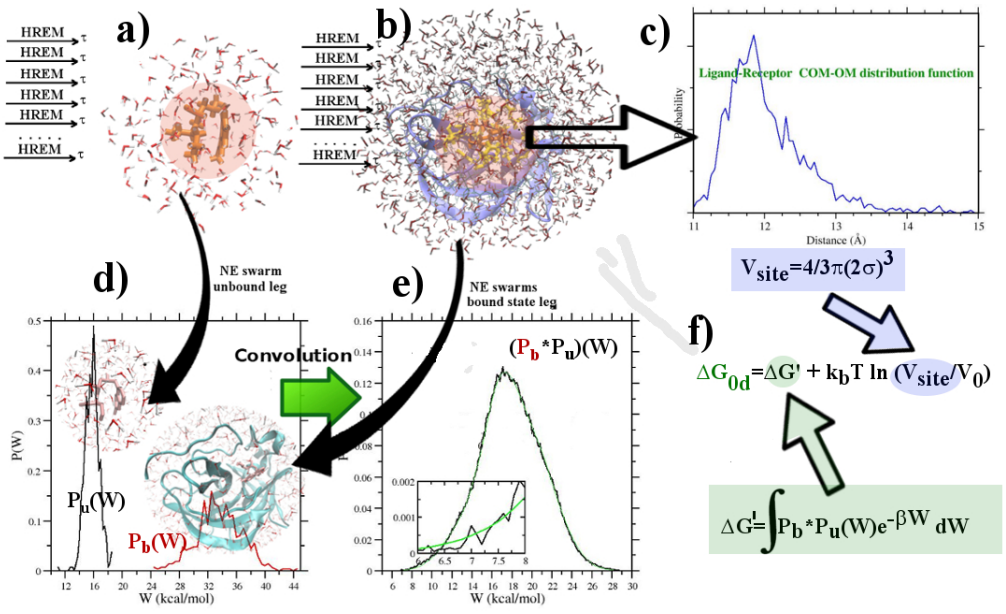
Graphical Abstract

Methodological uncertainties in drug-receptor binding free energy predictions based on classical molecular dynamics

Piero Procacci



Highlights

Methodological uncertainties in drug-receptor binding free energy predictions based on classical molecular dynamics

Piero Procacci

- Computational approaches are becoming an essential tool in modern drug design and discovery
- Automated workflows from triaging via docking to molecular dynamics (MD) approaches are being actively developed for drug discovery in a virtual hit-to-lead sprit.
- Reliable determination of absolute binding free energy (ABFE) via MD on HPC system is key requirement for virtual screening in industrial and academic settings
- Free energy perturbation methods (FEP) for ABFE of drug-receptor systems are plagued by uncertaintis related to sampling and protocols,
- The HPC-tailored nonequilibrium approach, combining multiple enhance sampling simulations with fast-swicthing alchemical methods, can deliver accurate estimates and credible confidence intervals for ABFE

Methodological uncertainties in drug-receptor binding free energy predictions based on classical molecular dynamics

Piero Procacci

Dipartimento di Chimica "Ugo Schiff", Università degli Studi di Firenze, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy

Abstract

Computational approaches are becoming an essential tool in modern drug design and discovery, with fast compound triaging using a combination of machine learning and docking techniques followed by molecular dynamics binding free energies assessment using alchemical techniques. The traditional MD-based alchemical free energy perturbation (FEP) method faces severe sampling issues that may limits its reliability in automated workflows. Here we review the major sources of uncertainty in FEP protocols for drug discovery, showing how the sampling problem can be effectively tackled by switching to nonequilibrium alchemical techniques.

Keywords:
Molecular dynamics, Free energy perturbation, drug design, drug discovery, nonequilibrium, Crooks theorem, Alchemical methods

1. Introduction

Molecular dynamics-based (MD) techniques using state of the art force fields such as AMBER/GAFF[1], CHARMM/CgenFF[2] and OPLS,[3], are considered as an essential and powerful tool for reliably predicting binding free energies in drug-receptor systems. In drug discovery projects, due to the high computational demand, MD methods are being increasingly used[4, 5, 6] in a post-docking refinement stage, via the implementation of partially automated virtual screening workflows in a hit-to-lead spirit.

In the last decades, MD methodologies have been devised for solvation and binding free energy calculations, with the so-called alchemical route emerging as one of the most powerful approaches.[5] Alchemical approaches, in essence, evaluates the *absolute* binding free energy (ABFE) as the difference of the solvation energy of the ligand in the bound state and in the bulk. These solvation energies, in turn, are computed by progressively decoupling the ligand from the environment, either along a stratification[7] of discrete *equilibrium* intermediate λ-states using free energy perturbation[8] (FEP) or, equivalently, thermodynamic integration[9] (TI), or by varying continuously the λ coupling alchemical parameter in a swarm of fast independent and concurrent trajectories, exploiting the Jarzynski[10] and Crooks[11] theorems on the resulting nonequilibrium work (NEW) distribution.

Most of the drug discovery alchemical applications, paralleling the medicinal chemistry practice, deal with the calculations (via FEP[12, 13, 6, 14, 15] or NEW[16, 17]) of *relative* binding free energies (RBFE), evaluating the free energy cost of *transmuting* a ligand into a strictly congeneric compound, a process involving, in general, a relatively small (few kcal/mol) perturbation. On the other hand, as recently noted[5], efficient ABFE approaches are urgently needed in the implementation of virtual screening funnel workflows from docking-based triaging to MD-based methodologies.[18] Docking campaigns on large compounds databases may in fact produce chemically distant hits that are not easily amenable for RBFE calculations.

