JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

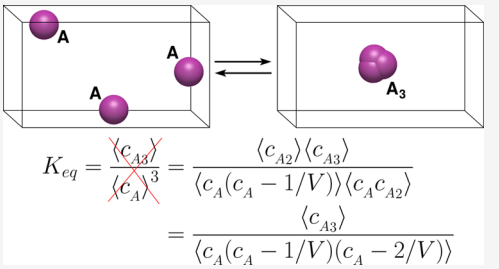
# Multimerizations, Aggregation, and Transfer Reactions of Small Numbers of Molecules

Ronen Zangi*

Cite This: J. Chem. Inf. Model. 2023, 63, 4383−4391 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** Chemical equilibria of multimerizations in systems with small numbers of particles exhibit a behavior seemingly at odds with that observed macroscopically. In this paper, we apply the recently proposed expression of equilibrium constant for binding, which includes cross-correlations in reactants' concentrations, to write an equilibrium constant for the formation of clusters larger than two (e.g., trimer, tetramer, and pentamer) as series of two-body reactions. Results obtained by molecular dynamics simulations demonstrate that the value of this expression is constant for all concentrations and system sizes, as well as at an onset of a phase transition to an aggregated state, where densities in the system change discontinuously. In contrast, the value of the commonly utilized expression of equilibrium constant, which ignores correlations, is not constant and its variations can reach few orders of magnitude. Considering different paths for the same multimer formation, with elementary reactions of any order, yields different expressions for the equilibrium constant, yet, with exactly the same value. This is also true for routes with essentially zero probability to occur. Existence of different expressions for the same equilibrium constant imposes equalities between averages of correlated, along with uncorrelated, concentrations of participating species. Moreover, a relation between an average particle number and relative fluctuations derived for two-body reactions is found to be obeyed here as well despite couplings to additional equilibrium reactions in the system. Analyses of transfer reactions, where association and dissociation events take place on both sides of the chemical equation, further indicate the necessity to include cross-correlations in the expression of the equilibrium constant. However, in this case, the magnitudes of discrepancies of the uncorrelated expression are smaller, likely because of partial cancellation of correlations, which exist on both the reactant and product sides.




## INTRODUCTION

One of the most powerful postulations in chemistry is the ability to assign a constant, albeit temperature-dependent, to a chemical reaction from which the system's composition at equilibrium can be determined by, for example, amounts of reactants put in. This equilibrium constant, $K$, is normally defined as the ratio between concentrations (activities) of products over reactants, each of which is raised to the power of its stoichiometric coefficient. That is, for the reaction

$$\alpha A + \beta B \rightleftharpoons \gamma C + \delta D \tag{1}$$

the equilibrium constant takes the form

$$K \equiv \frac{c_C^{\gamma} c_D^{\delta}}{c_A^{\alpha} c_B^{\beta}} (c^{\circ})^{\alpha+\beta-\gamma-\delta} \tag{2}$$

where each concentration of each component, $c_X$, is divided by a reference concentration $c^{\circ}$ to render the ratio dimensionless. Alternatively, $K$ is defined by

$$K \equiv e^{-\Delta G^{\circ}/RT} \tag{3}$$

where $\Delta G^{\circ}$ is the standard Gibbs free energy change when $\alpha$ moles of $A$ react with $\beta$ moles of $B$ to form $\gamma$ moles of $C$ and $\delta$ moles of $D$, given all components are at their standard conditions. These two definitions for $K$ are usually presented in textbooks together without giving importance which has precedence.[1,2] The reason is that if we assume the chemical potential of each component, relative to that at standard state, is proportional to the logarithm of its concentration

$$\mu_X - \mu_X^{\circ} = RT \ln \frac{c_X}{c^{\circ}} \tag{4}$$

and apply the condition of equilibrium $\Delta G = 0$, these two definitions of $K$ are equivalent. However, the relation in eq 4 is not always valid. Although it can be derived for a mixture of noninteracting ideal gases and observed for macroscopic systems, it is ill-defined for chemical reactions with small numbers of molecules. In these cases, the system is subject to substantial fluctuations in composition, where configurations with $c_X = 0$ are possible, and therefore, an ensemble average of $\mu_X$

Received: May 22, 2023
Published: July 11, 2023







https://doi.org/10.1021/acs.jcim.3c00774
J. Chem. Inf. Model. 2023, 63, 4383−4391

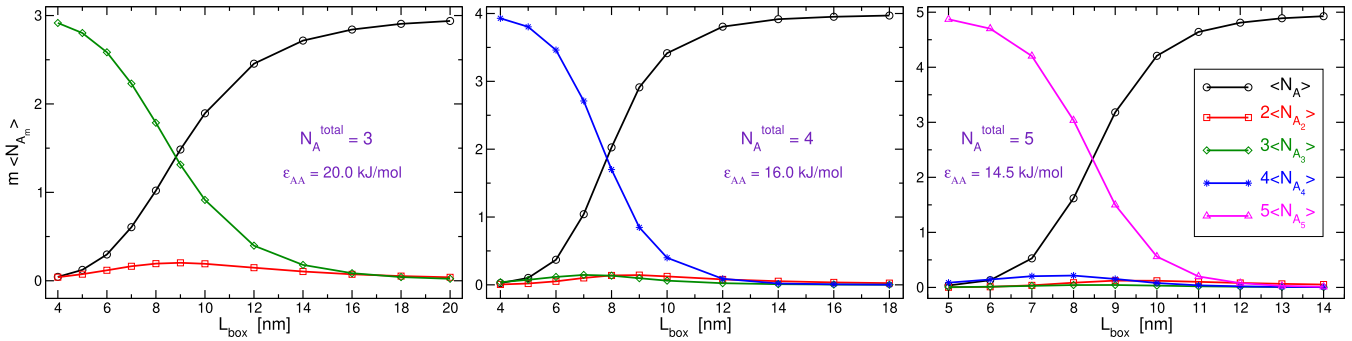



**Figure 1.** Average number of $m$-mers multiplied by $m$ as a function of length of the cubic simulation box for systems with $N_A^{\text{total}} = 3$, 4, and 5 of R1 series of simulations. The well-depth value of the Lennard-Jones potential acting between the particles, $\epsilon_{AA}$, is indicated for each system.

cannot be properly defined. To avoid these singularities, one can define $\langle \mu_X \rangle$ by eq 4 utilizing the ensemble average of the concentration, $\langle c_X \rangle$. Then, given the definition in eq 3, the expression of $K$ that follows is the same as in eq 2, where the concentration of each component is an ensemble average of that concentration. It turns out, the invalidity of eq 4 is not only at singular points of $c_X = 0$ but whenever the $X$ component in finite systems participates in a two-body (or higher order) reaction. As a consequence, the system's properties, such as concentrations, are not homogeneous functions, as observed in experiments[3−11] and simulations,[12−16] investigating binding processes with small numbers of molecules. Furthermore, for small systems, the conventional expression of $K$ (eq 2) is not constant upon changes in volumes and/or numbers of particles,[17−23] abolishing its importance in predicting the properties of chemical reactions. On the other hand, by definition, the value of $K$ in eq 3 is constant, however, taken on its own, this definition is of very limited use because it does not provide any information on reactions conducted at different concentrations or system sizes.

Recently, we utilized the definition of $K$ in eq 3 to derive an equivalent expression in terms of ensemble averages of concentrations, applicable also for reactions at different conditions. To this end, a simple, elementary, two-body reaction

$$A + B \rightleftharpoons AB \tag{5}$$

is considered. It is found that $K$ is not given by the expression widely applied in the literature,[24−33] $\langle c_{AB} \rangle c^{\circ} / (\langle c_A \rangle \langle c_B \rangle)$, but must include cross-correlations in reactants' concentrations, yielding the expression[34,35]

$$K = \frac{\langle c_{AB} \rangle}{\langle c_A (c_B - \delta_{AB}/V) \rangle} c^{\circ} \tag{6}$$

where $\delta_{AB}$ equals zero for hetero-dimerizations ($A$ and $B$ are distinct components) and equals one for homo-dimerizations ($A$ and $B$ are the same components, i.e., $A \equiv B$). The subtraction of the reciprocal of volume term, $1/V$, in homo-dimerizations, excludes self-correlations in particle number. For large systems, correlations between particles, as well as the term $1/V$, can be ignored; thus, eq 6 reduces to the commonly known expression. Using different arguments, Rubinovich and Polak[36] obtained the same expression of $K$ as that shown in eq 6.

Being derived for elementary reactions, the expression of $K$ for a multistep (complex) reaction, which can proceed through different mechanisms, depends on the path considered. This is because averages of correlated concentrations originating from different reaction steps cannot be canceled out in the expression of $K$ describing the total complex reaction. In this paper, we study multimerizations as models for such complex reactions. It is demonstrated that despite producing different expressions for the equilibrium constant, all mechanisms (paths) yield the same value of $K$. This is true for paths composed of several two-body reactions, as well as for routes composed of elementary reaction(s) with higher-order body correlations, which, in practice, are not probable to occur. Hence, in order for a system to be in equilibrium, several relations between averages of correlated, and also uncorrelated, concentrations of different components must hold.

