Stefano Angioletti-Uberti, Patrick Varilly, Bortolo M. Mognetti, Alexei V. Tkachenko, and Daan Frenkel

Additional information on J. Chem. Phys.

Communication: A simple analytical formula for the free energy of ligand–receptor-mediated interactions

Stefano Angioletti-Uberti,[1] Patrick Varilly,[1] Bortolo M. Mognetti,[1] Alexei V. Tkachenko,[2] and Daan Frenkel[1]

[1]*Department of Chemistry, University of Cambridge, Lensfield Road, CB2 1EW Cambridge, United Kingdom*
[2]*Center for Functional Nanomaterials, Brookhaven National Laboratory, Upton, New York 11973, USA*

Recently [P. Varilly, S. Angioletti-Uberti, B. M. Mognetti, and D. Frenkel, "A general theory of DNA-mediated and other valence-limited colloidal interactions," J. Chem. Phys. **137**, 094108 (2012)], we presented a general theory for calculating the strength and properties of colloidal interactions mediated by ligand–receptor bonds (such as those that bind DNA-coated colloids). In this Communication, we derive a surprisingly simple analytical form for the interaction free energy, which was previously obtainable only via a costly numerical thermodynamic integration. As a result, the computational effort to obtain potentials of interaction is significantly reduced. Moreover, we can gain insight from this analytic expression for the free energy in limiting cases. In particular, the connection of our general theory to other previous specialised approaches is now made transparent. This important simplification will significantly broaden the scope of our theory. © *2013 American Institute of Physics*. [http://dx.doi.org/10.1063/1.4775806]

I. INTRODUCTION

We consider a general system of many linkers, such as a solution of colloids coated with DNA strands that are capped with reactive sticky-ends. At any given time, each linker i can bind at most one other distinct linker j, with a free energy change ΔG_{ij} that depends on the polymer statistics of the linkers (e.g., length, flexibility, and grafting position). In many cases, including those of experimental relevance, the probability that linker i is unbound is approximately independent of whether or not any other linker is also unbound. Here, we show that in this limit, the free-energy of interaction of the system is given by

$$\beta F_{att} = \sum_i \ln p_i + \sum_{i<j} p_{ij},\qquad(1)$$

where p_i is the probability that linker i is unbound and p_{ij} is the probability that linkers i and j form a bond. Previously,[1] we showed that these quantities are given by the unique physical solution to the following set of self-consistent equations:

$$p_{ij} = p_i p_j e^{-\beta \Delta G_{ij}},\qquad(2)$$

$$p_i = 1 - \sum_j p_{ij}.\qquad(3)$$

In what follows, we first motivate the free energy expression in Eq. (1) through a calculation that closely resembles that for mixing entropy of solutions and gases. This free energy is minimised for the values of $\{p_i\}$ and $\{p_{ij}\}$ that solve the self-consistent conditions in Eqs. (2) and (3). We then show that the free energy in Eq. (1) is identical to that obtained through the costly and numerical thermodynamic integration previously proposed. In Sec. II, we compare the performance of Eq. (1) with that of the thermodynamic integration, establish the explicit connection between our

Eq. (1) and a previous treatment of DNA-mediated colloid interactions,[2] and state the analogous result to Eq. (1) for the mean-field system of plates discussed in Ref. 1.

II. DERIVATION OF THE MAIN RESULT

We consider here an ensemble of N independent copies of the real system. In each copy, a different set of bonds forms between the linkers (see Figure 1). Let N_i be the number of copies where linker i is unbound, and let N_{ij} be the number of copies where i and j are bound to each other. Conversely, let $N_{-i,-j}$ be the number of copies where both i and j are unbound. These quantities are not independent: each linker i is either bound or unbound, so

$$N_i + \sum_j N_{ij} = N.\qquad(4)$$

The fraction of copies where i is unbound (f_i), where i and j are bound to each other (f_{ij}), or where they are both unbound ($f_{-i,-j}$) follow immediately:

$$f_i = N_i/N,\qquad(5)$$

$$f_{ij} = N_{ij}/N,\qquad(6)$$

$$f_{-i,-j} = N_{-i,-j}/N.\qquad(7)$$

Let $Z(\{N_{ij}\})$ be the partition function of an ensemble under the constraint that each pair of linkers i and j is bound to each other in exactly N_{ij} copies. A closed-form expression for $Z(\{N_{ij}\})$ can be constructed recursively, by adding each bond one by one. For a given set of $\{N_{ij}\}$, we need to work out how $Z(\{N_{ij}\})$ changes upon adding one more i–j bond, which we do as follows. We call the set of realisations of the ensemble with $\{N_{ij}\}$ bonds the *old* ensemble, and that of realisations with one more i–j bond, the *new* ensemble. In the old ensem-
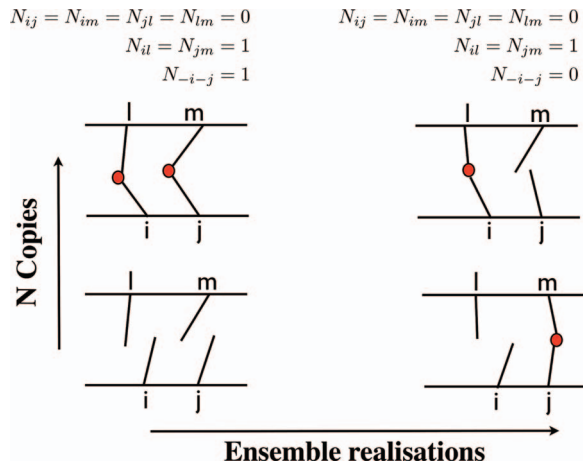
FIG. 1. Two different realisations of an ensemble of two copies. Linkers are depicted as straight lines, and bonds are shown as filled circles. Although the numbers of bonds formed in each ensemble, $\{N_{ij}\}$, are equal, the number of copies where both i and j are unbound, N_{-i-j}, differs.

ble, there are $N_{-i,-j}$ copies where an i–j bond can be added. In doing this, all the realisations of the new ensemble are generated, but not uniquely. For example, given two realisations, one with an i–j bond in copy X but not in Y or Z, and one with an i–j bond in Y but not in X or Z, the same final realisation can be obtained by adding an i–j bond to Y in the former and X in the latter. Conversely, in the new ensemble, we can generate a realisation of the old ensemble in $N_{ij} + 1$ ways by removing one of the i–j bonds. For example, the old realisation with an i–j bond in copy X but not Y or Z can be obtained by deleting the i–j bond from a new realisation with an i–j bond in X and Y but not Z, or from one with an i–j bond in X and Z but not Y. Since the number of ways of going from the old to the new ensemble is equal to the number of ways of going from the new to the old ensemble, it follows that

$$Z(\ldots, N_{ij}, \ldots) N_{-i,-j} e^{-\beta \Delta G_{ij}}$$
$$= Z(\ldots, N_{ij} + 1, \ldots)(N_{ij} + 1). \quad (8)$$

The value of $N_{-i,-j}$ depends not just on the values of $\{N_{ij}\}$ but on the details of how those bonds are distributed between copies (see Figure 1). To remove this complication, we approximate the probability of j being unbound as independent of whether or not i is also bound. Hence,

$$N_{-i,-j} = N f_{-i,-j} \approx N f_i f_j = \frac{N_i N_j}{N}. \quad (9)$$

Neatly, this approximation allows us to treat $N_{-i,-j}$ as a function of only $\{N_{ij}\}$. From the discussion above, we obtain an expression for the increase in $Z(\{N_{ij}\})$ upon adding one i–j bond to the system:

$$\frac{Z(\ldots, N_{ij} + 1, \ldots)}{Z(\ldots, N_{ij}, \ldots)} \approx \frac{e^{-\beta \Delta G_{ij}} N_i N_j}{N(N_{ij} + 1)}$$
$$= \frac{e^{-\beta \Delta G_{ij}} \left(N - \sum_k N_{ik} \right) \left(N - \sum_k N_{jk} \right)}{N(N_{ij} + 1)}. \quad (10)$$

This recursion relation, and the fact that $Z = 1$ when no bonds form, allows us to write an approximate closed-form expres-

sion for $Z(\{N_{ij}\})$, namely,

$$Z(\{N_{ij}\}) \approx \prod_i \frac{N!}{\left(N - \sum_k N_{ik} \right)!} \cdot \frac{1}{N^{\sum_{i<j} N_{ij}}} \cdot \prod_{i<j} \frac{e^{-\beta N_{ij} \Delta G_{ij}}}{N_{ij}!}. \quad (11)$$

Using Stirling's approximation and Eq. (11), the free energy per copy $\beta f_{\text{att}}^* = -(1/N) \ln Z(\{N_{ij}\})$ is then given by

$$\beta f_{\text{att}}^*(\{f_{ij}\}) = \sum_{i<j} f_{ij} \beta \Delta G_{ij} + \sum_i \left(1 - \sum_j f_{ij} \right)$$
$$\times \ln \left(1 - \sum_k f_{ik} \right) + \sum_{i<j} f_{ij} \ln f_{ij} + \sum_{i<j} f_{ij}. \quad (12)$$

