Supplemental Information for "Selective solvent conditions influence sequence development and supramolecular assembly in step-growth copolymerization"

Ryan L. Hamblin, Nhu Q. Nguyen, and Kateri H. DuBay

S1 Simulation details

All simulations were performed in LAMMPS[1,2]. Additional specifics on the original model are described in Ref. [3]. Alterations specific to this work are described below.

Dimensionless Units All simulations used dimensionless LJ units within LAMMPS and are defined in terms of a chosen mass, m, length, σ, and energy, ϵ. From these parameters, a characteristic time scale $\tau \equiv \sqrt{\frac{m\sigma^2}{\epsilon}}$ is defined. In this work, we take $m = 200$ Da, $\sigma = 5$ Å, and $\epsilon = 100$ K $\cdot\, k_\mathrm{B} = 1.38 \times 10^{-}21$ J, resulting in $\tau = 7.75$ ps. For every monomer, the two linking particles have a mass of 0.25 m and the central particle has a mass of 0.5 m. The total monomer mass is then 1.0 m, which is roughly the mass of a benzene ring with a 9-carbon alkyl substituent. The temperature for all simulations was set to 300 K. All simulations were run within a $50\sigma \times 50\sigma \times 50\sigma$ cubic box with periodic boundary conditions. Each simulation contained 7200 monomers, yielding an overall density of $\rho = 0.0576\ \sigma^{-3}$.

Persistence length and the angular potential. Intramolecular angles are governed by the harmonic potential:

$$E_\mathrm{angle}(\theta_{ijk}) = K_{ijk}^\mathrm{angle}(\theta_{ijk} - \theta_0)^2, \tag{S1}$$

where θ_{ijk} is the angle between particles i, j, and k, θ_0 is the value at the minimum of the angular potential, and the constant, K_{ijk}^angle, is the spring constant for that angle. For the intramonomer angle, **2**–**1**–**2**, $\theta_0 = 180°$ and K_{212}^angle was set to 5 $\epsilon \cdot \mathrm{rad}^{-2}$ for flexible chains and 50 $\epsilon \cdot \mathrm{rad}^{-2}$ for stiff chains. Previous work with this model has determined that these values correspond to persistence lengths of $l_p = 3.5$ for flexible chains and $l_p = 16.5$ for stiff chains.[3] For the intermonomer angle, **1**–**2**–**2′**, $\theta_0 = 180°$ and $K_{122'}^\mathrm{angle} = 100\ \epsilon \cdot \mathrm{rad}^{-2}$ in all cases.

Langevin dynamics Langevin dynamics was implemented in LAMMPS using the "fix langevin" command. A *damp* parameter is used to control the diffusion rate and the relaxation rate of the temperature, which is discussed in detail in the Supplemental Information of Ref. [3]. In the simulations in this work, the *damp* parameter is set to 0.1 τ, resulting in a viscosity of $\gamma = 0.1$ mPa $\cdot$ s. For further discussion of this parameter, see Ref. [4].

S2 Attractive interaction strengths and Lorentz-Berthelot combining

After choosing ε_AA and ε_BB as described in Methods, we utilized Lorentz-Berthelot combining rules[4] to define an attraction strength between our more and less solvophobic monomers, **A** and **B**, respectively, setting $\varepsilon_\mathrm{AB} = \sqrt{\varepsilon_\mathrm{AA}\varepsilon_\mathrm{BB}}$. As we are utilizing dimensionless, reduced unit values for σ_AA and σ_BB, this expression is also equivalent to Kong combining rules.[5] Table S1 shows all combinations of the ε_att values explored in this work.

ε_{AA} ($k_{\mathrm{B}}T$)	ε_{AB} ($k_{\mathrm{B}}T$)	ε_{BB} ($k_{\mathrm{B}}T$)
1.0	0.32	0.1
1.0	0.55	0.3
1.0	0.71	0.5
0.75	0.27	0.1
0.75	0.47	0.3
0.75	0.61	0.5
0.5	0.22	0.1
0.5	0.39	0.3
0.5	0.5	0.5