Despite the last decades progresses, FEP-based ABFE for drug design still constitutes an awesome challenge as these methodologies involve large perturbations (tens of kcal mol) facing hurdles and entanglements related to the need for equilibrium sampling on *each* λ-state of the discrete alchemical stratification. As acutely pointed out in Ref. [19], at low coupling, alchemical simulations experience order-disorder transi-

Email address: procacci@unifi.it (Piero Procacci)

tions to entropically favored states where the weakly coupled ligand can freely rotate and translate within the binding site region as opposed to the ordered high enthalpy states observed at the full coupling. These order-disorder transitions cause entropic bottlenecks which hinder the equilibration and convergence of binding free energy estimates. These entanglements are commonly bypassed by enforcing a set of restraint potentials[20] limiting the conformational activity of the ligand and hence the need for extensive sampling, shifting *de facto* the "sampling issue"[21] to the highly non trivial evaluation of the free energy cost of imposing and releasing those restraints[4, 22].

Quite curiously, in the authoritative perspective paper by Cournia and coworkers,[5] while FEP methodologies are amply and critically discussed, nonequilibrium (NE) alchemical techniques are not even mentioned, despite their recent success in the recent SAMPL rounds,[23, 24] as well in retrospective applications[25, 26]. In the SAMPLing challenge,[24], in particular, where a variety of methodologies were systematically compared for ABFE predictions, a NEW-based approach, the nonequilibrium switch double system single box (NS-DSSB),[27] remarkably obtained the overall highest efficiency and accuracy in the CB8-quinine system where both the host and the guest exhibited long correlation times and sampling challenges.

In the present contribution, we will review the FEP and NEW alchemical approach concerning the crucial and undervalued aspect of the reproducibility of the ABFE and of the determination of a credible confidence interval, a quantity of no less importance of the prediction itself, being strictly related to the investment risk in industrial drug discovery projects. We shall focus here on the statistical uncertainty deriving from the methodological and computational protocol, disregarding the systematic errors that may arise from force field deficiencies, a matter that is the object of a continuous and intense specialistic research.[28, 29, 30]

2. Uncertainties in alchemical FEP methodologies

In modern FEP applications, the bound and unbound ligand decoupling (solvation) free energies are computed as a sum of the $n-1$ individual free energy contributions along the λ-stratification as

$$\Delta G_i^{i+1} = \mathcal{E}[P_i(\Delta V_i^{i+1}), P_{i+1}(\Delta E)_{i+1}^i] \quad (1)$$

$$\Delta G = \sum_i^{n-1} \Delta G_i^{i+1} \quad (2)$$

where $P_i(\Delta V_i^{i+1}) = \langle \delta(\Delta V - (V_{i+1}(x) - V_i(x))) \rangle_i$, $P_{i+1}(\Delta V_{i+1}^i) = \langle \delta(\Delta V - (V_i(x) - V_{i+1}(x))) \rangle_{i+1}$ are the distributions of the potential energy difference between thermodynamic states with the alchemical parameters λ_{i+1} and λ_i and $\langle\cdot\rangle_i$,$\langle\cdot\rangle_{i+1}$ indicate canonical averages using the λ_i and λ_{i+1} Hamiltonian, respectively. $\mathcal{E}(\cdot)$ is a functional representing the estimate of the logarithm of the ratio two contiguous partition functions, $\Delta G_i^{i+1} = k_B T \ln Z_i/Z_{i+1}$.[31]

$\mathcal{E}(\cdot)$ corresponds, in the vast majority of FEP applications, to the Bennett Acceptance Ratio (BAR)[32, 33] BAR provides an accurate and precise estimate for the individual ΔG_i^{i+1}, so long that the contiguous potential energy distributions $P_i(\Delta V_i^{i+1})$ and $P_{i+1}(-\Delta V_{i+1}^i)$ do have a significant *overlap*. The λ protocol (i.e. the number and spacing of the λ states) should be chosen to yield a significant and approximately constant overlap along the alchemical stratification, to correspondingly make the uncertainty on each of the $n-1$ ΔG_i^{i+1} approximately constant, minimizing the overall uncertainty.[34] ΔG_i^{i+1} could be also computed as a functional of *all* the forward and reverse $n-1$ energy distributions, an estimator known as multiple Bennett Acceptance ratio (MBAR),[35] yielding however an only marginal increase in precision and accuracy.[31, 36, 37, 14]

Ultimately, the FEP *dissociation* free energy estimate is given by

$$\Delta G'_d = \Delta G_b - \Delta G_u \quad (3)$$

where the suffix b, u refers to the bound and unbound state for the ligand and where $\Delta G_b, \Delta G_u$ are computed according to Eqs. 1, 2. The prime indicates that the dissociation free energy must be corrected by a standard state dependent term related to the *binding site volume*.[38, 39] We will return on this subtle point later on in this section.

Given a mean to compute the individual uncertainties $\delta\Delta G_i^{i+1}$, the overall uncertainty of the estimate can be obtained by *summing in quadrature* the errors along the bound and unbound alchemical stratification. In most FEP calculations, individual uncertainty, $\delta\Delta G_i^{i+1}$, are evaluated either by computing the variance on block averages or by bootstrapping techniques.[7] This is accomplished usually by way of black-box post-processing application scripts (e.g. `gmx bar` in gromacs[40] or `pymbar`[35] for AMBER[1]). Summation of the errors in quadrature is based on the tacit assumption that repeated calculations of the individual ΔG_i^{i+1} yield n-1 *independent* and normally distributed random variables (RV).