Treating $\{f_{ij}\}$ as continuous in the range $[0, 1]$, the overall free energy per copy of the ensemble, F_{att}^*, follows from a saddle-point approximation:

$$\beta F_{\text{att}}^* \equiv -\frac{1}{N} \ln \left[\int \left(\prod_{i<j} N \, \mathrm{d} f_{ij} \right) e^{-N \beta f_{\text{att}}^*(\{f_{ij}\})} \right]$$
$$= \beta f_{\text{att}}^*(\{\overline{f_{ij}}\}) + \mathcal{O}(\ln N / N) \approx \beta f_{\text{att}}^*(\{\overline{f_{ij}}\}), \quad (13)$$

where the integration is over all positive values of $\{f_{ij}\}$ satisfying $\sum_j f_{ij} \leq 1$ for all i and the values $\{\overline{f_{ij}}\}$ are obtained by minimising the free energy per copy,

$$\left. \frac{\partial \beta f_{\text{att}}^*}{\partial f_{ij}} \right|_{\{\overline{f_{ij}}\}} = 0. \quad (14)$$

When $N \to \infty$, the values $\{\overline{f_{ij}}\}$ are precisely the average values of $\{f_{ij}\}$. Equation (14) implies that $\{\overline{f_{ij}}\}$ and $\{p_{ij}\}$ obey identical equations (Eq. (2)), and so are equal:

$$\overline{f_{ij}} = p_{ij}. \quad (15)$$

Connection to thermodynamic integration

The free energy F_{att}^*, defined in Eq. (13), is equal to the free energy of the real system to the extent that the approximation in Eq. (9) is valid. Since this is the same approximation that we used previously[1] to calculate the free energy in terms of a thermodynamic integral, F_{att}, it is reasonable to suppose that F_{att}^* and F_{att} are equal. We now show this explicitly.

In our original paper, we calculated the exact attractive free energy for the real system of linkers using thermodynamic integration. Specifically, we replaced $\beta \Delta G_{ij}$ by $\beta \Delta G_{ij} + \lambda$, whereupon the probabilities $\{p_i\}$ and $\{p_{ij}\}$ become functions of λ. We then integrated the appropriate free energy derivative over the range $0 \leq \lambda < \infty$, and obtained

$$\beta F_{\text{att}} = -\int_0^\infty \mathrm{d}\lambda \, \frac{\mathrm{d}\beta F_{\text{att}}}{\mathrm{d}\lambda} = -\int_0^\infty \mathrm{d}\lambda \sum_{i<j} p_{ij}(\lambda). \quad (16)$$

The same replacement of $\beta \Delta G_{ij}$ with $\beta \Delta G_{ij} + \lambda$ can be made in the ensemble of N copies. In that case, using Eqs. (12) and (13), we find that

$$\frac{\mathrm{d}\beta F_{\text{att}}^*}{\mathrm{d}\lambda} = \sum_{i<j} \left. \frac{\partial \beta f_{\text{att}}^*}{\partial f_{ij}} \right|_{\{p_{ij}\}} \frac{\mathrm{d} p_{ij}}{\mathrm{d}\lambda} + \left. \frac{\partial \beta f_{\text{att}}^*}{\partial \lambda} \right|_{\{p_{ij}\}}. \quad (17)$$

The first term vanishes owing to Eqs. (14) and (15), and the second term follows immediately from Eq. (12). We then have

$$\frac{\mathrm{d}\beta F_{\mathrm{att}}^{*}}{\mathrm{d}\lambda} = \sum_{i<j} p_{ij} = \frac{\mathrm{d}\beta F_{\mathrm{att}}}{\mathrm{d}\lambda}. \tag{18}$$

Moreover, both F_{att}^{*} and F_{att} are zero when λ is infinite, so the two quantities are equal for all λ.

