JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Efficient Nonequilibrium Method for Binding Free Energy Calculations in Molecular Dynamics Simulations

Robert B. Sandberg,[†,§] Martina Banchelli,[†,‡] Carlo Guardiani,[∥,⊥] Stefano Menichetti,[†] Gabriella Caminati,*[,†,‡] and Piero Procacci*[,†]

[†]Department of Chemistry and [‡]Center for Colloid and Surface Science (CSGI), University of Florence, 50121 Firenze, Italy

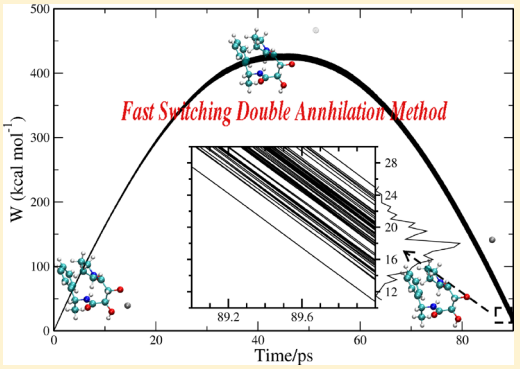
[§]Department of Chemistry, State University of New York at Binghamton, Binghamton, New York 13902, United States

[∥]Department of Physics, University of Cagliari, 09124 Cagliari, Italy

[⊥]IOM Institute, CNR, 09042 Cagliari, Italy

Supporting Information

ABSTRACT: We introduce an effective technique for the calculation of the binding free energy in drug-receptor systems using nonequilibrium molecular dynamics and application of the Jarzynski theorem. In essence, this novel methodology constitutes the nonequilibrium adaptation of an ancient free energy perturbation technique called Double Annihilation Method, invented more than 25 years ago [*J. Chem. Phys.* **1988**, *89*, 3742−3746] and upon which modern approaches of binding free energy computation in drug-receptor systems are heavily based. The proposed computational approach, termed Fast Switching Double Annihilation Methods (FS-DAM) in honor of its ancient ancestor, is applied to a prototypical example system with multiple binding sites, proving its computational potential and versatility in unraveling multiple site or allosteric binding processes.



INTRODUCTION

The last two decades have witnessed tremendous progress in the in silico determination of accurate absolute binding free energies of drug-receptor complexes using molecular dynamics (MD) simulations and explicit solvent molecules. Computational methods have now reached the point where making accurate predictions in drug−receptor systems is commonplace, and hence, these techniques are expected to contribute significantly to the process of drug discovery and optimization in the near future. In this context, the so-called alchemical double decoupling method[1−3] (DCM) is probably the most popular approach for binding free energy calculations being integrated in many widespread MD suites including GROMACS[4] and NAMD.[5] DCM was originally derived as a simple variant of the Double Annihilation Method invented by William L. Jorgensen and co-workers back in 1988.[6,7] In a nutshell, DAM or DCM proceeds by constructing ad hoc thermodynamic cycles whereby the ligand, in two distinct thermodynamic processes, is reversibly decoupled from the environment in the bulk and in the binding site. Reversible decoupling is achieved by implementing the free energy perturbation or thermodynamic integration approaches, that is, by breaking up the decoupling process in several contributions to be computed in a series of equilibrium simulations referring to progressively bulk or site decoupled states. The binding free energy is finally recovered by evaluating the difference between the site and bulk decoupling free energies.[1,3] DCM fundamentally differs from its ancestor DAM in the fact that a traslational restraint is imposed to the decoupling process of the bound ligand in order to avoid the difficulties related to the sampling issue of a wandering partially decoupled ligand.

A second important class of techniques for free energy calculations is based on the definition of appropriately selected docking collective variables for evaluating the associated potential of mean force (PMF) using advanced methodologies such as metadynamics,[8−11] thermodynamic integration,[2,7] and nonequilibrium steered MD.[12−15] Enhanced sampling techniques such as the Binding Energy Distribution Analysis Method (BEDAM)[16] or the Energy Driven Undocking via Hamiltonian Replica Exchange Method EDU-HREM[17] have also been recently proposed. In short, both these techniques implement, with different flavors, a single alchemical decoupling of the bound state via potential energy scaling along a replica progression in a multicanonical generalized ensemble (GE).

The common background of all these disparate techniques is the fact they all use the equilibrium or the quasi equilibrium simulation approach to evaluate the free energy differences between bound and unbound state. In BEDAM or EDU, for example, results are reliable only if the GE is at equilibrium, an expectation that cannot always be given for granted, especially in the presence of secondary binding sites even for a very long

Received: October 29, 2014

DOI: 10.1021/ct500964e
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

simulation time span.[17,18] In the DCM method, where the ligand in the bound state is translationally restrained,[1,3] the issue of convergence is less acute, but the technique, similar to the enhanced sampling methods, still has limitations in treating the problem of secondary binding sites. In fact, while in DCM it is in principle possible to individually screen all putative binding sites in a receptor, such an approach would be highly unpractical, because for *each* binding site, series equilibrium simulations in the nanosecond time scale of the intermediate alchemical states of the bound complex are required. PMF methodologies, such as metadynamics or steered molecular dynamics, suffer of the same limitations in dealing with multiple binding sites. The former would require a computationally intractable expansion of the defined collective coordinates space, while the latter, as in DCM, would require a costly individual screening of the binding sites.

Such practical limitations of the available advanced binding free energy computation techniques are indeed important as multiple site binding plays a relevant role in many crucial biological process. For example, in hemoglobin, the oxygen molecule can bind competitively or sequentially (at low and high oxygen pressure in blood, respectively) in four distinct binding sites. Having multiple binding sites, complicated by the interaction with the receptors metastable conformational states, is a relevant issue in, for example, kinase inhibition and in allosteric inhibition in general.[19,20] As a matter of fact, allosteric regulation provides a promising strategy for novel drug design not only in kinase inhibition for cancer therapy but also for the specific regulation and modulation of biological target in general.[21−23]

In this paper, we present an extension of the recently developed fast switching method for evaluating the solute hydration free energies applied to the calculation of the binding free energies. It has been recently shown[24,25] that the dissipation during fast subnanosecond alchemical annihilation/creation of solutes in solvent of small and rapidly diffusing molecules such as water or methanol is surprisingly modest, yielding fairly Gaussian work distributions. According to this methodology, rather than performing long (nanosecond time scale) equilibrium runs to produce reversible decoupling annihilation processes, many fast (picosecond time scale) site-decoupling and bulk-decoupling nonequilibrium simulations are run in parallel starting from canonically sampled configurations. The binding free energy is recovered by straightforwardly applying the Jarzynski theorem to the resulting normal distributions. The proposed strategy is basically a nonequilibrium variant of the old DAM and, as such, is termed Fast Switching Double Decoupling method (FS-DAM). Because all of the fast decoupling switching trajectories can be run independently with no communication overhead, FS-DAM provides a very powerful alternative route to binding free energy calculations allowing, through an embarrassingly parallel approach,[26] to compute the binding free energy of ligand-medium size receptor in a matter of tens of minutes on modern massively parallel platforms.[25] Besides its extraordinary efficiency, the method is of general validity providing a viable means to the computation of the binding free energy in multiple site systems and in treating allostery in general. In order to show this, we practically apply the methodology to the evaluation of the binding free energy for a prototypical ligand−receptor pair where the receptor contains multiple binding sites and exhibits multiple tautomers/conformers, namely, the complexes of Zinc(II) with the molecule MBET306, that is, a synthetic precursor of a family of recently discovered tartrate-based Tumor Necrosis Factor α converting enzyme (TACE) inhibitors[27−29] with experimentally known overall dissociation constant in water.

The article is organized as follows. In the Theoretical Methods section, we briefly review the thermodynamic foundation of the equilibrium constant of ligand−receptor reactions with emphasis on receptors with multiple binding sites. In the same section, we discuss the statistical mechanics basis of DAM and of FS-DAM. In the Materials and Methods section, we provide the basic computational details such as simulation setup and force field parameters. Here, we also describe the structural properties and binding characteristics of our prototypical multisite system MBET306, recollecting the available experimental data on Zinc(II) binding in water. In the Results section, we present the practical application of the technique to the determination of the binding affinity of Zinc(II) for MBET306 in water and compare the simulation outcome to the experimental counterpart. A brief discussion of the efficiency of the FS-DAM on parallel platforms can also be found in this section. Finally, in the Conclusions section, we envisage some possible future developments of the technology.

THEORETICAL METHODS

Receptors with Multiple Binding Sites and Conformations. In the process of binding of a small ligand L (e.g., a drug) to a receptor R (e.g., a protein), one has the chemical equilibrium

$$\mathrm{RL} \rightleftharpoons \mathrm{R} + \mathrm{L} \tag{1}$$

with the standard dissociation free energy and the dissociation constant defined as

$$\Delta G_0 = -k_B T \ln(K/C_0) \quad K_d = \frac{[\mathrm{R}]\,[\mathrm{L}]}{[\mathrm{RL}]} \tag{2}$$

where $C_0 = 1\ M$ is the standard concentration. Equations 1 and 2 generally refer to the special case of a single binding site (i.e., the receptor) in a unique and well-defined conformation, which has only one relevant binding pocket where the ligand can bind. For the case of competitive binding in multiple binding sites of distinct conformers/tautomers, in the ideal limit of infinite dilution, the physical observable [RL] refers to the bound species no matter what the conformational state of the receptor is or in which particular binding site the ligand L is placed, that is,

$$[\mathrm{RL}] = \sum_{k}^{N_{\mathrm{states}}} \sum_{i}^{N_{\mathrm{sites}}} [(\mathrm{R}^{(k)}\mathrm{L})_i] \tag{3}$$

where with the notation $(\mathrm{R}^{(k)}\mathrm{L})_i$ we indicate a bound state RL with the ligand L in the i-th binding site and with the receptor R in the k-th tautomeric/conformational structure, and the sum is extended to all possible N_{sites} receptor binding sites and N_{states} receptor tautomers/conformers. Also, the total concentration of the receptor in eq 2 is given by the sum of the concentrations of all its tautomeric or conformational interconverting metastable structures, that is,

