A general theory of kinetics and thermodynamics of steady-state copolymerization

Yao-Gen Shu\textsuperscript{1,2}, Yong-Shun Song\textsuperscript{2}, Zhong-Can Ou-Yang\textsuperscript{1} and Ming Li\textsuperscript{2}

\textsuperscript{1} State Key Laboratory of Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, Zhong Guan Cun East Street 55, PO Box 2735, Beijing 100190, People’s Republic of China
\textsuperscript{2} School of Physics, University of Chinese Academy of Sciences, No 19A Yuquan Road, Beijing 100049, People’s Republic of China

E-mail: liming@ucas.ac.cn

Received 9 January 2015, revised 2 April 2015
Accepted for publication 16 April 2015
Published 20 May 2015

Abstract

Kinetics of steady-state copolymerization has been investigated since the 1940s. Irreversible terminal and penultimate models were successfully applied to a number of comonomer systems, but failed for systems where depropagation is significant. Although a general mathematical treatment of the terminal model with depropagation was established in the 1980s, a penultimate model and higher-order terminal models with depropagation have not been systematically studied, since depropagation leads to hierarchically-coupled and unclosed kinetic equations which are hard to solve analytically. In this work, we propose a truncation method to solve the steady-state kinetic equations of any-order terminal models with depropagation in a unified way, by reducing them into closed steady-state equations which give the exact solution of the original kinetic equations. Based on the steady-state equations, we also derive a general thermodynamic equality in which the Shannon entropy of the copolymer sequence is explicitly introduced as part of the free energy dissipation of the whole copolymerization system.

Keywords: steady-state copolymerization, depropagation, higher-order terminal model, kinetics, thermodynamics

Online supplementary data available from stacks.iop.org/JPCM/27/235105/mmedia

(Some figures may appear in colour only in the online journal)

1. Introduction

Understanding the kinetics of copolymerization and thus controlling the copolymer sequence statistics (e.g., copolymer composition, sequence distribution) are the key subjects in the study of copolymer, since the sequence statistics significantly affect the chemical and physical properties of the copolymers [1]. Therefore, it has become an important issue to theoretically model the copolymerization kinetics and estimate the rate constants of all the involved polymerization reactions. This has drawn a lot of attention, both experimentally and theoretically, since the 1940s. In order to study the kinetics, the experiments are usually conducted at low conversion conditions to maintain the copolymerization process at steady state (i.e., the monomer concentrations in the environment are almost unchanged during the process), which actually much simplifies the modeling and analysis of experiment data. Based on the steady-state assumption, different theoretical models have been suggested for different systems. Early works assumed the so-called terminal effects, i.e., the last monomer unit at the growing end of the copolymer influences the chain growth and thus the copolymer composition. Several terminal models were developed in the 1940s and successfully applied to experiments [2–5]. Besides the assumptions of terminal effect, these early models also assumed that the copolymerization reactions are irreversible, which ensures
the corresponding kinetic equations to be solved analytically. These two assumptions were shown insufficient to explain later experimental results, which leads to the development of two other categories of models.

The first category was proposed to account for the so-called penultimate effect, i.e., the next-to-last (penultimate) monomer unit at the growing end can have substantial influence called penultimate effect, i.e., the next-to-last (penultimate) monomer in the polymer sequence. This effect was suggested by Merz et al. [8], and then was revised and developed (for a review, see [9]). Besides the terminal (also called first-order terminal in this article) and penultimate (the second-order terminal) effects, higher-order terminal effects are also possible (e.g., antepenultimate effect [10]). But such cases have not been systematically investigated.

The second category was proposed to account for depropagation effect which brings substantial mathematical difficulty to the studies of copolymerization kinetics. Depropagation was noticed very early in the 1960s. It originated from the thermodynamic argument, i.e., all the reactions pathways are essentially reversible and depropagation may become significant at some elevated temperature. A few copolymerization systems do exhibit depropagation which shows substantial impacts on the copolymerization kinetics and copolymer composition (e.g., [11–13]). Such temperature effects can only be described by reversible models. However, depropagation always leads to hierarchically-coupled and unclosed kinetic equations which are hard to solve analytically (as will be clear in later sections). Because of this mathematical difficulty, it was not until 1987 that the first systematic treatment of first-order terminal models with depropagation was given by Kruger et al. [14]. Kruger’s approach was based on the key assumption that the copolymer sequence can be described as a first-order Markov chain (equation (12) in [14]). By using this assumption, Kruger et al. succeeded in reducing the original kinetic equations into closed steady-state equations. However, the validity of this assumption has not been proven rigorously or verified numerically. Moreover, how to generalize Kruger’s approach to higher-order terminal models was unclear. As far as we know, the only attempt to extend Kruger logic to penultimate models with depropagation has been made by Li et al [15, 16]. In their works, however, the first-order Markov chain assumption which is valid only for terminal models, was inappropriately employed. This makes their penultimate model mathematically self-inconsistent (detailed discussion will be given in section 2.3). To date there are no well-established penultimate models with depropagation available in the literature.

Recently, the study on steady-state copolymerization also attracted attention from physicists who were interested in visualizing the nonequilibrium copolymerization as information-generating process. In [17–19], the zero-order copolymerization model with depropagation (named as Bernoullian model in this article, see section 2.1) was introduced, without giving the derivation of the steady-state equations, to discuss some interesting issues (e.g., fidelity of DNA replication). In [20], the first-order terminal model was discussed, similar to Kruger, under the assumption of first-order Markov chain. These works also put an emphasis on the thermodynamics of steady-state copolymerization and gave very general and interesting relations between the copolymer sequence entropy and the thermodynamic entropy production of the copolymerization system. However, there still lacks a systematic investigation on the steady-state kinetics and thermodynamics of any-order terminal model with depropagation.

In this article, we will generalize Kruger’s Markov-chain assumption of the copolymer sequence distribution and suggest a unified mathematical approach to solve the steady-state kinetic equations of any-order terminal model with depropagation. Based on the solution, we will also present a detailed discussion on the steady-state thermodynamics of copolymerization.

