Appendix A Numerical Approach for Chain Statistics of Self-Associating Chains at Infinite Dilution in θ-Solvent

A.1 Model Description

Our objective is to determine the size of a linear chain of N monomers, f of which are modified to act as stickers capable of forming pair-wise only, physical associations. The stickers are assumed to be equidistantly spaced l monomers apart along the chain, and the energy of association is εkT. We will assume Gaussian chain statistics for any segment of the chain whose configuration is unrestrained by reversible crosslinks.

To calculate the size of the chain in the very dilute regime (all associations are intramolecular), we define a semi-Markov process X(t), $t > 0$ such that each state i is fully specified by identifying which pairs of stickers form bonds. (Note here that a given state has an infinite number of chain configurations.) The chain goes from one state to the next by either breaking or forming a bond, as illustrated in Figure A.1.

This semi-Markov process is completely specified if we know both (a) the distribution of times T_i the chain spends in any given state i and (b) the probabilities P_{ij} that once it leaves state i it next enters state j. Thus, if we can determine the distributions of T_i, compute the transition probabilities P_{ij}, and calculate relevant properties of the chain in any state i, then we can estimate the long-run average of chain properties such as size by simply running the Markov process for a sufficiently long time. Fortunately, although the total number of states is extremely large for f as small as 20, the number of states that are *accessible in one step* from any given state is much more manageable, that is, $P_{ij} = 0$ for most values of j for a given state i.

Assume the polymer chain enters state i at time t. Clearly, the state it enters next is determined by which bond is broken or formed first; and the time spent in the present state is the time it took for that bond-breaking or bond-forming event to take place. Because the times for bond breaking and bond forming are random variables, in order to solve the problem we need to determine the distribution of the breaking time and the forming time of all the possible bonds, for any state i.

Consider the breaking of a single bond. A bond that has been "alive" for any given time s is just as strong as a bond which has just formed; in other words the expected time to break a bond given that is has been "alive" for time s is independent of s. Accordingly, bond breaking is a memoryless (hence exponential) process, and the time to break a bond is given by the exponential density function:

$$f_{T_b}(t) = \frac{1}{\tau_b}\exp\left(-\frac{t}{\tau_b}\right) \qquad t \geq 0$$

where τ_b is the expected time to break a bond, and has the same value for all bonds. Bond forming can also be argued to be an exponential process, as discussed below, with expected time τ_f dependent on the number of monomers $L_{qq'}$ in the shortest connected path between the two free stickers q and q'. Observe that the shortest path $L_{qq'}$ between stickers depends on the pair of stickers q and q' and on the current state of the chain; for example, the shortest distance between stickers 1 and 6 in state (b) of Figure A.1 is $L_{1,6} = 2l$.

Let's now see if we can determine expressions for τ_b and τ_f. Rubinstein and Semenov[1] give $\tau_b = \tau_0 \exp(\varepsilon + \varepsilon_a)$, where τ_0 is the monomer relaxation time and $\varepsilon_a kT$ is a potential barrier for bond breaking and also the activation energy for bond forming. I get $\tau_f \cong \tau_0/p_c p_s$ to a first approximation (as discussed below), where $p_c = (6/\pi L^3)^{1/2}$ is the contact probability that the two stickers separated by L monomers be within distance b of each other (b is the Kuhn length, assumed to be the maximum distance over which the stickers can associate), and p_s is the sticking probability that a bond is formed at any given "visit." If $p_s \approx (V_b/b^3)\exp(-\varepsilon_a)$ where V_b is the bond volume, then we get $\tau_f \approx \alpha\,\tau_0 \exp(\varepsilon_a)\, L^{3/2}$ where $\alpha = (b^3/V_b)(\pi/6)^{1/2}$.

