

## Binding polynomial in molecular self-assembly

Andrew A. Mosunov,<sup>1</sup> Kristina A. Rybakova,<sup>1</sup> Olga V. Rogova,<sup>1</sup> and Maxim P. Evstigneev<sup>1,2,\*</sup>

<sup>1</sup>*Department of Physics, Sevastopol National Technical University, Sevastopol 99053, Russia*

<sup>2</sup>*Department of Biology and Chemistry, Belgorod State University, Belgorod 308015, Russia*

(Received 8 April 2014; published 27 June 2014)

In the present work the concept of a binding polynomial is revisited for the most widely used case of self-assembly of identical molecular units and results in the re-construction of a link to the grand partition function of such a system. It is found that if the self-assembly process is not pronounced (i.e., the product of the equilibrium constant and the monomer concentration is close to zero), the binding polynomial has the meaning of a molecular partition function that is given by the summation over energy levels of any molecule in the system. In other cases the concept of a binding polynomial may be misleading.

DOI: [10.1103/PhysRevE.89.062138](https://doi.org/10.1103/PhysRevE.89.062138)

PACS number(s): 05.70.-a, 61.43.Hv, 87.15.K-

### I. INTRODUCTION

Noncovalent molecular self-assembly resulting in the formation of ordered supramolecular structures is a key physicochemical process lying at the heart of modern supramolecular chemistry. The simplest but the most widely used partial case of molecular self-assembly is the aggregation of molecules. It is often modeled by the linear polymerization scheme, featuring a relatively simple mathematical description that enables a link to experimental observations to be defined for the majority of cases (for reviews see [1–3]).

It has now become a common practice to use the partition function formalism when dealing with linear polymerizations. The central point here is the concept of a binding polynomial  $Z_b$ , often referred to as a quasipartition function or simply a partition function, extensively reviewed in the literature [4,5]. Here  $Z_b$  is given by the summation over statistical weights (or sometimes over concentrations) of all possible types of complexes present in the system; for instance, for the simplest case of indefinite linear aggregation,  $Z_b$  is written as

$$Z_b = 1 + (Kc_1)^2 + \dots + (Kc_1)^i + \dots, \quad (1)$$

where  $c_1$  is the concentration of monomers,  $K$  is the equilibrium aggregation constant, and  $i \in 1, \dots, \infty$  is the number of molecules in an  $i$ -mer aggregate.

It is important to note that the construction of the binding polynomial uses the law of mass action allowing the setup of a link between the concentrations  $c_i$  of various complexes, containing  $i$  molecules

$$c_i = c_1(Kc_1)^{i-1}. \quad (2)$$

In Eq. (2) the equilibrium constant  $K$  is defined as  $K = \frac{c_i}{c_1 c_{i-1}}$  and is assumed to be independent of the number of molecules  $i$  in the given  $i$ -mer. The magnitudes of  $c_i$  are interconnected by the mass conservation law, i.e.,  $c_0 = \sum_{i=1}^{\infty} i c_i$ . It is noteworthy that in this equation [as well as in the binding polynomial (1)] the total number of molecules is always finite, whereas  $i \rightarrow \infty$ , which formally contains a contradiction. Such imprecision is quite common in the theory

of self-assembling systems [1–3] and is artificially introduced, enabling us to evaluate the equations for  $c_0$  and  $Z_b$  as a geometric series. It allows us to represent all expressions in a closed and simple analytical form. In order to preserve the finiteness of  $c_0$  (and  $Z_b$ ), the limit  $\lim_{i \rightarrow \infty} c_i = 0$  and the associated condition  $Kc_1 < 1$  should always be valid, thus incorporating a negligible error into Eq. (1) (as a consequence of the artificial condition  $i \rightarrow \infty$ ). The definition of  $K$ , the use of the mass action law, and the related assumptions around it and discussed above lie at the heart of the concept of a binding polynomial.

If the manual construction of  $Z_b$  is difficult, special methods can be used to obtain  $Z_b$  such as transfer-matrix formalism [6], the method of sequence-generating functions [7], an algorithmic approach [8], and some others [9,10]. Knowledge of  $Z_b$  allows us to define all necessary equilibrium characteristics of the system, e.g., to evaluate the total number of molecules  $c_0$  and to evaluate the total number of interfaces (contacts) between molecules, which are required to set up a link to experimental observables [10,11].

