

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

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2015 J. Phys.: Condens. Matter 27 235105

(<http://iopscience.iop.org/0953-8984/27/23/235105>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 124.16.143.184

This content was downloaded on 26/11/2016 at 12:33

Please note that [terms and conditions apply](#).

You may also be interested in:

[Proofreading of DNA polymerase: a new kinetic model with higher-order terminal effects](#)

Yong-Shun Song, Yao-Gen Shu, Xin Zhou et al.

[Open-system thermodynamic analysis of DNA polymerase fidelity](#)

Field Cady and Hong Qian

[Force-velocity relation for copolymerization processes](#)

Pierre Gaspard

[Thermodynamics of accuracy in kinetic proofreading: dissipation and efficiency trade-offs](#)

Riccardo Rao and Luca Peliti

[GROWTH OF COPOLYMER CHAINS AND MIXED CRYSTALS—TRIAL-AND-ERROR STATISTICS](#)

A A Chernov

[NESM: A paradigm and applications](#)

T Chou, K Mallick and R K P Zia

[Kinetics and thermodynamics of reversible polymerization in closed systems](#)

Sourabh Lahiri, Yang Wang, Massimiliano Esposito et al.

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

Yao-Gen Shu^{1,2}, Yong-Shun Song², Zhong-Can Ou-Yang¹ and Ming Li²

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

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



CrossMark

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 on the copolymerization kinetics (e.g., [6, 7]). The original penultimate model 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.

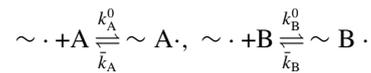
In this article, we will generalize Kruger's Markov-chain assumption of the copolymer sequence distribution and suggest a unified mathematical approach to solve the steady-state kinetic equations of any-order terminal model with depropagation. Based on the solution, we will also present a detailed discussion 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 \bar{k}_A or \bar{k}_B , we have the reaction scheme below



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

$$\begin{aligned} \dot{N}_A &\equiv J_A = k_A - \bar{k}_A P_A \\ \dot{N}_B &\equiv J_B = k_B - \bar{k}_B P_B \\ \dot{N} &\equiv J_{\text{tot}} = J_A + J_B \end{aligned} \quad (1)$$

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, $N_A/N = \dot{N}_A/\dot{N} = J_A/J_{\text{tot}}$ and $N_B/N = \dot{N}_B/\dot{N} = J_B/J_{\text{tot}}$. So the overall occurrence probability of A or B in the copolymer can be expressed as $Q_A \equiv N_A/N = J_A/J_{\text{tot}}$, $Q_B \equiv N_B/N = J_B/J_{\text{tot}}$.

Higher order of chain-end sequence distribution $P_{i_n \dots i_1}$ ($i_m = A$ or B , $m = 1, 2, \dots, n$. $i_n \dots i_1$ denotes the chain-end sequence, with i_1 representing the terminal unit), and the total number of sequence $i_n \dots i_1$ occurring in the

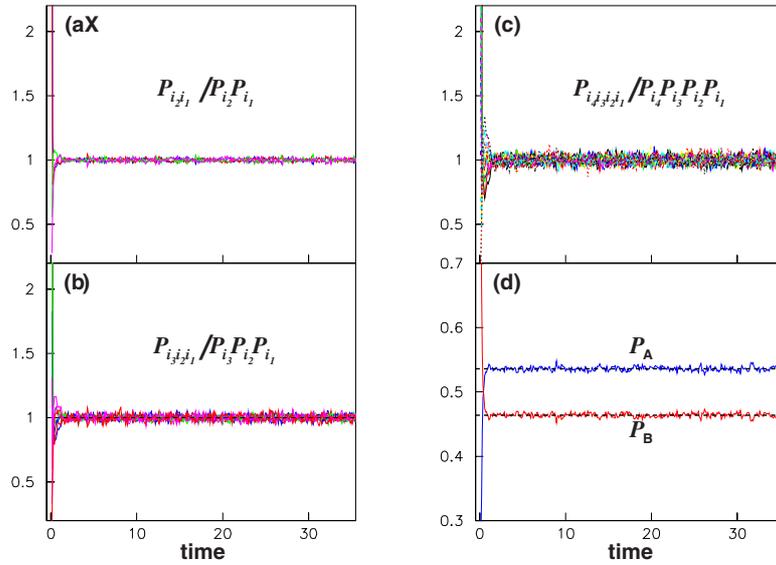


Figure 1. Simulation verification of the zero-order factorization conjecture (a,b,c) and steady-state equation (d), with illustrative rate parameters $k_A = 4.0$, $k_B = 3.0$, $\bar{k}_A = 2.0$, $\bar{k}_B = 1.0$. The theoretical values of P_A , P_B given by the steady-state equations are indicated as dashed lines in (d). Here $i_m = A, B$, $m = 1, 2, 3, 4$.

copolymer chain $N_{i_n \dots i_1}$ can be similarly defined. $\dot{N}_{i_n \dots i_1} \equiv J_{i_n \dots i_1} = k_{i_1} P_{i_n \dots i_2} - \bar{k}_{i_1} P_{i_n \dots i_1}$. In general, we have $P_{i_n \dots i_1} = P_{A i_n \dots i_1} + P_{B i_n \dots i_1}$, $J_{i_n \dots i_1} = J_{A i_n \dots i_1} + J_{B i_n \dots i_1}$. We also define $\tilde{J}_{i_n \dots i_1} \equiv J_{i_n \dots i_1 A} + J_{i_n \dots i_1 B}$ and the overall sequence distribution $Q_{i_n \dots i_1} \equiv N_{i_n \dots i_1} / N = J_{i_n \dots i_1} / J_{\text{tot}}$.

