Binding Reactions at Finite 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*

March 4, 2022

*r.zangi@ikerbasque.org

Abstract

A perpetual yearn exists among computational scientists to scale-down the size of physical systems, a desire shared as well with experimentalists able to track single molecules. A question then arises whether averages observed at small systems are the same as those observed at large, or macroscopic, systems. Utilizing statistical-mechanics formulations in ensembles in which the total numbers of particles are fixed, we demonstrate that properties of binding reactions are not homogeneous functions. That means averages of intensive parameters, such as the concentration of the bound-state, at finite-systems are different than those at large-systems. The discrepancy increases with decreasing temperature, volume, and to some extent, numbers of particles. As perplexing as it may sound, despite variations in average quantities, extracting the equilibrium constant from systems of different sizes does yield the same value. The reason is that correlations in reactants' concentrations are ought be accounted for in the expression of the equilibrium constant, being negligible at large-scale but significant at small-scale. Similar arguments pertain to the calculations of the reaction rate-constants, more specifically, the bimolecular rate of the forward reaction is related to the average of the product (and not to the product of the averages) of the reactants' concentrations. Furthermore, we derive relations aiming to predict the composition only from the equilibrium constant and the system's size. All predictions are validated by Monte-Carlo and molecular dynamics simulations. An important consequence of these findings is that the expression of the equilibrium constant at finite systems is not dictated solely by the chemical equation of the reaction but requires knowledge of the elementary processes involved.

Consider the following association reaction,

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

in which N_A° gas molecules of A and N_B° gas molecules of B are placed in an empty closed container with fixed volume, V, and temperature, T, to reach equilibrium with the bound product, AB. The experiment is then repeated at the same conditions but with λN_A°, λN_B°, and λV instead. Are the concentrations of AB particles in the two experiments equal? In the thermodynamic limit the answer is yes, because intensive and extensive properties are homogeneous zero-order and first-order functions, respectively[1],

$$X(T, \lambda V, \lambda N_A^\circ, \lambda N_B^\circ) = \lambda^\alpha X(T, V, N_A^\circ, N_B^\circ) \qquad , \tag{2}$$

where $\alpha = 0$ if X is an intensive property and $\alpha = 1$ if X is extensive. However, would Eq. 2 hold if we scale-down the system to a regime not belonging to the thermodynamic limit (hereafter, referred to as small- or finite-system), for a example that of $N_A^\circ = N_B^\circ = 1$? Currently accepted dogma assumes the validity of Eq. 2 for all system sizes[2–12], provided sufficient statistics is collected (yet, it is understood that relative magnitudes of fluctuations are inversely proportional to the system size). In this paper we argue that for bimolecular reactions, the homogeneous function character of the system's properties stated in Eq. 2 breaks-down at finite systems.

Results

I. Statistical Mechanical Derivation of the Equilibrium Constant for Association

The process in Eq. 1 is chosen to be described by the canonical (N_A°, N_B°, V, T) ensemble, where $N_A^\circ = N_A + N_{AB}$ and $N_B^\circ = N_B + N_{AB}$ are the total numbers of A and B particles. The particle labels are arranged to satisfy $N_A^\circ \leq N_B^\circ$. All three components on both sides of Eq. 1 are assumed to be gases with ideal behavior, that means, apart from the reaction described they are not interacting

with one another (this also excludes interactions between three or more particles). Upon formation of one bound AB particle, the potential energy of the system changes by an amount of ϵ_{AB} (i.e., ϵ_{AB} is negative). In this model, the energy states of the system due to interparticle interactions can be uniquely mapped onto N_{AB}. Thus, the canonical partition function of the system can be written as,

$$Q = \sum_{i=0}^{N_A^\circ} \frac{q_A^{N_A^\circ - i}}{(N_A^\circ - i)!} \cdot \frac{q_B^{N_B^\circ - i}}{(N_B^\circ - i)!} \cdot \frac{q_{AB}^i}{i!} = \sum_{i=0}^{N_A^\circ} W^i_{N_A^\circ, N_B^\circ}\ q_A^{N_A^\circ - i} q_B^{N_B^\circ - i} q_{AB}^i \qquad , \tag{3}$$

where the summation over index i ($i \equiv N_{AB}$) includes all possible numbers of bound AB particles and thereby all possible energy states. q_A and q_B are single-particle partition functions of unbound A and B particles (hence, obtained by summing only internal energies) and q_{AB} is the pair-particle partition function of one bound AB particle (which includes the Boltzmann factor $e^{-\beta\epsilon_{AB}}$). These partition functions can be expressed in different forms and they are described in details in section SI-1 of the Supporting Information. Given that A and B, each, being indistinguishable particles, $W^i_{N_A^\circ, N_B^\circ}$ in Eq. 3,

$$W^i_{N_A^\circ, N_B^\circ} \equiv \frac{1}{(N_A^\circ - i)!(N_B^\circ - i)!i!} \qquad , \tag{4}$$

corrects the over-counting when raising the partition functions to the power of the particle numbers. Alternatively, the value of $W^i_{N_A^\circ, N_B^\circ}$ can be obtained by first correcting all particles to be indistinguishable, i.e. dividing by $N_A^\circ! N_B^\circ!$, and then multiplying by the number of ways to form from N_A° and N_B° distinguishable particles, i pairs (where the order in each of the formed groups is not important), that is $N_A^\circ! N_B^\circ! / \left[(N_A^\circ - i)!(N_B^\circ - i)!i!\right]$.

The equilibrium constant is defined by,

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

where $\Delta G^\varnothing$ is the standard Gibbs free energy change of the association reaction. This means, $\Delta G^\varnothing$ is the change in Gibbs energy when one mole of A react with one mole of B to form one mole of AB, given all components are at their standard reference conditions. Here we choose

the reference state of component x to be the concentration $c_x^{\varnothing}$ at temperature T. Measurements of $\Delta G^{\varnothing}$ are normally not performed on exactly one mole of particles, N_{Avogadro}, but are scaled to correspond to this number. For generality, we set the volume of the macroscopic reference system to $V^{\varnothing}$ from which the numbers of particles undergoing the association reaction in the reference system can be obtained by $N_x^{\varnothing} = V^{\varnothing} c_x^{\varnothing}$. In writing the partition function of the reference system, $Q^{\varnothing}$, we can still use Eq. 3 but substitute N_A° with $N_A^{\varnothing}$ and N_B° with $N_B^{\varnothing}$. Additionally, because $V^{\varnothing}$ is not equal to V, the single- and pair-particle partition functions in the reference system, $q_x^{\varnothing}$, are different than those in our system, q_x. The dependency of these functions on volume is due to the translational partition function and is therefore linear. Thus the following equality,

$$\frac{q_{AB}^{\varnothing}/V^{\varnothing}}{q_A^{\varnothing}/V^{\varnothing} \cdot q_B^{\varnothing}/V^{\varnothing}} = \frac{q_{AB}/V}{q_A/V \cdot q_B/V} \qquad , \tag{6}$$

between the particle partition functions in the two systems, having the same temperature, exists and will be used below.