The simplicity in construction and evaluation of the binding polynomial has resulted in extensive use of this approach in various physicochemical applications, including multicomponent heterocomplexations [10–13], drug-biopolymer binding [4,5,7,9], and supramolecular polymerizations [14]. However, it is worth recalling that  $Z_b$  is not a true partition function  $Z$  and analysis of the literature suggests that its connection to  $Z$  seems to be lost. Blind use of  $Z_b$  has resulted in several confusing points, currently being discussed in the literature, e.g., the role of mixing entropy in molecular self-assembly (for a review see [15]) and the overestimation of certain types of complexes in heteroassociation (for a review see [11,16]). The following questions therefore arise. (i) What is the link between  $Z_b$  and  $Z$ ? (ii) Is there any physical meaning of  $Z_b$  as a partition function? (iii) What are the limitations of applying the concept of a binding polynomial to a particular system? In the present paper we shall try to answer these questions.

### II. RESULTS AND DISCUSSION

#### A. Grand partition function of the self-assembling system

We start by searching for the exact form of the true grand partition function  $Z$  of noncooperatively self-assembling

\*Author to whom correspondence should be addressed: max\_evstigneev@mail.ru

system. Let us define the self-assembling system as the system at isobaric-isothermal equilibrium containing  $N_0$  identical molecular units able to interact with each other and form complexes with an unrestricted number of molecules in them. Hence the equilibrium distribution of various oligomers is being formed; let the number of  $i$ -type oligomers containing  $i$  molecules be given by  $N_i$ , which are interconnected by the mass conservation law, i.e.,

$$N_0 = \sum_{i=1}^{\infty} i N_i. \quad (3)$$

Each interface between the molecules in any oligomer may be characterized by the Gibbs free-energy change (taking the monomer as a reference state) or binding constant  $K$ , i.e.,

$$\Delta G_1 = -kT \ln K, \quad (4)$$

where  $k$  is the Boltzmann constant. This allows us to define the net energy of the system as a sum of  $\Delta G_1$  over all types of oligomers, i.e.,  $\Delta G = \sum_{i=1}^{\infty} (i-1)N_i \Delta G_1$ . The latter enables us to write the grand partition function of the system as

$$Z = \sum_{\text{over all microstates}}, \dots, \sum \exp \left[ - \sum_{i=1}^{\infty} (i-1)N_i \frac{\Delta G_1}{kT} \right], \quad (5)$$

where the summation is executed over all microstates of the system that are isoenergetic.

Further evaluation of Eq. (5) requires finding the total number of isoenergetic microstates, which is equal to the number of ways needed to make up an equilibrium of various oligomers with the distribution  $N_1, N_2, \dots, N_i, \dots$ . In order to define this number we shall reevaluate  $Z$  as given below.

Let the partition function of an arbitrary selected oligomer made of  $i$  molecules be equal to

$$\exp \left( - \frac{(i-1)\Delta G_1}{kT} \right), \quad (6)$$

so the partition function of  $N_i$  indistinguishable oligomers is

$$\frac{[\exp(-\frac{(i-1)\Delta G_1}{kT})]^{N_i}}{N_i!}. \quad (7)$$

It is worth noting that the molecules forming the given oligomer now become distinguishable within this oligomer and their positions are not interchangeable. This means that the overall number of ways possible to form any  $i$ -mer is not accounted for in Eqs. (6) and (7) and any cooperative interaction of the molecules in aggregates is ignored. This is important in view of possible dependence of  $K$  on the number of molecules  $i$  in an aggregate and will be discussed below.

Multiplication of Eq. (7) over  $i \in 1, \dots, \infty$  yields

$$\frac{\exp(-\sum_{i=1}^{\infty} (i-1)N_i \frac{\Delta G_1}{kT})}{N_1! N_2! \dots}. \quad (8)$$

However, Eq. (8) is not yet a grand partition function. A constant formation and disruption of oligomers (i.e., the dynamic equilibrium) due to thermal fluctuations (e.g., in solution) must be taken into account. This means that the content of oligomers is constantly changing and every molecule during

the course of time can be found in any oligomer. Hence, Eq. (8) must be multiplied by  $N_0!$ , which accounts for all possible configurations of the system containing  $N_0$  molecules. Eventually, the grand partition function takes the final form

$$Z = \frac{N_0!}{N_1! N_2! \dots} \exp \left( - \sum_{i=1}^{\infty} (i-1)N_i \frac{\Delta G_1}{kT} \right). \quad (9)$$

A comparison of Eqs. (5) and (9) enables us to conclude that the factor  $\frac{N_0!}{N_1! N_2! \dots}$  may be associated with the total number of isoenergetic microstates cycled over in Eq. (5). It is simple to show that the partition function in Eq. (9) can be exactly evaluated into the grand partition function of linearly polymerizing systems derived a long time ago in [17] if the solvent molecules are explicitly taken into consideration [in Eq. (9) the interaction with the solvent is implicitly accounted for in  $\Delta G_1$ ].