The kinetic equations of $P_{i_n \dots i_1}$ ($n \geq 1$) can be written as

$$\dot{P}_{i_n \dots i_1} = J_{i_n \dots i_1} - \tilde{J}_{i_n \dots i_1} \quad (2)$$

For example,

$$\begin{aligned} \dot{P}_A &= J_A - \tilde{J}_{A^*} = J_{BA} - J_{AB} \\ &= k_A P_B - k_B P_A + \bar{k}_B P_{AB} - \bar{k}_A P_{BA} \end{aligned} \quad (3)$$

The existence of depropagation rates \bar{k}_A, \bar{k}_B makes these equations hierarchically coupled and hard to solve. Fortunately, for steady-state copolymerization $\dot{P}_{i_n \dots i_1} = 0$ (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, $P_{i_n \dots i_1} = \prod_{m=1}^n P_{i_m}$, which leads to

$$J_{i_n \dots i_1} = \left(\prod_{m=2}^n P_{i_m} \right) J_{i_1}, \quad \tilde{J}_{i_n \dots i_1} = \left(\prod_{m=1}^n P_{i_m} \right) J_{\text{tot}} \quad (4)$$

From the steady-state kinetic equation $0 = \dot{P}_{i_n \dots i_1} = J_{i_n \dots i_1} - \tilde{J}_{i_n \dots i_1}$, we can get

$$\frac{J_{i_1}}{P_{i_1}} = J_{\text{tot}} \quad (5)$$

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

$$\frac{J_A}{P_A} = \frac{J_B}{P_B} \quad (6)$$

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 $P_{i_1}, P_{i_2 i_1}$, etc) from a number of simulations. For simplicity, we only check the factorizations $P_{i_2 i_1} = P_{i_2} P_{i_1}$, $P_{i_3 i_2 i_1} = P_{i_3} P_{i_2} P_{i_1}$ and $P_{i_4 i_3 i_2 i_1} = P_{i_4} P_{i_3} P_{i_2} P_{i_1}$. As shown in figures 1(a)–(c), for arbitrary choice of rate parameters $k_A, k_B, \bar{k}_A, \bar{k}_B$ (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 depropagation of monomer

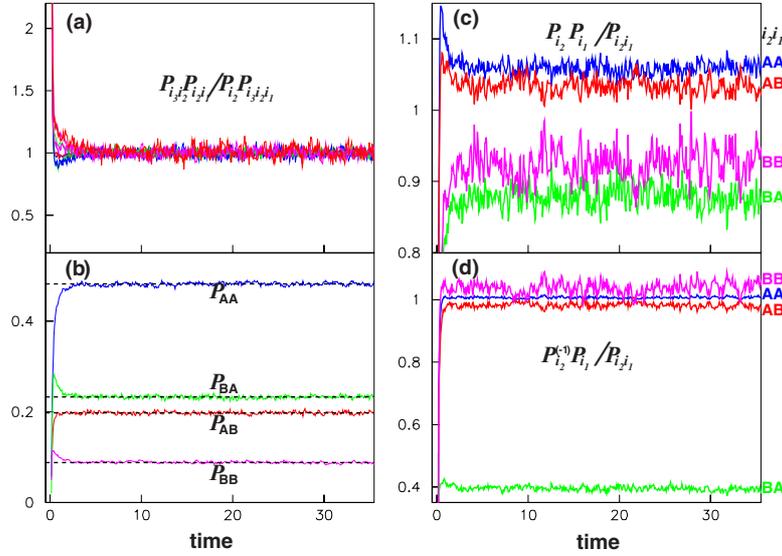
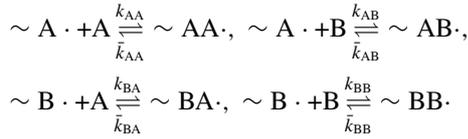


Figure 2. Simulation verification of the first-order factorization conjecture (a) and steady-state equations (b), with illustrative rate parameters $k_{AA} = 6.0, k_{AB} = 2.0, k_{BA} = 4.0, k_{BB} = 1.0, \bar{k}_{AA} = 7.0, \bar{k}_{AB} = 5.0, \bar{k}_{BA} = 3.0, \bar{k}_{BB} = 1.0$. Theoretical values of $P_{AA}, P_{AB}, P_{BA}, P_{BB}$ given by steady-state equations are indicated as dashed lines in (b). The zero-order factorization used in Bernoullian model fails in terminal model, as shown in (c). Direct factorization (see the text) also fails in terminal model, as shown in (d). Here $i_m = A, B, m = 1, 2, 3$.

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



Defining

$$\begin{aligned} J_{i_n \dots i_2 i_1} &\equiv k_{i_2 i_1} P_{i_n \dots i_2} - \bar{k}_{i_2 i_1} P_{i_n \dots i_2 i_1} \\ \tilde{J}_{i_n \dots i_2 i_1} &\equiv J_{i_n \dots i_2 i_1 A} + J_{i_n \dots i_2 i_1 B} \end{aligned} \quad (7)$$

where $i_m = A, B (m = 1, 2, \dots, n; n \geq 2)$, we can write the corresponding kinetic equations for $P_{i_n \dots i_2 i_1} (n \geq 1)$ as below

$$\dot{P}_{i_n \dots i_2 i_1} = J_{i_n \dots i_2 i_1} - \tilde{J}_{i_n \dots i_2 i_1} \quad (8)$$

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

$$P_{i_n \dots i_2 i_1} = \prod_{m=2}^n P_{i_m i_{m-1}} \left[\prod_{m=3}^n P_{i_{m-1}} \right]^{-1}, \quad n \geq 3 \quad (9)$$

Then the steady-state kinetic equations $\dot{P}_{i_n \dots i_2 i_1} = 0 (n \geq 2)$ are reduced to

$$\frac{J_{i_2 i_1}}{P_{i_2 i_1}} = \frac{\tilde{J}_{i_1}}{P_{i_1}} \quad (10)$$

or equivalently,

$$\frac{J_{Ai}}{P_{Ai}} = \frac{J_{Bi}}{P_{Bi}} \quad (11)$$

where $i = A, B$.