Actually, repeating a FEP computation for a complex drug-receptor system, starting, e.g., from differ-

ently prepared initial conditions or using a slightly different FEP protocol, may often produce a free energy estimate that differs from the original by a quantity largely exceeding the uncertainty evaluated using the data of a single FEP simulation.[36, 41, 42, 14] This is due to the way canonical (ensemble) averages, $\bar{A} = \langle A \rangle$, are estimated as *time* averages $\bar{A} = \frac{1}{\tau}\int_0^\tau A(t)dt$, in the assumption that ergodicity holds in the time τ, or stated in other terms, that the sampling with respect to all relevant coordinates is canonical in the time τ.

In the practice of FEP applications, τ is generally and arbitrarily chosen equal for all λ states in a range from few to few tens of nanoseconds, while convergence rates may differ substantially with different ligand coupling.[43, 44, 19] Side chains conformational motions, on the other hand, are observed in NMR experiments on a *microsecond* time range.[45] A simple movement such as the DFG flip, marking the active and inactive state in the *apo* state of kinases, is believed to occur on a *millisecond* time scale,[46] which means that just *one* sudden DFG flip per millisecond is observed on the average in a single molecule. Paradoxically, the advent of GPUs in scientific calculations, that allows simulating a typical drug-receptor system for up to hundreds ns/day, has strengthened the illusion that a *single* sufficiently long MD trajectory can achieve correct sampling in FEP applications of complex biomolecular systems.

More than two decades ago, it was authoritatively[48] recognized that "individual trajectories of length up to 5 ns [at that time, 5 ns sounded like an eternity] sample only a fraction of the conformational distribution generated by ten independent 120 ps trajectories at 300 K". This fact can be understood using the ball maze vintage game metaphor (see Figure 1). The holes in the board are akin to attractors (e.g. conformational states) in a protein system. While many short trajectories started from infinitesimally different initial conditions can sample several attractors like an ejected ball from the same spot can end up each time in a different hole, a single trajectory may get stuck in one of the attractors for a long time before it can jump to a different conformational states. Stated in other terms, staring at just *one* molecule in the hope of observing a rare event (e.g. a conformational transition) while such event is occurring in many of the molecules of the thermodynamic ensemble right behind the observer, is a rather tenuous approach. As a result, "one-off" FEP simulations are in general poorly reproducible.[49, 50]

These concepts have been recently formalized in terms of ergodicity, sensitivity to initial conditions of deterministic iterators[51], equilibrium and chaos in Hamiltonian systems[42] and discussed in a series of

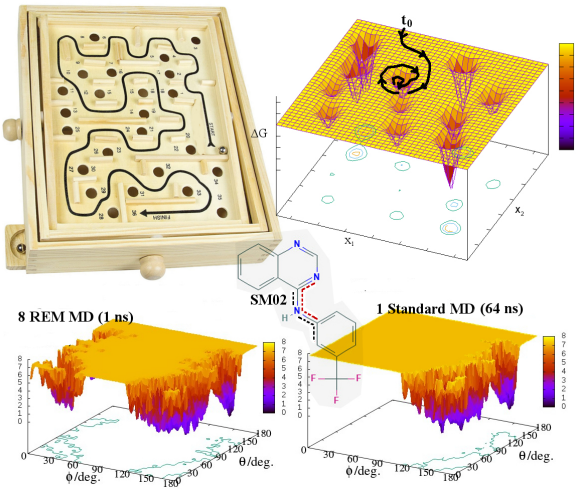


Figure 1: Upper panels: vintage ball maze gameboard (left) and hypothetical 2D free energy surface of a complex system with different attractors (metastable free energy minima). Lower panel: 2D free energy surface (FES) obtained for the two indicated dihedral angles marking the gauche ($|\theta|, |\phi| \leq 60°$), anti ($120° \leq |\theta|, |\phi| \leq 180°$) states of a synthetic precursor of kinase inhibitors (taken from the SAMPL6 challenge[47]) in water in standard conditions ; on the left, the 2D-FES computed using 8 batteries of HREM simulations (with eight replicas each) lasting 1 ns (for a total of 64 ns); on the right the same surface as obtained by the sampling of one single standard MD simulation in 64 ns. While the 1 ns HREM simulations effectively sampled all 4 conformational attractors in SM02, the standard MD, was incapable of sampling the gauche-gauche and gauche-anti attractors in 0.06 μs of simulation.[44]

remarkable papers on MD applications for drug design by Coveney and coworkers[42, 49, 37] In Ref. [37], in particular, the uncertainty of FEP-determined ABFE for some BRD4, FGFR1 and thrombin complexes was quantified and assessed using various enhanced sampling approaches, including multiplets of λ-hopping Hamiltonian replica exchange (HREM) simulations with solute tempering (REST)[52] at intermediate λ's (a methodology also known as FEP+)[53, 15] and multiplets of REST for each λ states. In REST, the "heating" via energy scaling (and hence kinetic boosting) is limited to the so-called[12] "hot zone" including the ligand and the nearby residues. Despite the tremendous CPU time invested, results for the ABFE were still affected by large and highly system-dependent uncertainties evaluated using random combinations of multiple simulations that were found to "vary by as much as 2.6, 6.5, and 7.6 kcal/mol for BRD4, FGFR1, and thrombin".