2. Basic theory of steady-state copolymerization kinetics

2.1. Bernoullian model

As the simplest case of copolymerization, Bernoullian model (i.e., zero-order terminal model) assumes that the propagation and depropagation of monomers are independent of the terminal monomer unit. Although it is not a good model for real copolymerization systems, it can serve as a starting point of our discussion. Below we investigate a two-component (A, B) system. Generalization to more complex cases (e.g., multi-component systems) will be given in later sections.

Denoting the propagation rate constants as $k_A^0$ or $k_B^0$, and depropagation rate constants as $k_A$ or $k_B$, we have the reaction scheme below

\[ \sim \cdot + A \underset{k_A}{\overset{k_A^0}{\rightleftharpoons}} A \cdot, \sim \cdot + B \underset{k_B}{\overset{k_B^0}{\rightleftharpoons}} B \cdot \]

Imaging a single growing copolymer. $\sim \cdot$ represents the reactive end (i.e., the growing end, being either A or B), and the occurrence probability of A- or B- at the terminal is denoted as $P_A$ and $P_B$, respectively. We define $k_A \equiv k_A^0[A], k_B \equiv k_B^0[B], [A], [B]$ as monomer concentrations in the environment which are constants during steady-state copolymerization. Supposing at some moment the copolymer contains $N_A$ monomer A and $N_B$ monomer B, the total number of monomers $N = N_A + N_B$. They all increase with time during copolymerization, and the corresponding kinetic equations are

\[
\begin{align*}
N_A & \equiv J_A = k_A - \bar{k}_A P_A \\
N_B & \equiv J_B = k_B - \bar{k}_B P_B \\
N & \equiv J_{tot} = J_A + J_B
\end{align*}
\]

$J_A$, $J_B$ are, respectively, the overall incorporation rates of A and B. In steady-state copolymerization, $d(N_A/N)/dt = d(N_B/N)/dt = 0$, or equivalently, $N_A/N = N_A/N = J_A/J_{tot}$ and $N_B/N = N_B/N = J_B/J_{tot}$. So the overall occurrence probability of A or B in the copolymer can be expressed as $Q_A \equiv N_A/N = J_A/J_{tot}, Q_B \equiv N_B/N = J_B/J_{tot}$.

Higher order of chain-end sequence distribution $P_{i_0 \cdots i_1}$ ($i_m = A \text{ or } B, m = 1, 2, \cdots, n, i_n \cdots i_1$ denotes the chain-end sequence, with $i_1$ representing the terminal unit), and the total number of sequence $i_n \cdots i_1$ occurring in the
The kinetic equations of \( P_{i,j} \) are \( \dot{P}_{i,j} = k_A P_{i,j} - k_B P_{i,j} + k_A P_{i,j-1} + k_B P_{i,j+1} \). We also define \( \tilde{J}_{i,j} = J_{i,j} + J_{i,j}^\perp \) and the overall sequence distribution \( Q_{i,j} = N_{i,j}/N = J_{i,j}/J_{\text{tot}} \).

The kinetic equations of \( \tilde{J}_{i,j} \) (for any \( n \geq 1 \)) can be written as

\[
P_{i,j} = J_{i,j} - \tilde{J}_{i,j} = J_{i,j} - J_{i,j}^\perp = k_A P_{i,j} - k_B P_{i,j} + \tilde{k}_A P_{i,j} - \tilde{k}_B P_{i,j}.
\]

The existence of depopagation rates \( \tilde{k}_A, \tilde{k}_B \) makes these equations hierarchically coupled and hard to solve. Fortunately, for steady-state copolymerization \( \tilde{J}_{i,j} = 0 \) (for any \( n \geq 1 \)), we can use the following truncation method to solve these equations.

In Bernoullian model, the steady-state copolymerization kinetics is determined only by \( P_A, P_B \). This means that the coupled equations are redundant and can be reduced to equations of the two basic variables \( P_A, P_B \). This reduction can be achieved by the following zero-order factorization conjecture of the chain-end sequence distribution, \( P_{i,j} = P_{i,j}^\perp \), which leads to

\[
J_{i,j} = \left( \prod_{m=2}^n P_{i,m} \right) J_{i,1}, \quad \tilde{J}_{i,j} = \left( \prod_{m=1}^n P_{i,m} \right) J_{\text{tot}}
\]

From the steady-state kinetic equation \( \dot{J}_{i,j} = J_{i,j} - \tilde{J}_{i,j} \), we can get

\[
\frac{J_{i,j}}{J_{i}} = \frac{J_{\text{tot}}}{P_{i,1}}
\]

Therefore, each of the coupled equations is reduced to the same steady-state equation of \( P_A, P_B \),

\[
\frac{J_A}{P_A} = \frac{J_B}{P_B}
\]

Combining the normalization condition \( P_A + P_B = 1 \), we now obtain a set of closed equations which gives the exact solution of the original kinetic equations (these steady-state equations have been used without derivation in [18, 19]). However, [19] also provides another independent method, i.e., the infinite-state Markov chain model, to give an exact calculation of \( J_A/J_B \) which is proven to be identical to that obtained from equation (6). This method can be generalized to higher-order terminal model and shown to be equivalent to our approach. Details can be found in the supplement (see stacks.iop.org/JPCM/27/235105/mmedia).

Support of the factorization conjecture comes from the Monte carlo simulations by using Gillespie algorithm [21, 22] (here the rate parameters are arbitrarily chosen). One can directly simulate the steady-state copolymerization from any given initial condition of \( P_A, P_B \), and obtain all the sequence statistics (e.g., the chain-end sequence distribution \( P_{i,1}, P_{i,j} \), etc) from a number of simulations. For simplicity, we only check the factorizations \( P_{i,1} = P_{i}, P_{i,j} = P_{i,j}, P_{i,j} = P_{i,j}, P_{i,j} = P_{i,j} \), and \( P_{i,j} = P_{i,j} \). As shown in figures 1(a)–(c), for arbitrary choice of rate parameters \( k_A, k_B, \tilde{k}_A, \tilde{k}_B \) (the only constraint on the parameters is that they should ensure \( J_{\text{tot}} > 0 \), i.e., the copolymer is growing), all the equalities hold when copolymerization reaches the unique steady state (which is determined only by rate parameters and is independent of the choice of initial conditions). In figure 1(d), we plot the time-evolution trajectories of \( P_A, P_B \) given by the simulation and also indicate the steady-state values (shown by dashed lines) of \( P_A, P_B \) obtained by numerically solving equation (6), which also shows good agreement between the simulation and the theory.

