Statistical Mechanics of Dimerizations and its Consequences for Small Systems

Ronen Zangi*[1,2]

[1] *POLYMAT & Department of Organic Chemistry I, University of the Basque Country UPV/EHU, Avenida de Tolosa 72, 20018, Donostia-San Sebastián, Spain*

[2] *IKERBASQUE, Basque Foundation for Science, Plaza Euskadi 5, 48009 Bilbao, Spain*

September 26, 2022

*r.zangi@ikerbasque.org

Abstract

Utilizing a statistical mechanics framework, we derive the expression of the equilibrium constant for dimerization reactions. An important feature arising from the derivation is the necessity to include two-body correlations between monomer's particles, reminiscent to those recently found crucial for binding reactions. However in (homo-) dimerizations, particles of the same type associate, and therefore, self-correlations are excluded. As a result, the mathematical form of the equilibrium constant differs from the well-known expression given in textbooks. For systems with large number of particles the discrepancy is negligible, whereas, for finite systems it is significant. Rationalized by collision probability between monomers, the bimolecular rate for dimer formation is proportional to concentration the same way correlations are accounted for. That is average of squared, and not square of averaged, monomer concentration should be considered in such a way that inconceivable collisions between a tagged particle with itself are excluded. Another consequence emerging from these two-body correlations, is an inhomogeneous function behavior of system's properties when scaling-down the system to a regime smaller than the thermodynamic limit. Thus, averages of properties observed at small systems are different than those observed at macroscopic systems. When applied to the size-dependent composition of the system, we further demonstrate the equilibrium concentration of the dimer (or monomer) can be obtained from only the magnitudes of fluctuations in the system. All predictions are verified by Monte Carlo and molecular dynamics simulations.

Introduction

Many, if not all, physical laws formulated for chemical reactions are deduced from macroscopic observations. A named example is the relation between the rate of an elementary process and the concentrations of the participating reactants[1]. As an underlying principle in chemical kinetics, it laid the foundation of another paramount example, the discovery of the law of mass action[2], wherein the equality between the forward and backward rates at equilibrium was demonstrated. In its turn, the law of mass action was linked to one more central concept in chemistry, the equilibrium constant[3], K. A case in point, to determine K for the following binding reaction,

$$A + B \rightleftharpoons AB \qquad , \tag{1}$$

the average concentration, at equilibrium, of the product and that of each of the reactants are obtained and then the ratio $\langle c_{AB} \rangle / (\langle c_A \rangle \langle c_B \rangle)$ is computed, where the brackets indicate either an average over measurement time or an ensemble average. This expression of K and the corresponding definition of the rate constant of the forward reaction, $k_{fw} = \langle \textit{fw-rate} \rangle / (\langle c_A \rangle \langle c_B \rangle)$, have been working faithfully for several generations of chemists without raising any suspicion they might be only special cases applicable for large enough systems.

Yet with recent advancements of technology experimental studies, able to conduct and monitor associations of the type shown in Eq. 1 in systems with small numbers of reactants, reported that bound products are observed at higher concentrations than predicted by the expression of K mentioned above[4–11]. Different explanations were put forward that include conformational changes of the unbound molecule(s), non-fluorescent bindings, and missed events due to transient interactions[12]. Several theoretical studies proposed that small systems, attributed to be stochastic in nature, are characterized by an equilibrium constant different than that observed for the thermodynamic limit, a system attributed to be deterministic[13–19]. Size-dependent equilibrium constant was also advocated by introducing 'nanoconfinement entropic effects on chemical equilibrium' applied only to systems with small number of molecules[20,21]. Other computational works also reported

deviations of the bound product's concentration from that anticipated by the above-mentioned expression of K[22,23]. In these cases, the anomalous behavior was explained by artefacts due to applications of periodic boundary conditions in finite simulation boxes[24–26] or due to neglected concentration fluctuations in small simulations in the canonical ensemble[27,28].

In tackling this issue, we recently argued[29] that any intensive property related to a two-body interaction (such as the concentration of the bound product AB in Eq. 1) changes its average value upon scaling-down homogeneously the size of the system (i.e., scaling all extensive parameters specifying the system by the same factor) to, or within, a regime outside the thermodynamic limit. The reason for this, unlooked for, behavior is the existence of two-body correlations in the system, and the known expressions of the equilibrium- and rate-constants mentioned above should actually take the form of $K = \langle c_{AB} \rangle / \langle c_A \cdot c_B \rangle$ and $k_{fw} = \langle \textit{fw-rate} \rangle / \langle c_A \cdot c_B \rangle$. In both cases it is the average of the product, and not the product of the averages, of reactants' concentrations that need to be considered. It is likely this concept has been overlooked in the literature because in all statistical-mechanics textbooks[30–32], the ensemble constructed to derive K ignores fluctuations in the numbers (or densities) of the chemical components, and thereby, can yield an expression valid only for the thermodynamic limit. Accordingly, literature works that followed ignored these correlations in reactants' concentrations when calculating K[33–42].

Girded with knowledge of the mathematical form of K and k_{fw} for the reaction in Eq. 1, it seems only trivial to write down the corresponding expressions for the following dimerization reaction,

$$2A \rightleftharpoons A_2 \qquad , \tag{2}$$

where reactant B in Eq. 1 is substituted with another reactant of A in Eq. 2, as,

$$K'' = \frac{\langle c_{A_2} \rangle}{\langle c_A^2 \rangle} \cdot c^{\varnothing} \qquad , \tag{3}$$

for the equilibrium constant, where for consistency with the definition of K stated in Eq. 5 below

we multiplied the ratio by the standard concentration $c^{\varnothing}$, and as,

$$k''_{fw} = \frac{\langle \textit{fw-rate} \rangle}{\langle c_{A}^{2} \rangle} \qquad , \tag{4}$$

for the bimolecular rate constant. That said, it appears unconsciously that the way chemists, including the writer of these lines[29], practice chemistry is deeply rooted in the behavior of macroscopic systems. More concretely, the expressions in Eq. 3 and Eq. 4 are incorrect, and although for systems with large number of particles the errors are negligible, at finite systems they are significant. Indeed, correlations between reactant's particles ought to be accounted for in these expressions. In the binding reactions of Eq. 1 the correlations are between two different types of particles and the term $\langle N_A \cdot N_B \rangle$, or alternatively $\langle c_{A} \cdot c_{B} \rangle$, properly counts these two-body correlations. However, in Eq. 2 the correlations are between the same type of particles and a term of the form $\langle N_A^2 \rangle$, or $\langle c_{A}^{2} \rangle$, counts not only correlations between different particles of A but also correlations of a labeled particle with itself. These latter N_A self-correlations are irrelevant for two-body interactions and should be subtracted to yield a term proportional to $\langle N_A(N_A - 1) \rangle$ or $\langle c_{A}(c_{A} - 1/V) \rangle$. Nevertheless, this subtraction is not performed actively but emerges naturally when deriving K as shown below.

