Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supplementary Information: Binding Reactions at Finite Systems

Ronen Zangi*[1,2]

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

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

March 4, 2022

*r.zangi@ikerbasque.org

SI-1 Comparisons with Analytical/Numerical Methods

We now compare the value of the equilibrium constant using Eq. 16 to two well-known analytical expressions derived from evaluations of the single-particle, $q_{_A}$ and $q_{_B}$, and pair-particle, $q_{_{AB}}$, partition functions. In the first method, these partition functions are evaluated by integration over the coordinates of the particles, whereas in the second method, $q_{_{AB}}$ is obtained by integrations over the coordinates and momenta of the center-of-mass and relative motions of the bound state.

For the purpose of comparisons, we choose a finite, $N_{_A}^{\circ} = N_{_B}^{\circ} = 1$, model system at $c_{_A}^{\circ} = c_{_B}^{\circ} = 0.00462963\ molecule/nm^3$ (corresponding to $L_{box} = 6.0\ nm$) in which the reference expressions (see below) can be easily calculated analytically or numerically if we describe A and B as single-site particles*, $A \equiv a$ and $B \equiv b$. We also modified the well-depth of the Lennard-Jones potential to $\epsilon_{_{AB}}^{LJ} = 22.15\ kJ/mol$ so that its magnitude is similar to the effective attraction between A and B in the simulations with diatomic monomers described above. All other simulation parameters are unchanged.

MC simulations of 10^{12} trial moves, with same relevant characteristics as described above, were performed to yield an acceptance-ratio of 0.44 and an equilibrium constant, calculated by Eq. 16, of 51.09. In addition two MD simulations, using Nosé-Hoover and velocity-rescaling thermostats, were ran for $48\ \mu s$ and $160\ \mu s$ resulting with a value of K of 50.93 and 52.20, respectively (see Table SI-1.2).

I. K from Integration over Particle's Coordinates

If $\mathcal{T}$ and $\mathcal{U}$ are the kinetic and potential parts of the Hamiltonian, the pair-particle partition function can be written as,

$$q_{_{AB}}(\vec{p_{_A}}, \vec{p_{_B}}, \vec{r_{_A}}, \vec{r_{_B}}) = \frac{1}{h^6} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\beta \mathcal{T}(\vec{p_{_A}}, \vec{p_{_B}})} d\vec{p_{_A}} d\vec{p_{_B}} \int_{\vec{r_{_A}}} d\vec{r_{_A}} \int_{0}^{r_c} e^{-\beta \mathcal{U}(r)} d\vec{r} \quad , \quad \text{(SI-1.1)}$$

where h is Planck's constant and r_c the cutoff distance defining the bound state. The integrals over the momenta of each particle are of three dimensions, as is the integral over $\vec{r_{_A}}$, i.e. over all

*The potential danger of products other than the AB bound state is now removed by restricting the simulations to a system with $N_{_A}^{\circ} = N_{_B}^{\circ} = 1$.

possible coordinates of particle A, which yields V. In Eq. SI-1.1 we assumed the potential energy of the system depends only on the relative distance between A and B particles, $r = |\vec{r_A} - \vec{r_B}|$. Further assuming $\mathcal{U}$ vanishes for $r > r_c$, the single-particle partition function, q_A, is

$$q_A(r) = \frac{1}{h^3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\beta\mathcal{T}(\vec{p_A})} d\vec{p_A} \int_{\vec{r_A}} d\vec{r_A} = \frac{V}{h^3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\beta\mathcal{T}(\vec{p_A})} d\vec{p_A} \quad , \quad \text{(SI-1.2)}$$

and a corresponding expression holds for q_B. In the ratio for K, the integrals over momenta cancel-out and we are left with,

$$K = \frac{q_{AB}V}{q_A \cdot q_B} \cdot c^{\varnothing} = c^{\varnothing}\int_0^{r_c} e^{-\beta\mathcal{U}(r)} d\vec{r} = c^{\varnothing}\int_0^{r_c} e^{-\beta\mathcal{U}(r)} 4\pi r^2 dr \quad . \quad \text{(SI-1.3)}$$

If there had been other degrees of freedom in the system, integratable at fixed values of r, then instead of $\mathcal{U}(r)$ we would have had $w(r)$, the potential of the averaged force acting between A and B due to those other degrees of freedom[1]. Thus, the need for additional simulations to calculate the potential of mean force is avoided here because A and B are mono-atomic particles, and we can solve Eq. SI-1.3 numerically using the Lennard-Jones potential described above. This gives $K = 51.04$.

II. K from a Molecular Partition Function

The Hamiltonian of the pair-particle partition function can also be written in terms of generalized coordinates and momenta describing translation of the center-of-mass, as well as, rotations and vibrations of the bound AB state. If the rotational and vibrational modes are decoupled, the expression of K becomes,

$$K = \frac{q_{\text{trans}(AB)}\ q_{\text{rot}}\ q_{\text{vib}}\ e^{-\beta\epsilon_{AB}}}{q_{\text{trans}(A)}\ q_{\text{trans}(B)}} V c^{\varnothing} \quad , \quad \text{(SI-1.4)}$$

where ϵ_{AB} equals $-\epsilon_{AB}^{LJ}/N_{\text{Avogadro}}$ set above. We use textbooks[2] results for the translational and high-temperature (rigid-rotor) rotational partition functions. These are,

$$q_{\text{trans}} = \left(\frac{2\pi m k_B T}{h^2}\right)^{3/2} V \quad , \quad \text{(SI-1.5)}$$

and,

$$q_{\text{rot}} = \frac{8\pi^2 I k_B T}{h^2} \quad , \quad \text{(SI-1.6)}$$

where m is the mass of the translating body, $I = \mu R_{eq}^2$ is the moment of inertia with μ the reduced mass and $R_{eq} = 2^{1/6}\sigma_{_{AB}} = 0.2020\ nm$, the equilibrium distance between A and B particles in the bound state. We also assume high-temperatures for the vibrational partition function, arising from the oscillatory motion around the minimum of the LJ potential, and perform numerical integration instead of discrete summation. Here, the Hamiltonian includes a one-dimensional kinetic term of a body with a reduced mass μ and the Lennard-Jones potential is shifted by ϵ_{AB}^{LJ} so its minimum is at zero energy. We therefore have,

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

We calculate these different elements of the molecular partition function for the system introduced above and present the results in Table SI-1.1. Inserting these values in Eq. SI-1.4 we obtain

Table SI-1.1: The value of different elements in the molecular partition function of a diatomic gas, along with the corresponding monoatomic partition functions and the Boltzmann's factor, necessary to compute the equilibrium constant in Eq. SI-1.4, at $T = 300\ K$. The quantities q_{trans}/V are given in units of m^{-3}.

$q_{\text{trans}}(AB)/V$	q_{rot}	q_{vib}	$e^{-\beta\epsilon_{AB}}$	$q_{\text{trans}}(A)/V = q_{\text{trans}}(B)/V$
$8.734 \cdot 10^{31}$	252.4	0.4854	7187	$3.088 \cdot 10^{31}$

$K = 48.57$.