2.2. Terminal model

The so-called terminal model (i.e., the first-order terminal model), where the propagation and depopagation of monomer
A and B are dependent on the identity of the terminal monomer unit, is a much more realistic model than Bernoullian model for real copolymerization systems. The reaction pathways in terminal model are

\[ \sim A \cdot +A \xrightarrow{k_{AA}} \sim AA \sim A \cdot +B \xrightarrow{k_{AB}} \sim AB \sim \]

\[ \sim B \cdot +A \xrightarrow{k_{BA}} \sim BA \sim B \cdot +B \xrightarrow{k_{BB}} \sim BB. \]

Defining

\[ J_{i_m \rightarrow i_{j_1}} \equiv k_{i_{j_1}j_{i_m}} P_{i_m \rightarrow i_{j_1}} \]

\[ \bar{J}_{i_m \rightarrow i_{j_1}} \equiv J_{i_m \rightarrow i_{j_1}A} + J_{i_m \rightarrow i_{j_1}B} \]

(7)

where $i_m = A, B (m = 1, 2, \ldots, n; n \geq 2)$, we can write the corresponding kinetic equations for $P_{i_m \rightarrow i_{j_1}} (n \geq 3)$ as below

\[ \dot{P}_{i_m \rightarrow i_{j_1}} = J_{i_m \rightarrow i_{j_1}} - \bar{J}_{i_m \rightarrow i_{j_1}} \]

(8)

The basic variables here are $P_{AA}, P_{AB}, P_{BA}, P_{BB}$, rather than $P_A, P_B$. Following the same logic in the previous section, we can reduce the hierarchically coupled equations equation (8) to an equivalent set of closed equations of $P_{AA}, P_{AB}, P_{BA}, P_{BB}$, by using the first-order factorization conjecture

\[ P_{i_m \rightarrow i_{j_1}} = \frac{n}{m=2} P_{i_m \rightarrow i_{m-1}} \left[ \sum_{m=3}^{n} P_{i_m \rightarrow i_{m-1}} \right]^{-1}, n \geq 3 \]

(9)

Then the steady-state kinetic equations $\dot{P}_{i_m \rightarrow i_{j_1}} = 0 (n \geq 3)$ are reduced to

\[ \frac{J_{i_{j_1}}}{P_{i_{j_1}}} = \frac{\bar{J}_{i_{j_1}}}{P_{i_{j_1}}} \]

(10)

or equivalently,

\[ \frac{J_{Ai}}{P_{Ai}} = \frac{J_{Bi}}{P_{Bi}} \]

(11)

where $i = A, B$.

The validity of the factorization $P_{i_{j_1}i_{j_2}} = P_{i_{j_1}} P_{i_{j_2}i_{j_1}} / P_{i_{j_1}}$ and the steady-state equations equation (12) can be checked by Monte Carlo simulation. Figure 2(a) shows the factorization holds when copolymerization reaches steady state, and figure 2(b) shows that the steady-state values of $P_{AA}, P_{AB}, P_{BA}, P_{BB}$ obtained by directly solving equation (12) are in good agreement with that given by the simulation (rate parameters used in the simulations are arbitrarily chosen as long as $J_{tot} > 0$).

It seems also possible in principle to use the zero-order factorization conjecture to reduce the original kinetic equations to closed steady-state equations. However, as shown by figure 2(c), the zero-order factorization fails in terminal model, meaning that it is not applicable to terminal models. One may also suggest other factorization conjectures, for instance, the direct factorization $P_{i_{j_1}} = P^{(-1)}_{i_{j_1}} P_j (P^{(-1)}_i$ refers to the occurrence probability of monomer unit $i$ at the penultimate position. In fact, this conjecture does not result in closed steady-state equations). As indicated by figure 2(d), this factorization also fails in terminal model, meaning that the correlation between the terminal unit and the penultimate unit cannot be decoupled. In other words, one should take $P_{i_{j_1}}$ as the basic variables to describe the terminal effect.

The first-order factorization conjecture $P_{i_{j_1}i_{j_2}} = P_{i_{j_1}} P_{i_{j_2}i_{j_1}} / P_{i_{j_1}}$ is actually equivalent to the first-order Markov-chain assumption used in [14] and [20]. Defining transition probability $p(i_{j_1}i_{j_2}) \equiv P_{i_{j_1}i_{j_2}} / P_{i_{j_1}}$, $p(A|i_{j_1}) + p(B|i_{j_1}) = 1$,
we now can rewrite the factorization conjecture as \( P_{i_2;i_1} = p(i_3;i_2) p(i_2;i_1) P_{i_1} \). It is worth noting that we have chosen \( P_{i_2} \), rather than \( p_2 \) and \( p(i_2;i_1) \), as basic variables so as to represent the steady-state equations in a much simpler and more intuitive form.

### 2.3. Penultimate model

The reaction pathway of penultimate model (i.e., the second-order terminal model) can be expressed as

\[
\begin{align*}
&\sim AA + A \xrightarrow{k_{\text{AAA}}} \sim AAA, \sim AA + B \xrightarrow{k_{\text{AB}}} \sim AAB, \\
&\sim AB + A \xrightarrow{k_{\text{ABA}}} \sim ABA, \sim AB + B \xrightarrow{k_{\text{ABB}}} \sim ABB, \\
&\sim BA + A \xrightarrow{k_{\text{BAA}}} \sim BAA, \sim BA + B \xrightarrow{k_{\text{BBB}}} \sim BBB, \\
&\sim BB + A \xrightarrow{k_{\text{BBA}}} \sim BBAA, \sim BB + B \xrightarrow{k_{\text{BBB}}} \sim BBAB.
\end{align*}
\]