Let $\{pp'\}_i$ be the set of all pairs of stickers that are bonded in state i, and $\{T_{b,pp'}\}_i$ be the set of random variables corresponding to their breaking time. Among the unpaired stickers in state i, let $\{qq'\}_i$ be the set of all possible pairings for formation of a new bond, and $\{T_{f,qq'}\}_i$ be the set of random variables corresponding to their expected bond-forming time. For a given state i, then, we have independent, exponential random variables $\{T_{b,pp'}\}_i$ and $\{T_{f,qq'}\}_i$, with expected values $E[T_{b,pp'}] = \tau_b$ and $E[T_{f,qq'}] = \alpha\,\tau_0 \exp(\varepsilon_a)\,(L_{qq'[i]})^{3/2}$, where $L_{qq'[i]}$ is the number of monomer in the shortest connected path between the two free stickers p and p' in state i. The probability that the next state is achieved by breaking a specific bond pp', or by

forming a specific bond *qq'* is the probability that $T_{b,pp'}$ or $T_{f,qq'}$ is the shortest of all the breaking and forming times. For independent exponential random variables this probability is (the rate of the given exponential variable)/(the sum of all the rates). Thus, the probability that the next state is obtained from forming a bond between any two free stickers *q* and *q'* is:

$$P = \frac{\alpha^{-1} e^{\varepsilon} L_{qq'[i]}^{-3/2}}{Q + \alpha^{-1} e^{\varepsilon} \cdot \sum_{qq'} L_{qq'[i]}^{-3/2}}$$

where *Q* is the number of bonds in the present state *i* and the sum is over all the possible pairs of unpaired stickers in the present state. The probability that the next state is obtained from breaking a given bond is:

$$P = \frac{1}{Q + \alpha^{-1} e^{\varepsilon} \cdot \sum_{qq'} L_{qq'[i]}^{-3/2}} .$$

Note that these probabilities are independent of the activation energy. Given that all the breaking times and forming times are independent exponential random variables, the time T_i to transition from any state *i* to the next is also exponentially distributed, with rate equal to 1/(sum of all the rates), so with mean:

$$\tau_i = \frac{\tau_b}{Q + \alpha^{-1} e^{\varepsilon} \cdot \sum_{qq'} L_{qq'[i]}^{-3/2}} .$$

In this case (exponentially distributed transition times), the semi-Markov process is a continuous-time Markov chain. Continuous-time Markov chains are characterized by the Markovian property that, given the present state, the future is independent of the past. This result is intuitive for the bond-breaking and bond-forming processes which determine the time-evolution of our chains: given the present state, the next state and the time to transition into the next state are both independent of what states were visited previously or how long the chain has been in the present state already. As a result of the memoryless property of a continuous-time Markov chain, the amount of time T_i spent in state *i*, and the next state visited are *independent* random variables. The reader is referred to a text by Ross[2] for an excellent description of Markov chains, semi-Markov chains, and continuous-time Markov chains.

A.2 Distribution Function for the Time to Form a Bond

Consider a strand of L monomers with 2 stickers at its ends. What is the probability density function of the random variable T_f, the time for the stickers to form a bond (given that they are not bonded at the present time)?

Clearly there is an infinite number of configurations for the strand, where a configuration is specified by specifying the position vectors $\mathbf{r_1}$, $\mathbf{r_2}$,..., $\mathbf{r_L}$ of all the monomers. However, if we break up space into a 3D lattice with arbitrarily small but finite volume elements, then there is now a finite number of strand configurations. If we further define a macroscopic time τ_0 over which the polymer configuration does not change (and renormalize time in units of τ_0), then the configurations the strand takes over time constitute an irreducible (all the states communicate), positive recurrent (the expected time to return to the present state is finite for all the states) Markov chain. Therefore, there exist stationary probabilities $\pi_v = \lim_{n\to\infty} P_{uv}^n$ independent of initial state u for all states v.