Results

I. Derivation of the Equilibrium Constant for Dimerizations

We consider the dimerization process shown in Eq. 2 to take place in the gas phase, where the behavior of all components is assumed ideal. This means except of the reaction described, the particles do not interact with one another and no higher-order clustering occurs. The equilibrium constant, K, is defined by,

$$K = e^{-\Delta G^{\varnothing}/RT} \quad , \tag{5}$$

where $\Delta G^{\varnothing}$, the standard Gibbs energy change of the reaction, is the change in Gibbs free energy when one mole of A dimerize with another mole of A to form one mole of A_2, under conditions in which both the reactant and product are at their standard state of temperature and (partial) pressure. For all gases, almost always, same values of temperature and pressure define the standard state. Instead of a standard pressure we will often indicate the corresponding standard concentration, $c^{\varnothing}$. Although reported per mole of dimer formation, $\Delta G^{\varnothing}$ is usually measured for a different (yet macroscopic) number of particles. Given the volume of this reference system, $V^{\varnothing}$, the number of dimers formed in a complete transformation of this reference reaction is $N^{\varnothing}_{A_2} \equiv N^{\varnothing} = c^{\varnothing}V^{\varnothing}$.

For convenience, we choose to perform our derivation in the canonical ensemble. However in contrast to the binding reaction in Eq. 1, the canonical ensemble for dimerization can not connect directly monomers at standard conditions to dimers at the same standard conditions. If the volume on both side of the chemical equation in Eq. 2 is the same, the pressure and thereby concentration of the $2N^{\varnothing}$ monomers will necessarily be twice those of the dimers. To rectify this situation, the reaction ought to start with a system of monomers in double the volume, thus $2V^{\varnothing}$, where the pressure and concentration have their standard values, followed by a reversible isothermal compression to a volume of $V^{\varnothing}$. The work of this hypothetical compression[43],

$$W^{\text{reversible}}_{P^{\varnothing},2V^{\varnothing}\rightarrow 2P^{\varnothing},V^{\varnothing}} = -2N^{\varnothing}k_{_B}T\ln\frac{V^{\varnothing}}{2V^{\varnothing}} \quad , \tag{6}$$

should then be accounted for when calculating $\Delta G^{\varnothing}$ (see Fig. 1). To put it another way, this

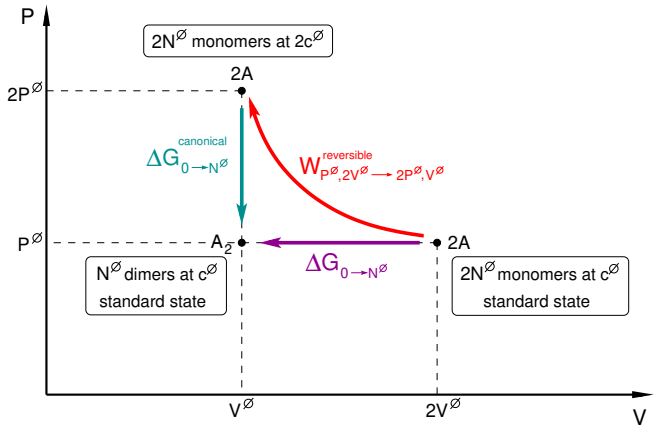

Figure 1: Projection of the dimerization reaction of the reference system onto an isothermal pressure-volume diagram. The figure illustrates that in this case, connecting the reactant ($2A$ monomers) to the product (A_2 dimers), both at standard state conditions (purple arrow), via a description in the canonical ensemble (green arrow) requires an additional process in which the reactant is reversibly compressed to a volume $V^\varnothing$ (red arrow).

additional compression step had to be introduced when utilizing the canonical ensemble because the stoichiometric coefficients of reactant and product in the dimerization reaction (Eq. 2) are not equal, whereas the conditions, in particular the pressure (or concentration), defining the standard states are the same.

Once the $2N^\varnothing$ monomers are compressed to $V^\varnothing$, we proceed to describe the dimerization in the canonical ($N^\varnothing$, $V^\varnothing$, T) ensemble. Upon the formation of one dimer, the energy of the system changes by an amount of ϵ (i.e., $\epsilon < 0$). Owing to the ideal behavior of the chemical components, the (interparticle) energy states of the system are uniquely defined by the number of dimers, $N_{A_2} \equiv i$, and the canonical partition function of the reference system can be written as,

$$Q^\varnothing = \sum_{i=0}^{N^\varnothing} \frac{\left(q_A^\varnothing\right)^{2(N^\varnothing - i)}}{[2(N^\varnothing - i)]!} \cdot \frac{\left(q_{A_2}^\varnothing\right)^i}{i!} \quad , \tag{7}$$

where the number of monomers, N_A, equals $2(N^\varnothing - i)$. As it should, the sum in Eq. 7 takes into account all possible energy states of the system. $q_A^\varnothing$ is the single-particle partition function

of one monomeric particle (which includes only summation over internal energies) and $q^{\varnothing}_{A_2}$ is the pair-particle partition function of one dimer A_2 (incorporating the exponential $e^{-\beta\epsilon}$). These partition functions can be expressed in different forms and are described in details in the Supporting Information. Because the particles are indistinguishable, the factorials in the denominators of Eq. 7 correct the over-counting when raising the single/pair partition functions to the power of the particle numbers. Utilizing $Q^{\varnothing}$, we calculate the Helmholtz free energy change, $\Delta F^{\text{canonical}}_{0\to N^{\varnothing}}$, for the formation of $N^{\varnothing}$ dimers (i.e., at a concentration of $c^{\varnothing}$) from $2N^{\varnothing}$ monomers (at a concentration of $2c^{\varnothing}$). The superscript 'canonical' denotes this free energy change is calculated only for the process at constant volume. This change in Helmholtz energy is obtained from the ratio of the probability to find all particles in the system as dimers, p^{A_2} (i.e., the fraction of the state $i = N^{\varnothing}$ in the sum of Eq. 7), to the probability to find all particles as free monomers, p^{2A} (the fraction of the state $i = 0$),

$$\Delta F^{\text{canonical}}_{0\to N^{\varnothing}} \equiv F^{\text{canonical}}_{i=N^{\varnothing}} - F^{\text{canonical}}_{i=0} = -k_{_B}T \ln \frac{p^{A_2}}{p^{2A}} = -k_{_B}T \ln \left[\frac{\left(q^{\varnothing}_{A_2}\right)^{N^{\varnothing}}}{N^{\varnothing}!} \cdot \frac{(2N^{\varnothing})!}{\left(q^{\varnothing}_{A}\right)^{2N^{\varnothing}}} \right] \quad , \tag{8}$$

where $k_{_B}$ is Boltzmann constant. The corresponding Gibbs free energy change is then,

$$\Delta G^{\text{canonical}}_{0\to N^{\varnothing}} = \Delta F^{\text{canonical}}_{0\to N^{\varnothing}} + V^{\varnothing}\Delta P^{\text{canonical}}_{0\to N^{\varnothing}} = -N^{\varnothing}k_{_B}T \ln \frac{q^{\varnothing}_{A_2}}{\left(q^{\varnothing}_{A}\right)^2} - k_{_B}T \ln \frac{(2N^{\varnothing})!}{N^{\varnothing}!} + V^{\varnothing}\Delta P^{\text{canonical}}_{0\to N^{\varnothing}} \quad , \tag{9}$$