Here the basic variables are \( P_{i_3;i_2;i_1} (i_3, i_2, i_1 = A, B) \). As in previous sections, we still define

\[
\begin{align*}
J_{i_3;i_2;i_1} &\equiv k_{i_3;i_2;i_1} P_{i_3;i_2;i_1} - \tilde{J}_{i_3;i_2;i_1} P_{i_3;i_2;i_1}, \\
\tilde{J}_{i_3;i_2;i_1} &\equiv J_{i_3;i_2;i_1} + J_{i_3;\text{AB}i_1} + J_{i_3;\text{BB}i_1},
\end{align*}
\]

where \( i_m = A, B \) (\( m = 1, 2, \ldots, n \); \( n \geq 3 \)). The kinetic equation of \( P_{i_3;i_2;i_1} \) is

\[
\dot{P}_{i_3;i_2;i_1} = J_{i_3;i_2;i_1} - \tilde{J}_{i_3;i_2;i_1} + \cdot
\]

To solve these equations, we take the following second-order factorization conjecture

\[
P_{i_3;i_2;i_1} = \prod_{m=3}^{n} P_{i_3;i_2;i_1;i_m} \left( \prod_{m=4}^{n} P_{i_3;i_2;i_1;i_2} \right)^{-1}, \quad n \geq 4
\]
is invalid for penultimate model, as indicated by figure 3(b). In a recent paper, however, Li et al invoked the first-order Markov chain assumption to solve their penultimate model [15]. According to our theory, Li’s mathematical treatment is inappropriate. To account for penultimate effects, 8 variables \( P_{i,j,i} \) (i.e., \( i_1, i_2, i_3 = A, B \)) are required. Li’s model oversimplifies the problem to 6 variables (the two-terminal probability \( P_i \) and the four-transition probability \( p(i_2|i_1), i_2, i_1 = A, B \), in terms of first-order Markov-chain assumption), and derived closed but over-determined equations (equation (8-10) in [15]) from the steady-state kinetic equations \( P_{i,j} = 0 \). These equations are doubtful: two of the 16 rate parameters \( k_{AAA}, k_{BBB} \) are totally absent from the original four kinetic equations \( P_{i,j} = J_{i,j} - J_{i,j^*} \), meaning that these equations of 6-variables are inadequate to describe the penultimate effect. Moreover, if Li’s treatment is extended to higher-order kinetic equations (e.g., \( P_{i,j} = 0 \)), self-inconsistency of their theory can be further uncovered. For instance, in the extreme case \( k_{i,j} = 0 \). Under the first-order Markov chain assumption, \( P_{AAA} = k_{AAA}P_{BAA} - k_{AAB}P_{AAA} = 0 \) yields \( k_{AAA}P_{BAA} = k_{AAB}P_{AAA} \). \( P_{BAA} = k_{BAA}P_{BA} - (k_{AAA} + k_{AAB})P_{BAA} = 0 \) yields \( k_{BAA}P_{BA} = (k_{AAA} + k_{AAB})P_{AAA} \). Combining these two equations will lead to the wrong conclusion \( k_{BAA} = k_{AAA} \). Therefore, the second-order but not first-order factorization conjecture is required for the penultimate model (however, for very special cases such as the implicit penultimate model treated by Li et al, the first-order factorization conjecture may sometimes be a good approximation. Detailed analysis can be found in the supplement (see stacks.iop.org/JPhysCM/27/235105/mmedia)).

In general, if \( m < s \) is applied to the steady-state kinetic equations \( P_{i,j} = 0 \) (\( n = 1, 2, \ldots \)) of \( s \)-th order model (see the next section), one can always obtain an overdetermined set of equations which is mathematically self-inconsistent. On the other hand, higher-order \((m > s)\) factorization conjecture is redundant for the \( s \)-th order model. We therefore conclude that the \( s \)-th order model can only be appropriately described by \( s \)-th order factorization conjecture.

### 2.4. Higher-order terminal models

The logic presented in previous sections can be directly generalized to higher-order terminal models. Below we list the major results for \( s \)-th order terminal model, i.e., the propagation and depagination of \( A/B \) depend on the last \( s \) monomer units of the copolymer. Here the basic variables are \( P_{i,j} = 0 \) (\( 2^s+1 \) in total). We denote the propagation rates as \( k_{i,j} \) and depagation rates as \( \bar{k}_{i,j,k} \), and also

\[
J_{i,j} = k_{i,j}P_{i,j} - \bar{k}_{i,j,k}P_{i,j,k}
\]

where \( k_{i,j} = A, B \); \( m = 1, 2, \ldots, n \); \( n \geq s + 1 \).

The \( s \)-th order factorization conjecture is

\[
P_{i,j} = \prod_{m=s+1}^{n} P_{i_{m-1},i_{m}} \left[ \prod_{m=s+2}^{n} P_{i_{m-1},i_{m}} \right]^{-1}, n \geq s + 2
\]

The closed steady-state equations derived from \( \dot{J}_{i,j} = 0 \) (\( n \geq s + 1 \)) are the following \( 2^s \) equations

\[
\frac{J_{i,j}}{P_{i,j}} = \frac{J_{i,j}}{P_{i,j+1}}
\]

or equivalently,

\[
\frac{J_{i,j}}{J_{i,j+1}} = \frac{J_{i,j+1}}{P_{i,j+1}}
\]

The kinetic equations \( \dot{J}_{i,j} = J_{i,j} - J_{i,j^*} \) give other \( 2^s \) steady-state equations, from which any \( 2^s - 1 \) equations can be chosen. Combining the normalization condition \( \sum P_{i,j} = 1 \), we finally obtain a closed set of \( 2^s+1 \) equations for \( 2^s+1 \) variables.

The \( s \)-th order factorization conjecture can be rewritten equivalently as \( s \)-th order Markov chain, by defining the transition probability \( p(i_1,i_2,i_3;i_4) = P_{i_1,i_2,i_3}/P_{i_1,i_4} \), \( p(A|i_1,i_2,i_3,i_4) + p(B|i_1,i_2,i_3,i_4) = 1 \). Noting that \( J_{i_1,i_2,i_3,i_4} = J_{i_1,i_2,i_3} \), the steady-state equations (23) can be transformed into

\[
\frac{J_{i,j}}{J_{i,j+1}} = \frac{P_{i,j}}{P_{i,j+1}}
\]

Since the overall sequence distribution \( Q_{i,j} = \frac{Q_{i,j}}{Q_{i,j+1}} \) of \( J_{i,j} = J_{i,j} / J_{i,j+1} \), the steady-state equations can be rewritten as

\[
\frac{Q_{i,j}}{Q_{i,j+1}} = \frac{P_{i,j}}{P_{i,j+1}} = p(i_1,i_2,i_3,i_4)
\]

This simply means that the overall sequence distribution and chain-end sequence distribution can be described by the same \( s \)-th order Markov chain.