$$[\mathrm{R}] = \sum_{k}^{N_{\mathrm{states}}} [R^{(k)}] \tag{4}$$

DOI: 10.1021/ct500964e
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

For each of the $(R^{(k)}L)_i$ and $R^{(k)}$ species, one can define the simultaneous equilibria and associated equilibrium constants, for example,

$$(R^{(k)}L)_i \rightleftharpoons R^{(k)} + L \quad K_i^{(k)} = \frac{[R^{(k)}]\,[L]}{[(R^{(k)}L)_i]} \tag{5}$$

$$R^1 \rightleftharpoons R^{(k)} \qquad G_k = \frac{[R^k]}{[R^{(1)}]} \tag{6}$$

with G_k being the equilibrium constant for the interconversion of the 1 and k receptor conformational/tautomeric metastable species. Equations 5 and 6 are both satisfied when the system is at equilibrium (i.e., when the chemical potentials of all species are equal). With trivial algebra it can be shown that the overall dissociation constant can be written in terms of the individual site-conformer dissociation constants as

$$\frac{1}{K_d} = \sum_k^{N_{states}} G'_k \sum_i^{N_{sites}} \frac{1}{K_i^{(k)}} \tag{7}$$

where $G'_k = G_k/(\Sigma_l G_l)$. In terms of the standard dissociation free energy, eq 7 may be written as

$$e^{\beta\Delta G^\circ} = \sum_k^{N_{states}} G'_k \sum_i^{N_{sites}} e^{\beta\Delta G^0_{ki}} \tag{8}$$

where $\Delta G_0 = -k_B T \ln(K_d/C_0)$ is the overall standard dissociation free energy (eq 2) of the reaction eq 1 and $\Delta G^0_{ki} = -k_B T \ln(K_i^{(k)}/C_0)$ is the standard dissociation free energy relative to the i-th bound state type with the k-th metastable conformational/tautomeric state of the receptor. For tightly bound complexes, the dissociation free energy is strongly positive and the weighted exponential sum on the right is usually dominated by one of few $e^{\beta\Delta G^0_{ki}}$ terms, in the assumption that the G'_k weights are not disparate (i.e., in our case, the free energies of the receptor tautomers/conformers are comparable). The calculation of binding free energies via computer simulations for multiple binding sites systems poses no difficulties provided that a viable method is available to compute the individual standard free energy terms ΔG^0_{ki}. As briefly reviewed in the introduction, many techniques aimed at so doing have been devised in the past. In the following, we shall discuss with some detail the two popular schemes based on alchemical transformations, namely the reversible Double Annihilation (DAM)[6] or Double Decoupling (DCM)[1] methods.

Reversible (Equilibrium) Alchemical Decoupling Schemes. Both DAM and DCM proceed by computing the standard free energy of binding (usually referred to a single binding site) via computer simulation defining ad hoc thermodynamic cycles. Quoting Jorgensen,[7] "all that is needed is two simulations, one in which [the ligand] disappears by itself and one in which it disappears from the complex." Actually, as pointed out in ref 1, the ligand does not "disappear" but it is merely decoupled from the environment. In DAM the dissociation free energy difference is thus given by the difference between (1) the reversible work to decouple the ligand when bound to the solvated receptor (ΔG_1) and (2) the reversible work to decouple the ligand from the bulk solvent ((ΔG_2, i.e., its solvation free energy). These two processes are usually schematized as

(1) ΔG_1:

$$(RL)_{sol} \rightarrow R_{sol} + L_{gas} \tag{9}$$

(2) ΔG_2:

$$L_{sol} \rightarrow L_{gas} \tag{10}$$

and the dissociation free energy referring to process $(RL)_{sol} \rightleftharpoons R_{sol} + L_{sol}$ is defined as $\Delta G = \Delta G_1 - \Delta G_2$. The reversible work in the two processes exemplified in eqs 9 and 10 can be standardly computed in molecular dynamics simulations using thermodynamic integration (TI)[30] or free energy perturbation techniques (FEP).[31−34] In both cases, this is achieved by defining a potential energy function that interpolates smoothly between the potential energy functions (including the PV terms if the simulations are done at constant pressure) of the initial and the final states. The precise functional form of the interpolation may be adjusted to optimize convergence of the simulations. Typically, the interpolation is controlled by a parameter $\lambda \in [0, 1]$, where 0 and 1 correspond to the starting and final energy functions, respectively. In eq 9, referring to process 1, the initial state is $(RL)_{sol}$, and the final state is $R_{sol} + L_{gas}$,while in eq 10, referring to process 2, the initial state is L_{sol} and the final state is L_{gas} and the pure bulk solvent. In the practice, the [0,1] λ interval is discretized such that several independent simulations are run for each λ_i in [0,1], and the reversible work of the whole process is simply obtained by appropriately summing up the individual contributions evaluated from the λ_i equilibrium simulations. In TI, for example, the reversible work is given by computing the discretized integral

$$W_{rev} = \Delta G(sim) = \int_0^1 \left\langle \left(\frac{\partial H(\lambda)}{\partial \lambda} \right) \right\rangle_\lambda d\lambda \tag{11}$$

As thoroughly discussed in refs 1 and 35, the original formulation by Jorgensen of the DAM method for the $(RL)_{sol} \rightarrow R_{sol} + L_{gas}$ reaction is an incomplete theory. While the computation of ΔG_2 in process 2 poses no difficulties with no standard state dependency, in process 1 the number of species in the initial and final states differs and therefore ΔG_1 should bear a dependence on the standard state concentration that is not apparent in the DAM formulation. From a practical standpoint, the DCM method is hence merely a restatement of the DAM approach with the implementation, in the process $(RL)_{sol} \rightarrow R_{sol} + L_{gas}$, of a constraint that forces the ligand to occupy a region near the binding site.[1] This introduces a standard state correction to ΔG_1 such that

$$\Delta G_1^0 = \Delta G_1(sim) + k_B T \ln(V_1/V_0) \tag{12}$$

where $\Delta G_1(sim)$ is the reversible work evaluated in the simulation, $V_0 = 1660$[3] is the molecular volume of the standard state, and V_1 is the site volume imposed by the artificial constraint. For example, if the imposed constraint is a simple three-dimensional square well potential, then $V_1 = b^3$ where b is the width of the square well. If the constraint is enforced using a 3D harmonic potential, then $V_1 = (2\pi RT/k)^{3/2}$ where k is the force constant of the harmonic potential. As pointed out by Roux,[3] for DAM the actual V_1 volume correcting for the standard state should be the volume of the simulation box V_{box}, introducing an apparent size dependence of ΔG_1^0. Actually, provided that the equilibrium simulations are all at convergence for each λ state in the interval [0,1], the DAM with $V_1 = V_{box}$ is

DOI: 10.1021/ct500964e
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

perfectly correct and bears no size dependence. In fact, by, for example, increasing the box volume, while the standard state correction becomes more positive, the dissociation free energy obtained by the equilibrium simulations $\Delta G_1(\mathrm{sim}) = -k_B T \ln(P_u/P_b)$ becomes more negative as the probability of the unbound state P_u becomes large with respect to that of the bound state P_b, thus compensating for the standard state box size dependent correction.[36] The same "size dependence" can be found in DCM, where a different arbitrary choice of the site volume V_1 leads to apparently different values of the standard free energy. As discussed in ref 1, if V_1 is, for example, made smaller, the negative standard state term will be balanced by a larger reversible work integral, because the ligand will not be so free during the simulation to drift off into regions where the probability to find the ligand is small. From another viewpoint, as pointed out in refs 17 and 37, the use of a constraint to keep the ligand close to the receptor is equivalent to fix the ligand concentration in the environment by setting its molecular volume to V_1. The standard state correction may be then viewed as the reversible work needed to bring the ligand concentration from that imposed by the constraint to that of the standard state. Therefore, the use of a periodic box, where the ligand is obliged to reside, is nothing else than a volume constraint, exactly like the square well potential or the harmonic tethering potential used in DCM. In DCM, the strong constraint sets a ligand concentration that is much higher than that of the standard state, while in DAM the loose constraint imposed by the periodicity yields a ligand concentration that is much larger than that of the standard state. To quantify this statement, in DCM for a typical square well potential of radius $b = 6$ Å and $V_1 = 216$ Å^3, one gets an effective ligand concentration of $[L] \simeq 8$ M, while in DAM, for an MD box of 50 Å side-length, one obtains $[L] \simeq 0.01$ M.

DCM and DAM (with the standard state correction proposed by Roux) are hence perfectly equivalent *provided that equilibrium is reached in all simulations and all relevant states, either bound or unbound, are sampled for all values of the λ discretized coupling parameter*. The real difficulty in DAM, where the ligand is unrestrained except for the imposition of the periodic boundary conditions, lies in the fact that in order to compute ΔG_1 correctly, the ligand would have to sample every possible position and orientation relative to the protein within the MD box.[38] This difficulty has been in part circumvented by Pande et al.[35] by discarding, at loosely coupled Hamiltonians, the conformations that were outside an arbitrarily defined binding site region (defining V_1) as not being part of the bound state. Such a procedure is not satisfactory from a theoretical point of view because, as we have seen above, none of the configurations should be discarded if equilibrium has been attained in a given λ state. However, from a practical standpoint, the Pande et al. approach proved to be effective. They stated that the so defined standard state correction ranged in their case (the ligands of FKBP12[39]) from −0.04 to −0.64 kcal/mol depending on the definition of the binding region, that is, values that are in all cases much smaller than the error of the method (see Figure 3 of ref 35). As we shall see below, the rather obscure approach proposed in ref 35 can be viewed under a much brighter light in nonequilibrium experiments.

As previously discussed, in DCM, the sampling issue is bypassed by enforcing a strong constraint (i.e., imposing a huge ligand concentration) that always keeps the ligand close to the binding site. Such an approach is definitely more coherent than that proposed by Pande in unrestrained DAM, although the assumed independence of the underlying and unknown binding site volume on the constraint (i.e., on the effective ligand concentration) should be verified case by case by varying the imposed V_1 at least within a certain range. For example, with a too tight constraint, some important configurations contributing to the bound state may be missed[18] or rarely sampled, while with a too loose one, the pathology concerning poor convergence of the DAM may re-emerge.