$\dot{P}_A = J_A - \tilde{J}_{A^*}$ or $\dot{P}_B = J_B - \tilde{J}_{B^*}$ (they are equivalent since $P_A + P_B = 1$) leads to another steady-state equation $J_{AB} = J_{BA}$. Finally, we get four equations for four variables

$$\begin{aligned} \frac{J_{AA}}{P_{AA}} &= \frac{J_{BA}}{P_{BA}}, \quad \frac{J_{AB}}{P_{AB}} = \frac{J_{BB}}{P_{BB}}, \quad J_{AB} = J_{BA}, \\ P_{AA} + P_{AB} + P_{BA} + P_{BB} &= 1 \end{aligned} \quad (12)$$

which gives the solution of the original kinetic equations.

The validity of the factorization $P_{i_3 i_2 i_1} = P_{i_3 i_2} P_{i_2 i_1} / P_{i_2}$ 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 $J_{\text{tot}} > 0$).

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 $P_{ij} = P_i^{(-1)} P_j (P_i^{(-1)})$ 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_{ij} as the basic variables to describe the terminal effect.

The first-order factorization conjecture $P_{i_3 i_2 i_1} = P_{i_3 i_2} P_{i_2 i_1} / P_{i_2}$ is actually equivalent to the first-order Markov-chain assumption used in [14] and [20]. Defining transition probability $p(i_2 | i_1) \equiv P_{i_2 i_1} / P_{i_1}, p(A | i_1) + p(B | i_1) = 1$,

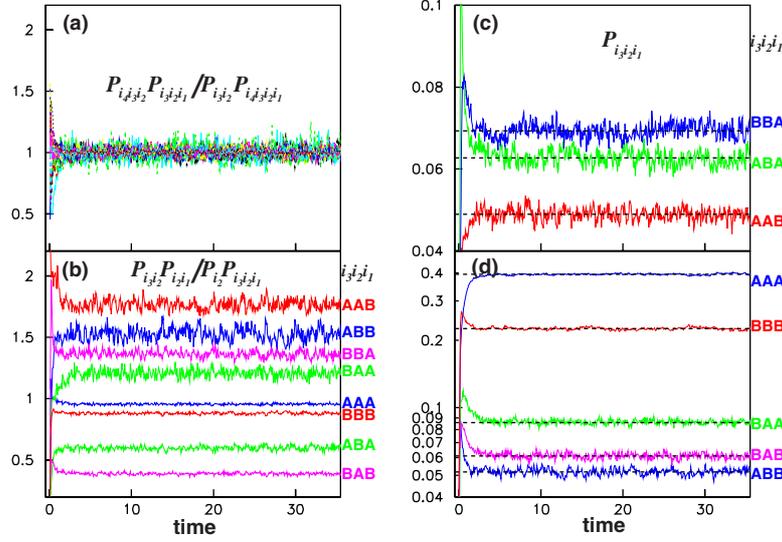


Figure 3. Simulation verification of the second-order factorization conjecture (a) and steady-state equations (c,d), with illustrative rate parameters $k_{AAA} = 8.0$, $k_{ABA} = 4.0$, $k_{BAA} = 7.0$, $k_{BBA} = 2.0$, $k_{AAB} = 1.0$, $k_{ABB} = 3.0$, $k_{BAB} = 5.0$, $k_{BBB} = 6.0$, $\bar{k}_{AAA} = 7.0$, $\bar{k}_{ABA} = 3.0$, $\bar{k}_{BAA} = 8.0$, $\bar{k}_{BBA} = 4.0$, $\bar{k}_{AAB} = 5.0$, $\bar{k}_{ABB} = 1.0$, $\bar{k}_{BAB} = 6.0$, $\bar{k}_{BBB} = 2.0$. (b) implies that first-order factorization $P_{i_3 i_2 i_1} = P_{i_3 i_2} P_{i_2 i_1} / P_{i_2}$ is invalid in penultimate model. In (c,d), the theoretical values of $P_{i_3 i_2 i_1}$ given by steady-state equations are indicated by dashed lines.

we now can rewrite the factorization conjecture as $P_{i_3 i_2 i_1} = p(i_3|i_2)p(i_2|i_1)P_{i_1}$. It is worth noting that we have chosen $P_{i_2 i_1}$, rather than P_{i_1} 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



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

$$\begin{aligned}
 J_{i_n \dots i_3 i_2 i_1} &\equiv k_{i_3 i_2 i_1} P_{i_n \dots i_3 i_2} - \bar{k}_{i_3 i_2 i_1} P_{i_n \dots i_3 i_2 i_1} \\
 \tilde{J}_{i_n \dots i_2 i_1} &\equiv J_{i_n \dots i_2 i_1 A} + J_{i_n \dots i_2 i_1 B}
 \end{aligned} \quad (13)$$

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

$$\dot{P}_{i_n \dots i_3 i_2 i_1} = J_{i_n \dots i_3 i_2 i_1} - \tilde{J}_{i_n \dots i_3 i_2 i_1} \quad (14)$$

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

$$P_{i_n \dots i_3 i_2 i_1} = \prod_{m=3}^n P_{i_m i_{m-1} i_{m-2}} \left[\prod_{m=4}^n P_{i_{m-1} i_{m-2}} \right]^{-1}, \quad n \geq 4 \quad (15)$$

The steady-state kinetic equation $\dot{P}_{i_n \dots i_3 i_2 i_1} = 0$ ($n \geq 3$) can thus be reduced to

$$\frac{J_{i_3 i_2 i_1}}{P_{i_3 i_2 i_1}} = \frac{\tilde{J}_{i_2 i_1}}{P_{i_2 i_1}} \quad (16)$$

or equivalently

$$\frac{J_{A i_2 i_1}}{P_{A i_2 i_1}} = \frac{J_{B i_2 i_1}}{P_{B i_2 i_1}} \quad (17)$$