Throughout the manuscript, we will refer to multimerization as a process in which like-particles form a cluster larger than two and are in equilibrium with smaller clusters or monomers present at appreciable concentrations. This is to be distinguished from aggregation, in which like-particles of different cluster sizes phase-transformed abruptly to a state of one large cluster that is by far the predominant component in the system. It is interesting to note the expressions of $K$ derived here produce values that stay constant also at an onset of a phase transition to an aggregated state, where concentrations of smaller-sized clusters change discontinuously.

## RESULTS AND DISCUSSION

Consider a system at constant temperature, $T$, and volume, $V$, in which $N_A^{\text{total}}$ particles of $A$ are able to bind with, and dissociate from, one another to form clusters of different $m$-mer sizes, $A_m$, where $1 \le m \le N_A^{\text{total}}$ (hereafter, monomers, $A_1$, will be denoted as $A$). The behavior of all components in the system is assumed ideal, that means, except for association and dissociation events, interparticle interactions do not exist or can be ignored. All potential reactions between the $A_m$ species are possible and, depending on the system investigated, quantities characterizing specific chemical equilibria can be of interest. Nevertheless, a quantity often sought for is the equilibrium constant, $K_m$, of the multimerization reaction

$$mA \rightleftharpoons A_m \tag{7}$$

where $m$ monomers of $A$ associate to form an $m$-mer, from which the corresponding free energy can be obtained. The reaction in eq 7 is not necessarily elementary, i.e., does not proceed via single step. More specifically, for $m \ge 3$, there are different chemical paths to realize the formation of the $A_m$ product, however, the most probable are likely those which require the lowest-order correlations between particles. For associations, this means sequences of two-body reactions. This is because concerted reactions of three or more particles are not likely to occur, especially in the ideal-gas approximation (low concentrations), due to the small probability to find all $m$ particles simultaneously next to one another.



https://doi.org/10.1021/acs.jcim.3c00774
J. Chem. Inf. Model. 2023, 63, 4383−4391

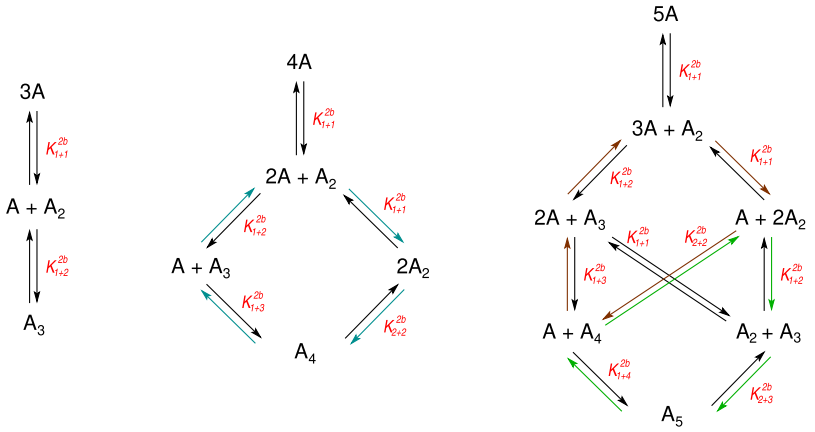



**Figure 2.** Chemical equilibria, assuming two-body-reactions, for the $N_A^{\text{total}} = 3$, 4, and 5 systems (left, middle, and right panels). The two-body binding constants are indicated in red and correspond to the net chemical equations.

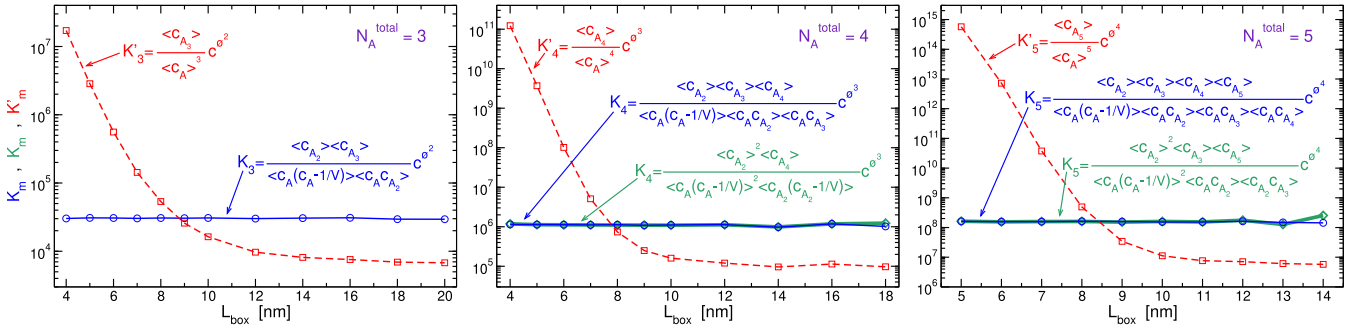



**Figure 3.** Equilibrium constant of $m$-mer formation from $m$ monomers, $K_m$, for $m = 3$ (trimer), 4 (tetramer), and 5 (pentamer) considering a sequence of bimolecular reactions for the three systems shown in Figure 1, respectively. Blue lines (circle symbols) are constructed by successive growths of the cluster by one monomer (eq 10). For $N_A^{\text{total}} = 4$ and 5, there are additional independent routes of two-body reactions; therefore, in green lines (diamond symbols), we present also, $K_4 = (K_{1+1}^{2b})^2 K_{2+2}^{2b}$ and $K_5 = (K_{1+1}^{2b})^2 K_{1+2}^{2b} K_{2+3}^{2b}$, which correspond to the paths on the right side in Figure 2. For comparison, the values of the corresponding uncorrelated expressions, $K'_m$ defined in eq 12, are shown by dashed red lines (square symbols).

In the first series of simulations, R1, we considered three systems, $N_A^{\text{total}} = 3$, 4, and 5, at various concentrations. The pair interaction energy between $A$ particles, in each of these three systems, was chosen to support an almost complete transformation from an $A_m$ state, where $m = N_A^{\text{total}}$, to an all-monomeric state upon augmenting $V$ within a desired range (see the Computational Details section for more information). This behavior of the systems is shown in Figure 1 by plotting the compositions at equilibrium as a function of length of the cubic simulation box, $L_{\text{box}}$.

**Two-Body Reaction Paths.** Chemical equilibria of all possible elementary two-body reactions are shown in Figure 2 for systems with $N_A^{\text{total}} = 3$, 4, and 5, which, among others, specify the different paths connecting reactants and products of the multimerization described in eq 7. For a two-body binding reaction between an $i$-mer and a $j$-mer to form a $(i + j)$-mer

$$A_i + A_j \rightleftharpoons A_{i+j} \tag{8}$$

we define the corresponding two-body binding constant $K_{i+j}^{2b}$, which equals

$$K_{i+j}^{2b} = \frac{\langle c_{A_{i+j}} \rangle}{\langle c_{A_i}(c_{A_j} - \delta_{ij}/V) \rangle} c^{\circ} \tag{9}$$

where $\delta_{ij}$ is the Kronecker delta (i.e., it equals 1 if $i = j$ and 0 otherwise). These bimolecular association constants are marked in Figure 2 and refer to the net specified equilibrium reactions. Because free energy changes are state functions, not all of the $K_{i+j}^{2b}$'s are independent. Thus, there is a smaller, irreducible, set of $K_{i+j}^{2b}$'s (whose size depends on $m$) from which the other $K_{i+j}^{2b}$'s can be derived from. It is also to be noted that the not drawn horizontal equilibria in Figure 2 for $N_A^{\text{total}} = 4$ and 5 are transfer (or exchange) reactions and will be discussed below.

The equilibrium constant of eq 7, $K_m$, obtained by sequences of bimolecular reactions sketched in Figure 2 are displayed in Figure 3 for the three systems. In all cases, the resulting $K_m$ is constant, independent of concentration, as it should be. For $N_A^{\text{total}} = 4$ and 5, there is more than one path to form a tetramer and a pentamer, respectively. For example, the left route in Figure 2, describing successive two-body additions of monomers, leads to the following equilibrium constant

$$K_m = \prod_{j=1}^{m-1} K_{1+j}^{2b} = \frac{\prod_{l=2}^{m} \langle c_{A_l} \rangle}{\prod_{l=1}^{m-1} \langle c_A(c_{A_l} - \delta_{1l}/V) \rangle} c^{\circ^{m-1}} \tag{10}$$

which is different than that obtained from the route on the right. Different reaction paths produce different expressions for $K_m$, some of which with no trivial dependency. However, because free energy changes are state functions, the different expressions for $K_m$ must have the same value. As exhibited in Figure 3, this is indeed the case, a property which imposes dependencies on the distribution of species in the system. For example, the concentrations of, and second-order correlations between, monomer, dimer, and trimer obey the relation



https://doi.org/10.1021/acs.jcim.3c00774
J. Chem. Inf. Model. 2023, 63, 4383−4391

$$\frac{\langle c_{A_2}\rangle}{\langle c_{A_3}\rangle} = \frac{\langle c_A(c_A - 1/V)\rangle\langle c_{A_2}(c_{A_2} - 1/V)\rangle}{\langle c_A c_{A_2}\rangle\langle c_A c_{A_3}\rangle} \tag{11}$$

which becomes trivial for large systems, where the terms $1/V$ and correlations can be ignored. In Figure 3, the commonly applied expression for equilibrium constant, $K'_m$, which ignores all correlations between particles

$$K'_m = \frac{\langle c_{A_m}\rangle}{\langle c_A\rangle^m} c^{\circ m-1} \tag{12}$$

is also shown. In contrast to the behavior of $K_m$ evaluated by any path, $K'_m$ is not constant as a function of concentration, especially in a regime of small volumes when clustering is substantial and the discrepancies reach several orders of magnitude. Even at large volumes, when it becomes constant, $K'_m$ does not approach $K_m$ because self-correlations are significant for systems with small values of $N_A^{\text{total}}$. For completeness, in Figure SI-1 of the Supporting Information (SI), we present the two-body binding constants of reactions with at least one monomer, $K^{2b}_{1+j}$. Also here, all equilibrium constants are constant for all values of $L_{\text{box}}$, whereas the corresponding expressions not accounting for two-body correlations between particles, $K'^{2b}_{1+j}$, are not. Furthermore, the discrepancy of $K'^{2b}_{1+j}$ from $K^{2b}_{1+j}$ at large volumes decreases with $N_A^{\text{total}}$ and increases with $j$.

