



UNIVERSITÀ
DEGLI STUDI
FIRENZE

FLORE

Repository istituzionale dell'Università degli Studi di Firenze

SAMPL7 blind predictions using nonequilibrium alchemical approaches

Questa è la Versione finale referata (Post print/Accepted manuscript) della seguente pubblicazione:

Original Citation:

SAMPL7 blind predictions using nonequilibrium alchemical approaches / Procacci P.; Guarnieri G.. - In: JOURNAL OF COMPUTER-AIDED MOLECULAR DESIGN. - ISSN 0920-654X. - STAMPA. - (2021), pp. 1-11. [10.1007/s10822-020-00365-3]

Availability:

The webpage <https://hdl.handle.net/2158/1221722> of the repository was last updated on 2021-01-18T10:12:57Z

Published version:

DOI: 10.1007/s10822-020-00365-3

Terms of use:

Open Access

La pubblicazione è resa disponibile sotto le norme e i termini della licenza di deposito, secondo quanto stabilito dalla Policy per l'accesso aperto dell'Università degli Studi di Firenze (<https://www.sba.unifi.it/upload/policy-oa-2016-1.pdf>)

Publisher copyright claim:

La data sopra indicata si riferisce all'ultimo aggiornamento della scheda del Repository FloRe - The above-mentioned date refers to the last update of the record in the Institutional Repository FloRe

(Article begins on next page)

Query Details[Back to Main Page](#)

No Queries

SAMPL7 blind predictions using nonequilibrium alchemical approaches

Piero Procacci, ¹✉Email procacci@unifi.itGuido Guarnieri, ²

¹ University of Florence, Department of Chemistry, Via Lastruccia n. 3, 50019 Sesto Fiorentino, FI, Italy

² ENEA, Portici Research Centre, DTE-ICT-HPC, P.le E. Fermi, 1, 80055 Portici, NA, Italy

Received: 1 July 2020 / Accepted: 27 November 2020

Abstract

In the context of the SAMPL7 challenge, we computed, employing a non-equilibrium (NE) alchemical technique, the standard binding free energy of two series of host-guest systems, involving as a host the Isaac's TrimerTrip, a Cucurbituril-like open cavitand, and the Gilson's Cyclodextrin derivatives. The adopted NE alchemy combines enhanced sampling molecular dynamics simulations with driven fast out-of-equilibrium alchemical trajectories to recover the free energy via the Jarzynski and Crooks NE theorems. The GAFF2 non-polarizable force field was used for the parametrization. Performances were acceptable and similar in accuracy to those we submitted for Gibb's Deep Cavity Cavitands in the previous SAMPL6 host-guest challenge, confirming the reliability of the computational approach and exposing, in some cases, some important deficiencies of the GAFF2 non-polarizable force field.

Keywords

SAMPL7

Binding free energy

Non-equilibrium

Crooks theorem

Fast switching

Hamiltonian Replica Exchange

HREX

Solute Tempering

Torsional tempering

Electronic supplementary material

The online version of this article (<https://doi.org/10.1007/s10822-020-00365-3>) contains supplementary material, which is available to authorized users.

Introduction

The SAMPL initiative [1, 2, 3] periodically proposes community-wide blind challenges aimed at advancing and assessing computational techniques as standard predictive tools in rational drug design. The SAMPL systems generally consist of a series of host-guest pairs for which the standard binding free energy must be predicted, given the chemical structure of the partners and the experimental conditions (pH, temperature, and pressure) used in the measurements. No information on the binding pose or protonation states of the guest/host molecules is in general provided. In this challenge, the organizers included three host systems, namely the Triptycene walled glycoluril trimer (codename TrimerTrip) [4], various mono-3-substituted β -cyclodextrin analogues [5] (codename CD), and the Gibb Deep Cavity CavitanDs or Octa-acids [6] (codename GDCC). SAMPL7 [7] was the first challenge of the series where the participants were asked to provide *just one* prediction set for *formal* ranking for each of the three challenges, while still allowing to submit multiple prospective (not to be ranked) prediction sets for the same system. Submitting only one ranked submission, “intended to be the single entry each participant expects to be best performing”, [8] forced to make uneasy choices regarding the selected computational methodology and protocol. The practice of filing multiple submissions, widely adopted in past challenges, with, e.g., slight variants of the same computational approach or using similar force fields, often resulted in “lucky shots on goal” with no clear scientific explanation.

Although multiple submissions were in most cases used by researchers for *bona fide* testing purposes, the practice somehow blurred the outcome of the challenge, making the analysis of the data less insightful and, ultimately, less useful.

In this study, we present our ranked predictions for the TrimerTrip and CD host-guest systems, done using a classical molecular dynamics (MD) approach, relying on enhanced sampling of the fully coupled end-states (bound and unbound) followed by a swarm of nonequilibrium alchemical trajectories where the guest-environment interaction is rapidly switched off by way of an alchemical coupling parameter. From the resulting work distribution, the free energy is recovered exploiting the Crooks [9] or Jarzynski [10] NE theorems. The adopted computational protocol is identical to that described in our previous paper on SAMPL6 host-guest challenge, [11] termed as fast-switching (FS) double annihilation method [12, 13, 14] (FSDAM). Here, we select to focus on the CD and TrimerTrip systems, as, unlike the relatively rigid GDCC host, the TrimerTrip and CD hosts were characterized by highly flexible substituents bearing a variety of polar groups, likely to interact and interfere with the “normal binding” modes of the guests in the CD or CB-like cavities and hence posing challenging sampling issues.

The paper is organized as follows. In section “Methods” we describe the FSDAM computational approach providing technical details concerning the MD simulation protocols. In section “Results”, we critically discuss our prediction sets for CD and TrimerTrip host-guest pairs, comparing our data to those of the other ranked submissions. In section “Discussion”, we make some general considerations on the MD-based techniques adopted in the challenge with a focus on uncertainty determination. In Section “Potential pitfalls and weak points in the FSDAM”, we describe pitfalls and entanglements of FSDAM theory for standard binding free energy calculations. Conclusions and perspectives are presented in the last section.

Methods

Structural details of the host-guest systems for the TrimerTrip and CD SAMPL7 challenges can be found at Ref. [8]. All simulations were performed with the ORAC6.1 program [15] using the fast switching double annihilation method as described in Refs. [11, 12, 13, 14] Briefly, in FSDAM the dissociation free energy is determined in two consecutive stages: in the first stage, the bound and unbound states are sampled at full coupling using Hamiltonian Replica Exchange (HREM) with Solute Tempering (REST) [16, 17]. In the second stage (fast switching, FS), starting from the canonical sampling at full coupling, we launch a swarm of independent and concurrent non-equilibrium trajectories where the ligand-

environment alchemical coupling parameter is rapidly switched off to zero coupling (ligand in the gas phase). The annihilation free energy in the two branches of the alchemical thermodynamic cycle is recovered from the work distribution exploiting the Crooks and Jarzynski theorems.

The Force Field (FF) parameters (bonded parametrization, fixed atomic charges and Lennard-Jones) and topology of the host and guests molecules were prepared using the PrimaDORAC interface [18]. PrimaDORAC uses the GAFF2 parameter set and AM1/BCC atomic charges. Protonation states of guest and host molecules are assigned using OpenBabel [19] at pH 7.4. No counter-ions were used. A background neutralizing plasma was assumed within the PBC Ewald method (PME). Each system was solvated in about 1000 OPC3 [20] water molecules in a cubic box of ≈ 31 side-length.

The initial configurations were prepared by molecular docking using a solvent-accessible surface area approach for implicit solvation [21] (SASA), according to the following scheme: the guest molecule was inserted with random orientations by placing its center of mass (COM) with uniform probability within a sphere of 6 Å centered at the COM of the host in the configuration provided by the organizer. From these randomly generated starting configurations (480), the system was minimized assuming full flexibility for both ligand and host using conjugate gradient. The starting configurations for the challenge is that with the smallest dissociation energy among the 500 docking runs. For the CD host-guest starting poses, the polar groups (OH and NH_3^+) of the guest molecules ~~and~~ are pointing upward towards the wider edge of the truncated cone. In six cases, g2–g24, g2–g34, g2–g36, g1–bcd g1–g09 g1–g35, we tested also the other pose with the polar groups pointing downward on the narrow edge, obtaining in most cases a negligible contribution to the binding free energy.

Production MD simulations (both REST and FS) were run in the isothermal-isobaric ensemble, with temperature control ($T=298$ K) using a Nose-Hoover thermostat and pressure control ($P=1$ atm) provided a Parrinello-Rahman barostat with isotropic constraints [22]. Lennard-Jones non-bonded interactions were truncated with a 13.0 Angstrom cutoff, whereas long-range electrostatics were handled with the PME method using $\alpha = 0.37 \text{ \AA}^{-1}$, 1 Å spacing for the gridded charge array and a 4-th order B-spline interpolation. SHAKE constraints were applied to bonds involving hydrogen atoms, and the simulation was integrated using a five-step RESPA integrator [22].

In the REST stage, we scaled the torsional and 14 non-bonded interactions of the solute (host+guest) up to 0.1 (i.e. 3000 K) using 8 replicas with the scaling protocol described in Ref. [23]. The average exchange rate was 30%. For the bound state, the REST simulations collected 480 configurations at regular intervals in eight 8-replicas independent HREM replicates (64 MPI processes), each lasting 2.2 ns, for a total of 18 ns on the target state. For the unbound state we launched eight independent 8-replicas HREM simulations (64 MPI processes), each lasting 1.2 ns for a total of 9.6 ns on the target state, sampling again 480 configurations at regular intervals. In the bound state, a weak COM-COM harmonic tethering potential ($k=0.052 \text{ kcal mol}^{-1} \text{ \AA}^2$) was imposed to prevent the guest to drifting off in the solvent.

In the FS stage, the guest, in the bound and unbound state, was annihilated in 0.36 and 0.24 ns, respectively, in 480 independent trajectories starting from the corresponding points sampled in the REST stage. The annihilation protocol, optimized in previous studies [24, 25] for minimizing the dissipation, for the bound state stipulates that atomic charges are linearly switched off in the first 120 ps, 70% of Lennard-Jones (LJ) interactions are linearly switched off in the following 120 ps, and finally the remainder of the LJ potential is annihilated using a soft-core regularization [26] for low coupling. For the unbound state, we used a similar protocol where charges are linearly annihilated in the first 90 ps, followed by the switching off of the LJ interaction in the final 150 ps with soft-core regularization.

Annihilation free energy estimates (bound and unbound states) are based on the work distribution produced in the FS stage. Before processing the work distribution for the bound state, the correlation between work values and corresponding starting COM-COM distance was checked to ensure that in none of the initial REST states the ligand left the binding pocket (when this happens an outlier lower work value is associated to an unbound state with outlier COM-COM distance). For the CD systems, in three cases (g2-35, g1-19, g1-36) we discarded a small fraction work values corresponding to over-sampled secondary poses (see Section 5). In this regard, we point out that REST with torsional tempering (TT) enhances *conformational* sampling of the ligand and the host, but does not accelerate translational passive diffusion. If the work distribution passed the Anderson-Darling (AD) [27] and the Jarque-Bera (JB) [28] normality tests, the annihilation free energy is calculated using the Gaussian estimator [29, 30, 31]. When the work distribution is found to be non-normal, the statistically boosted Jarzynski average is used, exploiting the decorrelation between discharging work and Lennard-Jones annihilation work [25]. Error on the free energy of annihilation estimates were

computed by bootstrapping with re-sampling in all cases. Gaussian estimates are more accurate (when normality tests are passed) than Jarzynski boosted averages but are less precise. In the TrimerTrip submission, several work distributions (either bound or unbound) failed the normality tests with the binding free energies of g2,g3,g5-g7,g9,g10,g12,g17 being estimated using the Jarzynski boosted averages. In the CD ranked submissions, all estimates are based on the Gaussian assumption as all distribution (bound or unbound) passed the AD test normality tests after checking the COM-COM distance work correlation.

Convergence analysis for bound and unbound ligand annihilation free energy estimates can be straightforwardly and reliably assessed by using the work values obtained from phase space points sampled in each of the independent HREM replicates.

A finite-size correction to the dissociation free energy due to net charges on the ligand has been calculated as described in Ref. [11]. The standard state correction to the dissociation free energy for translational restraint is given by

$\Delta G_{\text{SSC}} = RT \ln(V_{\text{site}}/V_0)$, where V_{site} is the binding site volume [14, 32]. V_{site} is computed from the variance of the guest-host COM-COM distances monitored during the REST stage as $V_{\text{site}} = 4\pi[(2\sigma)^3/3]$.

Results

In the following, we shall discuss the results obtained in this challenge using the FSDAM approaches for the TrimerTrip and the CD host-guest binding free energies. Details on how the predicted free energies are computed are given in Tables S1 and S2 of the Supporting Information. The Tables in the SI include the character of the bound and unbound distribution (AD and JB test outcomes), correlation level between electrostatic and LJ work data, the binding site volume as resulting from the bound state HREM simulations, multiple pose assessment (for the CD system), and finite charge corrections.

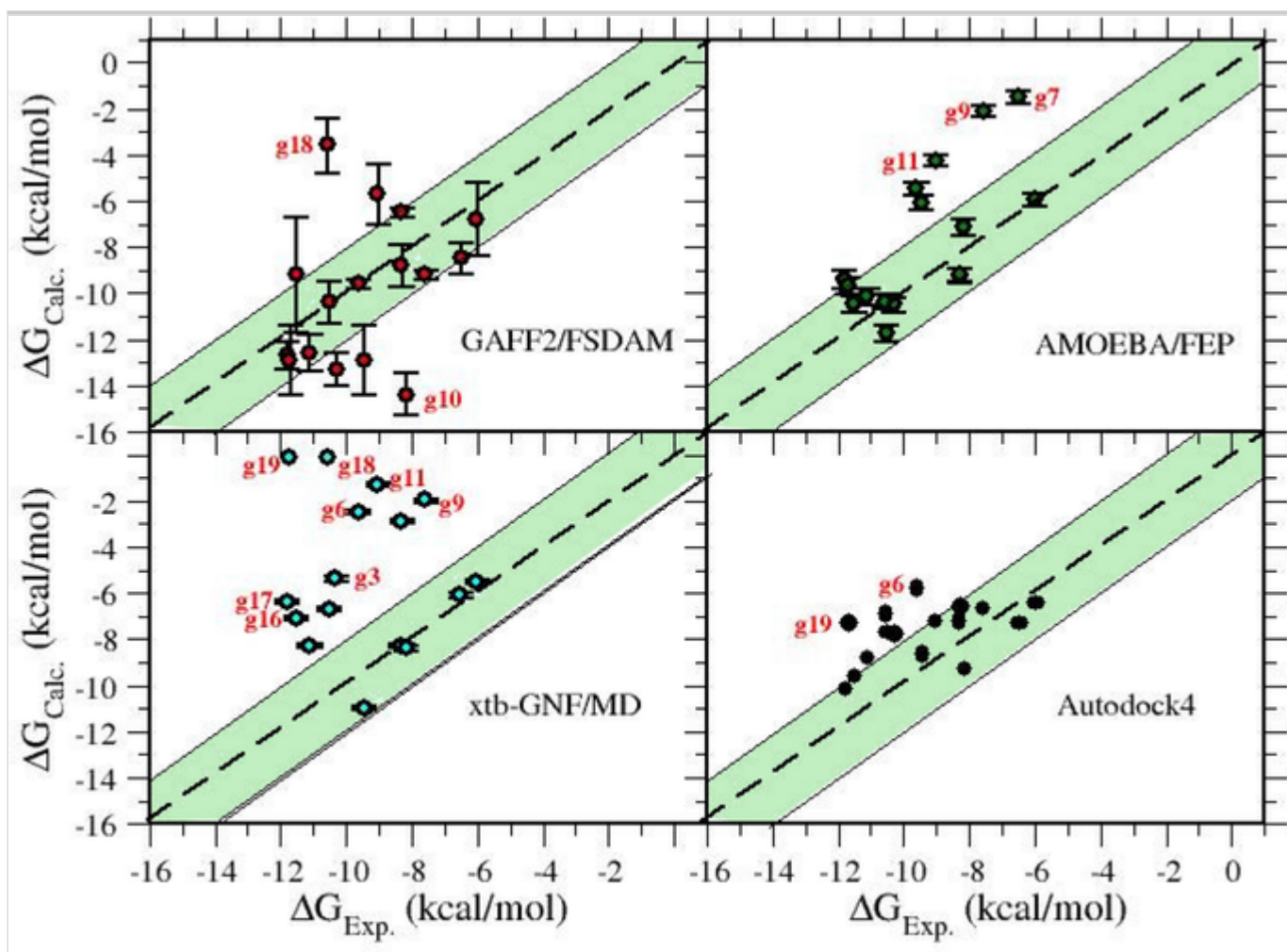
TrimerTrip

Three ranked submissions for the TrimerTrip were filed, namely the present submission (GAFF2/FSDAM), a prediction set using the AMOEBA polarizable force field [33] and [Free Energy Perturbation \(FEP\)](#), and a submission using a mixed approach, with sampling via standard MD and binding free energy calculation using the semi-empirical tight-binding xtb-GNF program developed by S. Grimme [34]. Correlation plots for the ranked submissions are reported in Fig. 1

and quality metrics of the corresponding predictions are shown in Table 1. In the Fig. 1 and in the Table 1 we also report an (unsubmitted) calculation of binding free energies using Autodock4, [35] assuming full flexibility of the ligand and rigidity of the host (using the conformation provided by the organizers at Ref. [8]). Outliers, differing by more than 4 kcal/mol with respect to the experimental value, are marked in red color. While AMEOPA/FEP appears to be the best correlated set to the experimental measures according to the Pearson correlation coefficient and to the Kendall rank coefficient τ , mean unsigned errors (MUE) are surprisingly minimal for the Autodock set, with GAFF2/FSDAM and AMOEBA/FEP exhibiting similar MUEs. The prediction set based on a mixed MD/QM (semi-empirical) approach is consistently the worst for all quality metrics. GAFF2/FSDAM and AMOEBA/FEP do not have outliers in common. Very likely, discrepancies in g18 and g10 for GAFF2/FSDAM should be ascribed to force field deficiencies, related to the fixed charge approach [36]. In the case of g18, the AM1/BCC charges could underestimate the polarization induced by the host's carboxy groups in the bound state, leading to an extra charge accumulation on the aromatic nitrogen.

Fig. 1

TrimerTrip correlation plots; Data in the shaded region are within 2.0 kcal/mol of experimental counterpart

**Table 1**

Salient data metrics for the assessment of the ranked submissions and of the Autodock prediction set (not submitted) in the TrimerTrip and CD SAMPL7 challenges: ρ : Pearson correlation coefficient; a : slope of the regression line; b : intercept of the regression line; MUE: mean unsigned error; τ : Kendall's rank coefficient

Method	ρ	a	b	MUE	τ
TrimerTrip					
Autodock	0.50	0.35	- 4.45	2.00	0.38
FEP/AMOEBA	0.71	1.24	3.94	2.10	0.47
GAFF2/FSDAM	0.35	0.61	- 4.06	2.23	0.23
xtb-GNF/MD	- 0.06	- 0.10	- 6.07	4.49	- 0.05
CD					
GAFF2/FSDAM	0.19	0.17	- 3.87	1.01	0.22

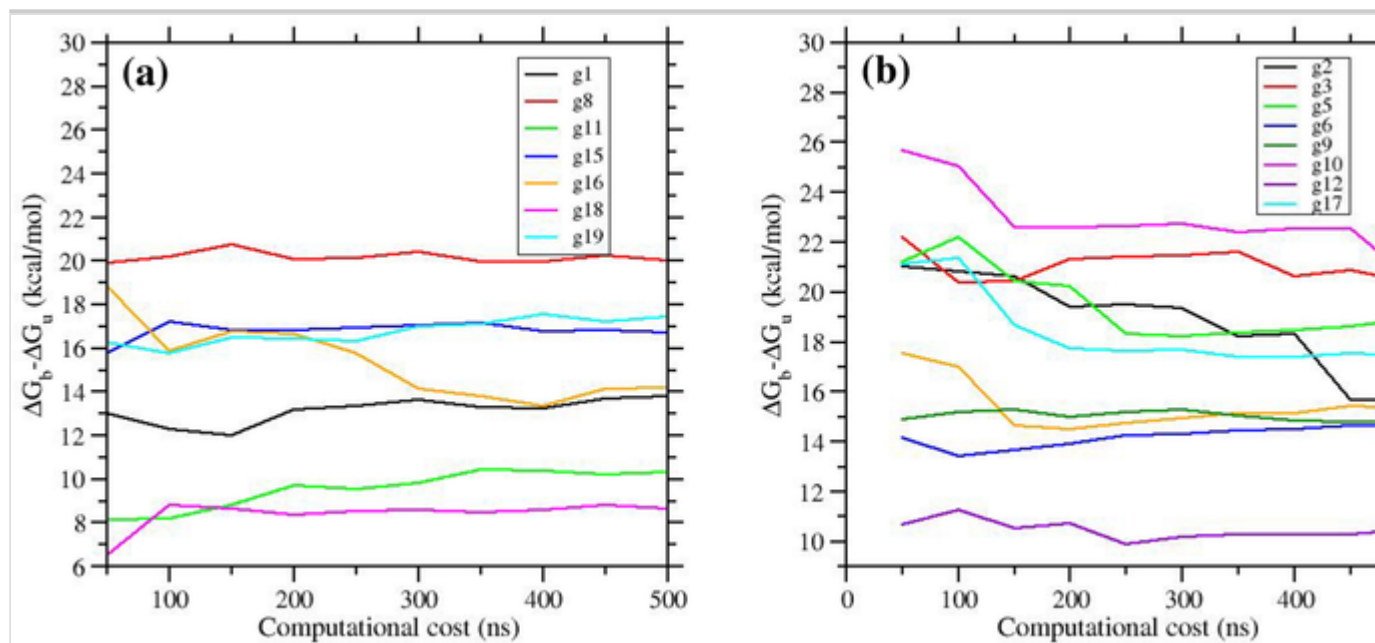
Method	ρ	a	b	MUE	τ
DSSB/NE	0.13	0.13	− 4.61	1.43	0.02
Autodock	0.13	0.17	− 4.78	1.66	0.07
QM/MM	0.10	3.68	− 20.36	32.00	0.22

The neglect of this likely polarization effect can lead to underestimation of the electrostatic contribution to the g18 decoupling in the bound state and hence to the binding affinity. For the diamondoid derivative g10, the AM1/BCC approach using MOPAC6 [37] assigns to each amino polar hydrogen a fixed charge of $0.31 e$, probably leading in this case to a systematic overestimation of the carboxy-amino electrostatic interactions in the bound state and hence to an overestimation of the binding affinity. Purging g10 and g18 from the GAFF2/FSDAM data-set results in $R = 0.67$ and $MUE = 1.61$.

Remarkably, neither g18 nor g10 are outlier in the AMEOBA/FEP prediction set which exhibit the largest discrepancies for g11 g9 and g7, due to sampling problems concerning two competing host structures [38]. While the poor performances of the xtb-GNF approach are in line with the poor performances of QM-based methods in past [host-guest](#) SAMPL challenges, [2, 3] the good results of the Autodock prediction set are surprising. Had we submitted the Autodock binding free energy data for TrimerTrip, the prediction would have ranked *first* for the MUE and second for the Pearson and Kendall coefficients.

Fig. 2

Free energy estimates of TrimerTrip host-guest systems as a function of the computational cost. The latter includes the cost of the REST and FS stages. In the y-axis, we report the difference $\Delta G_b - \Delta G_u$. In the panel (a) we report the estimates of the host-guest systems evaluated using the Gaussian assumption. In the panel (b) we show the Jarzynski boosted estimates for those pairs that failed the Anderson Darling normality tests (see section Methods)



In Fig. 2 we show the behavior of the estimates as a function of the total computational cost (on a host-guest pair basis). When using the Gaussian assumption (Fig. 2a), most of the estimates stabilize at 300 ns. Beyond 300 ns, the variability of the estimate is in general below 0.5 kcal/mol and is well within the confidence intervals computed using the full 500 ns of data (final confidence intervals are reported in Table S1 of the Supporting information). The boosted Jarzynski estimates converge less quickly, stabilizing, in most cases, beyond 400 ns. This is due to the fact that the number of combined Lennard-Jones and discharging work values used in the convolution estimate depends quadratically on the computational cost.

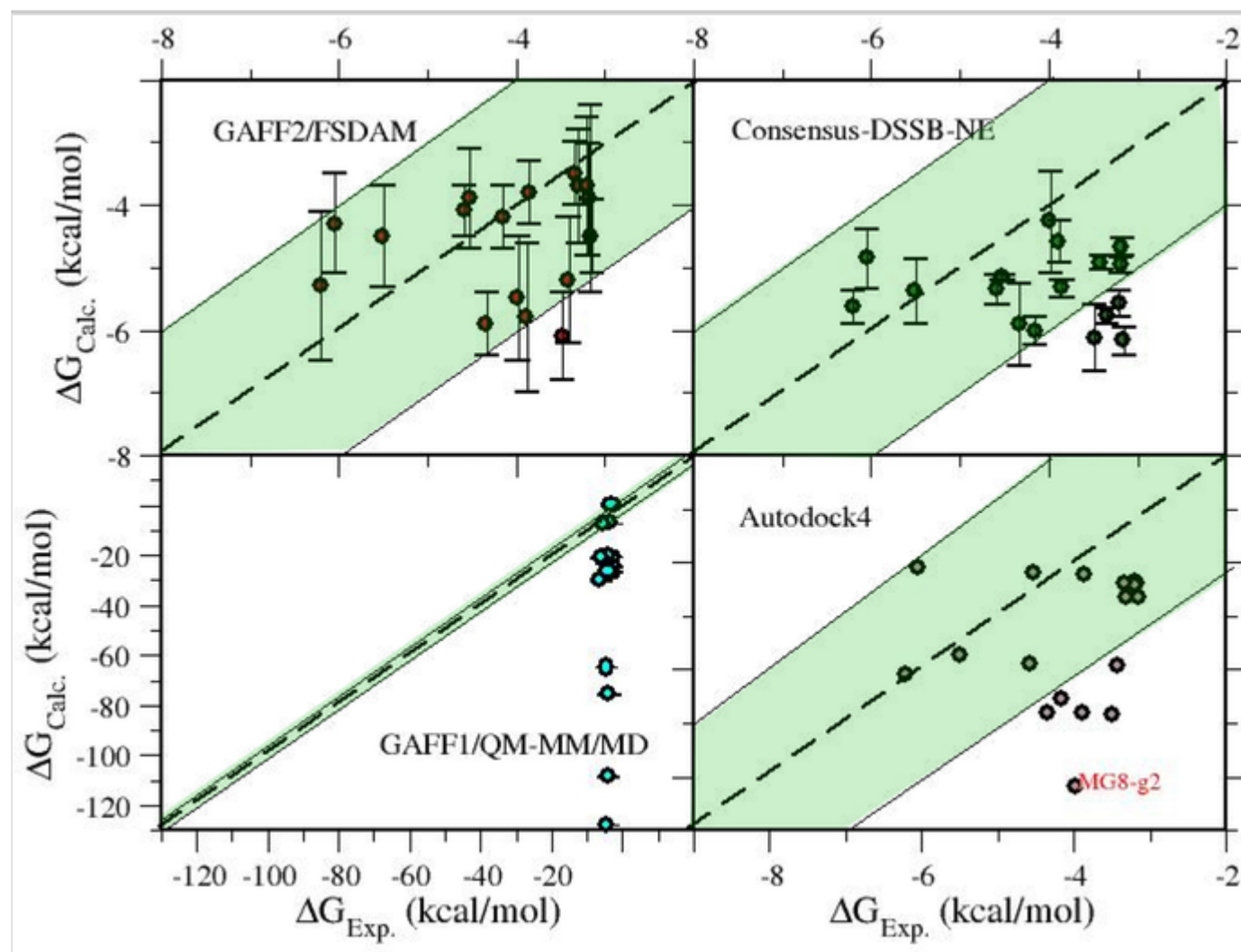
CD

In Fig. 3 we show the correlation plots for the three ranked submissions, namely that based on our GAFF2/FSDAM approach, a prediction set relying again on NE alchemical technology using the so-called double-system-single-box approach [39] (DSSB) and finally predicted binding free energies using a QM/MM approach. The DSSB set is actually the result of an arithmetic average of the free energies obtained with two sets using the GAFF1 [40] and the CgenFF [41] parametrization, exhibiting similar MUE but quite disparate Pearson and Kendall coefficients. The experimental data are clustered in a range of less than 3 kcal/mol and both GAFF2/FSDAM and DSSB correctly and remarkably predicts binding affinities within approximately the same range with no outlier. Given this small experimental range, and given that the systematic uncertainty in fixed charges force fields for

solvation free energies are of the order of 2 kcal/mol, [42] MUE appears to be the most meaningful metrics, with GAFF2/FSDAM resulting the best performing method.

Fig. 3

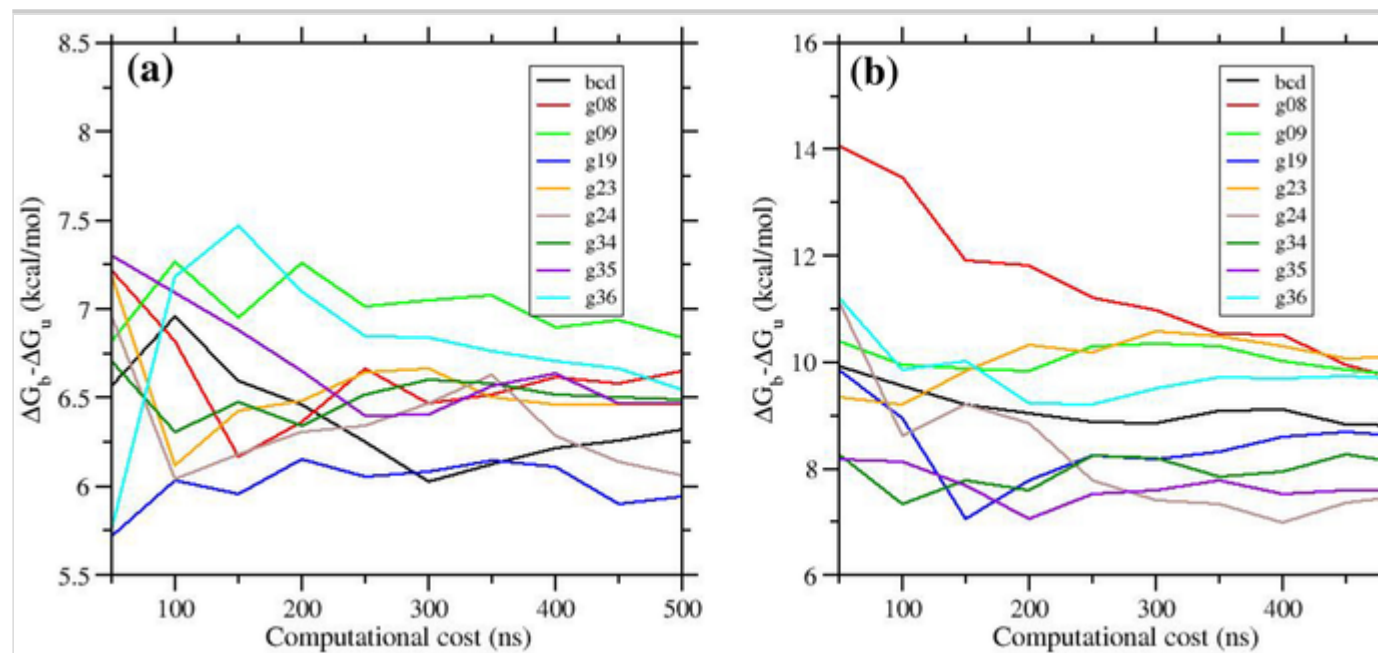
CD correlation plots; Data in the shaded region are within 2.0 kcal/mol of experimental counterpart



While the QM/MM ranked prediction set is, quite expectedly, [2, 3] totally off-the-mark (note the expanded scale in the left-bottom correlation plot of Fig. 3), the Autodock4 calculation is again in decent agreement with the experimental data, performing only slightly worse than the MD-based NE approaches do (see Table 1).

Fig. 4

Free energy estimates of CD host-guest systems as a function of the computational cost. In the y-axis, we report the difference $\Delta G_b - \Delta G_u$ computed using the Gaussian assumption. Panel a) and b) refers to the trans-4-methylcyclohexanol and R-rimantadine host-guest systems



In Fig. 4 we report the convergence of the estimates in the CD host-guest systems. As discussed in the Methods section, in evaluating the bound and unbound leg contributions for the CD systems, ΔG_b and ΔG_u , we used the Gaussian assumption for all work distributions. As seen for the TrimerTrip case (see Fig. 2a), the Gaussian estimate is quite accurate also for at relatively low computational cost (> 300 ns), with oscillation of less than 0.5 kcal/mol and in most cases within the confidence intervals (see Table S2 in the Supporting Information).

Discussion

Participants were asked by the organizers to provide uncertainties of the binding free energy estimates. Confidence intervals are quantities of no less importance than the free energy estimates themselves, since in an industrial drug design context they are related to the investment risk. In our case, the confidence interval refers to the uncertainty of the methodology alone, independently of the force field systematic (and unknown) uncertainty. FSDAM is a *unidirectional* technique using unidirectional estimators such as that based on the Gaussian assumption [31] or that based on the Jarzynski exponential average. In FSDAM, the latter estimator, which is notoriously biased [43], is used (in the statistically boosted variant) only when

the work distributions failed the AD and JB tests. In all cases, in our TrimerTrip and CD submissions, we used the unbiased Gaussian estimates for bound and unbound annihilation free energies. Unidirectional estimators are inherently less precise than bidirectional estimators when the forward and reverse work distribution exhibit a non negligible overlap [24, 44]. In unidirectional estimators, the confidence interval for the annihilation free energy estimate is in essence proportional to the *variance* of the work distribution, i.e. to the *dissipation* of the process, and inversely proportional to the square root of the number of collected work values [11, 14, 25, 45]. The bound and unbound errors can be rigorously summed in quadrature as they refer to two independent processes. Nonetheless, unidirectional confidence intervals can be large in case of high dissipation for either the bound or the unbound or both NE processes. For example, g1, g8, and g16 (TrimerTrip) systems and MG23-g2 (CD) show 95% confidence intervals from a minimum of ± 1.5 kcal/mol up to ± 2.5 kcal/mol. Large dissipation can be a consequence of the high conformational activity of the ligand in the binding cavity of the hosts or of a too fast annihilation protocol for the given solvation free energy level. In any case, FSDAM provides by design a *credible* methodological confidence interval, since the work distribution relies on an enhanced sampling of the starting states, done with a weak COM-COM restraint allowing for ligand orientation/conformation exploration in the binding site. In FEP calculations, the errors are estimated by summing in quadrature the individual errors based on block averages taken within the λ windows along the alchemical stratification, usually by way of post-processing application scripts (e.g. `gmx bar` in gromacs [46]). λ -windows MD simulations in FEP for absolute binding free energy calculations are in general done standardly (no enhanced sampling). Correlation between λ windows are rarely checked and “inner” block averages cannot provide reliable confidence intervals in case of serious sampling issues. As a result, FEP predictions are often heavily dependent [31] on the initial setup and on the stratification protocol [47] with non uniform convergence rates jeopardizing the equal-time assumption, [48] eventually leading to a systematic underestimation of the confidence interval [49]. An example of this pathology can be found also in the present challenge. For the AMOEBA/FEP prediction sets (ranked and not ranked), the confidence intervals are all consistently well below 0.5 kcal/mol. Given the disparity between the free energy estimates in some cases for the AMOEBA/FEP ranked and not ranked submissions, such uncertainty level does not apparently include the uncertainty due to the competing conformational states in the TrimerTrip host (see methodological discussion in Ref. [38])

Concerning DSSB in the CD systems, the prediction [50] was done using a *bidirectional approach*, allowing for the use of the accurate and precise maximum-likelihood Bennett Acceptance Ratio estimator (BAR) [51, 52]. Bidirectional estimators in the context of NE approaches are feasible only using strong translational/orientational/conformational *volume* restraints for the bound decoupled ligand, limiting the accessible conformational space of the system [53] and so yielding reverse and forward work distributions with discernible overlap. The enforcement of conformational restraints, whose contribution to the binding free energy must be somehow assessed by re-weighting procedures, [54] implies the *a priori* selection of a pose that in real-world drug-receptor systems can be sub-optimal leading to systematic (methodological) and unknown errors in spite of an apparent high precision as measured by the inverse of the Fischer information in BAR [52].

Potential pitfalls and weak points in the FSDAM

Gaussian assumption

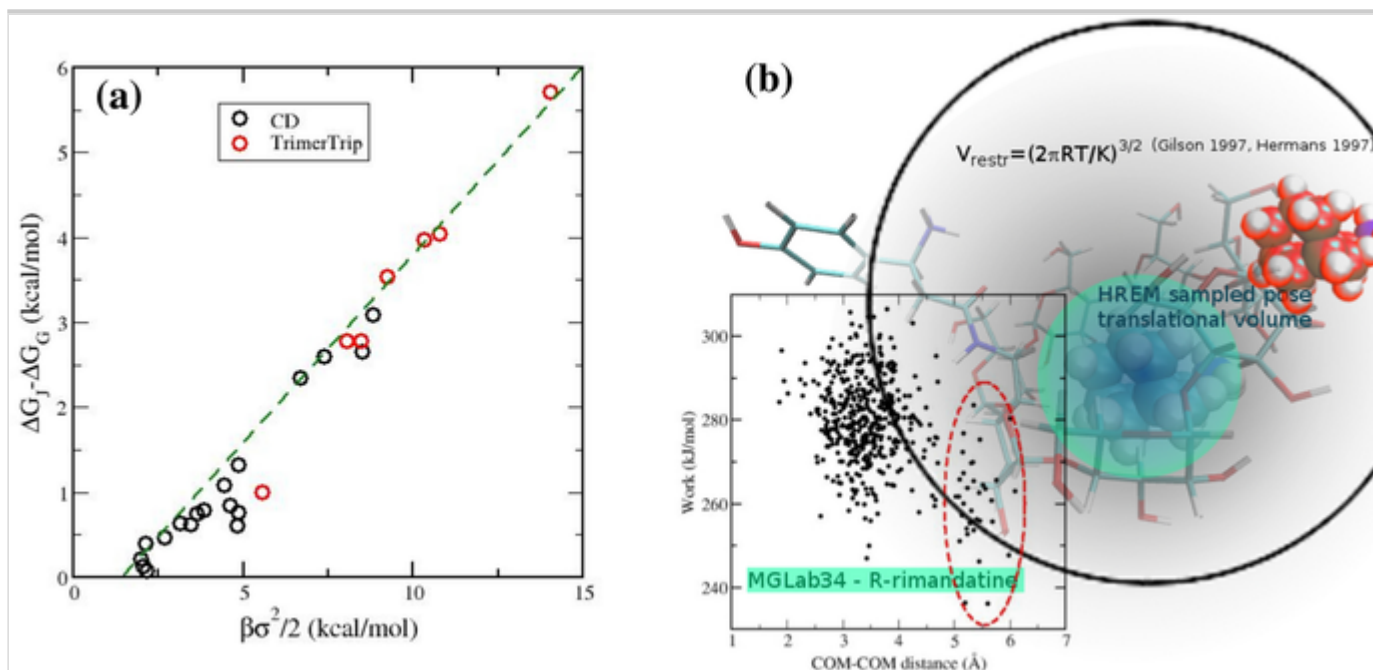
As explained in Sec. Methods, in our approach we have used, in both the HREM and **FSNS** stages of the bound state, a *weak* restraint potential with a large allowance volume of the order $V_r \simeq 600 \text{ \AA}^3$. This choice, while affording a basically unrestrained orientational/conformational sampling of the bound state with the fully coupled guest, prevents the use a bidirectional method. In fact, in the reverse process, the equilibrium states for the decoupled (gas-phase) ligand would be sampled from a random position/orientation within V_r , with the subsequent FS alchemical trajectories probing mostly bulk states or suboptimal poses, hence yielding a reverse work histogram with no overlap with the forward distribution and making the BAR estimate unreliable. Nonetheless, in the forward direction, when the work histogram is normal, the Gaussian assumption can still be used with a volume-related correction derived by applying the Crooks theorem to a Gaussian mixture (see Section 5.2 further below for details)

The bound and unbound ligand annihilation (forward process) free energies are estimated using either the Gaussian assumption or the Jarzynski boosted exponential average obtained by the combination of the decorrelated Lennard-Jones and discharging work values. The latter estimate is implemented when the work distribution fails the AD and JB normality tests. It should be however pointed out that all normality tests allows *to exclude*, with a given confidence, the null hypothesis (i.e., the distribution is Gaussian) [55]. However, when the tests are

passed, *no certainty* can be guaranteed for the null hypothesis and the distribution is only *likely* to be normal. Wide work histograms, constructed using small samples, can be easily mistaken as normal distributions while in effect they are given by, e.g., an undetected superposition of Gaussian mixtures, with catastrophic consequences on the accuracy of the annihilation free energy estimates. A Gaussian estimate is in general more reliable when the square root of variance is relatively low (below 2-3 kcal/mol). The accuracy of the Gaussian estimate can be assessed by comparing it with the standard Jarzynski estimate. The latter is known to be affected by a systematic *positive* bias, proportional to the dissipation and inversely proportional to the square root of the number of work values [43]. If the gap between the unbiased Gaussian estimate and the biased Jarzynski estimate is too wide (e.g. much larger than the *presumed* dissipation $\beta\sigma^2/2$ using the Gaussian assumption), then the work distribution is very likely the results of a mixture.

Fig. 5

(a) Difference of Gaussian and Jarzynski estimate for the bound state in the TrimerTrip and CD systems as a function of the dissipation. Differences have been evaluated for the bound work distribution passing the AD and JB tests. (b) Binding site volume V_{site} (cyan-shaded spherical region) and COM-COM restraint V_r volume (grey-shaded spherical region) in FSDAM. In the inset plot, we show the correlation between the COM-COM distance and the corresponding work obtained in the FS stage in the MGLab35-R-rimantadine system. The cluster at large COM-COM distance ($\geq 5\text{\AA}$), highlighted in the dashed red ellipse, refers to a secondary over-sampled pose (red-highlighted guest). These values, as part of the shadow components, are excluded in the computation of the annihilation free energy contribution of the primary pose using the Gaussian assumption



In Fig. 5 we show the difference, $B = \Delta G_J - \Delta G_G$ between the biased Jarzynski and unbiased Gaussian estimate as a function of the presumed dissipation $W_d = \beta\sigma^2/2$ obtained from the variance of the bound state distribution in the TrimerTrip (red circles) and CD (black circles) host-guest pairs whose binding free energies were estimated using the Gaussian assumption. Remarkably, B (i.e. the bias) appears to depend linearly on the dissipation, as suggested in Eq. 10 of Ref. [43], hence confirming the validity of the Gaussian assumption for the bound state work distributions in all reported systems (all pairs in CD and 7 pairs in the TrimerTrip).

Standard state correction

The standard state correction is a second issue of concern and in general a weak point in the FSDAM approach as in any other alchemical theory relying on the definition of the complex [32, 54, 56, 57]. The three MD-based **alchemical** predictions, AMOEBA/FEP for the TrimerTrip, the DSSB/NE for the CD and the GAFF2/FSDAM used three different approaches for implementing the standard state correction.

In AMOEBA/FEP, the bound state was subject in all λ -windows to a flat-bottom COM-COM potential, [58] “chosen such that [these restraints] were not violated during long unrestrained MD runs on the bound host-guest complex”, i.e. satisfying the condition $V_r \geq V_{\text{site}}$ [32]. Following Refs. [56], the standard state correction to the dissociation free energy was estimated as $\Delta G_{\text{corr}} = RT \ln(V_r/V_0)$ where V_r is

the allowance volume of the chosen flat-bottom COM-COM restraint potential and V_0 is the standard state volume.

In DSSB/NE, strong pose restraints [54] were imposed in the bound state, involving ligand-receptor coordinates of arbitrarily selected atoms on the ligand and the host. The standard state correction was in this case estimated as prescribed in Refs. [57, 59], from the difference between the cost of imposing the restraint at full coupling in the bound state and that of releasing it for the decoupled (bound) ligand.

A thorough discussion on potential pitfalls of the above approaches addressing the binding site volume issue and the related standard state correction is clearly outside the scope of the present contribution, and we refer to Ref. [32] for an in-depth analysis.

As previously stated, in FSDAM, the harmonic restraint is weak, with a rather large allowance volume [56, 60] $V = (2\pi RT/K)^{3/2}$ of the order of $\simeq 600 \text{ \AA}^3$, corresponding to a *ligand concentration* [61] around 3M. The combination of the weak COM-COM restraint and of the HREM with torsional tempering imposing only *intrasolute* scaling affords an (almost) unrestrained orientational-conformational sampling of the ligand in the binding site. On the other hand, the HREM torsional tempering algorithm *do not enhance passive diffusion* (the solvent stays cold and solute-solvent interactions are not scaled), [17] so that unbound states or secondary low energy poses, while in principle made accessible by the selected ligand concentration $1/V_r$, will not be observed during simulations lasting few or **few** tens of nanoseconds on the target state. HREM-TT is hence designed to sample a low free energy, long-lived *metastable* state corresponding to the bound system. In this regard, we recall that the average half-life of the bound state is given by $\tau_{1/2} = 1/k_{\text{off}} \equiv 1/(K_d K_{\text{on}})$, where k_{off} and k_{on} are the first-order dissociation rate of the host-guest complex and the second-order rate constant of the binding reaction, respectively. Given a typical value of $k_{\text{on}} = 10^8 \text{ M}^{-1} \text{ s}^{-1}$ due to passive diffusion, even for a weak millimolar ligand (as in several cases in the CD systems), we expect $\tau_{1/2}$ of the order *microseconds*. As illustrated in Fig. 5b for the case of MGLab35-R-rimantadine system, when departure of the ligand from the binding site occasionally occurs, it will result in general in the oversampling of *shadow* low-free energy metastable states [44]. These phase-space points, being part of the shadow metastable component due to secondary poses or bulk states, must be purged. In fact, even for a weak millimolar ligand, the unbound or weakly bound states at the restraint-imposed concentration of $C = 1/V_r \simeq 3\text{M}$, are

expected with a cumulative Boltzmann weight of $1/1000$ with respect to that of the principal component referring to the bound state.

Exploiting the fact that any work distribution can be expressed as a linear combination of independent normal distributions, and using the Crooks theorem for Gaussian mixtures, it can be shown [12, 44] that the unknown weight of the metastable shadow component when sampling the bound state can be worked out as $w_s = V_{\text{site}}/V_r$, where V_{site} is the effective *translational* binding site volume spanned by the ligand in the binding site. It is a matter of simple algebra [12] (see also Section “Allowance/restraint volume in FSDAM theory” in the Supporting Information of Ref. [14]), to derive the cumulative standard state free energy correction due to the restraint volume and to the shadow components as $\Delta G_{\text{vol}} = RT \ln V_{\text{site}}/V_0$. Note that this correction coincides approximately with that used in the AMOEBA/FEB submission for the TrimerTrip if the flat-bottom potential spans a volume of the order of V_{site} .

Finally, we must warn that the quantity V_{site} , which is used to define the re-entrance probability V_{site}/V_r at the ligand concentration $1/V_r$ in an hypothetical reverse process, [12] is introduced in a non equilibrium context and is hence in principle dependent on the way the restraint is imposed (hard wall or harmonic) and, most importantly, on the duration time τ of the NE experiments. In all our blind predictions, we have *estimated* such elusive binding site volume as $V_{\text{site}} = 4\pi[(2\sigma)^3/3]$, where σ is the half-width of the COM-COM distribution in the bound state, under the assumption that V_{site} is weakly dependent on τ so long that $\tau \ll t_{\text{on}}$, where $t_{\text{on}} = V_r/k_{\text{on}}$ is the average ~~mean~~ time for the ligand to bind the target at the concentration $1/V_r$.

Conclusion

In this contribution, we have presented and critically discussed our ranked prediction sets for the TrimerTrip and CD systems in the context of the latest SAMPL7 challenge, using FSDAM, a nonequilibrium alchemical approach combined with enhanced sampling end-state simulations where the fully coupled ligand is allowed to explore the conformational space in the bound and unbound states. FSDAM is tailored for the implementation on high performing computing facilities with nearly ideal parallel efficiency, due to the limited communication overhead required in the HREM and FS stages. The performances of our MD-based technique, that uses conventional fixed-charge force fields, is in line with our previous submissions in the SAMPL6 host-guest challenge [3] (done using the very

same technology) yielding binding free energies estimates within 2 kcal/mol in most of the cases. Outliers are rare and likely to be ascribed to structural deficiencies of the force field due to the neglect of important polarization effects in the anisotropic environment of some host-guest complexes. Thanks to the enhanced sampling of the fully coupled end-states, FSDAM provides by design a credible confidence interval that depends in essence on the dissipation level of the process.

Concerning specifically the SAMPL7 challenge ranked submissions, in the TrimerTrip, a much more computationally demanding approach based on a polarizable force field, did better than FSDAM, while in the CD system, a nonequilibrium bidirectional fixed-charges approach produced comparable MUEs. We finally must honestly point out the surprising and excellent results for both the TrimerTrip and CD obtained using a simple and inexpensive Docking approach. While MD simulations certainly provide valuable information on entropic and conformational effects in ligand-receptor association that Docking cannot simply deliver by design, modern docking score functions, such as those provided by Autodock4 software, appear to be remarkably predictive given the limited computational cost of the approach.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements

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. CRESCO/ENEAGRID High Performance Computing infrastructure is funded by ENEA, the Italian National Agency for New Technologies, Energy and Sustainable Economic Development and by Italian and European research programmes (see www.cresco.enea.it for information).

Electronic supplementary material

Below is the link to the electronic supplementary material.

Electronic supplementary material 1 (PDF 127 kb)

Electronic supplementary material 2 (PDF 7156 kb)

References

1. Muddana HS, Fenley AT, Mobley DL, Gilson MK (2014) The sampl4 host-guest blind prediction challenge: an overview. *J Comput Aided Mol Des* 28(4):305–317
2. Yin J, Henriksen NM, Slochower DR, Shirts MR, Chiu MW, Mobley DL, Gilson MK (2016) Overview of the sampl5 host-guest challenge: are we doing better? *J Comput Aided Mol Des*, pp 1–19
3. Rizzi A, Murkli S, McNeill JN, Yao W, Sullivan M, Gilson MK, Chiu MW, Isaacs L, Gibb BC, Mobley DL, Chodera JD (2018) Overview of the sampl6 host-guest binding affinity prediction challenge. *J Comput Aided Mol Des* 32(10):937–963
4. Ndendjio SZ, Liu W, Yvanez N, Meng Z, Zavalij PY, Isaacs L (2020) Synthesis and recognition properties Triptycene walled glycoluril trimer. *New J Chem* 44:338–345
5. Kellett K, Duggan BM, Gilson MK (2019) Facile synthesis of a diverse library of mono-3-substituted β -cyclodextrin analogues. *Supramol Chem* 31(4):251–259
6. Gibb Corinne LD, Gibb Bruce C (2014) Binding of cyclic carboxylates to octa-acid deep-cavity cavitand. *J Comput Aided Mol Des* 28(4):319–325
7. Amezcua Martin, Mobley David (2020) SAMPL7 challenge overview: assessing the reliability of polarizable and non-polarizable methods for host-guest binding free energy calculations. *ChemRxiv* 8 :12768353.v1
8. <https://samplchallenges.github.io/roadmap/submissions/> , Accessed 23 June 2020

9. Crooks GE (1998) Nonequilibrium measurements of free energy differences for microscopically reversible markovian systems. *J Stat Phys* 90:1481–1487
10. Jarzynski C (1997) Nonequilibrium equality for free energy differences. *Phys Rev Lett* 78:2690–2693
11. Procacci P, Guarrasi M, Guarnieri G (2018) Sampling host-guest binding predictions using a non equilibrium alchemical approach. *J Comput Aided Mol Des* 32(10):965–982
12. Procacci P (2016) I. dissociation free energies of drug-receptor systems via non-equilibrium alchemical simulations: a theoretical framework. *Phys Chem Chem Phys* 18:14991–15004
13. Nerattini F, Chelli R, Procacci P (2016) 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
14. Procacci P (2018) Myeloid cell leukemia 1 inhibition: An in silico study using non-equilibrium fast double annihilation technology. *J Chem Theory Comput* 14(7):3890–3902
15. Procacci P (2016) Hybrid MPI/OpenMP Implementation of the ORAC Molecular Dynamics Program for Generalized Ensemble and Fast Switching Alchemical Simulations. *J Chem Inf Model* 56(6):1117–1121
16. Liu P, Kim B, Friesner RA, Berne BJ (2005) Replica exchange with solute tempering: a method for sampling biological systems in explicit water. *Proc Acad Sci* 102:13749–13754
17. Marsili S, Signorini GF, Chelli R, Marchi M, Procacci P (2010) Orac: a molecular dynamics simulation program to explore free energy surfaces in biomolecular systems at the atomistic level. *J Comput Chem* 31:1106–1116
18. Procacci P (2017) Primadorac: a free web interface for the assignment of partial charges, chemical topology, and bonded parameters in organic or drug molecules. *J Chem Inf Model* 57(6):1240–1245

19. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR (2011) Open babel: an open chemical toolbox. *J Cheminf* 3(1):33
20. Izadi S, Onufriev AV (2016) Accuracy limit of rigid 3-point water models. *J Chem Phys* 145(7):074501
21. Hasel W, Hendrickson TF, Clark SW (1988) A rapid approximation to the solvent accessible surface areas of atoms. *Tetrahedron Comput Methodol* 1(2):103–116
22. Marchi M, Procacci P (1998) Coordinates scaling and multiple time step algorithms for simulation of solvated proteins in the npt ensemble. *J Chem Phys* 109:5194–5202
23. Procacci P (2019) Solvation free energies via alchemical simulations: let's get honest about sampling, once more. *Phys. Chem. Chem Phys* 25:13826–13834
24. Procacci P (2019) Accuracy, precision, and efficiency of nonequilibrium alchemical methods for computing free energies of solvation. i. bidirectional approaches. *J Chem Phys* 151(14):144113
25. Piero P (2019) Precision and computational efficiency of nonequilibrium alchemical methods for computing free energies of solvation. ii. unidirectional estimates. *J Chem Phys* 151(14):144115
26. Beutler TC, Mark AE, van Schaik RC, Gerber PR, van Gunsteren WF (1994) Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem Phys Lett* 222:5229–539
27. Anderson TW, Darling DA (1954) A test of goodness of fit. *J Am Stat Assoc* 49:765–769
28. Jarque CM, Bera AK (1980) Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Econ Lett* 6(3):255–259
29. Hummer G (2001) Fast-growth thermodynamic integration: error and efficiency analysis. *J Chem Phys* 114:7330–7337

30. Procacci P, Marsili S, Barducci A, Signorini GF, Chelli R (2006) Crooks equation for steered molecular dynamics using a nosé-hoover thermostat. *J Chem Phys* 125:164101
31. Pohorille A, Jarzynski C, Chipot C (2010) Good practices in free-energy calculations. *J Phys Chem B* 114(32):10235–10253
32. Procacci P, Chelli R (2017) 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
33. Zhang C, Chao L, Jing Z, Chuanjie W, Piquemal J-P, Ponder JW, Ren P (2018) Amoeba polarizable atomic multipole force field for nucleic acids. *J Chem Theory Comput* 14(4):2084–2108
34. Bannwarth C, Ehlert S, Grimme S (2019) Gfn2-xtb—an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *J Chem Theory Comput* 15(3):1652–1671
35. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) Autodock4 and autodocktools4: automated docking with selective receptor flexibility. *J Comput Chem* 30(16):2785–2791
36. Vassetzki D, Pagliai M, Procacci P (2019) Assessment of gaff2 and opls-aa general force fields in combination with the water models tip3p, spce, and opc3 for the solvation free energy of druglike organic molecules. *J Chem Theory Comput* 15(3):1983–1995
37. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) Am 1: a new general purpose quantum mechanical model. *J Am Chem Soc* 107:3902–3909
38. See comments on OVERLAP and INDENT host conformation in the AMOEBA submission file `Clip-ponder.txt` at https://github.com/samplchallenges/SAMPL7/tree/master/host_guest/Analysis/Submis Accessed 23 June 2020
39. Gapsys V, Michielssens S, Peters JH, de Groot BL, Leonov H (2015) Calculation of binding free energies. In: *Molecular modeling of protein*. Humana

Press, pp 73–209

40. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. *J Comp Chem* 25:1157–1174
41. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, Mackerell AD (2010) Charmm general force field: a force field for drug-like molecules compatible with the charmm all-atom additive biological force fields. *J Comput Chem* 31(4):671–690
42. Mobley DL, Bannan CC, Rizzi A, Bayly CI, Chodera JD, Lim VT, Lim NM, Beauchamp KA, Slochower DR, Shirts MR, Gilson MK, Eastman PK (2018) Escaping atom types in force fields using direct chemical perception. *J Chem Theory Comput* 14(11):6076–6092 PMID: 30351006
43. Gore J, Ritort F, Bustamante C (2003) Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proc Natl Acad Sci USA* 100(22):12564–12569
44. Procacci P (2015) Unbiased free energy estimates in fast nonequilibrium transformations using gaussian mixtures. *J Chem Phys* 142(15):154117
45. Procacci P (2020) A remark on the efficiency of the double-system/single-box nonequilibrium approach in the sampl6 sampling challenge. *J Comput Aided Mol Des* 34(6):635–639
46. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E (2015) Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2:19–25
47. Naden Levi N, Shirts Michael R (2015) Linear basis function approach to efficient alchemical free energy calculations. 2. inserting and deleting particles with coulombic interactions. *J Chem Theory Comput* 11:2536–2549
48. Sun ZX, Wang XH, Zhang JZH (2017) Bar-based optimum adaptive sampling regime for variance minimization in alchemical transformation. *Phys Chem Chem Phys* 19:15005–15020

49. Yildirim A, Wassenaar TA, van der Spoel D (2018) Statistical efficiency of methods for computing free energy of hydration. *J Chem Phys* 149(14):144111
50. Khalak, Y., Tresadern, G., de Groot, B.L., Gapsys, V. Non-equilibrium approach for binding free energies in cyclodextrins in SAMPL7: force fields and software. *J Comput Aided Mol Des* (2020).
<https://doi.org/10.1007/s10822-020-00359-1>
51. Bennett CH (1976) Efficient estimation of free energy differences from monte carlo data. *J Comput Phys* 22:245–268
52. Shirts MR, Bair E, Hooker G, Pande VS (2003) Equilibrium free energies from nonequilibrium measurements using maximum likelihood methods. *Phys Rev Lett* 91:140601
53. Heinzelmann G, Gilson MK (2020) Automated docking refinement and virtual compound screening with absolute binding free energy calculations. *bioRxiv*. <https://doi.org/10.1101/2020.04.15.043240>
54. Boresch S, Tettinger F, Leitgeb M, Karplus M (2003) Absolute binding free energies: a quantitative approach for their calculation. *J Phys Chem B* 107(35):9535–9551
55. Tanweer Ul Islam (2017) Stringency-based ranking of normality tests. *Commun Stat Simul Comput* 46(1):655–668
56. Gilson MK, Given JA, Bush BL, McCammon JA (1997) The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys J* 72:1047–1069
57. Deng Y, Roux B (2006) Calculation of standard binding free energies: aromatic molecules in the t4 lysozyme l99a mutant. *J Chem Theory Comput* 2(5):1255–1273
58. Shi Y, Laury ML, Wang Z, Ponder JW (2020) Amoeba binding free energies for the sampl7 trimertrip host-guest challenge. *J Comput Aided Mol Des* 1–15

59. Deng Y, Roux B (2009) Computations of standard binding free energies with molecular dynamics simulations. *J Phys Chem B* 113:2234–2246
60. Hermans J, Wang L (1997) Inclusion of loss of translational and rotational freedom in theoretical estimates of free energies of binding. Application to a complex of benzene and mutant t4 lysozyme. *J Am Chem Soc* 119(11):2707–2714
61. Zhou H-X, Gilson MK (2009) Theory of free energy and entropy in noncovalent binding. *Chem Rev* 109:4092–4107