Table S1: Lennard-Jones attractions. Columns show $\varepsilon_{\mathrm{AA}}$, $\varepsilon_{\mathrm{AB}}$, and $\varepsilon_{\mathrm{BB}}$, which are the well-depths for the attractive portion of the LJ interactions (see Eq. 1 in the main text) for **A** to **A**, **A** to **B** and **B** to **B** monomers, respectively. Values for $\varepsilon_{\mathrm{AA}}$ and $\varepsilon_{\mathrm{BB}}$ were chosen to reflect a copolymerization in which monomer species **A** is more solvophobic than species **B**. From the values of $\varepsilon_{\mathrm{AA}}$ and $\varepsilon_{\mathrm{BB}}$, $\varepsilon_{\mathrm{AB}}$ values were calculated according to Lorentz-Berthelot rules.[4] For all combination of attraction strengths, the strength of repulsive interactions, $\varepsilon_{\mathrm{rep}}$, is kept at a constant value of 1.33 $k_{\mathrm{B}}T$, irrespective of monomer identity.

S3 Activation energies

To ensure that having a fixed value for $\varepsilon_{\mathrm{rep}}$, plus the additional soft repulsive potential between type **2** particles, was maintaining a consistent activation energy across monomer type pairs for our range of simulation parameters, we tabulated the activation energies for successful bond formations for a set of ~ 52000 bonding events drawn from the full range of simulation parameters explored in this work. Activation energies are calculated from the total potential energy of the two bonding monomers arising from the LJ potential between type **1** particles, the soft repulsive potential between type **2** particles, and the intramonomer bond lengths and angles. The values of each of these contributions above their minimum energy are summed at the time of bond formation to obtain the total activation energy for the bonding event. Figure S1 shows the results of this comparison, broken down by monomer type pairs, demonstrating consistent activation energies across monomer type pairing.

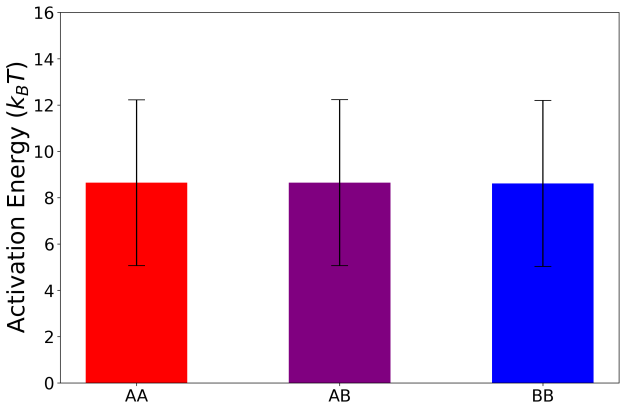

Figure S1: Observed activation Energies. The average activation energies for bond formation events of **A** to **A**, **A** to **B**, and **B** to **B** monomer pairs for the full range of simulation parameters explored. Error bars represent the standard deviation in the recorded values for each monomer type pairing.

S4 Calculation of the first coordination number from g(r)

In liquids the first coordination number is commonly defined in terms of the radial distribution function, $g(r)$, and is taken as the number of atoms within the first coordination shell.[6] With this definition, we calculate the first coordination number for a given monomer species combination as:

$$n_1^{ij} = 4\pi\rho \int_0^{r'} r^2 g_{ij}(r) dr, \tag{S2}$$

where ρ is the number density of the system, r is the radial distance from the center of a fixed reference particle, and r' is the value of r at which $g_{ij}(r)$ reaches the first minimum after the initial peak. Indices i and j correspond to the possible monomer species pairs within the system, namely **AA**, **AB**, and **BB**.

S5 Chain length and block length statistics

In this section we derive a prediction for the block length distribution among the oligomers based on a Markovian statistical model.

We begin with the probability distribution for chain lengths in a step-growth polymerization, which is given by the Flory-Schulz distribution.[7] That is, the probability of finding a chain of length n is given by

$$P(n) = (1-p)p^{n-1}, \tag{S3}$$

where p is the reaction extent, or the probability that a given monomer has reacted. Distributions calculated from Equation S3 are shown in Fig. 3 b&c and Fig. 6a in the main text for comparison to the observed distributions of oligomer chain lengths.