It is also worth noting that the \( s \)-th order model can reproduce the \((s - 1)\)-th order model if assuming \( k_{i_1,i_2,i_3} = k_{i_1,i_2,i_3} = k_{i_1,i_2,i_3} = k_{i_1,i_2,i_3} \). By the \( s \)-th order model, we have

\[
\frac{J_{i,j}}{J_{i,j+1}} = \frac{P_{i,j}}{P_{i,j+1}} = \frac{P_{i,j}}{P_{i,j+1}}
\]

which yields

\[
\frac{P_{i,j}}{P_{i,j+1}} = \frac{P_{i,j}}{P_{i,j+1}}
\]

or equivalently

\[
\frac{P_{i,j}}{P_{i,j+1}} = \frac{P_{i,j}}{P_{i,j+1}} = \frac{P_{i,j}}{P_{i,j+1}}
\]

This means \( P_{i,j} = P_{i,j} \), which is exactly the \((s - 1)\)-th order factorization conjecture.

Equation (27) also leads to

\[
\frac{J_{i,j}}{J_{i,j+1}} = \frac{J_{i,j}}{J_{i,j+1}} = \frac{J_{i,j}}{J_{i,j+1}} = \frac{J_{i,j}}{J_{i,j+1}}
\]

which is exactly the steady-state equations of \((s - 1)\)-th order model.
3. Generalization to multi-component systems

In real copolymerization systems, there might be multiple species of monomers. Generalization of the above kinetic theory to multi-component system is direct. Suppose there are \( l \) species of monomer in the system \((M_1, M_2, \cdots, M_l)\). In Bernoullian model, for instance, the basic variables are \( P_{M_i} \) \((i = 1, 2, \cdots, l)\). We have the following \( l \) equations for the \( l \) variables,

\[
\frac{J_{M_i}}{P_{M_i}} = \frac{J_{M_i}}{P_{M_i}} = \cdots = \frac{J_{M_l}}{P_{M_l}} = \cdots = \sum_{i=1}^{l} P_{M_i} = 1
\]

(30)

where \( J_{M_i} = k_{M_i} - \bar{k}_{M_i} P_{M_i} \).

Generalization of higher-order models to multi-component system is similar (details not given here).

4. Generalization to multi-step process

Different from cases discussed above where the propagation and depropagation are regarded as single-step process (e.g., in free radical copolymerization [13]), bio-copolymerization and depropagation are regarded as single-step process (e.g., equation (19) in [19]). In the latter case, the factorization conjectures such as DNA replication are often multi-step processes (e.g., equation (15) in [17]). Here we give explicit examples of such thermodynamic relations, based on the kinetic theory presented in previous sections. It should be pointed out first that the thermodynamics can only be well defined for special cases where propagation and depropagation are microscopically reversible (i.e., they proceed along the same reaction pathway), whereas the kinetic theory is generally applicable even to cases in which propagation and depropagation proceed in different reaction pathways (e.g., in DNA replication, propagation is catalyzed by the synthesis domain of DNA polymerase, and depropagation is catalyzed by the editing domain of DNA polymerase). Below we assume that propagation and depropagation are microscopically reversible.

We start from the Bernoullian model, given that \( A, B \) are of identical concentration, i.e., \([A] = [B] = [M]\). Details of higher-order models are given in appendix.

In Bernoullian model, the averaged free energy dissipation per incorporation can be expressed as \((RT \text{ is omitted for simplicity})\)

\[
\Delta G = \frac{\dot{S}}{J_{tot}}
\]

\[
\dot{S} = J_A \ln \left( \frac{\bar{k}_A P_A}{k_A P_A} \right) + J_B \ln \left( \frac{\bar{k}_B P_B}{k_B P_B} \right)
\]

(32)

\[
J_A = k_A^0 [M] - \bar{k}_A P_A, J_B = k_B^0 [M] - \bar{k}_B P_B
\]

\[
J_{tot} = J_A + J_B
\]

\( \dot{S} \) is the instantaneous entropy production of the reaction system. It is non-negative by definition.

The unique equilibrium state can be defined by \( J_A = k_A^0 [M]^{eq} - \bar{k}_A P_A^{eq} = 0 \) and \( J_B = k_B^0 [M]^{eq} - \bar{k}_B P_B^{eq} = 0 \), which yield \([M]^{eq} = (K_A + K_B)^{-1}, P_A^{eq} = K_A [M]^{eq} \) and \( P_B^{eq} = K_B [M]^{eq} \), here \( K_A \equiv k_A^0 / \bar{k}_A \) and \( K_B \equiv k_B^0 / \bar{k}_B \) are the equilibrium constants of the A, B reaction pathway, respectively. To drive the copolymerization process out of equilibrium, one should have \([M] > [M]^{eq} \).

In steady-state copolymerization, \( J_A / J_{tot} = P_A, J_B / J_{tot} = P_B \), so we have

\[
\Delta G = P_A \ln \left( \frac{P_A^{eq} [M]}{P_A [M]^{eq}} \right) + P_B \ln \left( \frac{P_B^{eq} [M]}{P_B [M]^{eq}} \right)
\]

(33)

this leads to the following equality

\[
\Delta G = \Delta \Psi - \Delta I
\]

\[
\Delta \Psi \equiv \ln \frac{[M]}{[M]^{eq}}
\]

(34)

\[
\Delta I \equiv P_A \ln \left( \frac{P_A}{P_A^{eq}} \right) + P_B \ln \left( \frac{P_B}{P_B^{eq}} \right)
\]

\( \Delta I \) is in the form of mutual information [23]. It is in fact the sequence information generated in the copolymerization process. The information of a sequence of length \( N \) is usually
Information from the equilibrium state to the steady state is

\[ J = \Delta_1 \ln \frac{Q(S_N)}{Q^S(S_N)} \]

Here, the occurrence probability of \( S_N \) is the overall occurrence probability of \( S_N \) in the copolymer sequence.