Now we have had five independent equations (equation (17), along with the normalization condition) for the eight variables. The rest of the three equations come from the remaining kinetic equations of $P_{i_2 i_1}$

$$\dot{P}_{i_2 i_1} = J_{i_2 i_1} - \tilde{J}_{i_2 i_1} = 0 \quad (18)$$

These four equations are not independent, due to the normalization condition $\sum P_{i_2 i_1} = 1$. So any three of them can be selected to form a closed set of equations of $P_{i_3 i_2 i_1}$, for instance,

$$\begin{aligned}
 \frac{J_{AAA}}{P_{AAA}} &= \frac{J_{BAA}}{P_{BAA}}, \quad \frac{J_{AAB}}{P_{AAB}} = \frac{J_{BAB}}{P_{BAB}}, \\
 \frac{J_{ABA}}{P_{ABA}} &= \frac{J_{BBA}}{P_{BBA}}, \quad \frac{J_{ABB}}{P_{ABB}} = \frac{J_{BBB}}{P_{BBB}}, \\
 J_{AA} &= \tilde{J}_{AA*}, \quad J_{AB} = \tilde{J}_{AB*}, \\
 J_{BB} &= \tilde{J}_{BB*}, \quad \sum_{i_3, i_2, i_1=A, B} P_{i_3 i_2 i_1} = 1
 \end{aligned} \quad (19)$$

We checked the validity of $P_{i_4 i_3 i_2 i_1} = P_{i_4 i_3 i_2} P_{i_3 i_2 i_1} / P_{i_3 i_2}$ and the steady-state equations by Monte Carlo simulations. Figure 3(a) shows the second-order factorization holds when copolymerization reaches the steady state, figure 3(c,d) shows a good agreement between the simulated values and the theoretical values of $P_{i_3 i_2 i_1}$.

It is worth noting that, as the zero-order factorization is inapplicable to first-order model, the first-order factorization

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 $P_{i_3 i_2 i_1}$ ($i_3, i_2, i_1 = A, B$) are required. Li's model oversimplifies the problem to 6 variables (the two-terminal probability P_{i_1} 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 $\dot{P}_{i_2 i_1} = 0$. These equations are doubtful: two of the 16 rate parameters k_{AAA}, k_{BBB} are totally absent from the original four kinetic equations $\dot{P}_{i_2 i_1} = J_{i_2 i_1} - \tilde{J}_{i_2 i_1^*}$, 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., $\dot{P}_{i_3 i_2 i_1} = 0$), self-inconsistency of their theory can be further uncovered. For instance, in the extreme case $\bar{k}_{i_3 i_2 i_1} = 0$. Under the first-order Markov chain assumption, $\dot{P}_{AAA} = k_{AAA} P_{BAA} - k_{AAB} P_{AAA} = 0$ yields $k_{AAA} P_{BA} = k_{AAB} P_{AA}$, $\dot{P}_{BAA} = k_{BAA} P_{BA} - (k_{AAA} + k_{AAB}) P_{BAA} = 0$ yields $k_{BAA} P_A = (k_{AAA} + k_{AAB}) P_{AA}$. 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 th-order factorization conjecture ($m < s$) is applied to the steady-state kinetic equations $\dot{P}_{i_n \dots i_1} = 0$ ($n = 1, 2, \dots$) 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 depropagation of A/B depend on the last s monomer units of the copolymer. Here the basic variables are $P_{i_{s+1} i_s \dots i_1}$ (2^{s+1} in total). We denote the propagation rates as $k_{i_{s+1} i_s \dots i_1}$ and depropagation rates as $\bar{k}_{i_{s+1} i_s \dots i_1}$, and also

$$\begin{aligned} J_{i_n \dots i_3 i_2 i_1} &\equiv k_{i_{s+1} \dots i_3 i_2 i_1} P_{i_n \dots i_3 i_2} - \bar{k}_{i_{s+1} \dots i_3 i_2 i_1} P_{i_n \dots i_3 i_2 i_1} \\ \tilde{J}_{i_n \dots i_2 i_1^*} &\equiv J_{i_n \dots i_2 i_1 A} + J_{i_n \dots i_2 i_1 B} \end{aligned} \quad (20)$$

where $i_m = A, B$; $m = 1, 2, \dots, n$; $n \geq s + 1$.

The s th-order factorization conjecture is

$$P_{i_n \dots i_1} = \prod_{m=s+1}^n P_{i_m i_{m-1} \dots i_{m-s}} \left[\prod_{m=s+2}^n P_{i_{m-1} \dots i_{m-s}} \right]^{-1}, \quad n \geq s + 2 \quad (21)$$

The closed steady-state equations derived from $\dot{P}_{i_n \dots i_1} = J_{i_n \dots i_1} - \tilde{J}_{i_n \dots i_1^*} = 0$ ($n \geq s + 1$) are the following 2^s equations

$$\frac{J_{A i_s i_{s-1} \dots i_1}}{P_{A i_s i_{s-1} \dots i_1}} = \frac{J_{B i_s i_{s-1} \dots i_1}}{P_{B i_s i_{s-1} \dots i_1}} \quad (22)$$

or equivalently,

$$\frac{J_{i_{s+1} i_s \dots i_1}}{P_{i_{s+1} i_s \dots i_1}} = \frac{J_{i_s i_{s-1} \dots i_1}}{P_{i_s i_{s-1} \dots i_1}} \quad (23)$$