Under Flory's assumption of equal reactivity, the probability of having a like or unlike bond is only based on the identities of the two bonding monomers, as reactivity does not depend on the length or properties of a molecular chain attached to either of the two reacting monomers. This dependence of reactivity and bond formation on only the identities of the most recent monomer and next monomer to be added means that the sequence can be treated as a Markov chain. To obtain an expression for the probability distribution of contiguous blocks of either **A** or **B**, we define analogous equations to Equation S3, but specific to **A** or **B**. Just as p in Eq. S3 is the probability that a given monomer has reacted with any other given monomer, we now define p_A as the probability that a given monomer of **A** has reacted with another **A** monomer. Similarly, p_B the probability that a given monomer of **B** has reacted with another **B** monomer. Then, for the distributions of **A** blocks of length n_A and **B** blocks of length n_B, we obtain:

$$\begin{aligned} P(n_\mathrm{A}) &= (1-p_\mathrm{A})p_\mathrm{A}^{n_\mathrm{A}-1}; \\ P(n_\mathrm{B}) &= (1-p_\mathrm{B})p_\mathrm{B}^{n_\mathrm{B}-1}. \end{aligned} \tag{S4}$$

The resulting Markovian expression, Equation S4, is used as a point of comparison for our observed sequence behavior in Fig. 4e-g and Fig. 6b, as discussed in the main text.

From the Flory-Schulz distribution it is possible to derive a corresponding expression for the system dispersity, $Đ$, for a given reaction extent p, which also provides an important comparison for the simulation results. We start with $Đ$ defined as:

$$Đ = \frac{M_w}{M_n}, \tag{S5}$$

which is the ratio of the weight-averaged molar mass, M_w, to the number-averaged molar mass, M_n. These are each obtainable from the Flory-Schulz distribution, $P(n)$ in Equation S3, using the definitions[8,9]:

$$M_n = M_0 \sum_n nP(n); \tag{S6}$$

$$M_w = \sum_n w_n M_n = M_0 \sum_n w_n nP(n). \tag{S7}$$

Here w_n is the weight-fraction of n-mers given by $w_n = \frac{(nM_0)N_n}{M_0 N_0} = \frac{nN_n}{N_0}$, for N_n total n-mers and N_0 initial monomers with a molar mass of M_0. From these definitions the following identities are used:

$$\sum_x x p^{x-1} = (1-p)^{-2} \text{ for } p < 1; \tag{S8}$$

$$\sum_x x^2 p^{x-1} = (1+p)(1-p)^{-3} \text{ for } p < 1. \tag{S9}$$

Using the expressions in Equation S8 and S9, the mass-averaged and weight-averaged molar mass expressions reduce to:

$$M_n = M_0 \frac{1}{1-p}; \tag{S10}$$

$$M_w = M_0 \frac{1+p}{1-p}. \tag{S11}$$

Finally, combining Equations S10 and S11 with the definition of $Đ$ in Equation S5 we obtain the expression:

$$Đ = 1 + p. \tag{S12}$$

The resulting expression in Equation S12 is plotted vs. p in Fig. 3a in the main text and is compared to the observed dispersities in our simulations.

For the derivation of Equation S12 we made use of the Flory-Schulz distribution for the entire chain given in S3. To obtain an expression for the expected dispersity in block lengths, we replace the Flory-Schulz distribution for chain lengths with the expected distribution of block lengths using a Markovian model, namely $P(n_\mathrm{A})$ or $P(n_\mathrm{B})$ as defined in Equation S4. This leads to the following expressions for block length dispersity:

$$\begin{aligned} Đ_A &= 1 + p_\mathrm{A}; \\ Đ_B &= 1 + p_\mathrm{B}. \end{aligned} \tag{S13}$$

The predicted block dispersities from Equation S13 are utilized to examine the deviation from Markovian behavior in our system, as described in Section S6 and shown in Fig. S2a below.

S6 Quantifying deviations from Markovian block statistics

To quantitatively explore the deviation from Markov statistics in our system, we compared the Markovian predictions for dispersity and block length distribution with the results of our simulations. We determined the mean standard error (MSE) between the observed and predicted block dispersities (Equation S13) and calculated the Wasserstein metric [10,11], W_p, as a measure of the statistical distance between the Markovian distributions and our observed block length distributions. Fig. S2 shows the results of each of these metrics, as a function of $\Delta\varepsilon$. With the increase in attraction asymmetry between **A** and **B** species, characterized by $\Delta\varepsilon$, the difference between the Markovian distribution and the observed distribution of **A** block lengths also increases. This deviation demonstrates the non-Markovian nature of **A**-block formation under conditions in which attractive interactions drive monomer self-assembly. Notably, however, the Markovian distribution is a better match to the observed distribution of **B** block lengths, as W_p for the **B** blocks is lower than that for the **A** blocks and also largely independent of $\Delta\varepsilon$. Indeed, the dispersities and length distributions of the **B** blocks are still well described by Markovian chain statistics for each interaction strength combination explored in this work.