The fact that not all $K^{2b}_{i+j}$'s are independent can also be illustrated by constructions of thermodynamic cycles, where the sum of free energy changes in a closed cycle should be equal to zero. The magnitude of deviation from zero is an estimate to error in the computation. When applied to our systems, as shown in Figure 4, the results indicate that the deviations are smaller than 0.2 kJ/mol except for the largest two $L_{\text{box}}$ values, where they reach 1.3 kJ/mol in closing one cycle of the $N_A^{\text{total}} = 5$ system. Obviously, obtaining converged results is more difficult for systems with a larger volume and number of particles.

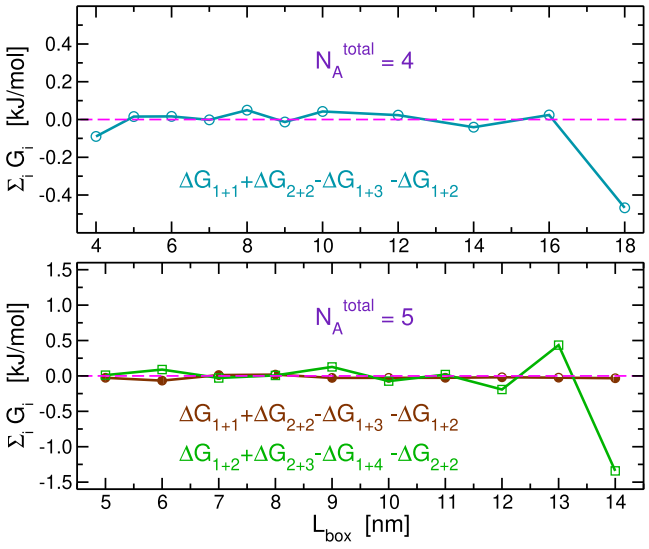



**Figure 4.** Thermodynamic cycle closures of the two-body-reaction model for $N_A^{\text{total}} = 4$ and 5 of R1 series of simulations. For the former, there is only one independent cycle involving two-body association constants, whereas, for the latter, there are two independent cycles. The cycles are depicted in Figure 2 with colors of the arrows matching the colors of the curves here. The free energy changes are defined by $\Delta G_{i+j} = -RT \ln K^{2b}_{i+j}$. The dashed magenta lines at $y = 0$ denote perfect closure.

**Transfer Reactions.** As mentioned above, the unmarked horizontal reactions in Figure 2 of the type

$$A_i + A_{j+1} \rightleftharpoons A_{i+1} + A_j \tag{13}$$

correspond to transfer reactions, where in both forward and backward directions, association and dissociation occur. Taking cross-correlations into account, the expression of the equilibrium constant is given by

$$K^{\text{trans}}_{i+(j+1)} = \frac{\langle c_{A_{i+1}}[c_{A_j} - \delta_{(i+1)j}/V]\rangle}{\langle c_{A_i}[c_{A_{j+1}} - \delta_{i(j+1)}/V]\rangle} \tag{14}$$

where the standard concentration, $c^{\circ}$, cancels out because the number of particles, thus phase space, is the same on both sides of the chemical equation. Alternatively, the same reaction can be described by subtracting a two-body binding reaction, $A_j + A \rightleftharpoons A_{j+1}$, from another two-body binding reaction, $A_i + A \rightleftharpoons A_{i+1}$, and consequently its equilibrium constant equals

$$K^{\text{trans}}_{i+(j+1)} = \frac{K^{2b}_{i+1}}{K^{2b}_{j+1}} = \frac{\langle c_{A_{i+1}}\rangle\langle c_{A_j}(c_A - \delta_{j1}/V)\rangle}{\langle c_{A_{j+1}}\rangle\langle c_{A_i}(c_A - \delta_{i1}/V)\rangle} \tag{15}$$

Clearly, the expressions in eqs 14 and 15 must be equal and independent of concentration. As shown in Figure 5, such is the

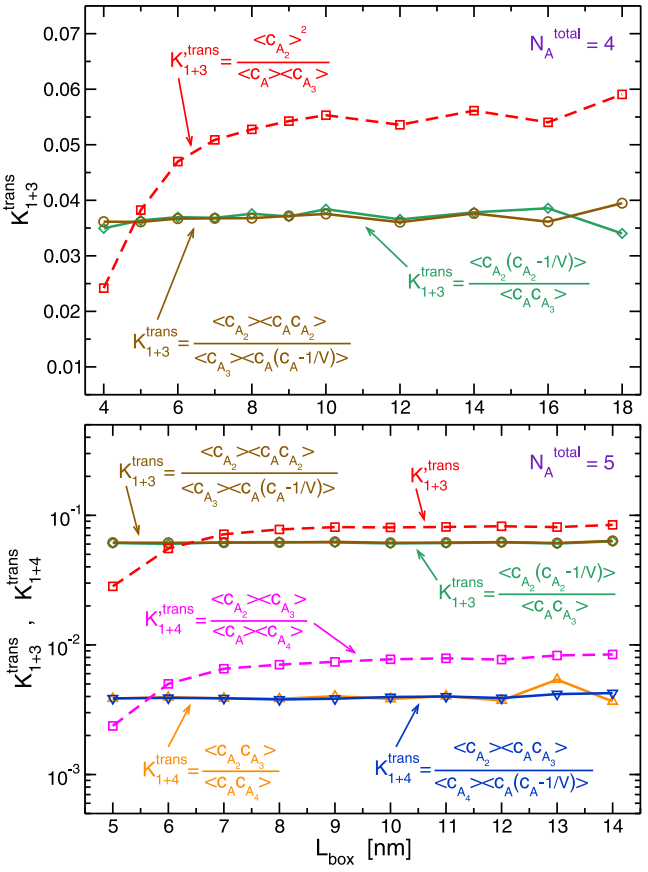



**Figure 5.** Equilibrium constants, $K^{\text{trans}}_{1+3}$ and $K^{\text{trans}}_{1+4}$, for the transfer reactions, $A + A_3 \rightleftharpoons 2A_2$ and $A + A_4 \rightleftharpoons A_2 + A_3$, respectively, of the $N_A^{\text{total}} = 4$ and 5 systems. Green (diamonds) and orange (triangles pointing up) curves were calculated by eq 14, whereas brown (circles) and blue (triangles pointing down) were calculated by eq 15. The corresponding expressions ignoring correlations, $K'^{\text{trans}}_{1+3}$ and $K'^{\text{trans}}_{1+4}$, are presented by dashed lines (squares) in red and magenta, respectively.



https://doi.org/10.1021/acs.jcim.3c00774
J. Chem. Inf. Model. 2023, 63, 4383−4391

case for the three unmarked transfer reactions of the equilibrium schemes in Figure 2. Again discrepancies, which are small in magnitude, appear only at large volumes. Furthermore, the corresponding uncorrelated expressions

$$K^{'\mathrm{trans}}_{i+(j+1)} = \frac{\langle c_{A_{i+1}}\rangle\langle c_{A_j}\rangle}{\langle c_{A_i}\rangle\langle c_{A_{j+1}}\rangle} \tag{16}$$

are not constant, especially at small values of $L_{box}$. It should be noted though, their deviations from the value of the equilibrium constant (either eqs 14 or 15) are much smaller, less than an order of magnitude, than for binding reactions (see Figures 3 and SI-1), which can reach several orders of magnitudes. It is likely that cross-correlations, present on both sides of transfer reactions (eq 13), partially cancel each other.