Nonequilibrium Alchemical Decoupling Scheme. In this contribution, as a natural follow-up of the our previous work on fast switching alchemical simulations,[25] we propose a nonequilibrium (NE) variant of the unrestrained DAM originally proposed by Jorgensen.[1] We shall name such methodology Fast Switching Double Annihilation Method (FS-DAM) in honor of its ancient ancestor, DAM.

In ref 25, a rigorous implementation of continuous alchemical transformations for molecular systems (i.e., the coupling λ parameter varying continuously while the system is evolving) is discussed. The technique was specifically devised for systems where electrostatic interactions are treated using the efficient and widespread Smooth Particle Mesh Ewald Method.[40] The stumbling block of the lack of knowledge of the Ewald potential in the FFT based approach, which prevented the straightforward implementation of alchemical schemes in many MD codes, is bypassed by judiciously subtracting in the dynamics the unwanted intrasolute Erf modulated charge−charge interactions in the direct lattice automatically included in the SPME reciprocal lattice contribution. Crooks-based[41] bidirectional fast switching alchemical transformations were then used to compute the solvation free energies for various solutes in water and octanol. This technique proved to be highly competitive with respect to conventional free energy perturbation approaches based on equilibrium simulations. For further details of the implementation of driven alchemical transformations within the framework of molecular dynamics simulations, we refer to ref 25.

We now come back to FS-DAM. We start by considering that the two processes (1) and (2) of eqs 9 and 10 in DAM are two alchemical transformations that can be irreversibly driven by externally varying a λ parameter that rapidly annihilates the ligand in two different environments, that is, that of the binding site in the solvated receptor and that of the bulk solvent. The free energies ΔG_1 and ΔG_2 can be therefore recovered in both cases using the Jarzynski theorem, that is,

$$e^{-\beta \Delta G} = \langle e^{-\beta W} \rangle \tag{13}$$

where W is the work produced in the NE λ driven experiment from the fully coupled system to the decoupled system (either 1 or 2) and the average is taken over many independent nonequilibrium trajectories, all done according to an identical time schedule, and that are started from the canonically sampled configurations of the coupled system. When dealing with multiple binding site systems, by canonical sampling, we mean conventional sampling within each basin defined by the given bound state. The relative importance of the basins, that is, their mean free energy difference, as we shall see in the Results section, will be a simple byproduct of the method. The exponential average in eq 13 can be written in terms of the work distribution $P(W)$ as $\langle e^{-\beta W} \rangle = \int P(W) \exp(-\beta W) dW$. It must be remarked that, in the fast switching method, the *equilibrium is required only on one end of the process*, that is, the coupled starting states must be canonically sampled, but the arrival decoupled states are by definition NE configurations

DOI: 10.1021/ct500964e
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

bearing an excess energy related to the dissipation of the irreversible transformation. We stress again this point because it is central in the theory: in NE alchemical simulations *there is no sampling issue whatsoever for all NE intermediate coupling or ending decoupled states*. The problem of statistical sampling is "shifted" in the number of independent trajectories that must be produced to obtain a stationary work distribution $P(W)$. This number is larger the wider the distribution, that is, as we shall see later on, the more dissipative the NE process. For a reversible transformation, that is, for a process of infinite duration, the dissipation is zero and only one single trajectory is needed, yielding (except for standard state corrections) a work equal to ΔG.

In constructing the FS-DAM theory, we shall exploit here an important fact that was already noted in previous studies on fast switching alchemical transformations of solvation energies, namely, the Gaussian nature of the alchemical work distribution along with the low dissipation in water environment ever for duration or the experiments as fast as 50−100 ps (see also Table S4 and related discussion in the SI on this point). The Gaussian nature of the work distribution is a consequence of the fact that the NE decoupling process form the environment is essentially Markovian. For the process of eq 9, for example, as long as the ligand remains in the same binding site and the receptor remains in the same conformational state, a Markovian behavior is indeed a reasonable expectation, as the energy change at a given time t in the fast decoupling process should depend only on the alchemical state (i.e., the instantaneous value of $\lambda(t)$) at that given time. The probability for a conformational or binding site change during the process grows with the duration of the process, which may hence exhibit non-Markovian character resulting in, for example, multimodal work distributions. The Gaussian character of the work distribution should therefore be enhanced for shorter duration of the NE process. A Gaussian distribution, on the other hand, constitutes an enormous advantage when applying eq 13. The strongly statistically biased[42] exponential average in eq 13 can in fact be expanded in terms of cumulants.[43] As shown in ref 44, the cumulant expansion is either of order two for a Gaussian distribution or of infinite order. In case of Gaussian distribution, the Jarzynski theorem then reads

$$\Delta G = \langle W \rangle - \frac{\beta\sigma^2}{2} \tag{14}$$

where $\langle W \rangle$, the mean NE work, is the first cumulant, and σ^2, the square of the standard deviation, is the second order cumulant of the work distribution. The term $\beta\sigma^2/2$ represents the dissipated work during the NE transformation. We stress here that for a Gaussian distribution of the work, eq 14 is an *exact result*, implying no approximation whatsoever.[43,45,46] For fast switching decoupling processes, the Gaussian nature of the distribution can be speedily assessed by evaluating the moments of order higher than two of the work distribution, that is, the standardized skewness and the excess kurtosis that (similar to all other higher cumulants) should be close to zero within the standard error.

The unrestrained FS-DAM therefore reduces to the application of eq 14 to two distinct fast NE experiments, namely (1) the annihilation (or decoupling) of the ligand from the receptor in the bound state and (2) the annihilation (or decoupling) of the ligand in bulk solvent and application of eq 14, to recover the free energy, subject to the previous verification of the Gaussian character of the distributions. The procedure is general and (as explained in the Results section) perfectly parallelizable, bearing no communication overhead at any stage, thus delivering the result with a constant wall clock time (i.e., that needed to run serially a conventional simulation of the system under scrutiny lasting few tens of picosecond), no matter how many binding sites need to be examined for the determination of the binding affinity provided that unlimited CPU resources are available.

It remains to assess the crucial question of the standard state in the decoupling of the bound state. We first note that, unlike in equilibrium DAM, the quantity $\Delta G_1 = \langle W_1 \rangle - (\beta\sigma^2/2)$, measured by collecting the work distribution from a set of independent NE trajectories starting from a canonically sampled bound state, by construction cannot have any dependence on the box size, under some minimal provisions that we discuss below. If the swarm of fast NE simulations is driven sufficiently fast, then the decoupling ligand remains (on average) bound to the receptor up to a certain alchemical decoupling stage. Beyond this point, the ligand starts to diffusively drift off in the bulk in a random direction reaching the ballistic regime only when is perfectly decoupled from the environment, that is, at the end of the simulation, generally traveling for a distance within 1 nm of the receptor depending on the speed with which the NE trajectories are conducted. In the Supporting Information (SI), we report the average receptor−ligand distance as a function of time measured in the 90 ps NE decoupling of one specific case of ligand−receptor systems. The extent of mean free path of the ligand relative to the receptor depends only on the speed of the transformation and on the λ-dependent diffusion coefficient of the ligand in the solvent, and it must be, hence, clearly independent of the box size. In conclusion, at fixed decoupling speed, the MD box in the NE decoupling can be made large at will with no change on the resulting ΔG_1. In all evidence, this is at variance with standard equilibrium DAM where, as explained above, the box size, being related to the effective ligand concentration, affects the value of the resulting free energy in the loosely coupled intermediate (see also ref 36). The key factor is that in FS-DAM, the final decoupled state is a nonequilibrium state, with a nonequilibrium ligand concentration. In such a NE state, the ligand can be found only within some fixed distance from the receptor (controlled by the decoupling speed), and outside this sphere, the mean ligand concentration is practically zero. *The volume of this NE sphere, where the decoupled ligand is concentrated, must be used in the standard state correction*. For defining this volume, which we call $V(\tau)$ to make it clear that it is a function of the duration τ of the NE experiment, one can safely use the procedure adopted by Pande[35] by defining some threshold distance to set the volume binding site. For example, one can choose the distance yielding exactly the standard state volume (i.e., ligand within a maximum distance of $r \simeq 8$ A from some point in the binding pocket of the receptor) and erase from the distribution the trajectories that do not satisfy this condition at all times during the NE transformation. In practice, if the alchemical decoupling is done sufficiently fast, this condition is verified for nearly all trajectories and can even be approximately imposed by simply tuning the speed of the transformations. All of the above can be summarized in one single equation, which constitutes the central result of our paper, that is, the expression of the (Markovian) decoupling free energy of the ligand−receptor system:

DOI: 10.1021/ct500964e
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

$$\Delta G_1 = \langle W \rangle - \frac{\beta\sigma^2}{2} + k_B T \ln(V(\tau)/V_0) \quad (15)$$

Note that eq 15 is valid also in the infinite duration limit of a single reversible alchemical transformation. In this case, in fact, $W = \langle \Delta G_1(\mathrm{sim}) \rangle_{\mathrm{TI}}$, $\sigma = 0$, and $V(\tau) = V_{\mathrm{box}}$, thus recovering the DAM results with the Roux standard state correction. As in any other method, the ligand NE concentration $1/V(\tau)$ that corrects for the standard state in FS-DAM has an intrinsic uncertainty because the NE sphere has somewhat undefined boundaries. It should never be forgotten, however, that an error of 100% in such a volume would yield a difference of only $k_B T \ln(2) = 0.4$ kcal/mol in the dissociation free energy.