Some remarks are in order on the calculation of the standard state volume term, that when added to Eq. 3, should yield the standard dissociation energy. In the FEP practice, such term is estimated from the difference between the free energy of imposing a restraint potential (usually a harmonic function involving translational, orientational and conformational degrees of freedom of the ligand[20]) in the binding site at the full ligand coupling, minus the free energy of releasing that restraint at zero coupling (or *viceversa*). These restraints are introduced to alleviate the sampling issues characterizing flexible ligands and protein side chains.[4] In the strong restraint limit, the contribution of the translational restraint can be shown[39, 54] to be equal to $RT\ln(V_{\rm site}/V_0)$ (where $V_{\rm site}$ is the allowance volume of the ligand in the binding pocket) constituting a penalty for the dissociation free energy, $\Delta G_0 = \Delta G' + RT\ln(V_{\rm site}/V_0)$, so long that $V_{\rm site} < V_0$, with $V_0 = 1661$ Å^3 being the standard state volume. While the zero-coupling restraint contribution is computed analytically, the free energy cost of the restraints at full coupling in virtually all FEP applications for ABFE determination is inappropriately computed again via an FEP-like approach where the restraints are progressively switched on (or off), in few windows and in few tens of ns *in total* at best, with the ligand lingering in the *presumed* binding site with the *presumed* conformation/orientation. Needless to say that if the presumptions were wrong, the prediction is also wrong. Besides, with no or very weak restraint potential (i.e. in the first strata of the FEP-restraint approach) for the fully coupled ligand in the bound state, a *true* equilibrium sampling would directly allow the estimate the binding affinity by simply evaluating the probability ratio of the bound and unbound states.[55] A more sensible approach of this questionable FEP-restraint computational practice has been recently proposed by Heinzelmann and Gilson[4] where the authors evaluated the binding free energies ΔG_i for *multiple* ligand-receptor poses (i.e. with no assumption on the "best" ligand pose), recovering the free energy as $\Delta G = -RT\ln(\sum_i e^{-\beta\Delta G_i})$.

3. NEW-based approach for ABFE calculations

In NEW-based techniques, the connection between the end-states is performed by a swarm of fast nonequilibrium trajectories rather than by a stratification of equilibrium λ states. These λ-driven NE trajectories are started from a canonical (equilibrium) sampling of one end-state ending up in nonequilibrium configurations of the other end-state. As such, at variance with FEP, the *sense* of the transformation is important in NEW, that is, the distributions of the work values (that are related to the free energy via the Jarzynski or Cooks theorems) can be markedly different performing the process in one sense or the other.

In this respect, it has been shown[31, 24] that whenever the NE transformation involves the entrance into or escape from a free energy funnel (such as the folding of a protein or the formation of a drug-receptor complex), the process is much less *dissipative* in the escape direction. The dissipation is defined as the difference between the mean NE work done in the independent driven processes and the underlying free energy. If the induced-fit upon binding involves important conformational reorganization in the protein pocket, recoupling a ghost ligand starting from the equilibrium *apo* state of a protein via fast (few hundreds picoseconds) NE trajectories can be a tremendously dissipative process, with a high probability of producing a manifold of suboptimal NE poses characterized by negligible Boltzmann weight. In the escape direction, the fast NE decoupling of a well fit bound ligand, yields, in general, less dissipation as the NE end-states (free protein and gas-phase ligand) involve no mutual conformational clashes.

In NEW, the alchemical free energies are a functional of the work distributions and the uncertainty is strictly related to the dissipation of the process, proportional to the inverse of the duration time τ of the NE trajectories[56]. For normal work distributions, it can be shown[57, 16, 36, 25] that the leading term of the uncertainty in the NEW estimates, is proportional to $\sigma^2/(k_BTn^{1/2})$ where n is the number of collected NE work values and σ^2 is the variance of the work distribution.

While the unbound state leg of the NEW alchemical thermodynamic cycle can be performed in either direction or both using bidirectional estimators such as BAR, the bound state leg should be performed in the less dissipative *annihilation* direction, unless imposing restraints that would imply, as we have discussed previously, strong (and possibly wrong) assumptions on the binding poses.

The NEW thermodynamic cycle can be effectively unified[58] in a protocol whereby the ligand undergoes n_b fast-annihilation and n_u fast-growth in the bound state and in the bulk, respectively, implementing a sort of "virtual" DSSB approach.[27] The resulting independent n_u and n_b-sized work histograms, $P_u(W)$, $P_B(W)$, are then convoluted to yield a statistically boosted work histogram $(P_B * P_u)(W)$ constructed using $n_u \times n_b$ independent values $W = W_b + W_u$, lowering the uncertainty to $(\sigma_u^2 + \sigma_b^2)/[k_B T(n_u n_b)^{1/2}]$. These concepts are illustrated in Figure 2. The standard state correction in NEW is implemented by estimating the translational volume V_{site} from the fluctuation of the COM-COM distance[59]. As pointed out in Ref. [59], such correction can be a source of uncertainty, partially mitigated by the logarithmic dependency of the volume ratio. We recall that binding site volume determination is the rather undervalued weak point of *any* computational approach based on the definition of "bound state", including of course FEP-based techniques for ABFE *and* RBFE, as well. The latter implicitly (and arbitrarily) assume the constancy of the binding site volume upon the transmutation of the bound ligand into another bound parent compound.