We start by expressing the Gibbs free energy change, $\Delta G_{0\to N_A^{\varnothing}}$, when $N_A^{\varnothing}$ particles of A associate with $N_A^{\varnothing}$ (out of $N_B^{\varnothing}$) particles of B, i.e., when all components are at their reference conditions. Then, we will obtain $\Delta G^{\varnothing}$ by scaling $\Delta G_{0\to N_A^{\varnothing}}$ to the stoichiometric number of moles of the reaction. The corresponding change in Helmholtz free energy, $\Delta F_{0\to N_A^{\varnothing}}$, can be calculated from the ratio of the probability to find the system in the bound state, p^{AB} (i.e., the fraction of the state $i = N_A^{\varnothing}$ in the sum of the partition function for the reference system, $Q^{\varnothing}$), to the probability of the unbound state, p^{A+B} (the fraction of the state $i = 0$). Thus, $\Delta G_{0\to N_A^{\varnothing}}$ can be written as,

$$\begin{aligned}\Delta G_{0\to N_A^{\varnothing}} \equiv G_{i=N_A^{\varnothing}} - G_{i=0} &= \Delta F_{0\to N_A^{\varnothing}} + V^{\varnothing}\Delta P_{0\to N_A^{\varnothing}} = -k_B T \ln \frac{p^{AB}}{p^{A+B}} + V^{\varnothing}\Delta P_{0\to N_A^{\varnothing}} \\ &= -k_B T \ln \left[\frac{\left(q_{AB}^{\varnothing}\right)^{N_A^{\varnothing}} \left(q_B^{\varnothing}\right)^{N_B^{\varnothing}-N_A^{\varnothing}}}{N_A^{\varnothing}!(N_B^{\varnothing}-N_A^{\varnothing})!} \frac{N_A^{\varnothing}!N_B^{\varnothing}!}{\left(q_A^{\varnothing}\right)^{N_A^{\varnothing}} \left(q_B^{\varnothing}\right)^{N_B^{\varnothing}}} \right] + V^{\varnothing}\Delta P_{0\to N_A^{\varnothing}} \qquad ,\end{aligned} \tag{7}$$

where $\Delta P_{0\to N_A^{\varnothing}}$ is the change in the pressure of the system accompanied the reaction. Almost without exception, the reference concentrations are chosen to be the same for all components

($c_x^\varnothing = c^\varnothing$ for all x), thus, Eq. 7 reduces to,

$$\Delta G_{0\to N_A^\varnothing} = -N_A^\varnothing k_B T \ln \frac{q_{AB}^\varnothing}{q_A^\varnothing q_B^\varnothing} - k_B T \ln N_A^\varnothing! + V^\varnothing \Delta P_{0\to N_A^\varnothing} \quad . \tag{8}$$

Applying Stirling's approximation to evaluate $\ln N_A^\varnothing!$, thus, requiring $N_A^\varnothing$ to be large, as is always the case for the standard state realized by a macroscopic measurement of $\Delta G^\varnothing$, and subsequently substituting $N_A^\varnothing$ with $V^\varnothing c^\varnothing$ gives,

$$\Delta G_{0\to N_A^\varnothing} = -N_A^\varnothing k_B T \ln \frac{q_{AB}^\varnothing / V^\varnothing}{q_A^\varnothing / V^\varnothing \cdot q_B^\varnothing / V^\varnothing} - N_A k_B T \ln c^\varnothing + N_A^\varnothing k_B T + V^\varnothing \Delta P_{0\to N_A^\varnothing} \quad . \tag{9}$$

Now we will evaluate the ratio inside the first logarithm in Eq. 9. Given the equality in Eq. 6, we can do that using a different system, convenient for us to study, at the same temperature but with arbitrary numbers of particles N_A°, N_B° and volume V, thus at arbitrary concentrations, as long as the ideal behavior of the system is maintained. That means, we chose the system for which the partition function in Eq. 3 was written for. We begin by multiplying and dividing this ratio by the term,

$$\sum_{i=0}^{N_A^\circ - 1} (i+1)\ W_{N_A^\circ, N_B^\circ}^{i+1}\ q_A^{N_A^\circ - i} q_B^{N_B^\circ - i} q_{AB}^i \quad , \tag{10}$$

and obtain,

$$V^\varnothing \frac{q_{AB}^\varnothing}{q_A^\varnothing q_B^\varnothing} = V \frac{q_{AB}}{q_A q_B} = V \frac{\sum_{i=0}^{N_A^\circ - 1} (i+1)\ W_{N_A^\circ, N_B^\circ}^{i+1}\ q_A^{N_A^\circ - (i+1)} q_B^{N_B^\circ - (i+1)} q_{AB}^{i+1}}{\sum_{i=0}^{N_A^\circ - 1} \frac{(i+1)}{[N_A^\circ - (i+1)]!\ [N_B^\circ - (i+1)]!\ (i+1)!}\ q_A^{N_A^\circ - i} q_B^{N_B^\circ - i} q_{AB}^i} \quad . \tag{11}$$

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

$$V \frac{q_{AB}}{q_A q_B} = V \frac{\sum_{j=1}^{N_A^\circ} j\ W_{N_A^\circ, N_B^\circ}^{j}\ q_A^{N_A^\circ - j} q_B^{N_B^\circ - j} q_{AB}^j}{\sum_{i=0}^{N_A^\circ - 1} (N_A^\circ - i)(N_B^\circ - i)\ W_{N_A^\circ, N_B^\circ}^{i}\ q_A^{N_A^\circ - i} q_B^{N_B^\circ - i} q_{AB}^i} \quad . \tag{12}$$

Without changing the value of the sum in the numerator, we can let index j start from zero. The same is true if we let index i in the denominator end at N_A°. This yields,

$$V \frac{q_{AB}}{q_A q_B} = V \frac{\frac{1}{Q}\sum_{j=0}^{N_A^\circ} j\ W_{N_A^\circ, N_B^\circ}^{j}\ q_A^{N_A^\circ - j} q_B^{N_B^\circ - j} q_{AB}^j}{\frac{1}{Q}\sum_{i=0}^{N_A^\circ} (N_A^\circ - i)(N_B^\circ - i)\ W_{N_A^\circ, N_B^\circ}^{i}\ q_A^{N_A^\circ - i} q_B^{N_B^\circ - i} q_{AB}^i} = V \frac{\langle N_{AB} \rangle}{\langle N_A N_B \rangle} = \frac{\langle c_{AB} \rangle}{\langle c_A c_B \rangle} \quad , \tag{13}$$

where the sum in the numerator is identified as the ensemble average of the number of bound particles, $\langle N_{AB} \rangle$, and the sum in the denominator is the average of the product of the numbers of unbound particles, $\langle N_A N_B \rangle$, both in our chosen arbitrary system under equilibrium conditions. Inserting this result into Eq. 9 gives,

$$\Delta G_{0 \to N_A^\varnothing} = -N_A^\varnothing k_B T \ln \frac{\langle c_{AB}/c^\varnothing \rangle}{\langle c_A/c^\varnothing \cdot c_B/c^\varnothing \rangle} + N_A^\varnothing k_B T + V^\varnothing \Delta P_{0 \to N_A^\varnothing} \quad . \tag{14}$$

For ideal gases, the term $V^\varnothing \Delta P_{0 \to N_A^\varnothing}$ equals $-N_A^\varnothing k_B T$, so the last two terms in Eq. 14 cancel each other. In addition, the value of $\Delta G^\varnothing$ is reported per mole of chemical equation,

$$\Delta G^\varnothing = \frac{N_{\text{Avogadro}}}{N_A^\varnothing} \Delta G_{0 \to N_A^\varnothing} \quad . \tag{15}$$

Considering the definition of K in Eq. 5 we obtain,

$$K = \frac{\langle c_{AB} \rangle}{\langle c_A c_B \rangle} \cdot c^\varnothing = \frac{\langle P_{AB} \rangle}{\langle P_A P_B \rangle} \cdot P^\varnothing \quad , \tag{16}$$

stating the equilibrium constant of binding reactions must include cross correlations in the reactants' concentrations. The second equality relates K to the corresponding ratio of the partial pressures of the different components where $P^\varnothing = c^\varnothing k_B T$ is the standard reference pressure.

Notice that during the entire derivation there were no conditions imposed specifying a finite system. In fact the definition of K implies a reference system with stoichiometric numbers of moles of particles, justifying the application of Stirling's approximation. It is only for convenience that we might use a small system to evaluate the ratio of the partition functions (of single- and pair-particle natures) encountered in Eq. 9. Yet it is in particular then, that the resulting equilibrium constant expressed in Eq. 16, is substantially different than an analogous expression neglecting correlations, K',

$$K' = \frac{\langle c_{AB} \rangle}{\langle c_A \rangle \langle c_B \rangle} \cdot c^\varnothing \quad . \tag{17}$$

It is also essential to note that in statistical mechanics textbooks[13–15], the equilibrium constant is derived using an ensemble at constant N_A, N_B, N_{AB}, V, T, where the numbers of particles are identified as those at equilibrium upon imposing the macroscopic condition of chemical equilibrium.

Fixing the numbers of particles of all components in the system, and inevitably their corresponding chemical potentials (conjugated parameters), render the description of chemical equilibrium of the reaction macroscopic. Not surprisingly, the resulting equilibrium constant is obtained in its thermodynamic form, $K = c_{AB}c^{\varnothing}/(c_A c_B)$, even if along the derivation statistical mechanics relations were applied.

II. Computational Validation

To test our derivation we constructed a simple system of Lennard-Jones A and B molecules able to establish the equilibrium binding reaction of Eq. 1. Three out of the four parameters specifying the system in the canonical ensemble, N_A°, N_B°, and V, were changed systematically at constant temperature, producing three different series of simulations, R1, R2, and R3. The first two series were subject to three different simulation methods; Monte-Carlo (MC), molecular dynamics with Nosé-Hoover thermostat (MD-NH), and molecular dynamics with velocity-rescaling thermostat (MD-VR). The third series of simulations (R3) was conducted only by MC. Section SI-2 provides further details on the model system and computational methodologies.

Figure 1 displays the equilibrium constant, K, calculated by Eq. 16, as well as, the value of K' defined in Eq. 17. Clearly, inclusion of cross-correlations in the reactants' concentrations are crucial for the equilibrium constant to stay constant at finite systems. In contrast, K' depends on the numbers of particles and/or volume of the system studied, where its deviation from K increases with decreasing the size of the system. For systems large enough, where correlations become negligible, K' approaches K. The fact that K is constant for all sizes of the system, even for the smallest system possible, indicates the law of mass action[16] holds not only for macroscopic- but for finite-systems as well, contrary to arguments found in the literature[17,18]. In Section SI-1 we consider even a simpler model system, of single-site reactants, to facilitate an easy comparison between the value of K obtained by Eq. 16 and analytical/numerical calculations. Excellent agreements, with all three simulations methods, are attained.

As might be expected, the extent of divergence of K' from K is also a function of temperature. To demonstrate this, we conducted additional simulations of R1 series at different temperatures. Figure 2, as well as Fig. SI-3.1 in Section SI-3, indicate this divergence of K' increases with decreasing temperature (or with increasing $-\epsilon_{AB}/k_{B}T$). For example, for $N_A^\circ = N_B^\circ = 1$, K' is larger than K by a factor of 300 at $T = 200\ K$, whereas it is nearly equal to K at $T = 1200\ K$.

A direct consequence of taking the average of the product of reactants' concentrations in calculating K, and not the product of their averages, is on the condition for equilibrium. Using the relation between the chemical potentials of component x at c_x and at $c^\varnothing$ at the same temperature, $\mu_x = \mu_x^\varnothing + RT\ln(c_x/c^\varnothing)$, and identifying $\Delta G^\varnothing$ with $\mu_{AB}^\varnothing - \mu_A^\varnothing - \mu_B^\varnothing$, it follows from Eq. 16 that the condition for equilibrium is,

$$\langle \mu_{AB} \rangle - \langle \mu_A + \mu_B \rangle = 0 \qquad , \tag{18}$$

and not that expressed by the stoichiometric sum of the average of each component,

$$\langle \mu_{AB} \rangle - \langle \mu_A \rangle - \langle \mu_B \rangle = 0 \qquad , \tag{19}$$

unless the system is large enough to render the correlations negligible.

The statistical thermodynamics expression of the equilibrium constant (Eq. 16) can also be rationalized from dynamics. At equilibrium, the average (over replica or over time) net change in the product's and reactants' concentrations is zero, thus we have,

$$\left\langle \frac{dc_{AB}}{dt} \right\rangle = \left\langle -\frac{dc_A}{dt} \right\rangle = \left\langle -\frac{dc_B}{dt} \right\rangle = \langle k_{fw} c_A c_B - k_{bw} c_{AB} \rangle = 0 \qquad , \tag{20}$$

where k_{fw} and k_{bw} are the rate constants of the forward and backward reactions, respectively. The backward reaction is a simple unimolecular process, while the forward reaction is a bimolecular process and its rate is proportional to the collision probability between A and B. In turn, this collision probability at each point in time is proportional to the product of the corresponding instantaneous concentrations. That is, averaging the rate of the forward reaction in finite systems requires the cross-correlations of the two reactants' concentrations. By defining K as the ratio

between forward and backward rate constants, and rendering its value dimensionless via $c^{\varnothing}$, we recover Eq. 16. We calculated k_{fw} and k_{bw} from the MD simulations (Section SI-2) and the results corroborate k_{fw} at finite systems must include correlations between $c_{_A}$ and $c_{_B}$ (Fig. 3). Clearly, ignoring these correlations will produce rate constants that depend on concentrations (Fig. 3b) as evidenced when analyzing single-molecule fluorescence binding experiments[19].

In constructing R1 series of simulations, we multiplied all extensive parameters specifying the system by the same factor, exactly as described by Eq. 2. This means, intensive properties are expected to have the same average values for all system sizes if the system's properties were homogeneous in character. However, Fig. 4 demonstrates this is not the case at finite systems. In particular, the concentration of the bound state, as well as the inter-particle energy per particle, exhibit rising divergence from a horizontal line as the number of particles decreases. We also plot the radial distribution function between a and b sites. Again Eq. 2 predicts overlapping curves for all system sizes, however, different distributions are obtained where the maxima describing the bound state for small-sized systems are higher in accordance with their larger concentrations. It is worth mentioning, these changes in average properties at finite systems are not emerging from artefacts due to neglected concentration fluctuations in small simulations[20] or application of periodic boundary conditions in finite simulation boxes[17,21–23], but are a consequence of incompatibility between two-body interactions and linear scaling.

III. Calculating Concentrations from Fluctuations

It is well known that fluctuations are related to susceptibilities. In our system, the incessant transitions at equilibrium between reactants and product force the number of particles of each component to fluctuate. We now show that the composition of the system (particle numbers, or concentrations) can be determined only from the magnitudes of these fluctuations.

Adopting the notation of Lebowitz et al.[24], we define the cross fluctuations between quantities

ζ and η as,

$$L(\zeta,\eta) = \langle \zeta\eta \rangle - \langle \zeta \rangle \langle \eta \rangle \qquad , \tag{21}$$

and their relative magnitude by,

$$l(\zeta,\eta) = \frac{L(\zeta,\eta)}{\langle \zeta \rangle \langle \eta \rangle} \qquad . \tag{22}$$

We now look at the following difference,

$$l(N_{AB},N_{AB}) - l(N_{AB},N_A N_B) = \frac{1}{\langle N_{AB} \rangle} \left[\frac{\langle N_{AB}^2 \rangle}{\langle N_{AB} \rangle} - \frac{\langle N_{AB} N_A N_B \rangle}{\langle N_A N_B \rangle} \right] \qquad , \tag{23}$$

and concentrate on evaluating the term inside the square brackets. We start by evaluating the term,

$$\frac{\langle N_{AB}^2 \rangle}{\langle N_{AB} \rangle} = \frac{\frac{1}{Q} \sum_{i=0}^{N_A^\circ} \; i^2 \; W^i_{N_A^\circ,N_B^\circ} \; q_A^{N_A^\circ - i} q_B^{N_B^\circ - i} q_{AB}^i}{\frac{1}{Q} \sum_{i=0}^{N_A^\circ} \; i \; W^i_{N_A^\circ,N_B^\circ} \; q_A^{N_A^\circ - i} q_B^{N_B^\circ - i} q_{AB}^i} = \frac{\sum_{j=0}^{N_A^\circ - 1} \; (j+1)^2 \; W^{j+1}_{N_A^\circ,N_B^\circ} \; q_A^{N_A^\circ - j} q_B^{N_B^\circ - j} q_{AB}^j}{\sum_{j=0}^{N_A^\circ - 1} \; (j+1) \; W^{j+1}_{N_A^\circ,N_B^\circ} \; q_A^{N_A^\circ - j} q_B^{N_B^\circ - j} q_{AB}^j} \qquad , \tag{24}$$

where we skipped the terms corresponding to $i = 0$ and changed the index of the summation to $j = i - 1$. In the second equality we also multiplied and divided the ratio by $q_A q_B / q_{AB}$. Similarly, we can express the second term inside the square brackets in Eq. 23 by,

$$\begin{aligned} \frac{\langle N_{AB} N_A N_B \rangle}{\langle N_A N_B \rangle} &= \frac{\frac{1}{Q} \sum_{i=0}^{N_A^\circ} \; i(N_A^\circ - i)(N_B^\circ - i) \; W^i_{N_A^\circ,N_B^\circ} \; q_A^{N_A^\circ - i} q_B^{N_B^\circ - i} q_{AB}^i}{\frac{1}{Q} \sum_{i=0}^{N_A^\circ} \; \frac{(N_A^\circ - i)(N_B^\circ - i)}{(N_A^\circ - i)!(N_B^\circ - i)! i!} \; q_A^{N_A^\circ - i} q_B^{N_B^\circ - i} q_{AB}^i} \\ &= \frac{\sum_{i=0}^{N_A^\circ - 1} \; i(i+1) \; W^{i+1}_{N_A^\circ,N_B^\circ} \; q_A^{N_A^\circ - i} q_B^{N_B^\circ - i} q_{AB}^i}{\sum_{i=0}^{N_A^\circ - 1} \; (i+1) \; W^{i+1}_{N_A^\circ,N_B^\circ} \; q_A^{N_A^\circ - i} q_B^{N_B^\circ - i} q_{AB}^i} \qquad . \end{aligned} \tag{25}$$

The second equality is realized by letting index i in the sum end at $N_A^\circ - 1$ and rewriting the coefficients of the single/pair-particle partition functions in terms of $W^{i+1}_{N_A^\circ,N_B^\circ}$. Note the denominators of Eq. 24 and Eq. 25 are the same, so the difference of the two terms inside the square brackets in Eq. 23 can be easily evaluated,

$$\frac{\langle N_{AB}^2 \rangle}{\langle N_{AB} \rangle} - \frac{\langle N_{AB} N_A N_B \rangle}{\langle N_A N_B \rangle} = \frac{\sum_{i=0}^{N_A^\circ - 1} \; (i+1) \; W^{i+1}_{N_A^\circ,N_B^\circ} \; q_A^{N_A^\circ - i} q_B^{N_B^\circ - i} q_{AB}^i}{\sum_{i=0}^{N_A^\circ - 1} \; (i+1) \; W^{i+1}_{N_A^\circ,N_B^\circ} \; q_A^{N_A^\circ - i} q_B^{N_B^\circ - i} q_{AB}^i} = 1 \qquad , \tag{26}$$

which actually reduces to one. This means Eq. 23 becomes,

$$l(N_{AB}, N_{AB}) - l(N_{AB}, N_A N_B) = \frac{1}{\langle N_{AB} \rangle} \quad , \tag{27}$$

from which the average concentration of the bound AB particles can be expressed by,

$$\langle c_{AB} \rangle = \frac{1}{[l(N_{AB}, N_{AB}) - l(N_{AB}, N_A N_B)]\, V} \quad . \tag{28}$$

Equation 27 can also be derived by a more conventional procedure, i.e. by partially differentiating the partition function of the system with respect to temperature. However in this case we need to assume K is given by Eq. 16 (see Section SI-4).

It is interesting to comment that whereas $l(N_{AB}, N_{AB})$ is necessarily positive, the relative fluctuations in $l(N_{AB}, N_A N_B)$ measure correlations between two quantities that are anti-correlated and hence always negative. Thus, the quantity inside the square brackets of the denominator in Eq. 28 is a summation of two positive terms with magnitude that reduces with increasing system size. For large systems, this reduction is proportional to the reciprocal of the volume so that $\langle c_{AB} \rangle$ approaches a constant.

The relation in Eq. 27 was tested for all simulations performed. In Fig. 5a we plot results at $T = 300\ K$, and in Fig. SI-3.2 results of the R1 series at different temperatures. All data points, independent of temperature, fall on the same straight line as predicted. The correlation coefficients of the linear regressions turned-out perfect, within the accuracy of the analyzing software, likely because comparison is made between two quantities calculated from the same simulation allowing elimination of certain errors.

IV. Calculating Concentrations from K

A drawback of Eq. 27 or Eq. 28 is when the system simulated or studied experimentally is not of the same size as the target system. In this case, relative fluctuations of the target system are needed in order to compute composition. Thus it would be more practical if we can determine the concentrations from K and the parameters defining the target system.

We start by rewriting the expression of K,

$$\begin{aligned}\langle N_{AB}\rangle &= \frac{K}{Vc^{\varnothing}}\langle N_A N_B\rangle = \frac{K}{Vc^{\varnothing}}\langle (N_A^{\circ}-N_{AB})(N_B^{\circ}-N_{AB})\rangle \\ &= \frac{K}{Vc^{\varnothing}}\left[N_A^{\circ}N_B^{\circ}-(N_A^{\circ}+N_B^{\circ})\langle N_{AB}\rangle + L(N_{AB},N_{AB}) + \langle N_{AB}\rangle^2\right] \quad , \end{aligned} \tag{29}$$

and solve the quadratic equation to obtain,

$$\langle c_{AB}\rangle = \frac{\left(c_A^{\circ}+c_B^{\circ}+\frac{c^{\varnothing}}{K}\right) - \sqrt{\left(c_A^{\circ}+c_B^{\circ}+\frac{c^{\varnothing}}{K}\right)^2 - 4\left[l(N_{AB},N_{AB})+1\right]c_A^{\circ}c_B^{\circ}}}{2\left[l(N_{AB},N_{AB})+1\right]} \quad . \tag{30}$$

If we performed simulations at a finite size and wish to know the concentrations in the thermodynamic limit ($V, N_A^{\circ} \to \infty$), we simply set $l(N_{AB},N_{AB}) = 0$ and recover general chemistry textbooks' result[25],

$$\langle c_{AB}\rangle_{\infty} = \frac{1}{2}\left[c_A^{\circ}+c_B^{\circ}+\frac{c^{\varnothing}}{K} - \sqrt{\left(c_B^{\circ}-c_A^{\circ}\right)^2 + \frac{2c^{\varnothing}\left(c_A^{\circ}+c_B^{\circ}\right)}{K} + \frac{c^{\varnothing 2}}{K^2}}\right] \quad , \tag{31}$$

and because $L(N_{AB},N_{AB}) = L(N_A,N_B)$, we can substitute the correlated reactants' concentrations appearing in the expression for K with the uncorrelated concentrations, i.e. $K' \longrightarrow K$.

Another important case, especially for simulation studies and single-molecule experiments, is that of $N_A^{\circ} = 1$ (where $N_B^{\circ} \geq N_A^{\circ}$). Here it is easy to show the relation, $\langle N_A N_B\rangle = N_B^{\circ}(N_A^{\circ} - \langle N_{AB}\rangle)$, is satisfied which leads to,

$$\langle c_{AB}\rangle_{N_A^{\circ}=1} = \frac{N_A^{\circ}N_B^{\circ}K}{V(Vc^{\varnothing}+N_B^{\circ}K)} \quad . \tag{32}$$

In addition, the relative fluctuations obey,

$$l(N_{AB},N_{AB})_{N_A^{\circ}=1} = \frac{Vc^{\varnothing}}{KN_B^{\circ}} \quad , \tag{33}$$

by noting that in this case ($N_A^{\circ} = 1$), $l(N_{AB},N_AN_B) = -1$ and $\langle N_{AB}^2\rangle = \langle N_{AB}\rangle$. R2 and R3 series of simulations fall within this special case, therefore, in Fig. 5b and Fig. 5c we predict $\langle c_{AB}\rangle$ as stated by Eq. 32, and in Fig. SI-3.3 we predict $l(N_{AB},N_{AB})$ according to Eq. 33. In both cases, the agreement is almost perfect.

In contrast to the thermodynamic limit and systems with $N_A^\circ = 1$, predicting $\langle c_{_{AB}} \rangle$ from the value of K for other finite systems, thus with $N_A^\circ > 1$, is less simple because of the difficulty of predicting $l(N_{AB}, N_{AB})$. Obviously, the magnitude of $l(N_{AB}, N_{AB})$ for $N_A^\circ > 1$ must be smaller than that for $N_A^\circ = 1$. A plausible guess can be that it is inversely proportional to the system size. We therefore express $l(N_{AB}, N_{AB})$ for $N_A^\circ > 1$ by scaling the corresponding value at $N_A^\circ = 1$ according to,

$$l(N_{AB}, N_{AB})_{_{N_A^\circ \geq 1}} = l(N_{AB}, N_{AB})_{_{N_A^\circ = 1}} \cdot \frac{1}{(N_A^\circ)^\lambda} = \frac{Vc^\varnothing}{KN_B^\circ} \cdot \frac{1}{(N_A^\circ)^\lambda} \quad , \quad \text{where} \quad 0 \leq \lambda \leq 1 \quad . \tag{34}$$

When $\lambda = 0$, Eq. 34 reduces to Eq. 33, whereas for the thermodynamic limit it turns out from the simulations that $\lambda = 1$. Empirically we find λ can be approximated by,

$$\lambda \simeq \frac{1}{1 + K/(Vc^\varnothing \ln N_B^\circ)} \qquad . \tag{35}$$

This approximation is investigated in Fig. SI-3.4 for the R1 series of simulations. Very good agreement with simulation data is obtained where the accuracy of the prediction increases with temperature. Armed with the ability to estimate $l(N_{AB}, N_{AB})$, we proceed to predict the concentrations via Eq. 30 in Fig. SI-3.5. Although not perfect at lower temperatures, the approximation yields satisfactory agreement with concentrations observed in the simulations. Due to the asymmetric roles of N_A° and N_B° in Eq. 34, we examined the approximation also on another series of simulations, R4, in which N_A° and N_B° are not equal (Section SI-2). Here the accuracy of the prediction, shown in Fig. SI-3.6, is even better. Moreover, we scale $g_{ab}(r)$ obtained at finite systems to the corresponding distribution of a macroscopic system, as shown in Fig. SI-5.1 and discussed in Section SI-5.

Discussion

An Example

We now exemplify and discuss the calculation of K for the association reaction in Eq. 1 for the smallest system possible, $N_A^\circ = N_B^\circ = 1$. In this case, there are only two possible macroscopic states in the system, one corresponding to a bound AB particle and the other to unbound $A + B$ particles. Suppose the fraction of independent configurations in which the bound state is observed is f^{AB} (thus, the fraction of the unbound state is $f^{A+B} = 1 - f^{AB}$). Applying the expression of K with uncorrelated reactants' concentrations defined in Eq. 17 yields,

$$K' = \frac{f^{AB}/V}{\left[\left(1 - f^{AB}\right)/V\right]^2} \cdot c^{\varnothing} \qquad . \tag{36}$$

Although this is currently the most employed expression in the literature[3–12], it provides erroneous results at finite systems as demonstrated throughout the manuscript. This is because correlations in reactants' concentrations, that must be taken into account, are augmented as the system size is decreased. Yet, for this system it is possible to calculate K from the ratio of f^{AB} to f^{A+B}. However application of the plain ratio,

$$K'' = \frac{f^{AB}}{1 - f^{AB}} \qquad , \tag{37}$$

which equal the ratio of probabilities to find this particular system in the bound and unbound states, *does not* correspond to K. The reason is that this ratio is size-dependent. There are many more possible microstates for the unbound state than for the bound state, and scalings with system-size follow different power-laws for the two states. In our derivation (Eq. 7), this is expressed in the corresponding translational partition functions; the number of possible states is proportional to the volume for the bound particles whereas it is proportional to the square of the volume for the unbound particles. It is only when q_A, q_B, and q_{AB} are, each, divided by V that the ratio becomes size-independent as argued in Eq. 6. Thus normalizations of the probabilities (or number of configurations) in Eq. 37 by a factor of V and V^2, to obtain probability densities, are

necessary, albeit the introduction of a dimension of volume to the ratio. It is therefore for these cases, i.e. when the number of particles on the reactant side is not equal to that on the product side, that a reference to a standard system is necessary to render K dimensionless. In Eq. 9 the standard concentration, $c^{\varnothing}$, emerged from the term $N_A^{\varnothing}!$ that was not canceled-out in the ratio of probabilities of the two states (in Eq. 7). Dividing the probabilities in Eq. 37 by the normalization factors and eliminating the dimension of the ratio by $c^{\varnothing}$ gives,

$$K = \frac{f^{AB}\, V}{1 - f^{AB}} \cdot c^{\varnothing} \qquad , \tag{38}$$

an expression identical to that obtained had we used Eq. 16. Obviously this simple direct counting of configurations of the two states to obtain K, or even just the free energy difference ΔG for the studied system[26], can only work for $N_A^{\circ} = N_B^{\circ} = 1$. The reason is, in this case $\langle N_A N_B \rangle = \langle N_A \rangle = \langle N_B \rangle$ and the value of the term $W^i_{N_A^{\circ}, N_B^{\circ}}$, defined in Eq. 4, is the same for the different macroscopic states.

The Difference with Unimolecular Processes

It is important to emphasize the arguments presented in this paper are pertinent to bimolecular reactions or two-body properties. Consider the chemical equation,

$$A + B \rightleftharpoons 2C \qquad , \tag{39}$$

representing, for example, the recombination of hydronium and hydroxide ions to form two water molecules. As it is a bimolecular process, the expression of the equilibrium constant is,

$$K = \frac{\langle c_C^2 \rangle}{\langle c_A c_B \rangle} \qquad . \tag{40}$$

Another process that can also be represented by exactly the same chemical equation is, for example, the transitions between different conformations of a peptide, where A, B, and C denote α-helix, β-sheet, and random-coil. In this case, Eq. 39 is actually a sum of two chemical equations ($A \rightleftharpoons C$ and $B \rightleftharpoons C$) in which α-helix and β-sheet, separately, form equilibrium with coil conformation.

The transitions between the different conformations are unimolecular in nature and the expression of K in Eq. 40 is not appropriate. As no correlations exist between the α-helix and β-sheet conformations, the equilibrium constant should be computed by,

$$K = \frac{\langle c_C \rangle^2}{\langle c_A \rangle \langle c_B \rangle} \quad , \tag{41}$$

that is, the product of the equilibrium constants of the two unimolecular reactions. The outcome of these two examples contradicts the principle upon which chemists view chemical equilibrium, that is, K is dictated only by the chemical equation of the reaction, irrespective of its nature. This is indeed true as long as the system is macroscopic or large enough. However at finite systems, the expression of the equilibrium constant of two reactions with the same chemical equation can be different. The distinction emerges because different averaging applies for calculating K depending on the order of the elementary process(es) involved, as is the case when determining rate constants.

Magnitude of the Two-Body Correlations

As pointed-out above, the magnitude of correlations between the reactants, which can be represented also by the deviation of the ratio K'/K from 1, is influenced by temperature (or by the 'reduced' temperature, $k_B T/\epsilon_{AB}$). Figure 2 shows this clearly, yet it indicates the correlations are affected as well by the numbers of particles and/or volume, because in R1 series when $N_A^\circ = N_B^\circ$ increases, V increases by the same factor to keep $c_A^\circ = c_B^\circ$ constant. In R2 series, the volume is the only parameter changing and from Fig. 1b it is evident it affects the magnitude of correlations. Given the well-known expression of fluctuations in the number of particles in the grand-canonical ensemble, it is tempting to assume the correlations in our system would decay inversely with the numbers of the particles. In Fig. SI-3.7a we plot K'/K for R4 series where all simulations had the same T, V, and N_B° and only N_A° was increased from 1 to 8. However, this increase in the value of N_A° did not have an effect on the ratio of K'/K. In contrast, the value of N_B° does influence the correlations. This can be seen in Fig. SI-3.7b where we compare R1 and R2 series. In both series, $N_A^\circ = N_B^\circ$, however in R2 these numbers equal 1 for all simulations whereas in R1 they vary. The

curves, plotted as a function of V, indicate the value of K'/K at fixed V is lower for R1 where the numbers of particles are larger, however, the decay is much smaller (less strong) than $1/N_B^\circ$.

Conclusions

In this paper we demonstrated that equilibrium constants, as well as rate constants, of binding reactions at finite systems must include correlations in reactants' concentrations. That being the case, equilibrium is achieved when the average chemical potential of the bound product is equal to the average of the sum, and not to the sum of the averages, of the chemical potentials of the unbound reactants. This point has never been considered in the literature before, likely because the working assumption followed an outcome presented in statistical mechanics textbooks based on an ensemble, claimed here inappropriate, which leads only to the well-known expression applicable for macroscopic systems. Instead, a different derivation is offered in which the constructed ensemble fixes only the total number of each particle-type in the system. This allows the numbers of reactants and product(s) of the reaction to experience fluctuations, with magnitude dictated by the parameters specifying the system. Accordingly, the resulting expression of K provides information on how to perform averaging over the ensemble utilized. A key step in the derivation is the evaluation of the ratio $Vq_{AB}/(q_Aq_B)$. By applying a sequence of algebraic operations, we showed this ratio to be equal to $\langle c_{AB}\rangle/\langle c_Ac_B\rangle$, where the brackets indicate ensemble average under equilibrium conditions. This inclusion of correlations in calculating the equilibrium constant, can produce values that differ by few orders of magnitude compared with those neglecting them. Because correlations become less important with increasing system size, for macroscopic systems, the statistical mechanical expression of K reduces to that obtained from thermodynamics.

Conflicts of interest

There are no conflicts of interest to declare

Acknowledgments

I would like to thank Zohar Nussinov for stimulating and insightful discussions. 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.