The kinetic equations $\dot{P}_{i_s \dots i_1} = J_{i_s \dots i_1} - \tilde{J}_{i_s \dots i_1^*} = 0$ give other 2^s steady-state equations, from which any $2^s - 1$ equations can be chosen. Combining the normalization condition $\sum P_{i_{s+1} i_s \dots i_1} = 1$, 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 $p(i_{s+1}|i_s \dots i_1) \equiv P_{i_{s+1} i_s \dots i_1} / P_{i_s \dots i_1}$, $p(A|i_s \dots i_1) + p(B|i_s \dots i_1) = 1$. Noting that $J_{i_s i_{s-1} \dots i_1^*} = J_{i_s i_{s-1} \dots i_1}$, the steady-state equations equation (23) can be transformed into

$$\frac{J_{i_{s+1} i_s \dots i_1}}{J_{i_s \dots i_1}} = \frac{P_{i_{s+1} i_s \dots i_1}}{P_{i_s \dots i_1}} \quad (24)$$

Since the overall sequence distribution $Q_{i_{s+1} i_s \dots i_1} / Q_{i_s \dots i_1} = J_{i_{s+1} i_s \dots i_1} / J_{i_s \dots i_1}$, the steady-state equations can be rewritten as

$$\frac{Q_{i_{s+1} i_s \dots i_1}}{Q_{i_s \dots i_1}} = \frac{P_{i_{s+1} i_s \dots i_1}}{P_{i_s \dots i_1}} = p(i_{s+1}|i_s \dots i_1) \quad (25)$$

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 $(s - 1)$ th-order model if assuming $k_{A i_s \dots i_1} = k_{B i_s \dots i_1} = k_{i_s \dots i_1}$ and $\bar{k}_{A i_s \dots i_1} = \bar{k}_{B i_s \dots i_1} = \bar{k}_{i_s \dots i_1}$. By the s th-order model, we have

$$\frac{P_{A i_s \dots i_2 i_1}}{P_{B i_s \dots i_2 i_1}} = \frac{J_{A i_s \dots i_2 i_1}}{J_{B i_s \dots i_2 i_1}} = \frac{k_{i_s \dots i_2 i_1} P_{A i_s \dots i_2} - \bar{k}_{i_s \dots i_2 i_1} P_{A i_s \dots i_2 i_1}}{k_{i_s \dots i_2 i_1} P_{B i_s \dots i_2} - \bar{k}_{i_s \dots i_2 i_1} P_{B i_s \dots i_2 i_1}} \quad (26)$$

which yields

$$\frac{P_{A i_s \dots i_2 i_1}}{P_{B i_s \dots i_2 i_1}} = \frac{P_{A i_s \dots i_2}}{P_{B i_s \dots i_2}} \quad (27)$$

or equivalently

$$\frac{P_{A i_s \dots i_2 i_1}}{P_{A i_s \dots i_2}} = \frac{P_{B i_s \dots i_2 i_1}}{P_{B i_s \dots i_2}} = \frac{P_{i_s \dots i_2 i_1}}{P_{i_s \dots i_2}} \quad (28)$$

This means $P_{i_{s+1} i_s \dots i_2 i_1} = P_{i_{s+1} \dots i_2} P_{i_s \dots i_1} / P_{i_s \dots i_2}$ which is exactly the $(s - 1)$ th-order factorization conjecture.

Equation (27) also leads to

$$\frac{P_{A i_s \dots i_2}}{P_{B i_s \dots i_2}} = \frac{J_{A i_s \dots i_2 A}}{J_{B i_s \dots i_2 A}} = \frac{J_{A i_s \dots i_2 B}}{J_{B i_s \dots i_2 B}} = \frac{\tilde{J}_{A i_s \dots i_2^*}}{\tilde{J}_{B i_s \dots i_2^*}} = \frac{J_{A i_s \dots i_2}}{J_{B i_s \dots i_2}} \quad (29)$$

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, \dots, M_l). In Bernoullian model, for instance, the basic variables are P_{M_j} ($j = 1, 2, \dots, l$). We have the following l equations for the l variables.

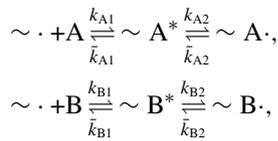
$$\frac{J_{M_1}}{P_{M_1}} = \frac{J_{M_2}}{P_{M_2}} = \dots = \frac{J_{M_i}}{P_{M_i}} = \dots = \frac{J_{M_l}}{P_{M_l}}, \quad \sum_{i=1}^l P_{M_i} = 1 \quad (30)$$

where $J_{M_i} \equiv k_{M_i} - \bar{k}_{M_i} P_{M_i}$.

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 such as DNA replication are often multi-step processes (e.g., [19]). In the latter case, the factorization conjectures can also be applied and similar steady-state equations can be derived. For simplicity and without loss of generality, we only discuss a simple case in Bernoullian model, as below



here EA^* , EB^* represent the intermediate states. Now there are four possible states A, B, A^* , B^* at the terminal, the corresponding probabilities are $P_A, P_{A^*}, P_B, P_{B^*}$. We now have kinetic equations very similar to equation (2), for instance,

$$\begin{aligned} \dot{P}_m &= J_m^* - \tilde{J}_{m^*}, \quad i, m = A, B \\ \dot{P}_{m^*} &= J_{m^*} - J_m \\ J_{im^*} &\equiv k_{m1} P_i - \bar{k}_{m1} P_{im^*} \\ J_m^* &\equiv J_{Am^*} + J_{Bm^*} \\ \tilde{J}_{m^*} &\equiv J_{mA^*} + J_{mB^*} \\ J_m &\equiv k_{m2} P_{m^*} - \bar{k}_{m2} P_m \end{aligned} \quad (31)$$

In steady state, $\dot{P}_{A^*} = 0, \dot{P}_{B^*} = 0$, i.e., we get two steady-state equations $J_{A^*} = J_A, J_{B^*} = J_B$. Furthermore, assuming $P_{ij} = \pi_i P_j$ and $P_{ij^*} = \pi_i P_{j^*}$, where $\pi_i \equiv P_i / (P_A + P_B)$, one can again derive $J_A / J_B = \pi_A / \pi_B$, or equivalently $J_A / P_A = J_B / P_B$, from $\dot{P}_A = 0$ or $\dot{P}_B = 0$. Combining the normalization condition $P_A + P_{A^*} + P_B + P_{B^*} = 1$, we thus have four equations for $P_A, P_{A^*}, P_B, P_{B^*}$. It is worth noting that the above zero-order factorization conjectures generally hold for any multi-step processes in Bernoullian model.