For the Bernoullian model, \( Q(S_N) = \prod_{i=1}^{N} P_i \). One can prove \( I_N = -N (P_A \ln P_A + P_B \ln P_B) \). Change of the sequence information from the equilibrium state to the steady state is defined as \( \Delta I_N = \sum Q(S_N) \ln \frac{Q(S_N)}{Q^S(S_N)} \), namely,

\[ \Delta I_N = N \left( P_A \ln \frac{P_A}{P_A^0} + P_B \ln \frac{P_B}{P_B^0} \right) = N \Delta I \quad (35) \]

or

\[ \Delta I = \lim_{N \to \infty} \frac{1}{N} \sum Q(S_N) \ln \frac{Q(S_N)}{Q^S(S_N)} \quad (36) \]

Hence, \( \Delta I \) can be understood as information gain per incorporation. \( \Delta \Psi \) can be regarded as the driving force to maintain the steady-state condition (i.e., to maintain the constant and nonequilibrium monomer concentration). The equality \( \Delta \Psi = \Delta G + \Delta I \), which also holds for higher-order terminal models (details can be found in appendix), has an intuitive meaning that the overall driving force \( \Delta \Psi > 0 \) is partitioned into two parts, one to keep the polymerization reaction out of equilibrium \( \Delta G > 0 \), the other to generate sequence information \( \Delta I > 0 \).

The above theory can be readily generalized to multi-component systems (details not given here).

For multi-step processes, one can similarly define the free energy dissipation \( \Delta G \) per incorporation. In the two-step process discussed in section 4, for instance, \( \Delta G \) can be written as (using the same notations as in section 4)

\[ \Delta G = \dot{S} / J_{tot} \]

\[ J_{tot} = J_A + J_B \]

\[ \dot{S} = J_A \ln \left( \frac{k_A k_{A2} (P_A + P_B)}{k_A k_{A2} P_A} \right) + J_B \ln \left( \frac{k_B k_{B2} (P_A + P_B)}{k_B k_{B2} P_B} \right) \]

\[ = J_A \ln \left( \frac{k_A (P_A + P_B)}{k_A P_A} \right) + J_B \ln \left( \frac{k_B (P_A + P_B)}{k_B P_B} \right) \]

It is obvious that \( \dot{S} \) is the instantaneous entropy production of the whole reaction system, which is non-negative by definition. Noting that \( J_{tot} \) is not defined as \( J_{tot} = J_A + J_A + J_B + J_B \), since the latter is the total flux of all the involved reactions but not the flux of incorporation, we have

\[ \Delta G = \pi_A \ln \left( \frac{k_A k_{A2}}{k_A k_{A2} \pi_A} \right) + \pi_B \ln \left( \frac{k_B k_{B2}}{k_B k_{B2} \pi_B} \right) \quad (38) \]

One can similarly define the information gain \( \Delta I \) and the driving force \( \Delta \Psi \)

\[ \Delta I = \pi_A \ln \left( \frac{\pi_A}{\pi_A^0} \right) + \pi_B \ln \left( \frac{\pi_B}{\pi_B^0} \right) \]

\[ \Delta \Psi = \ln \left( \frac{[M]}{[M]^0} \right) \quad (39) \]

and also have \( \Delta \Psi = \Delta G + \Delta I \).

The above discussion is based on the existence of the uniquely defined equilibrium state since \( [A] = [B] \).

If \([A] \neq [B]\), one can still define the equilibrium state by \( J_A = k_A^0 [A]^0 - k_{A2} P_A^0 = 0 \) and \( J_B = k_B^0 [B]^0 - k_{B2} P_B^0 = 0 \), but the equilibrium state is not unique. Instead, there are an infinite number of equilibrium states which satisfy \( J_A = 0 \) and \( J_B = 0 \). In such cases, one can arbitrarily choose \([A]^0\), \([B]^0\), and then determine the corresponding \( P_A^0\), \( P_B^0\) to define the equilibrium state. The thermodynamic equality still holds, only with some modifications. For instance, for single-step process, we have

\[ \Delta \Psi = \Delta G + \Delta I \]

\[ \Delta G \equiv \frac{J_A}{J_{tot}} \ln \left( \frac{k_A^0}{k_A P_A} \right) + \frac{J_B}{J_{tot}} \ln \left( \frac{k_B^0}{k_B P_B} \right) > 0 \]

\[ \Delta I \equiv \frac{P_A}{P_A^0} \ln \left( \frac{P_A}{P_A^0} \right) + \frac{P_B}{P_B^0} \ln \left( \frac{P_B}{P_B^0} \right) > 0 \quad (40) \]

Last but not least, the information interpretation of \( \Delta I \) is totally based on the 0th-order factorization conjecture of the 0th-order terminal model. Any lower-order factorization conjecture is incompatible with such an information interpretation. This gives an extra support of the factorization conjectures we used in the kinetic theory.

6. Summary

In this article, we proposed a systematic approach, based on Markov chain assumptions of the copolymer sequence distribution, to solve the unclosed kinetic equations of any-order terminal models with propagation of steady-state copolymerization. The Markov chain assumptions were directly validated by Monte Carlo simulations, and the original kinetic equations were then reduced to closed steady-state equations which give the exact solution of the original equations. The derived steady-state equations were presented in a unified and intuitive form (e.g., equation (23)) which provides convenient tools to fit or explain experimental data. This approach was also successfully generalized to more complex cases, e.g., multi-component systems and multi-step processes. Furthermore, based on the steady-state equations and Markov chain assumptions, we derived a general thermodynamic equality in which the Shannon entropy of the copolymer sequence is explicitly introduced as part of the free energy dissipation of the whole copolymerization system. This not only offers extra support to the validation of the Markov chain assumptions, but also provides new insights to understand the copolymerization process from the perspective of information theory.