Figures

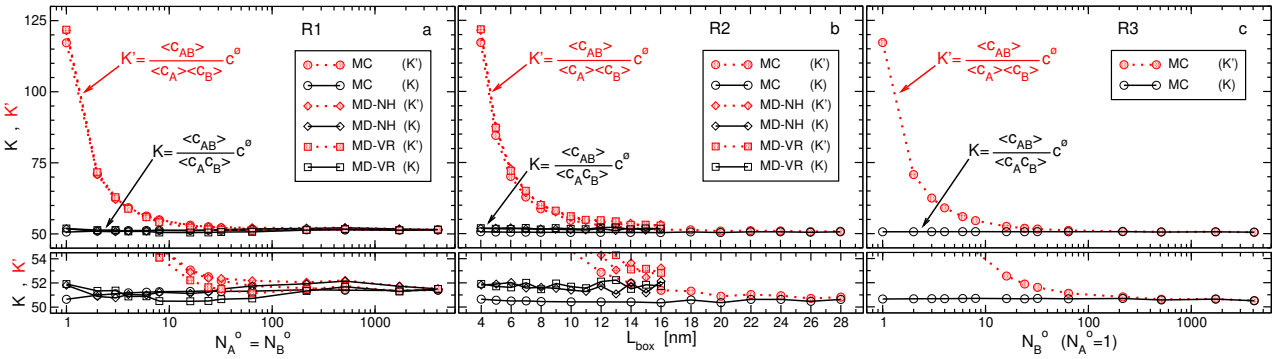

Figure 1: The equilibrium constant K defined by Eq. 16 ($c^{\varnothing} \equiv 1\ M$) for three series of simulations at: (a) constant $c_A^{\circ} = c_B^{\circ} = 0.026\ M$ (R1), (b) constant $N_A^{\circ} = N_B^{\circ} = 1$ (R2), and (c) constant $N_A^{\circ} = 1$ and $c_B^{\circ} = 0.026\ M$ (R3), all performed in the canonical ensemble at $T = 300\ K$. The value of K' defined by Eq. 17 is shown in red for comparison. The lower panels are magnified plots around the value of K. The simulations were performed by three methods: Monte-Carlo (MC), molecular-dynamics with a Nosé-Hoover thermostat (MD-NH), and molecular-dynamics with a Velocity-Rescaling thermostat (MD-VR). The left-most point in all series represents the same system ($N_A^{\circ} = N_B^{\circ} = 1$, $L_{box} = 4\ nm$). The estimated errors for the values of K are smaller or about the size of the symbols. Results from simulations at lower and higher temperatures are shown in Fig. SI-3.1 in the Supporting Information.

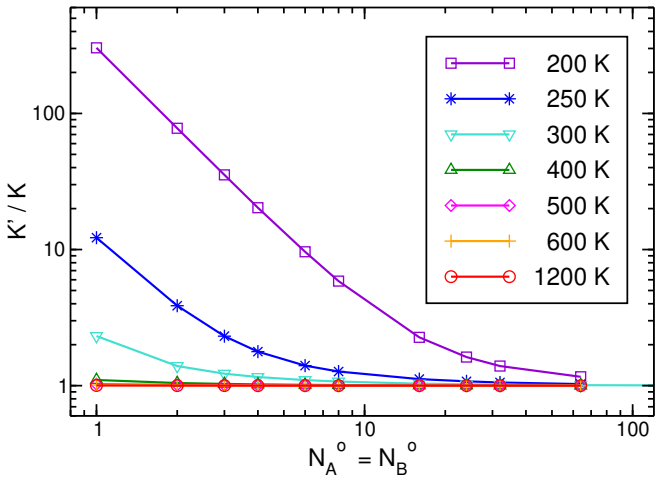

Figure 2: The ratio of the equilibrium constant in which correlations in the reactants concentrations are ignored to that in which they are accounted for, K'/K, from MC R1 series of simulations at different temperatures.

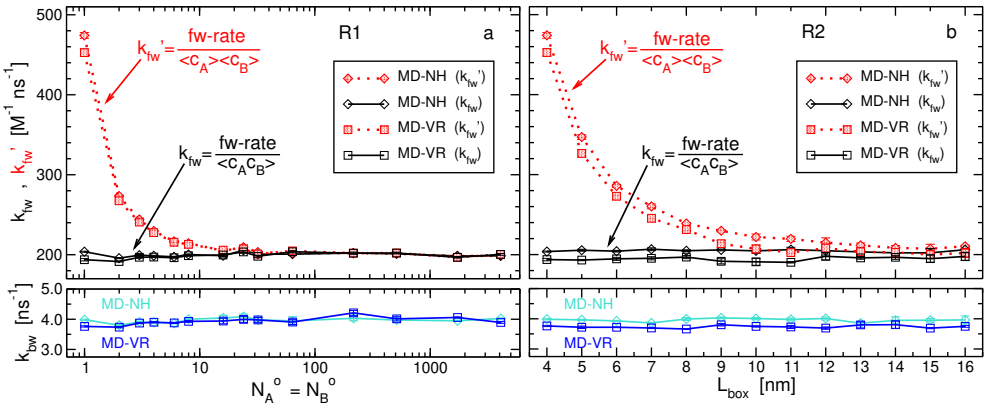

Figure 3: The rate constants of the binding reaction for simulation series R1 (a) and R2 (b) obtained from molecular dynamics simulations. The top panels show the rate constant in the forward direction, k_{fw}, whereas the lower panels the rate constant in the backward direction, k_{bw}. For comparison we present also k'_{fw} calculated by uncorrelated reactants' concentrations.

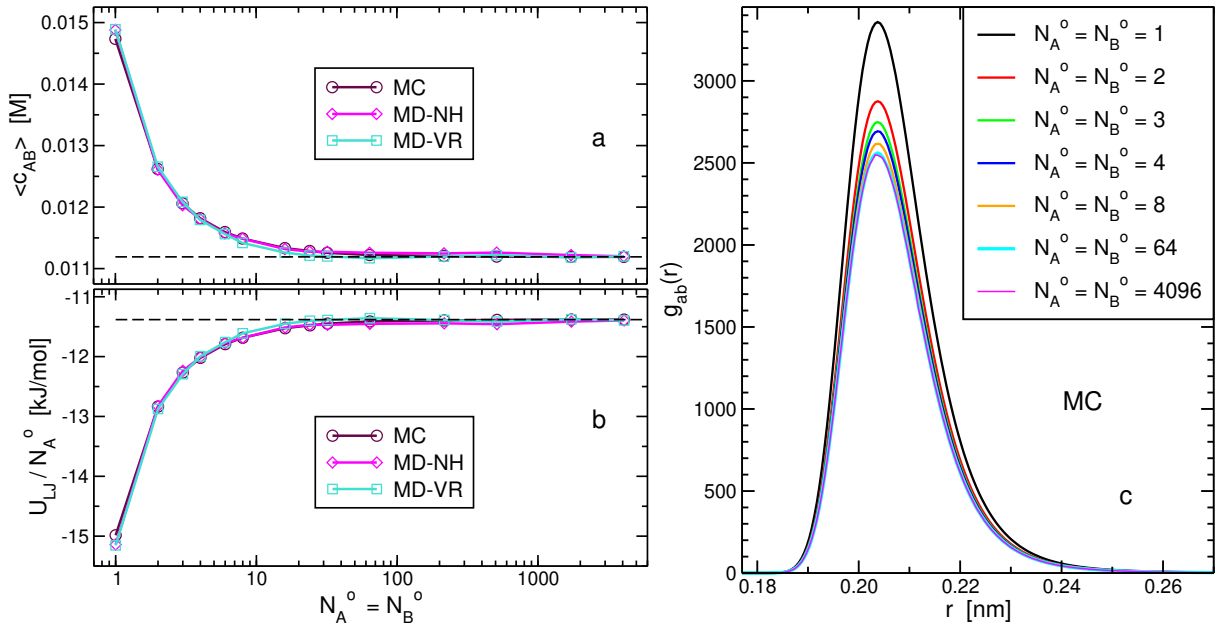

Figure 4: Results exhibiting the inhomogeneous character of properties of bimolecular reactions upon scaling-down system size. The analyzes were performed on R1, i.e. the series of simulations generated by scaling all extensive parameters specifying the system ($N_A^\circ = N_B^\circ, V$) by the same factor. (a) The concentration of bound molecules, $\langle c_{AB} \rangle$, (b) the inter-particle energy per particle, and (c) the radial distribution function between a and b sites for different system sizes. (a) and (b) are almost perfect mirror-image of each other, and the estimated errors are smaller than the size of the symbols. In (c), only results from MC simulations are shown, however, very similar figures are obtained for MD-NH and MD-VR. If average quantities of the system were homogeneous functions, the data points in (a) and (b) would follow the horizontal dashed line, and the pair-distribution functions in (c) would collapse on the curve of the largest system.