Similarly, for the terminal model of multi-step processes, one can still assume the first-order factorization conjectures $P_{ijk} = P_{ij} P_{jk} / P_j$ and $P_{ijk^*} = P_{ij} P_{jk^*} / P_j$, etc. This logic can be directly extended to any higher-order models.

5. Steady-state thermodynamics

Recently, the steady-state thermodynamics of copolymerization has been discussed from the perspective of information theory [17, 18, 20]. In these works, the authors proposed very general relations between the information of the copolymer sequence and the entropy production of the copolymerization system (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 is omitted for simplicity)

$$\begin{aligned} \Delta G &= \frac{\dot{S}}{J_{\text{tot}}} \\ \dot{S} &\equiv J_A \ln \left(\frac{k_A^0 [M]}{\bar{k}_A P_A} \right) + J_B \ln \left(\frac{k_B^0 [M]}{\bar{k}_B P_B} \right) \\ J_A &\equiv k_A^0 [M] - \bar{k}_A P_A, \quad J_B \equiv k_B^0 [M] - \bar{k}_B P_B \\ J_{\text{tot}} &\equiv J_A + J_B \end{aligned} \quad (32)$$

\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 $J_A = k_A^0 [M]^{\text{eq}} - \bar{k}_A P_A^{\text{eq}} = 0$ and $J_B = k_B^0 [M]^{\text{eq}} - \bar{k}_B P_B^{\text{eq}} = 0$, which yield $[M]^{\text{eq}} = (K_A + K_B)^{-1}$, $P_A^{\text{eq}} = K_A [M]^{\text{eq}}$ and $P_B^{\text{eq}} = K_B [M]^{\text{eq}}$, 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 $[M] > [M]^{\text{eq}}$.

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

$$\Delta G = P_A \cdot \ln \left(\frac{P_A^{\text{eq}} [M]}{P_A [M]^{\text{eq}}} \right) + P_B \cdot \ln \left(\frac{P_B^{\text{eq}} [M]}{P_B [M]^{\text{eq}}} \right) \quad (33)$$

this leads to the following equality

$$\begin{aligned} \Delta G &= \Delta \Psi - \Delta I \\ \Delta \Psi &\equiv \ln \left(\frac{[M]}{[M]^{\text{eq}}} \right) \\ \Delta I &\equiv P_A \ln \left(\frac{P_A}{P_A^{\text{eq}}} \right) + P_B \ln \left(\frac{P_B}{P_B^{\text{eq}}} \right) \end{aligned} \quad (34)$$

ΔI 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

defined as Shannon entropy $I_N = -\sum_{S_N} Q(S_N)\ln Q(S_N)$, here the summation runs on all possible sequence $S_N = i_N i_{N-1} \cdots i_1$ ($i_m = A, B$) of length N , $Q(S_N)$ is the overall occurrence probability of S_N in the copolymer sequence.

For the Bernoullian model, $Q(S_N) = \prod_{n=1}^N P_{i_n}$. 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 $\Delta I_N = \sum Q(S_N) \ln [Q(S_N)/Q^{\text{eq}}(S_N)]$, namely,

$$\Delta I_N = N \left(P_A \ln \frac{P_A}{P_A^{\text{eq}}} + P_B \ln \frac{P_B}{P_B^{\text{eq}}} \right) = N \Delta I \quad (35)$$

or

$$\Delta I = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{S_N} Q(S_N) \ln \frac{Q(S_N)}{Q^{\text{eq}}(S_N)} \quad (36)$$

Hence, ΔI 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 ΔG per incorporation. In the two-step process discussed in section 4, for instance, ΔG can be written as (using the same notations as in section 4)

$$\begin{aligned} \Delta G &= \dot{S}/J_{\text{tot}} \\ J_{\text{tot}} &\equiv J_A + J_B \\ \dot{S} &\equiv J_A \ln \left[\frac{k_{A1} k_{A2} (P_A + P_B)}{\bar{k}_{A1} \bar{k}_{A2} P_A} \right] + J_B \ln \left[\frac{k_{B1} k_{B2} (P_A + P_B)}{\bar{k}_{B1} \bar{k}_{B2} P_B} \right] \\ &= J_{A^*} \ln \left[\frac{k_{A1} (P_A + P_B)}{\bar{k}_{A1} P_{A^*}} \right] + J_A \ln \left[\frac{k_{A2} P_{A^*}}{\bar{k}_{A2} P_A} \right] \\ &+ J_{B^*} \ln \left[\frac{k_{B1} (P_A + P_B)}{\bar{k}_{B1} P_{B^*}} \right] + J_B \ln \left[\frac{k_{B2} P_{B^*}}{\bar{k}_{B2} P_B} \right] \end{aligned} \quad (37)$$

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 $J_{\text{tot}} = J_{A^*} + J_A + J_{B^*} + J_B$, since the latter is the total flux of all the involved reactions but not the flux of incorporation, we have

$$\Delta G = \pi_A \ln \left[\frac{k_{A1} k_{A2}}{\bar{k}_{A1} \bar{k}_{A2} \pi_A} \right] + \pi_B \ln \left[\frac{k_{B1} k_{B2}}{\bar{k}_{B1} \bar{k}_{B2} \pi_B} \right] \quad (38)$$

One can similarly define the information gain ΔI and the driving force $\Delta \Psi$,

$$\begin{aligned} \Delta I &\equiv \pi_A \ln \left(\frac{\pi_A}{\pi_A^{\text{eq}}} \right) + \pi_B \ln \left(\frac{\pi_B}{\pi_B^{\text{eq}}} \right) \\ \Delta \Psi &\equiv \ln \left(\frac{[M]}{[M]^{\text{eq}}} \right) \end{aligned} \quad (39)$$

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 $J_A = k_A^0 [A]^{\text{eq}} - \bar{k}_A P_A^{\text{eq}} = 0$ and $J_B = k_B^0 [B]^{\text{eq}} - \bar{k}_B P_B^{\text{eq}} = 0$, 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 $[A]^{\text{eq}}$, $[B]^{\text{eq}}$, and then determine the corresponding P_A^{eq} , P_B^{eq} to define the equilibrium state. The thermodynamic equality still holds, only with some modifications. For instance, for single-step process, we have

$$\begin{aligned} \Delta \Psi &= \Delta G + \Delta I \\ \Delta G &\equiv \frac{J_A}{J_{\text{tot}}} \ln \left(\frac{k_A^0 [A]}{\bar{k}_A P_A} \right) + \frac{J_B}{J_{\text{tot}}} \ln \left(\frac{k_B^0 [B]}{\bar{k}_B P_B} \right) > 0 \\ \Delta I &\equiv P_A \ln \left(\frac{P_A}{P_A^{\text{eq}}} \right) + P_B \ln \left(\frac{P_B}{P_B^{\text{eq}}} \right) > 0 \\ \Delta \Psi &\equiv P_A \ln \left(\frac{[A]}{[A]^{\text{eq}}} \right) + P_B \ln \left(\frac{[B]}{[B]^{\text{eq}}} \right) \end{aligned} \quad (40)$$

Last but not least, the information interpretation of ΔI is totally based on the s th-order factorization conjecture of the s th-order terminal model. Any lower-order factorization conjectures are 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 depropagation 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

$$J_{i_{s+1}i_s \dots i_1} = k_{i_{s+1}i_s \dots i_1}^0 [M]^{eq} P_{i_{s+1}i_s \dots i_2}^{eq} - \bar{k}_{i_{s+1}i_s \dots i_1} P_{i_{s+1}i_s \dots i_1}^{eq} = 0 \quad (41)$$

where $i_m = A, B$ ($m = 1, 2, \dots, s+1$). $[M]^{eq}$, $P_{i_{s+1}i_s \dots i_1}^{eq}$ can be directly solved from these equations. They are functions only of equilibrium constants.

So we have

$$\frac{k_{i_{s+1}i_s \dots i_1}^0}{\bar{k}_{i_{s+1}i_s \dots i_1}} = \frac{P_{i_{s+1}i_s \dots i_1}^{eq}}{[M]^{eq} P_{i_{s+1}i_s \dots i_2}^{eq}} \quad (42)$$

The averaged free energy dissipation is defined

$$\Delta G \equiv \frac{1}{J_{tot}} \sum_{\substack{i_m=A,B \\ m=1, \dots, s+1}} J_{i_{s+1} \dots i_1} \ln \left(\frac{k_{i_{s+1} \dots i_1}^0 [M] P_{i_{s+1} \dots i_2}}{\bar{k}_{i_{s+1} \dots i_1} P_{i_{s+1} \dots i_1}} \right) \quad (43)$$

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

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

$$\Delta \Psi \equiv \frac{1}{J_{tot}} \sum J_{i_{s+1} \dots i_1} \ln \left(\frac{[M]}{[M]^{eq}} \right) = \ln \left(\frac{[M]}{[M]^{eq}} \right) \quad (44)$$

$$\Delta I \equiv \frac{1}{J_{tot}} \sum J_{i_{s+1} \dots i_1} \ln \left(\frac{P_{i_{s+1} \dots i_1} / P_{i_{s+1} \dots i_2}}{P_{i_{s+1} \dots i_1}^{eq} / P_{i_{s+1} \dots i_2}^{eq}} \right)$$

ΔI is also the information gain per incorporation step

$$\Delta I = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i_N \dots i_1} Q_{i_N \dots i_1} \ln \frac{Q_{i_N \dots i_1}}{Q_{i_N \dots i_1}^{eq}} \quad (45)$$

To prove this, we transform equation (44) into

$$\Delta I = \frac{1}{J_{tot}} \sum J_{i_{s+1} \dots i_1} \ln \left(\frac{P_{i_{s+1} \dots i_1} / P_{i_s \dots i_1}}{P_{i_{s+1} \dots i_1}^{eq} / P_{i_s \dots i_1}^{eq}} \right) + \frac{1}{J_{tot}} \sum J_{i_{s+1} \dots i_1} \ln \left(\frac{P_{i_s \dots i_1} / P_{i_{s+1} \dots i_2}}{P_{i_s \dots i_1}^{eq} / P_{i_{s+1} \dots i_2}^{eq}} \right) \quad (46)$$

In the second term

$$\begin{aligned} & \sum_{i_{s+1}, \dots, i_1} J_{i_{s+1} \dots i_1} \ln \left(\frac{P_{i_s \dots i_1}}{P_{i_{s+1} \dots i_2}} \right) \\ &= \sum_{i_{s+1}, \dots, i_1} J_{i_{s+1} \dots i_1} \ln P_{i_s \dots i_1} - \sum_{i_{s+1}, \dots, i_1} J_{i_{s+1} \dots i_1} \ln P_{i_{s+1} \dots i_2} \\ &= \sum_{i_s, \dots, i_1} J_{i_s \dots i_1} \ln P_{i_s \dots i_1} - \sum_{i_{s+1}, \dots, i_2} \tilde{J}_{i_{s+1} \dots i_2} \ln P_{i_{s+1} \dots i_2} \\ &= \sum_{i_s, \dots, i_1} J_{i_s \dots i_1} \ln P_{i_s \dots i_1} - \sum_{j_s, \dots, j_1} \tilde{J}_{j_s \dots j_1} \ln P_{j_s \dots j_1} \\ &= \sum_{i_s, \dots, i_1} (J_{i_s \dots i_1} - \tilde{J}_{i_s \dots i_1}) \ln P_{i_s \dots i_1} = 0 \end{aligned} \quad (47)$$

In the third step, we have substituted i_m by j_{m-1} ($m = 2, 3 \dots, s+1$). In the last step, we have used the steady-state conditions $J_{i_s \dots i_1} = \tilde{J}_{i_s \dots i_1}$.

Similarly, one can show

$$\sum_{i_{s+1}, \dots, i_1} J_{i_{s+1} \dots i_1} \ln \left(\frac{P_{i_s \dots i_1}^{eq}}{P_{i_{s+1} \dots i_2}^{eq}} \right) = 0 \quad (48)$$

Therefore, the second summation term in equation(46) is zero. So we have

$$\Delta I = \frac{1}{J_{tot}} \sum J_{i_{s+1} \dots i_1} \ln \left(\frac{P_{i_{s+1} \dots i_1} / P_{i_s \dots i_1}}{P_{i_{s+1} \dots i_1}^{eq} / P_{i_s \dots i_1}^{eq}} \right)$$

We define the transition probability as $p(i_{s+1}|i_s \dots i_1) \equiv P_{i_{s+1} \dots i_1} / P_{i_s \dots i_1}$. Since $Q_{i_{s+1} \dots i_1} = J_{i_{s+1} \dots i_1} / J_{tot}$ and $Q_{i_{s+1} \dots i_1} / Q_{i_s \dots i_1} = P_{i_{s+1} \dots i_1} / P_{i_s \dots i_1} = p(i_{s+1}|i_s \dots i_1)$ (equation (25)), we rewrite the above equation as

$$\begin{aligned} \Delta I &= \sum_{i_{s+1}, \dots, i_1} Q_{i_{s+1} \dots i_1} \ln \left(\frac{p(i_{s+1}|i_s \dots i_1)}{p^{eq}(i_{s+1}|i_s \dots i_1)} \right) \quad (49) \\ &= \sum_{i_{s+1}, \dots, i_1} Q_{i_s \dots i_1} p(i_{s+1}|i_s \dots i_1) \ln \left(\frac{p(i_{s+1}|i_s \dots i_1)}{p^{eq}(i_{s+1}|i_s \dots i_1)} \right) \\ &= \sum_{i_s \dots i_1} Q_{i_s \dots i_1} \left[p(A|i_s \dots i_1) \ln \left(\frac{p(A|i_s \dots i_1)}{p^{eq}(A|i_s \dots i_1)} \right) \right. \\ &\quad \left. + p(B|i_s \dots i_1) \ln \left(\frac{p(B|i_s \dots i_1)}{p^{eq}(B|i_s \dots i_1)} \right) \right] \end{aligned}$$

Since $p(A|i_s \dots i_1) + p(B|i_s \dots i_1) = 1$ and $p^{eq}(A|i_s \dots i_1) + p^{eq}(B|i_s \dots i_1) = 1$, it can be readily proven that ΔI is non-negative. Furthermore, one can show ΔI 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 $Q_{i_N \dots i_1} = \left[\prod_{n=s+1}^N p(i_n|i_{n-1} \dots i_{n-s}) \right] Q_{i_s \dots i_1}$, we have

$$\begin{aligned} \Delta I_N &= \sum_{i_N, \dots, i_1} Q_{i_N \dots i_1} \left[\sum_{n=s+1}^N \ln \frac{p(i_n|i_{n-1} \dots i_{n-s})}{p^{eq}(i_n|i_{n-1} \dots i_{n-s})} \right] \\ &\quad + \sum_{i_s, \dots, i_1} Q_{i_s \dots i_1} \ln \left(\frac{Q_{i_s \dots i_1}}{Q_{i_s \dots i_1}^{eq}} \right) \end{aligned} \quad (50)$$

The first summation term can be rewritten as

$$\begin{aligned} & \sum_{n=s+1}^N \left[\sum_{i_n, \dots, i_{n-s}} \left(\sum_{\substack{i_N, \dots, i_{n+1}, \\ i_{n-s-1}, \dots, i_1}} Q_{i_N \dots i_1} \right) \ln \frac{p(i_n|i_{n-1} \dots i_{n-s})}{p^{eq}(i_n|i_{n-1} \dots i_{n-s})} \right] \\ &= \sum_{n=s+1}^N \left[\sum_{i_n, \dots, i_{n-s}} Q_{i_n \dots i_{n-s}} \ln \frac{p(i_n|i_{n-1} \dots i_{n-s})}{p^{eq}(i_n|i_{n-1} \dots i_{n-s})} \right] \\ &= (N-s) \sum_{i_{s+1}, \dots, i_1} Q_{i_{s+1} \dots i_1} \ln \left(\frac{p(i_{s+1}|i_s \dots i_1)}{p^{eq}(i_{s+1}|i_s \dots i_1)} \right) \end{aligned} \quad (51)$$

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

$$\begin{aligned} \Delta I &= \lim_{N \rightarrow \infty} \frac{1}{N} \Delta I_N \\ &= \sum_{i_{s+1}, \dots, i_1} Q_{i_{s+1} \dots i_1} \ln \left(\frac{p(i_{s+1}|i_s \dots i_1)}{p^{eq}(i_{s+1}|i_s \dots i_1)} \right) \end{aligned} \quad (52)$$

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 O'Driscoll 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