Given a probability density function for the initial polymer configuration, there exists a distribution function for T_f, the time it will take for the stickers to bond *given* that they are not bonded at the present time $t = 0$. If the probability density of the initial chain configuration is the stationary probability density, and *given* that the sticker ends have not bonded after time s, the probability density to find the chain in any given configuration at time s is still equal to the stationary probability density. Therefore, nothing has changed after time s, so that the remaining time it will take to form a bond is independent of s. Accordingly, bond forming is a memoryless (hence exponential) process, and the time to break a bond is given by:

$$f_{T_b}(t) = \frac{1}{\tau_b}\exp\left(-\frac{t}{\tau_b}\right)$$

where τ_f is the expected value of T_f. To determine τ_f for a strand of L monomers with stickers at its ends, we need to know the time it takes for the stickers to come within close enough distance of each other to associate. Assume stickers form a bond with sticking probability p_s if they come within distance b (= Kuhn length) of each other. Consider another semi-Markov process with the following two states only: the strand is in state 1 if the stickers are within distance b of each other, and in state 2 otherwise. Let μ_1 and μ_2 be the mean times spent in states 1 and 2, respectively (we do not need to know the distribution of transition times). For

$\varepsilon = 0$ (corresponding to a non-associating strand of L monomers), the long-run fraction of time spent in state 1 is:

$$\frac{\frac{1}{2}\mu_1}{\frac{1}{2}\mu_1 + \frac{1}{2}\mu_2} = p_c = \left(\frac{6}{\pi \cdot L^3}\right)^{1/2}$$

where p_c is the contact probability for the chain ends to be within distance b of each other, and the only assumption is that of Gaussian statistics. But the mean time spent in state 1 is $\mu_1 \approx \tau_0$, from which we get $\mu_2 \approx \tau_0 (1-p_c)/p_c$. Note here that the above expression for μ_1 is also valid when $\varepsilon \neq 0$ if stickers fail to stick while the strand is in state 1, and that the above expression for μ_2 is likewise also valid when $\varepsilon \neq 0$. By conditioning on the present state of the chain, the expected time to form a bond is:

$$\begin{aligned} \mathrm{E}\left[\mathrm{T}_f\right] &= \mathrm{E}\left[\mathrm{T}_f \middle| 1\right]\mathrm{P}_1 + \mathrm{E}\left[\mathrm{T}_f \middle| 2\right]\mathrm{P}_2 \\ &= \mathrm{E}\left[\mathrm{T}_f \middle| 1\right] p_c + \left(\mu_2 + \mathrm{E}\left[\mathrm{T}_f \middle| 1\right]\right)(1 - p_c) \\ &= \mu_2 (1 - p_c) + \mathrm{E}\left[\mathrm{T}_f \middle| 1\right]. \end{aligned}$$

We find the expected time to bond given that the strand is in state 1, $\mathrm{E}[\mathrm{T_f}|1]$, by conditioning on whether the stickers stick the first time around:

$$\mathrm{E}\left[\mathrm{T}_f \middle| 1\right] \approx p_s \tau_0 + \left(\mu_1 + \mu_2 + \mathrm{E}\left[\mathrm{T}_f \middle| 1\right]\right)(1 - p_s)$$

where $\mathrm{p_s} \approx (\mathrm{V_b}/\mathrm{b}^3) \exp(-\varepsilon_a)$ is the probability that the stickers form a bond while the chain segment is in state 1. So after rearranging:

$$\mathrm{E}\left[\mathrm{T}_f \middle| 1\right] \approx \frac{\mu_1 + \mu_2}{p_s} - \mu_2 .$$

Substituting, we get: $$\mathrm{E}\left[\mathrm{T}_f\right] \approx \frac{\tau_0 (1 - p_c)(1 - p_c p_s)}{p_c p_s} \approx \frac{\tau_0}{p_c p_s}.$$

A.3 Simplifying Assumptions of the Model

We made two simplifying assumptions in our construction of the model. First, we have assumed Gaussian statistics for any strand whose configuration is unrestrained by reversible crosslinks. This assumption is reasonable in θ-solvents only to the extent that congestion is not an issue. In reality there is a limit to the number of monomers that can be collapsed into

a given volume, and for a high-enough number density of paired stickers and long-enough chains, congestion is expected to become important.