4. Conclusion

There are several aspects in favor of the NEW alchemical approach for ABFE in drug design. In NEW, the equilibrium sampling is required only for the end-states and such sampling can be effectively obtained, as we have seen (see Figure 1), using batteries of concurrent relatively short enhanced sampling simulations, an algorithm that is perfectly tailored for modern homogeneous or heterogeneous parallel computing (HPC) platforms. Such enhanced sampling of the end-state in the bound state can be performed by "heating" along the HREM progression all atoms in the binding site (REST), imposing a *weak* harmonic restraint between the centers of mass (COM) of the fully coupled ligand and the receptor and hence allowing an *unrestrained* sampling of the conformational/orientational states of the bound ligand and nearby residues. The end-states of the unbound state can be generated essentially at no cost by performing multiplets of HREM on an isolated (gas-phase) molecule and combining the so sampled gas-phase states of the ghost ligand with pre-equilibrated samples of the solvent. The second step in NEW corresponds to the embarrassingly parallel production of NE decoupling/recoupling alchemical trajectories, again a computation that can be efficiently implemented on an HPC platform. Given that the sampling of the end-states is accurate, accuracy and precision in NEW-based ABFE estimates depend only on the dissipation and on the resolution of the convoluted work histogram $(P_B * P_u)(W)$. This is strikingly at variance with FEP techniques, where accuracy and precision are an unknown function of the energy distributions in *all* λ windows, *de facto* preventing a reliable estimate of the confidence interval of the prediction in "one-off" calculations.

NEW efficiency, accuracy, and precision has been amply assessed in recent studies.[36, 56, 60, 16, 17] Besides producing accurate and reproducible results for solvation energies[36, 61], RBFE[16, 17] and ABFE[59] estimates, NEW provides by design a credible methodological confidence interval, a fundamental quantity in an industrial setting. Despite these features, and despite its consistently good performances in recent blind challenges for ABFE predictions[24, 62], NEW is still scantly used compared to FEP-based approaches, both in academic and pharmaceutical contexts. Popular MD engines, such as gromacs, AMBER or OpenMM, already support HREM and fast switching alchemical schemes, the two key ingredients in NEW. What is probably deterring end-users in selecting NEW for ABFE and RBFE calculations is the lack of software tools for the complex pre- and post-processing of the two computational steps in NEW, namely the preparation of the enhanced sampling of the equilibrium end-states and the manipulation of the work data resulting from the fast-switching stage. Such tools, such as `pmx`[63] (gromacs) or `BAT.py`[4] (AMBER) or `Flare`[6] (OpenMM) have been recently developed and tailored for FEP-based alchemical applications and could be easily adapted for NEW alchemy as well.

5. Conflict of interest statement

Nothing declared

6. References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

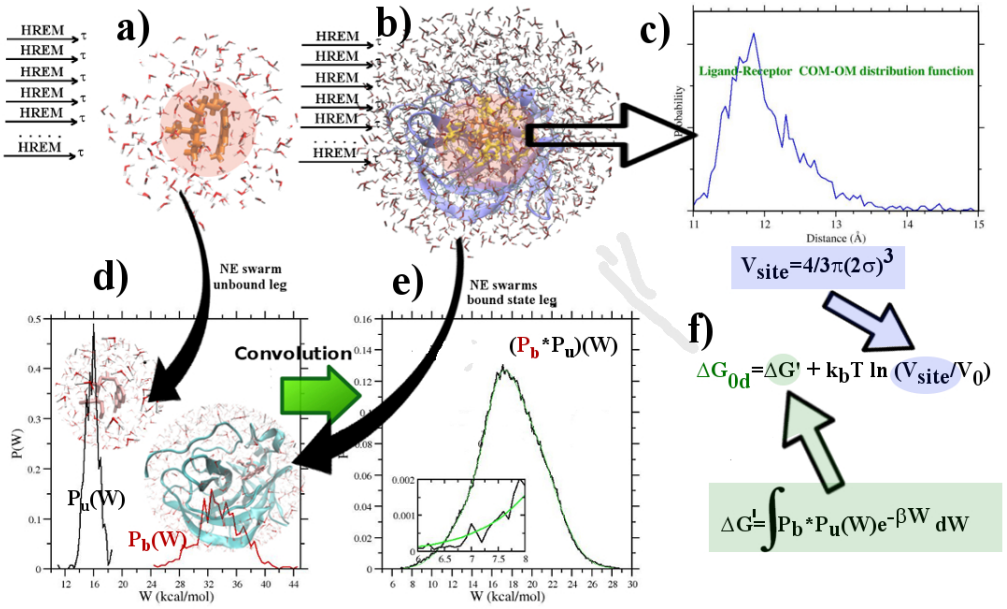


Figure 2: *Virtual* DSSB workflow in NEW. Enhanced sampling of the **a)** unbound state and of the **b)** bound state (the "hot zone in HREM is highlighted in the orange spheres). **c)** Volume COM-COM fluctuations in the bound state HREM. **d)** NE recoupling (unbound) and decoupling (bound) runs yielding the $P_u(W)$, $P_b(W)$ work histograms. **e)** $(P_B * P_u)(W)$ convolution process. **f)** calculation of the standard dissociation free energy using V_{site} and $(P_B * P_u)(W)$. Confidence interval is computed by bootstrapping on the W_u and W_b collection prior to convolution.

- • of special interest
- •• of outstanding interest

References

[1] Romelia Salomon-Ferrer, David A. Case, and Ross C. Walker. An overview of the amber biomolecular simulation package. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 3(2):198–210, 2013.

[2] Sunhwan Jo, Wei Jiang, Hui Sun Lee, Benoit Roux, and Wonpil Im. Charmm-gui ligand binder for absolute binding free energy calculations and its application. *J. Chem. Inf. Model.*, 53(1):267–277, 2013.

[3] Edward Harder, Wolfgang Damm, Jon Maple, Chuanjie Wu, Mark Reboul, Jin Yu Xiang, Lingle Wang, Dmitry Lupyan, Markus K. Dahlgren, Jennifer L. Knight, Joseph W. Kaus, David S. Cerutti, Goran Krilov, William L. Jorgensen, Robert Abel, and Richard A. Friesner. Opls3: A force field providing broad coverage of drug-like small molecules and proteins. *J. Chem. Theory Comput.*, 12(1):281–296, 2016.

[4] •• Germano Heinzelmann and Michael K. Gilson. Automated docking refinement and virtual compound screening with absolute binding free energy calculations. *bioRxiv*, 2020.

Rigorous FEP calculations of ABFE on drug-receptor complexes using a multiple poses approach.

[5] •Zoe Cournia, Bryce K. Allen, Thijs Beuming, David A. Pearlman, Brian K. Radak, and Woody Sherman. Rigorous free energy simulations in virtual screening. *J. Chem. Inf. Model.*, Jun 2020.

Complete and accurate review on the current status and role of FEP-based ABFE incalculations in virtual screening.

[6] •Maximilian Kuhn, Stuart Firth-Clark, Paolo Tosco, Antonia S. J. S. Mey, Mark Mackey, and Julien Michel. Assessment of binding affinity via alchemical free-energy calculations. *J. Chem. Inf. Model.*, 60(6):3120–3130, 2020.

Development of automated workflows for assessing the relability in relative binding free energy calculations.

[7] Andrew Pohorille, Christopher Jarzynski, and Christophe Chipot. Good practices in free-energy calculations. *J. Phys. Chem. B*, 114(32):10235–10253, 2010.

[8] R. W. Zwanzig. High-temperature equation of state by a perturbation method. i. nonpolar gases. *J. Chem. Phys.*, 22:1420–1426, 1954.

[9] J. G. Kirkwood. Statistical mechanics of fluid mixtures,. *J. Chem. Phys.*, 3:300–313, 1935.

[10] C. Jarzynski. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, 78:2690–2693, 1997.

[11] G. E. Crooks. Nonequilibrium measurements of free energy differences for microscopically reversible markovian systems. *J. Stat. Phys.*, 90:1481–1487, 1998.

[12] Lingle Wang, Yujie Wu, Yuqing Deng, Byungchan Kim, Levi Pierce, Goran Krilov, Dmitry Lupyan, Shaughnessy Robinson, Markus K. Dahlgren, Jeremy Greenwood, Donna L. Romero, Craig Masse, Jennifer L. Knight, Thomas Steinbrecher, Thijs Beuming, Wolfgang Damm, Ed Harder, Woody Sherman, Mark Brewer, Ron Wester, Mark Murcko, Leah Frye, Ramy Farid, Teng Lin, David L. Mobley, William L. Jorgensen, Bruce J. Berne, Richard A. Friesner, and Robert Abel. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.*, 137(7):2695–2703, 2015.

[13] Lin Frank Song, Tai-Sung Lee, Chun Zhu, Darrin M. York, and

Kenneth M. Merz. Using amber18 for relative free energy calculations. *J. Chem. Inf. Model.*, 59(7):3128–3135, 2019.

[14] •David F. Hahn, Gerhard Knig, and Philippe H. Hnenberger. Overcoming orthogonal barriers in alchemical free energy calculations: On the relative merits of λ-variations, λ-extrapolations, and biasing. *J/ Chem. Theory Comput.*, 16(3):1630–1645, 2020.

Thorough study on the merits of various enhanced sampling techniques in FEP application, highlighting the importance of the biasing approaches (e.g. REST).

[15] •Guanglei Cui, Alan P. Graves, and Eric S. Manas. Gram: A true null model for relative binding affinity predictions. *J. Chem. Inf. Model.*, 60(1):11–16, 2020.

Assessment of RBFE predictions from an industrial perspective.

[16] Matteo Aldeghi, Vytautas Gapsys, and Bert L. de Groot. Accurate estimation of ligand binding affinity changes upon protein mutation. *ACS Central Science*, 4(12):1708–1718, Dec 2018.

[17] •Vytautas Gapsys, Laura Prez-Benito, Matteo Aldeghi, Daniel Seeliger, Herman van Vlijmen, Gary Tresadern, and Bert L. de Groot. Large scale relative protein ligand binding affinities using non-equilibrium alchemy. *Chem. Sci.*, 11:1140–1152, 2020.

An extensive assessment of nonequilibroum techniques in RBFE calculations.

[18] ••John Chodera, Alpha A. Lee, Nir London, and Frank von Delft. Crowdsourcing drug discovery for pandemics. *Nature Chemistry*, 2020.

A remarkable example of the power of automated computational approaches in the fight against the Covid-19 pandemic threough open science and worldwide collaboration.

[19] ••Rajat K. Pal and Emilio Gallicchio. Perturbation potentials to overcome order/disorder transitions in alchemical binding free energy calculations. *J. Chem. Phys.*, 151(12):124116, 2019.

Very insightful and rigorous study on the origin, nature and consequences of the disordered state of the bound ligand at low alchemical coupling.

[20] Stefan Boresch, Franz Tettinger, Martin Leitgeb, and Martin Karplus. Absolute binding free energies: a quantitative approach for their calculation. *J. Phys. Chem. B*, 107(35):9535–9551, 2003.

[21] David L. Mobley. Let's get honest about sampling. *J. Comput. Aided Mol. Des.*, 26(1):93–95, Jan 2012.

[22] Niel M. Henriksen, Andrew T. Fenley, and Michael K. Gilson. Computational calorimetry: High-precision calculation of host-guest binding thermodynamics. *J. Chem. Theory and Comput.*, 11(9):4377–4394, Sep 2015.

[23] •Andrea Rizzi, Steven Murkli, John N. McNeill, Wei Yao, Matthew Sullivan, Michael K. Gilson, Michael W. Chiu, Lyle Isaacs, Bruce C. Gibb, David L. Mobley, and John D. Chodera. Overview of the sampl6 host–guest binding affinity prediction challenge. *J. Comput. Aided Mol. Des.*, 32(10):937–963, Oct 2018.

Overview on merits and critical issues of FEP-based approaches in the blind prediction of hots-guest ABFE.

[24] •• Andrea Rizzi, Travis Jensen, David R. Slochower, Matteo Aldeghi, Vytautas Gapsys, Dimitris Ntekoumes, Stefano Bosisio, Michail Papadourakis, Niel M. Henriksen, Bert L. de Groot, Zoe Cournia, Alex Dickson, Julien Michel, Michael K. Gilson, Michael R. Shirts, David L. Mobley, and John D. Chodera. The sampl6 sampling challenge: assessing the reliability and efficiency of binding free energy calculations. *J. Comput.-Aided Mol. Des.*, 34(5):601–633, 2020.

Latest SAMPLing challenge results for host-guest ABFE predictions including a highly performant NEW approach (DSSB).

[25] Piero Procacci. Myeloid cell leukemia 1 inhibition: An in silico study using non-equilibrium fast double annihilation technology. *J. Chem. Theory Comput.*, 14(7):3890–3902, 2018.

[26] Francesca Nerattini, Riccardo Chelli, and Piero Procacci. Ii. dissociation free energies in drug-receptor systems via nonequilibrium alchemical simulations: application to the fk506-related immunophilin ligands. *Phys. Chem. Chem. Phys.*, 18:15005–15018, 2016.

[27] V. Gapsys, S. Michielssens, J.H. Peters, B.L. de Groot, and H. Leonov. *Molecular Modeling of Protein*, chapter Calculation of Binding Free Energies, pages 173–209. Humana Press, 2015.

[28] Carl Caleman, Paul J. van Maaren, Minyan Hong, Jochen S. Hub, Luciano T. Costa, and David van der Spoel. Force field benchmark of organic liquids: Density, enthalpy of vaporization, heat capacities, surface tension, isothermal compressibility, volumetric expansion coefficient, and dielectric constant. *J. Chem. Theory Comput.*, 8(1):61–74, Jan 2012.

[29] Haiyang Zhang, Chunhua Yin, Yang Jiang, and David van der Spoel. Force field benchmark of amino acids: I. hydration and diffusion in different water models. *J. Chem. Inf. Model.*, 58(5):1037–1052, 2018.

[30] David L. Mobley, Caitlin C. Bannan, Andrea Rizzi, Christopher I. Bayly, John D. Chodera, Victoria T. Lim, Nathan M. Lim, Kyle A. Beauchamp, David R. Slochower, Michael R. Shirts, Michael K. Gilson, and Peter K. Eastman. Escaping atom types in force fields using direct chemical perception. *J. Chem. Theory Comput.*, 14(11):6076–6092, 2018. PMID: 30351006.

[31] Piero Procacci. Unbiased free energy estimates in fast nonequilibrium transformations using gaussian mixtures. *J. Chem. Phys.*, 142(15):154117, 2015.

[32] C. H. Bennett. Efficient estimation of free energy differences from monte carlo data. *J. Comput. Phys.*, 22:245–268, 1976.

[33] M. R. Shirts, E. Bair, G. Hooker, and V. S. Pande. Equilibrium free energies from nonequilibrium measurements using maximum likelihood methods. *Phys. Rev. Lett.*, 91:140601, 2003.

[34] Levi N. Naden and Michael R. Shirts. Linear basis function approach to efficient alchemical free energy calculations. 2. inserting and deleting particles with coulombic interactions. *J. Chem. Theory Comput.*, 2015. in press, DOI:10.1021/ct501047e.

[35] M. R. Shirts and J. D. Chodera. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.*, 129:124105, 2008.

[36] Ahmet Yildirim, Tsjerk A. Wassenaar, and David van der Spoel. Statistical efficiency of methods for computing free energy of hydration. *J. Chem. Phys.*, 149(14):144111, 2018.

[37] ••Agastya P. Bhati, Shunzhou Wan, Yuan Hu, Brad Sherborne, and Peter V. Coveney. Uncertainty quantification in alchemical free energy methods. *J. Chem. Theory Comput.*, 14(6):2867–2880, 2018.

A thorough, rigorous and insightful study for quantifying uncertainty in the calculation of absolute and relative binding free energy in protein ligand complexes using equilibrium approaches with enhanced sampling and simulations repeats.

[38] Mihail Mihailescu and Michael K. Gilson. On the theory of noncovalent binding. *Biophys. J.*, 87(1):23 – 36, 2004.

[39] Piero Procacci and Riccardo Chelli. Statistical Mechanics of Ligand-Receptor Noncovalent Association, Revisited: Binding Site and Standard State Volumes in Modern Alchemical Theories. *J. Chem. Theory Comput.*, 13(5):1924–1933, may 2017.

[40] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilrd Pll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–

25, 2015.

[41] Francesco Manzoni and Ulf Ryde. Assessing the stability of free-energy perturbation calculations by performing variations in the method. *J. Comput.-Aided Mol. Des.*, 32(4):529–536, 2018.

[42] Peter V. Coveney and Shunzhou Wan. On the calculation of equilibrium thermodynamic properties from molecular dynamics. *Phys. Chem. Chem. Phys.*, 18:30236–30240, 2016.

[43] Zhao X. Sun, Xiao H. Wang, and John Z. H. Zhang. Barbased optimum adaptive sampling regime for variance minimization in alchemical transformation. *Phys. Chem. Chem. Phys.*, 19:15005–15020, 2017.

[44] P. Procacci. Solvation free energies via alchemical simulations: let's get honest about sampling, once more. *Phys. Chem., Chem. Phys.*, 21:13826–13834, 2019.

[45] Petra Rovó, Colin A. Smith, Diego Gauto, Bert L. de Groot, Paul Schanda, and Rasmus Linser. Mechanistic insights into microsecond time-scale motion of solid proteins using complementary 15n and 1h relaxation dispersion techniques. *J. Am. Chem. Soc.*, 141(2):858–869, 2019.

[46] Yilin Meng, Yen lin Lin, and Benoit Roux. Computational study of the dfg-flip conformational transition in c-abl and c-src tyrosine kinases. *J. Phys. Chem. B*, 119(4):1443–1456, 2015.

[47] Mehtap Isik, Dorothy Levorse, David L. Mobley, Timothy Rhodes, and John D. Chodera. Octanol-water partition coefficient measurements for the sampl6 blind prediction challenge. *bioRxiv*, 2019. DOI: 10.1101/757393.

[48] Leo S. D. Caves, Jeffrey D. Evanseck, and Martin Karplus. Locally accessible conformations of proteins: Multiple molecular dynamics simulations of crambin. *Protein Science*, 7(3):649–666, 1998.

[49] Agastya P. Bhati, Shunzhou Wan, David W. Wright, and Peter V. Coveney. Rapid, accurate, precise, and reliable relative free energy prediction using ensemble based thermodynamic integration. *J. Chem. Theory Comput.*, 13(1):210–222, 2017.

[50] •Bernhard Knapp, Luis Ospina, and Charlotte M. Deane. Avoiding false positive conclusions in molecular simulation: The importance of replicas. *Journal of Chemical Theory and Computation*, 14(12):6127–6138, 2018.

On the reproducibility issues in the simulations of solvated proteins using single trajectories.

[51] Bruce M. Boghosian, Peter V. Coveney, and Hongyan Wang. A new pathology in the simulation of chaotic dynamical systems on digital computers. *Advanced Theory and Simulations*, 2(12):1900125, 2019.

[52] Wei Jiang and Benoit Roux. Free energy perturbation hamiltonian replica-exchange molecular dynamics (fep/h-remd) for absolute ligand binding free energy calculations. *J. Chem. Theory Comput.*, 6(9):2559–2565, Jul 2010.

[53] Lingle Wang, Yuqing Deng, Jennifer L. Knight, Yujie Wu, Byungchan Kim, Woody Sherman, John C. Shelley, Teng Lin, and Robert Abel. Modeling local structural rearrangements using fep/rest: Application to relative binding affinity predictions of cdk2 inhibitors. *J. Chem. Theory Comput.*, 9(2):1282–1293, 2013.

[54] Yuqing Deng and Benot Roux. Calculation of standard binding free energies: aromatic molecules in the t4 lysozyme l99a mutant. *J. Chem. Theory Comput.*, 2(5):1255–1273, 2006.

[55] Huan-Xiang Zhou and Michael K. Gilson. Theory of free energy and entropy in noncovalent binding. *Chem. Rev.*, 109:4092–4107, 2009.

[56] Piero Procacci. Accuracy, precision, and efficiency of nonequilibrium alchemical methods for computing free energies of solvation. i. bidirectional approaches. *J. Chem. Phys.*, 151(14):144113, 2019.

[57] Piero Procacci. I. dissociation free energies of drug-receptor systems via non-equilibrium alchemical simulations: a theoretical framework. *Phys. Chem. Chem. Phys.*, 18:14991–15004, 2016.

[58] Piero Procacci, Marina Macchiagodena, Marco Pagliai, Guido Guarnieri, and Francesco Iannone. Interaction of hydroxychloroquine with sars-cov2 functional proteins using all-atoms non-equilibrium alchemical simulations. *Chem. Comm.*, pages – , 2020.

[59] Piero Procacci, Massimiliano Guarrasi, and Guido Guarnieri. Sampl6 host–guest blind predictions using a non equilibrium alchemical approach. *J. Comput. Aided Mol. Des.*, Aug 2018.

[60] Piero Procacci. Precision and computational efficiency of nonequilibrium alchemical methods for computing free energies of solvation. ii. unidirectional estimates. *J. Chem. Phys.*, 151(14):144115, 2019.

[61] Piero Procacci and Guido Guarnieri. Sampl6 blind predictions of water-octanol partition coefficients using nonequilibrium alchemical approaches. *Journal of Computer-Aided Molecular Design*, 2019.

[62] The sampl7 blind prediction challenges for computational chemistry, 2019. `https://github.com/samplchallenges/SAMPL7` (accessed 23 June 2020).

[63] Vytautas Gapsys and Bert L. de Groot. pmx webserver: A user friendly interface for alchemistry. *J. Chem. Inf. Model.*, 57(2):109–114, 2017.