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*

December 8, 2021

[*]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 system's properties of binding reactions are not homogeneous functions. That means averages of intensive properties, such as the concentration of the bound-state, at finite-systems are different than those at large-systems. The discrepancy increases with decreasing numbers of particles, temperature, and volume. As perplexing as it may sound, despite these 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 of the system only from the value of the equilibrium constant. All predictions are validated by Monte-Carlo and molecular dynamics simulations. An important significance of these findings is that the expression of the equilibrium constant at finite systems is not dictated solely by the chemical equation 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[2–15], 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

We start by deriving the expression of the equilibrium constant in the canonical ensemble. The total numbers of A and B particles, $N_A^{\circ} = N_A + N_{AB}$ and $N_B^{\circ} = N_B + N_{AB}$, are fixed and labeled to satisfy $N_A^{\circ} \leq N_B^{\circ}$. All three components on both sides of Eq. 1 are assumed to be 'ideal gases' in the sense that apart from the reaction described they are not interacting with one another. When an A particle associates with a B particle to form a bound AB particle, there is a change in the potential energy of the system by an amount of ϵ_{AB}. In this model, all possible interparticle energy

states can be uniquely mapped onto N_{AB}, and the partition function of the system can be written as,

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

where index i represents N_{AB}. q_A and q_B are single-particle partition functions of unbound A and B particles and q_{AB} is the pair-particle partition function of one bound AB particle (see Section 7 in Supporting Information for explicit expressions of these functions). A key step in the derivation is the evaluation of the ratio $q_{AB}/(q_A q_B)$. By applying a sequence of algebraic operations, we show (Section 1 in SI) this ratio to be equal to $\langle N_{AB} \rangle / \langle N_A N_B \rangle$, where the brackets indicate ensemble average under equilibrium conditions. From here it follows, the equilibrium constant of the reaction is related to the concentrations of the different components by,

$$K = \frac{\langle c_{AB} \rangle}{\langle c_A c_B \rangle} \cdot c^{\varnothing} \quad , \tag{4}$$

where $c^{\varnothing}$ is a standard reference concentration. Hence, K includes cross-correlations in the reactant concentrations, a feature neglected had we used the following expression,

$$K' = \frac{\langle c_{AB} \rangle}{\langle c_A \rangle \langle c_B \rangle} \cdot c^{\varnothing} \quad , \tag{5}$$

as currently performed in the literature. It is worth noting that in statistical mechanics textbooks[16–18], 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.

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, S1, S2, and S3. 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 (S3) was conducted only by MC. Section 5 in SI provides further details on the model system and computational methodologies.

Figure 1 displays the equilibrium constant, K, calculated by Eq. 4, as well as, the value of K' defined in Eq. 5. 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 volume of the system studied, where its deviation from K increases with decreasing the size of the system. Because correlations become less important at larger systems, in our model K' approaches K for systems with $N_x^\circ > 100$ (S1 and S3) or $L_{box} > 28\ nm$ (S2). In Section 7 in SI we consider even a simpler model system, of single-site reactants, to facilitate an easy comparison between the value of K obtained by Eq. 4 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 S1 series at different temperatures. Figure 2, as well as Fig. 6.1 in Section 6 in SI, indicate this divergence of K' increases with decreasing temperature (or with increasing $-\epsilon_{AB}/k_BT$). 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 $T = 1200\ K$. Moreover, the number of particles above which $K' \longrightarrow K$, increases with decreasing T.

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, $\mu_x = \mu_x^\varnothing + RT \ln\left(c_x/c^\varnothing\right)$, and identifying $\Delta G^\varnothing$ with $\mu_{AB}^\varnothing - \mu_A^\varnothing - \mu_B^\varnothing$, it follows from Eq. 4 that

the condition for equilibrium is,

$$\langle \mu_{AB} \rangle - \langle \mu_A + \mu_B \rangle = 0 \qquad , \tag{6}$$

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

$$\langle \mu_{AB} \rangle - \langle \mu_A \rangle - \langle \mu_B \rangle = 0 \qquad , \tag{7}$$

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

The statistical thermodynamics expression of the equilibrium constant (Eq. 4) 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,

$$\left\langle \frac{dc_{AB}}{dt} \right\rangle = \left\langle -\frac{dc_A}{dt} \right\rangle = \left\langle -\frac{dc_B}{dt} \right\rangle = \langle k_{fw} c_A c_B - k_{bw} c_{AB} \rangle = 0 \qquad , \tag{8}$$

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. 4. We calculated k_{fw} and k_{bw} from the MD simulations (Section 5 in SI) and the results corroborate k_{fw} at finite systems must include correlations between c_A and c_B (Fig. 3).