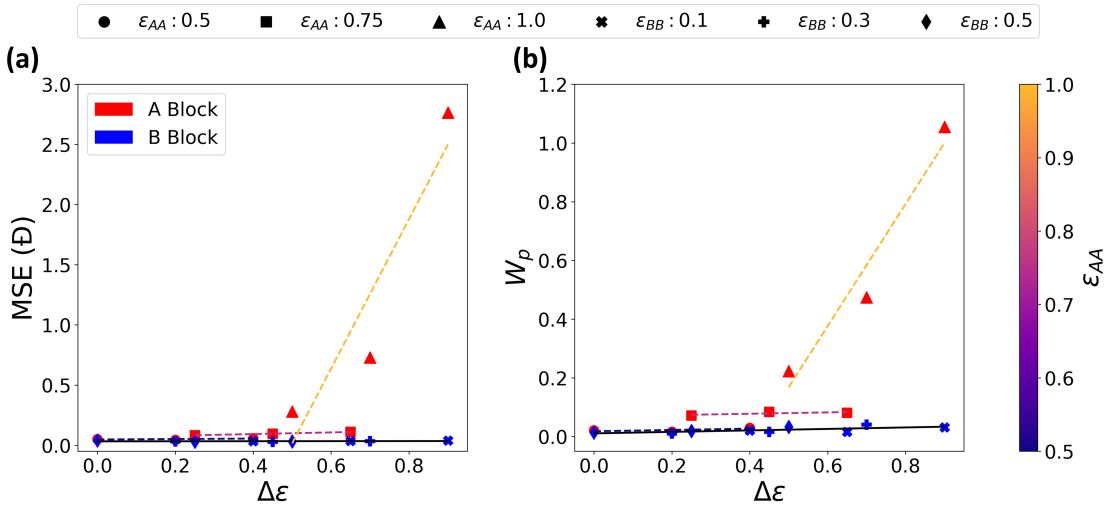

Figure S2: Error metrics for the predicted and observed block dispersities and block length distributions. (a) The mean standard error (MSE) between the observed dispersity, $Đ$, in block lengths, and the associated prediction for block length $Đ$ based on Markovian statistics. (b) The Wasserstein metric, W_p, a statistical distance between our observed block distribution and the Markovian distribution. In both (a) and (b), the block type, **A** or **B**, is shown by marker coloration. The marker shape indicates the value of attractions – $\varepsilon_{\mathrm{AA}}$ for **A** blocks and $\varepsilon_{\mathrm{BB}}$ for **B** blocks. Dashed lines show trends in the values for **A** blocks for cases with equal $\varepsilon_{\mathrm{AA}}$, with the value of $\varepsilon_{\mathrm{AA}}$ indicated by line coloration. The black line is the trend in **B** block values for all simulation parameters. Observed distributions were collected from three independent trials for each simulation parameter set.

S7 Calculating the local nematic order parameter

To calculate the local nematic order parameter, we begin with the standard definition of a nematic order parameter:

$$\overline{S} = \langle \frac{3\cos^2\theta - 1}{2} \rangle, \tag{S14}$$

where θ is the angle between the orientation of a single monomer and the director, namely the average orientation of all monomers in the system. Since we wish to quantify the degree of alignment within each aggregate rather than across the entire simulation, we calculate the angle θ to be the angle between the orientation of one monomer and the average orientation of all monomers within the local environment, defined as monomers within a distance of 2.5σ. Moreover, since we are interested in the alignment within an aggregate, we only include monomers in the overall ensemble average if there are at least 12 neighboring monomers within the 2.5σ distance. Our final expression for the local order parameter is therefore given by:

$$\overline{S_{\text{local}}} = \langle \frac{3\cos^2\theta - 1}{2} \cdot \mathbf{1}_{\text{aggregates}} \rangle, \tag{S15}$$

which takes the standard order parameter in Equation S14 (using the localized θ) and multiplies by an indicator function, $\mathbf{1}_{\text{aggregates}}$, that takes the value 1 if the neighboring monomer criteria is met and 0 otherwise.