It should be noted that the  $Z$  quantity in Eq. (9) in fact represents the grand subpartition function of specifically the equilibrium state, which is characterized by a certain distribution of oligomers  $N_1, N_2, \dots, N_i, \dots$  in the system, corresponding to the free-energy minimum (to be referred to in the following as the grand partition function). Evaluation of the grand partition function including all possible equilibrium and nonequilibrium states of the system will require summation of  $Z$  over all allowed values of  $N_i$  in Eq. (9) and is beyond the scope of the concept of the equilibrium binding polynomial in Eq. (1) considered in the present work.

The grand partition function in the form of Eq. (9) is the true partition function and is the central object that allows us to determine all equilibrium properties of the self-assembling systems. We shall further discuss some of these properties that are of importance to the understanding of the physical meaning of the binding polynomial.

## B. Equilibrium distribution of oligomers

Let us use the common assumption that the self-assembly process proceeds with negligible volume change. Taking the set of nonaggregated monomer molecules as a reference state, one can write the Gibbs free energy of the system as

$$\begin{aligned} G &= G_1 N_0 - kT \ln Z \\ &= G_1 N_0 + \sum_{i=1}^{\infty} (i-1)N_i \Delta G_1 - kT \ln \frac{N_0!}{N_1! N_2! \dots}. \end{aligned} \quad (10)$$

Recall that all oligomers are interconnected by the mass conservation law in Eq. (3), which allows us to derive the monomer concentration in the form  $N_1 = N_0 - \sum_{i=2}^{\infty} i N_i$ . The remaining parameters  $N_i$  can now be considered as independent variables.

In the equilibrium state the Gibbs free energy has a minimum, hence the set of partial derivatives by  $N_i$  from Eq. (10) must determine the equilibrium:

$$\frac{\partial G}{\partial N_2} = 0, \dots, \frac{\partial G}{\partial N_i} = 0, \dots \quad (11)$$

Applying further the Stirling approximation  $\ln N_i! \approx N_i \ln N_i - N_i$ , evaluation of Eq. (11) gives

$$\frac{\partial G}{\partial N_i} = (i-1)\Delta G_1 + kT \ln \frac{N_i}{(N_1)^i} = 0. \quad (12)$$

Solution of Eq. (12) results in the expression

$$N_i = N_1 (K N_1)^{i-1}, \quad (13)$$

which coincides with the law of mass action [see Eq. (2)].

This result is probably the most important consequence of the grand partition function (9) because it provides the physical justification for using the law of mass action when constructing the binding polynomial. However, the derivation given above highlights the tight link between the explicit form of the grand partition function  $Z$  and the law of mass action. If  $Z$  for a particular system is different from that written in Eq. (9), the validity of Eq. (13) for this system must always be verified according to the derivation given in Eqs. (10)–(13). Hence, one may conclude that great caution should be given to the construction of the binding polynomial by means of Eq. (13) for each particular system.

### C. Link between the grand partition function and the binding polynomial

Let us substitute Eq. (13) into Eq. (9) and evaluate the grand partition function using the Stirling approximation for the factorial and Eq. (4) for the link between  $\Delta G_1$  and  $K$ :

$$Z = \frac{N_0!}{N_1! N_2! \dots} \exp\left(-\sum_{i=1}^{\infty} (i-1)N_i \frac{\Delta G_1}{kT}\right) \approx \left[\frac{N_0}{N_1} \exp\left(-1 + \sum_{i=1}^{\infty} \frac{N_i}{N_0}\right)\right]^{N_0}. \quad (14)$$

Further simplification of Eq. (14) demands the introduction of an assumption of a relatively small contribution from all oligomers except the monomers, i.e.,

$$\sum_{i=1}^{\infty} \frac{N_i}{N_0} \approx 1 \Rightarrow K N_1 \approx 0, \quad (15)$$

which allows reduction of the exponent down to

$$Z \approx \left(\sum_{i=1}^{\infty} \frac{N_i}{N_1}\right)^{N_0} = [1 + K N_1 + (K N_1)^2 + \dots + (K N_1)^i + \dots]^{N_0} = (Z_b)^{N_0}. \quad (16)$$

Equation (16) provides an explicit link between the grand partition function  $Z$  and the binding polynomial  $Z_b$ , which is the answer to question (i) formulated in the Introduction.

In order to understand the physical meaning of  $Z_b$  we rewrite Eq. (16) in the form

$$Z \approx (Z_b)^{N_0} = \left[\frac{N_1 + N_1^2 e^{-\Delta G_1/kT} + N_1^3 e^{-2\Delta G_1/kT} + \dots + N_1^i e^{-(i-1)\Delta G_1/kT} + \dots}{N_1}\right]^{N_0}. \quad (17)$$

Analysis of Eq. (17) allows us to state that the sum in square brackets is the sum over all possible *energetic states* of a given molecule inside any type of oligomer that may be present in the system. The  $(N_1)^i$  quantity represents the overall number of ways needed to make up an oligomer from  $i$  molecules, which is known as the degeneracy of the given energy level. Hence, if the expressions in Eqs. (13) and (15) are valid,  $Z_b$  has the meaning of a molecular partition function that, when raised to the power of the total number of molecules  $N_0$ , becomes equal to the grand partition function  $Z$  of the self-assembling system. This knowledge is important for the correct application of the concept of the binding polynomial in a particular system and answers questions (ii) and (iii) formulated in the Introduction.

### D. Summary and discussion of basic assumptions behind the concept of a binding polynomial

Based on the derivation of the  $Z$  and  $Z_b$  quantities provided above, it is possible to formulate a complete set of assumptions behind the concept of the binding polynomial that are always inferred when applying  $Z_b$  to any particular system.

(a) The values for the pressure, temperature, and number of molecules are constant. This assumption is incorporated by writing the partition function in Eq. (9).

(b) There are negligible volume changes on the aggregate or complex formation.

(c) The system is in a state of equilibrium. This assumption enables us to operate with the subpartition function [Eq. (9)] associated with a unique distribution of oligomers  $N_1, \dots, N_i, \dots$  corresponding to the free-energy minimum [Eq. (11)] and results in the law of mass action [Eqs. (2) and (13)], which is the principal component in the concept of the binding polynomial.

(d) The self-assembly process is not pronounced, i.e.,  $K N_1 \approx 0$ . This assumption allows us to reduce  $Z$  down to  $Z_b$  and provides the latter with physical meaning.

(e) The magnitudes of the equilibrium constant  $K$  are equal for any stage of aggregation or complexation, i.e., independent of the number of molecules  $i$ . This assumption neglects any cooperativity in complexation and is incorporated depending on how the partition function (6) of an  $i$ -mer is written [i.e., the summation of the energies  $\Delta G_1$  over all  $(i-1)$  interfaces of the molecules in an aggregate]. Taking into account that, as shown above, the law of mass action is tightly linked to the way the grand partition function  $Z$  is defined, the application of the binding polynomial  $Z_b$  to a particular system with cooperativity must be treated with caution. The cooperative self- and heteroassemblies have attracted a great deal of attention in supramolecular chemistry (see, e.g., [10,18,19]);

this issue will not be discussed in the present paper. We must note, however, a specific partial case of cooperativity, reviewed in [20], which is fully compliant with the oligomer partition function (6) and hence with the binding polynomial  $Z_b$ . This is a fundamental anticooperativity, intrinsic to any complexation process, originating solely from the loss of degrees of freedom on complexation and resulting in the dependence of  $K$  on  $i$ . This dependence has been established for the self-assembly process [20] and ligand-polymer complexation [21].

(f) Only a single way of formation of any  $i$ -mer  $X_i$  is considered, viz., by sequential addition of a monomer  $X_1$  to an  $(i - 1)$ -mer  $X_{i-1}$ , commonly written in the form of an equilibrium reaction  $X_1 + X_{i-1} \xrightleftharpoons{K} X_i$  [1,3,18,19]. An alternative way, often called the random association  $X_i + X_j \xrightleftharpoons{K} X_{i+j}$  reviewed in [3], is ignored. Interestingly, of the six assumptions listed here, this one seems to be the most contradictory. At least two consequences of this assumption have been discussed in the literature.

(i) The mixing of molecules constituting any  $i$ -mer should additionally be taken into account [22], resulting in the so-called mixing entropy contribution to the assembly process. If so, the molecules in an aggregate become indistinguishable, i.e., they are allowed to freely interchange their positions within an aggregate in analogy to an ideal gas in a tube. Such a view requires correction of Eq. (6) by introducing terms accounting for the molecules' interchanging and is not compliant with molecular aggregate formation [15].

(ii) Although some experimental evidence of random association has been reported several decades ago [23], the overwhelming majority of published applications of the binding polynomial to aggregation or complexation processes use the sequential model of aggregation as an axiom. Recently, indirect [22] and direct [3] statistical-thermodynamical proof of the sequential mode of aggregation has been reported,

thereby supporting the appropriateness of using a binding polynomial in the form of Eq. (1).

### III. CONCLUSION

The use of the concept of a binding polynomial has become a routine method in various applications dealing with noncovalent molecular self-assembly. However, nearly all currently known applications of this method use it axiomatically without a reference to a true grand partition function of the system and to fundamental limitations of this approach.

In the present work we revisited the concept of binding polynomial for the most widely used case of self-assembly (or aggregation) of identical molecular units and reconstructed a link to the grand partition function of such a system. It was found that the common law of mass action (as a fundamental basis of any binding polynomial) has a direct link to the grand partition function. Hence, the validity of this law must always be checked before constructing the binding polynomial for each particular system. If the law of mass action is valid and if the self-assembly process is not pronounced (i.e.,  $KN_1 \approx 0$ ), the binding polynomial has the meaning of a molecular partition function that is given by the summation over energy levels of any molecule in the system (the energy levels are related to the type of oligomer where the molecule may sit). In such a case the use of the binding polynomial is justified. If the above-stated conditions are not valid, the concept of the binding polynomial may be misleading. This understanding is important in the correct application of the concept of a binding polynomial to particular systems.

### ACKNOWLEDGMENT

Dr. Ivan Laponogov (King's College, University of London) is thanked for helpful comments.

- 
- [1] T. F. A. De Greef, M. M. J. Smulders, M. Wolffs, A. P. H. J. Schenning, R. P. Sijbesma, and E. W. Meijer, *Chem. Rev.* **109**, 5687 (2009).
- [2] P. Thordarson, *Chem. Soc. Rev.* **40**, 1305 (2011).
- [3] M. P. Evstigneev, A. S. Buchelnikov, and V. P. Evstigneev, *Phys. Rev. E* **85**, 061405 (2012).
- [4] W. Bujalowski, *Chem. Rev.* **106**, 556 (2006).
- [5] E. Freire, A. Schön, and A. Velazquez-Campoy, *Methods Enzymol.* **455**, 127 (2009).
- [6] V. B. Teif, *Nucleic Acids Res.* **35**, e80 (2007).
- [7] Y. Kong, *J. Phys. Chem. B* **105**, 10111 (2001).
- [8] M. P. Evstigneev, V. P. Evstigneev, and D. B. Davies, *J. Chem. Phys.* **127**, 154511 (2007).
- [9] V. B. Teif and K. Rippe, *J. Phys.: Condens. Matter* **22**, 414105 (2010).
- [10] A. J. Markvoort, H. M. M. ten Eikelder, P. A. J. Hilbers, T. F. A. de Greef, and E. W. Meijer, *Nat. Commun.* **2**, 509 (2011).
- [11] V. P. Evstigneev, A. A. Mosunov, A. S. Buchelnikov, A. A. Hernandez Santiago, and M. P. Evstigneev, *J. Chem. Phys.* **134**, 194902 (2011).
- [12] M. Zdunek, J. Piosik, and J. Kapuscinski, *Biophys. Chem.* **84**, 77 (2000).
- [13] A. S. Buchelnikov, V. P. Evstigneev, and M. P. Evstigneev, *Chem. Phys.* **421**, 77 (2013).
- [14] S. Reith, Z. Li, C. E. Hinkle, C. X. Guzman, J. J. Lee, S. I. Nehme, and A. B. Braunschweig, *J. Phys. Chem. C* **117**, 11347 (2013).
- [15] M. P. Evstigneev and A. S. Buchelnikov, *Chem. Phys. Lett.* **567**, 48 (2013).
- [16] Y. Kong, *J. Chem. Phys.* **135**, 156101 (2011).
- [17] A. V. Tobolsky and P. J. Blatz, *J. Chem. Phys.* **13**, 379 (1945).
- [18] D. Zhao and J. S. Moore, *Org. Biomol. Chem.* **1**, 3471 (2003).
- [19] H. M. M. ten Eikelder, A. J. Markvoort, T. F. A. de Greef, and P. A. J. Hilbers, *J. Phys. Chem. B* **116**, 5291 (2012).
- [20] D. A. Beshnova, A. O. Lantushenko, D. B. Davies, and M. P. Evstigneev, *J. Chem. Phys.* **130**, 165105 (2009).
- [21] D. A. Beshnova, A. O. Lantushenko, and M. P. Evstigneev, *Biopolymers* **93**, 932 (2010).
- [22] I. Turcu and M. Bogdan, *J. Phys. Chem. B* **116**, 6488 (2012).
- [23] D. Thusius, *J. Mol. Biol.* **94**, 367 (1975).