We conclude this section with some comments regarding the very significance of the bound complex, of the associated volume binding site, and hence of K_d on ΔG. As explained in ref 1, the arbitrariness in the binding site definition is an inescapable fact in the statistical mechanics of binding. The binding site enters in the theory through the partition function of the bound state that must be integrated only in the binding region. This is mathematically accomplished, as explained in refs 1, 36, and 47, by introducing a Heavyside step function of the ligand receptor relative coordinates that defines somehow the bound state. The function is equal to one when the bound state is formed and zero otherwise. There are some guiding lines in defining the domain of the bound state[1] (i.e., where the step function is one) but such a definition essentially remains an external decision. Gilson et al. have shown that the integral reported in ref 36 or in eq 12 of ref 1 depends weakly on the size of the binding region provided that the binding is sufficiently strong ($\Delta G > 6$ kcal/mol) and that the size of the bound state is not too big. Such arbitrariness is implicit also in the experimental measure. The physical observable referring to the dissociation constant is generally obtained by analyzing some signal that exhibits a different interaction with the bound and unbound state. For example, at a given wavelength, fluorescence excitation spectra of the bound state may be different from the excitation of the unbound receptor. This measure, however, may miss completely some important conformations of the bound state with an impact on the dissociation constant. A paradigmatic example of this experimental uncertainty related to the binding site arbitrariness is the difference between inhibition constant (typically obtained by assessing the effect of an inhibitor on the function of a target) and of dissociation constant typically measured, for example, by intrinsic fluorescence quenching.[48] The former probes, in principle, the bound state (or a manifold of bound states) that cripples the activity of the target protein, while the latter probes the bound state that reduces the receptor intrinsic fluorescence signal due to a chromophore localized near one binding region. Although in principle referring to the same physical observable, inhibition constant and dissociation constant may differ in some cases of orders of magnitude.[48,49]

MATERIAL AND METHODS

MBET306 and Zinc(II)-MBET306 Complexes in Water. In this paper, we shall present the application of the FS-DAM to the calculation of the dissociation constant of the complex $Zn(MBET306)^{-1}$ in water. The MBET306 molecule (see Figure 1) is a bis-amide of the L-tartaric acid (IUPAC name 2,3-dihydroxy-4-oxo-N-((R)-1-phenylethyl)-4-(pyrrolidin-1-yl)-butanamide (**5**))[27] featuring two planar or quasi planar hydrophobic substituents for optimal inhibition of the Zinc(II)

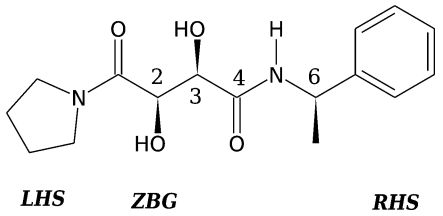


Figure 1. Chemical structure of MBET306.

protease TACE.[28,29] In water solution, MBET306 behaves similar to a weak acid[27] with experimentally determined[27] pK_a of 7.0 ± 0.3 referring to the first ionization constant with loss of one of the hydroxyl protons of the central tartaric moiety (see Figure 1). At a pH = 8.0, the experimental dissociation constant for the equilibrium $Zn(MBET306)^+ + H^+ \rightleftharpoons MBET306 + Zn^{2+}$ is found to be $K = 12 \pm 1$ μM, and it has been determined in ref 27 with great accuracy using as a probe for the formation of the $Zn(MBET306)^+$ complex in excess of $ZnCl_2$ the increase with MBET306 concentration of the weak absorption band peaked at about $\lambda = 330$ nm. As stated above, the experimental pK_a for the ionization of the neutral species, $MBET306 \rightleftharpoons MBET306^- + H^+$, is equal to 7.0 ± 0.3. Given the equilibrium constant of $K = 12$ μM for the reaction involving the neutral species, the dissociation constant for the equilibrium

$$Zn(MBET306)^+ \rightleftharpoons MBET306^- + Zn^{2+} \quad (16)$$

referred to the binding processes involving the anion $MBET306^-$ and the cation Zn(II) can be calculated as

$$\begin{aligned} K_d &= \frac{[MBET306^-]\,[Zn^{2+}]}{[Zn(MBET306)^+]} = KK_a \\ &= (1.2 \pm 0.1 \times 10^{-5})\,(1.0 \pm 0.9 \times 10^{-7}) \\ &= 1.2 \pm 1 \times 10^{-12} \end{aligned} \quad (17)$$

yielding a standard dissociation free energy, $\Delta G_0^{(\mathrm{exp})} = 16.2 \pm 0.6$ kcal mol^{-1}. Here, we are not interested in the biological relevance of MEBT306 with regard to TACE inhibition but rather we use this simple molecule as a first class guinea pig for modeling a binding process regulated by multiple site, multiple tautomer/conformer interactions, using it as a mean for assessing the accuracy and the efficiency of FS-DCM in the determination of the binding free energy in these systems. In the Zn(II)−$MBET306^-$ pair, Zn(II) is the ligand and MBET306 is the receptor. The experimentally determined equilibrium constant for the $ZnMBET306^+$ formation, eq 17, is the result of a complex scheme of multiple site conformations[27] that is schematized in Figure 2. The ligand tautomeric/conformational states can in fact be defined in terms of hydrolyzed hydroxyl groups in the tartaric moiety (either that involving the oxygen atom labeled 2 or the oxygen atom labeled 3 as shown in the top left part of the figure) and/or by the *cis*, *trans* conformations of the amide bond flanking the ethyl-phenyl moiety (see structures in the top right part of the figure). Each of these four possible conformers can bind the Zinc(II) atom in six possible bidentate complexes and four possible tridentate complexes (an oxygen atom is defined to be bound to the Zinc when the O−Zn distance is less than 2.6 Å[29]). For each conformer, these complexes are classified according to the zinc binding oxygen atoms. The conventional numbering of the oxygen atoms is shown in the top left structures. In the bottom part of the figure, as representative

DOI: 10.1021/ct500964e
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

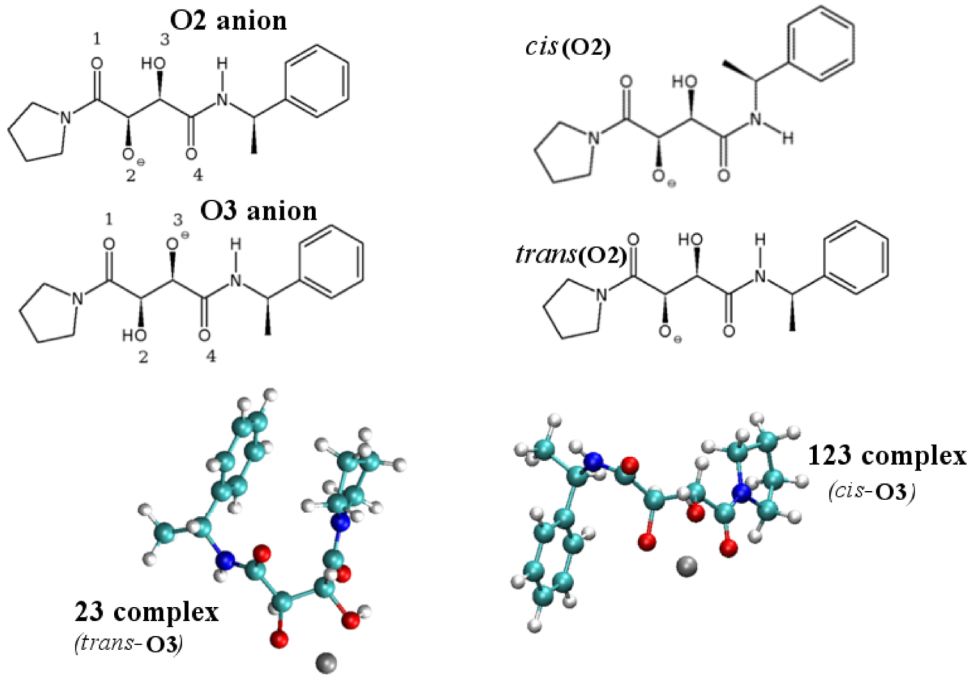


Figure 2. Conformational states for MBET306$^-$ and possible Zn-MBET306$^+$ complexes.

examples, the bidentate, (2,3)-*trans*-O3 and the (1,2,3,)-*cis*-O3 tridentate complexes are shown. Such richness of multiple conformations and competitive binding sites along with the modest size of the receptor make the MBET306-Zn(II) system an ideal candidate to assess and test from a practical standpoint the capability of the FS-DAM in multiple site binding systems.

Computer Simulation Methods. At physiological pH and in water, MBET306 is overwhelmingly present[27] as a tautomeric mixture of monoanions with hydrolyzation of the hydroxyl in position 2 or 3. The force fields for the MBET306 anionic species are taken from ref 27. In ref 27, the atomic charges of the two anion MBET306$^-$ were standardly computed using Density Functional Theory and the B3-LYP functional[50] on the gas phase optimized structures and applying the RESP model with 6-31G(d) basis set.[51] We adopted the TIP3P[52] model for water. The Lennard-Jones parameters for the zinc cation, σ = 1.247 Å and ε = 0.0125 kcal mol^{-1}, were modified with respect to the original amber99SB values[53] and adjusted so as to reproduce, in molecular mechanics minimization with AGBNP implicit solvent model,[54] the distances between the cation and the binding oxygen atoms in the optimized structures of bidentate and tridentate complexes computed using Polarizable Continuum Model[55] at the B3-LYP/631G(d) level of theory.[27,28] Constant pressure in MD was enforced isotropically using a modification of the Parrinello−Rahman Lagrangian,[56] and temperature control was achieved using the Nosé thermostat.[57] Electrostatic interactions were computed using the smooth particle mesh Ewald algorithm with the convergence parameter set to 0.43 Å^{-1} and grid spacing of ≃1 Å.[40] The equations of motion were integrated using a multiple time-step r-RESPA scheme[58] with a potential subdivision specifically tuned for biomolecular systems in the NPT ensemble.[56,59] Details of the integrator are provided in ref 59. Distance constraints on the MBET306 molecule were enforced only for bonds involving a hydrogen atom (X-H bonds). All calculations were done with the ORAC code.[59,60]

RESULTS

In FS-DAM, there are two distinct phases: (a) the preparatory stage corresponding to the canonical sampling of the fully coupled system 1, that is, the manifold of bound states of Zn(MBET306)$^+$, and of system 2 (the solvated Zn(II) atom); (b) the annihilation or decoupling stage concerning the actual NE simulations that bring system 1 and 2 into the final decoupled states. The preparatory stage is, from a practical standpoint, perfectly identical to the λ = 0 simulations in the reversible DAM. It differs from the latter in the scope. While in FS-DAM the preparatory stage is aimed at collecting the starting configurations of the bound state (or states) for the successive decoupling NE runs, in reversible DAM the λ = 0 point is just one of the many points in which the reversible path from coupled to uncoupled states is discretized. This section is hence divided into three parts: The first part deals with the preparation of the starting configurations (all sharing the same system topology) in the Zn(MBET306)$^+$ multiple binding site system and in the solvated Zinc(II) atom. In the second part, we present the results on the ΔG_{ki} free energies along with a discussion and a comparison with the experimental counterpart. In the final part, we compare FS-DAM and DAM (or DCM) in terms of computational effectiveness on massively parallel platforms.