Fig. S3 shows the value of this order parameter for flexible chains throughout the reaction at each of the attractive interactions explored in this work. Flexible polymers do not show significant nematic ordering regardless of attraction strength. Thus, the peak value for $\overline{S_{\text{local}}}$ observed in Fig. S3 is used a threshold value for nematic ordering in stiff chains, as shown in Fig. 5 and described in the main text.

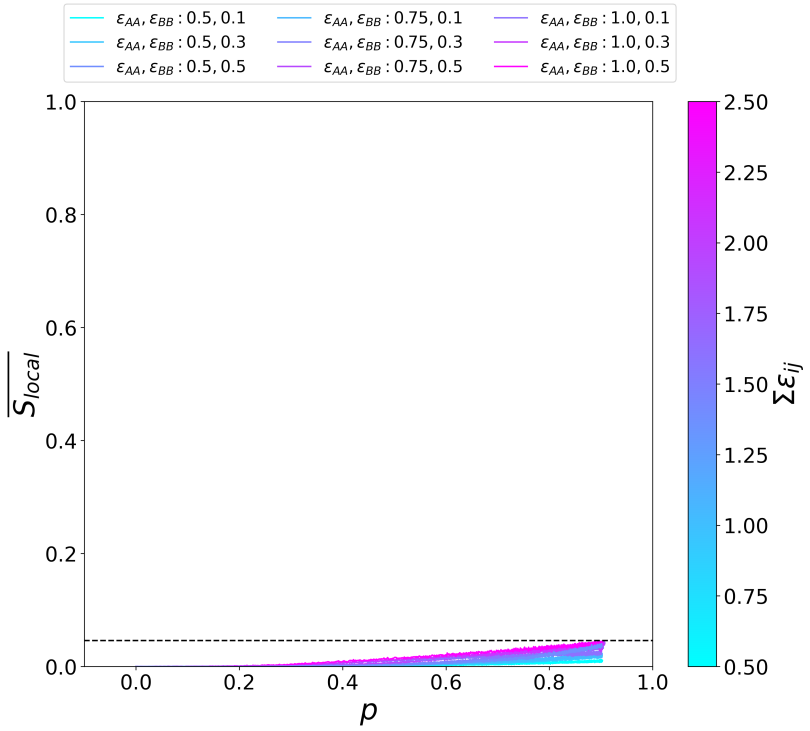

Figure S3: Local nematic order parameter for flexible chains. Local nematic order parameter, $\overline{S_{\text{local}}}$, as a function of reaction extent, p, over the range of attraction strengths explored for flexible-chain, $l_p = 3.5$, polymers. The dashed horizontal line indicates the highest value for $\overline{S_{\text{local}}}$ seen in this case. This line is reproduced in Fig. 5 in the main text.

S8 Polymerized Fraction

Asymmetric attractive interactions produce different rates of polymerization between monomer species. The fraction of each monomer type that has undergone polymerization depends on the balance of these interactions as well as the development of the emergent aggregate phase. We simulate a fixed quantity of monomers with an equal amount of **A** and **B**. Without a continuous reactant feed, the polymerized fractions of the monomer species are also impacted by the total availability of each monomer species. This impact is particularly notable late in the reaction, where the more rapidly polymerizing monomers near complete conversion. As such, the polymerized fractions of the two monomer species varies in a complex fashion throughout the reaction. This variation can be seen in Fig. S4 below.