**Paths Involving Three-, or Higher-Order, Body Reactions.** We argued above that three-, or higher-order, body reactions are not likely to occur for reactants with unrestricted motions at low concentrations. Still, due to the state function character of a change in free energy, it should be possible to compute equilibrium constants by these implausible paths. Even if these reactions would practically never proceed via these mechanisms, the probabilities of observing the states on both sides of the chemical equation (each in equilibrium with other viable reaction mechanisms) do permit to calculate the equilibrium constant. Here, we show that these calculations yield results equal to those obtained by any other possible path. We start by writing the expression of $K_m$ for the multimerization reaction described in eq 7 for the case when all m monomers react simultaneously, that is, when eq 7 describes an elementary reaction. Because this $m$-body reaction is concerted, cross-correlations in reactant concentrations should include $(m-1)$ successive subtractions of self-correlations, and therefore, the $m$-body equilibrium constant takes the form

$$K^{mb}_{\underbrace{1+\cdots+1}_{m-\mathrm{terms}}} = \frac{\langle c_{A_m}\rangle}{\langle \prod_{i=0}^{m-1}(c_A - i/V)\rangle} c^{\varnothing^{m-1}} \tag{17}$$

where the summation label of $m$-terms of 1's in the subscript of $K$ indicates a simultaneous reaction between $m$ monomers. The same expression was obtained by Rubinovich and Polak,[36] however, without the claim, it corresponds to a concerted $m$-body reaction. Definitely, other $n$-order body reactions ($3 \le n < N_A^{\mathrm{total}} - 1$) involving clusters of larger size are also possible, where the expressions of the corresponding equilibrium constants can be easily inferred. For example, the equilibrium constant of a three-body reaction between two monomers and a dimer to form a tetramer is $K^{3b}_{1+1+2} = \langle c_{A_4}\rangle c^{\varnothing^2}/\langle c_A(c_A - 1/V)c_{A_2}\rangle$. We compute $K_m$ for $m$ = 3, 4, and 5 in the systems $N_A^{\mathrm{total}}$ = 3, 4, and 5, respectively, using all possible $n$-order body reactions and present the results in Figure 6. To eliminate ambiguity, we specify the mechanisms by outlining the elementary reactions involved and the total expressions of $K_m$ in Table SI-1. Figure 6 demonstrates that all possible mechanisms, including those with different orders of body reactions, of the multimerization of eq 7 yield the same value of $K_m$. Here, again, different expressions representing the same quantity of $K_m$ establish relations between averages of correlated, or uncorrelated, concentrations of different species in the system. For example, equating eq 10 with eq 17 results in

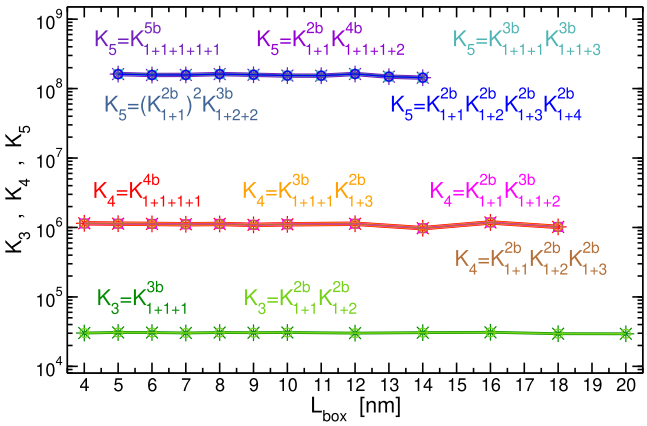



**Figure 6.** Equilibrium constant for $m$-mer formation, $K_m$, the same as that calculated in Figure 3 but here considering paths involving higher-than-two-body reactions. As before, $K_3$, $K_4$, and $K_5$ are calculated from systems with $N_A^{\mathrm{total}}$ = 3, 4, and 5, respectively. The results of different order body reactions are very similar, and as references, we also show those obtained by the two-body reaction route of gradual monomer additions, $K_m = \prod_{j=1}^{j=m-1} K^{2b}_{1+j}$ (eq 10), presented in Figure 3.

$$\left\langle \prod_{i=0}^{m-1}(c_A - i/V) \right\rangle = \frac{\prod_{i=1}^{m-1}\langle c_A(c_{A_i} - \delta_{1i}/V)\rangle}{\prod_{i=2}^{m-1}\langle c_{A_i}\rangle} \tag{18}$$

We note, even though all different paths (mechanisms) produce the same value for the equilibrium constant, they would not yield the same value for the corresponding reaction rates. Again, some (higher-order body reaction) mechanics are not likely to occur at all.

**Multimerizations in Systems with Different $N_A^{\mathrm{total}}$.** In the analyses above, we did not attempt to compare values of $K_m$ extracted from systems with different total numbers of monomers. The reason is because the well-depth values of the interaction energy, $\epsilon_{AA}$, are different (resulting from the aim to get continuous transformations between a predominantly monomeric state and a predominantly aggregated state within a certain range of $L_{box}$ as exhibited in Figure 1), and therefore, the values of $K_m$, for the same $m$, across different systems are different as well. To rectify this situation, we conducted a second series of simulations, R2, where the binding potential is the same for all systems, $\epsilon_{AA}$ = 10 kJ/mol, however, the total number of monomers was changed in the range of $2 \le N_A^{\mathrm{total}} \le 12$, keeping its total concentration constant (0.03245 $M$). In Figure 7a, we plot the value of $K_3$ and $K_4$ extracted from systems with different $N_A^{\mathrm{total}}$. $K_3$ is calculated by two mechanisms, one involving two-body and the other three-body reactions. $K_4$ is computed by five expressions, representing paths of two-, three-, and four-body reactions. The results indicate all different mechanisms, which lead to different expressions of $K_m$, give the same value, and are invariant of system size. The same conclusion is obtained by plotting the two-body equilibrium constants $K^{2b}_{1+j}$, defined by eq 9 for $i$ = 1, in Figure 7b. The significance of these results is that the equilibrium constant does not depend on whether the reaction in question is the only reaction taking place in the system or on whether the reactant and/or product engage with many other equilibrium reactions involving foreign compounds. This is obvious from the definition of the equilibrium constant in eq 3 and is illustrated, for example, by $K^{2b}_{1+1}$ ($\equiv K_2$) for $N_A^{\mathrm{total}}$ = 2, where dimerization of monomers and dissociation of dimers are the only two possible processes in the system, and for $N_A^{\mathrm{total}}$ = 12,



https://doi.org/10.1021/acs.jcim.3c00774
J. Chem. Inf. Model. 2023, 63, 4383−4391

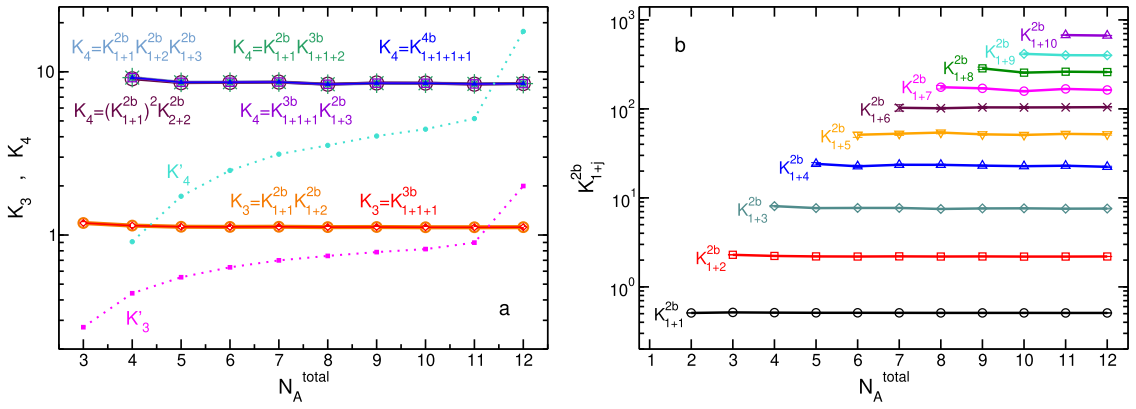



**Figure 7.** Results from R2 series of simulations (where the association energy between monomers, $\epsilon_{AA}$ = 10.0 kJ/mol, is the same for all systems). (a) Equilibrium constants for trimer ($K_3$) and tetramer ($K_4$) formation from the corresponding number of monomers as a function of $N_A^{total}$. The value of $K_3$ is calculated by a product of two two-body reactions ($K^{2b}_{1+1}K^{2b}_{1+2}$) as well as by a three-body reaction ($K^{3b}_{1+1+1}$). $K_4$ is calculated in five different ways: by a product of three two-body reactions [$K^{2b}_{1+1}K^{2b}_{1+2}K^{2b}_{1+3}$ and $(K^{2b}_{1+1})^2K^{2b}_{2+2}$], by a product of two-body and three-body reactions ($K^{2b}_{1+1}K^{3b}_{1+1+2}$ and $K^{3b}_{1+1+1}K^{2b}_{1+3}$), and by a four-body reaction ($K^{4b}_{1+1+1+1}$). See Table SI-1 of SI for explicit expressions of all equilibrium constants. We also present as dotted lines the value of the expressions ignoring correlations, $K'_3$ (magenta) and $K'_4$ (cyan), as defined in eq 12. (b) Equilibrium constants of two-body reactions involving a monomer, $K^{2b}_{1+j}$. In both plots, the magnitudes of the estimated errors are smaller or similar to the size of symbols.

where the reactant and product take part in 10 and 9, respectively, additional equilibrium reactions.

In Figure 7a, we also plot the values of $K'_3$ and $K'_4$, the expressions ignoring correlations in the systems as defined in eq 12. As expected, large deviations are observed for small $N_A^{total}$, which decrease with an increase in a total number of monomers. However, at approximately $N_A^{total}$ = 11, there is an abrupt increase in the values of $K'_3$ and $K'_4$. Analyses of the concentrations of monomers, trimers, and tetramers (Figure SI-2a,b) indicate a sharp decrease exactly around this value of $N_A^{total}$ for these three species, and as a result, the curves of $K'_3$ and $K'_4$ experience a discontinuous increase because the monomer concentration is raised to a power of 3 and 4 in the denominator. These sharp drops in concentrations are most likely due to an onset of a phase transition to an aggregated state. Indeed, there is an increase in the concentration of the largest possible cluster in the system (Figure SI-2c) at that point, however, apparently with a continuous character. It is worth emphasizing that this onset of phase transition (aggregation), although drastically affects $K'_3$ and $K'_4$, does not influence the values of $K_3$ and $K_4$, which stay constant also across this point.

