PCCP

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Phys. Chem. Chem. Phys.*, 2022, **24**, 28804

Statistical mechanics of dimerizations and its consequences for small systems†

Ronen Zangi [ab]

Received 23rd September 2022,
Accepted 1st November 2022

DOI: 10.1039/d2cp04450a

rsc.li/pccp

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 number of 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 upon 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. All predictions are verified by Monte Carlo and molecular dynamics simulations.

Published on 03 November 2022. Downloaded on 12/9/2022 8:44:40 AM.

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 concentrations of the participating reactants.[1] As an underlying principle in chemical kinetics, this 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,

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

the average concentration, at equilibrium, of the product and that of each of the reactants are obtained and then the ratio $\langle c_{\mathrm{AB}}\rangle_{\mathrm{eq}}/(\langle c_{\mathrm{A}}\rangle_{\mathrm{eq}}\langle c_{\mathrm{B}}\rangle_{\mathrm{eq}})$ 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, $k_{\mathrm{fw}} = \langle \text{fw-rate}\rangle_{\mathrm{eq}}/(\langle c_{\mathrm{A}}\rangle_{\mathrm{eq}}\langle c_{\mathrm{B}}\rangle_{\mathrm{eq}})$, 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 eqn (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.[4–12] Different explanations were put forward that include conformational changes of the unbound molecule(s), non-fluorescent bindings, and missed events due to transient interactions.[13,14] Several theoretical studies proposed that small systems, attributed to be stochastic in nature, are characterized by equilibrium constants different than that observed for a macroscopic system, which attributed to be deterministic.[15–22] Size-dependent equilibrium constant was also advocated by introducing 'nanoconfinement entropic effects on chemical equilibrium' applied only to systems with small number of molecules.[23,24] Other computational works also reported deviations of the bound product's concentration from that anticipated by the above-mentioned expression of K.[25,26] In these cases, the anomalous behavior was explained by artefacts due to applications of periodic boundary conditions in finite simulation

[a] *POLYMAT & Department of Organic Chemistry I, University of the Basque Country UPV/EHU, Avenida de Tolosa 72, 20018, Donostia-San Sebastián, Spain. E-mail: r.zangi@ikerbasque.org*
[b] *IKERBASQUE, Basque Foundation for Science, Plaza Euskadi 5, 48009 Bilbao, Spain*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d2cp04450a

 This journal is © the Owner Societies 2022

Published on 03 November 2022. Downloaded on 12/9/2022 8:44:40 AM.

boxes[27–29] or due to neglected concentration fluctuations in small simulations in the canonical ensemble.[30,31]

In tackling this issue, we recently argued[32] that any intensive property related to a two-body interaction (such as the concentration of the bound product AB in eqn (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 (hereafter, we omit the subscript 'eq' indicating averages are taken at equilibrium conditions) $K = \langle c_{AB}\rangle/\langle c_A{\cdot}c_B\rangle$ and $k_{fw} = \langle \text{fw-rate}\rangle/\langle c_A{\cdot}c_B\rangle$. 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,[33–35] 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, works in the literature that followed ignored these correlations in reactants' concentrations when calculating K.[36–45]

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

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

where reactant B in eqn (1) is substituted with another reactant of A in eqn (2), as,

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

for the equilibrium constant, where for consistency with the definition of K stated in eqn (5) below we multiplied the ratio by the standard concentration $c^{\varnothing}$ and as,

$$k''_{fw} = \frac{\langle \text{fw-rate}\rangle}{\langle c_A{}^2\rangle}, \tag{4}$$

for the bimolecular rate constant. That said, it appears unconsciously that the way chemists, including the writer of these lines,[32] practice chemistry is deeply rooted in the behavior of macroscopic systems. More concretely, the expressions in eqn (3) and (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 eqn (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 eqn (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 dimerization

We consider the dimerization process shown in eqn (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,

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

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 eqn (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 eqn (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,[46]

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

should then be accounted for when calculating $\Delta G^{\varnothing}$ (see Fig. 1). To put it another way, this additional compression step had to be introduced when utilizing the canonical ensemble because the stoichiometric coefficients of reactant and product in the dimerization reaction (eqn (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.*, $\varepsilon < 0$). Owing to the ideal behavior of the chemical components, the (interparticle) energy states of the system are uniquely defined by the number

Published on 03 November 2022. Downloaded on 12/9/2022 8:44:40 AM.

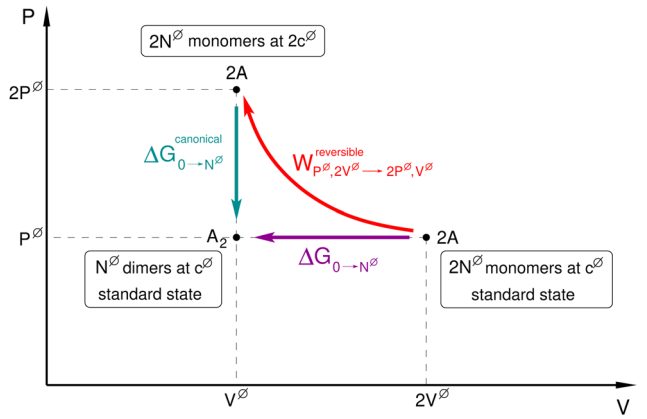

Fig. 1 Projection of the dimerization reaction of the reference system onto an isothermal pressure–volume diagram. The figure illustrates that in this case, connecting the reactant (2A monomers) to the product (A_2 dimers), both at standard state conditions (purple arrow), *via* a description in the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

of dimers, $N_{A_2} \equiv i$, and the canonical partition function of the reference system can be written as,