Acknowledgments

The authors are grateful for the financial support by the National Basic Research Program of China (973 program, No.2013CB932804) and National Natural Science Foundation of China (No.11105218, No.91027046).
Appendix

For $s$th-order terminal model ($s \geq 1$), we define the unique equilibrium state by

$$J_{i_{m+1} \cdots i_{m+s}} = k^0_{i_{m+1} \cdots i_{m+s}} [M]^{eq} P_{i_{m+1} \cdots i_{m+s}}^{eq} - \bar{k}^0_{i_{m+1} \cdots i_{m+s}} P_{i_{m+1} \cdots i_{m+s}}^{eq} = 0$$

(41)

where $i_m = A, B$ ($m = 1, 2, \cdots, s + 1$). $[M]^{eq}$, $P_{i_{m+1} \cdots i_{m+s}}^{eq}$ can be directly solved from these equations. They are functions only of equilibrium constants.

So we have

$$k^0_{i_{m+1} \cdots i_{m+s}} = \frac{P_{i_{m+1} \cdots i_{m+s}}^{eq}}{[M]^{eq} P_{i_{m+1} \cdots i_{m+s}}^{eq}}$$

(42)

The averaged free energy dissipation is defined

$$\Delta G = \frac{1}{J_{tot}} \sum_{i_{m+1} \cdots i_{m+s}} J_{i_{m+1} \cdots i_{m+s}} \ln \left( \frac{k^0_{i_{m+1} \cdots i_{m+s}} [M] P_{i_{m+1} \cdots i_{m+s}}^{eq}}{k_{i_{m+1} \cdots i_{m+s}} P_{i_{m+1} \cdots i_{m+s}}^{eq}} \right)$$

(43)

Substituting equation (42) into the above equation, we have

$$\Delta G = \Delta \Psi - \Delta I$$

$$\Delta \Psi = \frac{1}{J_{tot}} \sum_{i_{m+1} \cdots i_{m+s}} J_{i_{m+1} \cdots i_{m+s}} \ln \left( \frac{[M]^{eq}}{[M]^{eq}} \right)$$

$$\Delta I = \frac{1}{J_{tot}} \sum_{i_{m+1} \cdots i_{m+s}} J_{i_{m+1} \cdots i_{m+s}} \ln \left( \frac{P_{i_{m+1} \cdots i_{m+s}}^{eq} P_{i_{m+1} \cdots i_{m+s}}^{eq}}{P_{i_{m+1} \cdots i_{m+s}}^{eq} P_{i_{m+1} \cdots i_{m+s}}^{eq}} \right)$$

(44)

$$\Delta I$$ is also the information gain per incorporation step

$$\Delta I = \lim_{N \to \infty} \frac{1}{N} \sum_{i_{m+1} \cdots i_{m+s}} Q_{i_{m+1} \cdots i_{m+s}} \ln \frac{Q_{i_{m+1} \cdots i_{m+s}}^{eq}}{Q_{i_{m+1} \cdots i_{m+s}}}$$

(45)

To prove this, we transform equation (44) into

$$\Delta I = \frac{1}{J_{tot}} \sum_{i_{m+1} \cdots i_{m+s}} J_{i_{m+1} \cdots i_{m+s}} \ln \left( \frac{P_{i_{m+1} \cdots i_{m+s}}^{eq} P_{i_{m+1} \cdots i_{m+s}}^{eq}}{P_{i_{m+1} \cdots i_{m+s}}^{eq} P_{i_{m+1} \cdots i_{m+s}}^{eq}} \right)$$

(46)

$$\sum_{i_{m+1} \cdots i_{m+s}} J_{i_{m+1} \cdots i_{m+s}} \ln \left( \frac{P_{i_{m+1} \cdots i_{m+s}}^{eq} P_{i_{m+1} \cdots i_{m+s}}^{eq}}{P_{i_{m+1} \cdots i_{m+s}}^{eq} P_{i_{m+1} \cdots i_{m+s}}^{eq}} \right)$$

In the second term

$$\sum_{i_{m+1} \cdots i_{m+s}} J_{i_{m+1} \cdots i_{m+s}} \ln \left( \frac{P_{i_{m+1} \cdots i_{m+s}}^{eq} P_{i_{m+1} \cdots i_{m+s}}^{eq}}{P_{i_{m+1} \cdots i_{m+s}}^{eq} P_{i_{m+1} \cdots i_{m+s}}^{eq}} \right)$$

$$= \sum_{i_{m+1} \cdots i_{m+s}} J_{i_{m+1} \cdots i_{m+s}} \ln P_{i_{m+1} \cdots i_{m+s}} - \sum_{i_{m+1} \cdots i_{m+s}} J_{i_{m+1} \cdots i_{m+s}} \ln P_{i_{m+1} \cdots i_{m+s}}$$

$$= \sum_{i_{m+1} \cdots i_{m+s}} J_{i_{m+1} \cdots i_{m+s}} \ln P_{i_{m+1} \cdots i_{m+s}}$$

$$= \sum_{i_{m+1} \cdots i_{m+s}} J_{i_{m+1} \cdots i_{m+s}} \ln P_{i_{m+1} \cdots i_{m+s}}$$

In the third step, we have substituted $i_m$ by $j_{m-1}$ ($m = 2, 3, \cdots, s + 1$). In the last step, we have used the steady-state conditions $J_{i_{m+1} \cdots i_{m+s}} = \bar{J}_{i_{m+1} \cdots i_{m+s}}$.

Similarly, one can show

$$\sum_{i_{m+1} \cdots i_{m+s}} J_{i_{m+1} \cdots i_{m+s}} \ln \left( \frac{P_{i_{m+1} \cdots i_{m+s}}^{eq} P_{i_{m+1} \cdots i_{m+s}}^{eq}}{P_{i_{m+1} \cdots i_{m+s}}^{eq} P_{i_{m+1} \cdots i_{m+s}}^{eq}} \right) = 0$$

(48)

We define the transition probability as $p(i_{m+1} \cdots i_{m+s}) = P_{i_{m+1} \cdots i_{m+s}} / P_{i_{m+1} \cdots i_{m+s}}$.