**Relative Fluctuations in Particle Numbers.** Previously, we showed that for homo- and hetero-dimerizations, the average number of bound particles is related to relative fluctuations in the system.[34,35] Accordingly, for the two-body reaction described in eq 8, we can write the average number of a cluster of size $(i + j)$ as

$$\langle N_{A_{i+j}}\rangle = \frac{1}{l(N_{A_{i+j}}, N_{A_{i+j}}) - l[N_{A_{i+j}}, N_{A_i}(N_{A_j} - \delta_{ij})]} \tag{19}$$

where the relative fluctuations between two variables $\zeta$ and $\eta$ are defined by $l(\zeta, \eta) = \langle\zeta\eta\rangle/(\langle\zeta\rangle\langle\eta\rangle) - 1$. In Figure 8a, we test this equality on R2 series for systems with $N_A^{total} \geq 5$, 6, and 8 for dimer, trimer, and tetramer, respectively. This is because the relation is trivial for systems with a smaller number of particles (which includes the entire R1 series). The results obtained exhibit an almost perfect agreement.

For systems with $N_A^{total} \geq 4$, the sum $(i + j)$ can be decomposed into different sets of integers (not including zero). If $i + j = k + s$, then equating the expression of $\langle N_{A_{i+j}}\rangle$ with that of $\langle N_{A_{k+s}}\rangle$ (both using eq 19) yields the relative fluctuations $l[N_{A_{i+j}}, N_{A_i}(N_{A_j} - \delta_{ij})]$ equal the relative fluctuations $l[N_{A_{k+s}}, N_{A_k}(N_{A_s} - \delta_{ks})]$, or more explicitly

$$\frac{\langle N_{A_{i+j}} N_{A_i}(N_{A_j} - \delta_{ij})\rangle}{\langle N_{A_i}(N_{A_j} - \delta_{ij})\rangle} = \frac{\langle N_{A_{k+s}} N_{A_k}(N_{A_s} - \delta_{ks})\rangle}{\langle N_{A_k}(N_{A_s} - \delta_{ks})\rangle} \tag{20}$$

This relation is examined in Figure 8b for $N_A^{total}$ = 4 of R2 series of simulations. Although some discrepancies are noticeable, only for the smallest system, the agreement falls just outside the estimated error bars.

**Calculating $K_m$ from the Ratio of Probabilities of Occurrences.** Examining the expression of $K_m$ derived from the $m$-body reaction, eq 17, it is easy to show that for the private case of $N_A^{total} = m$, this expression can also be written as the ratio of probability to observe the system in the $m$-mer state ($A_m$) to probability to observe all $m$ $A$ particles as monomers. More explicitly, if $f^{A_m}$ and $f^{mA}$ are fractions of frames (or probabilities) in which all particles are clustered and in which all particles are monomers, respectively, $K_m$ takes the form

$$K_{m_{N_A^{total}=m}} = \frac{f^{A_m}}{m! f^{mA}} (Vc^{\circ})^{m-1} \tag{21}$$

which is equivalent to eq 17 and corresponds to expressions obtained by Ouldridge and co-workers for dimerization[37] and hexamer formation[38] for the private case discussed here.

Last, and likely least, we inspect again the densities of different species of R1 series of simulations presented in Figure 1. It is evident the predominant components at equilibrium are either monomer or largest-mer possible, and hence, it might be tempting to approximate the behavior of these systems by a two-state model. This means densities of other species (states) are ignored, and it is enough to determine the fraction of frames of only one state ($f^{A_m}$ or $f^{mA}$), where the fraction of frames of the second state is obtained by subtraction from one. Assessing this approximation (Figure SI-3) reveals very poor agreement with the exact results of eq 21. If we choose to determine $f^{A_m}$ ($f^{mA}$), it is only for a regime in which the state of $A_m$ ($mA$ monomers) is hardly observed that $K_m$ is adequately approximated.



https://doi.org/10.1021/acs.jcim.3c00774
J. Chem. Inf. Model. 2023, 63, 4383−4391

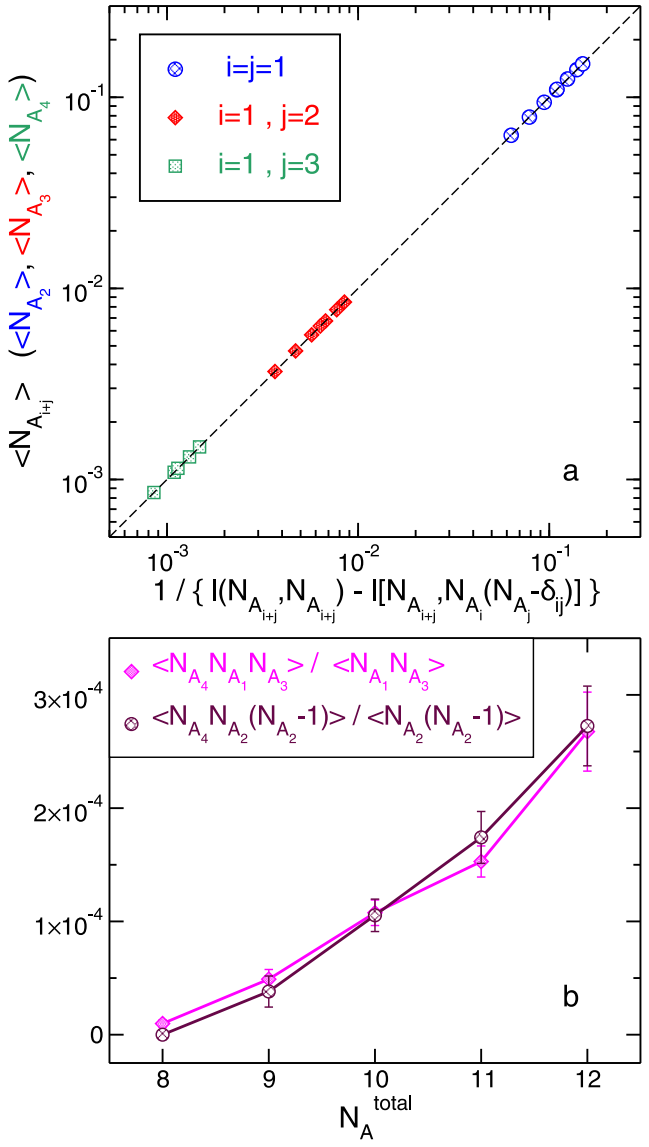



**Figure 8.** (a) Relation between the average number of dimer (blue), trimer (red), and tetramer (green) and relative fluctuations in the system as described by eq 19. The results are obtained from R2 series for systems with $N_A^{total} \geq 5$, 6, and 8 for dimer, trimer, and tetramer, respectively. The dashed black line, $y = x$, is plotted as a reference. (b) Validation of the equality described in eq 20 for the case $i + j = k + s = 4$ ($i = 1$, $j = 3$, and $k = s = 2$) for simulations with 8, or larger, number of total $A$ particles.

## CONCLUSIONS

For large systems, the expression of equilibrium constant for a complex reaction, which can proceed via several alternate mechanisms, is known to be independent of the mechanism considered. This is because cross-correlations between particle numbers are negligible, and therefore, average concentrations originating from different reaction steps can be canceled out in nominator and denominator of the total equilibrium constant, leaving only those corresponding to the net reaction. However, for finite systems, cross-correlations must be taken into account and cancellations of concentrations are, in general, not possible. This means, for each reaction mechanism (path), there is a unique expression of the equilibrium constant. At the same time, equilibrium constants are defined by free energy changes and as state functions, their values must be path independent. In this paper, we showed that different expressions of the equilibrium constant for a multimer formation, arising from different routes, all yield the same value. Thus, equalities between the different expressions characterize equilibrium, which translates to equalities between averages of concentrations (correlated or uncorrelated). Put another way, the condition for equilibrium includes additional restrictions on the distribution of particle numbers in the systems. At first glance, it might appear tantalizing that an expression of equilibrium constant corresponding to a path not likely to occur, such as a concerted approach of five monomers to form a pentamer, would give the same value as those derived from path(s) that do happen. However, this is indeed the case, and reaction routes which do not take place can also be utilized for calculating $K$ as long as there is at least one physically viable path connecting reactants with products.

Another characteristic of equilibrium constant demonstrated in this work is that its value is constant for all system sizes (down to the smallest system possible) and concentrations, as should be the case. In contrast, the commonly utilized expression applicable for the thermodynamic limit, which ignores correlations, produces values that vary significantly with changes in volume and the total number of particles. It is interesting to point out that for transfer reactions, the error of ignoring correlations is much smaller than that for corresponding association and dissociation reactions. This can be explained by partial cancellation of correlations, which appear on both sides of the chemical equation of transfer reactions.

For hetero- and homo-dimerizations, magnitudes of relative fluctuations are related to average concentrations. This equality is found to be valid also here, thus for larger-sized cluster formations, where reactants and products are coupled to other equilibrium reactions. When considering clusters of sizes equal to or larger than four, the different ways to partition the integer representing the cluster size into sum of smaller integers establish relations between correlated averages of different particle numbers in the system (eq 20).