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

where the number of monomers, N_A, equals $2(N^{\varnothing} - i)$. As it should, the sum in eqn (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 $q_{A_2}^{\varnothing}$ is the pair-particle partition function of one dimer A_2 (incorporating the exponential $e^{-\beta\varepsilon}$). These partition functions can be expressed in different forms and are described in details in the ESI.† Because the particles are indistinguishable, the factorials in the denominators of eqn (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, $\Delta F_{0\to N^{\varnothing}}^{\text{canonical}}$, 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 eqn (7)), to the probability to find all particles as free monomers, p^{2A} (the fraction of the state $i = 0$),

$$\begin{aligned}\Delta F_{0\to N^{\varnothing}}^{\text{canonical}} &\equiv F_{i=N^{\varnothing}}^{\text{canonical}} - F_{i=0}^{\text{canonical}} = -k_B T \ln \frac{p^{A_2}}{p^{2A}} \\ &= -k_B T \ln \left[\frac{\left(q_{A_2}^{\varnothing}\right)^{N^{\varnothing}}}{N^{\varnothing}!} \cdot \frac{(2N^{\varnothing})!}{\left(q_A^{\varnothing}\right)^{2N^{\varnothing}}} \right],\end{aligned} \tag{8}$$

where k_B is Boltzmann constant. The corresponding Gibbs free energy change is then,

$$\begin{aligned}\Delta G_{0\to N^{\varnothing}}^{\text{canonical}} &= \Delta F_{0\to N^{\varnothing}}^{\text{canonical}} + V^{\varnothing} \Delta P_{0\to N^{\varnothing}}^{\text{canonical}} \\ &= -N^{\varnothing} k_B T \ln \frac{q_{A_2}^{\varnothing}}{\left(q_A^{\varnothing}\right)^2} - k_B T \ln \frac{(2N^{\varnothing})!}{N^{\varnothing}!} \\ &\quad + V^{\varnothing} \Delta P_{0\to N^{\varnothing}}^{\text{canonical}},\end{aligned} \tag{9}$$

where $\Delta P_{0\to N^{\varnothing}}^{\text{canonical}}$ 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 $\Delta W_{P^{\varnothing},2V^{\varnothing}\to 2P^{\varnothing},V^{\varnothing}}^{\text{reversible}}$ (as computed in eqn (6)) to $\Delta G_{0\to N^{\varnothing}}^{\text{canonical}}$, and continue by applying Stirling's approximation to the numerator and denominator of the second term on the right hand side of eqn (9), that means requiring the reference system to be large,

$$\begin{aligned}\Delta G_{0\to N^{\varnothing}} &= \Delta W_{P^{\varnothing},2V^{\varnothing}\to 2P^{\varnothing},V^{\varnothing}}^{\text{reversible}} + \Delta G_{0\to N^{\varnothing}}^{\text{canonical}} \\ &= -N^{\varnothing} k_B T \left[\ln \frac{q_{A_2}^{\varnothing}}{\left(q_A^{\varnothing}\right)^2} + \ln N^{\varnothing} \right] \\ &\quad + N^{\varnothing} k_B T + V^{\varnothing} \Delta P_{0\to N^{\varnothing}}^{\text{canonical}}.\end{aligned} \tag{10}$$

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

$$\Delta G_{0\to N^{\varnothing}} = -N^{\varnothing} k_B T \left[\ln \frac{q_{A_2}^{\varnothing} / V^{\varnothing}}{\left(q_A^{\varnothing} / V^{\varnothing}\right)^2} + \ln c^{\varnothing} \right]. \tag{11}$$

We now evaluate the ratio of the partition functions in eqn (11). Due to translational degrees of freedom, $q_{A_2}^{\varnothing}$ and $q_A^{\varnothing}$ depend on the size of the system. Nonetheless assuming 'classical' behavior of translational energy states (eqn (SI-2.2), ESI†), as is the case in deriving eqn (SI-1.5) (ESI†), the single- and pair-particle partition function can be rendered size-independent upon division by the volume. Hence if we consider another system for the dimerization process in eqn (2), at the same temperature T but with an arbitrary total number of monomers, N_A^{total}, and an arbitrary volume, V, the following relation holds,

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

where q_A and q_{A_2} are the single- and pair-particle partition functions of this arbitrary system. We note the validity of the 'classical' translation approximation diminishes with decreasing temperature, mass, and volume. In Section SI-2 of the ESI† we analyze and check the equality of eqn (12). Although for the vast majority of molecular systems this equality seems to hold to an acceptable degree of accuracy, there are special cases of low molecular weight gases (such as hydrogen and helium) at low temperatures and confined to small volumes for which the 'classical' approximation yields large discrepancies. Returning to

 This journal is © the Owner Societies 2022

Published on 03 November 2022. Downloaded on 12/9/2022 8:44:40 AM.

the arbitrary system, its total partition function is analogous to that of the reference system (eqn (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_A^{total} 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 $N_A^{total} = N_A + 2N_{A_2} = 2N^{\circ} + \delta$, where N° is the maximum number of dimers that can hypothetically form, and δ equals 0 or 1 depending on whether N_A^{total} is even or odd, respectively. We then write the canonical partition function for the arbitrary system as,

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

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

In order to proceed with the evaluation of $\Delta G_{0\to N^{\varnothing}}$ (eqn (11)) we multiply and divide the right-hand side of eqn (12) by,

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

and obtain,

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

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

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

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,

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

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 eqn (11) gives,

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

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

$$\begin{aligned} \Delta G^{\varnothing} &= \frac{N_{Avogadro}}{N^{\varnothing}} \cdot \Delta G_{0\to N_A^{\varnothing}} \\ &= -RT \ln \frac{\langle c_{A_2}\rangle c^{\varnothing}}{\left\langle c_A \cdot \left(c_A - \frac{1}{V}\right)\right\rangle}, \end{aligned} \tag{19}$$

and comparing the resulting expression to the definition of K in eqn (5) we arrive at,

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

Therefore as for the case of binding reaction in eqn (1), the equilibrium constant for dimerization must include density 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 $\Delta G_{0\to N^{\varnothing}}^{canonical}$) the expression of K would be the same as that in eqn (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 $1/V \ll c_A$, and correlations between reactant particles are totally lost. In this case, K'' in eqn (3) and a related expression ignoring all correlations,

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

approach K in eqn (20). The second case is for the smallest system possible, $N_A^{total} = 2$, where the system has only two macroscopic states. Despite strong correlations in the system,

Published on 03 November 2022. Downloaded on 12/9/2022 8:44:40 AM.

the two-body average $\langle N_A(N_A - 1)\rangle$ reduces to a one-body average $\langle N_A\rangle$, and eqn (20) can be written as,

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

where $f^{A_2} \equiv \langle N_{A_2}\rangle$ is the fraction of frames, or probability, in which the dimer is observed. The relation in eqn (22) is valid only when there are two particles in the system and has already been employed by Ouldridge *et al.*[30]

II. Validation by computer simulation

To check our derivation, we consider a simple system of Lennard-Jones (LJ) molecules able to dimerize according to eqn (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 $c_A^{total} = N_A^{total}/V$ constant, whereas in the second series, R2, we fixed $N_A^{total} = 2$ 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.