where $\Delta P^{\text{canonical}}_{0\to N^{\varnothing}}$ is the change in the pressure of the system accompanied the dimerization reaction at constant volume. To get $\Delta G_{0\to N^{\varnothing}}$, we add $W^{\text{reversible}}_{P^{\varnothing},2V^{\varnothing}\to 2P^{\varnothing},V^{\varnothing}}$ (as computed in Eq. 6) to $\Delta G^{\text{canonical}}_{0\to N^{\varnothing}}$, and continue by applying Stirling's approximation to the numerator and denominator of the second term on the right hand side of Eq. 9, that means requiring the reference system to be large,

$$\begin{aligned} \Delta G_{0\to N^{\varnothing}} &= W^{\text{reversible}}_{P^{\varnothing},2V^{\varnothing}\to 2P^{\varnothing},V^{\varnothing}} + \Delta G^{\text{canonical}}_{0\to N^{\varnothing}} \\ &= -N^{\varnothing}k_{_B}T \left[\ln \frac{q^{\varnothing}_{A_2}}{\left(q^{\varnothing}_{A}\right)^2} + \ln N^{\varnothing} \right] + N^{\varnothing}k_{_B}T + V^{\varnothing}\Delta P^{\text{canonical}}_{0\to N^{\varnothing}} \quad . \end{aligned} \tag{10}$$

Substituting $N^{\varnothing}$ with $c^{\varnothing}V^{\varnothing}$, and noting for ideal gases the term $V^{\varnothing}\Delta P^{\text{canonical}}_{0\to N^{\varnothing}}$ equals $-N^{\varnothing}k_{B}T$,

$$\Delta G_{0\to N^{\varnothing}} = -N^{\varnothing}k_{B}T\left[\ln\frac{q^{\varnothing}_{A_2}/V^{\varnothing}}{\left(q^{\varnothing}_{A}/V^{\varnothing}\right)^2} + \ln c^{\varnothing}\right] \qquad . \tag{11}$$

We now evaluate the ratio of the partition functions in Eq. 11. Although $q^{\varnothing}_{A_2}$ and $q^{\varnothing}_{A}$ depend on the size of the system because of the translational partition functions, they can be rendered size-independent upon division by the volume. Hence if we consider another system for the dimerization process in Eq. 2, at the same temperature T but with an arbitrary total number of monomers, N^{total}_{A}, and an arbitrary volume, V, the following relation is obeyed,

$$\frac{q^{\varnothing}_{A_2}/V^{\varnothing}}{\left(q^{\varnothing}_{A}/V^{\varnothing}\right)^2} = \frac{q_{A_2}/V}{\left(q_{A}/V\right)^2} \qquad , \tag{12}$$

where q_{A} and q_{A_2} are the single- and pair-particle partition functions of this arbitrary system. Accordingly, the total partition function is analogous to that of the reference system (Eq. 7), however, we write it in a slightly different form. The reason is that in the reference system we assumed the total number of monomers, $2N^{\varnothing}$, to be an even number. This is a valid assumption for the reference system because the contribution of one particle out of an Avogadro's number of particles is negligible. Note also that Stirling's approximation was applied only to the reference system, and therefore, the arbitrary system can, in principle, be as small as possible (e.g., N^{total}_{A} equals 2 or 3). Thus, the assumption of the total number of monomers to be an even number is not correct for the arbitrary system. As a consequence we set $N^{\text{total}}_{A} = N_A + 2N_{A_2} = 2N^{\circ} + \delta$, where N° is the maximum number of dimers that can hypothetically form, and δ equals 0 or 1 depending on whether N^{total}_{A} is even or odd, respectively. We then write the canonical partition function for the arbitrary system as,

$$Q = \sum_{i=0}^{N^{\circ}} \frac{q_{A}^{2(N^{\circ}-i)+\delta}}{[2(N^{\circ}-i)+\delta]!}\cdot\frac{q_{A_2}^{i}}{i!} \qquad , \tag{13}$$

where as before, $i \equiv N_{A_2}$.

In order to proceed with the evaluation of $\Delta G_{0\to N^{\varnothing}}$ (Eq. 11) we multiply and divide the right-

hand side of Eq. 12 by,

$$\sum_{i=0}^{N^\circ-1} \frac{(i+1)}{[2(N^\circ-(i+1))+\delta]!\ (i+1)!}\ q_A^{2(N^\circ-i)+\delta}\ q_{A_2}^i \qquad , \tag{14}$$

and obtain,

$$V^\varnothing \frac{q_{A_2}^\varnothing}{\left(q_A^\varnothing\right)^2} = V\frac{q_{A_2}}{q_A^2} = V\frac{\sum_{i=0}^{N^\circ-1} \frac{(i+1)}{\left[2(N^\circ-(i+1))+\delta\right]!\ (i+1)!}\ q_A^{2[N^\circ-(i+1)]+\delta}\ q_{A_2}^{i+1}}{\sum_{i=0}^{N^\circ-1} \frac{(i+1)}{[2(N^\circ-(i+1))+\delta]!\ (i+1)!}\ q_A^{2(N^\circ-i)+\delta}\ q_{A_2}^i} \qquad . \tag{15}$$

We change the index of the sum in the numerator to $j = i+1$ and rewrite the factorials in the denominator,

$$V^\varnothing \frac{q_{A_2}^\varnothing}{\left(q_A^\varnothing\right)^2} = V\frac{\sum_{j=1}^{N^\circ} \frac{j}{\left[2(N^\circ-j)+\delta\right]!\ j!}\ q_A^{2(N^\circ-j)+\delta}\ q_{A_2}^j}{\sum_{i=0}^{N^\circ-1} \frac{[2(N^\circ-i)+\delta-1][2(N^\circ-i)+\delta]}{[2(N^\circ-i)+\delta]!\ i!}\ q_A^{2(N^\circ-i)+\delta}\ q_{A_2}^i} \qquad . \tag{16}$$

Given the form of the coefficients of the partition functions in the sum, index j in the numerator can start from zero and index i in the denominator can end at N° (remembering δ equals either 0 or 1). This yields,

$$\begin{aligned} V^\varnothing \frac{q_{A_2}^\varnothing}{\left(q_A^\varnothing\right)^2} &= V\frac{\frac{1}{Q}\sum_{j=0}^{N^\circ}\ j\ \frac{q_A^{2(N^\circ-j)+\delta}}{[2(N^\circ-j)+\delta]!}\frac{q_{A_2}^j}{j!}}{\frac{1}{Q}\sum_{i=0}^{N^\circ}\ [2(N^\circ-i)+\delta][2(N^\circ-i)+\delta-1]\ \frac{q_A^{2(N^\circ-i)+\delta}}{[2(N^\circ-i)+\delta]!}\frac{q_{A_2}^i}{i!}} \\ &= V\frac{\langle N_{A_2}\rangle}{\langle N_A(N_A-1)\rangle} = \frac{\left\langle c_{A_2}\right\rangle}{\left\langle c_A(c_A-\frac{1}{V})\right\rangle} \qquad , \end{aligned} \tag{17}$$

where the sum in the numerator is the ensemble average of the number of dimers, $\langle N_{A_2}\rangle$, and the sum in the denominator is the average of the product of $N_A(N_A-1)$, both in our chosen arbitrary system under equilibrium conditions. Inserting this result into Eq. 11 gives,

$$\Delta G_{0\to N^\varnothing} = -N^\varnothing k_{_B} T \ln \frac{\left\langle c_{A_2}\right\rangle\ c^\varnothing}{\langle c_{_A}\cdot(c_{_A}-1/V)\rangle} \qquad . \tag{18}$$

