2
X-X-111-X	cvff	1.1	−1	2
X-X-112-X	cvff	1.1	−1	2
X-X-115-X	cvff	1.1	−1	2
X-X-113-X	cvff	1.1	−1	2

S13.3   Solvent (DCB) parameters

All-atom simulations were conducted in an explicit solvent of *o*-dichlorobenzene (DCB). All parameters for this molecule were taken directly from the OPLS force field,[2–8] but we have reproduced them below for clarity. The structure of the molecule and the atom/bond types are given in Fig. S28 and the parameters in the following tables. Atom and bond type numbering follows the `oplsaa.prm` file distributed with Tinker.[32] The numbers for the atom and bond types of P(NDI2OD-T2) described above were chosen to not overlap with the existing OPLS atom types from this source.
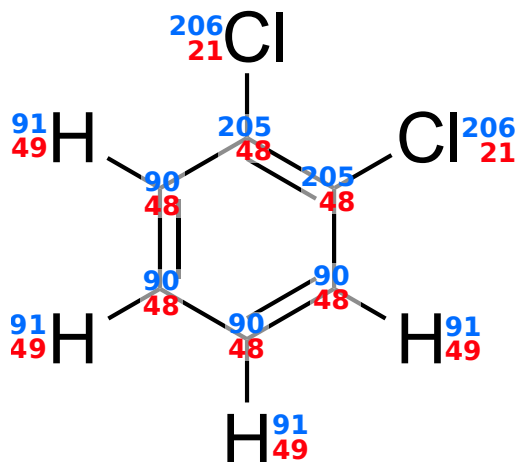
Fig. S28 Atom (blue) and bond (red) types for the AA model of DCB used as the solvent for parameterisation of the CG system. These atom/bond types are referenced in the tables below for the parameters of this molecule.

Table S10 Masses, partial charges q_i, and homonuclear LJ parameters ϵ_{ii} and σ_{ii} for the AA DCB model, with parameters taken directly from the OPLS force field.

atom type (i)	mass (g/mol)	q_i (e)	ϵ_{ii} (kcal/mol)	σ_{ii} (Å)
90	12.01	−0.115	0.07	3.55
91	1.008	0.115	0.03	2.42
205	12.01	0.18	0.07	3.55
206	35.453	−0.18	0.3	3.4

Table S11 Bond length parameters. All parameters come directly from the OPLS force field. The harmonic bond potential has the form given in eqn (S8a).

bond	bond style	k_2 (kcal/mol/Å^2)	l_0 (Å)
48–48	harmonic	469	1.4
48–21	harmonic	300	1.725
48–49	harmonic	367	1.08

Table S12 Bond angle parameters. All parameters come directly from the OPLS force field. The harmonic angle potential has the form given in eqn (S9a).

angle	angle style	k_2 (kcal/mol/rad^2)	θ_0 (°)
48–48–48	harmonic	63	120.0
48–48–49	harmonic	35	120.0
21–48–48	harmonic	75	120.0

Table S13 Dihedral parameters. All parameters come directly from the OPLS force field. The opls style dihedral potential has the form given in eqn (S3)

dihedral	dihedral style	k_0	k_1	k_2	k_3
X–48–48–X	opls	0	7.25	0	0

Table S14 Improper torsion parameters. All parameters come directly from the OPLS force field. The cosine (cvff) style potential has the form given in eqn (S11).

improper	improper style	k (kcal/mol)	d	n
X-X-48-X	cvff	1.1	-1	2

S14 Coarse-grained model parameters

The parameters for the coarse-grained model are given in this section. The forms of the potentials, and the definitions of the parameters are given in Section S3.1. Although it is included in the main text, we replicate the image of the site-type definitions and their relationship to the all-atom model for clarity, as we will refer to the sites by their number in the following tables.
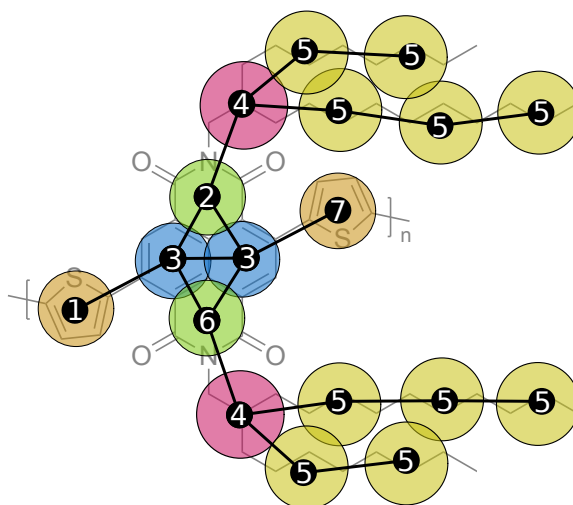
Fig. S29 Definition of site types for the CG P(NDI2OD-T2) model. Sites are coloured and labelled according to their site types, with sites that have the same non-bonded parameters (though not necessarily the same bonded parameters) shown as the same colour. To preserve the backbone structure in the AA representation, in which the thiophenes are connected at the 1 and 6 (rather than 1 and 7) carbon positions of the naphthalene group, two different site types for the thiophenes (1 and 7) and imides (2 and 6) are defined.

S14.1 Site masses

Table S15 Site masses for the CG representation of P(NDI2OD-T2). Masses correspond to the mass of the particles in the AA representation that make up the CG site. Where atoms are shared between multiple CG sites, the mass is split evenly among those sites.

site type	mass (g/mol)
1	84.116
2	86.050
3	45.040
4	55.096
5	42.078
6	86.050
7	82.116

S14.2 Non-bonded parameters

The non-bonded parameters for the P(NDI2OD-T2) model corresponding to the three solvent conditions are given in Table S16, and plotted in Fig. S30.

Table S16 Non-bonded interaction parameters for the three solvent conditions. ϵ_{DCB} is the as-parameterised parameters. ϵ_{good} and ϵ_{poor} are the same parameters scaled by $\pm$ 20 %. The same values of σ_{LJ} are used for all three sets of parameters. The WCA cutoff is the cutoff used for the setup simulations with repulsive particles. It is equal to $2^{1/6}\sigma_{\mathrm{LJ}}$ and the potential is shifted to zero at this point. Note that some parameters (e.g. 1–1 and 7–7) are identical, as the site types are chemically equivalent. They are assigned different site types as the bonded interactions differ, but the non-bonded interactions have been constrained to be the same. In contrast to the AA model, for which 1–4 interactions were included, 1–2, 1–3, and 1–4 interactions were all set to zero in the CG model.

pair	ϵ_{DCB} (kcal/mol)	ϵ_{good} (kcal/mol)	ϵ_{poor} (kcal/mol)	σ_{LJ} (Å)	WCA cutoff (Å)
1–1	0.364	0.291	0.437	3.987	4.475
1–2	0.023	0.018	0.028	5.089	5.712
1–3	0.331	0.265	0.397	4.605	5.169
1–4	0.070	0.056	0.084	5.425	6.089
1–5	0.027	0.022	0.032	5.556	6.236
1–6	0.023	0.018	0.028	5.089	5.712
1–7	0.364	0.291	0.437	3.987	4.475
2–2	0.586	0.469	0.703	3.778	4.241
2–3	0.551	0.441	0.661	3.449	3.871
2–4	0.555	0.444	0.666	4.141	4.648
2–5	0.050	0.040	0.060	5.230	5.870
2–6	0.586	0.469	0.703	3.778	4.241
2–7	0.023	0.018	0.028	5.089	5.712
3–3	0.369	0.295	0.443	3.798	4.263
3–4	0.338	0.270	0.406	4.153	4.662
3–5	0.047	0.038	0.056	5.165	5.798
3–6	0.551	0.441	0.661	3.449	3.871
3–7	0.331	0.265	0.397	4.605	5.169
4–4	0.243	0.194	0.292	5.977	6.709
4–5	0.068	0.054	0.082	5.182	5.817
4–6	0.555	0.444	0.666	4.141	4.648
4–7	0.070	0.056	0.084	5.425	6.089
5–5	0.076	0.061	0.091	5.068	5.689
5–6	0.050	0.040	0.060	5.230	5.870
5–7	0.027	0.022	0.032	5.556	6.236
6–6	0.586	0.469	0.703	3.778	4.241
6–7	0.023	0.018	0.028	5.089	5.712
7–7	0.364	0.291	0.437	3.987	4.475
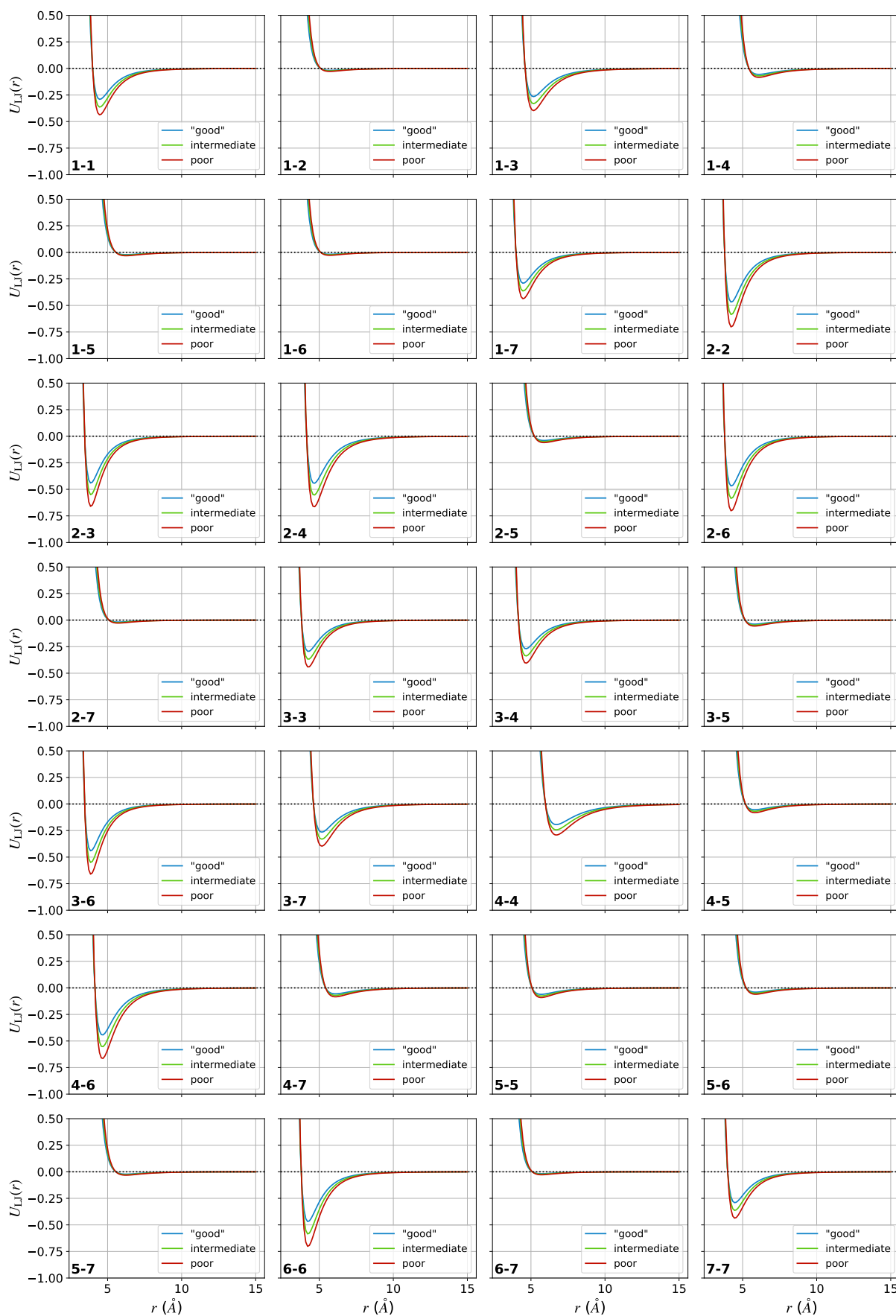
**Fig. S30** LJ non-bonded potentials (in kcal/mol) for all possible pairs of interactions in the CG P(NDI2OD-T2) model. The "good" solvent has parameters that are 20% weaker than those obtained in DCB, while the "poor" solvent interactions are 20% stronger. The as-parameterised model (from DCB) is labelled as "intermediate". The site types for the interactions are listed in the bottom left of each plot.

S14.3   Bonded parameters

For the bonds, two different potentials were used for the stretching parameters: a harmonic potential (eqn (S8a)), and a quartic potential that used the LAMMPS "class2" bond style (eqn (S8b)). Angles also made use of two different potential forms: a harmonic (eqn (S9a)) and a quartic potential (eqn (S9b)). All dihedrals were modelled with the dihedral style described in eqn (S10) with $m = 4$. The improper torsion was modelled with a cosine form (eqn (S11)).

In order to model the flexible backbone, the 1–7–3 and 7–1–3 angles, and the 3–7–1–3 dihedrals were reduced to 1% of their original stiffness (Fig. S31). The modified parameters are given in Tables S18 and S19. All other parameters are the same as for the regular backbone

(a) Dihedral 3–7–1–3   **(b)** Angle 1–7–3   **(c)** Angle 7–1–3
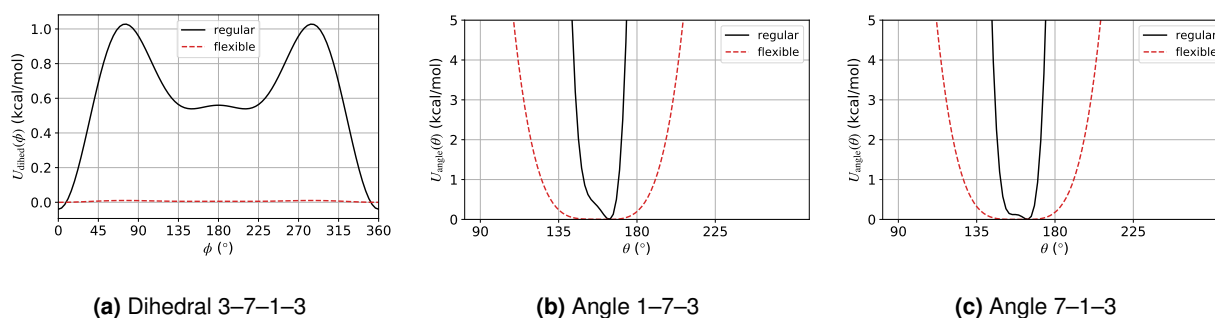
Fig. S31 (a) Dihedral and (b)–(c) angle potentials used to tune the flexibility of the polymer backbone from regular (black lines) to flexible (red dashed line). In all three cases, the coefficients were reduced to 1% of their original values.

Table S17 Bond length parameters for the CG model of P(NDI2OD-T2). Bond stretching coefficients k_i are in units of kcal/mol/Å^i. The forms of the harmonic and quartic bond styles are given in eqns (S8a) and (S8b), respectively. Note that the quartic bond potential is implemented in LAMMPS as the "class2" bond style.

bond	bond style	l_0 (Å)	k_2	k_3	k_4
1–3	harmonic	4.12	129.33		
1–7	harmonic	3.91	101.73		
2–3	harmonic	2.90	679.95		
2–4	harmonic	3.57	83.39		
3–3	harmonic	2.47	1206.79		
6–3	harmonic	2.89	525.52		
6–4	harmonic	3.57	80.73		
7–3	harmonic	4.12	138.60		
4–5	quartic	3.92	92.92	596.82	1014.01
5–5	quartic	3.86	88.96	556.24	890.46

Table S18 Bond angle parameters for the CG model of P(NDI2OD-T2). Angle bending coefficients k_i are in units of kcal/mol/rad^i. The parameters for the flexible backbone are also included (see Fig. S31). These replace the corresponding angle of the regular-flexibility backbone. The forms of the harmonic and quartic angle styles are given in eqns. (S9a) and (S9b), respectively.

angle	angle style	θ_0 (°)	k_2	k_3	k_4
1–3–2	quartic	138.60	62.00	−163.76	2.00
1–3–3	quartic	161.44	129.09	1062.34	2561.64
1–3–6	quartic	89.61	104.08	−1308.04	4778.21
1–7–3	quartic	164.05	60.76	426.09	967.28
2–4–5	quartic	110.48	5.69	−11.47	6.85
4–5–5	quartic	177.89	24.91	51.38	29.35
5–5–5	quartic	174.92	15.91	33.06	19.92

Continued on next page

angle	angle style	θ_0 (°)	k_2	k_3	k_4
6–4–5	quartic	112.21	5.88	−11.15	6.68
7–3–2	quartic	89.89	96.31	−1192.48	4517.46
7–3–3	quartic	161.55	108.44	946.50	2560.32
7–3–6	quartic	137.74	76.89	12.79	2.00
7–1–3	quartic	164.25	40.77	401.68	1106.43
3–2–4	harmonic	148.42	49.80		
3–6–4	harmonic	148.11	47.04		
1–7–3*	quartic	164.05	0.61	4.26	9.67
7–1–3*	quartic	164.25	0.41	4.02	11.06

* Flexible backbone parameters

Table S19 Dihedral parameters for the CG model of P(NDI2OD-T2). Energy parameters k_i are in units of kcal/mol. All dihedrals use the dihedral potential defined in eqn (S2) with $m = 4$ (implemented in LAMMPS as the "multi/harmonic" style). The parameters for the flexible backbone are also given (see Fig. S31). This replaces the corresponding dihedral of the regular-flexibility backbone.

dihedral	dihedral style	k_0	k_1	k_2	k_3	k_4
1–7–3–2	multi/harmonic	0.10	−0.11	−0.32	0.21	0.65
1–7–3–3	multi/harmonic	0.33	−0.44	−0.34	0.13	0.51
1–7–3–6	multi/harmonic	0.19	−0.51	0.38	−0.72	0.99
2–4–5–5	multi/harmonic	0.03	0.02	0.15	0.03	−0.02
3–2–4–5	multi/harmonic	0.46	−0.84	−0.31	0.56	0.34
3–6–4–5	multi/harmonic	0.16	−0.70	0.55	0.67	−0.32
3–7–1–3	multi/harmonic	0.96	0.48	−0.63	−0.76	−0.07
4–5–5–5	multi/harmonic	0.13	−0.26	0.10	0.12	−0.06
5–4–5–5	multi/harmonic	0.01	0.01	0.08	−0.00	0.04
6–4–5–5	multi/harmonic	0.03	0.05	0.01	−0.02	0.08
7–1–3–2	multi/harmonic	0.17	−0.30	−0.05	−0.74	1.37
7–1–3–3	multi/harmonic	0.42	−0.30	−0.71	−0.10	0.78
7–1–3–6	multi/harmonic	0.12	−0.14	−0.53	0.23	0.83
3–7–1–3*	multi/harmonic	0.010	0.005	−0.006	−0.008	−0.001

* Flexible backbone parameters

Table S20 Improper torsion parameters for the CG model of P(NDI2OD-T2). Energy parameter k is in units of kcal/mol. The form of the cosine style improper is given in eqn (S11).

improper	improper style	k	d	n
2–6–3–3	cosine	35.24	−1	2

References

[1] 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.

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

[3] W. L. Jorgensen and N. A. McDonald, *J. Mol. Struct., Theochem*, 1998, **424**, 145–155.

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

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

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

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

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

[9] L. Martínez, R. Andrade, E. G. Birgin and J. M. Martínez, *J. Comput. Chem.*, 2009, **30**, 2157–2164.

[10] W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695–1697.

[11] W. Shinoda, M. Shiga and M. Mikami, *Phys. Rev. B*, 2004, **69**, 134103.

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

[13] V. Boţan, V. D. Ustach, K. Leonhard and R. Faller, *J. Phys. Chem. B*, 2017, **121**, 10394–10406.

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

[15] R. Faller and D. Reith, *Macromolecules*, 2003, **36**, 5406–5414.

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

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

[18] 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.

[19] M. Rubinstein and R. H. Colby, *Polymer Physics*, Oxford University Press, 2003.

[20] A. Montesi, M. Pasquali and F. C. MacKintosh, *Phys. Rev. E*, 2004, **69**, 021916.

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

[22] R. A. Mansbach and A. L. Ferguson, *J. Phys. Chem. B*, 2017, **121**, 1684–1706.

[23] S. Chandrasekhar, *Rev. Mod. Phys.*, 1943, **15**, 1–89.

[24] B. Liu and B. Dünweg, *J. Chem. Phys.*, 2003, **118**, 8061–8072.

[25] J. G. Kirkwood and J. Riseman, *J. Chem. Phys.*, 1948, **16**, 565–573.

[26] M. Doi, S. F. Edwards and S. F. Edwards, *The Theory of Polymer Dynamics*, Clarendon Press, 1988.

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

[28] 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.

[29] W. M. Haynes, *CRC Handbook of Chemistry and Physics, 95th Edition*, CRC Press LLC, Oakville, United Kingdom, 2014.

[30] A. Marantan and L. Mahadevan, *Am. J. Phys.*, 2018, **86**, 86–94.

[31] N. E. Jackson, L. X. Chen and M. A. Ratner, *Proc. Natl. Acad. Sci. USA*, 2016, **113**, 8595–8600.

[32] J. A. Rackers, Z. Wang, C. Lu, M. L. Laury, L. Lagardère, M. J. Schnieders, J.-P. Piquemal, P. Ren and J. W. Ponder, *J. Chem. Theory Comput.*, 2018, **14**, 5273–5289.