Fig. 2 displays the equilibrium constant, K, calculated by eqn (20), together with the value of K' (eqn (21)) and K'' (eqn (3)). As indicated by the figure, inclusion of cross-correlations are needed in order to keep the equilibrium 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 value of 2 is always significant. Furthermore, in Section SI-1 of the ESI† we show the value of K calculated by eqn (20), for a system with a single-site reactant, agrees almost perfectly with that obtained by analytical calculations (Table SI-1.2, ESI†).

The expression of K in eqn (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,

$$\langle \text{fw-rate}\rangle = k_{fw}\langle c_A(c_A - 1/V)\rangle. \quad (23)$$

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,

$$\left\langle \frac{dc_{A_2}}{dt}\right\rangle = -\frac{1}{2}\left\langle \frac{dc_A}{dt}\right\rangle = \langle k_{fw}c_A(c_A - 1/V) - k_{bw}c_{A_2}\rangle = 0, \quad (24)$$

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

$$k'_{fw} = \frac{\langle \text{fw-rate}\rangle}{\langle c_A\rangle^2}, \quad (25)$$

mirrors the results presented for the corresponding expressions of K.

III. A relation between composition and fluctuation

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 we adopt the notation of Lebowitz *et al.*[47] and define the cross fluctuations between quantities ζ and η as,

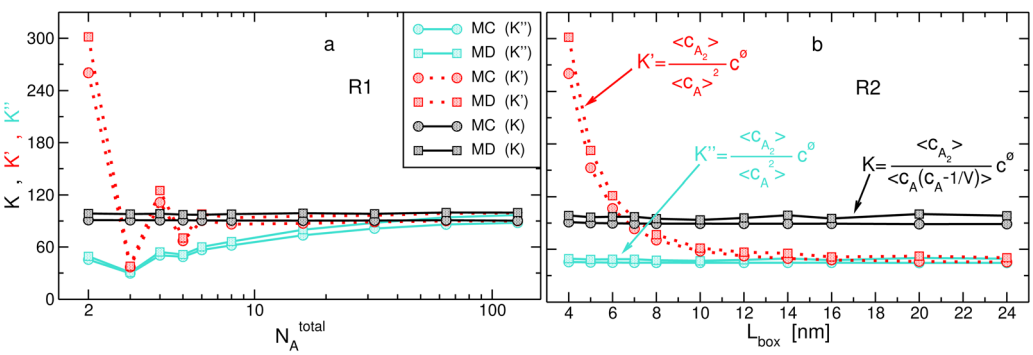

Fig. 2 The equilibrium constant K for dimerization defined by eqn (20) ($c^{\varnothing} \equiv 1$ M) for two series of simulations at: (a) constant $c_A^{total} = 0.052$ M and as a function of the total number of A particles (R1), as well as, at (b) constant $N_A^{total} = 2$ and as a function of the 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 eqn (21) and (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.

 This journal is © the Owner Societies 2022

Published on 03 November 2022. Downloaded on 12/9/2022 8:44:40 AM.

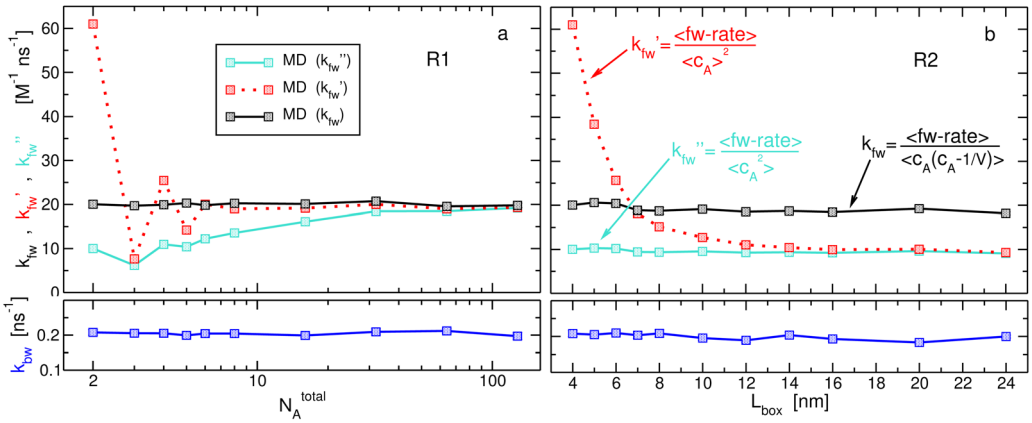

Fig. 3 Rate constants of dimerization (eqn (2)) for (a) R1 series, and (b) R2 series, obtained from MD simulations. The top panels show the rate constant in the forward direction, k_{fw} defined in eqn (23), whereas the lower panels the rate constant in the backward direction, $k_{bw} = \langle \text{bw-rate}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

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

and their relative magnitude by,

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

We now look at the following difference in our system,

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

and evaluate the term inside the square brackets. Utilizing the partition function defined in eqn (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,

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

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 eqn (28) by,

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

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 term in the square brackets of eqn (28) by noting the denominators of the two terms are equal,

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

and obtain a ratio that equals one. Consequently, eqn (28) reduces to,

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

Published on 03 November 2022. Downloaded on 12/9/2022 8:44:40 AM.

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

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

In Fig. 4 we examine the validity of eqn (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, 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[32] for the binding reaction of eqn (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.

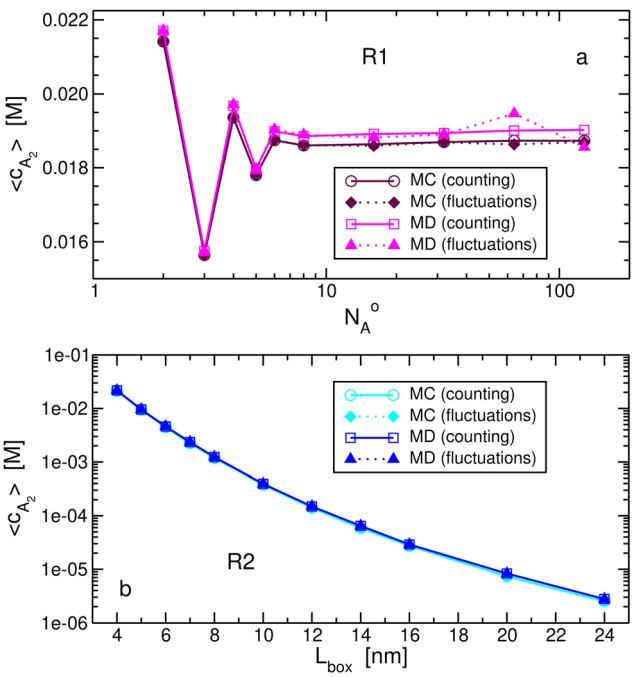

Fig. 4 Average concentrations of dimers, $\langle c_{A_2} \rangle$, for (a) R1 and (b) R2 series of simulations. Along results obtained from direct counting of the number of dimers, we also predict the concentrations from the relative fluctuations, $l(N_{A_2}, N_{A_2})$ and $l(N_{A_2}, N_A[N_A - 1])$,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

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 eqn (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 eqn (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[32] 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.*, eqn (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[32] 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, eqn (20), and of the rate of the forward bimolecular reaction, eqn (23), ought to be used in dimerization reactions of the type presented in eqn (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

 This journal is © the Owner Societies 2022

Published on 03 November 2022. Downloaded on 12/9/2022 8:44:40 AM.

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 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 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^{total}, V, T) at a temperature of $T = 300$ K. The total number of A molecules in the system, $N_A^{total} = N_A + 2N_{A_2}$, 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 $c_A^{total} = N_A^{total}/V$ is constant at 0.03125 molecules per nm^3 ($\sim$0.052 M). In the second series of simulations, R2, we considered only two molecules of A, $N_A^{total} = 2$, and increased V by increasing the box length, L_{box}, from 4.0 nm to 24.0 nm.

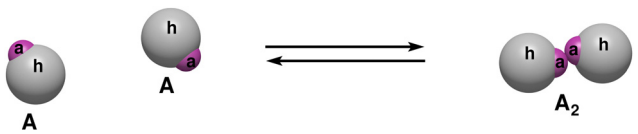

Fig. 5 A model system for dimerization between two A molecules. These A molecules consist of uncharged LJ, a and h, atom-sites covalently bonded to each other. The distance of this intramolecular bond is fixed in the MC simulations to a value of 0.25 nm whereas it oscillates around this value, due to 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).

Table 1 LJ parameters between the different atom sites for a system of A(a–h) molecules

	σ [nm]	ε [kJ mol^{-1}]
a···a	0.15	47.0
h···h	0.85	0.1
a···h	0.40	0.1

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,[48,49] which output configurations in the canonical ensemble, were performed by an in-house code executed in double-precision. The Metropolis acceptance criteria[50] 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 $\theta(0 \le \theta \le \pi)$, and 0.314 rad for rotations around angle ϕ $(0 \le \phi \le 2\pi)$. These trial moves resulted in acceptance-ratios that varied from 0.17, for the system $N_A^{total} = 2$ 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 4×10^{12} moves for $N_A^{total} = 2$ to 1.25×10^{11} moves for the largest system of $N_A^{total} = 128$.

The MD simulations were conducted by the software package GROMACS version 4.6.5[51] (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 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 2×10^5 kJ mol^{-1} nm^{-2}. The temperature was maintained by applying the Nosé–Hoover thermostat[52,53] with a chain-length[54] 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 μs was conducted prior to data collection for each system, whereas, the time period for collecting data ranged from 400 μs for $N_A^{total} = 2$ to 29.6 μs for $N_A^{total} = 128$.

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 $N_A^{total} = 2$ (or $L_{box} = 4.0$ nm) to a frequency of every step for $N_A^{total} \ge 8$ (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 ranged from 12 μs for $N_A^{total} = 2$ (or $L_{box} = 4.0$ nm) to 300 ns for the largest system in R1, or to 600 μs 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

Published on 03 November 2022. Downloaded on 12/9/2022 8:44:40 AM.

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, it turned out the effect of including this buffer zone is rather negligible.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

I would like thank one of the referees for pointing out the limitation of applying the 'classical' approximation to the translational partition function. 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 L. F. Wilhemy, Ueber das Gesetz, nach welchem die Einwirkung der Säuren auf den Rohrzucker stattfindet, *Ann. Phys.*, 1850, **81**, 413–433.
2 P. Waage and C. M. Guldberg, Studier over Affiniteten *Forhandlinger i Videnskabs-selskabet i Christiania*, 1864, pp. 35–45.
3 J. H. van 't Hoff, *Studies in Chemical Dynamics*, William & Norgate, London, 1896.
4 M. Morimatsu, H. Takagi, G. Kosuke, R. I. Ota, T. Yanagida and Y. Sako, Multiple-state reactions between the epidermal growth factor receptor and Grb2 as observed by using single-molecule analysis, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 18013–18018.
5 S. Polarz and A. Kuschel, Chemistry in confining reaction fields with special emphasis on nanoporous materials, *Chem. – Eur. J.*, 2008, **14**, 9816–9829.
6 T. Sawada, M. Yoshizawa, S. Sato and M. Fujita, Minimal nucleotide duplex formation in water through enclathration in self-assembled hosts, *Nat. Chem.*, 2009, **1**, 53–56.
7 M. J. Shon and A. E. Cohen, Mass action at the single-molecule level, *J. Am. Chem. Soc.*, 2012, **134**, 14618–14623.
8 S. Patra, A. N. Naik, A. K. Pandey, D. Sen, S. Mazumder and A. Goswami, Silver nanoparticles stabilized in porous polymer support: A highly active catalytic nanoreactor, *Appl. Catal., A*, 2016, **524**, 214–222.
9 C. J. Galvin, K. Shirai, A. Rahmani, K. Masaya and A. Q. Shen, Total capture, convection-limited nanofluidic immunoassays exhibiting nanoconfinement effects, *Anal. Chem.*, 2018, **90**, 3211–3219.
10 C. F. Megarity, B. Siritanaratkul, R. S. Heath, L. Wan, G. Morello, S. R. FitzPatrick, R. L. Booth, A. J. Sills, A. W. Robertson, J. H. Warner, N. J. Turner and F. A. Armstrong, Electrocatalytic volleyball: Rapid nanoconfined nicotinamide cycling for organic synthesis in electrode pores, *Angew. Chem., Int. Ed.*, 2019, **58**, 4948–4952.
11 A. M. Downs, C. McCallum and S. Pennathur, Confinement effects on DNA hybridization in electrokinetic micro- and nanofluidic systems, *Electrophoresis*, 2019, **40**, 792–798.
12 S. Jonchhe, S. Pandey, C. Beneze, T. Emura, H. Sugiyama, M. Endo and H. Mao, Dissection of nanoconfinement and proximity effects on the binding events in DNA origami nanocavity, *Nucleic Acids Res.*, 2022, **50**, 697–703.
13 J. Yang and J. E. Pearson, Origins of concentration dependence of waiting times for singlemolecule fluorescence binding, *J. Chem. Phys.*, 2012, **136**, 244506.
14 G. Longatte, F. Lisi, P. Bakthavathsalam, T. Böcking, K. Gaus, R. D. Tilley and J. J. Gooding, Biomolecular binding under confinement: Statistical predictions of steric influence in absence of long-distance interactions, *Comput. Phys. Commun.*, 2022, **23**, e202100765.
15 I. G. Darvey, B. W. Ninham and P. J. Staff, Stochastic models for second-order chemical reaction kinetics. The equilibrium state, *J. Chem. Phys.*, 1966, **45**, 2145–2155.
16 G. Rothenberger and M. Grätzel, Effects of spatial confinement on the rate of bimolecular reactions in organized liquid media, *Chem. Phys. Lett.*, 1989, **154**, 165–171.
17 I. J. Laurenzi, An analytical solution of the stochastic master equation for reversible bimolecular reaction kinetics, *J. Chem. Phys.*, 2000, **113**, 3315–3322.
18 D. Holcman and Z. Schuss, Stochastic chemical reactions in microdomains, *J. Chem. Phys.*, 2005, **122**, 114710.
19 K. Ghosh, Stochastic dynamics of complexation reaction in the limit of small numbers, *J. Chem. Phys.*, 2011, **134**, 195101.
20 R. Szymanski, S. Sosnowski and L. Maślanka, 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.
21 R. Szymanski and S. Sosnowski, Stochasticity of the transfer of reactant molecules between nano-reactors affecting the reversible association A + B $\rightleftharpoons$ C, *J. Chem. Phys.*, 2019, **151**, 174113.
22 W. Goch and W. Bal, Stochastic or not? method to predict and quantify the stochastic effects on the association reaction equilibria in nanoscopic systems, *J. Phys. Chem. A*, 2020, **124**, 1421–1428.
23 M. Polak and L. Rubinovich, Nanochemical equilibrium involving a small number of molecules: A prediction of a distinct confinement effect, *Nano Lett.*, 2008, **8**, 3543–3547.
24 M. Polak and L. Rubinovich, The intrinsic role of nanoconfinement in chemical equilibrium: Evidence from DNA hybridization, *Nano Lett.*, 2013, **13**, 2247–2251.
25 J. T. Kindt, Accounting for finite-number effects on cluster size distributions in simulations of equilibrium aggregation, *J. Chem. Theory Comput.*, 2013, **9**, 147–152.

 This journal is © the Owner Societies 2022

Published on 03 November 2022. Downloaded on 12/9/2022 8:44:40 AM.

26 L. A. Patel and J. T. Kindt, 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.
27 D. H. De Jong, L. V. Schäfer, A. H. De Vries, S. J. Marrink, H. J. C. Berendsen and H. Grubmüller, determining equilibrium constants for dimerization reactions from molecular dynamics simulations, *J. Comput. Chem.*, 2011, **32**, 1919–1928.
28 R. Cortes-Huerto, K. Kremer and R. Potestio, Communication: Kirkwood–Buff integrals in the thermodynamic limit from small-sized molecular dynamics simulations, *J. Chem. Phys.*, 2016, **145**, 141103.
29 N. Dawass, P. Krüger, S. K. Schnell, D. Bedeaux, S. Kjelstrup, J. M. Simon and T. J. H. Vlugt, Finite-size effects of Kirkwood–Buff integrals from molecular simulations, *Mol. Inf.*, 2018, **44**, 599–612.
30 T. E. Ouldridge, A. A. Louis and J. P. K. Doye, Extracting bulk properties of self-assembling systems from small simulations, *J. Phys.: Condens. Matter*, 2010, **22**, 104102.
31 T. E. Ouldridge, 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.
32 R. Zangi, Binding reactions at finite systems, *Phys. Chem. Chem. Phys.*, 2022, **24**, 9921–9929.
33 D. A. McQuarrie, *Statistical Thermodynamics*, University Science Books, Mill Valley, CA, 1973.
34 D. Chandler, *Introduction to Modern Statistical Mechanics*, Oxford University Press, New York, NY, 1987.
35 H. Gould and J. Tobochnik, *Statistical and Thermal Physics: With Computer Applications*, Princeton University Press, Princeton, NJ, 2010.
36 M. K. Gilson, J. A. Given, B. L. Bush and J. A. McCammon, The Statistical-thermodynamic basis for computation of binding affinities: A critical review, *Biophys. J.*, 1997, **72**, 1047–1069.
37 P. H. Hünenberger, J. K. Granwehr, J.-N. Aebischer, N. Ghoneim, E. Haselbach and W. F. van Gunsteren, Experimental and theoretical approach to hydrogen-bonded diastereomeric interactions in a model complex, *J. Am. Chem. Soc.*, 1997, **119**, 7533–7544.
38 H. Luo and K. Sharp, On the calculation of absolute macromolecular binding free energies, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 10399–10404.
39 Y. Zhang and J. A. McCammon, Studying the affinity and kinetics of molecular association with molecular-dynamics simulation, *J. Chem. Phys.*, 2003, **118**, 1821–1827.
40 E. Psachoulia, P. W. Fowler, P. J. Bond and M. S. P. Sansom, Helix–helix interactions in membrane proteins: Coarse-grained simulations of glycophorin a helix dimerization, *Biochemistry*, 2008, **47**, 10503–10512.
41 Y. Deng and B. Roux, Computations of standard binding free energies with molecular dynamics simulations, *J. Phys. Chem. B*, 2009, **113**, 2234–2246.
42 R. Skorpa, J.-M. Simon, D. Bedeaux and S. Kjelstrup, Equilibrium properties of the reaction $H_2 \rightleftharpoons 2H$ by classical molecular dynamics simulations, *Phys. Chem. Chem. Phys.*, 2014, **16**, 1227–1237.
43 J. J. Montalvo-Acosta and M. Cecchini, Computational approaches to the chemical equilibrium constant in protein-ligand binding, *Mol. Inf.*, 2016, **35**, 555–567.
44 N. D. Piccolo and K. Hristova, Quantifying the Interaction between EGFR Dimers and Grb2 in Live Cells, *Biophys. J.*, 2017, **113**, 1353–1364.
45 E. Duboué-Dijon and J. Hénin, Building intuition for binding free energy calculations: Bound state definition, restraints, and symmetry, *J. Chem. Phys.*, 2021, **154**, 204101.
46 H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics*, John Wiley & Sons, New York, NY, 1985.
47 J. L. Lebowitz, J. K. Percus and L. Verlet, Ensemble dependence of fluctuations with application to machine computations, *Phys. Rev.*, 1967, **153**, 250–254.
48 M. P. Allen and D. J. Tildesley, *Computer Simulations of Liquids*, Oxford Science Publications, Oxford, 1987.
49 D. Frenkel and B. Smit, *Understanding Molecular Simulations: From Algorithms to Applications*, Academic Press, London, 2002.
50 N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, Equation of state calculations by fast computing machines, *J. Chem. Phys.*, 1953, **21**, 1087–1092.
51 B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation, *J. Chem. Theory Comput.*, 2008, **4**, 435–447.
52 S. Nosé, A unified formulation of the constant temperature molecular dynamics methods, *J. Chem. Phys.*, 1984, **81**, 511–519.
53 W. G. Hoove, Canonical dynamics: Equilibrium phase-space distributions, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1985, **31**, 1695–1697.
54 G. J. Martyna, M. L. Klein and M. Tuckerman, Nosé–Hoover chains: The canonical ensemble via continuous dynamics, *J. Chem. Phys.*, 1992, **97**, 2635–2643.

Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.
This journal is © the Owner Societies 2022

Supplementary Information: 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*

November 2, 2022

*r.zangi@ikerbasque.org

SI-1 Analytical Evaluations of K

The equilibrium constant of dimerization derived in the main text and expressed in Eq. 20 in terms of ensemble average of reactant and product concentrations is now compared against two analytical evaluations based on the single-particle, q_A, and pair-particle, q_{A_2}, partition functions. To this end, we simplify our system and model the reactants, A, only as single-site particles, thus, removing the protecting site that prevented higher-order aggregation. To preclude the formation of aggregates larger than a dimer, we simply restrict this test system to $N_A^{\text{total}} = 2$. We choose the a cubic box with $L_{box} = 6.0\ nm$ thus $c_A^{\text{total}} = 0.00926\ molecule/nm^3$. To render the magnitude, as well as the location, of the first maximum of $g(r)$ in the single-site system and in the main-model system similar, we modified ϵ and σ parameters of the LJ potential to $\epsilon^{LJ} = 26.90\ kJ/mol$ and $\sigma = 0.152\ nm$. Other simulation parameters were unchanged. The MC simulation consisted of $8 \cdot 10^{12}$ trial moves whereas the MD simulation was run for 720 μs. The value of K obtained by Eq. 20, for each of these simulations, is listed in Table SI-1.2.

I. K from Integration over Particle's Coordinates

In this approach we completely separate the integrations over momenta from those over spatial coordinates. If $\mathcal{T}$ is the kinetic part of the Hamiltonian, the single-particle partition function of unbound A can be written as,

$$q_A(r) = \frac{1}{h^3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\beta\mathcal{T}(\vec{p_A})}d\vec{p_A}\int_{\vec{r_A}} d\vec{r_A} = \frac{V}{h^3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\beta\mathcal{T}(\vec{p_A})}d\vec{p_A} \quad , \quad \text{(SI-1.1)}$$

where h is Planck's constant and the integral over $\vec{r_A}$ is of three dimensions yielding the volume when the particle does not interact with its surrounding. If $\mathcal{U}$ is the potential part of the Hamiltonian and r_c the cutoff distance defining the bound state, the pair-particle partition function can be written as,

$$q_{A_2}(\vec{p_{A'}}, \vec{p_{A''}}, \vec{r_{A'}}, \vec{r_{A''}}) = \frac{1}{h^6}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{-\beta\mathcal{T}(\vec{p_{A'}},\vec{p_{A''}})}d\vec{p_{A'}}d\vec{p_{A''}}\int_{\vec{r_{A'}}} d\vec{r_{A'}}\int_0^{r_c} e^{-\beta\mathcal{U}(r)}2\pi r^2 dr \quad , \quad \text{(SI-1.2)}$$

where we labeled the first particle A', and the second A''. The relative distance is defined as $r = |\vec{r_{A''}} - \vec{r_{A'}}|$, and the usual volume element for integration over this relative distance, $4\pi r^2 dr$,

is divided by two because A' and A'' are indistinguishable. In addition, the spatial integration over the coordinates of the first particle A' gives V, thus, the equilibrium constant can be expressed as,

$$K = \frac{q_{A_2}}{q_A^2}\, Vc^{\varnothing} = \frac{1}{2}\, c^{\varnothing} \int_0^{r_c} e^{-\beta \mathcal{U}(r)} d\vec{r} = c^{\varnothing} \int_0^{r_c} e^{-\beta \mathcal{U}(r)} 2\pi r^2 dr \qquad , \tag{SI-1.3}$$

where the integrals over momenta cancel-out when taking the ratio of the partition functions. Equation SI-1.3 can be solved numerically and the result is shown in Table SI-1.2.

II. K from a Molecular Partition Function

We now evaluate q_{A_2} by integrations over coordinates and momenta of the center-of-mass of the dimer and over the relative motions therein. This is realized by writing the Hamiltonian of the pair-particle partition function in terms of generalized coordinates and momenta that describe translation of center-of-mass, as well as, rotation and vibration of the bound state. If the rotational and vibrational modes are decoupled, the expression of K becomes[1],

$$K = \frac{q_{\text{trans}}(A_2) \cdot q_{\text{rot}}(A_2) \cdot q_{\text{vib}}(A_2) \cdot e^{-\beta\epsilon}}{q^2_{\text{trans}}(A)}\, Vc^{\varnothing} \qquad , \tag{SI-1.4}$$

where ϵ equals $-\epsilon^{LJ}/N_{\text{Avogadro}}$. In the 'classical' approximation, where the sum over translational states can be substituted by an integral, the translational partition function has the form,

$$q_{\text{trans}}^{\text{classical}} = \left(\frac{2\pi m k_{_B} T}{h^2}\right)^{3/2} V \qquad , \tag{SI-1.5}$$

where m is the mass of the moving body. The rotational partition function of a homonuclear rigid-rotor dimer at high-temperatures is,

$$q_{\text{rot}} = \frac{8\pi^2 I k_{_B} T}{2h^2} \qquad , \tag{SI-1.6}$$

where the moment of inertia is $I = \mu R_{eq}^2$, μ the reduced mass, and $R_{eq} = 0.1707\ nm$ the equilibrium bond length of the dimer. The evaluation of the vibrational partition function is normally proceeded by an input of the vibrational frequency (or force-constant). Because the vibrations in our dimer are actually oscillatory motions around the minimum of the LJ potential, we also apply the high-temperature approximation in this case and evaluate the vibrational partition function by performing numerical integration instead of discrete summation. The Hamiltonian here includes a

one-dimensional kinetic term of a body with a reduced mass μ and the LJ potential is shifted by ϵ^{LJ} so its minimum is at zero energy. Consequently we get,

$$q_{\text{vib}} = \frac{1}{h}\int_{-\infty}^{\infty} e^{-\beta p^2/2\mu} d\vec{p} \int_0^{r_c} e^{-\beta[U_{LJ}(r)+\epsilon^{LJ}]} dr = \left(\frac{2\pi\mu k_B T}{h^2}\right)^{1/2} \int_0^{r_c} e^{-\beta[U_{LJ}(r)+\epsilon^{LJ}]} dr \quad , \tag{SI-1.7}$$

which can be easily calculated. The values of the different terms of the molecular partition function of the dimer are exhibited in Table SI-1.1.

Table SI-1.1: The values of different modes in the molecular partition function of the dimer, along with the corresponding monoatomic partition function and the Boltzmann's factor, necessary to compute the equilibrium constant of our test system ($V = 216.0\ nm^3$ and $T = 300.0\ K$) via Eq. SI-1.4.

$q^{\text{classical}}_{\text{trans}}(A_2)$	q_{rot}	q_{vib}	$e^{-\beta\epsilon}$	$q^{\text{classical}}_{\text{trans}}(A)$
$1.8866 \cdot 10^7$	90.103	0.36473	48261	$6.6702 \cdot 10^6$

The comparisons between the equilibrium constant, as well as of the standard Gibbs energy change, obtained by the four different (two simulation- and two analytical-) methods is shown in Table SI-1.2. The agreement between the MC simulation and the numerical integration over particles' coordinates (Eq. SI-1.3) is almost perfect. Relative to this, the agreement of K between the MC and MD simulations may seem compromised. However when considering the difference between the corresponding $\Delta G^{\varnothing}$, which equals $0.02\ kJ/mol$, the agreement is still very good, and the mild discrepancy can be attributed to application of a thermostat to a system with small number of degrees of freedom. By far, the largest deviation is observed when the calculation is performed using the molecular partition function (Eq. SI-1.4) where the difference in $\Delta G^{\varnothing}$ with the other methods is in the range $0.06 - 0.09\ kJ/mol$. As we argued before[2], this is not surprising given the several assumptions made in deriving this equation, and in particular, the neglect of coupling between vibrational and rotational modes for a bond formed by a 'soft', intermolecular, potential.

Table SI-1.2: Comparison between values of the equilibrium constant K computed by four different methods, for the dimerization described in Eq. 2 of the simplified model system of single-site monomers detailed in this section. In the two simulation methods, Monte-Carlo (MC) and Molecular Dynamics (MD), K was obtained by calculating the ratio between the product and correlated-reactants concentration according to Eq. 20. The analytical/numerical calculations were based on integration of the particles coordinates (Eq. SI-1.3), as well as on partition functions describing relative motions of a homonuclear diatomic molecule (Eq. SI-1.4). In addition to the value of K, we also display (in kJ/mol) the corresponding change in the standard Gibbs energy, $\Delta G^{\varnothing}$, using the definition in Eq. 5.

	Simulations (Eq. 20)		Analytical/Numerical Calculations	
	MC	MD	Eq. SI-1.3	Eq. SI-1.4
K	90.625 ± 0.005	89.73 ± 0.24	90.623	87.481
$\Delta G^{\varnothing}$	-11.2413 ± 0.0001	-11.217 ± 0.007	-11.2412	-11.1532

SI-2 Limits on the Relation between Reference and Finite Systems

The relation expressed in Eq. 12 between partition functions of the reference state and those of the arbitrary system assumes translational partition functions of monomer and dimer are linearly proportional to the volume. This is true if these translational partition functions can be described 'classically' as considered in Eq. SI-1.5. For macroscopic reference systems this assumption is clearly valid. However, would it also hold for a chosen system that is finite in size, thus, with a small volume?

In obtaining Eq. SI-1.5, quantum translational energy states are actually considered however the discrete sum, that in 1-dimension (along the x-axis) takes the form[1]

$$q_{\text{trans},x} = \sum_{n_x=1}^{\infty} \exp\left[-\beta h^2 n_x^2/(8mL_{box}^2)\right] \tag{SI-2.1}$$

with n_x a positive integer, is approximated by an integral over n_x,

$$q_{\text{trans},x} \approx \int_0^{\infty} \exp\left[-\beta h^2 n_x^2/(8mL_{box}^2)\right] dn_x \qquad . \tag{SI-2.2}$$

Because motion along each axes is independent, the translational partition function in 3-dimensions becomes,

$$q_{\text{trans}} = q_{\text{trans},x} \cdot q_{\text{trans},y} \cdot q_{\text{trans},z} \qquad . \tag{SI-2.3}$$

Approximating Eq. SI-2.1 by Eq. SI-2.2 requires successive terms in the sum to be spaced close enough. In fact, the spacing is constant with a value of an integer unit, nonetheless, it can be small *relative* to the range (width along the n_x axis) of significant terms that are summed. Given the Gaussian form of the terms inside the sum, the condition is that the width $\sigma = \sqrt{(8mL_{box}^2/(\beta h^2))}$ should be much larger than 1. For the single-site monomer system mentioned in Section SI-1 ($m = 10\ amu$, $L_{box} = 6.0\ nm$, and $T = 300\ K$), the value of σ is 212. Although this may be considered a large number compared to 1, we also assess the approximation directly by calculating q_{trans} (Eq. SI-2.3) using the discrete summation of energies as indicated in Eq. SI-2.1. The results are, $q_{\text{trans}(A)} = 6.6171 \cdot 10^6$ and $q_{\text{trans}(A_2)} = 1.8760 \cdot 10^7$, for the monomer and dimer respectively. The corresponding values using the 'classical' translation approximation (Eq. SI-1.5), shown in Table SI-1.1, exhibit relative deviations of $0.6\ \%$ and $0.8\ \%$. As a matter of fact, our aim is to

assess the 'classical' approximation applied to the ratio of the partition functions shown in Eq. 12. We therefore define,

$$R_q = \frac{q_{\text{trans}}(A_2)}{\left[q_{\text{trans}}(A)\right]^2} \quad , \tag{SI-2.4}$$

as well as the corresponding ratio of the 'classical' translational partition functions,

$$R_q^{\text{classical}} = \frac{\overset{\text{classical}}{q_{\text{trans}}(A_2)}}{\left[\overset{\text{classical}}{q_{\text{trans}}(A)}\right]^2} \quad , \tag{SI-2.5}$$

and quantify the relative error by,

$$\Delta = \frac{R_q - R_q^{\text{classical}}}{R_q} \quad , \tag{SI-2.6}$$

which also represents the relative error in determining K. This gives $\Delta = 0.010$, thus an error of $1\ \%$, an acceptable accuracy for many applications.

As apparent from Eq. SI-2.1, besides volume, the width of the translational energy sum is also affected by mass and temperature. To study the effect of these three parameters systematically, we consider the single-site monomer system again and vary each parameter while keeping the other two constants. We then plot Δ as a function of the parameter that is changed and display the results in Fig. SI-2.1. As expected Δ decreases for heavier masses, higher temperatures, and larger volumes.

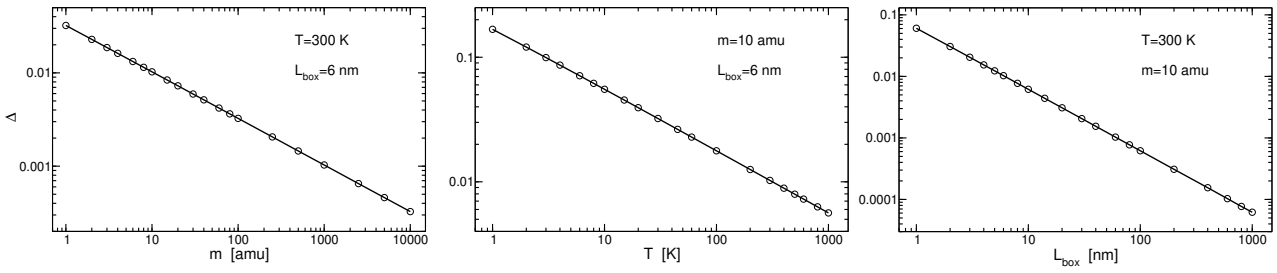

Figure SI-2.1: The relative error, Δ, defined in Eq. SI-2.6, of applying the 'classical' translation approximation to the ratio of the partition functions as a function of mass (left panel), temperature (middle panel), and box-length (right panel). The values of the parameters fixed in each plot correspond to the system defined in Section SI-1.

The smallest mass considered is $1\ amu$ which corresponds to the lightest (i.e., a hydrogen) atom.

For temperature, the lowest value shown is $1\ K$. Although attainable, is unlikely to be of interest for most association/dimerization reactions of molecular systems and higher temperatures are more relevant. The smallest value considered for volume is that which corresponds to $L_{box} = 1\ nm$. Again, this smallest system is not likely to be applicable here because it can not satisfy the ideal gas behavior assumed in the derivation of K. Given the excluded volume of one atom and range of interaction between two atoms, a larger system is required. For the two-site monomer system described in the main text, with a LJ potential acting between particles (i.e., the dispersions decay as $1/r^6$), we found that a box length of $3 - 4\ nm$ is probably the smallest for which ideal gas behavior can be observed. In any case in Fig. SI-2.1, the largest relative error observed is $17\ \%$ (middle panel at $T = 1\ K$) indicating the approximation in this case is not valid.

We now attempt to identify chemical systems for which the 'classical' translation approximation will exhibit the largest deviations. Very low temperatures are crucial, and systems operative under this condition are low molecular weight gases just above their boiling temperature. In Table SI-2.1 we list four gases (helium, hydrogen, neon, and nitrogen) having the lowest boiling points ($4 - 77\ K$). We then consider these gases in a small box, that in our computational experience is already too small to support ideal behavior, and calculate the relative error Δ. What should be considered an acceptable error? Because $\Delta G^{\varnothing}$ is related to K by a natural logarithm, a given error in the value of the latter translates to a much lower error of the former. We therefore propose, arbitrarily, relative errors lower than 0.05 to be acceptable and mark larger errors in table SI-2.1 by red color. For hydrogen gas, only at temperatures higher than $\sim 200\ K$ the 'classical' approximation can be applied, for helium, at temperatures higher than $\sim 100\ K$, whereas for neon and nitrogen, or for any other gas, at any temperature.

Table SI-2.1: The relative error, Δ, defined in Eq. SI-2.6, of applying the 'classical' translation approximation to the ratio of the partition functions for dimerization of hydrogen (H_2), Helium (He), Neon (Ne), and Nitrogen (N_2) gases confined to a cubic box with $L_{box} = 3.0\ nm$, at their corresponding boiling point T_b and at three higher temperatures. Discrepancies with relative magnitude larger than an arbitrary threshold of $5\ \%$ are marked in red.

gas	$m\ [amu]$	$T_b\ [K]$	$\Delta(T = T_b)$	$\Delta(T = 100\ K)$	$\Delta(T = 200\ K)$	$\Delta(T = 300\ K)$
H_2	2.0	20.3	0.17	0.078	0.055	0.045
He	4.0	4.2	0.25	0.055	0.039	0.032
Ne	20.2	27.1	0.047	0.025	0.018	0.014
N_2	28.0	77.4	0.024	0.021	0.015	0.012

References

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

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