Stage a: Sampling the Fully Coupled Starting Configurations. The following is an explanation of the preparatory step for the decoupling of the Zn(MBET306)$^+$ complexes. As stated in the preceding section, the Zn-(MBET306)$^+$ chemical species can assume various forms depending on the tautomeric/rotameric state of the MBET306$^-$ anion and on the binding mode of the Zinc(II) atom (mono-, bi-, or tridentate complexes, see Figure 2). Due to steric/energetic considerations, not all of the possible 40 bi or tridentate complexes should be considered in the evaluation of the dissociation constant through eq 7. For each rotameric/tautomeric state of the MBET306 ligand, the relevant Zinc(II) complexes may be identified by randomly sampling bound

DOI: 10.1021/ct500964e
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

states using a conjugate gradient minimization in vacuo (CG). In Table 1 we summarize the results of such search for

Table 1. Stability of the Zn(II)-MBET306 Complexes as Determined from Conjugate Gradient Minimization of MBET-306 Tautomers/Conformers In Vacuo[a]

	O2-*cis*		O2-*trans*		O3-*cis*		O3-*trans*	
type	*p*	*E*	*p*	*E*	*p*	*E*	*p*	*E*
12	0.01	−198	0.02	−245	–	–	–	–
13	–	–	–	–	–	–	<0.01	−232
14	0.15	−197	0.13	−223	0.16	−174	0.20	−205
23	0.17	−266	0.21	−262	0.20	−230	0.18	−226
34	0.02	−219	–	–	0.30	−242	0.19	−225
123	0.23	−285	0.19	−276	0.19	−259	0.41	−254
124	0.02	−294	0.09	−293	–	–	0.01	−244
134	–	–	–	–	–	–	–	–
234	0.32	−292	0.36	−290	0.01	−252	0.01	−252
nb	0.09	−86	–	–	0.14	−87	–	–

[a]The bound states are classified according to the conventions exemplified in Figure 2. For each entry in the table, 1000 starting configurations were generated with a randomly placed Zn(II) atom within a sphere of 8 Å from the center of mass of the ligand. *p* and *E* indicate the fraction of minimization runs out of 1000 that ended up in the corresponding MBET complex (see Figure 2) and the corresponding mean potential energy in kcal mol^{-1}, respectively. The entry "nb" indicates non-oxygen bound Zn−MBET306 complexes.

Zn(MBET306)$^+$ bound states. In general, the most populated complexes (i.e., the fraction of the CG minimization that end up in the given bound state) are also those characterized by a large mean electrostatic stabilization energy. Monodentate complexes are never found in the search for local minima. Among the bidentate complexes, the 23 bound state (i.e., the complex with the Zinc(II) atom bound to the two hydroxyl oxygen atoms of the tartaric moiety) is consistently found to be stable for all of the MBET306$^-$ tautomeric/rotameric forms. The 13 bidentate complexes are found in 3 cases out of 1000 minimization only for the O3-*trans*-tautomeric-rotameric state of MBET306$^-$ while the 134 tridentate and the 24 bidentate complexes are never found in the minimization search.

In the ensuing step, configurations of the low energy bound states identified via minimization in vacuo are tested in the explicit TIP3P water environment using, for each state, a standard unrestrained MD simulation of 0.6 ns duration. In total, we tested 26 Zn(MBET306)$^+$ complexes. For each of them, the initial configuration of the Zinc(II)−MBET306$^+$ bound state was prepared using the corresponding in vacuo minimized structure (see Table 1) surrounded by randomly generated solvent. These simulations were done in a cubic box of average side-length around 24.7 Å in all cases. The number of water molecules included in each of the 0.6 ns simulations (479) was identical so that all of the O3 and O2 complexes shared a common topology. This is not strictly necessary but makes technically much simpler the set up of the subsequent NE massively parallel step. Prior to the 0.6 ns production, the explicit solvent surrounding the bound complex was let to equilibrate for 0.4 ns. Of course, this second preparatory step can significantly alter the structural pattern obtained in the in vacuo minimization, reported in Table 1, and to some extent, it indeed does so. The new situation after inclusion of the solvent effect is summarized in Table 2. Of the 26 tested complexes,

Table 2. Stability of the Zn(MBET306)$^+$ Complexes in Explicit Water as Measured in a 0.6 ns Standard MD at T = 300 K and P = 1 atm[a]

	O2		O3	
	cis	*trans*	*cis*	*trans*
12	s	s	–	–
13	–	–	–	s
14	s	s	s	s
23	s	s	s	s
34	s	s	**24**	s
123	**23**	**23**	**23**	**23**
124	s	s		**24**
234	**23**	s	**23**	s

[a]In the first column, the starting bound states are reported for each of the four MBET306$^-$ tautomers/conformers. Entries marked with the letter "s" (stable) are bound states that lasted for the entire time span of the conventional simulation. Bold entries indicate the final evolution of starting bound state. Entries marked with a dash symbol were never found in the minimization step and were not hence tested.

none underwent full dissociation, 8 evolved to a different bound state, and 18 remained stable for the whole run. The simulations of the 18 stable complexes (14 bidentate and only 4 tridentate states) were used to collect the starting configurations (coordinates and velocities) for the subsequent NE fast switching decoupling stage.

The preparatory stage for the process in eq 10 is trivially achieved by sampling the equilibrium conformations of a solvated Zn(II) atom during a 2 ns simulation, performed with a single zinc(II) cation in a cubic box of average side-length of 21.72 Å with 322 rigid TIP3P water molecules in standard temperature and pressure conditions.

Stage b: Fast Annihilation of the Zinc(II) Atom in the Process 1 and Process 2. Regarding the annihilation of zinc(II) in the complex, for each of the 18 stable bound states found in the preceding step, we determine the corresponding dissociation free energy via FS-DCM applying eq 15. For a given *k*,*i* stable bound state, 256 nonequilibrium simulations lasting 90 ps are performed starting from configurations produced in the 0.6 ns conventional simulations (see Table 2). For a 90 ps duration of the decoupling of a simple monotonic ion in water solvent, we should be safely in the Gaussian regime, as demonstrated in Table S3 of the SI (see related discussion herein).

In these NE simulations, the bound zinc(II) atom is progressively annihilated. In order to avoid singularity problems, the fast annihilation process is performed according to the prescription given in ref 25, that is, by first discharging the Zinc cation with a linear scaling from 0 to 75 ps and then, in the last 15 ps, by bringing the Lennard-Jones potential to zero using the Beutler soft-core parametrization.[61] In the 18 complexes, ΔG_{ki}, computed from the corresponding work distribution using eq 14, ranged from 397 to 420 kcal mol^{-1} (see Table S2 of the SI) and the behavior of the decoupling Zinc(II) was very similar in all cases: as shown in the example reported in Figure S1 of the SI, the zinc(II) atom leaves the MBET306$^-$ partner when the discharging process is almost complete, drifting away from the anion. A common standard state correction volume $V(\tau)$ (see eq 15) of 4187 Å^3 was then applied to the ΔG_{ki} to get the corresponding standard energies ΔG^0_{ki}. Such volume was defined as the volume of the sphere centered on the O2 atom of MBET306$^-$ that included the

DOI: 10.1021/ct500964e
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

decoupled Zinc(II) in 95% of the ending configurations obtained in all NE $N_{sites} \times N_{states} \times 256$ trajectories and produces a standard state correction of +0.555 kcal mol^{-1}.

The fast annihilation of the zinc(II) cation in water is done in a similar way by running 256 independent NE trajectories using the same alchemical protocol adopted for the bound complexes. The decoupling free energy ΔG_2 corresponds to the negative of the zinc(II) solvation free energy. The final value is 403.626 ± 0.104 kcal mol^{-1}, including the so-called finite-size correction. This term has nothing to do with the standard state correction and arises naturally in the Ewald derivation,[62] because of the electroneutrality constraint of the unit cell. In the zinc(II) decoupling process, the starting and the ending system have a total charge of +2 and 0, respectively. For charged system, in the standard Ewald treatment one must add to the electrostatic energy a term corresponding to the interaction of the excess charge of the system with a uniform neutralizing background plasma. Such term is given by[63] $-|\Sigma_i q_i|^2\pi/(2\alpha^2 V)$ where V is the volume of the box and the sum is extended to all charges in the box. For the solvated zinc(II) in our simulations this term contributes +1.03 kcal/mol to the decoupling free energy ΔG_2. Note also that the neutralizing background plasma correction term is identical in the final and initial state for process 1 and hence does not contribute to the ΔG_1 free energy.