In Table SI-1.2 we summarize the results obtained from the simulations, as well as, from the two analytical/numerical methods. The agreement of the MC and MD-NH simulations with the numerical evaluation of Eq. SI-1.3 is excellent. The result of MD-VR is slightly less good where it converges to a different value than that determined by Eq. SI-1.3, nevertheless the discrepancy of $0.056\ kJ/mol$ in $\Delta G^{\varnothing}$ is rather small. A mild discrepancy, relative to the other four results, is also observed when we evaluate K by the molecular partition function using Eq. SI-1.4 with a magnitude that translates to $0.12 - 0.18\ kJ/mol$ for the value of $\Delta G^{\varnothing}$. This is not surprising given the assumptions made in Eq. SI-1.4 and is likely to be the least accurate method. The most questionable assumption is the neglect of coupling between vibrational and rotational degrees of

Table SI-1.2: Comparison between values of the equilibrium constant K computed by five different methods, for the reaction described in Eq. 1 using the simplified system of single-site monomers detailed in this section. Simulations utilizing Monte-Carlo (MC) and two Molecular Dynamics, one with a Nosé-Hoover (MD-NH) and one with a velocity-rescaling (MD-VR) thermostats, methods were performed. In these simulations, K was obtained by calculating the ratio between the product and correlated-reactants concentrations according to Eq. 16. 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 diatomic molecule (Eq. SI-1.4). In addition to the values of K, we also provide (in kJ/mol) the corresponding change in the standard Gibbs energy, $\Delta G^{\varnothing}$, using the definition in Eq. 5.

	Simulations (Eq. 16)			Analytical/Numerical Evaluations	
	MC	MD-NH	MD-VR	Eq. SI-1.3	Eq. SI-1.4
K	51.09 ± 0.06	50.93 ± 0.25	52.20 ± 0.13	51.04	48.57
$\Delta G^{\varnothing}$	-9.812 ± 0.003	-9.804 ± 0.012	-9.865 ± 0.006	-9.809	-9.686

freedom in a system of bound particles held together by an *intermolecular* potential that, for a rigid-rotor approximation, is rather soft.

Comparing the Radial Distribution Functions

We write the total partition function of the system of one A and one B particles based on the way we defined q_{AB} in Eq. SI-1.1 but with an upper bound of the integral over r that includes all possible values of the relative distances between A and B,

$$Q(\vec{p_A}, \vec{p_B}, \vec{r_A}, \vec{r_B}) = \frac{1}{h^6} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\beta \mathcal{T}(\vec{p_A}, \vec{p_B})} d\vec{p_A} d\vec{p_B} \int_{\vec{r_A}} d\vec{r_A} \int_0^{r_{box}} e^{-\beta \mathcal{U}(r)} 4\pi r^2 dr \quad . \tag{SI-1.8}$$

Note that computer simulations often use rectangular-shaped boxes which are not so convenient to integrate by a spherically symmetric coordinate system. This is easy to solve if we recall our assumption that $\mathcal{U}(r)$ vanishes for $r > r_c$, because for these values of r the integrand is 1 and we are integrating only the relative spatial coordinates. Thus all we need to do is to perform the

integration from 0 to r_c and add the remaining volume element, $V - 4\pi r_c^3/3$. Alternatively, we can integrate from 0 to r_{box}, as indicated in Eq. SI-1.8, where we set $r_{box} = (3V/4\pi)^{1/3}$, i.e., substituting the rectangular box with a sphere of the same volume.

The probability density of finding particle B at a distance r from particle A is,

$$P(r) = \frac{\frac{1}{h^6}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{-\beta\mathcal{T}(\vec{p_A},\vec{p_B})}d\vec{p_A}d\vec{p_B}\ V\ e^{-\beta\mathcal{U}(r)}4\pi r^2}{Q} \quad , \tag{SI-1.9}$$

whereas for a random distribution this probability density is,

$$P_{\text{random}}(r) = \frac{4\pi r^2}{V} \quad . \tag{SI-1.10}$$

The radial distribution function is exactly the ratio between these two probabilities,

$$g_{AB}(r)_{N_A^\circ=N_B^\circ=1} = \frac{P(r)}{P_{\text{random}}(r)} = \frac{V}{\int_0^{r_{box}} e^{-\beta\mathcal{U}(r)}4\pi r^2 dr}\ e^{-\beta\mathcal{U}(r)} \quad , \tag{SI-1.11}$$

which can be easily solved numerically given the above mentioned LJ potential. In Fig. SI-1.1 we compare this numerical result to the three different simulations. The MC simulations produce almost identical radial distribution function to that obtained from Eq. SI-1.11. The agreement of MD-NH is again excellent, however, MD-VR displays small but significant discrepancies in line with the slight overestimation of K (Table SI-1.2).

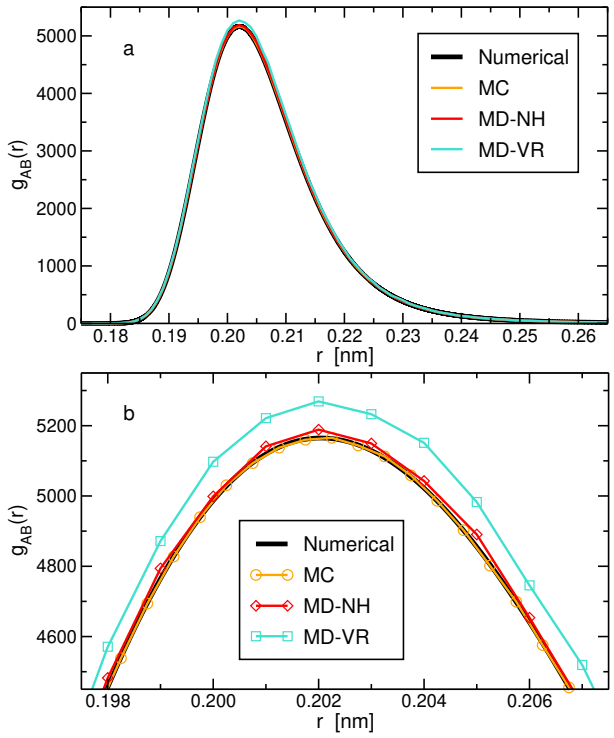

Figure SI-1.1: The radial distribution function between A and B in the single-site monomers model calculated numerically using Eq. SI-1.11, as well as, from the trajectories of the MC, MD-NH, and MD-VR simulations (a). In (b) we magnified a section around the maximum, representing the bound state, and added symbols to the plots of the simulations.

SI-2 Computational Details

The model system consists of two types of molecules where each molecule is represented by two sites, $A \equiv ah$ and $B \equiv bh$, 'covalently' bonded with a bond-length of $0.15\ nm$ as shown schematically in Fig. SI-2.1. The role of the h atoms is to prevent any clustering of the molecules, apart

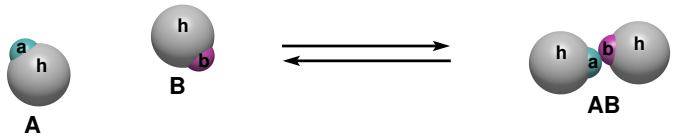

Figure SI-2.1: Simulation model for the association process between A and B molecules to produce the bound state AB. Molecule A and molecule B consist of, uncharged LJ, $a-h$ and $b-h$ atom-sites, respectively. The interaction between a and b is strongly attractive, whereas other intermolecular interactions are repulsive (see Table SI-2.1). Within each molecule, the intramolecular distance between the two atom-sites, having a value of $0.15\ nm$, is either fixed (Monte-Carlo simulations) or held together by a harmonic potential (molecular-dynamics simulations).

from product formation. All atom-sites have zero charge, $q_a = q_b = q_h = 0.0\ e$, and their intermolecular interactions are described by Lennard-Jones (LJ) potentials truncated at a distance of $2.0\ nm$. The different possible σ and ϵ parameters are specified in Table SI-2.1, yielding essentially repulsive interactions between all sites except for a strong attraction between the a and b atoms. This model results in a two-state system of unbound, $A + B$, and bound, AB, gas particles. Based on the location of the first minimum of $g_{ab}(r)$ (see Fig. SI-5.1b), the bound state is defined for $r_{ab} < 0.4\ nm$. We did not encounter any product other than this bound, AB, state in all frames of all simulations.