In constructing S1 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 applications of periodic boundary conditions in finite simulation boxes, but are a consequence of incompatibility between two-body interactions and linear scaling.

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. If the relative fluctuations between quantities ζ and η are defined by, $l(\zeta, \eta) = (\langle \zeta \eta \rangle - \langle \zeta \rangle \langle \eta \rangle) / (\langle \zeta \rangle \langle \eta \rangle)$, we show (Section 2 in SI) that,

$$l(N_{AB}, N_{AB}) - l(N_{AB}, N_A N_B) = \frac{1}{\langle N_{AB} \rangle} \qquad , \tag{9}$$

the particle numbers, or concentrations, of the reactants and product of Eq. 1 can be calculated from only the fluctuations in the system. Equation 9 can also be derived by a more conventional procedure (Section 3 in SI), however there we need to assume K is given by Eq. 4. The relation in Eq. 9 was tested for all simulations performed. In Fig. 5a we plot results at $T = 300\ K$, and in Fig. 6.2 results of the S1 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.

A drawback of Eq. 9 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. The concentration of the bound state can be expressed by K and $l(N_{AB}, N_{AB})$ (Section 4 in SI), transforming the problem to the prediction of $l(N_{AB}, N_{AB})$. For macroscopic systems $l(N_{AB}, N_{AB}) = 0$ and the well known result connecting

concentrations and equilibrium constant is recovered (Eq. 4.3). It turns out, another special case is that of $N_A^\circ = 1$ where $N_B^\circ \geq N_A^\circ$, for which the concentration of the bound state can be expressed by,

$$\langle c_{AB} \rangle_{N_A^\circ = 1} = \frac{N_A^\circ N_B^\circ K}{V(Vc^\varnothing + N_B^\circ K)} \quad , \tag{10}$$

and the relative fluctuations by,

$$l(N_{AB}, N_{AB})_{N_A^\circ = 1} = \frac{Vc^\varnothing}{KN_B^\circ} \quad . \tag{11}$$

S2 and S3 series of simulations fall within this special case, and in Fig. 5b and Fig. 5c we predict $\langle c_{AB} \rangle$ as stated by Eq. 10, whereas in Fig. 6.3 we predict $l(N_{AB}, N_{AB})$ according to Eq. 11. In both cases, the agreement is almost perfect.

For other finite systems, i.e. $N_A^\circ > 1$, we could not establish a link between $l(N_{AB}, N_{AB})$ and K, therefore we propose an approximation based on the case of $N_A^\circ = 1$,

$$l(N_{AB}, N_{AB})_{N_A^\circ \geq 1} \simeq \frac{Vc^\varnothing}{KN_B^\circ} \cdot \frac{1}{(N_A^\circ)^\lambda} \quad , \tag{12}$$

where λ ranges from zero to one and is given by, $\lambda = \left[1 + K/(Vc^\varnothing \ln N_B^\circ)\right]^{-1}$. This approximation is investigated in Fig. 6.4 for the S1 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. 4.2 in Fig. 6.5. 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. 12, we examined the approximation also on another series of simulations, S4, in which N_A° and N_B° are not equal (Section 5 in SI). Here the accuracy of the prediction, shown in Fig. 6.6, is even better. Finally, we scale $g_{ab}(r)$ obtained at finite systems to the corresponding distribution of a macroscopic system, as shown in Fig. 6.7 and discussed in Section 4 in SI.

Discussion

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

$$A + B \rightleftharpoons 2C \qquad , \tag{13}$$

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,

$$K = \frac{\langle c_C^2 \rangle}{\langle c_A c_B \rangle} \qquad . \tag{14}$$

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. 13 is actually a sum of two chemical equations 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. 14 is not appropriate. As no correlations exist between the α-helix and β-sheet conformations, the equilibrium constant should be computed by,

$$K = \frac{\langle c_C \rangle^2}{\langle c_A \rangle \langle c_B \rangle} \qquad . \tag{15}$$

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.

Acknowledgments
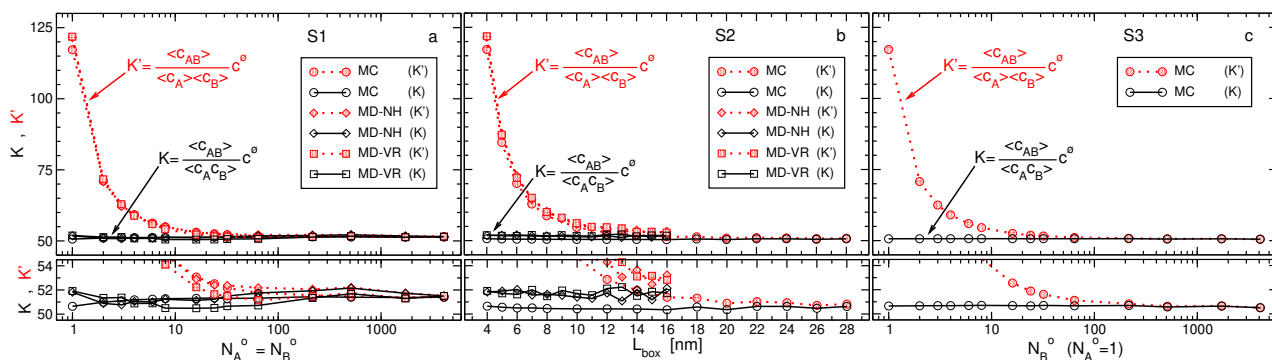
Figures

Figure 1: The equilibrium constant K defined by Eq. 4 $\left(c^{\varnothing} \equiv 1\ M\right)$ for three series of simulations at: (a) constant $c_A^{\circ} = c_B^{\circ} = 0.026\ M$ (S1), (b) constant $N_A^{\circ} = N_B^{\circ} = 1$ (S2), and (c) constant $N_A^{\circ} = 1$ and $c_B^{\circ} = 0.026\ M$ (S3), all performed in the canonical ensemble at $T = 300\ K$. The value of K' defined by Eq. 5 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 $\left(N_A^{\circ} = N_B^{\circ} = 1, L_{box} = 4\ nm\right)$. 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. 6.1 in the Supporting Information.
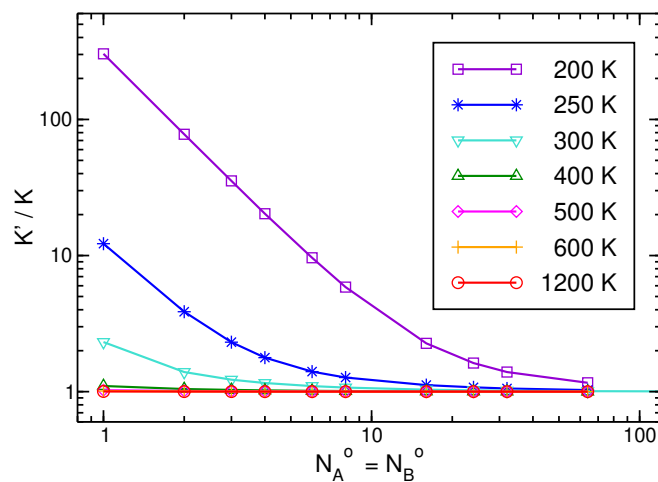
Figure 2: The ratio of the equilibrium constant in which correlations in the reactants concentrations are ignored to that in which they are accounted for, K'/K, from MC S1 series of simulations at different temperatures. This ratio equals $l(N_A, N_B) + 1$ by definition.
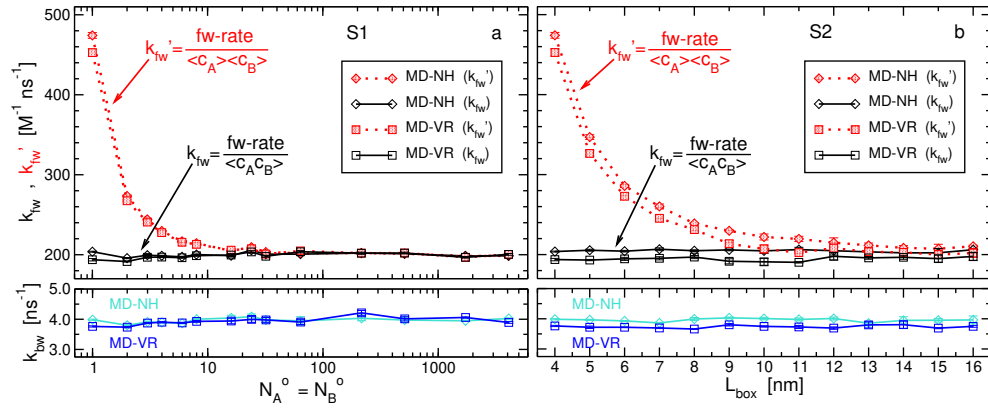
Figure 3: The rate constants of the binding reaction for simulation series S1 (a) and S2 (b) obtained from molecular dynamics simulations. The top panels show the rate constant in the forward direction, k_{fw}, whereas the lower panels the rate constant in the backward direction, k_{bw}. For comparison we present also k'_{fw} calculated by uncorrelated reactants' concentrations.
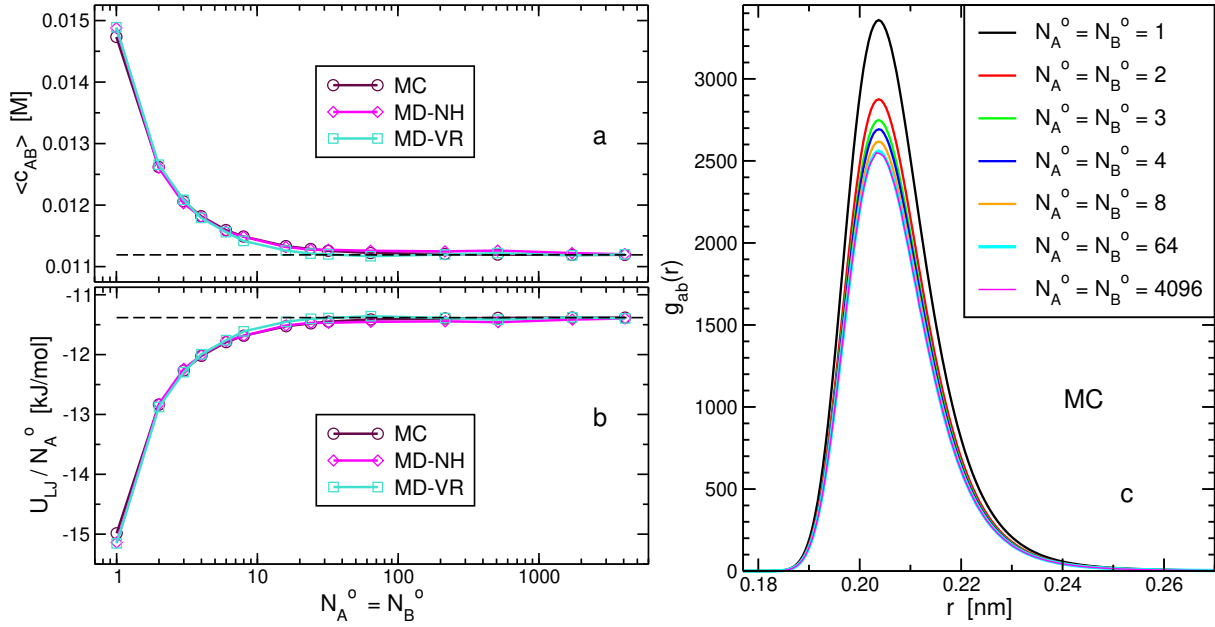
Figure 4: Results exhibiting the inhomogeneous character of properties of bimolecular reactions upon scaling-down system size. The analyzes were performed on S1, i.e. the series of simulations generated by scaling all extensive parameters specifying the system $(N_A^\circ = N_B^\circ, V)$ by the same 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.
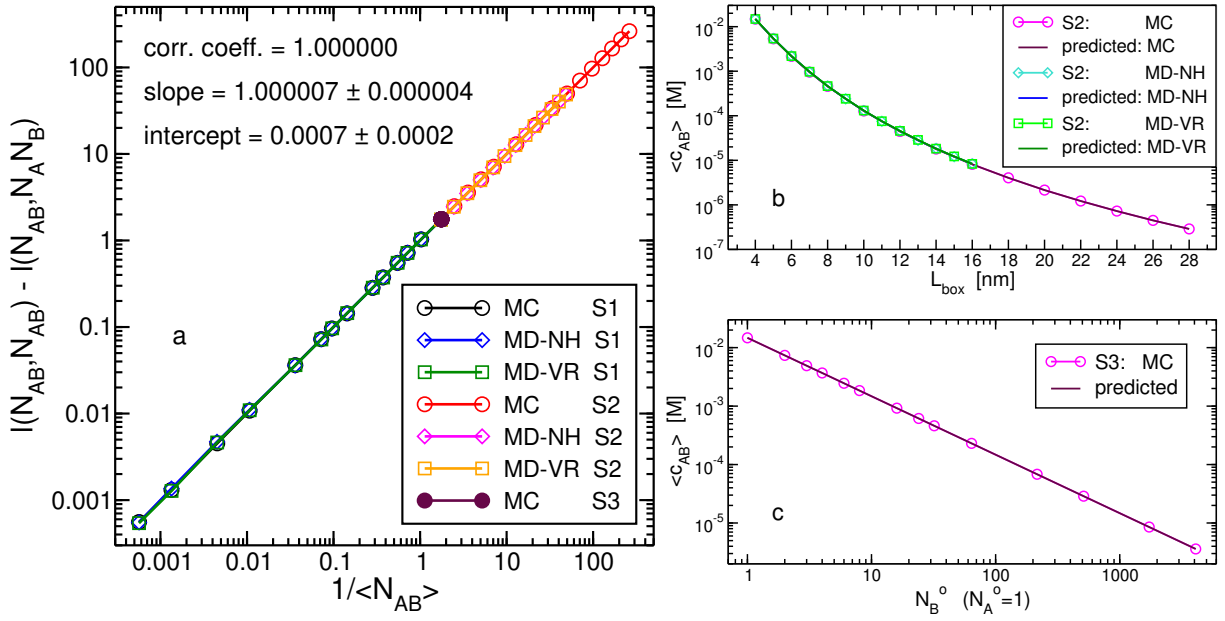
Figure 5: (a) A relation between two relative fluctuations and the reciprocal of average number of bound particles. All simulations results (here displayed for $T = 300\ K$, for other temperatures see Fig. 6.2 in SI) fall on a linear line crossing the origin with a slope of one as described in Eq. 9. Results obtained from linear regression (using xmgrace) of all data points are indicated. All points of S3 have the same x, y values. (b) The concentration of bound particles as a function of box length for S2 series. The results obtained from simulations are shown along predictions based on the value of the equilibrium constant (Eq. 10). (c) Same as (b) but for S3 series of simulations, in which case, the concentration is plotted as a function of the total number of B particles.

References

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

[2] 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.

[3] 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.

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

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

[6] 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.

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

[8] 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.

[9] 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.

[10] De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery, *J. Med. Chem.* **2016**, *59*, 4035–4061.

[11] 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.

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

[13] 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.

[14] Hall, R.; Dixon, T.; Dickson, A. On Calculating Free Energy Differences Using Ensembles of Transition Paths, *Front. Mol. Biosci.* **2020**, *7*, 106.

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

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

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

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