Scaling $\Delta G_{0\to N^\varnothing}$ to one mole of formed dimer yields $\Delta G^\varnothing$,

$$\Delta G^\varnothing = \frac{N_{\text{Avogadro}}}{N^\varnothing}\cdot\Delta G_{0\to N_A^\varnothing} = -RT\ln\frac{\left\langle c_{A_2}\right\rangle\ c^\varnothing}{\langle c_{_A}\cdot(c_{_A}-1/V)\rangle} \qquad , \tag{19}$$

and comparing the resulting expression to the definition of K in Eq. 5 we arrive at,

$$K = \frac{\left\langle c_{A_2} \right\rangle}{\left\langle c_A \cdot (c_A - \frac{1}{V}) \right\rangle} \, c^{\varnothing} \qquad . \tag{20}$$

Therefore as for the case of binding reaction in Eq. 1, the equilibrium constant for dimerization must include correlations between the unbound reactants (monomers), however here, self-correlations are subtracted, that is, the correlation between a tagged particle with itself. Note that if we did not consider the reversible work for compression (i.e., taking into account only $\Delta G^{\text{canonical}}_{0 \to N^{\varnothing}}$) the expression of K would be the same as that in Eq. 20 but multiplied by a factor of 4.

We would like to point out two special cases. The first is the thermodynamic limit, where $\langle N_A(N_A - 1) \rangle \to \langle N_A^2 \rangle$, or alternatively $1/V \to 0$, and correlations between the reactant are totally lost. In this case, K'' in Eq. 3 and a related expression ignoring all correlations,

$$K' = \frac{\left\langle c_{A_2} \right\rangle}{\langle c_A \rangle \langle c_A \rangle} \cdot c^{\varnothing} \qquad , \tag{21}$$

approach K in Eq. 20. The second case is for the smallest system possible, $N_A^{\text{total}} = 2$, where the system has only two macroscopic states. Despite strong correlations in the system, the two-body average $\langle N_A(N_A - 1) \rangle$ reduces to a one-body average $\langle N_A \rangle$, and Eq. 20 can be written as,

$$K_{N_A^{\text{total}}=2} = \frac{f^{A_2}}{2(1 - f^{A_2})} \, V c^{\varnothing} \qquad , \tag{22}$$

where $f^{A_2} \equiv \langle N_{A_2} \rangle$ is the fraction of frames in which the dimer is observed. The relation in Eq. 22 is identical to that derived by Ouldridge et al.[27].

II. Validation by Computer Simulations

To check our derivation, we consider a simple system of Lennard-Jones (LJ) molecules able to dimerize according to Eq. 2 and let the system propagate by Monte-Carlo (MC) and molecular dynamics (MD) algorithms. Two series of simulations were performed. In the first, R1, we increased N_A^{total} keeping the concentration $c_A^{\text{total}} = N_A^{\text{total}}/V$ constant, whereas in the second series, R2, we fixed $N_A^{\text{total}} = 2$ and increased V by increasing the length of the cubic simulation box, L_{box}. Detailed

information on the model system and computational methodologies are given in the Computational Details section below.

Figure 2 displays the equilibrium constant, K, calculated by Eq. 20, together with the value of K' (Eq. 21) and K'' (Eq. 3). As indicated by the figure, inclusion of cross-correlations are needed

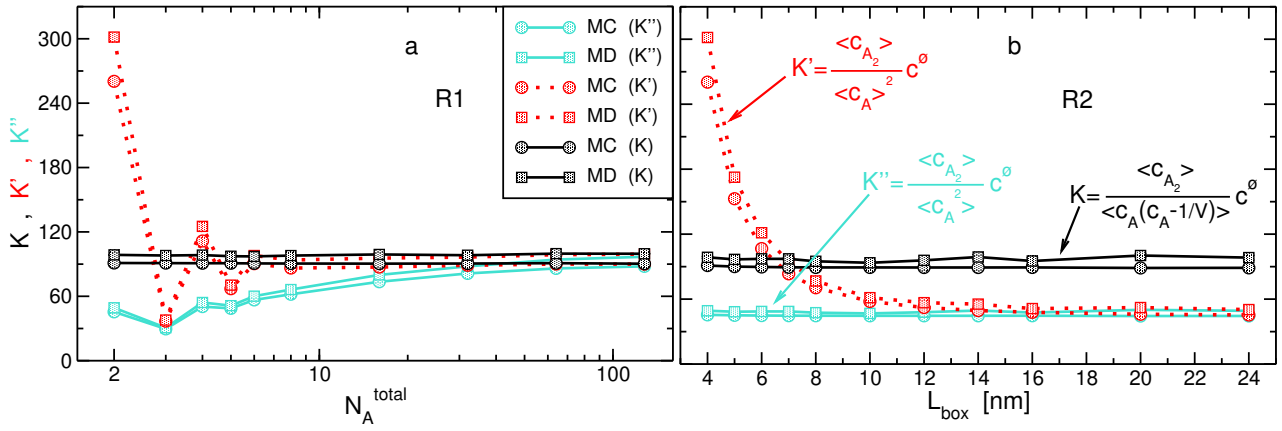

Figure 2: The equilibrium constant K for dimerization defined by Eq. 20 ($c^{\varnothing} \equiv 1\ M$) for two series of simulations at: (a) constant $c_A^{\text{total}} = 0.052\ M$ and as a function of the total number of A particles (R1), as well as, at (b) constant $N_A^{\text{total}} = 2$ and as a function of the length of the simulation box (R2). Both series were performed in the canonical ensemble at $T = 300\ K$ by Monte-Carlo (MC) and molecular-dynamics (MD) methods. The values of K' and K'' defined in Eq. 21 and Eq. 3 are also shown for comparison. The left-most points in R1 and R2 ($N_A^{\circ} = 2$, $L_{box} = 4\ nm$) represent the same system. The estimated errors for the values of K are smaller than the size of the symbols.

in order to keep the equilibrium constant constant for all simulations in both series. That means, self-correlations, $\langle c_A/V \rangle$, must be subtracted from the correlation term $\langle c_A^2 \rangle$. Notice, whereas K' and K'' approach K with increasing N_A^{total} in R1, they approach a different value than that of K with increasing L_{box} in R2. This is because in the former, subtracting 1 from an increasing number of N_A particles will eventually become negligible, whereas, subtracting 1 from a maximum value of 2 is always significant. Furthermore, in the Supporting Information we show the value of K calculated by Eq. 20, for a system with a single-site reactant, agrees almost perfectly with that

obtained by analytical calculations.