Since $Q_{i_{m+1} \cdots i_{m+s}} = J_{i_{m+1} \cdots i_{m+s}} / J_{tot}$ and $Q_{i_{m+1} \cdots i_{m+s}} / P_{i_{m+1} \cdots i_{m+s}} = p(i_{m+1} \cdots i_{m+s})$ (equation (25)), we rewrite the above equation as

$$\Delta I = \sum_{i_{m+1} \cdots i_{m+s}} Q_{i_{m+1} \cdots i_{m+s}} \ln \frac{p(i_{m+1} \cdots i_{m+s})}{p^{eq}(i_{m+1} \cdots i_{m+s})}$$

(49)

$$= \sum_{i_{m+1} \cdots i_{m+s}} Q_{i_{m+1} \cdots i_{m+s}} \ln \frac{p(Ai_{m+1} \cdots i_{m+s})}{p^{eq}(Ai_{m+1} \cdots i_{m+s})}$$

$$= \sum_{i_{m+1} \cdots i_{m+s}} Q_{i_{m+1} \cdots i_{m+s}} \ln \frac{p(Bi_{m+1} \cdots i_{m+s})}{p^{eq}(Bi_{m+1} \cdots i_{m+s})}$$

Since $p(Ai_{m+1} \cdots i_{m+s}) + p(Bi_{m+1} \cdots i_{m+s}) = 1$ and $p^{eq}(Ai_{m+1} \cdots i_{m+s}) + p^{eq}(Bi_{m+1} \cdots i_{m+s}) = 1$, it can be readily proven that $\Delta I$ is non-negative.

Furthermore, one can show $\Delta I$ is exactly the information gain per incorporation. The information of a sequence of length $N$ is $\Delta I_N = \sum Q(S_N) \ln (Q(S_N) / Q^{eq}(S_N))$. Since $Q(S_N) = \prod_{n=1}^{N} p(i_n | i_{n-1} \cdots i_{n-s}) \sum_{i_{n-1} \cdots i_{m+s}} \sum_{i_{m+1} \cdots i_{m+s}} Q_{i_{m+1} \cdots i_{m+s}} \ln \frac{Q_{i_{m+1} \cdots i_{m+s}}^{eq}}{Q_{i_{m+1} \cdots i_{m+s}}}$$

(50)

The first summation term can be rewritten as

$$\sum_{n=1}^{N} \left[ \sum_{i_{m+1} \cdots i_{m+s}} Q_{i_{m+1} \cdots i_{m+s}} \ln \frac{p(i_n | i_{n-1} \cdots i_{n-s})}{p^{eq}(i_n | i_{n-1} \cdots i_{n-s})} \right]$$

$$= \sum_{n=1}^{N} \left[ \sum_{i_{m+1} \cdots i_{m+s}} Q_{i_{m+1} \cdots i_{m+s}} \ln \frac{p(i_n | i_{n-1} \cdots i_{n-s})}{p^{eq}(i_n | i_{n-1} \cdots i_{n-s})} \right]$$

$$= (N - s) \sum_{i_{m+1} \cdots i_{m+s}} Q_{i_{m+1} \cdots i_{m+s}} \ln \frac{p(i_n | i_{n-1} \cdots i_{n-s})}{p^{eq}(i_n | i_{n-1} \cdots i_{n-s})}$$

(51)

The second term of equation (50) is comparatively negligible. Hence the information gain per incorporation is

$$\Delta I = \lim_{N \to \infty} \frac{1}{N} \Delta I_N$$

$$= \sum_{i_{m+1} \cdots i_{m+s}} Q_{i_{m+1} \cdots i_{m+s}} \ln \frac{p(i_n | i_{n-1} \cdots i_{n-s})}{p^{eq}(i_n | i_{n-1} \cdots i_{n-s})}$$

(52)

This is exactly equation (49).

References

[1] Matyjaszewski K and Davis T P 2002 Handbook Of Radical Polymerization (New York: Wiley)
[2] Wall F T 1941 J. Am. Chem. Soc. 63 1862
[3] Mayo F R and Lewis F M 1944 J. Am. Chem. Soc. 66 1594
[4] Alfrey T and Goldfinger G 1944 J. Chem. Phys. 12 205
[5] Alfrey T and Goldfinger G 1944 J. Chem. Phys. 12 322
[6] Fukuda T, Ma Y D and Inagaki H 1985 Macromolecules 18 17
[7] Davis T P and ODriscoll K F 1989 J. Polym. Sci. C: Polym. Lett. 27 181
[8] Merz E, Alfrey T Jr and Goldfinger G 1946 J. Polym. Sci. 1 75
[9] Coote M L, Davis T P and Radom L 2000 Models for free-radical copolymerization propagation kinetics Controlled/Living Radical Polymerization (ACS Symposium Series vol 768) (Washington, DC: American Chemical Society)
[10] Lowry G G 1960 J. Polym. Sci. 42 463
[11] O’Driscoll K F and Gasparro F P 1967 J. Macromol. Sci: Chem. A 1 643
[12] Martinet F and Guillot J 1997 J. Appl. Polym. Sci. 65 2297
[13] Leamen M J, McManus N T and Penlidis A 2005 J. Polym. Sci. A: Polym. Chem. 43 3868
[14] Kruger H, Bauer J and Rubner J 1987 Macromol. Chem. 188 2163
[15] Li D H, Leiza J R and Hutchinson R A 2005 Macromol. Theory Simul. 14 554
[16] Li D H, Li N and Hutchinson R A 2006 Macromolecules 39 4366
[17] Andrieux D and Gaspard P 2008 Proc. Natl Acad. Sci. USA 105 9516
[18] Esposito M, Lindenberg K and Van den Broeck C 2010 J. Stat. Mech. P01008
[19] Cady F and Qian H 2009 Phys. Biol. 6 036011
[20] Gaspard P and Andrieux D 2014 J. Chem. Phys. 141 044908
[21] Gillespie D T 1977 J. Phys. Chem. 81 2340
[22] Gillespie D T 2007 Annu. Rev. Phys. Chem. 58 35
[23] Kullback S and Leibler R A 1951 Ann. Math. Stat. 22 79