Second, in our derivation of the time for bond formation we assumed that the initial configuration of the strand between the stickers was chosen according to the stationary probabilities π_u, where a state u corresponds to a specific configuration of the strand. Justification for this assumption is derived from the fact that the probability density of the strand configuration *will* reach the stationary probability density, for any arbitrary initial probability density, after a sufficiently long time during which the stickers do not pair up. The assumption however presents limitations for a pair of stickers right after their bond is broken: due to spatial proximity, these may reassociate before the strand between them reaches its stationary probability density. In other words, the memoryless Markovian property may be violated to that extent in that the future is not completely independent of the past.

A.4 References and Notes

1. Rubinstein, M.; Semenov, A. N., Dynamics of entangled solutions of associating polymers. *Macromolecules* **2001,** 34, (4), 1058–1068.

2. Ross, S. M., *Introduction to Probability Models*. 8th ed.; Academic Press: 2003.

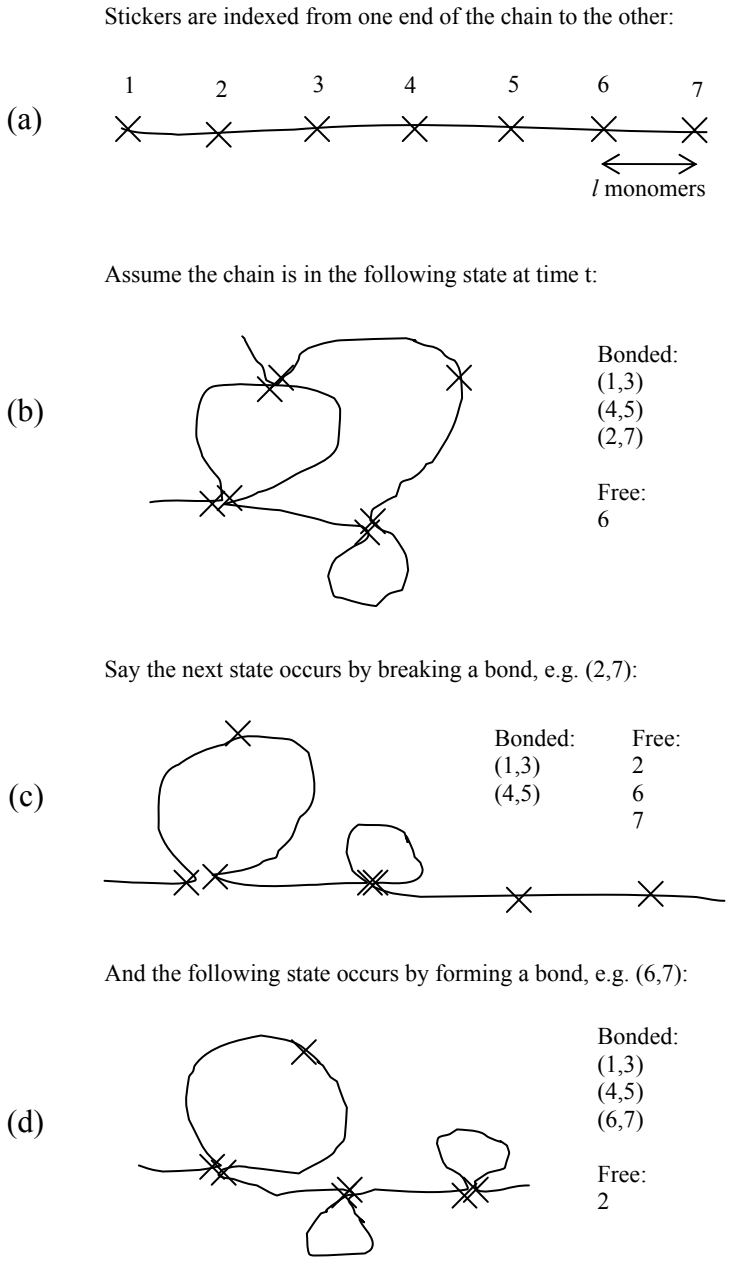

Figure A.1 Schematic illustration of the transition from one state of the chain to another by bond breaking and bond forming, for a chain with $f = 7$ stickers.