Periodic boundary conditions were applied along all three Cartesian axes. The total number of A molecules is denoted by $N_A^\circ = N_A + N_{AB}$ and that of B molecules by $N_B^\circ = N_B + N_{AB}$. Three main series of simulations were designed. In the first, labeled R1, we changed the value of $N_A^\circ = N_B^\circ$ from 1 to 4096, and concomitantly, the volume of the cubic simulation box, V, keeping the concentrations, $c_A^\circ = N_A^\circ/V$ and $c_B^\circ = N_B^\circ/V$, constant at $0.015625\ molecules/nm^3$

Table SI-2.1: Lennard-Jones parameters between all atom sites for a system with $A(ah)$ and $B(bh)$ molecules.

	σ $[nm]$	ϵ $[kJ/mol]$
$a\cdots a$	1.00	0.1
$b\cdots b$	1.00	0.1
$h\cdots h$	0.50	0.1
$a\cdots h$	0.35	0.1
$b\cdots h$	0.35	0.1
$a\cdots b$	0.18	30.0

($\sim 0.026\ M$). In the second series of simulations, R2, we considered only one molecule of A, $N_A^\circ = 1$, and one molecule of B, $N_B^\circ = 1$, and increased V by increasing the length of the cubic box from $L_{box} = 4.0\ nm$ to $L_{box} = 28.0\ nm$. The third series of simulations, R3, consisted of asymmetrical concentrations of the A and B molecules, in which $N_A^\circ = 1$ is fixed whereas N_B° varied from 1 to 4096, coupled to changes of V to satisfy $c_B^\circ = 0.015625\ molecules/nm^3$. In order to further examine the validity of the approximation to predict composition from K at finite systems where $N_A^\circ > 1$ (see below) a fourth series of simulations, R4, also with asymmetrical concentrations, was conducted. In this case, $N_B^\circ = 8$ and $V = 512\ nm^3$ (i.e., $c_B^\circ = 0.015625\ molecules/nm^3$) were kept constant whereas N_A° varied from 1 to 8.

All four series of simulations were performed by the Monte-Carlo (MC) technique (coded in-house in double-precision) where the canonical ensemble emerges naturally from the generated configurations[3,4]. The Metropolis acceptance criteria[5] was applied to either accept or reject trial moves. Each trial move is composed of randomly selecting one A and one B molecules which are 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, on each of the molecules separately. Their magnitudes and directions were determined randomly from a uniform distribution with maximum values of $0.4\ nm$ for displacements along each of the Cartesian-axes, 0.1 for $\cos\theta$ when rotating around angle θ ($0 \leq \theta \leq \pi$), and $0.314\ rad$ for rotations around angle ϕ ($0 \leq \phi \leq 2\pi$). These trial moves resulted in acceptance-ratios that varied from

0.313, for the system with the largest $N_A^\circ = N_B^\circ$ in R1, to 0.996, for the system with the largest L_{box} in R2. The number of trial moves applied for each simulation was inversely proportional to the size of the system. More specifically, the equilibration and data collection stages ranged from 10^4 and $1.4 \cdot 10^{12}$ moves, respectively, for the smallest system of $N_A^\circ = N_B^\circ = 1$, to 10^9 and $1.5 \cdot 10^{10}$ moves for the largest system of $N_A^\circ = N_B^\circ = 4096$.

Unless stated differently, the simulations were carried out at $T = 300\ K$. Nonetheless, we also performed the R1 series of simulations at temperatures of $200,\ 250,\ 400,\ 500,\ 600$, and $1200\ K$. Here $N_A^\circ = N_B^\circ$ ranged from 1 to 64, and the number of trial moves for data collection at the lowest two temperatures, $1.5 \cdot 10^{11}$, was three times larger than that at the highest four temperatures. Note that for the system $N_A^\circ = N_B^\circ = 1$, the average number of bound particles at the lowest temperature $(200\ K)$ is 0.997 whereas at the highest temperature $(1200\ K)$ it is 0.003, spanning a wide range of values for the equilibrium constant. This system $(N_A^\circ = N_B^\circ = 1)$ at $1200\ K$ exhibited the largest acceptance-ratio of 0.995 whereas the smallest acceptance-ratio, 0.0034, was recorded at $200\ K$ for the largest four systems. Larger systems at lower temperatures are more difficult to equilibrate and reach convergence. Nevertheless, the results presented here are converged as demonstrated in Fig. SI-2.2 and Table SI-2.2 for the most challenging system.

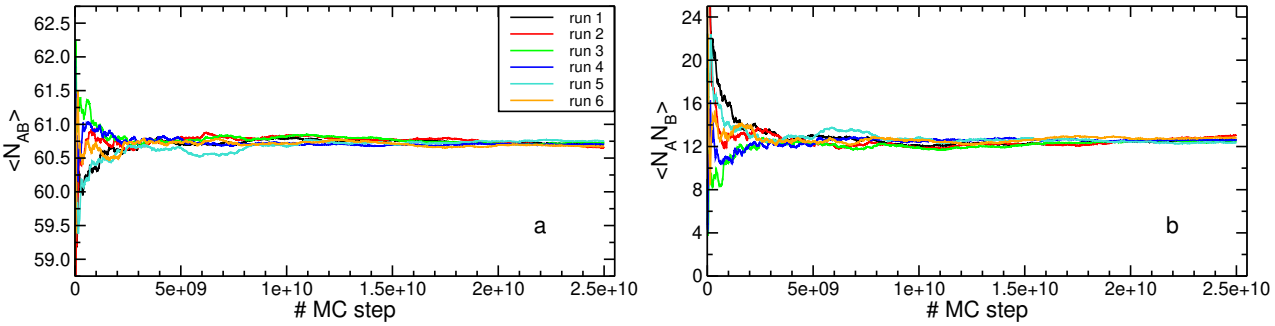

Figure SI-2.2: (a) The average number of bound particles as a function of MC steps for the system of $N_A^\circ = N_B^\circ = 64$ (R1 series) at $T = 200\ K$. Six curves corresponding to six different runs are shown where each spans $2.5 \cdot 10^{10}$ MC steps plotted every $5 \cdot 10^5$ steps. (b) The same as (a) but for the average of the product of the number of unbound A and unbound B particles.

We also attempted simulations at $T = 150\ K$, however, with the number of trial-moves specified

Table SI-2.2: Results obtained from six independent simulations, each with $2.5{\cdot}10^{10}$ MC trial-moves, for the system $N_A^\circ = N_B^\circ = 64$ of R1 series at $T = 200\ K$. The table presents the average number of bound particles, the average of the product between unbound A and unbound B particles, and the variance $\sigma^2 = L(N_{AB}, N_{AB}) = L(N_A, N_B)$ defined in Eq. 21. Unlike Fig. SI-2.2, here all averages are calculated over all MC steps.

Simulation #	$\langle N_{AB} \rangle$	$\langle N_A N_B \rangle$	σ^2
1	60.70	12.60	1.74
2	60.65	13.05	1.83
3	60.73	12.45	1.73
4	60.70	12.59	1.72
5	60.76	12.35	1.83
6	60.67	12.82	1.76

above convergence was not attained and therefore the results were not considered.

Besides MC, we also performed molecular-dynamics (MD) simulations for the R1 and R2 series utilizing the software package GROMACS version 4.6.5[6] (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 all atom-sites. The $a-h$ and $b-h$ 'covalent' bonds were represented by a harmonic potential with bond-length of $0.15\ nm$ and force-constant of $2 \cdot 10^5\ kJ/(mol \cdot nm^2)$. A temperature of $300.0\ K$ was maintained by applying either the Nosé-Hoover[7,8] (MD-NH) or the velocity-rescaling[9] (MD-VR) thermostats. In the first, 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 (see the Gromacs manual). Due to systems with very few degrees of freedom, we applied 10 chained Nose-Hoover thermostats[10] and the coupling strength determining the friction coefficient was set to 0.1. In simulations with the second thermostat, the leap-frog algorithm was used for integrating the equations of motion and the particles' velocities were scaled with a coupling-time of $0.1\ ps$. Note that for the systems described here, MD simulations were less efficient than MC, and therefore, we applied them only to R1 and R2 (up to a box length of $L_{box} = 16.0\ nm$) series of simulations.

Equilibration time of at least $1\ \mu s$ was conducted prior to data collection for each system. For R1, the time period for collecting data ranged from $224\ \mu s$ for the smallest system to $3.84\ \mu s$ for the largest system. For R2, data was collected for $224\ \mu s$.

In order to analyze the dynamics of the forward and backward reactions we performed the R1 and R2 series of simulations by the MD-NH and MD-VR techniques again. However, this time the trajectories were saved more frequently; from a frequency of every 200 steps for $N_A^\circ = 1$ to a frequency of every step for $N_A^\circ = N_B^\circ \geq 16$. 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 also smaller than those described above and ranged from $24\ \mu s$ for the smallest system to $12\ ns$ for the largest system. To keep the size of the trajectories manageable, each run was split into few shorter runs. The rates of the forward and backward reactions were calculated by counting the number of transitions per period of time divided by V. A transition between the two states is identified when the distance r_{ab} crossed the cutoff-value of $0.4\ nm$ plus, or minus, a distance of $0.1\ nm$ on either side of the cutoff (i.e., $0.3\ nm$ for an unbound-to-bound transition and $0.5\ nm$ for the opposite transition) to avoid counting return-trajectories originating from transient species in the proximity of the transition state.

SI-3 Supplementary Figures & Tables

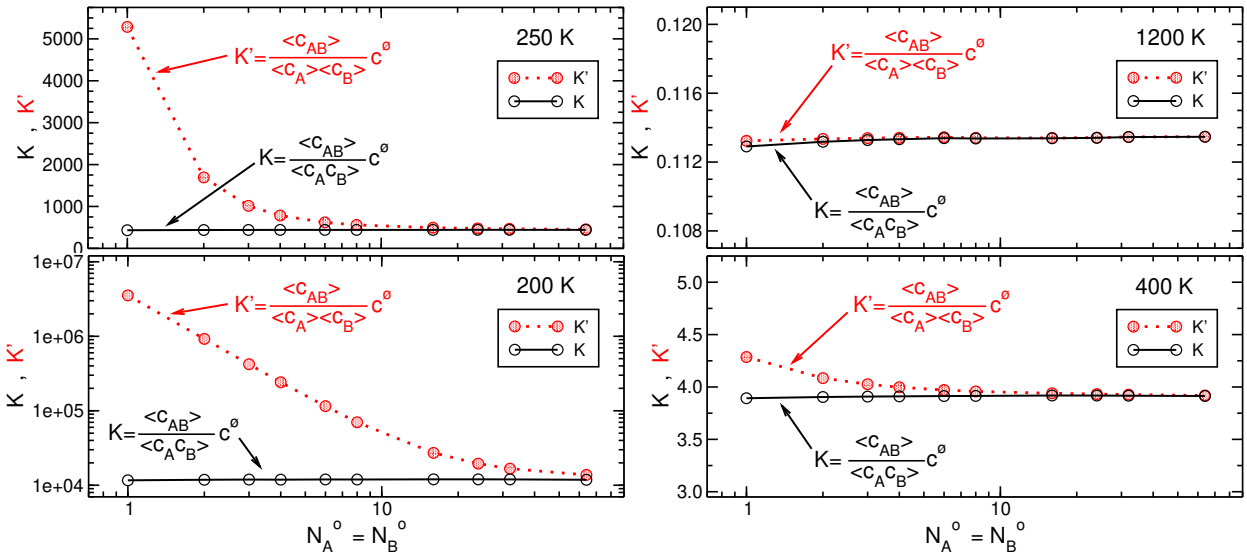

Figure SI-3.1: The equilibrium constant K defined by Eq. 16, as well as the value of K' defined by Eq. 17, from Monte-Carlo R1 series of simulations (i.e., constant $c_{_A}^{\circ} = c_{_B}^{\circ} = 0.026\ M$) at four different temperatures. Here the number of particles, $N_A^{\circ} = N_B^{\circ}$, ranges from 1 to 64. Note the scales of the y-axis are substantially different for the different temperatures and at the lowest temperature, $T = 200\ K$, is not linear but logarithmic.

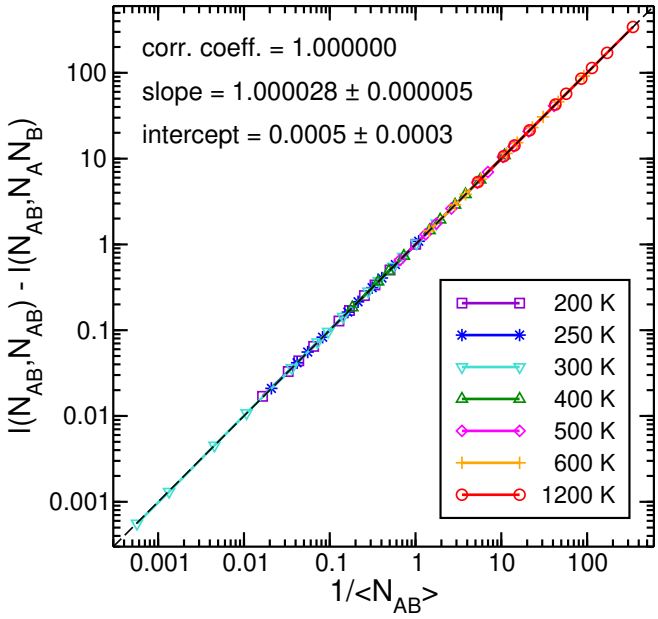

Figure SI-3.2: The difference of the relative correlations, $l(N_{AB}, N_{AB}) - l(N_{AB}, N_A N_B)$, as a function of the reciprocal average of bound AB particles for MC R1 series of simulations at different temperatures. The results at $T = 300\ K$ displayed in Fig. 5a are included here as well as a reference. Linear regression results (obtained by xmgrace) of all data points are indicated. The dashed black line is a $y = x$ line, plotted as a reference for perfect predictions.

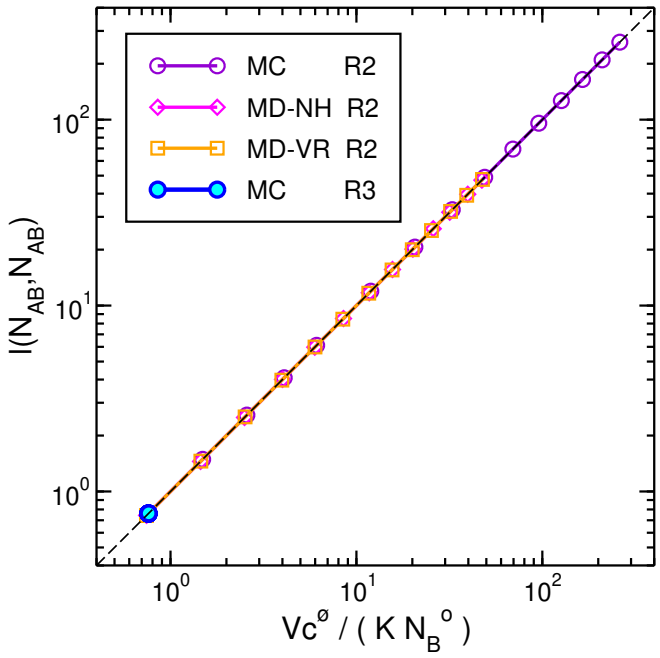

Figure SI-3.3: The expression for predicting relative fluctuations in the number of bound particles, $Vc^{\varnothing}/(KN_B^{\circ})$, for the case $N_A^{\circ} = 1$ ($N_B^{\circ} \geq N_A^{\circ}$) as described in Eq. 33, plotted against the fluctuations themselves for R2 ($N_A^{\circ} = N_B^{\circ} = 1$) and R3 ($N_A^{\circ} = 1$, $c_B^{\circ} = 0.026\ M$) series of simulations. Note that because in R3 series, the ratio V/N_B° is constant, all points in this series have the same value. Linear regression results are presented in Table SI-3.1 below.

Table SI-3.1: Linear-regression analyses (performed by xmgrace) of the predictions of the values of $l(N_{AB}, N_{AB})$ shown in Fig. SI-3.3 above for R2 series of simulations using three different simulation methods.

	Correlation coef.	Slope	Intercept
MC	1.000000	$1.000014 \pm 8 \cdot 10^{-6}$	-0.0001 ± 0.0009
MD-NH	1.000000	$1.00006 \pm 3 \cdot 10^{-5}$	0.0007 ± 0.0006
MD-VR	1.000000	$1.00011 \pm 3 \cdot 10^{-5}$	-0.0006 ± 0.0007

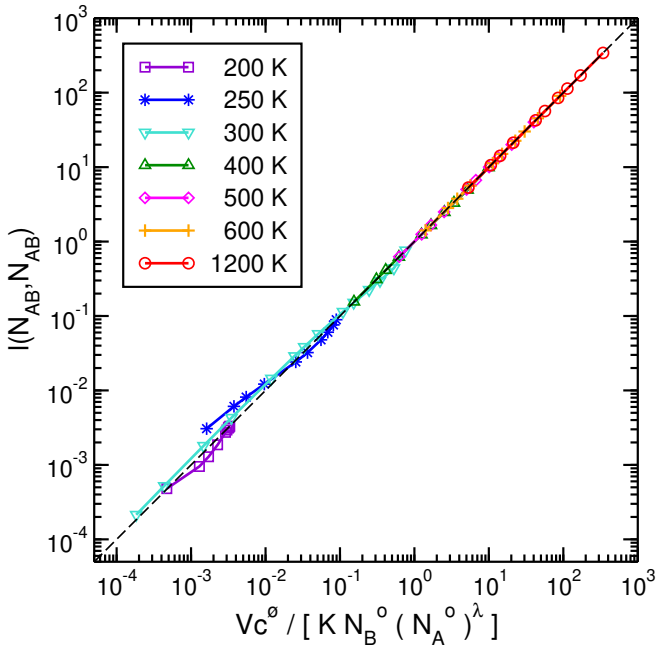

Figure SI-3.4: Approximation results obtained for MC R1 series of simulations at different temperatures. The graph displays the relative fluctuations, $l(N_{AB}, N_{AB})$, as a function of predicted values given by Eq. 34 with $\lambda = [1 + K/(Vc^{\varnothing} \ln N_B^{\circ})]^{-1}$. Linear regression results are presented in Table SI-3.2 below.

Table SI-3.2: Linear-regression analyses of the predictions of the values of $l(N_{AB}, N_{AB})$ shown in Fig. SI-3.4 above at each temperature.

	Correlation coef.	Slope	Intercept
200	0.9882321	1.07 ± 0.06	-0.0003 ± 0.0002
250	0.9948747	0.91 ± 0.03	0.001 ± 0.002
300	0.993245	0.93 ± 0.04	0.001 ± 0.013
400	0.9997421	0.992 ± 0.008	-0.02 ± 0.03
500	0.9999818	0.998 ± 0.002	-0.02 ± 0.03
600	0.9999963	0.999 ± 0.001	-0.03 ± 0.03
1200	0.9999997	0.9998 ± 0.0003	-0.03 ± 0.03

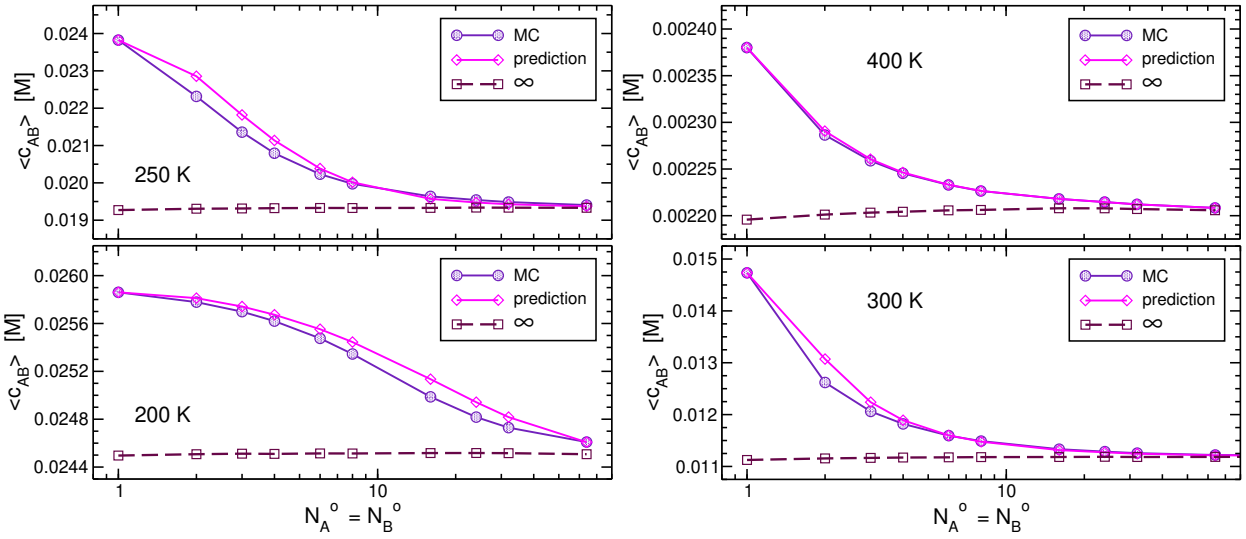

Figure SI-3.5: Concentrations of the bound AB particles calculated by Eq. 30 using approximated predictions for the values of $l(N_{AB}, N_{AB})$ as shown in Fig. SI-3.4, along with the concentrations obtained directly from the MC R1 simulations. The dashed maroon lines are the corresponding values at the thermodynamic limit, $l(N_{AB}, N_{AB}) \rightarrow 0$, calculated by Eq. 31 at each value of $N_A^\circ = N_B^\circ$. For temperatures in the rage $500 - 1200\ K$, the predictions are more accurate than those exhibited at $T = 400K$ (graphs not shown). At $T = 300\ K$, the actual curves end at $N_A^\circ = N_B^\circ = 4096$, however, the last four points are not shown because the predictions obtained are more accurate than that of the last point displayed at $N_A^\circ = N_B^\circ = 64$. At all temperatures, the predictions of the concentrations for $N_A^\circ = N_B^\circ = 1$ are almost identical to those found in the simulations because in this case, the value of the exponent ($\lambda = 0$) given in Eq. 35 makes the expression of $l(N_{AB}, N_{AB})$ in Eq. 34 exact.

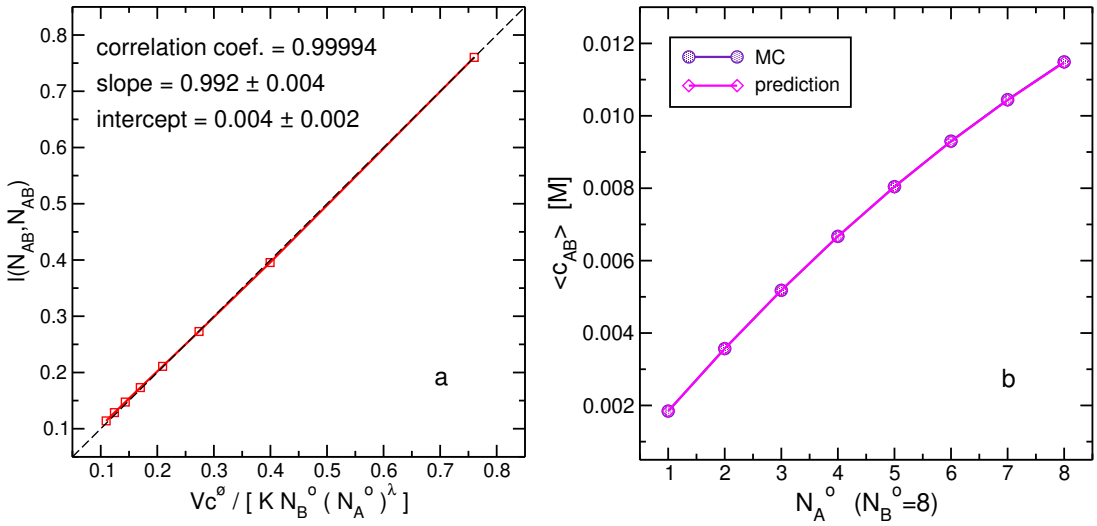

Figure SI-3.6: Approximation results obtained from MC R4 series of simulations. In this series, N_A° and N_B° are not equal and N_A° is not fixed at the value of 1. More specifically, N_A° varied from 1 to 8, whereas $N_B^\circ = 8$, $V = 512\ nm^3$, and $T = 300\ K$ are fixed. (a) The corresponding plot to that of Fig. SI-3.4 and (b) the corresponding plot to Fig. SI-3.5.

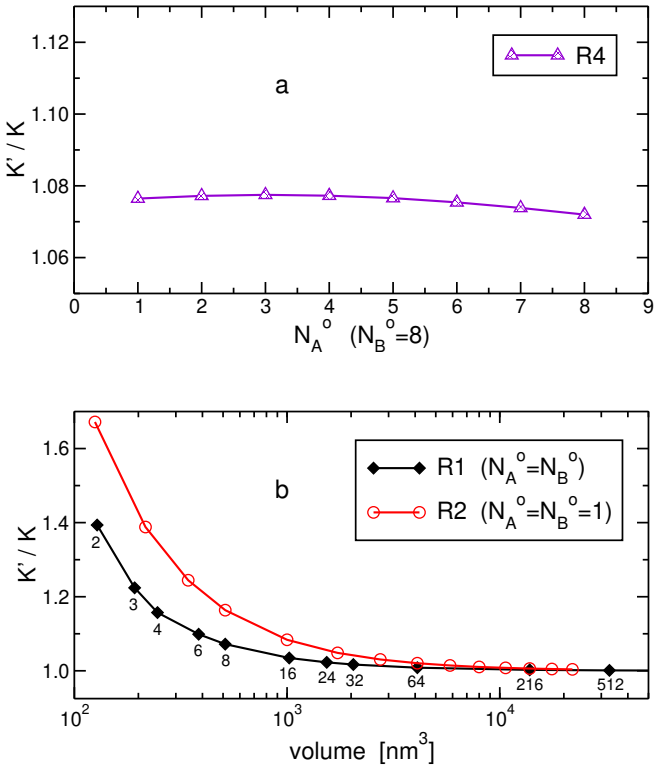

Figure SI-3.7: The ratio between K' and K, which equals $l(N_A, N_B)+1$ and thereby is a measure of correlations between the reactants, from MC simulations at $T = 300\ K$. (a) Results from R4 series where $V = 512\ nm^3$ and $N_B^\circ = 8$ are constants and only N_A° is varied. (b) Resutls from R1 and R2 series as a function of V. In both series $N_A^\circ = N_B^\circ$, however in R2 these numbers equal 1, whereas in R1 their value varies and is indicated below the symbols in the figure.

SI-4 An Alternative Derivation of the Relation between Concentrations and Fluctuations

Given the setup specified in the manuscript, i.e., a system subject to the process described in Eq. 1 in the canonical ensemble (N_A°, N_B°, V, T) where N_A° and N_B° are the total number of A and B particles, satisfying $N_A^\circ \leq N_B^\circ$. We then express the partition function of the system as,

$$Q = \sum_{i=0}^{N_A^\circ} W^i_{N_A^\circ,N_B^\circ} e^{-\beta\mathcal{H}(i)} = \sum_{i=0}^{N_A^\circ} W^i_{N_A^\circ,N_B^\circ} e^{-\beta[\mathcal{T}+\mathcal{U}(i)]} = \sum_{i=0}^{N_A^\circ} W^i_{N_A^\circ,N_B^\circ} e^{-\beta[\mathcal{T}+i\epsilon_{AB}]} \quad , \qquad \text{(SI-4.1)}$$

where as before, we mapped the sum over energy states onto the sum over $i \equiv N_{AB}$, the number of bound AB particles. The Hamiltonian of the system, $\mathcal{H}(i)$, along with its potential energy component, $\mathcal{U}(i)$, are functions of i, whereas the kinetic energy term, $\mathcal{T}$, is not. In the last equality, $\mathcal{U}(i)$ is expressed explicitly by the energy liberated upon the formation of i bound AB particles, and for simplicity we assume no other intra-molecular potential energy terms. The term $W^i_{N_A^\circ,N_B^\circ}$ which corrects the overcounting due to the indistinguishable character of the particles is defined in Eq. 4.

We start by expressing* $\langle N_{AB}^2 \rangle$,

$$\begin{aligned}
\langle N_{AB}^2 \rangle &= \frac{1}{Q}\sum_{i=0}^{N_A^\circ} i^2\, W^i_{N_A^\circ,N_B^\circ} e^{-\beta\mathcal{H}(i)} = -\frac{1}{Q}\frac{1}{\epsilon_{AB}}\left[\frac{\partial}{\partial\beta}\sum_{i=0}^{N_A^\circ} i\, W^i_{N_A^\circ,N_B^\circ} e^{-\beta\mathcal{H}(i)} + \sum_{i=0}^{N_A^\circ} i\,\mathcal{T} W^i_{N_A^\circ,N_B^\circ} e^{-\beta\mathcal{H}(i)}\right] \\
&= -\frac{1}{\epsilon_{AB}}\left[\frac{1}{Q}\frac{\partial}{\partial\beta}\left(\langle N_{AB}\rangle Q\right) + \langle N_{AB}\mathcal{T}\rangle\right] \\
&= -\frac{1}{\epsilon_{AB}}\left[\frac{\partial\langle N_{AB}\rangle}{\partial\beta} + \frac{\langle N_{AB}\rangle}{Q}\frac{\partial Q}{\partial\beta} + \langle N_{AB}\mathcal{T}\rangle\right] \\
&= -\frac{1}{\epsilon_{AB}}\left[\frac{\partial\langle N_{AB}\rangle}{\partial\beta} + \langle N_{AB}\rangle\frac{\partial \ln Q}{\partial\beta} + \langle N_{AB}\mathcal{T}\rangle\right] \\
&= -\frac{1}{\epsilon_{AB}}\left[\frac{\partial\langle N_{AB}\rangle}{\partial\beta} - \epsilon_{AB}\langle N_{AB}\rangle^2 - \langle N_{AB}\rangle\langle\mathcal{T}\rangle + \langle N_{AB}\mathcal{T}\rangle\right] \\
&= -\frac{1}{\epsilon_{AB}}\frac{\partial\langle N_{AB}\rangle}{\partial\beta} + \langle N_{AB}\rangle^2 \quad , \qquad \text{(SI-4.2)}
\end{aligned}$$

*When writing partial derivatives we will omit the specification of the parameters which are kept constant. That means, in our case of the canonical ensemble, partial derivatives with respect to temperature are taken when N_A°, N_B°, and V are constant.

where the last equality is obtained because, by definition, the value of the kinetic energy in the canonical ensemble is constant. Then, the fluctuations in the number of bound AB particles can be expressed by,

$$L(N_{AB}, N_{AB}) = -\frac{1}{\epsilon_{AB}} \frac{\partial \langle N_{AB} \rangle}{\partial \beta} \quad . \tag{SI-4.3}$$

Using the relation in Eq. 16 we write,

$$\begin{aligned} L(N_{AB}, N_{AB}) &= -\frac{1}{\epsilon_{AB}} \frac{\partial \left[K \langle N_A N_B \rangle \right]}{V c^{\varnothing} \partial \beta} = -\frac{K}{\epsilon_{AB} V c^{\varnothing}} \left[\langle N_A N_B \rangle \frac{1}{K} \frac{\partial K}{\partial \beta} + \frac{\partial \langle N_A N_B \rangle}{\partial \beta} \right] \\ &= -\frac{\langle N_{AB} \rangle}{\epsilon_{AB}} \frac{\partial \ln K}{\partial \beta} - \frac{K}{\epsilon_{AB} V c^{\varnothing}} \frac{\partial \langle N_A N_B \rangle}{\partial \beta} \quad . \end{aligned} \tag{SI-4.4}$$

We now evaluate the first partial derivative after the last equality by using the definition of K in Eq. 5,

$$\begin{aligned} \left(\frac{\partial \ln K}{\partial \beta} \right)_V &= -\frac{1}{R} \frac{\partial (\Delta G^{\varnothing}/T)}{\partial \beta} = -\frac{1}{R} \frac{\partial (\Delta F^{\varnothing}/T + V \Delta P^{\varnothing}/T)}{\partial \beta} = -\frac{1}{R} \frac{\partial (\Delta F^{\varnothing}/T)}{\partial \beta} \\ &= \frac{T^2}{N_{\text{Avogadro}}} \frac{\partial (\Delta F^{\varnothing}/T)}{\partial T} = -\frac{\Delta U^{\varnothing}}{N_{\text{Avogadro}}} = -\epsilon_{AB} \quad . \end{aligned} \tag{SI-4.5}$$

The third equality in Eq. SI-4.5 holds for ideal gases, $V \Delta P^{\varnothing}/T = R \Delta n^{\varnothing}$ (for reactions described by Eq. 1, the change in the number of moles of gas particles under standard conditions, $\Delta n^{\varnothing}$, equals 1) and for reactions in solution where the change in pressure is negligible, $V \Delta P^{\varnothing} \simeq 0$. It is worth pointing that Eq. SI-4.5 is the equivalent of the van't Hoff relation, which is applicable at constant pressure, to processes at constant volume.

Next, we evaluate the second partial derivative after the last equality in Eq. SI-4.4,

$$\begin{aligned} &\frac{\partial \langle N_A N_B \rangle}{\partial \beta} = \frac{\partial}{\partial \beta} \left[\frac{1}{Q} \sum_{i=0}^{N_A^\circ} (N_A^\circ - i)(N_B^\circ - i) \; W^i_{N_A^\circ, N_B^\circ} e^{-\beta \mathcal{H}(i)} \right] \\ &= -\frac{1}{Q} \sum_{i=0}^{N_A^\circ} \mathcal{H}(i) N_A(i) N_B(i) \; W^i_{N_A^\circ, N_B^\circ} e^{-\beta \mathcal{H}(i)} - \frac{1}{Q^2} \frac{\partial Q}{\partial \beta} \sum_{i=0}^{N_A^\circ} N_A(i) N_B(i) \; W^i_{N_A^\circ, N_B^\circ} e^{-\beta \mathcal{H}(i)} \quad , \end{aligned} \tag{SI-4.6}$$

where $N_A(i) = (N_A^\circ - i)$ designates the number of unbound A particles, and a corresponding

notation, $N_B(i)$, designates the unbound B particles. We continue,

$$\begin{aligned}\frac{\partial\langle N_A N_B\rangle}{\partial\beta} &= -\langle\mathcal{H}N_A N_B\rangle - \frac{\partial \ln Q}{\partial\beta}\langle N_A N_B\rangle = -\langle(\mathcal{T}+\mathcal{U})N_A N_B\rangle + \langle\mathcal{H}\rangle\langle N_A N_B\rangle \\ &= -\langle\mathcal{T}N_A N_B\rangle - \langle\epsilon_{AB}N_{AB}N_A N_B\rangle + \langle\mathcal{T}+\epsilon_{AB}N_{AB}\rangle\langle N_A N_B\rangle \\ &= -\epsilon_{AB}\langle N_{AB}N_A N_B\rangle + \epsilon_{AB}\langle N_{AB}\rangle\langle N_A N_B\rangle = -\epsilon_{AB}L(N_{AB}, N_A N_B) \quad . \quad \text{(SI-4.7)}\end{aligned}$$

Again, because the kinetic energy in the canonical ensemble is constant, the two terms containing the value of $\mathcal{T}$ cancel each other. Now we take the results obtained in Eq. SI-4.5 and Eq. SI-4.7 and insert them into Eq. SI-4.4 to calculate $L(N_{AB}, N_{AB})$,

$$L(N_{AB}, N_{AB}) = \langle N_{AB}\rangle + \frac{K}{Vc^{\varnothing}}L(N_{AB}, N_A N_B) = \langle N_{AB}\rangle + \frac{\langle N_{AB}\rangle}{\langle N_A N_B\rangle}L(N_{AB}, N_A N_B) \quad . \quad \text{(SI-4.8)}$$

If we divide both sides of Eq. SI-4.8 by $\langle N_{AB}\rangle^2$ we can express a relation between two relative deviations as,

$$l(N_{AB}, N_{AB}) = \frac{1}{\langle N_{AB}\rangle} + l(N_{AB}, N_A N_B) \quad , \quad \text{(SI-4.9)}$$

which is identical to Eq. 27.

SI-5 Transformations between g(r) of Systems with Different Sizes

As demonstrated in Fig. 4c, the radial distribution function of the product, $g_{ab}(r)$, depends on the system size even if the total concentrations of the A and B particles (c_A° and c_B°) are not altered. Obviously, this is because the equilibrium concentrations do depend on the size of the system. However, because we can predict $\langle c_{_{AB}} \rangle$ for a macroscopic system from a finite-system (Eq. 31), we can perform the corresponding transformation for $g_{ab}(r)$. If the formation of trimers can be ignored, as we actively prevented in our model, the transformation of $g_{ab}(r)$ for the bound state, i.e. for distances around the first-minumum and lower, $r < r_{_{fm}}$, can be performed by using the ratio of the concentrations as a scaling-factor,

$$g_{ab}(r)_\infty = g_{ab}(r)_{finite} \cdot \frac{\langle c_{_{AB}} \rangle_\infty}{\langle c_{_{AB}} \rangle_{finite}} \qquad \text{for} \quad r < r_{_{fm}} \qquad . \tag{SI-5.1}$$

On the other hand, the scaling-factor for larger distances, $r \geq r_{_{fm}}$, is different. To obtain it, we calculate the probability of finding a and b sites at distances $r \geq r_{_{fm}}$ apart,

$$P_{ab}(r \geq r_{_{fm}}) = \frac{N_A^\circ N_B^\circ - \langle N_{AB} \rangle}{N_A^\circ N_B^\circ} = 1 - \frac{\langle N_{AB} \rangle}{N_A^\circ N_B^\circ} \qquad , \tag{SI-5.2}$$

where we subtracted in the numerator the average number of bound particles from the overall possible number of pairs. We consider this probability, for both, finite and macroscopic systems. For the latter case, given $-\beta\epsilon_{_{AB}}$ is not too large, we have $P_{ab}(r \geq r_{_{fm}})_\infty \to 1$, thus $g_{ab}(r)_\infty$ for distances larger than the first-minimum, $r \geq r_{_{fm}}$, can be obtained by,

$$g_{ab}(r)_\infty = g_{ab}(r)_{finite} \cdot \frac{P_{ab}(r \geq r_{_{fm}})_\infty}{P_{ab}(r \geq r_{_{fm}})_{finite}} = g_{ab}(r)_{finite} \cdot \frac{1}{1 - \frac{\langle N_{AB} \rangle_{finite}}{(N_A^\circ N_B^\circ)_{finite}}} \qquad \text{for} \quad r \geq r_{_{fm}} \qquad . \tag{SI-5.3}$$

This conversion of $g_{ab}(r)_{finite}$ obtained at a finite system to that of a macroscopic system is demonstrated in Fig. SI-5.1 utilizing Eq. 31 to calculate $\langle c_{_{AB}} \rangle_\infty$. Although the region describing the bound state and the unbound state are very well reproduced, the transition region, not surprisingly, is not. In addition in this transition region, the conversion from the system of $N_A^\circ = N_B^\circ = 1$ exhibits larger deviations compare to those from any other finite systems. Plausibly because this system, $N_A^\circ = N_B^\circ = 1$, is the only one that does not contain the pure repulsion between the like-type sites (thus between $a \cdots a$ or between $b \cdots b$ sites).

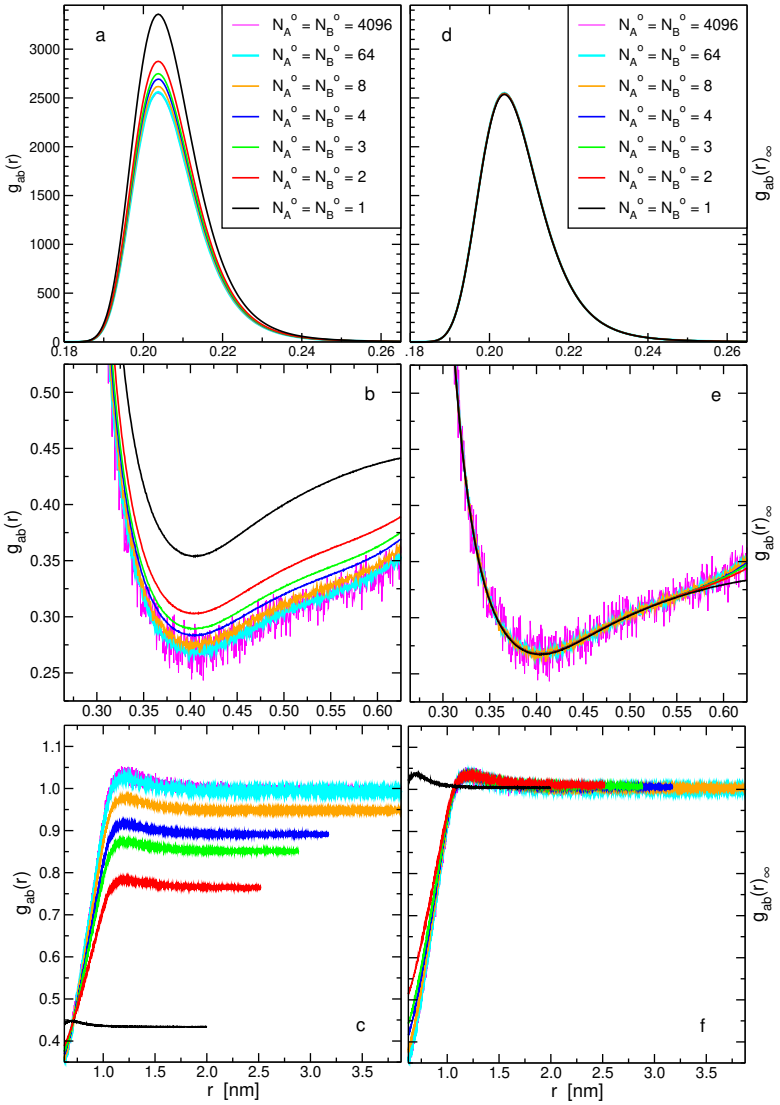

Figure SI-5.1: Transformations of the radial distribution functions, $g_{ab}(r)$ (of R1 MC simulations, also displayed in Fig. 4c), obtained at different system sizes and shown on the left panel (a)-(c), to a corresponding distribution of a system with an infinite-size, $g_{ab}(r)_\infty$, shown on the right panel (d)-(f). The segment of the distribution up to around the minimum defining the bound state, $r < 0.625\ nm$ thus (a) and (b), is converted by applying the ratio of the bound-state concentrations in the two systems as the scaling factor (Eq. SI-5.1). The segment of the distribution with larger values of r, (c), is converted according to Eq. SI-5.3. These transformations break-down in the range, $0.57\ nm < r < 1.00\ nm$, whereas for $N_A^\circ = 1$ it is not valid for a wider range, up to $r \sim 1.5\ nm$. The x-axes in (a) and (d) end at a point where the x-axes of (b) and (e) start, and the latter end at a point where the x-axes of (c) and (f) start.

References

[1] W. G. McMillan and J. E. Mayer, *J. Chem. Phys.*, 1945, **13**, 276–305.

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

[3] M. P. Allen and D. J. Tildesley, *Computer Simulations of Liquids*, Oxford Science Publications, Oxford, 1987.

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

[5] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, *J. Chem. Phys.*, 1953, **21**, 1087–1092.

[6] B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435–447.

[7] S. Nosé, *J. Chem. Phys.*, 1984, **81**, 511–519.

[8] W. G. Hoove, *Phys. Rev. A*, 1985, **31**, 1695–1697.

[9] G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.

[10] G. J. Martyna, M. L. Klein and M. Tuckerman, *J. Chem. Phys.*, 1992, **97**, 2635–2643.