The expression of K in Eq. 20 can also be justified from kinetics. The rate of the forward reaction is proportional to the collision probability between a tagged particle A_1 and any other particle A_i (where $i \neq 1$), summed over all N_A particles. This yield a collision probability that is a function of the term $\langle N_A(N_A-1)\rangle$, thereby excluding the impossible event of a collision of a particle with itself. Hence we write,

$$\langle \textit{fw-rate} \rangle = k_{fw}\langle c_{_A}(c_{_A}-1/V)\rangle \qquad . \tag{23}$$

The backward reaction is a simple first-order kinetics and its rate is proportional linearly to dimer concentration. At equilibrium, there is no change in average concentration of any of the chemical components,

$$\left\langle \frac{dc_{_{A_2}}}{dt} \right\rangle = -\frac{1}{2}\left\langle \frac{dc_{_A}}{dt} \right\rangle = \left\langle k_{fw}c_{_A}(c_{_A}-1/V) - k_{bw}c_{_{A_2}} \right\rangle = 0 \qquad , \tag{24}$$

and if we define K as the ratio between forward and backward rate constants and render its value dimensionless via $c^{\varnothing}$, we recuperate Eq. 20. In fact, plotting (Fig. 3) the rate constant of the forward reaction, k_{fw}, defined in Eq. 23, together with k''_{fw} (Eq. 4) which includes self-correlations, and that ignoring correlations all together,

$$k'_{fw} = \frac{\langle \textit{fw-rate} \rangle}{\langle c_{_A}\rangle^2} \qquad , \tag{25}$$

mirrors the results presented for the corresponding expressions of K.

III. Predicting Compositions from Fluctuations

The main difference between thermodynamics and statistical mechanics is that the latter incorporates fluctuations in the values of the system's properties. The magnitudes of these fluctuations depend on the parameters specifying the system, and generally, can be used to extract information on the system. For the dimerization reaction considered here, we demonstrate now the information that can be extracted from fluctuations is the composition of the system. To represent fluctuations

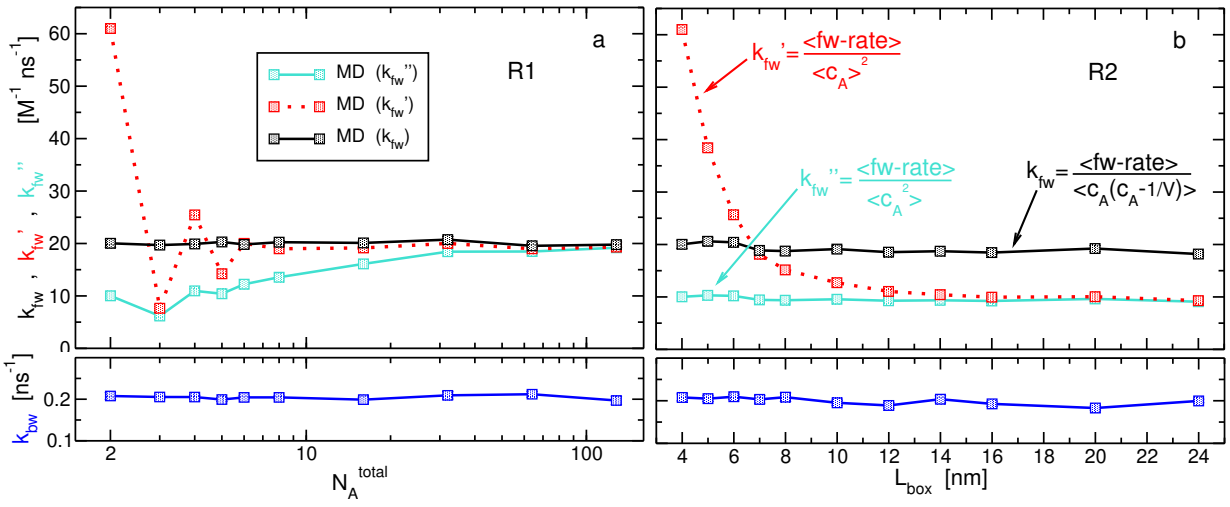

Figure 3: Rate constants of dimerization (Eq. 2) for (a) R1 series, and (b) R2 series, obtained from MD simulations. The top panels show the rate constant in the forward direction, k_{fw} defined in Eq. 23, whereas the lower panels the rate constant in the backward direction, $k_{bw} = \langle bw\text{-}rate \rangle / \langle c_{A_2} \rangle$. For comparison, we also present k'_{fw} and k''_{fw} defined in Eq. 25 and Eq. 4.

we adopt the notation of Lebowitz et al.[44] and define the cross fluctuations between quantities ζ and η as,

$$L(\zeta, \eta) = \langle \zeta\eta \rangle - \langle \zeta \rangle \langle \eta \rangle \quad , \tag{26}$$

and their relative magnitude by,

$$l(\zeta, \eta) = \frac{L(\zeta, \eta)}{\langle \zeta \rangle \langle \eta \rangle} \quad . \tag{27}$$

We now look at the following difference in our system,

$$l(N_{A_2}, N_{A_2}) - l[N_{A_2}, N_A(N_A - 1)] = \frac{1}{\langle N_{A_2} \rangle} \left[\frac{\langle N_{A_2}^2 \rangle}{\langle N_{A_2} \rangle} - \frac{\langle N_{A_2} N_A (N_A - 1) \rangle}{\langle N_A (N_A - 1) \rangle} \right] \quad , \tag{28}$$

and evaluate the term inside the square brackets. Utilizing the partition function defined in Eq. 13 and recalling that $i \equiv N_{A_2}$ and $N_A = N_A^{\text{total}} - 2N_{A_2} = 2N^\circ + \delta - 2i$, the first term can be written as,

$$\frac{\langle N_{A_2}^2 \rangle}{\langle N_{A_2} \rangle} = \frac{\frac{1}{Q} \sum_{i=0}^{N^\circ} i^2 \, \frac{q_A^{2(N^\circ - i) + \delta}}{[2(N^\circ - i) + \delta]!} \cdot \frac{q_{A_2}^i}{i!}}{\frac{1}{Q} \sum_{i=0}^{N^\circ} i \, \frac{q_A^{2(N^\circ - i)}}{[2(N^\circ - i)]!} \cdot \frac{q_{A_2}^i}{i!}} = \frac{\sum_{j=0}^{N^\circ - 1} (j+1)^2 \, \frac{q_A^{2(N^\circ - j) + \delta}}{[2(N^\circ - (j+1)) + \delta]!} \cdot \frac{q_{A_2}^j}{(j+1)!}}{\sum_{j=0}^{N^\circ - 1} (j+1) \, \frac{q_A^{2(N^\circ - j) + \delta}}{[2(N^\circ - (j+1)) + \delta]!} \cdot \frac{q_{A_2}^j}{(j+1)!}} \quad , \tag{29}$$

where in the second equality we skipped the terms corresponding to $i = 0$, changed the index of the summation to $j = i - 1$, and multiplied and divided the ratio by q_A^2/q_{A_2}.

Similarly, we can express the second term inside the square brackets in Eq. 28 by,

$$\frac{\langle N_{A_2} N_A(N_A-1)\rangle}{\langle N_A(N_A-1)\rangle} = \frac{\frac{1}{Q}\sum_{i=0}^{N^\circ} i[2(N^\circ-i)+\delta][2(N^\circ-i)+\delta-1]\ \frac{q_A^{2(N^\circ-i)+\delta}}{\left[2(N^\circ-i)+\delta\right]!}\cdot\frac{q_{A_2}^i}{i!}}{\frac{1}{Q}\sum_{i=0}^{N^\circ}[2(N^\circ-i)+\delta][2(N^\circ-i)+\delta-1]\ \frac{q_A^{2(N^\circ-i)+\delta}}{\left[2(N^\circ-i)+\delta\right]!}\cdot\frac{q_{A_2}^i}{i!}}$$
$$= \frac{\sum_{i=0}^{N^\circ-1} i(i+1)\frac{q_A^{2(N^\circ-i)+\delta}}{\left[2(N^\circ-(i+1))+\delta\right]!}\cdot\frac{q_{A_2}^i}{(i+1)!}}{\sum_{i=0}^{N^\circ-1}(i+1)\frac{q_A^{2(N^\circ-i)+\delta}}{\left[2(N^\circ-(i+1))+\delta\right]!}\cdot\frac{q_{A_2}^i}{(i+1)!}} \quad , \tag{30}$$

where the second equality is realized by letting index i in the sum end at $N^\circ - 1$ (again, δ is a binary parameter of 0 or 1) and rewriting the factorials. Now we subtract the second term from the first tern in the square brackets of Eq. 28 by noting the denominators of the two terms are equal,

$$\frac{\langle N_{A_2}^2\rangle}{\langle N_{A_2}\rangle} - \frac{\langle N_{A_2} N_A(N_A-1)\rangle}{\langle N_A(N_A-1)\rangle} = \frac{\sum_{i=0}^{N^\circ-1}(i+1)\frac{q_A^{2(N^\circ-i)+\delta}}{\left[2(N^\circ-(i+1))+\delta\right]!}\cdot\frac{q_{A_2}^i}{(i+1)!}}{\sum_{i=0}^{N^\circ-1}(i+1)\frac{q_A^{2(N^\circ-i)+\delta}}{\left[2(N^\circ-(i+1))+\delta\right]!}\cdot\frac{q_{A_2}^i}{(i+1)!}} = 1 \quad , \tag{31}$$

and obtain a ratio that equals one. Consequently, Eq. 28 reduces to,

$$l(N_{A_2}, N_{A_2}) - l[N_{A_2}, N_A(N_A-1)] = \frac{1}{\langle N_{A_2}\rangle} \quad , \tag{32}$$

or to a similar expression specifying the average concentration of the dimer,

$$\langle c_{A_2}\rangle = \frac{1}{\{l(N_{A_2}, N_{A_2}) - l[N_{A_2}, N_A(N_A-1)]\}\, V} \quad . \tag{33}$$

In Fig. 4 we examine the validity of Eq. 33 by computing these relative fluctuations and compare the predicted values of $\langle c_{A_2}\rangle$ to those obtained by direct counting of dimers. The agreement is excellent, nevertheless, there are two points with noticeable discrepancies. They appear in R1 series by MD simulations for the two largest N_A^{total} values (64 and 128), which we conjecture to arise due to insufficient simulation time to yield accurate averages for the relative fluctuations. Note in R1 series all extensive parameters specifying the system are scaled by the same factor, and therefore,

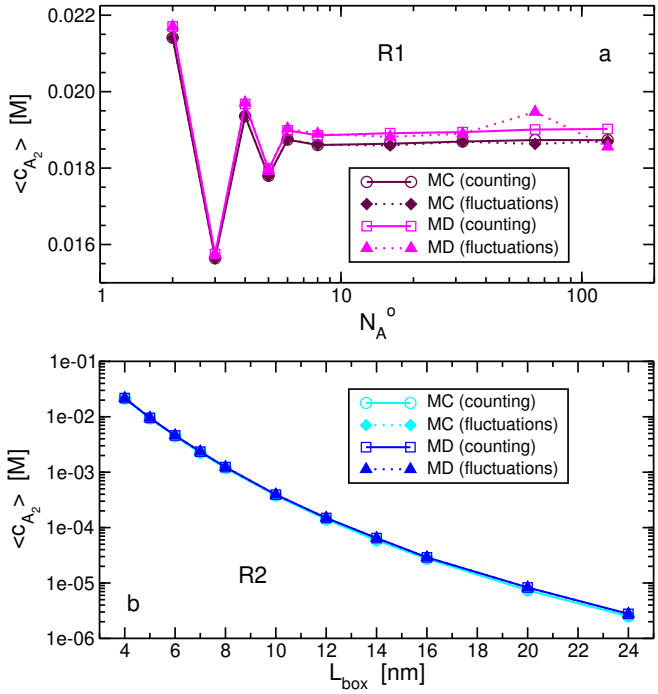

Figure 4: Average concentrations of dimers, $\langle c_{A_2} \rangle$, for (a) R1 and (b) R2 series of simulations. Along results obtained from direct counting of the number of dimers, we also predict the concentrations from the relative fluctuations, $l(N_{A_2}, N_{A_2})$ and $l(N_{A_2}, N_A[N_A - 1])$, in the system as described in Eq. 33.

if average quantities of the system were homogeneous functions then $\langle c_{A_2} \rangle$ would be constant. This is not the case at finite systems and instead there is a rising divergence from a horizontal line with scaling-down the size of the system, the same as that observed[29] for the binding reaction of Eq. 1. However for (homo-) dimerization, odd values of N_A^{total} (3 and 5) exhibit strong reduction in $\langle c_{A_2} \rangle$, breaking up the continuous curve, simply because in these cases it is unfeasible to pair all particles simultaneously whereas for even numbers of N_A^{total} it is. Although it is clear this odd effect diminishes with increasing numbers of particles, these two points are the only evidence we have because other odd numbers were not considered.

Discussion

There are two points we would like to discuss. The first concerns the concept reported in the literature of a size-dependent equilibrium constant. As follows from our treatment, K is a quantity defined not for the system we have at hand but for a macroscopic system at agreed conditions of temperature and pressure (or concentration) as specified in Eq. 5. Thus at constant temperature, it has a fixed, or constant, value regardless of the size of system we choose to work with. It has been known for a long time that K can be extracted utilizing other systems, for example with different concentrations, by applying a relation such as the one shown in Eq. 21. For finite systems this relation yields different values, yet, there is no justification to claim the value of K is now different. One might argue that it is not possible to obtain K from systems that are too small. Contrary to this statement, a main conclusion of current and previous[29] papers is that K can be retrieved from a system of any size, including a system with the smallest possible number of particles. To this end, the employment of a general relation between K and equilibrium properties of the chosen system (e.g., Eq. 20) is required.