## COMPUTATIONAL DETAILS

The model system consists of $N_A^{total}$ single-site $A$ particles interacting solely via Lennard-Jones (LJ) potential with a diameter $\sigma_{AA}$ = 0.20 nm and a well-depth $\epsilon_{AA}$. Newton's equations of motion were applied to propagate the system in time in such a way the resulting trajectory generated a canonical ensemble ($N_A^{total}$, $V$, $T$) at a temperature $T$ = 300 K and a volume $V = L_{box}^3$, with $L_{box}$ the length of the cubic simulation box.

Two series of simulations were designed. In the first, R1, we studied systems with $N_A^{total}$ = 3, 4, and 5, and modified systematically $L_{box}$. For each $N_A^{total}$, the value of $\epsilon_{AA}$ was chosen such that within the range of $L_{box}$ considered, an almost complete transformation occurred from a monomeric state to an aggregated state (see Figure 1). The values of $\epsilon_{AA}$, together with the range of $L_{box}$ investigated, are presented in Table 1. In the second series of simulations, R2, the well-depth of the interactions between the particles was constant, $\epsilon_{AA}$ = 10 kJ/mol, and we changed the total number of $A$ particle in the system in the range $2 \leq N_A^{total} \leq 12$ and concomitantly $L_{box}$, such that the concentration $N_A^{total}/V$ equals 0.03245 $M$ (i.e., 0.01954 molecules/nm$^3$). This resulted in a box length of 4.678 and 8.500 nm for the smallest and largest systems, respectively.

Molecular dynamics (MD) simulations were conducted by the software package GROMACS version 4.6.5[39] (single-



https://doi.org/10.1021/acs.jcim.3c00774
J. Chem. Inf. Model. 2023, 63, 4383−4391

Table 1. Well-Depth of Lennard-Jones Potential and Range of Lengths of Cubic Simulation Boxes for the Three Systems of R1

| $N_A^{total}$ | $\epsilon_{AA}$ [kJ/mol] | $L_{box}$ [nm] |
|---|---|---|
| 3 | 20.0 | 4.0−20.0 |
| 4 | 16.0 | 4.0−18.0 |
| 5 | 14.5 | 5.0−14.0 |

precision). A time step of 0.002 ps was employed for integrating the equations of motion and a mass of 10.0 amu was assigned to each *A* particle. The temperature was maintained by applying a Nosé−Hoover thermostat[40,41] with a chain length[42] of 2 and a coupling strength set to 0.1. The equations of motion were propagated by the velocity Verlet algorithm, in which the kinetic energy is determined by the average of the two half-steps. Periodic boundary conditions were applied along all three Cartesian axes, and at every step, motion of center of mass of the system was removed. The LJ potential was evaluated up to a cutoff distance of 2.0 nm. Based on the location of the first minimum of the radial distribution function, two *A* particles are defined as clustered (bonded) for $r_{AA} < 0.35$ nm.

For all simulations, equilibration for approximately 5 $\mu s$ was conducted prior to data collection. In R1 series for $N_A^{total} = 3$ and 4, at each value of $L_{box}$, data were collected for 400 $\mu s$. However, for $N_A^{total} = 4$ with $L_{box} \geq 14$ and for $N_A^{total} = 5$, the data collection time was doubled, thus for 800 $\mu s$. In R2 series, the data collection step for each $N_A^{total}$ was run for 1600 $\mu s$.

## ■ ASSOCIATED CONTENT

### Data Availability Statement

Input (parameter, configuration, and topology) files of R1 and R2 series of simulations necessary to reproduce all data presented in this paper are available at https://zenodo.org/record/7956480#.ZGs3vxlBzmF. Gromacs version 4.6.5, used for all simulations in this work, is available to the public via the link https://manual.gromacs.org.

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.3c00774.

(1) Figure of the set of $K_{1+j}^{2b}$, (2) table specifying different expressions to calculate $K_m$, (3) figure of average concentrations in R2 series of simulations, and (4) figure of $K_m$ approximated by a two-state model (PDF)

## ■ AUTHOR INFORMATION


### Corresponding Author

**Ronen Zangi** − *Donostia International Physics Center (DIPC), 20018 Donostia-San Sebastián, Spain; Department of Organic Chemistry I, University of the Basque Country UPV/EHU, 20018 Donostia-San Sebastián, Spain; IKERBASQUE, Basque Foundation for Science, 48009 Bilbao, Spain;* orcid.org/0000-0001-5332-885X; Email: r.zangi@ikerbasque.org

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.3c00774


### Notes

The author declares no competing financial interest.

## ■ ACKNOWLEDGMENTS

Technical and human support of the computer cluster provided by IZO-SGI SGIker of UPV/EHU and the European fundings, ERDF and ESF, are greatly acknowledged.

## ■ REFERENCES


(1) Daniels, F.; Alberty, R. A. *Physical Chemistry*, 4th ed.; John Wiley & Sons, Inc.: New York, NY, 1975.

(2) Castellan, G. W. *Physical Chemistry*, 3rd ed.; Addison-Wesley: Reading, MA, 1983.

(3) Morimatsu, M.; Takagi, H.; Kosuke, G.; Ota, R. I.; Yanagida, T.; Sako, Y. Multiple-state reactions between the epidermal growth factor receptor and Grb2 as observed by using single-molecule analysis. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 18013−18018.

(4) Polarz, S.; Kuschel, A. Chemistry in Confining Reaction Fields with Special Emphasis on Nanoporous Materials. *Chem. - Eur. J.* **2008**, *14*, 9816−9829.

(5) Sawada, T.; Yoshizawa, M.; Sato, S.; Fujita, M. Minimal nucleotide duplex formation in water through enclathration in self-assembled hosts. *Nat. Chem.* **2009**, *1*, 53−56.

(6) Shon, M. J.; Cohen, A. E. Mass Action at the Single-Molecule Level. *J. Am. Chem. Soc.* **2012**, *134*, 14618−14623.

(7) Patra, S.; Naik, A. N.; Pandey, A. K.; Sen, D.; Mazumder, S.; Goswami, A. Silver nanoparticles stabilized in porous polymer support: A highly active catalytic nanoreactor. *Appl. Catal., A* **2016**, *524*, 214−222.

(8) Galvin, C. J.; Shirai, K.; Rahmani, A.; Masaya, K.; Shen, A. Q. Total Capture, Convection-Limited Nanofluidic Immunoassays Exhibiting Nanoconfinement Effects. *Anal. Chem.* **2018**, *90*, 3211−3219.

(9) Megarity, C. F.; Siritanaratkul, B. R. S.; Heath, L. W.; Wan, L.; Morello, G.; Morello, G.; FitzPatrick, S. R.; FitzPatrick, S. R.; Booth, R. L.; Booth, R. L.; Sills, A. J.; Sills, A. J.; Robertson, A. W.; Robertson, A. W.; Warner, J. H.; Warner, J. H.; Turner, N. J.; Turner, N. J.; Armstrong, F. A. Electrocatalytic Volleyball: Rapid Nanoconfined Nicotinamide Cycling for Organic Synthesis in Electrode Pores. *Angew. Chem., Int. Ed.* **2019**, *58*, 4948−4952.

(10) Downs, A. M.; McCallum, C.; Pennathur, S. Confinement effects on DNA hybridization in electrokinetic micro- and nanofluidic systems. *Electrophoresis* **2019**, *40*, 792−798.

(11) Jonchhe, S.; Pandey, S.; Beneze, C.; Emura, T.; Sugiyama, H.; Endo, M.; Mao, H. Dissection of nanoconfinement and proximity effects on the binding events in DNA origami nanocavity. *Nucleic Acids Res.* **2022**, *50*, 697−703.

(12) Schnell, S. K.; Liu, X.; Simon, J.-M.; Bardow, A.; Bedeaux, D.; Vlugt, T. J. H.; Kjelstrup, S. Calculating Thermodynamic Properties from Fluctuations at Small Scales. *J. Phys. Chem. B* **2011**, *115*, 10911−10918.

(13) De Jong, D. H.; Schäfer, L. V.; De Vries, A. H.; Marrink, S. J.; Berendsen, H. J. C.; Grubmüller, H. Determining Equilibrium Constants for Dimerization Reactions from Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, *32*, 1919−1928.

(14) Cortes-Huerto, R.; Kremer, K.; Potestio, R. Communication: Kirkwood-Buff integrals in the thermodynamic limit from small-sized molecular dynamics simulations. *J. Chem. Phys.* **2016**, *145*, No. 141103.

(15) Dawass, N.; Krüger, P.; Schnell, S. K.; Bedeaux, D.; Kjelstrup, S.; Simon, J. M.; Vlugt, T. J. H. Finite-size effects of Kirkwood-Buff integrals from molecular simulations. *Mol. Simul.* **2018**, *44*, 599−612.

(16) Bråten, V.; Wilhelmsen, O.; Schnell, S. K. Chemical Potential Differences in the Macroscopic Limit from Fluctuations in Small Systems. *J. Chem. Inf. Model.* **2021**, *61*, 840−855.

(17) Ghosh, K. Stochastic dynamics of complexation reaction in the limit of small numbers. *J. Chem. Phys.* **2011**, *134*, No. 195101.

(18) Szymanski, R.; Sosnowski, S.; Maślanka, L. Statistical effects related to low numbers of reacting molecules analyzed for a reversible association reaction A + B = C in ideally dispersed systems: An apparent violation of the law of mass action. *J. Chem. Phys.* **2016**, *144*, No. 124112.

(19) Patel, L. A.; Kindt, J. T. Cluster Free Energies from Simple Simulations of Small Numbers of Aggregants: Nucleation of Liquid MTBE from Vapor and Aqueous Phases. *J. Chem. Theory Comput.* **2017**, *13*, 1023−1033.
(20) Zhang, X.; Patel, L. A.; Beckwith, O.; Schneider, R.; Weeden, C. J.; Kindt, J. T. Extracting Aggregation Free Energies of Mixed Clusters from Simulations of Small Systems: Application to Ionic Surfactant Micelles. *J. Chem. Theory Comput.* **2017**, *13*, 5195−5206.
(21) Szymanski, R.; Sosnowski, S. Stochasticity of the transfer of reactant molecules between nano-reactors affecting the reversible association A + B ⇄ C. *J. Chem. Phys.* **2019**, *151*, No. 174113.
(22) Goch, W.; Bal, W. Stochastic or Not? Method To Predict and Quantify the Stochastic Effects on the Association Reaction Equilibria in Nanoscopic Systems. *J. Phys. Chem. A* **2020**, *124*, 1421−1428.
(23) Jost Lopez, A.; Quoika, P. K.; Linke, M.; Hummer, G.; Köfinger, J. Quantifying Protein-Protein Interactions in Molecular Simulations. *J. Phys. Chem. B* **2020**, *124*, 4673−4685.
(24) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. *Biophys. J.* **1997**, *72*, 1047−1069.
(25) Hünenberger, P. H.; Granwehr, J. K.; Aebischer, J.-N.; Ghoneim, N.; Haselbach, E.; van Gunsteren, W. F. Experimental and Theoretical Approach to Hydrogen-Bonded Diastereomeric Interactions in a Model Complex. *J. Am. Chem. Soc.* **1997**, *119*, 7533−7544.
(26) Luo, H.; Sharp, K. On the calculation of absolute macromolecular binding free energies. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 10399−10404.
(27) Zhang, Y.; McCammon, J. A. Studying the affinity and kinetics of molecular association with molecular-dynamics simulation. *J. Chem. Phys.* **2003**, *118*, 1821−1827.
(28) Psachoulia, E.; Fowler, P. W.; Bond, P. J.; Sansom, M. S. P. Helix-Helix Interactions in Membrane Proteins: Coarse-Grained Simulations of Glycophorin A Helix Dimerization. *Biochemistry* **2008**, *47*, 10503−10512.
(29) Deng, Y.; Roux, B. Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. *J. Phys. Chem. B* **2009**, *113*, 2234−2246.
(30) Skorpa, R.; Simon, J.-M.; Bedeaux, D.; Kjelstrup, S. Equilibrium properties of the reaction $H_2$ ⇌ 2H by classical molecular dynamics simulations. *Phys. Chem. Chem. Phys.* **2014**, *16*, 1227−1237.
(31) Montalvo-Acosta, J. J.; Cecchini, M. Computational Approaches to the Chemical Equilibrium Constant in Protein-ligand Binding. *Mol. Inf.* **2016**, *35*, 555−567.
(32) Del Piccolo, N.; Hristova, K. Quantifying the Interaction between EGFR Dimers and Grb2 in Live Cells. *Biophys. J.* **2017**, *113*, 1353−1364.
(33) Duboué-Dijon, E.; Hénin, J. Building intuition for binding free energy calculations: Bound state definition, restraints, and symmetry. *J. Chem. Phys.* **2021**, *154*, No. 204101.
(34) Zangi, R. Binding Reactions at Finite Systems. *Phys. Chem. Chem. Phys.* **2022**, *24*, 9921−9929.
(35) Zangi, R. Statistical Mechanics of Dimerizations and its Consequences for Small Systems. *Phys. Chem. Chem. Phys.* **2022**, *24*, 28804−28813.
(36) Rubinovich, L.; Polak, M. Unraveling the Distinct Relationship between the Extent of a Nanoconfined Reaction and the Equilibrium Constant. *J. Phys. Chem. C* **2021**, *125*, 452−457.
(37) Ouldridge, T. E.; Louis, A. A.; Doye, J. P. K. Extracting bulk properties of self-assembling systems from small simulations. *J. Phys.: Condens. Matter* **2010**, *22*, No. 104102.
(38) Ouldridge, T. E. Inferring bulk self-assembly properties from simulations of small systems with multiple constituent species and small systems in the grand canonical ensemble. *J. Chem. Phys.* **2012**, *137*, No. 144105.
(39) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(40) Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **1984**, *81*, 511−519.
(41) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695−1697.
(42) Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nosé-Hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.* **1992**, *97*, 2635−2643.


https://doi.org/10.1021/acs.jcim.3c00774
*J. Chem. Inf. Model.* 2023, 63, 4383−4391

# Supporting Information: Multimerizations, Aggregation, and Transfer Reactions of Small Numbers of Molecules

Ronen Zangi*[1,2,3]

[1] *Donostia International Physics Center (DIPC), 20018 Donostia-San Sebastián, Spain*

[2] *Department of Organic Chemistry I, University of the Basque Country UPV/EHU, 20018 Donostia-San Sebastián, Spain*

[3] *IKERBASQUE, Basque Foundation for Science, 48009 Bilbao, Spain*

June 22, 2023

*r.zangi@ikerbasque.org

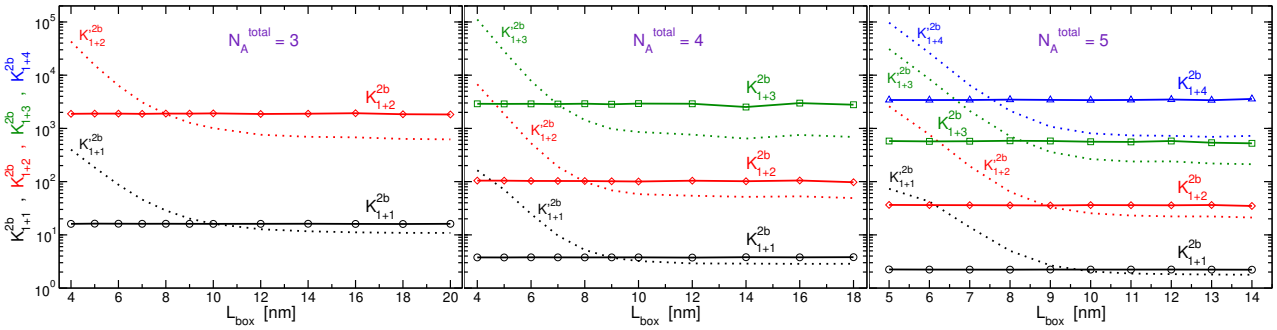



Figure SI-1: The set of equilibrium constants for two-body elementary reactions involving at least one monomer, $K^{2b}_{1+j}$, defined by Eq. 9 with $i = 1$, for the three systems of R1 series of simulations (solid lines with symbols). The values of the corresponding uncorrelated expressions, $K'^{2b}_{1+j} = \langle c_{A_{1+j}} \rangle c^{\varnothing} / (\langle c_A \rangle \langle c_{A_j} \rangle)$, are shown as well (dotted lines).

Table SI-1: Different expressions to calculate $K_m$ (for $m = 3$, 4, and 5) depending on the elementary reaction(s) considered. As shown in Fig. 6, all expressions yield the same value.

| $K_m$ | highest order correlations | elementary reaction(s) | total expression of $K_m$ |
|---|---|---|---|
| $K_3$ | two-body | $A + A \rightleftharpoons A_2$<br>$A + A_2 \rightleftharpoons A_3$ | $K^{2b}_{1+1}K^{2b}_{1+2} = \frac{\langle c_{A2}\rangle\langle c_{A3}\rangle\ c^{\varnothing 2}}{\langle c_A(c_A - 1/V)\rangle\langle c_A c_{A2}\rangle}$ |
| | three-body | $A + A + A \rightleftharpoons A_3$ | $K^{3b}_{1+1+1} = \frac{\langle c_{A3}\rangle\ c^{\varnothing 2}}{\langle c_A(c_A - 1/V)(c_A - 2/V)\rangle}$ |
| $K_4$ | two-body | $A + A \rightleftharpoons A_2$<br>$A + A_2 \rightleftharpoons A_3$<br>$A + A_3 \rightleftharpoons A_4$ | $K^{2b}_{1+1}K^{2b}_{1+2}K^{2b}_{1+3} = \frac{\langle c_{A2}\rangle\langle c_{A3}\rangle\langle c_{A4}\rangle\ c^{\varnothing 3}}{\langle c_A(c_A - 1/V)\rangle\langle c_A c_{A2}\rangle\langle c_A c_{A3}\rangle}$ |
| | three-body | $A + A + A \rightleftharpoons A_3$<br>$A + A_3 \rightleftharpoons A_4$ | $K^{3b}_{1+1+1}K^{2b}_{1+3} = \frac{\langle c_{A3}\rangle\langle c_{A4}\rangle\ c^{\varnothing 3}}{\langle c_A(c_A - 1/V)(c_A - 2/V)\rangle\langle c_A c_{A3}\rangle}$ |
| | three-body | $A + A \rightleftharpoons A_2$<br>$A + A + A_2 \rightleftharpoons A_4$ | $K^{2b}_{1+1}K^{3b}_{1+1+2} = \frac{\langle c_{A2}\rangle\langle c_{A4}\rangle\ c^{\varnothing 3}}{\langle c_A(c_A - 1/V)\rangle\langle c_A(c_A - 1/V)c_{A2}\rangle}$ |
| | four-body | $A + A + A + A \rightleftharpoons A_4$ | $K^{4b}_{1+1+1+1} = \frac{\langle c_{A4}\rangle\ c^{\varnothing 3}}{\langle c_A(c_A - 1/V)(c_A - 2/V)(c_A - 3/V)\rangle}$ |
| $K_5$ | two-body | $A + A \rightleftharpoons A_2$<br>$A + A_2 \rightleftharpoons A_3$<br>$A + A_3 \rightleftharpoons A_4$<br>$A + A_4 \rightleftharpoons A_5$ | $K^{2b}_{1+1}K^{2b}_{1+2}K^{2b}_{1+3}K^{2b}_{1+4} = \frac{\langle c_{A2}\rangle\langle c_{A3}\rangle\langle c_{A4}\rangle\langle c_{A5}\rangle\ c^{\varnothing 4}}{\langle c_A(c_A - 1/V)\rangle\langle c_A c_{A2}\rangle\langle c_A c_{A3}\rangle\langle c_A c_{A4}\rangle}$ |
| | three-body | $A + A + A \rightleftharpoons A_3$<br>$A + A + A_3 \rightleftharpoons A_5$ | $K^{3b}_{1+1+1}K^{3b}_{1+1+3} = \frac{\langle c_{A3}\rangle\langle c_{A5}\rangle\ c^{\varnothing 4}}{\langle c_A(c_A - 1/V)(c_A - 2/V)\rangle\langle c_A(c_A - 1/V)c_{A3}\rangle}$ |
| | three-body | $A + A \rightleftharpoons A_2$ x 2<br>$A + A_2 + A_2 \rightleftharpoons A_5$ | $(K^{2b}_{1+1})^2K^{3b}_{1+2+2} = \frac{\langle c_{A2}\rangle^2\langle c_{A5}\rangle\ c^{\varnothing 4}}{\langle c_A(c_A - 1/V)\rangle^2\langle c_A c_{A2}(c_{A2} - 1/V)\rangle}$ |
| | four-body | $A + A \rightleftharpoons A_2$<br>$A + A + A + A_2 \rightleftharpoons A_5$ | $K^{2b}_{1+1}K^{4b}_{1+1+1+2} = \frac{\langle c_{A2}\rangle\langle c_{A5}\rangle\ c^{\varnothing 4}}{\langle c_A(c_A - 1/V)\rangle\langle c_A(c_A - 1/V)(c_A - 2/V)c_{A2}\rangle}$ |
| | five-body | $A+A+A+A+A \rightleftharpoons A_5$ | $K^{5b}_{1+1+1+1+1} = \frac{\langle c_{A5}\rangle\ c^{\varnothing 4}}{\langle c_A(c_A - 1/V)(c_A - 2/V)(c_A - 3/V)(c_A - 4/V)\rangle}$ |

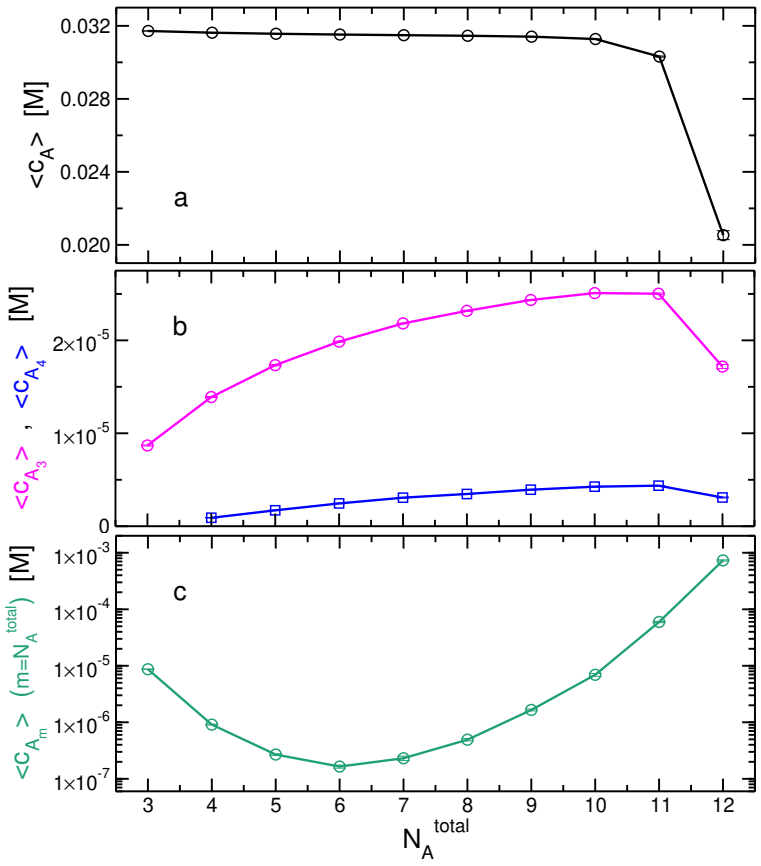



Figure SI-2: Average concentrations of several $m$-mers as a function of $N_A^{\text{total}}$ for R2 series of simulations. (a) For monomer, (b) for trimer and tetramer, (c) for the largest multi-mer (cluster) possible, thus for $A_m$ with $m = N_A^{\text{total}}$. Note that for all simulations in this series, the concentration of $N_A^{\text{total}}/V$ is constant at $0.03245\ M$.

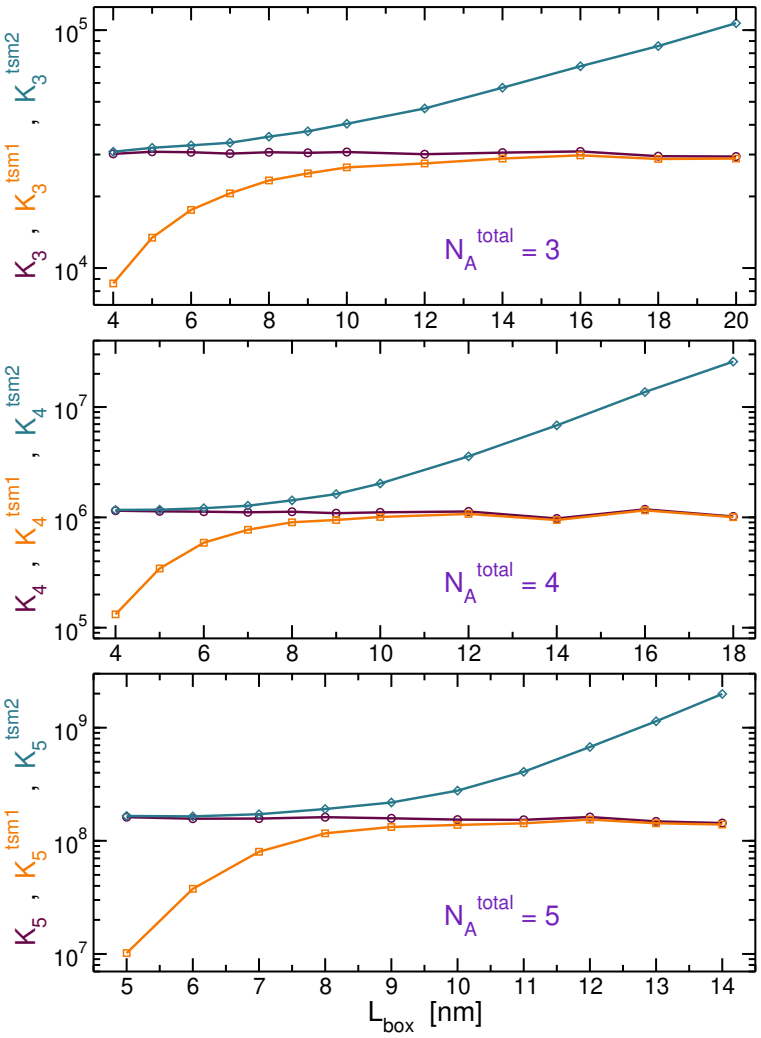



Figure SI-3: Approximations of $K_m$ for a private case of $m = N_A^{\text{total}}$ assuming a two-state model. The curves in orange (square symbols) are calculated by $K_m^{tsm1} = f^{A_m}(Vc^{\varnothing})^{m-1}/[m!(1-f^{A_m})]$, whereas the curves in blue (diamonds) are computed by $K_m^{tsm2} = (1-f^{mA})(Vc^{\varnothing})^{m-1}/[m!f^{mA}]$, where $f^{A_m}$ and $f^{mA}$ are fractions of frames, or probabilities, of observing the system in $A_m$ and $mA$ (i.e., all $m$ $A$ particles are monomers) states, respectively. As references for exact results, we plot in maroon (circles) $K_m = f^{A_m}(Vc^{\varnothing})^{m-1}/[m!f^{mA}]$ (Eq. 21).