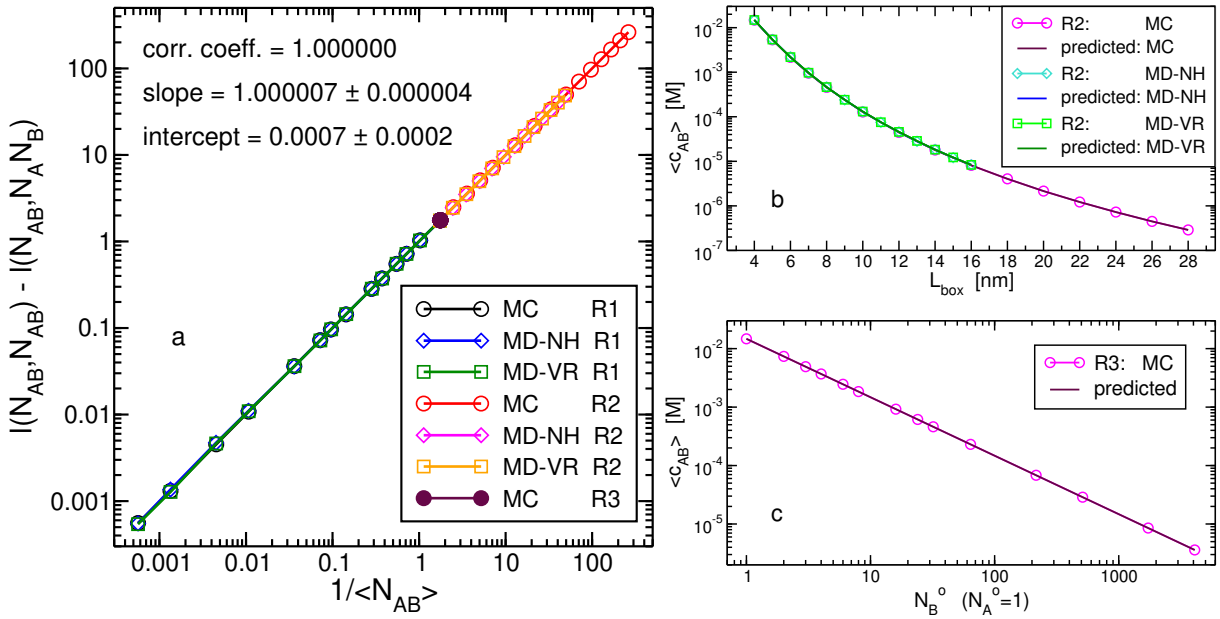

Figure 5: (a) A relation between two relative fluctuations and the reciprocal of average number of bound particles. All simulations results (here displayed for $T = 300\ K$, for other temperatures see Fig. SI-3.2) fall on a linear line crossing the origin with a slope of one as described in Eq. 27. Results obtained from linear regression (using xmgrace) of all data points are indicated. All points of R3 have the same x, y values. (b) The concentration of bound particles as a function of box length for R2 series. The results obtained from simulations are shown along predictions based on the value of the equilibrium constant (Eq. 32). (c) Same as (b) but for R3 series of simulations, in which case, the concentration is plotted as a function of the total number of B particles.

References

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

[2] M. K. Gilson, J. A. Given, B. L. Bush and J. A. McCammon, *Biophys. J.*, 1997, **72**, 1047–1069.

[3] P. H. Hünenberger, J. K. Granwehr, J.-N. Aebischer, N. Ghoneim, E. Haselbach and W. F. van Gunsteren, *J. Am. Chem. Soc.*, 1997, **119**, 7533–7544.

[4] H. Luo and K. Sharp, *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 10399–10404.

[5] Y. Zhang and J. A. McCammon, *J. Chem. Phys.*, 2003, **118**, 1821–1827.

[6] E. Psachoulia, P. W. Fowler, P. J. Bond and M. S. P. Sansom, *Biochemistry*, 2008, **47**, 10503–10512.

[7] Y. Deng and B. Roux, *J. Phys. Chem. B*, 2009, **113**, 2234–2246.

[8] R. Skorpa, J.-M. Simon, D. Bedeaux and S. Kjelstrup, *Phys. Chem. Chem. Phys.*, 2014, **16**, 1227–1237.

[9] M. De Vivo, M. Masetti, G. Bottegoni and A. Cavalli, *J. Med. Chem.*, 2016, **59**, 4035–4061.

[10] J. J. Montalvo-Acosta and M. Cecchini, *Mol. Inform.*, 2016, **35**, 555–567.

[11] L. A. Patel and J. T. Kindt, *J. Chem. Theory Comput.*, 2017, **13**, 1023–1033.

[12] E. Duboué-Dijon and J. Hénin, *J. Chem. Phys.*, 2021, **154**, 204101.

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

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

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

[16] P. Waage and C. M. Guldberg, *Forhandlinger i Videnskabs-selskabet i Christiania*, 1864, 35–45.

[17] D. H. De Jong, L. V. Schäfer, A. H. De Vries, S. J. Marrink, H. J. C. Berendsen and H. Grubmüller, *J. Comp. Chem.*, 2011, **32**, 1919–1928.

[18] X. Zhang, L. A. Patel, O. Beckwith, R. Schneider, C. J. Weeden and J. T. Kindt, *J. Chem. Theory Comput.*, 2017, **13**, 5195–5206.

[19] J. Yang and J. E. Pearson, *J. Chem. Phys.*, 2012, **136**, 244506.

[20] T. E. Ouldridge, A. A. Louis and J. P. K. Doye, *J. Phys.: Condens. Matter*, 2010, **22**, 104102.

[21] R. Cortes-Huerto, K. Kremer and R. Potestio, *J. Chem. Phys.*, 2016, **145**, 141103.

[22] N. Dawass, P. Krüger, S. K. Schnell, D. Bedeaux, S. Kjelstrup, J. M. Simon and T. J. H. Vlugt, *Mol. Sim.*, 2018, **44**, 599–612.

[23] R. Hall, T. Dixon and A. Dickson, *Front. Mol. Biosci.*, 2020, **7**, 106.

[24] J. L. Lebowitz, J. K. Percus and L. Verlet, *Phys. Rev.*, 1967, **153**, 250–254.

[25] D. W. Oxtoby and N. H. Nachtrieb, *Principles of Modern Chemistry*, Saunders College Publishing, Orlando, FL, 3rd edn, 1996.

[26] W. F. van Gunsteren, X. Daura and A. Mark, *Helv. Chim. Acta*, 2002, **85**, 3113–3129.

A Table of Contents Entry