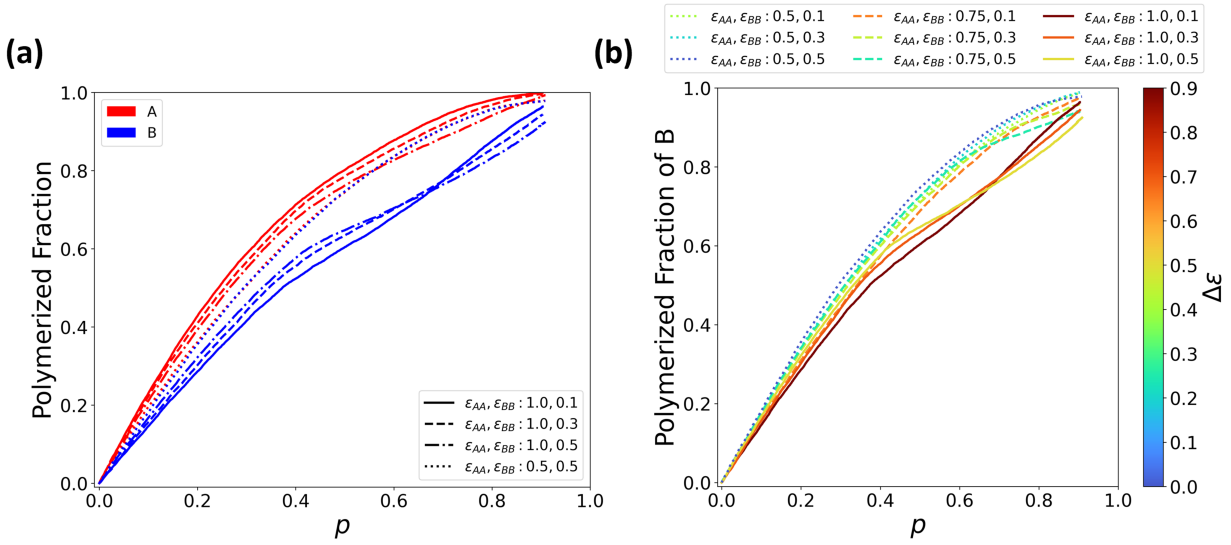

Figure S4: Polymerized fraction of monomer species. (a) The fraction of **A** and **B** monomers which have polymerized as a function of the reaction extent, p, for stiff chains of $l_p = 16.5$ and select attraction values. Monomer species is indicated by line color, and line styles indicate the set of attractive interactions. (b) The fraction of **B** monomers which have polymerized as a function of the reaction extent, p, for stiff chains of $l_p = 16.5$ at all attraction combinations. Color indicates the degree of attraction asymmetry, $\Delta\varepsilon$, and line styles indicate $\varepsilon_{\mathrm{AA}}$ value. All data was obtained from three independent simulation trials for each parameter set. In each case, similar results were observed for flexible chains.

S9 Supplemental Kinetics and Length Distributions

Additional figures providing kinetic parameters, chain and block length distributions, and characteristic snapshots of the final system state for select ε and l_p parameter sets are shown in Fig. S5 and Fig. S6 below.