In Table 3, we collect the results for the dissociation free energies of the 18 complexes ΔG_{ki}, and of the four rotameric/tautomeric states of the MBET306$^-$ receptor (these values lack the common +0.555 kcal mol^{-1} standard state correction that is applied to the final value). These values were obtained by summing the ΔG_1 values for Zinc(II) decoupling in the complexes and the ΔG_2 values for the Zinc(II) decoupling in bulk solvent reported in Table S1 of the SI obtained using eq 14. We note incidentally that the strongly statistically biased Jarzynski exponential averages, eq 13, also reported in Table S2 of the SI, are very close to the Gaussian average in most cases. This is due to the low mean dissipation of the fast alchemical decoupling, $W_d = \beta\sigma^2/2$, which reduces the statistical bias and constitutes a confirmation of the assumed Gaussian character of the distributions. In the last row of Table 3, we finally report the overall standard dissociation constant of the Zn-(MBET306)$^+$ complex obtained using eqs 15 and 8. The weights G'_k have been obtained using the ab initio energies of the optimized structures with Polarizable Continuum Model (PCM) reported in Table 1 of refs 27 and 64. The agreement with the experimental value is satisfactory although it should not be emphasized. In the NE simulations, we are, in fact, using for Zn(II) a nonpolarizable force field that, while reproducing correctly the Zn−O distances in the MBET306 complexes and in water (2.0−2.1 Å), completely neglects short-range polarization effects, both in the MBET306 complexes and in the $Zn(H_2O)_n$ species.[65,66] In the final FS-DAM free energy difference, error compensation may therefore play a relevant role. Basically, the dissociation free energy in Zn(MBET306)$^+$ is controlled by the four 23 complexes that have comparable ΔG_{ki} values in the range 16:17 kcal mol^{-1} with one of these states, namely, the (23)-O2-*trans* complex, prevailing over all the others. The stability of the Zinc(II) bidentate complexes in water in comparison to tridentate structures is in accord with experimental indications obtained from absorption and fluorescence excitation spectra.[27]

We now come to the issue of the moments of the work distributions, that is, to the Gaussian nature of $P(W)$. Equation 15 can be used as such if the work distribution is Gaussian. In

Table 3. Top: Standard Dissociation Free Energy ΔG^0_{ki} for Zn(II)MBET306$^+$ Complexes in Water[a] Bottom: Weight and the Zn(II) Dissociation Free Energies $\Delta G^0_k = \Sigma_i \Delta G^0_{ki}$ for the Four Conformers/Tautomers of the Receptor MBET306$^-$[b]

anion	conf.	c-type	$\Delta G_1^{(ki)} - \Delta G_2$
O2	*cis*	124	7.141 ± 0.111
O2	*cis*	12	4.254 ± 0.449
O2	*cis*	14	−0.667 ± 0.051
O2	*cis*	23	16.888 ± 0.102
O2	*cis*	34	−5.194 ± 0.059
O2	*trans*	124	7.568 ± 0.102
O2	*trans*	12	6.713 ± 0.169
O2	*trans*	14	−0.425 ± 0.092
O2	*trans*	234	7.228 ± 0.286
O2	*trans*	23	16.229 ± 0.157
O2	*trans*	34	−3.458 ± 0.047
O3	*cis*	14	2.850 ± 0.086
O3	*cis*	23	16.849 ± 0.153
O3	*trans*	13	−1.490 ± 0.083
O3	*trans*	14	1.782 ± 0.083
O3	*trans*	234	7.182 ± 0.150
O3	*trans*	23	17.276 ± 0.195
O3	*trans*	34	−0.757 ± 0.236
anion	conf.	$G_k/\Sigma_l G_l$	$\Delta G_1^{(k)} - \Delta G_2$
O2	*cis*	0.003	16.894 ± 0.320
O2	*trans*	0.996	16.226 ± 0.400
O3	*cis*	0.0003	16.845 ± 0.395
O3	*trans*	0.0001	17.280 ± 0.445
ΔG_{Exp}	$k_BT \ln(V(\tau)/V_0)$	$\Delta G_1 - \Delta G_2 + k_BT \ln(V(\tau)/V_0)$	
16.2 ± 0.6	0.555	**16.782 ± 0.399**	

[a]Column 1 defines the anionic species. Column 2 defines the conformational state of the MBET306 anionic ligand relative to the dihedral angle C3−C4−N5−C6. Column 3 defines the zinc(II) complex binding type (see main text for definition). The errors on the free energies have been obtained evaluating the distribution by block bootstrapping, that is, using 50 sets each containing 128 works randomly sampled out of the 256 set of works. [b]In the last row, the overall dissociation free energy for the reaction of eq 16 is reported along with the standard state correction. All energies are in kcal mol^{-1}.

Figure 3 we report the distribution of the work obtained in the alchemical decoupling of the zinc(II) atom in 8 of the 18 Zn(MBET306)$^+$ complexes, including the four 23-type complexes contributing the most to the overall free energy. The light lines are the numerical distribution with a bin size of 1 kcal/mol resulting from the 256 NE alchemical works. The thick smooth curves are not the result of a fit but simply correspond to the normal distributions obtained from the first two moments of the raw work data, that is, $n(W) = (1/(\sigma 2\pi^1/_2))e^{-(W-\langle W\rangle)/2\sigma^2}$. In Table S2 of the SI we further report the skewness, $\langle (W - \langle W\rangle)^3/\sigma^3\rangle$, and excess kurtosis, $\langle (W - \langle W\rangle)^4/\sigma^4\rangle - 3$, for all NE work distributions. The values of the higher moments of $P(W)$ are in most cases extremely small, and in many cases, they can be considered zero with standard error. Figure 3 and Table S2 of the SI represent a convincing manifestation of the inherent Gaussian nature of the distribution, that is, of the Markovian character of the alchemical decoupling of the bound state in FS-DAM. As a further check of the validity of eq 14 in the case of fast switching alchemical transformation in water, in Table S4 of the SI, we compare, for the small molecules analyzed in ref 25, the

DOI: 10.1021/ct500964e
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

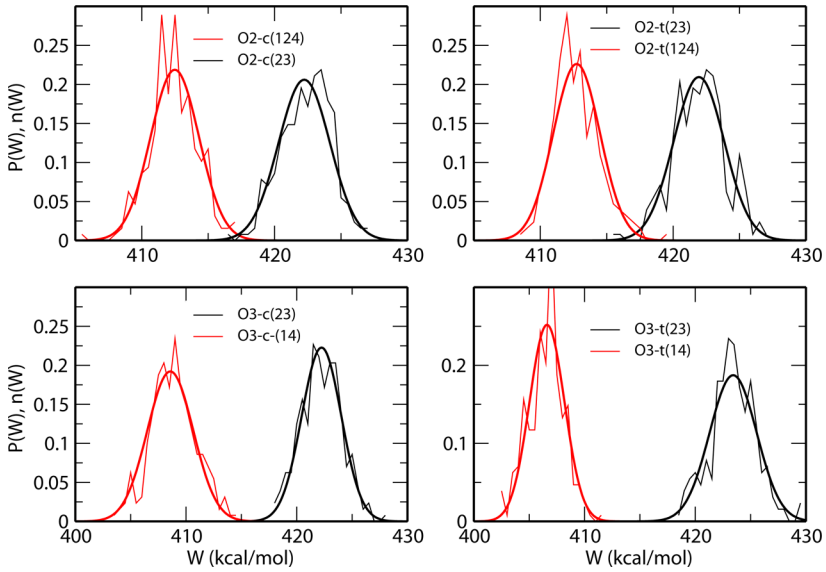


Figure 3. Work distributions in the zinc(II) decoupling of eight $Zn(MBET306)^+$ complexes in water. The light lines refer to the numerical distributions $P(W)$ obtained from the 256 NE trajectories. The thick lines refer to the normal distributions $n(W)$ obtained from the first two moments of $P(W)$.

solvation free energies in water computed according to the Gaussian assumption eq 14 and that computed by applying the Crooks theorem to bidirectional experiments (i.e., using the forward and backward distributions as they are, with no assumption whatsoever on their nature). These FS simulations were performed using the same alchemical protocol used in the present study.

Efficiency of DAM and FS-DAM on Parallel Platforms. Equation 15 underpins the extraordinary computational efficiency of FS alchemical decoupling in recovering ΔG. The underlying work distribution may be in fact constructed in parallel in a single run of the order of few tens of picoseconds, lasting, for medium size ligand−receptor systems, from a few minutes to a few tens of minutes assuming one processor per each of the N_{traj} trajectories. This single embarrassingly parallel[26] step substitutes all the intermediate simulations with the discretized $\lambda_i > 0$ parameters in the reversible DAM. This fact immediately suggests a simple and perfectly viable way to evaluate ΔG in case of a receptor with multiple binding sites. It suffices, in fact, to prepare with conventional MD the starting canonically sampled configurations for the NE process using all possible bound complexes. Such step is identical to the $\lambda = 0$ calculation in reversible DAM. From a technical standpoint, as the starting bound states in multiple binding site systems all share, in general, the same topology,[67] the computation of ΔG_{ki} terms in eq 8 implies a workload in terms of CPU that is $N_{sites} \times N_{states}$ larger with respect to the single site computation, with no impact on the wall clock time that remains identical to that of the single site computation if $N_{traj} \times N_{sites} \times N_{states}$ processors are available for the computation. In the case of Zn-$(MEBT306)^+$ we need 18×256 processors for computing all work distributions in a single parallel run lasting a few minutes.

The great advantage of FS-DAM with respect to reversible DCM (r-DCM) is an inherently (thermodynamically it should be said) parallel computation. On the basis of what we have learned so far for FS-DAM, we can quantify this statement by evaluating a realistic estimate of the ratio of the theoretical wall clock times needed for the determination of the binding free energy in FS-DAM and r-DAM as a function of the number of atoms N and of the bound conformations $N_{sites} \times N_{states}$ in the system. We exclude from this evaluation the $\lambda = 0$ point that is performed in the same way in the two methods. We start by defining the rate S (i.e., the CPU time needed to run serially 1 ns of simulation of a given system), n_{traj}, n_λ (i.e., the number of NE trajectories and of λ_i points in $[0,1]$ per binding site in FS-DAM and r-DAM, respectively), and $N_s = N_{sites} \times N_{states}$, representing the manifold of bound conformations.