The previous result, together with Eqs. (12)–(15), yields the following closed-form expression for F_{att}:

$$\beta F_{\mathrm{att}} = \sum_{i<j} p_{ij}\beta\Delta G_{ij} + \sum_{i}\left(1 - \sum_{j} p_{ij}\right)\ln\left(1 - \sum_{k} p_{ik}\right)$$
$$+ \sum_{i<j} p_{ij}\ln p_{ij} + \sum_{i<j} p_{ij}. \tag{19}$$

This expression is, in fact, equivalent to the much more compact Eq. (1). Concretely,

$$\beta F_{\mathrm{att}} = \sum_{i<j} p_{ij}\beta\Delta G_{ij} + \sum_{i}\left(1 - \sum_{j} p_{ij}\right)\ln p_i$$
$$+ \sum_{i<j} p_{ij}\ln(p_i p_j e^{-\beta\Delta G_{ij}}) + \sum_{i<j} p_{ij}, \tag{20a}$$

$$= \sum_{i}\left(1 - \sum_{j} p_{ij}\right)\ln p_i + \frac{1}{2}\sum_{i,j} p_{ij}\ln p_i$$
$$+ \frac{1}{2}\sum_{i,j} p_{ij}\ln p_j + \sum_{i<j} p_{ij}, \tag{20b}$$

$$= \sum_{i}\left(1 - \sum_{j} p_{ij}\right)\ln p_i + \sum_{i,j} p_{ij}\ln p_i + \sum_{i<j} p_{ij}, \tag{20c}$$

$$= \sum_{i}\ln p_i + \sum_{i<j} p_{ij}. \tag{20d}$$

III. DISCUSSION

Figure 2 reports the typical computational speedups obtained by using Eq. (1) versus our original thermodynamic
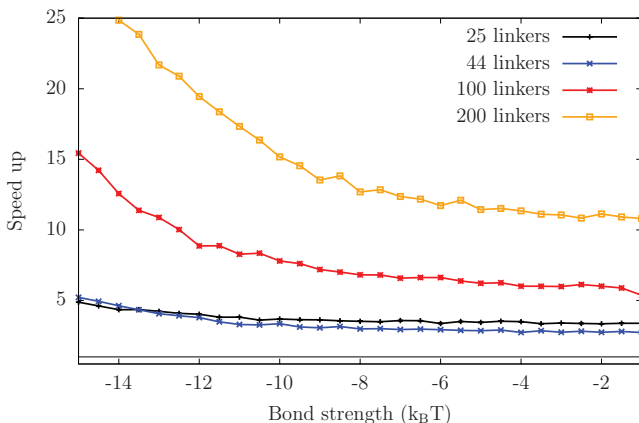

FIG. 2. Computational speedups obtained by using Eq. (1) vs. Eq. (16), for a system of parallel plates coated with complementary linkers.

integral, Eq. (16), for systems of M linkers. The speedup is higher for larger M and for stronger bonds because each evaluation of the thermodynamic integrand involves solving a system of M equations, and the size of the integration domain scales linearly with the bond strength. Typically, experimentally relevant regimes deal with hundreds to tens of thousands of strands, a regime that can now be treated exactly with Eq. (1).

In the limit of weak-bonds, where p_{ij} is close to 0, we find that

$$\beta F_{\mathrm{att}} = \sum_{i}\ln p_i + \sum_{i<j} p_{ij} = \sum_{i}\ln\left(1 - \sum_{j} p_{ij}\right) + \sum_{i<j} p_{ij},$$
$$\approx -\sum_{i,j} p_{ij} + \sum_{i<j} p_{ij} = -\sum_{i<j} p_{ij}.$$

This approximate result has been widely used by previous authors[2–5] under the name of the "Poisson approximation" or the "weak binding regime." However, in experiments with micron-sized DNA-coated colloids, this approximation can be significantly inaccurate.[6] At the nanoscale, where high bond strengths are commonly used, the Poisson approximation is expected to break down (see Figure 3 for comparison). Equation (1) is instead quantitatively accurate for bonds of any strength.[1]

Equations (1)–(3) also make explicit the connection between our theory and previous treatments. For example, Dreyfus *et al.*[2] model the attraction between two DNA-coated spheres by first estimating the maximum number N_p of linkers on each sphere that could form a bond with a linker on the second sphere, and then assuming that each such linker can independently bind any of k linkers with an average free energy $\Delta F_{\mathrm{tether}}$. For later convenience, we define a small expansion parameter x as

$$x \equiv k e^{-\beta\Delta F_{\mathrm{tether}}}, \tag{21}$$

and note that the experiments in Ref. 2 take place mostly in what the author call the "weak-binding regime," where
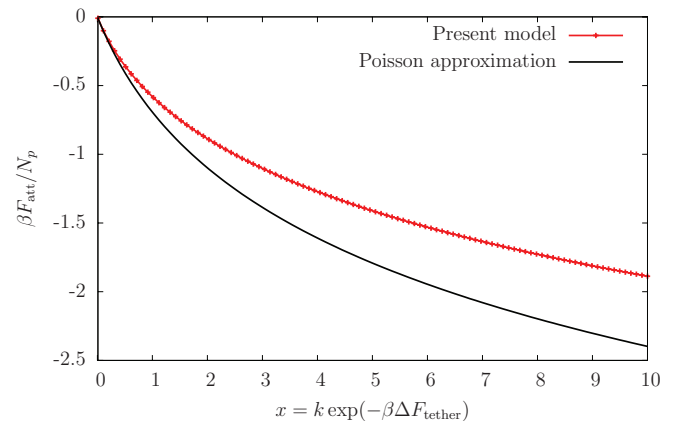

FIG. 3. Free energy per linker in the "Poisson approximation" and the present model. Higher values of $x = k\exp(-\beta\Delta F_{\mathrm{tether}})$ lead to higher bonding probabilities, either because bonds are stronger or because linkers have more binding partners. The two models agree in the "weak binding regime" ($x \ll 1$), but disagree when correlations between neighbouring strands become significant.

$x \ll 1$ and $N_p x \gg 1$. In their model, the partition function of the system is

$$Z \approx (1+x)^{N_p}, \tag{22}$$

from which the free energy F_{att} follows:[7]

$$\beta F_{att} \approx -\ln[(1+x)^{N_p}] = -N_p \ln(1+x) \tag{23}$$

$$\approx -N_p(x - x^2/2 + \cdots). \tag{24}$$

In the present framework, which does not treat the linkers as binding independently, every linker has the same probability p of being bound, given by the solution to

$$p = \frac{1}{1+xp} \Rightarrow p = \frac{\sqrt{1+4x}-1}{2x}. \tag{25}$$

The free energy then follows from Eqs. (1) and (2):

$$\beta F_{att} = N_p p^2 x + 2N_p \ln p \tag{26}$$

$$\approx -N_p(x - x^2 + \cdots). \tag{27}$$

Thus, our theory recovers the results of Dreyfus *et al.*[2] in the weak binding regime. However, there is significant disagreement already at second order in x, where linkers begin to compete for binding partners.

Finally, using the same procedure as in our original paper, we can directly write an attractive free energy density for a pair of plates, treated at a more approximate, *spatial* mean-field level. In the notation of Ref. 1,

$$\beta f_{att} = \frac{1}{2} \sum_{\alpha, \beta} \sigma_\alpha p_\alpha K_{\alpha\beta} p_\beta \sigma_\beta + \sum_\alpha \sigma_\alpha \ln p_\alpha. \tag{28}$$

This result also follows from a large-area limit of Eq. (19) with random grafting points.[1,8] We expect the simplification provided in this Communication will boost the use of our model for calculating interactions free-energy for general ligand–receptor-mediated systems.

[1]P. Varilly, S. Angioletti-Uberti, B. M. Mognetti, and D. Frenkel, "A general theory of DNA-mediated and other valence-limited colloidal interactions," J. Chem. Phys. **137**, 094108 (2012).

[2]R. Dreyfus, M. Leunissen, R. Sha, A. Tkachenko, N. C. Seeman, D. J. Pine, and P. M. Chaikin, "Aggregation-disaggregation transition of DNA-coated colloids: Experiments and theory," Phys. Rev. E **81**, 041404 (2010).

[3]P. L. Biancaniello, A. J. Kim, and J. C. Crocker, "Colloidal interactions and self-assembly using DNA hybridization," Phys. Rev. Lett. **94**, 058302 (2005).

[4]N. A. Licata and A. Tkachenko, "Statistical mechanics of DNA-mediated colloidal aggregation," Phys. Rev. E **74**, 041408 (2006).

[5]W. B. Rogers and J. C. Crocker, "Direct measurements of DNA-mediated colloidal interactions and their quantitative modeling," Proc. Natl. Acad. Sci. U.S.A. **108**, 15687–15692 (2011).

[6]B. M. Mognetti, P. Varilly, S. Angioletti-Uberti, F. J. Martinez-Veracoechea, J. Dobnikar, M. Leunissen, and D. Frenkel, "Predicting DNA-mediated colloidal pair interactions," Proc. Natl. Acad. Sci. U.S.A. **109**, E378–E379 (2012).

[7]In Ref. 2, the free energy of attraction includes only states where at least one bond is formed, thus βF_{att} there is $-\ln[(1+x)^{N_p} - 1]$. Our definition of F_{att} is the free energy of the system of linkers, and includes the configuration where no bonds are formed.

[8]B. M. Mognetti, M. E. Leunissen, and D. Frenkel, "Controlling the temperature sensitivity of DNA-mediated colloidal interactions through competing linkages," Soft Matter **8**, 2213 (2012).