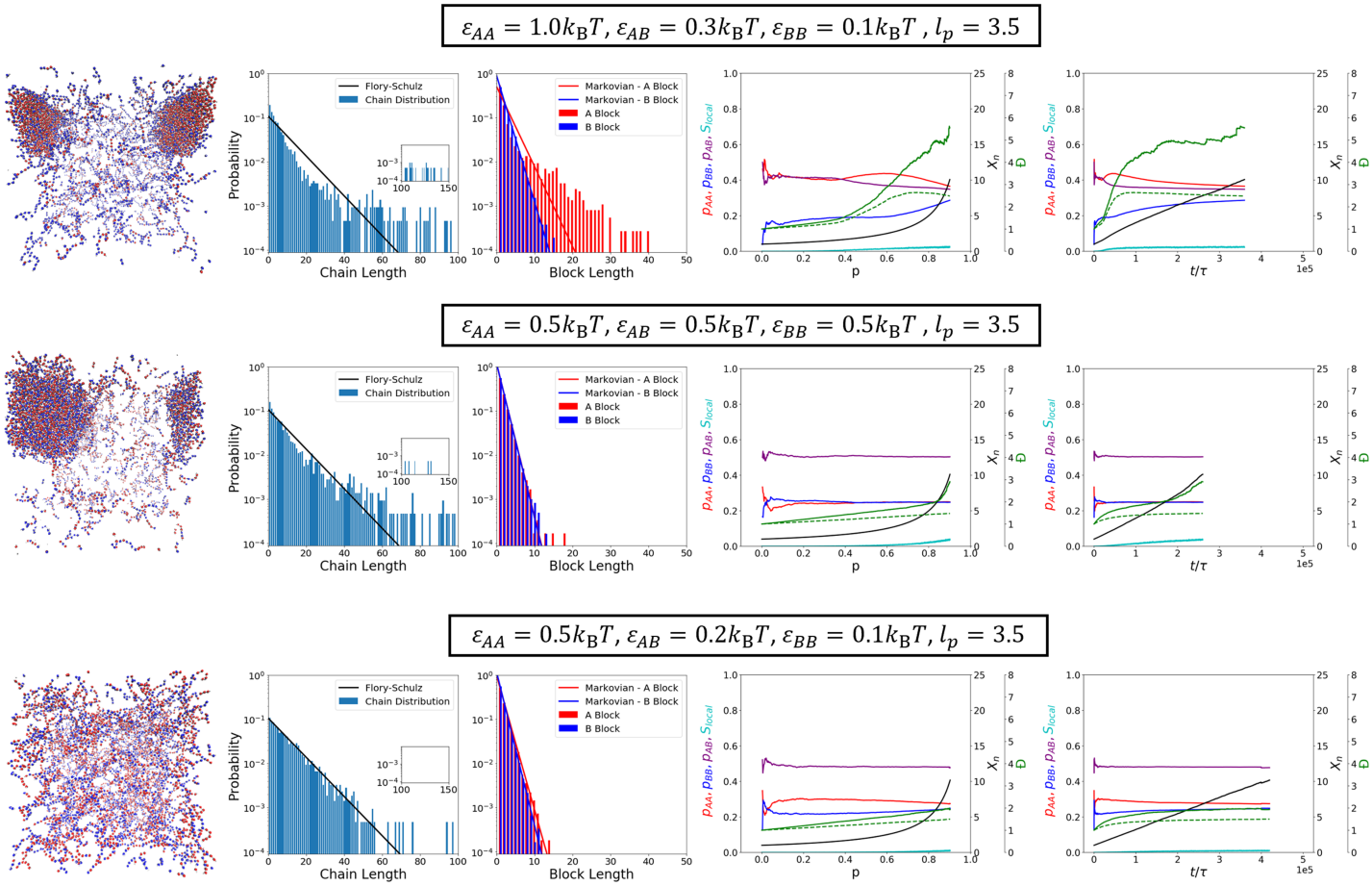

Figure S5: Length distributions and kinetic parameters for flexible-chain copolymerizations. Columns from left to right: the final system snapshot, chain length distribution, block length distributions, and several kinetic parameters versus reaction extent and simulation time. The chain length and block length distributions are represented as histograms, with the predicted distributions from Flory-Schulz and Markovian statistics plotted as lines. Kinetic plots include the pairwise bonding probabilities (p_{AA}, p_{AB}, and p_{BB}), $\overline{S_{\text{local}}}$, X_n, and $Đ$. For $Đ$ a dashed line is calculated for blocks, while the solid line is calculated for the entire chain. Rows represent different attraction strengths, as indicated, with flexible chains of $l_p = 3.5$. All data is collected from three independent simulation trials at each parameter set.

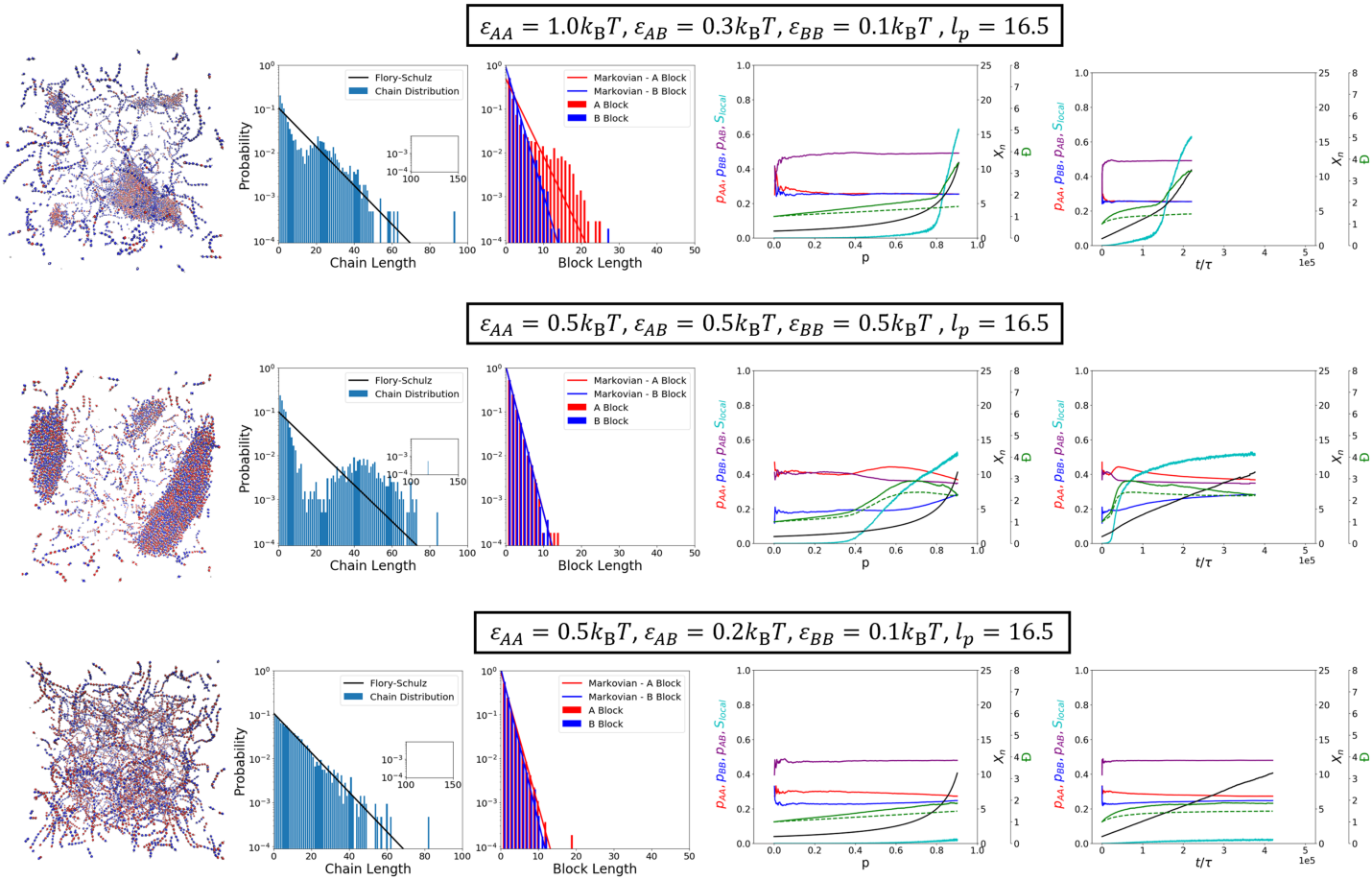

Figure S6: Length distributions and kinetic parameters for stiff-chain copolymerizations. Columns from left to right: the final system snapshot, chain length distribution, block length distributions, and several kinetic parameters versus reaction extent and simulation time. The chain length and block length distributions are represented as histograms, with the predicted distributions from Flory-Schulz and Markovian statistics plotted as lines. Kinetic plots include the pairwise bonding probabilities (p_{AA}, p_{AB}, and p_{BB}), $\overline{S_{\text{local}}}$, X_n, and $Đ$. For $Đ$ a dashed line is calculated for blocks, while the solid line is calculated for the entire chain. Rows represent different attraction strengths, as indicated, with stiff chains of $l_p = 16.5$. All data is collected from three independent simulation trials at each parameter set.

References

[1] S. J. Plimpton, Journal of Computational Physics, 1995, **117**, 1–19.

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

[3] Z. Zhang and K. H. DuBay, The Journal of Physical Chemistry B, 2021, **125**, 3426–3437.

[4] H. A. Lorentz, Annalen der Physik, 1881, **248**, 127–136.

[5] C. L. Kong, The Journal of Chemical Physics, 1973, **59**, 2464–2467.

[6] D. Chandler, Introduction to Modern Statistical Mechanics, Oxford University Press, 1987, p. 200.

[7] P. J. Flory, Journal of the American Chemical Society, 1936, **58**, 1877–1885.

[8] P. J. Flory, Principles of Polymer Chemistry, Cornell University Press, 1953.

[9] R. Young and P. Lovell, Introduction to Polymers: Third Edition, CRC Press, United States, 3rd edn, 2011.

[10] R. L. Dobrushin, Theory of Probability & Its Applications, 1970, **15**, 458–486.

[11] S. S. Vallender, Theory of Probability & Its Applications, 1974, **18**, 784–786.