Then, total CPU time needed in FS-DAM and in r-DAM to evaluate the decoupling free energy is given by

$$t_{CPU}(\text{FS-DAM}) = f_1 S N_s n_{traj} \tag{18}$$

$$t_{CPU}(\text{r-DAM}) = f_2 S N_s n_l \tag{19}$$

where f_1 and f_2 are the simulation times (in ns) needed for running one NE trajectory in FS-DAM and one equilibrium simulation of a λ_i point in r-DAM, respectively. We now assume that the number of available processors, N_p, largely exceeds both n_{traj} and n_λ, as is commonplace on HPC platforms. The wall-clock time with N_p processors can then be evaluated as

$$t_{wall}(\text{FS-DAM}) = N_s \frac{f_1 S}{\alpha(n_1) n_1} \tag{20}$$

$$t_{wall}(\text{r-DAM}) = N_s \frac{f_2 S}{\alpha(n_2) n_2} \tag{21}$$

where we have defined the integers $n_2 = N_p/(N_s n_\lambda)$ and $n_1 = N_p/(N_s n_{traj})$ ($n_1 > 1$ and $n_2 > 1$) and where $\alpha(n) = 1/(an + (1 - a))$ is the parallel efficiency, that is, the ratio between the actual and the theoretical speed-up at n processors according to Amdhal law[68] (with a being the serial fraction of time in standardly parallelized MD codes). In eq 21 we have tacitly assumed that the independent $N_s n_\lambda$ r-DAM simulations and the independent $N_s n_{traj}$ FS-DAM simulations can be run perfectly in parallel with no communication overhead or parallel degradation ($\alpha = 1$). We can eliminate the size dependent S parameter by evaluating the ratio

DOI: 10.1021/ct500964e
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

$$r = \frac{t_{\text{wall}}(\text{r} - \text{DAM})}{t_{\text{wall}}(\text{FS} - \text{DAM})} = \frac{f_2 \alpha(n_1) n_1}{f_1 \alpha(n_2) n_2} \tag{22}$$

We stress that in eq 22 there is still an explicit N_s dependence in the integers n_1 and n_2. We now estimate r in a worst-case scenario for FS-DAM. We have seen that in FS-DAM, for the case of the Zn(MBET306)$^+$ complex, the NE simulation lasted 90 ps. For drug-sized molecules, this simulation time depends only on the binding strength and not on the receptor's size and can be even made shorter for ligand−receptor pairs that are less tightly bound than the complex ZnMBET306$^+$. We set in any case $f_1 = 0.1$. For the r-DAM, we have, in general, about 50−100 λ points[2,3,38] for an accuracy below 0.5 kcal mol^{-1}. n_λ is essentially size independent but is a function of the underlying free energy derivative. A larger n_λ is needed for steep $\Delta G(\lambda)$ profiles. Unlike in FS-DAM, in r-DAM f_2 depends on the size of the system in a nontrivial way. In general, large systems require longer simulation times. If we assume linear scaling, for a typical ligand−receptor system (say 10000 atoms) we may very generously set $f_2 = 1$ ns of serial time simulation per λ_i point. Actually, f_2 should also include a size-dependent equilibration stage usually of a comparable duration, to the least. In any case, we set $f_2 = 10^{-4}N$. Using typical values for $n_{\text{traj}}, n_\lambda$, we set, once for all, $n_{\text{traj}} = 256$ (as in the MBET306 tight binding Zinc(II) complexes) and $n_\lambda = 64$.[3] The number of needed processors to perform an MD simulation of a given length in a given fixed wall time can be taken to be a linear function of N, say $N_p = 10^{-2}N$, such that for $N = 1000$ atoms one gets a reasonable optimal choice of $N_p = 10$ (i.e., 100 atoms per cell in the standard domain decomposition parallel algorithms).[69] Below this limit, the interdomain communication overhead dominates over the time spent in the parallelized computation of the forces within each domain and parallelization becomes counterproductive. With this regard, the use of the Amdhal law in evaluating the parallel efficiency $\alpha(n)$, where the communication overhead is assumed to be zero, is also unfair for FS-DAM whose computational efficiency relies on the fact that no communications are need to run the independent n_{traj} trajectories. Using the above-defined parameters, we finally get an expression of the ratio as a function of N and N_s:

$$r(N, N_s) = 10^{-3} N \frac{aN + 0.64(1 - a)N_s}{aN + 0.256(1 - a)N_s} \tag{23}$$

Equation 23 shows that on parallel platforms the ratio of the wall clock time of FS-DAM over r-DAM is a linear function of N. For a system of 100 000 atoms and $N_s = 1$, FS-DAM is expected to be about 2 orders of magnitude times faster than r-DAM. In Figure 4 we show the function $r(N,N_s)$ for various values of N_s. Clearly, in the limit $N_s \to \infty$, the impact of the serial fraction reduces to zero and the function r goes to the limiting ratio of 10^{-3} N. The linear regime is reached faster for large values of the unparallelized fraction a. There are two factors that contribute to the superior efficiency of FS-DAM over r-DAM. The first is that the equilibrium-based techniques on large systems requires, in general, longer simulation times to attain convergence of the statistical averages. In NE techniques, such system size dependency affects only the sampling of the starting configurations of the $\lambda = 0$ while the length of the fast NE run depends on the type of binding, on solvent diffusivity, but not on the receptor's size. The second crucial issue concerns accuracy. A very simple way to increase the accuracy

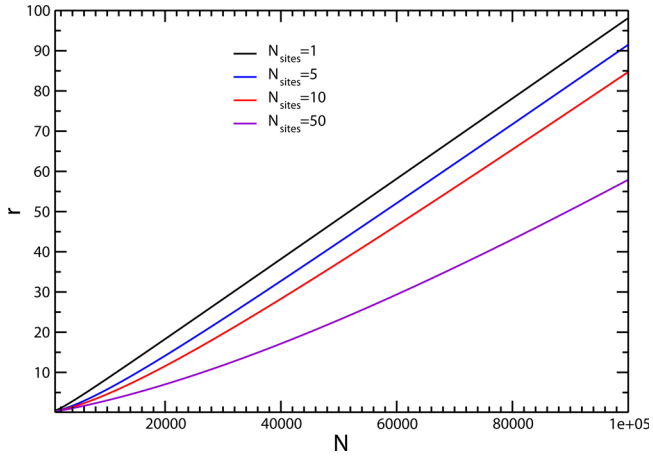


Figure 4. Wall clock time ratio between r-DAM and FS-DAM for various values of N_s (eq 23). 256 NE runs and 64 λ_i points are assumed in FS-DAM and r-DAM, respectively. The fraction of serial code with respect to conventional MD parallelization is assumed to be $a = 0.001$. Independent NE runs in FS-DAM and independent λ_i simulations in r-DAM are assumed to be perfectly parallelizable.

in FS-DAM simply consists in raising the number of NE trajectories, with no impact on the wall clock time. In r-DAM, accuracy is affected by two conflicting factors from a computational standpoint (i.e., by the number n_λ of points in $[0,1]$ and by the length of the simulations at fixed λ_i). A choice must hence be made on how the available computational resources should be optimally distributed, taking into account that raising n_λ reduces the parallel instances devoted to the equilibrium MD at fixed λ_i, with a negative impact on the wall clock time.

CONCLUSIONS

In this study, we have presented a nonequilibrium approach for binding free energy calculation in molecular dynamics simulation. The method, FS-DAM, is an irreversible variant of the equilibrium Double Annihilation Method[6] and is hence based on the production of many fast unidirectional alchemical decoupling processes of the ligand in the bound state and in the bulk solvent. The decoupling free energies of these two processes are recovered from the computed nonequilibrium alchemical decoupling works by applying the Jarzynski theorem. In order to avoid the calculation of strongly statistically biased exponential averages, FS-DAM exploits the Markovian nature of the fast decoupling process, yielding a Gaussian work distribution and hence allowing the application of the unbiased Jarzynski estimate $\Delta G = \langle W \rangle - (\beta\sigma^2/2)$ for Gaussian distributions. The nontrivial issue of the standard state correction in the context of nonequilibrium experiments has been addressed using an approach proposed in the past for unrestrained DAM.[35] The proposed methodology, being based on an unrestrained decoupling of the bound state, does not require any prior knowledge of the binding site volume.

The algorithm has been successfully tested in the calculation of the overall dissociation free energy in a multiple binding site system model (i.e., the zinc(II) complexes of the bis-amide of the L-tartaric acid MBET306 in explicit water). The efficiency of the FS-DAM on parallel platforms has been discussed and compared to the conventional methods based on equilibrium simulations. FS-DAM nonequilibrium technique, by releasing the system size dependency of the simulation time needed to

DOI: 10.1021/ct500964e
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

achieve converged statistical averages in equilibrium-based techniques, possesses the computational potential for unraveling multiple site or allosteric binding processes in complex biomolecular systems.

A straightforward evolution of the proposed FS-DAM could be that of evaluating the decoupling free energies of the ligand using bidirectional rather than unidirectional experiments with exploitation of the Crooks theorem[41] and of the accurate Bennett acceptance ratio.[32] This approach can be used effectively and without accuracy degradation also when dealing with non-Markovian decoupling process producing multimodal non-Gaussian work distributions. The price to pay in setting up a Crooks-based bidirectional scheme is that of doubling the computational effort and to impose a restraint in the bound state such that the decoupling process can be rigorously inverted. In order to do so, the restraint (e.g., in the form of a simple harmonic potential that keeps the ligand in the binding site) should be fully enforced in the final decoupled state of the ligand−receptor system and let progressively to vanish while the ligand is alchemically switched on. The contribution of such restraint to the decoupling free energy can be then evaluated ex post analytically or numerically, as done in the Double Decoupling Scheme method.

■ ASSOCIATED CONTENT

Supporting Information

Time record of the Zn−O_i distances in the Zinc(II) decoupling process for the 23 Zn(MBET306)$^+$ complexes in water. Free energy data for the decoupling of all Zn(MBET306)$^+$ complexes in water and of Zinc(II) in water with standard errors as obtained from FS-DAM simulations. Skewness and kurtosis data for the decoupling of all Zn(MBET306)$^+$ complexes in water with standard errors. Calculation of the G_k weights by means of H-REM and fast switching annihilation/creation of the O_2 and O_3 anionic species. Assessment of the Gaussian nature in the work distribution in the fast decoupling of some polar and apolar molecules in water. This material is available free of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: caminati@unifi.it.
*E-mail: procacci@unifi.it.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The computing resources and the related technical support used for this work have been provided by CRESCO/ENEAGRID High Performance Computing infrastructure and its staff; see http://www.cresco.enea.it for information. This work was funded in part by a Fulbright Research Fellowship to Mr. Robert Sandberg.