The second point concerns the ascription made in the literature of small systems as stochastic, and of macroscopic systems as deterministic, in character. It is likely this attribution is not related to the forces/algorithm propagating the system, but to the fact that if we measure a property of a small system at different points in time we obtain different values, whereas, for a macroscopic system the results are always almost the same. This is obvious; an average over space, or number of particles, in macroscopic systems is sufficient to yield converged quantities, whereas, finite systems require an ensemble large enough, or repetitive instantaneous measurements spanned over long-enough period of time, to yield convergence. That means, sufficient statistical data is necessary, however even when this condition is met, it is not to say average values obtained from large and small systems are the same. On the contrary, and in contrast to the thermodynamic limit, another main conclusion of current and previous[29] works is that properties of chemical equilibriums involving two-body interactions are not homogeneous functions.

Conclusions

In this paper we derive the expressions of the equilibrium constant, Eq. 20, and of the rate of the forward bimolecular reaction, Eq. 23, ought to be used in dimerization reactions of the type presented in Eq. 2. These expressions account for cross-correlations between reactant particles and are, therefore, different from those presented in textbooks. Nevertheless, they do reduce to the textbooks' well-known expressions for large enough (macroscopic) systems. In this case, correlations between reactant particles vanish and the contribution of self-correlations becomes negligible. An important effect of the underlying two-body interactions, is that in a regime outside the thermodynamic limit (thus, for small systems), scaling the system homogeneously will change the average values of intensive properties, such as the concentration of dimer or monomer. We further derive a relation connecting these size-dependent concentrations to relative fluctuations in the systems.

Computational Details

The model system consists of A molecules where each molecule is represented by two sites, a and h, 'covalently' bonded with a bond-length of $0.25\ nm$ as shown schematically in Fig. 5. The role

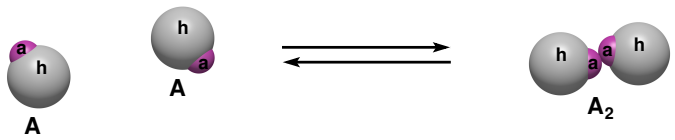

Figure 5: A model system for dimerization between two A molecules. These A molecules consist of uncharged LJ, a and h, atom-sites covalently bonded to each other. The distance of this intramolecular bond is fixed in the MC simulations to a value of $0.25\ nm$ whereas it oscillates around this value, due to a harmonic potential, in the MD simulations. The interaction between the a sites is strongly attractive, whereas the other two intermolecular interactions are repulsive (see Table 1).

of the h atoms is to prevent any clustering of the molecules, apart from dimer formation. All atom-sites have zero charge, $q_a = q_h = 0.0\ e$, and their intermolecular interactions are modeled by Lennard-Jones (LJ) potentials truncated at a distance of $2.0\ nm$. The different σ and ϵ LJ parameters in this system, specified in Table 1, describe essentially repulsive interactions between all sites except for a strong attraction between the a atoms. Based on the location of the first

Table 1: LJ parameters between the different atom sites for a system of $A(a-h)$ molecules.

	$\sigma\ \ [nm]$	$\epsilon\ \ [kJ/mol]$
$a\cdots a$	0.15	47.0
$h\cdots h$	0.85	0.1
$a\cdots h$	0.40	0.1

minimum of the radial distribution function between the a atoms, the dimeric state is defined for $r_{aa} < 0.3\ nm$. Despite the introduction of the protective site h in each molecule A, we did encounter, albeit seldomly, clusters larger than two. These higher order clusters occurred more often in the MD than in the MC simulations, due to the flexibility of the covalent bond in the former, with a percentage of particles involved in these aggregates lower than $0.1\ \%$ and $0.03\ \%$, respectively.

All simulations were conducted in the canonical ensemble $(N_A^{\text{total}}, V, T)$ at a temperature of $T = 300\ K$. The total number of A molecules in the system, $N_A^{\text{total}} = N_A + 2N_{A_2}$, and/or the volume V of the cubic box, varied systematically within two series of simulations. In the first series, labeled R1, we increased N_A^{total} from 2 to 128 and, concomitantly, V such that the concentration $c_A^{\text{total}} = N_A^{\text{total}}/V$ is constant at $0.03125\ molecules/nm^3$ ($\sim 0.052\ M$). In the second series of simulations, R2, we considered only two molecules of A, $N_A^{\text{total}} = 2$, and increased V by increasing the box length, L_{box}, from $4.0\ nm$ to $24.0\ nm$. Periodic boundary conditions were applied along all three Cartesian axes.

Both series of simulations were conducted by Monte-Carlo (MC) and molecular dynamics (MD) techniques. The MC simulations[45,46], which output configurations in the canonical ensemble, were performed by an in-house code executed in double-precision. The Metropolis acceptance criteria[47] was applied to either accept or reject trial moves. Each trial move starts by randomly selecting one A molecule which is then displaced, in each of the three Cartesian-axes, and rotated around each of the two axes perpendicular to the molecular axis. The displacements and rotations are performed as rigid bodies. Their magnitudes and directions were determined randomly from a uniform distribution with maximum values of $0.4\ nm$ for displacements along each of the Cartesian-axes, 0.1 for $\cos\theta$ when rotating around angle θ ($0 \leq \theta \leq \pi$), and $0.314\ rad$ for rotations around angle ϕ ($0 \leq \phi \leq 2\pi$). These trial moves resulted in acceptance-ratios that varied from 0.17, for the system $N_A^{\text{total}} = 2$ in R1, to 0.98, for the system with $L_{box} = 24.0\ nm$ in R2. The number of trial moves applied for each simulation was inversely proportional to the size of the system. For

example the data collection stages ranged from $4 \cdot 10^{12}$ moves for $N_A^{\text{total}} = 2$ to $1.25 \cdot 10^{11}$ moves for the largest system of $N_A^{\text{total}} = 128$.

The MD simulations were conducted by the software package GROMACS version 4.6.5[48] (single-precision). A time step of $0.002\ ps$ was employed to integrate the equations of motion and a mass of $10.0\ amu$ was assigned to the a and h atom sites. The a–h 'covalent' bond was modeled by a harmonic potential with bond-length of $0.25\ nm$ and force-constant of $2{\cdot}10^5\ kJ/(mol{\cdot}nm^2)$. The temperature was maintained by applying the Nosé-Hoover thermostat[49,50] with a chain-length[51] 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. Equilibration time of at least $1\ \mu s$ was conducted prior to data collection for each system, whereas, the time period for collecting data ranged from $400\ \mu s$ for $N_A^{\text{total}} = 2$ to $29.6\ \mu s$ for $N_A^{\text{total}} = 128$.

To analyze the dynamics of the forward and backward reactions we had to simulate again R1 and R2 series by MD. However, this time the trajectories were saved more frequently; from a frequency of every 20 steps for $N_A^{\text{total}} = 2$ (or $L_{box} = 4.0\ nm$) to a frequency of every step for $N_A^{\text{total}} \geq 8$ (R1) or to a frequency of every 1000 steps for $L_{box} = 24.0\ nm$ (R2 series). These frequencies corresponded to, approximately, the lowest frequencies for which trial calculations of the rate constants were not affected upon an increase of the trajectory-saving frequency. At the same time, the duration of trajectories were shorter than those mentioned above and for R1 ranged from $12\ \mu s$ for the smallest system to $300\ ns$ for the largest system in R1, or to $600\ ns$ for the largest system in R2. To keep the size of the trajectories manageable, each run was split into multiple $(10 - 60)$ runs. The rates of the forward and backward reactions were calculated by counting the number of transitions per period of time divided by V. A transition between the two states is identified when the distance between a sites of two molecules crossed the cutoff-value of $0.3\ nm$. To avoid counting return-trajectories originating from transient species in the proximity of the transition state, we introduced a buffer-zone of $0.05\ nm$ on either side of the cutoff such that if a particle is already bound, r_{aa} needs to be larger than $0.35\ nm$ to consider a transition,

whereas if it is unbound, r_{aa} needs to be smaller than $0.25\ nm$ to count a transition. Nevertheless, the effect of including this buffer zone is rather negligible.

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] Wilhemy, L. F. Ueber das Gesetz, nach welchem die Einwirkung der Säuren auf den Rohrzucker stattfindet , *Annalen der Physik und Chemie* **1850**, *81*, 413–433.

[2] Waage, P.; Guldberg, C. M. Studier over Affiniteten, *Forhandlinger i Videnskabs-selskabet i Christiania* **1864**, pages 35–45.

[3] van 't Hoff, J. H. *Studies in Chemical Dynamics;* William & Norgate: London, 1896.

[4] 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. USA* **2007**, *104*, 18013–18018.

[5] Sawada, T.; Yoshizawa, M.; Sato, S.; Fujita, M. Minimal nucleotide duplex formation in water through enclathration in self-assembled hosts, *Nature 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-Gen* **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.; Morello, G.; FitzPatrick, S. R.; Booth, R. L.; Sills, A. J.; Robertson, A. W.; Warner, J. H.; Turner, N. J.; Armstrong, F. A. Electrocatalytic Volleyball: Rapid Nanoconfined Nicotinamide Cycling for Organic Synthesis in Electrode Pores, *Angew. Chemie 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] Yang, J.; Pearson, J. E. Origins of concentration dependence of waiting times for single-molecule fluorescence binding, *J. Chem. Phys.* **2012**, *136*, 244506.

[13] Darvey, I. G.; Ninham, B. W.; Staff, P. J. Stochastic Models for Second-Order Chemical Reaction Kinetics. The Equilibrium State, *J. Chem. Phys.* **1966**, *45*, 2145–2155.

[14] Rothenberger, G.; Grätzel, M. Effects of spatial confinement on the rate of bimolecular reactions in organized liquid media, *Chem. Phys. Lett.* **1989**, *154*, 165–171.

[15] Laurenzi, I. J. An analytical solution of the stochastic master equation for reversible bimolecular reaction kinetics, *J. Chem. Phys.* **2000**, *113*, 3315–3322.

[16] Holcman, D.; Schuss, Z. Stochastic chemical reactions in microdomains, *J. Chem. Phys.* **2005**, *122*, 114710.

[17] 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*, 124112.

[18] Szymanski, R.; Sosnowski, S. Stochasticity of the transfer of reactant molecules between nano-reactors affecting the reversible association $A + B \rightleftarrows C$, *J. Chem. Phys.* **2019**, *151*, 174113.

[19] 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.

[20] Polak, M.; Rubinovich, L. Nanochemical Equilibrium Involving a Small Number of Molecules: A Prediction of a Distinct Confinement Effect, *Nano Lett.* **2008**, *8*, 3543–3547.

[21] Polak, M.; Rubinovich, L. The Intrinsic Role of Nanoconfinement in Chemical Equilibrium: Evidence from DNA Hybridization, *Nano Lett.* **2013**, *13*, 2247–2251.

[22] Kindt, J. T. Accounting for Finite-Number Effects on Cluster Size Distributions in Simulations of Equilibrium Aggregation, *J. Chem. Theory Comput.* **2013**, *9*, 147–152.

[23] 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.

[24] 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. Comp. Chem.* **2011**, *32*, 1919–1928.

[25] 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*, 141103.

[26] 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. Sim.* **2018**, *44*, 599–612.

[27] 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*, 104102.

[28] 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*, 144105.

[29] Zangi, R. Binding Reactions at Finite Systems, *Phys. Chem. Chem. Phys.* **2022**, *24*, 9921–9929.

[30] McQuarrie, D. A. *Statistical Thermodynamics;* University Science Books: Mill Valley, CA, 1973.

[31] Chandler, D. *Introduction to Modern Statistical Mechanics;* Oxford University Press: New York, NY, 1987.

[32] Gould, H.; Tobochnik, J. *Statistical and Thermal Physics: With Computer Applications;* Princeton University Press: Princeton, NJ, 2010.

[33] 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.

[34] 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.

[35] Luo, H.; Sharp, K. On the calculation of absolute macromolecular binding free energies, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 10399–10404.

[36] Zhang, Y.; McCammon, J. A. Studying the affinity and kinetics of molecular association with molecular-dynamics simulation, *J. Chem. Phys.* **2003**, *118*, 1821–1827.

[37] 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.

[38] Deng, Y.; Roux, B. Computations of Standard Binding Free Energies with Molecular Dynamics Simulations, *J. Phys. Chem. B* **2009**, *113*, 2234–2246.

[39] Skorpa, R.; Simon, J.-M.; Bedeaux, D.; Kjelstrup, S. Equilibrium properties of the reaction $H_2 \rightleftharpoons 2H$ by classical molecular dynamics simulations, *Phys. Chem. Chem. Phys.* **2014**, *16*, 1227–1237.

[40] Montalvo-Acosta, J. J.; Cecchini, M. Computational Approaches to the Chemical Equilibrium Constant in Protein-ligand Binding, *Mol. Inform.* **2016**, *35*, 555–567.

[41] Piccolo, N. D.; Hristova, K. Quantifying the Interaction between EGFR Dimers and Grb2 in Live Cells, *Biophys. J.* **2017**, *113*, 1353–1364.

[42] Duboué-Dijon, E.; Hénin, J. Building intuition for binding free energy calculations: Bound state definition, restraints, and symmetry, *J. Chem. Phys.* **2021**, *154*, 204101.

[43] Callen, H. B. *Thermodynamics and an Introduction to Thermostatistics;* John Wiley & Sons: New York, NY, 1985.

[44] Lebowitz, J. L.; Percus, J. K.; Verlet, L. Ensemble Dependence of Fluctuations with Application to Machine Computations, *Phys. Rev.* **1967**, *153*, 250–254.

[45] Allen, M. P.; Tildesley, D. J. *Computer Simulations of Liquids;* Oxford Science Publications: Oxford, 1987.

[46] Frenkel, D.; Smit, B. *Understanding Molecular Simulations: From Algorithms to Applications;* Academic Press: London, 2002.

[47] Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of State Calculations by Fast Computing Machines, *J. Chem. Phys.* **1953**, *21*, 1087–1092.

[48] 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.

[49] Nosé, S. A unified formulation of the constant temperature molecular dynamics methods, *J. Chem. Phys.* **1984**, *81*, 511–519.

[50] Hoove, W. G. Canonical dynamics: Equilibrium phase-space distributions, *Phys. Rev. A* **1985**, *31*, 1695–1697.

[51] Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nosé-Hoover chains: The canonical ensemble via continuous dynamics, *J. Chem. Phys.* **1992**, *97*, 2635–2643.