■ REFERENCES

(1) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047−1069.
(2) Gumbart, J. C.; Roux, B.; Chipot, C. *J. Chem. Theory Comput.* **2013**, *9*, 974−802.
(3) Deng, Y.; Roux, B. *J. Phys. Chem. B* **2009**, *113*, 2234−2246.
(4) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(5) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(6) Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; TiradoRives, J. *J. Chem. Phys.* **1988**, *89*, 3742−3746.
(7) Jorgensen, W. L. *Acc. Chem. Res.* **1989**, *22*, 184−189.
(8) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562−12566.
(9) Barducci, A.; Bussi, G.; Parrinello, M. *Phys. Rev. Lett.* **2008**, *100*, 020603.
(10) Fidelak, J.; Juraszek, J.; Branduardi, D.; Bianciotto, M.; Gervasio, F. L. *J. Phys. Chem. B* **2010**, *114*, 9516−9524.
(11) Vargiu, A. V.; Ruggerone, P.; Carloni, A. M. P. *Nucleic Acids Res.* **2008**, 5910−5921.
(12) Patel, J. S.; Branduardi, D.; Masetti, M.; Rocchia, W.; Cavalli, A. *J. Chem. Theory Comput.* **2011**, *7*, 3368−3378.
(13) Colizzi, F.; Perozzo, R.; Scapozza, L.; Recanatini, M.; Cavalli, A. *J. Am. Chem. Soc.* **2010**, *132*, 7361−7371.
(14) Favia, A. D.; Masetti, M.; Recanatini, M.; Cavalli, A. *PLoS One* **2011**, *6*, e25375.
(15) Nicolini, P.; Frezzato, D.; Gellini, C.; Bizzarri, M.; Chelli, R. *J. Comput. Chem.* **2013**, *34*, 1561−1576.
(16) Gallicchio, E.; Lapelosa, M.; Levy, R. M. *J. Chem. Theory Comput.* **2010**, *6*, 2961−2977.
(17) Procacci, P.; Bizzarri, M.; Marsili, S. *J. Chem. Theory Comp.* **2014**, *10*, 439−450.
(18) Lapelosa, M.; Gallicchio, E.; Levy, R. M. *J. Chem. Theory Comput.* **2012**, *8*, 47−60.
(19) Chioccioli, M.; Marsili, S.; Bonaccini, C.; Procacci, P.; Gratteri, P. *J. Chem. Inf. Model.* **2012**, *52*, 483−491.
(20) Hari, S.; Merritt, E.; Maly, D. *Chemistry and Biology* **2013**, *20*, 806−815.
(21) Nussinov, R.; Tsai, C. *Trends Pharmacol. Sci.* **2014**, *35*, 256−264.
(22) Su, J. G.; Qi, L. S.; Li, C. H.; Zhu, Y. Y.; Du, H. J.; Hou, Y. X.; Hao, R.; Wang, J. H. *Phys. Rev. E* **2014**, *90*, 022719.
(23) Hardy, J. A.; Wells, J. A. *Curr. Opin. Struct. Biol.* **2004**, *14*, 706−715.
(24) Gapsys, V.; Seeliger, D.; de Groot, B. *J. Chem. Teor. Comp.* **2012**, *8*, 2373−2382.
(25) Procacci, P.; Cardelli, C. *J. Chem. Theory, Comp.* **2014**, *10*, 2813−2823.
(26) Foster, I. *Designing and Building Parallel Programs: Concepts and Tools for Parallel Software Engineering*; Addison Wesley Longman Publishing Co.: Boston MA, 1995.
(27) Banchelli, M.; Guardiani, C.; Tenori, E.; Menichetti, S.; Caminati, G.; Procacci, P. *Phys. Chem. Chem. Phys.* **2013**, *15*, 18881−18893.
(28) Guardiani, C.; Procacci, P. *Phys. Chem. Chem. Phys.* **2013**, *15*, 9186−9196.
(29) Dai, C.; Li, D.; Popovici-Muller, J.; Zhao, L.; Girijavallabhan, V.; et al. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 3172−3176.
(30) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300−313.
(31) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420−1426.
(32) Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245−268.
(33) Shirts, M.; Mobley, D.; Chodera, J. *Annu. Rep. Comput. Chem.* **2007**, *3*, 41−59.
(34) Chelli, R. *J. Chem. Theory Comput.* **2010**, *6*, 1935−1950.
(35) Jayachandran, G.; Shirts, M. R.; Park, S.; Pande, V. S. *J. Chem. Phys.* **2006**, *125*, 084901.
(36) In other words, in the statistical average of the quantity $I(\mathbf{r},\Omega)$ defining the complex R−L tends to zero such that the product of the $\langle I(\mathbf{r},\Omega)\rangle$ times V tends to the equilibrium constant as V tends to infinity:[47]

DOI: 10.1021/ct500964e
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

$$\langle I(\mathbf{r}, \Omega)\rangle V|_{\lim\to\infty} = V\frac{\int I(\mathbf{r}, \Omega)e^{-\beta w(\mathbf{r},\Omega)}d\mathbf{r}\, d\Omega}{\int e^{-\beta w(\mathbf{r},\Omega)}d\mathbf{r}\, d\Omega} = \frac{1}{8\pi^2}\int I(\mathbf{r}, \Omega)e^{-\beta w(\mathbf{r},\Omega)}d\mathbf{r}\, d\Omega = K \quad (24)$$

In the above equation, $K = 1/K_d$, $\mathbf{r}$ and Ω are the translational and orientational coordinates of the ligand relative to the receptor, $w(\mathbf{r},\Omega)$ is the associated potential of mean force obtained by integrating (averaging) over all solvent coordinates as well as receptor and ligand internal coordinates and $I(\mathbf{r},\Omega)$ is a step function that is equal to 1 when the complex is formed and 0 otherwise.

(37) Zhou, H.-X.; Gilson, M. K. *Chem. Rev.* **2009**, *109*, 4092−4107.
(38) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. *J. Phys. Chem. B* **2003**, *107*, 9535−9551.
(39) Holt, D. A.; Konialian-Beck, A. L.; Oh, H. J.; Yen, H. K.; Rozamus, L. W.; Krog, A. J.; Erhard, K. F.; Ortiz, E.; Levy, M. A.; Brandt, M.; Bossard, M. J.; Luengo, J. I. *Bioorg. Med. Chem. Lett.* **1994**, *4*, 315−320.
(40) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(41) Crooks, G. E. *J. Stat. Phys.* **1998**, *90*, 1481−1487.
(42) Gore, J.; Ritort, F.; Bustamante, C. *Proc. Nat. Acad. Sci. U.S.A.* **2003**, *100*, 12564−12569.
(43) Hummer, G. *J. Chem. Phys.* **2001**, *114*, 7330−7337.
(44) Marcinkiewic, J. *Math. Z.* **1939**, *44*, 612−618.
(45) Procacci, P.; Marsili, S.; Barducci, A.; Signorini, G. F.; Chelli, R. *J. Chem. Phys.* **2006**, *125*, 164101.
(46) Park, S.; Schulten, K. *J. Chem. Phys.* **2004**, *120*, 5946−5961.
(47) Luo, H.; Sharp, K. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 10399−10404.
(48) Armistead, D.; Boger, J.; Meyers, H.; Saunders, J.; Tung, R. Immunosuppressive compounds. U.S. Patent No. 5,330,993, 1994; http://www.google.com.ar/patents/US5330993.
(49) Armistead, D.; Boger, J.; Meyers, H.; Saunders, J.; Tung, R. Suppressing graft rejection. U.S. Patent No. 5,516,797, 1996; https://www.google.com.ar/patents/US5516797.
(50) Becke, A. D. *Phys. Rev. A* **1988**, *33*, 3098.
(51) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(52) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926−935.
(53) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712−725.
(54) Gallicchio, E.; Levy, R. M. *J. Comput. Chem.* **2004**, *25*, 479−499.
(55) Tomasi, J.; Mennucci, B.; Cancés, E. *J. Mol. Struct.: THEOCHEM* **1999**, *464*, 211−226.
(56) Marchi, M.; Procacci, P. *J. Chem. Phys.* **1998**, *109*, 5194−5202.
(57) Nosé, S. *Mol. Phys.* **1984**, *52*, 255−268.
(58) Tuckerman, M.; Berne, B. J. *J. Chem. Phys.* **1992**, *97*, 1990−2001.
(59) Procacci, P.; Darden, T.; Paci, E.; Marchi, M. *J. Comput. Chem.* **1997**, *18*, 1848−1862.
(60) Marsili, S.; Signorini, G.; Chelli, R.; Marchi, M.; Procacci, P. *J. Comput. Chem.* **2010**, *31*, 1106−1116.
(61) Beutler, T.; Mark, A.; van Schaik, R.; Gerber, P.; van Gunsteren, W. *Chem. Phys. Lett.* **1994**, *222*, 5229−539.
(62) de Leeuw, S. W.; Perram, J. W.; Smith, E. R. *Proc. R. Soc. London A* **1980**, *373*, 27−56.
(63) Darden, T.; Pearlman, D.; Pedersen, L. G. *J. Chem. Phys.* **1998**, *109*, 10921−10935.
(64) The weights G'_k can also be computed directly from the MD simulations by combining the H-REM data taken from Banchelli et al., *Phys. Chem. Chem. Phys.* **2013**, *15*, 18881−18893 and the fast switching alchemical simultaneous creation/annhilation of the entire O_3 and O_2 species. Details of these calculations are given in the Supporting Information.
(65) Pavlov, M.; Siegbahn, P. E. M.; Sandström, M. *J. Phys. Chem. A* **1998**, *102*, 219−228.
(66) Tam, H. H.; Asthagiri, D.; Paulaitis, M. E. *J. Chem. Phys.* **2012**, *137*, 164504.
(67) Actually, the 11 complexes of the O_2 anion have the same topology. The 7 complexes of the O_3 tautomeric anion have a different topology.
(68) Amdahl, G. M. Validity of the Single Processor Approach to Achieving Large Scale Computing Capabilities. *Proc. AFIPS '67 Spring Joint Computer Conference* **1967**, 483−485 http://doi.acm.org/10.1145/1465482.1465560.
(69) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.

DOI: 10.1021/ct500964e
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX