SAMPL6 host-guest blind predictions using a non equilibrium alchemical approach

Piero Procacci · Massimiliano Guarrasi · Guido Guarnieri

Abstract In this paper, we compute, by means of a non equilibrium alchemical technique, called Fast Switching Double Annihilation Methods (FSDAM), the absolute standard dissociation free energies of the the octa acids hots-guest systems in the SAMPL6 challenge initiative. FSDAM is based on the production of canonical configurations of the bound and unbound states via enhanced sampling and on the subsequent generation of hundreds of fast non-equilibrium ligand annihilation trajectories. The annihilation free energies of the ligand when bound to the receptor and in bulk solvent are obtained from the collection of work values using an estimate based on the Crooks theorem for driven non equilibrium processes. The FSDAM blind prediction, relying on the normality assumption for the annihilation work distributions, ranked fairly well among the submitted blind predictions that were not adjusted with a linear corrections obtained from retrospective data on similar host guest systems. Improved results for FSDAM can be obtained by post-processing the work data assuming mixtures of normal components.

Piero Procacci
University of Florence, Department of Chemistry
Via Lastruccia n. 3, I-50019 Sesto Fiorentino (FI) (Italy)
Tel.: +39 055 4573081
E-mail: procacci@unifi.it

Massimiliano Guarrasi
CINECA-Interuniversity Consortium
via Magnanelli 6/3, 40033 Casalecchio di Reno, Bologna, Italy
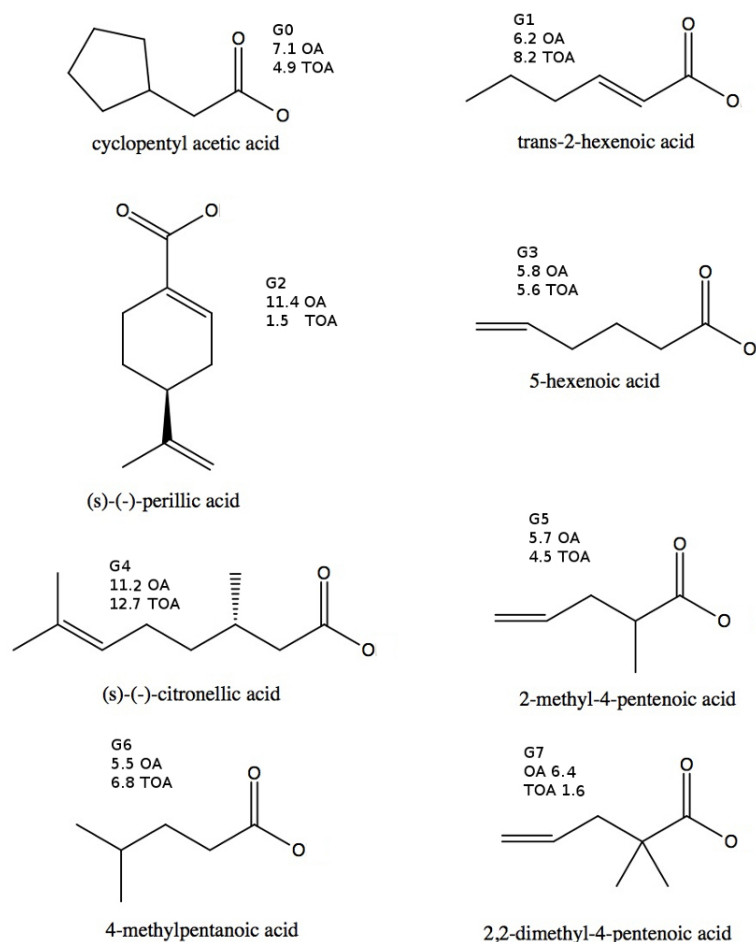Tel.: 39 051 6171560
E-mail: m.guarrasi@cineca.it

Guido Guarnieri
ENEA, Portici Research Centre, DTE-ICT-HPC
P.le E. Fermi, 1, I-80055 Portici (NA) (Italy)
Tel.: +39 081 7723576
E-mail: guido.guarnieri@enea.it

1 Introduction

Host-guest systems have been proposed in the SAMPL challenges[1–3] as effective substitutes for protein ligand systems in the evaluation of computational methods of predicting binding affinities. Typical host molecules in the SAMPL initiatives are cavitands like the octa acids (OA) and the tetra-endo-methyl octa-acids (TEMOA)[4–6]. These systems are supposed to mimic just the binding site pockets in proteins for hosting small flexible guest molecules, sparing the demanding task of simulating the whole protein and/or identifying the relevant binding sites on the protein surface. While SAMPL6 guests (see Figure 1) may exhibit a complex conformational pattern (e.g. G2 with axial/equatorial conformations of the 1-4 substituents on the six-membered ring, or G4 with five rotable sp_3 bonds and, in principle, 243 different conformations), OA and TEMOA are relatively rigid systems, hence further alleviating the problem of canonically sampling the conformations of the binding sites in real proteins.

Binding free energies predictions for SAMPL6 host-guest systems were tackled using disparate computational strategies,[1] from quantum mechanical to semiclassical or classical techniques, with explicit or implicit solvation models. Classical methodologies of fully atomistic representations included the host-guest potential of mean force determination via Umbrella Sampling[7] and double annihilation alchemical simulations, based either on free energy perturbation[8] (FEP) or thermodynamic integration[9] (TI). Many of these studies adjusted their raw free energy predictions with a linear corrections obtained from previous SAMPLx data on similar host guest molecules. Strictly speaking, the latter knowledge-based results, while definitely interesting, are not genuine blind predictions as they somehow rely on the availability of *retrospective* host-guest binding free energy measurements. In this contribution we present results from a single submission (shown in Figure 1) of a blind prediction with no retrospective corrections on the host-guest dissociation free energy. We used the so-called Fast Switching Double Annihilation technique (FSDAM).[10–13] FSDAM is the only non-equilibrium technique used in the SAMPL6 challenge. This method is based on the production of canonical configurations of the bound and unbound states via enhanced sampling (Hamiltonian Replica Exchange with Solute Tempering, HREM stage) and on the subsequent generation of hundreds of fast non-equilibrium (NE) ligand annihilation trajectories (Fast non equilibrium annihilation, FNEA stage) producing a bound and unbound work distributions. The annihilation free energies of the ligand when bound to the receptor and in bulk solvent, can be obtained from the collection of NE work values using an estimate based on the Crooks theorem for driven NE processes [14], in the assumption that the observed annihilation work distributions can be described by a normal distribution or by a mixture of Gaussian components.[15, 12] The absolute binding free energies are recovered by the differences of the annihilation free energies of the ligand in bulk and in the bound state, plus a standard volume and finite-size corrections.

FSDAM has been here applied using a completely automatic procedure for the generation of the topology and parameter files of the host and guest molecules with the PrimaDORAC assignment tool[16], as well as for the input files of the HREM and FNEA stages and for the corresponding batch submission scripts on high performance computing platforms and for the post-processing of the HREM and FNEA data. FSDAM is specifically tailored[17] for non uniform memory ac-

Host-guest dissociation free energies (kcal mol^{-1})

G0
7.1 OA
4.9 TOA

cyclopentyl acetic acid

G1
6.2 OA
8.2 TOA

trans-2-hexenoic acid

G2
11.4 OA
1.5 TOA

(s)-(-)-perillic acid

G3
5.8 OA
5.6 TOA

5-hexenoic acid

G4
11.2 OA
12.7 TOA

(s)-(-)-citronellic acid

G5
5.7 OA
4.5 TOA

2-methyl-4-pentenoic acid

G6
5.5 OA
6.8 TOA

4-methylpentanoic acid

G7
OA 6.4
TOA 1.6

2,2-dimethyl-4-pentenoic acid

Fig. 1 Blind prediction for the SAMPL6 challenge using FSDAM

cess systems, implemented via a two layer OpenMP/MPI parallelism, in such a way that the distributed memory layer manages the production of the simultaneous HREM or FNEA trajectories, corresponding to weakly communicating MPI instances, each parallelized with a strong scaling scheme implemented on the OpenMP layer within the intra-node multi-cores shared memory environment. A single FNEA job can engage thousands of cores with nearly ideal parallel efficiency, producing a total simulation time of hundreds of nanonseconds in few wall clock minutes. On a per host-guest pair basis, the whole calculation required, for the HREM stage, $\simeq$ 35 ns sampling (on the target state) using 8 replicas for both the bound and unbound systems and, for the FNEA stage, a cumulative $\simeq$ 250 ns

for the bound and unbound state for a total simulation time on the sixteen host-guest system of $\simeq 8.5$ μs of simulation performed in few wall-clock days on on the ENEA-CRESCO3[18] cluster and the Marconi Broadwell High Performing Computing (HPC) systems[19]. The present blind prediction for SAMPL6 constitutes hence in first instance a prototypical test for industrial applications of FSDAM in ligand-receptor systems and in second instance one of the first significant tryouts for the recently released GAFF2 force field[20] and for the GAFF2-based PrimaDORAC assignment tools.[16] The paper is organized as follows. In the Section 2 we describe in detail the methodology used in the FSDAM SAMPL6 prediction. In Subsection 2.1, we briefly revise the theoretical basis of the FSDAM. The basic simulations details are provided in the subsection 2.2. We then proceed to illustrate the HREM stage (subsection 2.3) with a relevant example concerning the two enantiomers of the chiral ligand G2 (perillic acid) in their unbound state. The standard volume and finite-size corrections (subsections 2.4, 2.5) are discussed when evaluating the annihilation free energies in the FNEA stage of the G2 ligand in the OA-G2 complex. FSDAM results for binding free energies based on the normality assumption of the work distributions are presented in the Section 3. In the subsection 3.1, we first report in detail the data concerning the single prediction that was submitted to the SAMPL6 challenge, critically analyzing the results with respect to the various source of errors. In subsection 3.2, we show how in some cases the prediction can be improved by using the expectation-maximization algorithm for the determination of the the hidden normal components in a mixture. A possible force field issue is finally outlined in subsection 3.3. Conclusive remarks are presented in the last section.

2 Theoretical and methodological aspects of FSDAM

The theoretical background of FSDAM has been thoroughly described elsewhere[10–13,17,21,22]. Here we provide a brief theoretical summary, focusing mostly on the technical details used in implementing this methodology on a HPC systems.

2.1 FSDAM theory

Likewise its equilibrium counterpart[23] (FEP- or TI-based), this alchemical methodology accomplishes the determination of the binding free energy by computing the difference between the decoupling free energies of the ligand in the solvated complex and in bulk solvent. As outlined in the introduction, each of these two independent calculations is in turn done in two steps: i) a replica exchange simulation with solute torsional tempering[24] (HREM stage) of the fully coupled ligand state, aimed at harvesting canonical (equilibrium) configurations of the systems; ii) a transformative stage based on NE simulations (FNEA stage), whereby the ligand, starting from the configurations sampled in the corresponding HREM stages, is rapidly annihilated in a swarm of concurrent and independent NE trajectories each yielding a NE work and eventually an annihilation work distribution.

In both HREM and FNEA bound state simulations, a weak harmonic restraint between the centers of mass (COM) of the host and guest molecules is

imposed. A COM-COM restraint in the bound state simulations is equivalent[25–27] to fix the ligand concentration or, equivalently, a ligand allowance volume $V_r = (2\pi RT/K_h)^{3/2}$, with K_h being the harmonic force constant of the restraint potential.[28] If no restraints are imposed, the allowance volume is clearly that of the MD box.[29,30] As discussed in Refs. [25,31], the *standard* dissociation free energy of the restrained system is given by the equation

$$\Delta G_0 = (\Delta G_b - \Delta G_u) + RT \ln(V_r/V_0) \tag{1}$$

where ΔG_b and ΔG_u are the (COM-COM restrained) bound ligand and free ligand annihilation free energy. Eq. 1 simply says that while ΔG_0 is constant, the ratio between dissociated and undissociated host-guest species $[H]/[HG] \equiv e^{-\beta(\Delta G_b - \Delta G_u)}$, depends on the imposed guest concentration $[G]_r$, i.e. on the restrained ligand reference volume V_r. In a HREM simulation of a bound state, with the guest molecule G restrained in the reference volume $V_r = 1/[G]_r$ with respect to the COM of the host H, the ratio of dissociated and bound states is given by $[H]/[HG] = K_d/[G]_r$, where $K_d = e^{-\beta \Delta G_0}$ is the dissociation constant. If one chooses for the guest reference volume, the value of 1661 Å^3, corresponding to a ligand concentration $[G]_r = 1$ M, then the ratio $[G]/[HG] \simeq K_d$. Therefore, choosing the force constant K_h so that $V_r = 1661\mathring{A}^3$, we expect that the number of dissociated states in an equilibrium HREM run, sampling N configurations, are of the order of $K_d N$. If N is of the order of few hundreds and $K_d < 10^{-3}$M , then we basically expect that *all* sampled HREM states of the restrained bound state should be of the associated type. In Figure 3, we show the COM-COM distance probability distributions of the OA and TEMOA complexes obtained in a 24 ns HREM simulation with a restraint potential so that $V_r = 1661$ Å^3. As it will be discussed further on, the guests remains in the host cavity in all sampled target state configurations, in agreement with the fact that $K_d < 10^{-3}$ M for all host-guest pair if the SAMPL6 challenge.

Provided that the annihilation NE work distributions $P_b(W)$ and $P_u(W)$ from the FNEA stages are normal, according to the Crooks theorem[14] the corresponding annihilation free energies can be straightforwardly recovered[32,33] using an unbiased estimator based on the mean and variance of the work distributions, i.e.

$$\Delta G_{b/u} = \langle W_{b/u} \rangle - \frac{1}{2}\beta \sigma_{b/u}^2 \tag{2}$$

where the indices b, u refer to the bound or unbound states and where $\langle W_b \rangle$ and $\langle W_u \rangle$ are mean value of the ligand decoupling work in the bound state and in bulk solvent and σ_b^2, σ_u^2 are the corresponding variances. Most importantly, the confidence interval for the estimates based on Eq. 2 can be easily assessed taking into account that the mean and variance for normally distributed samples are *independent* random variables and follow the t-statistics and chi-square distribution, respectively.[34] By inverting these distributions, we can get an estimate of the confidence intervals, thus providing an overall error for the Gaussian estimator of Eq. 2 as

$$\delta \Delta G_{b/u} = z_{\alpha/2} \left| \frac{\sigma_{b/u}}{n_{b/u}^{1/2}} + \frac{1}{2}\beta \left(\frac{2}{n_{b/u}}\right)^{1/2} \sigma_{b/u}^2 \right| \tag{3}$$

where $n_{b/u}$ is the number of sampled work values and with $1 - \alpha$ being the confidence level, meaning that the true value of $\Delta G_{b/u}$ falls within the given range of

Eq. 3 with probability $1 - \alpha$. Choosing $\alpha = 0.05$, for $n > 100$ we can safely take $z_{0.025} \simeq 2$ thus defining a confidence level for the interval of Eq. 3 of 95%.

Actually, as discussed above, the bound state annihilation work distribution should be contaminated on the left tail by a so-called[15] shadow normal component $n_d(W)$ corresponding to the dissociated states obtained with the imposed reference volume V_r, so that $P_b(W) = (1 - c)n_b(W) + cn_d(W)$ with c of the order of $c \simeq K_d V_r / V_0$. In Ref. [12, 15], we showed that in the hypothetical fast-growth reverse NE process of re-coupling an initially decoupled ligand from a random position in the allowance volume V_r, the principal bound state component is basically suppressed while the shadow component $cn_d(W)$, corresponding to an unbound or weakly bound guest molecule, gets exponentially amplified. The Crooks theorem for such two-component mixture yields

$$\Delta G_b^r = \langle W_b \rangle - \frac{1}{2}\beta\sigma_b^2 + RT\ln(V_{\text{site}}/V_r) \tag{4}$$

where V_{site}/V_r represents the weight of principal component in the hypothetical reverse fast growth process, corresponding to the probability of re-entrance in the binding pocket, defined in terms of a *binding site volume* V_{site}. ΔG_b^r in Eq. 4 is the V_r-dependent annihilation free energy of the restrained bound state. In deriving Eq. 4, it has been implicitly assumed that $V_r \gg V_{\text{site}}$,[27]. We further assume that the volume V_{site} is weakly dependent on the duration τ of the NE process, so long that $\tau \ll \tau_{\text{on}}$, where $\tau_{\text{on}} = V_{\text{ref}}/k_{on}$ is the average mean time for the ligand to bind the target at the concentration $1/V_{\text{ref}}$.[35, 36] On this basis, combining Eqs. 4, 2 and 1, we obtain the FSDAM expression for the standard dissociation free energy

$$\Delta G_0 = \Delta G_b - \Delta G_u + RT\ln\left(\frac{V_{\text{site}}}{V_0}\right) \tag{5}$$

where V_0 is the standard state volume. In Eq. 2, the σ-related energies $\frac{1}{2}\beta\sigma_{b/u}^2$ have a straightforward physical interpretation: they represent the dissipation in the NE process of the annihilation of the ligand in the bound and unbound state. Hence, the wider are the normal distributions, the more dissipative the NE annihilation process is.

The evaluation of the elusive V_{site} volume will be discussed further on in subsection 2.5. We conclude this subsection, with a brief outline of how the finite-size effects, involved in the annihilation of charged ligands, are handled in FSDAM. In the Particle Mesh Ewald treatment[37] used in this study for treating long range electrostatics, when computing the electrostatic energy of a charged system, one implicitly adds to the Coulomb energy a term corresponding to the contribution of the total charge in the MD box interacting with a uniform neutralizing background plasma. Such term, called the Wigner self energy, is automatically included in the reciprocal lattice when using PME,[38] while in the direct lattice is given by $E_w^d(\alpha, V) = -|\sum_i q_i|^2 \pi/(2\alpha^2 V)$, where V is the volume of the zero-cell, the sum is extended to all charges in the box and α is the Ewald convergence parameter. If the ligand bears a net charge, a finite size correction must hence be added to account for the change in the free energy related to the annihilation of the net charge. This direct space correction to the dissociation free energy can be

computed as

$$\Delta G_{\mathrm{fs}} = -\frac{\pi}{2\alpha^2}\left\{\frac{[Q_H^2 - (Q_H + Q_G)^2]}{V_{\mathrm{BOX}}^{(b)}} - \frac{Q_G^2}{V_{\mathrm{BOX}}^{(u)}}\right\} \tag{6}$$

where Q_H and Q_G are the net charge on the host and guest molecule, respectively and $V_{\mathrm{BOX}}^{(b/u)}$ are the MD box volume of the bound and unbound states.

2.2 Simulation details and sample preparations

Atomic type assignment and partial atomic charges on the ligands were computed using the PrimaDORAC web interface.[16] PrimaDORAC computes the AM1-BCC charges[39] on the AM1 optimized geometry[40] of the ligand and assigns the atomic types according to the recently released GAFF2 general parameterization for organic molecules.[20]. Following the indications of the organizers, all the carboxylate groups on the host molecules OA and TEMOA and on the guest molecules were assumed to be deprotonated, so that the guest and the host molecules bear a net charge of -1e and -8e respectively. The solvent was treated explicitly using the TIP3P model.[41] Long range electrostatic were treated using the Smooth Particle Mesh Ewald (SPME) method,[37] with an α parameter of 0.37 Å^{-1}, a grid spacing in the direct lattice of about 1 Å and a fourth order B-spline interpolation for the gridded charge array. As no counterions were included, charge neutralization in charged bound and unbound systems is implicitly done in SPME using a uniform neutralizing background plasma. Bonds constraints were imposed to X-H bonds only, where X is an heavy atom. All other bonds were assumed to be flexible. The pressure was set to 1 atm using a Parrinello-Rhaman Lagrangian[42] with isotropic stress tensor[43] while temperature was held constant to 298 K using three Nosé Hoover-thermostats coupled to the translational degrees of freedom of the systems and to the rotational/internal motions of the solute and of the solvent. The equations of motion were integrated using a multiple time-step r-RESPA scheme[44] with a potential subdivision specifically tuned for bio-molecular systems in the NPT ensemble.[45,43]. The long range cut-off for Lennard-Jones interactions was set to 13 Å in all cases.

Using the starting host and guest structures provided by the organizers, the preparation of the bound states for each host-guest pair was done by generating one hundred random host-guest structures within a COM-COM docking radius of 5 Å, followed by energy minimization in implicit solvent using the AGBNP model.[46] The starting structure of the complex corresponds to that of lowest energy found in the docking stage. The so obtained least energy host-guest structures were then oriented along the inertia frame of the guest molecule and explicit water molecules at a density of 1 g/cm^3 were added in a cubic MD box whose side-length was computed so that the minimum distance between host or guest ligand atoms belonging to neighboring replicas was larger than 24 Å in any direction. After removal of overlapping water molecules, the bound state systems contained about 1250 solvent molecules for a volume of approximately 40000 Å^3 corresponding to a side-length of $\simeq$ 34 Å. For water and box volume equilibration, a preliminary 50 ps constant pressure, constant temperature simulation was run for each of the so prepared solvated complexes. The starting structures of the guest molecules in

bulk solvent were prepared by inserting the structures provided by the organizers in a box of side-length 26 Å, containing TIP3P molecules at the density of $1\mathrm{g/cm}^3$.

The above procedure for generating the bound and unbound starting structures and the corresponding input file for subsequent HREM and FNEA stages was completely automatized by an application script program. All simulations in this study are done using the program ORAC.[17]

2.3 HREM stage

The HREM simulations of the bound state was run on the HPC system by launching, in a single parallel job, four independent Hamiltonian replica exchange[47] simulation with eight solute tempered replicas in the Generalized Ensemble (GE) for a total of 32 MPI instances. Each HREM battery sampled 96 configurations taken at regular interval in a simulation time of 7.8 ns, hence accumulating 384 solvated bound state configurations in a total simulation time of 31.2 ns. For the free guest HREM simulation, we used again 4 independent batteries of eight exchanging replicas, each lasting 1 ns, for a total simulation time of 4.0 ns, sampling 240 solvated guest molecules configurations. Each GE walker in the bound and unbound states used 6 threads on the OpenMP strong scaling layer, so that the hybrid OpenMP/MPI parallel HREM jobs engaged a total 192 cores. Only the torsional potential of host and guest system was scaled throughout the GE up to
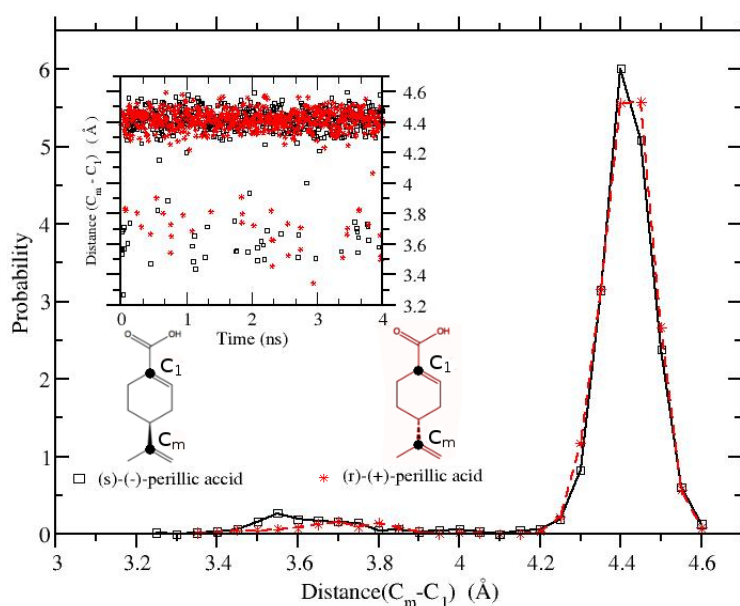

Fig. 2 Probability distribution of the distance between C1 and the sp_2 carbon C_m identifying the axial-equatorial conformation in the two enantiomers of the G2 guest molecule as obtained from the HREM simulations (see text for details). In the inset, the time record of the distance C1-C_m in the GE target state

0.1 in both cases, corresponding to a solute "torsional" temperature of 3000 K. The scaling protocol in the eight replicas GE range [1-0.1] was set according to the scheme described in Ref. [24]. Exchanges were attempted between neighboring replicas at each long time step (i.e. each 15 fs), yielding a mean acceptance ratio throughout the GE of no less that 50%, in all cases. All HREM computations were completed on the CRESCO3-ENEA HPC platform[48] in few wall clock days.

The adopted scaling protocol allows to effectively sample in the target state the whole accessible conformational space of the flexible guest molecules in their free state and when embedded in the cavitand. As a relevant example of HREM, we show the time record and the corresponding distribution of the axial-equatorial conformations in the R- and S- enantiomers of the G2 guest molecule in bulk as obtained in two 4 ns REM simulations referring to the two enantiomers solvated in TIP3P water. As it can be seen, axial-equatorial conformations (identified by the distance between the sp_2 carbon of the isopropenyl moiety and the ring carbon in position 1) easily interconverts in the G2 molecule, in spite of the high barrier ($\simeq 4$ kcal mol^{-1}) separating these conformational states. According to the GAFF2 force field, the conformational distributions, show a clear prevalence of the equatorial conformer and are essentially identical in the two enantiomers, and so should be their affinities towards the symmetrical OA and TEMOA host molecules, as assumed by the organizers.
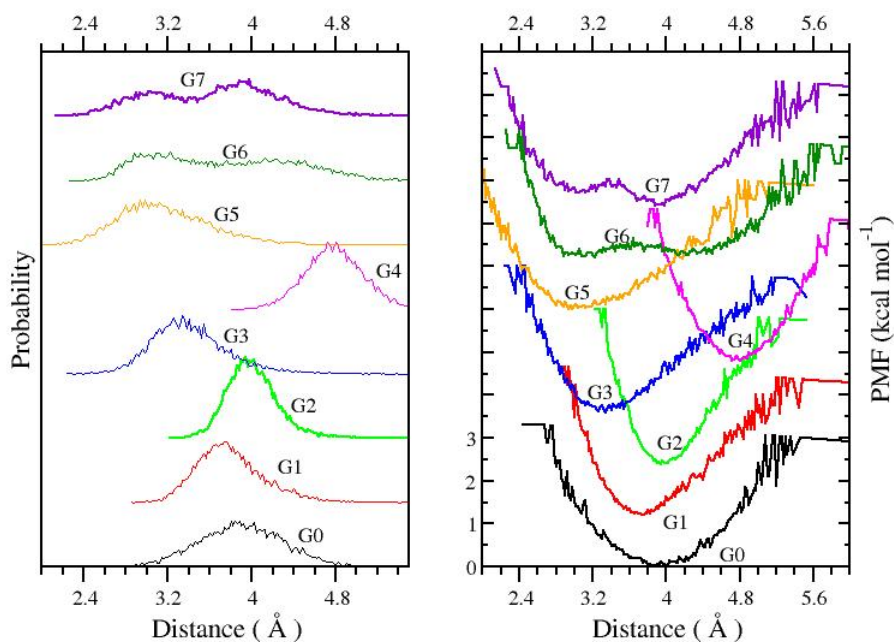

Fig. 3 COM-COM distance distribution function and and PMF for the OA-guest systems

The effectiveness of HREM approach in sampling conformational states in the bound state can be appreciated in the Figures 3 and 4, where we report the distribution of the host-guest COM-COM distance (left panels) and the corresponding potential of mean force (right panels) for the guest-OA and guest-TEMOA systems, respectively. In general, although the imposed weak restraint potential implies an allowance radius of more than 7 Å, we can see that the ligand lingers around the binding pocket of the cavitand. In most cases, the COM-COM distance distribution exhibits a positively skewed single peak with half-width rarely exceeding 1 Å. In some cases (G6 and G7 in OA and in TEMOA), the COM-COM distributions have a complex structure due to multiple poses. These can be highlighted by evaluating the potential of mean force along the COM-COM distances, shown in the right panels of Figures 3 and 4. Concerning the PMF computation via HREM, some comments are in order. In principle, setting the restraint so that
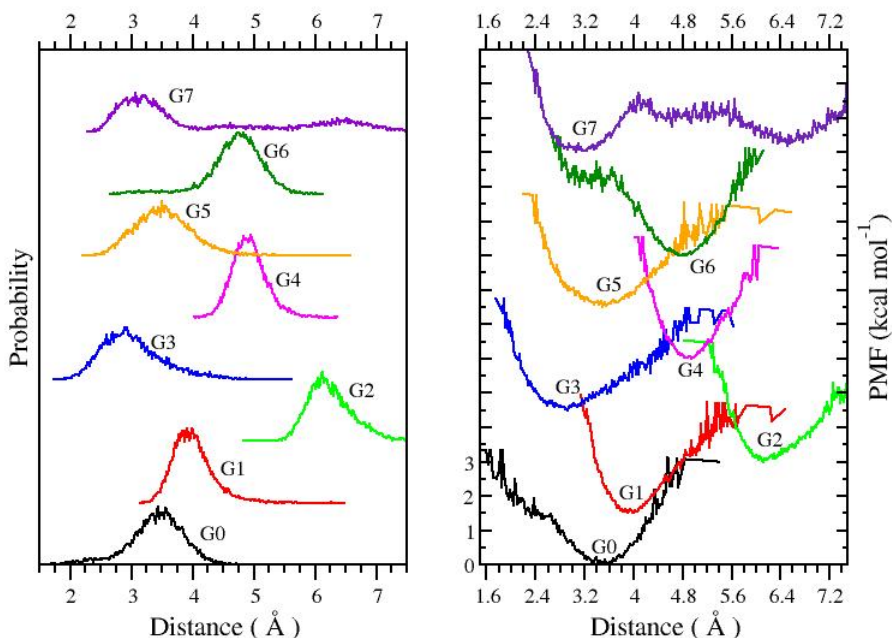

Fig. 4 COM-COM distance distribution function and and PMF for the TEMOA-guest systems

$[G]_r = 1$M, a HREM simulation could be used to compute directly the dissociation[49] constant as $K_d = [H]/[HG] = P_u/P_b$ where P_u and P_b are the probability of observing the dissociated and associated state.[50] In practice, while the bound state sampling in the HREM stage allows quite an accurate reconstruction of the PMF in the bottom of the well, zones at larger distances/higher energies towards unbound states are statistically noisy. This is so since solute torsional tempering

with no water-solute rescaling[49] (as done in our HREM approach) does not accelerate relative host-guest diffusion, in a such a way that the expected shortest τ_{off} for the weakest binder among all host-guest pairs (TEMOA-G5, $K_d \simeq 10^{-3}\text{M}$) should be of the order of the microseconds[35, 36] in any GE state, making *de facto* unattainable the brute force sampling of unbound states in the HREM stage and hence the direct calculation of K_d.

2.4 FNEA stage

For each host-guest pair, in the FNEA stage a swarm of fast independent ligand annihilation trajectories were started from the phase-space points sampled in the HREM stage for the unbound and bound guest molecules. The total number of independent NE trajectories were 384 for the decoupling of the ligand in the bound state and 240 for the decoupling of the ligand in bulk solvent. The annihilation protocol of the τ lasting NE processes was common to all host-guest systems and stipulates that the electrostatic interactions between the ligand and the environment are linearly brought to zero at $t = \tau/2$, while the Lennard-Jones interactions are switched off in the range $\tau/2 < t < \tau$ using a soft-core Beutler potential[51] regularization as λ is approaching to 1 corresponding to the decoupled state. The alchemical work along the alchemical path was computed as described in Ref. [10]. We stress here that in FSDAM there is no need for the optimization of the so-called thermodynamic length,[52] that is of choosing the alchemical protocol so that the total uncertainty for the transformation is the one which has an equal contribution to the uncertainty across every point along the alchemical path.[53] As previously discussed (see Eq. 3), a confidence intervals for the FSDAM free energy value can be estimated directly from the moments of the final work distributions.

As done in HREM, also in the FNEA stage parallel execution is performed using a straightforward hybrid OpenMP-MPI approach, with NE non communicating annihilation trajectories handled at the MPI level and a force decomposition scheme implemented with 9 threads on the shared memory OpenMP layer. As in past FSDAM studies on ligand-receptor or host-guest, the duration of each of NE independent decoupling trajectories adopted in this paper are of few hundreds of picoseconds (from 90 to 720 ps) for both the solvated bound state and the free ligand in bulk solvent. Each host-guest FNEA computation required 3240 and 2160 cores (bound and unbound state, respectively) on the Marconi-Broadwell HPC system[19] for few to few tens of wall clock minutes depending on the selected annihilation time. From the collection of annihilation works, the bound state and unbound state normalized work histograms probability were computed, namely $P_b(W)$ and $P_u(W)$, respectively. In the Figure 5 we show these distributions for the duration time $\tau_{b/u} = 360$ ps for all sixteen host-guest pairs. The normality of these distribution is assessed using the Anderson-Darling (AD) quadratic test.[54–56] The Anderson-Darling test A^2 is defined as $A^2 = \sum_{i=1}^n \frac{2i-1}{n}[\ln(\Phi(w_i) + \ln(1 - \Phi(w_{n+1-i})]$, where Φ is the Gaussian cumulative distribution function with sample mean and variance and w_i are the work values sorted in ascending order. The critical value of A^2 at the level $\alpha = 0.05$ is 0.752.[55] AD has been recently shown[57] to be the most stringent normality test among many popular alternatives including the Kolmogorov-Smirnov and the Wilk-Shapiro tests. As shown in Figure 5 and detailed in Table 1 further on, the
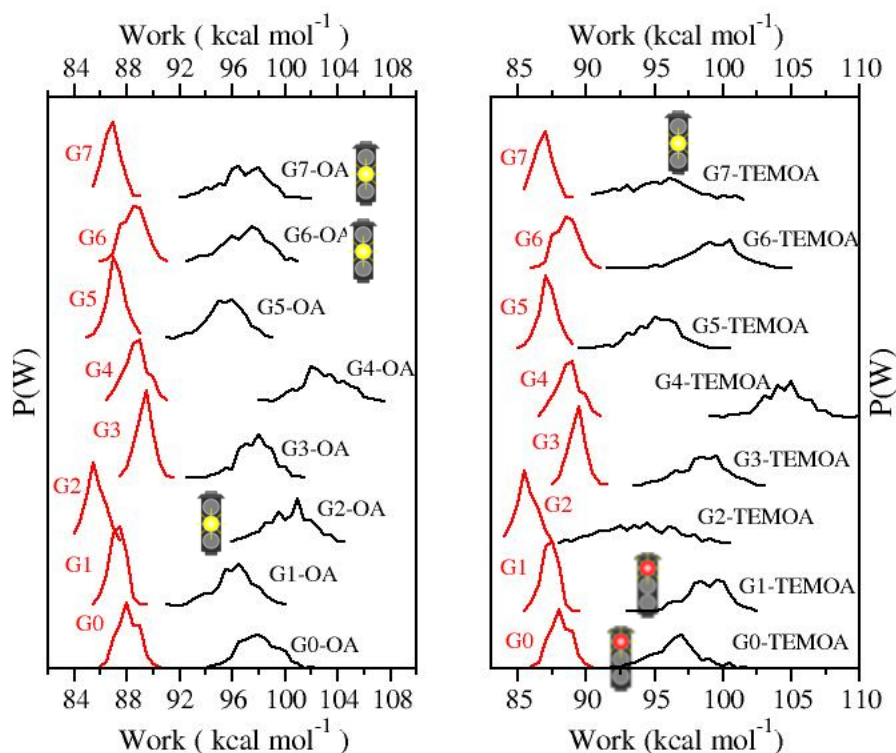
Fig. 5 Annihilation work distributions for the bound state (black trait, on the right) and unbound state (red trait, on the left) for the SAMPL6 host-guest pairs (OA complexes on the left panel and TEMOA complexes on the right panel) as obtained in the FNEA stage with an annihilation duration time of 360 ps for the unbound bound state. For the bound state $K_h = 0.003$ was set to kcal mol^{-1} Å^{-2} in both the HREM and FNEA stages. Red/top and orange/mid traffic-light symbols signal a failed AD test, with A^2 exceeding the critical value at the 0.05 α level by more than 1 and less than 0.5, respectively

work distributions for the unbound state of all eight ligands amply passed the AD test for normality. The AD test was passed with a confidence level of 95% for 10 out of 16 bound state distributions and exceeded the critical value by more than 0.5 in only two cases for the G0 and G1 TEMOA ligands. As shown in Table 1, we observe in general higher A^2 values for the TEMOA bound state, indicating that the extra methyl moieties decorating the crown of the host do have a significant impact on the host-guest free energy surface and, correspondingly, in the subsequent NE guest annihilation work. In the HREM stage, for example, the TEMOA-G7 complex is clearly characterized by a bimodal COM-COM distance distribution as shown in Figure 3. This bimodal distribution of the starting canonically sampled initial states is somehow reverberated in (see Figure 5) in the negatively skewed NE work distribution of the G7-TEMOA complex, yielding a A^2 value close to 1, pointing to a non normally distributed sample.
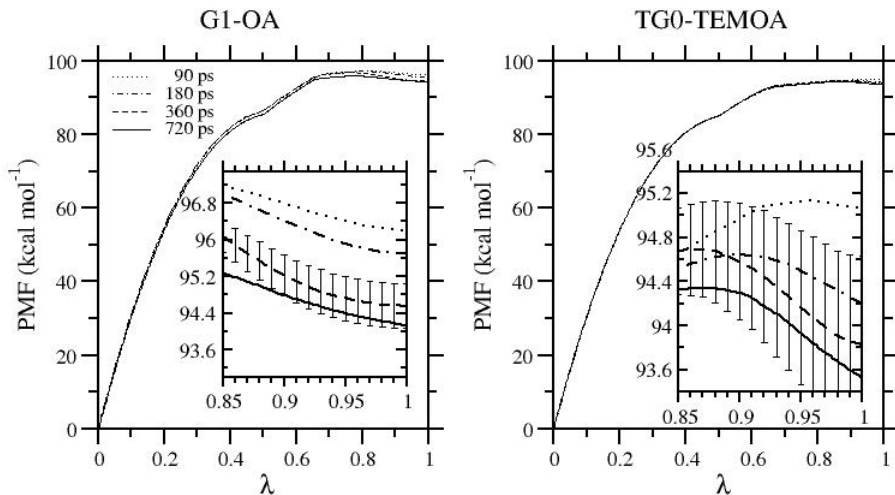
Fig. 6 PMF for the annihilation of the guest molecule in the G1-OA complex (left) and in the G0-TEMOA complex (right), obtained using different annihilation rates and a force constant of $K_h = 0.03$ kcal mol^{-1} Å^{-2}. The insets are an enlarged view of the PMF in the final stages $\lambda \simeq 1$ of the alchemical parameter λ. Error bars have been computed using Eq. 3 and are shown only for the guest annihilation with $\tau = 360$ ps for clarity

.

We stress here that errors in computing the annihilation free energies via the NE formula Eq. 2 are not due, by any means, to insufficient sampling of the intermediate alchemical states. These are crossed at fast speed in the concurrent NE trajectories and their transient distribution is of no concern whatsoever for the end-user. In the context of the Crooks theorem, what ought to be stationary is the annihilation work distribution. For normal distributions, this comes down to the stationarity of the first two moments, when the FNEA computation is replicated starting from a different set of initial canonical configurations sampled in the HREM stage and using the same annihilating protocol. Even more, one can repeat the FNEA experiment using a *different duration time* and evaluate the alchemical potential of mean force (PMF) as $\Delta G_{b/u}(\lambda) = \langle W_{b/u}(\lambda) \rangle - \frac{1}{2}\beta\sigma_{b/u}^2(\lambda)$ along the alchemical coordinate $0 < \lambda < 1$, with $\lambda = 0$ and $\lambda = 1$ representing the fully coupled and fully decoupled state, respectively. If the work distributions are normal at all λ and for all τ, this quantity should be stationary at any λ irrespective of the selected duration τ of the NE experiment or of the adopted annihilation protocol. As an example, we show for the $\Delta G_b(\lambda)$ the annihilation of the bound G1 and G0 in the OA and TEMOA system. Both the G1-OA and TG0-TEMOA work distributions for $\tau = 360$ and for $K = 0.03$ passed the AD test (data not shown). As it can be seen on the full energy scale, the PMFs computed on the basis of the normality assumption on the whole λ interval using different duration times are barely distinguishable. Differences can be appreciated only in the enhanced view of the insets. Note that increasing the duration time of the FNEA runs, leads to annihilation free energy estimates differing at most 1 kcal mol^{-1} at full guest annihilation $\lambda = 1$. The $\tau = 720$ ps estimates for both the G1-OA and G0-TEMOA

fall within the error bar of the corresponding $\tau = 360$, with the 95% confidence interval computed according Eq. 3. This fact is remarkable indeed considering that
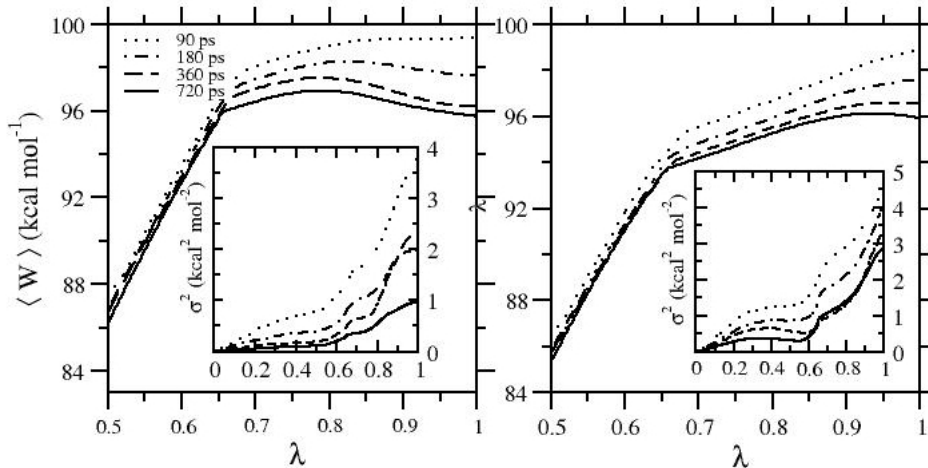

Fig. 7 Mean (main plots) and variance (insets) as a function of the λ alchemical parameter of the annihilation work distribution of the guest molecule in the G1-OA complex (left) and in the G0-TEMOA complex (right), obtained using different annihilation rates.
.

the final work distributions are characterized by strongly τ-dependent moments, as show in the Figure 7. For the two G1-OA and G0-TEMOA complexes, the mean, $\langle W \rangle$, drops by $\simeq 4$ kcal mol^{-1} with duration time passing from $\tau = 90$ ps to $\tau = 720$ ps. A similar trend is observed in the variance σ^2.

2.5 Standard state volume and finite size corrections

As discussed in Section 2.1, the FSDAM estimate should be independent of the imposed restraint volume V_r in the HREM and FNEA stages (see on is issue Table 3 further on). However, according to Eq. 5, the estimate does depend on the binding site volume V_{site}. In the following we provide some details on the computation of V_{site} appearing in the FSDAM dissociation free energy estimate, Eq. 5. In the submitted blind prediction we have evaluated the V_{site} from the the distributions of the host-guest COM-COM *vector* distance. This has been accomplished by numerically evaluating in the HREM equilibrium simulations the volume enclosed in a Van Der Waals-like surface defined by spheres with centers at the sampled COM-COM vector points and radius 0.1 Å. As such, this choice for the elusive V_{site}, while apparently reasonable, still remains arbitrary to some extent. Fortunately, the logarithmic dependence of V_{site} in Eq. 5 significantly tames the uncertainty of the volume correction (see Table 1 further on). The binding site volume determination is undoubtedly the weak point of the FSDAM approach.

However, the binding site volume determination is the rather undervalued weak point of *any* computational approach based on the definition of "bound state".[27] These include also methodologies aimed at the determination of *relative* binding free energy which all implicitly (and arbitrarily) assume a constancy of the binding site volume upon transmutation of the bound ligand into another bound parent compound.

As discussed above, the finite size correction in FSDAM using PME can be straightforwardly computed *a priori* by means of Eq. 6. This box volume dependent correction defines the direct lattice contribution to the Wigner self energy.[38] In the SAMPL6 challenge, this correction affects significantly the annihilation free energy of the complex, namely $\Delta G_{\mathrm{fs}}^{(b)} = -2\alpha^2[Q_H^2 - (Q_H + Q_G)^2]/V_{\mathrm{BOX}}^{(b)}$, since in that case the net charge goes from -9e to -8e, with $\Delta G_{\mathrm{fs}}^{(b)}$ amounting to approximately 1.3 kcal mol^{-1} while $\Delta G_{\mathrm{fs}}^{(u)}$ is negligible in the other leg of the FNEA process, i.e. the annihilation of the the guest molecule in bulk. Note that, being the Wigner self energy more negative for the fully coupled state, the finite size correction for the guest annihilation in the bound state is *positive*. The finite size effect can be numerically tested by repeating the HREM and FNEA calculation of the annihilation free energy in the complex using a much larger box at constant Ewald α convergence parameter. If the work distribution of the larger sample are normal for any λ, we do expect to observe an uncorrected PMF, $\Delta G_b(\lambda)$, consistently larger with respect to that of the smaller sample. The final uncorrected annihilation free energies should differ by the positive quantity

$$\Delta G_b(V_L) - \Delta G_b(V_s) = \frac{\pi}{2\alpha^2}[Q_H^2 - (Q_H + Q_G)^2]\left(\frac{1}{V_L} - \frac{1}{V_s}\right) \qquad (7)$$

where V_s and V_L are the small and large box volume respectively. We tested the finite size correction in the G2-OA system (that passed the AD normality test), by repeating the HREM and FNEA stages using a mean box volume V_L of about 150000 Å^3 corresponding to a side-length of about 53 Å containing about 5000 water molecules. Due to high computational cost of the large sample computation, we produced only 120 annihilation trajectories, with an expected 95% confidence interval increasing by a factor of $\sqrt{2}$ according to Eq. 3. As V_L is about 4 times larger than V_s, according to the Eq. 7 we do expect the difference $\Delta G_b(V_L) - \Delta G_b(V_s)$ to be approximately equal to $3/4 \times \Delta G_{\mathrm{fs}}^{(b)}$, i.e. 3/4 times the finite size correction for the smaller sample. This is happily confirmed in Figure 8, where the two PMFs computed with the large and small box differ almost exactly by the expected quantity computed according to Eq. 7. We stress that such test, besides numerical assessing the finite size effect verifying Eq. 7, constitutes yet an other strong confirmation of the extraordinary robustness and soundness of the Gaussian based estimate 2, given that two PMFs refers to two independent HREM and following FNEA computations done on samples of different size.

3 Results and discussion

3.1 The FSDAM blind prediction on OA and TEMOA systems

All data for delivering our blind prediction of the octa-acids SAMPL6 challenge are collected in Table 1.
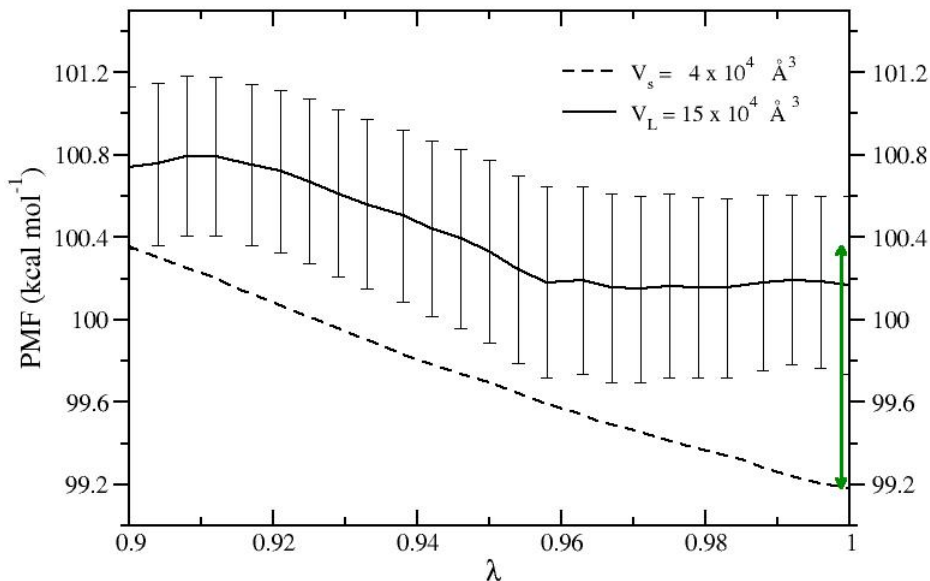
Fig. 8 PMFs as a function of the λ alchemical coordinate for the annihilation of G2 in complex with OA, obtained on samples of different size. The arrowed bar on the right corresponds to the expected difference between the two PMFs on the basis of Eq. 7. Error bars, computed according to Eq. 3, have been plotted only for the PMF of the large system.

The Table reports all the computed quantities that are needed in the FSDAM expression 5 for the evaluation of dissociation free energy of a guest-hots pair. The errors in the evaluation of the moments $\langle W \rangle$ and σ^2 are standardly computed using bootstrap with resampling and, as it can be easily verified using the variances reported in the Table, are remarkably similar to those predicted by Eq. 3 based on the normality assumption. The last column in the Table refers to the A^2 value computed with the AD test form normality on the corresponding work distributions. Since Eq. 5 has been used for all computed work distributions, we may expect a better performance for the OA prediction set with respect to the TEMOA host-guest pairs, the latter exhibiting in general higher A^2 values and wider work ditributions (see Figure 5). In Figure 9, we show the correlation diagram between experimental[1] and FSDAM computed standard dissociation free energies for the OA (left) and TEMOA (right) systems. We may say in general that our FSDAM blind prediction, with a RMSD of less than 2.5 kcal mol^{-1}, has a good ranking among all submitted genuine blind predictions (i.e without adjusted free energies based on retrospective data). FSDAM ranked in 4-th place for OA with a mean RMSD below 2 kcal mol^{-1} and in 5-th place for TEMOA with a mean RMSD below 3 kcal mol^{-1}.

OA bound states					
system	$\langle W \rangle$	$\frac{1}{2}\beta\sigma^2$	$RT\ln(V_{\text{site}}/V_0)$	ΔG_{fs}	AD test
G0-OA	97.9 ± 0.13	1.9 ± 0.28	-2.73	1.26	0.383
G1-OA	96.2 ± 0.14	1.7 ± 0.29	-2.57	1.26	0.489
G2-OA	100.3 ± 0.18	2.2 ± 0.26	-2.72	1.26	1.121
G3-OA	97.7 ± 0.10	1.6 ± 0.22	-2.69	1.26	0.443
G4-OA	102.9 ± 0.13	2.3 ± 0.31	-2.46	1.26	0.444
G5-OA	95.5 ± 0.14	1.5 ± 0.18	-2.78	1.26	0.401
G6-OA	97.0 ± 0.12	2.0 ± 0.22	-2.70	1.26	0.891
G7-OA	96.8 ± 0.19	2.4 ± 0.31	-2.88	1.26	1.082

TEMOA bound states					
system	$\langle W \rangle$	$\frac{1}{2}\beta\sigma^2$	$RT\ln(V_{\text{site}}/V_0)$	ΔG_{fs}	AD test
G0-TEMOA	96.5 ± 0.15	2.7 ± 0.31	-2.62	1.26	1.342
G1-TEMOA	98.8 ± 0.14	2.0 ± 0.27	-2.85	1.26	1.378
G2-TEMOA	93.8 ± 0.21	5.8 ± 0.90	-2.63	1.26	0.452
G3-TEMOA	98.5 ± 0.15	2.3 ± 0.43	-2.89	1.26	0.739
G4-TEMOA	104.6 ± 0.11	2.0 ± 0.25	-2.85	1.26	0.392
G5-TEMOA	95.1 ± 0.15	2.0 ± 0.29	-3.08	1.26	0.512
G6-TEMOA	99.4 ± 0.26	2.9 ± 0.47	-2.83	1.26	0.395
G7-TEMOA	95.1 ± 0.25	5.7 ± 1.09	-2.59	1.26	0.937

unbound states					
system	$\langle W \rangle$	$\frac{1}{2}\beta\sigma^2$	$RT\ln(V_{\text{site}}/V_0)$	ΔG_{fs}	AD test
G0	88.1 ± 0.11	0.6 ± 0.09	-	0.10	0.414
G1	87.3 ± 0.08	0.3 ± 0.04	-	0.10	0.437
G2	85.6 ± 0.09	0.4 ± 0.06	-	0.10	0.286
G3	89.3 ± 0.07	0.3 ± 0.06	-	0.10	0.396
G4	88.7 ± 0.11	0.6 ± 0.11	-	0.10	0.287
G5	87.2 ± 0.08	0.4 ± 0.08	-	0.10	0.503
G6	88.6 ± 0.10	0.6 ± 0.09	-	0.10	0.320
G7	86.9 ± 0.09	0.4 ± 0.08	-	0.10	0.330

Table 1 Raw FSDAM data for the SAMPL6 octa-acids challenge. All energy entries are in kcal mol^{-1}. Standard dissociation free energy can be computed, assuming normality for all work distributions, from the reported energy contribution using Eqs. 2, 5, 7. In the last column the A^2 value for the AD test for normality is reported.

As in many MD-based SAMPL6 approaches, also in our case results are definitely better for the OA host with respect to TEMOA cavitand. For the OA host, only G2 and G4 FSDAM dissociation free energies differ by more that 3.5 kcal mol^{-1} with respect to the experimental counterpart, all others differing at most for 1.5 kcal mol^{-1}. The fact that the outlier G2 and G4 (in both the OA and TEMOA FSDAM prediction sets) are the only chiral species in the SAMPL6 challenge can instill the doubt that the experimental measurements could actually refer to a non enantiopure sample or to the racemic mixture. Organizers assurances and Figure 2 for the case of G2 definitely rules out this wrong inference. As shown by our HREM simulation, the two G2 enantiomers, according to the GAFF2 force field, both exhibit overwhelmingly the extended (equatorial) conformation in bulk. Therefore, because of the symmetry of the host species, the two enantiomers should have the same affinity towards OA and TEMOA. Similar considerations applies to the G4 chiral guest molecule as well, where the asymmetric carbon has a minimal impact on the conformational landscape of the molecule.
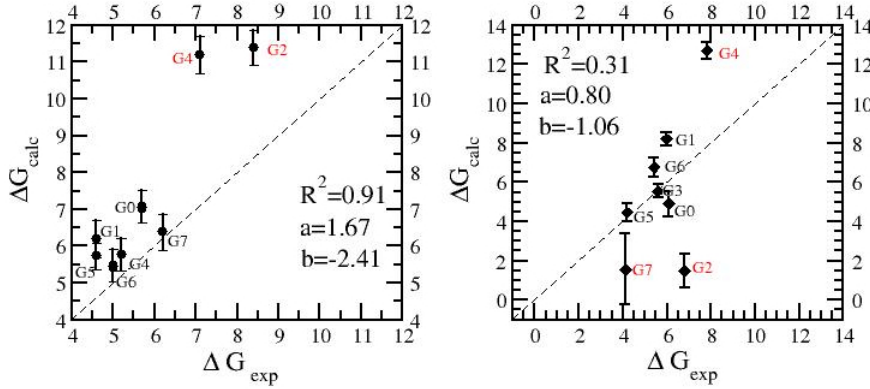
Fig. 9 Correlation diagram between experimental and calculated dissociation free energies (kcal mol^{-1}) for the OA (left panel) and TEMOA (right panel) host-guest pairs. The normal assumption and Eq. 5 was used in the estimate and τ was set to 360 ps in the FNEA stage for both the bound and bulk state. $K_h = 0.003$ was set to kcal mol^{-1} Å^{-2} in both the HREM and FNEA stages. The quantities a and b are the slope and intercept, respectively, of the regression line. The dashed line with $a = 1$ and $b = 0$ corresponds to perfect match.

For the TEMOA system, the correlation coefficient is indeed unsatisfactory dropping from 0.91 in the OA guests to only to 0.31 (see Figure 9, right panel). This small value is due to three FSDAM predictions that are completely off the mark, i.e. those involving again the chiral species G2 and G4 and that referring to G7. While in the case of G2 and G4 discrepancies could be due to a deficiency of the GAFF2 force field and/or to an erroneous PrimaDORAC atomic type assignment, for G7 the problem could be due to the normality assumption implied in the estimate of the annihilation free energies in Eq. 5. As a matter of fact, the G7-TEMOA bimodal work distribution failed the AD test (see Table 1) and the error bar using a single Gaussian is abnormally large. Moreover, we recall that the AD test gives only the probability for *rejecting* the null hypothesis (i.e. the work are normally distributed) but does not provide any *certitude* on the correctness of the null hypothesis.

3.2 FSDAM with Gaussian mixtures

In the following, we generalize the FSDAM approach assuming that the principal component of the observed bound state distribution is given by a *mixture* of normal distributions, i.e.

$$P_b(W) = (1 - c) \sum_i^{K_b} w_i^b n_b(W, \mu_i^b, \sigma_i^b) + c n_d(W) \tag{8}$$

where $n_d(W)$ is the shadow component distribution corresponding to the dissociated states with the guest allowance volume V_r, μ_i^b and σ_i^b are the mean and variance of the i-*th* normal sub-component in the bound state mixture, w_i^b, and

K_b are the normalized weights and the number of sub-components, respectively. Under this assumption it can be easily shown, using again the Crooks theorem, that the estimate Eq. 5 still holds with ΔG_b determined as

$$\Delta G_b = -RT \ln \left[\sum_i^{K_b} w_i^b e^{\mu_i^b - \frac{1}{2}\beta(\sigma_i^b)^2} \right] \qquad (9)$$

Note that Eqs. 10 comes down to Eq. 2 when K_b is equal to 1, i.e. when $P_b(W)$ is normal. A similar generalization can be extended to the work distribution of the unbound state that can be assumed to be given by a mixture of the kind $P_u(W) = \sum_i^{K_u} w_i^u n_u(W, \mu_i^u, \sigma_i^u)$ so that

$$\Delta G_u = -RT \ln \left[\sum_i^{K_u} w_i^u e^{\mu_i^u - \frac{1}{2}\beta(\sigma_i^u)^2} \right] \qquad (10)$$

We stress that the variant of the FSDAM based on mixtures concerns only work data post-processing and not the production stages HREM and FNEA.

In order to deliver a dissociation free energy estimate based on Eq. 5, where the observed work distributions are assumed to be a mixture of normal components, one has first to identify the number K_b and K_u of normal components in the bound state and unbound state distributions produced by the ordinary FNEA stages and then, for each normal component in a given distribution one has to determine the weight w_i and the parameters μ_i and σ_i. As shown in Ref. [15], when dealing with NE processes that are characterized by strongly asymmetric forward and reverse work distributions, as in the folding/unfolding of proteins or in the docking/undocking of a ligand, it is convenient to keep the number of components as small as possible in order to minimize the uncertainty in the mixture-based estimate Eqs. 10 and **??**. The minimum number of component $K_{b/u}$ in the mixtures can be determined by computing the standardized skewness $\gamma = E[((x-\mu)/\sigma))^3]$ and kurtosis $\kappa = E[((x-\mu)/\sigma))^4] - 3$ of the work distributions. For normal distribution these higher moments should be equal to zero within the confidence intervals $\delta\gamma$ and $\delta\kappa$ determined by bootstrapping with resampling. Then, we set K as

$$
\begin{aligned}
|\gamma| < \delta\gamma, \quad & |\kappa| < \delta\kappa| \quad & \rightarrow K = 1 \\
|\gamma| > \delta\gamma, \quad & |\kappa| < \delta\kappa| \quad & \rightarrow K = 2 \\
|\gamma| < \delta\gamma, \quad & |\kappa| > \delta\kappa| \quad & \rightarrow K = 2 \\
|\gamma| > \delta\gamma, \quad & |\kappa| > |\delta\kappa| \quad & \rightarrow K = 3
\end{aligned} \qquad (11)
$$

Once K has been selected, the $3K-1$ independent parameters of a K-component mixture, namely $\theta_K = b_1..b_{K-1}, \mu_1..\mu_K, \sigma_1..\sigma_K$ can be effectively determined using the Expectation Maximization (EM) algorithm. For a detailed explanation of the EM algorithm see Refs. [58,59]. Here it suffices to say that EM is a maximum likelihood iterative algorithm that starts from some initial estimate of the parameter θ_K (e.g., random), and then proceeds to iteratively update θ_K until convergence is detected. Each iteration consists of an expectation-step (E-step) and maximization step (M-Step). In the E-step, the membership weights $w_{ij}, i = 1..N, j = 1..K$ of each work value in the sample of size N (i.e. the probability of the work W_i to

belong to the j-th normal component) is determined using the guess θ_K. In the M-step, the membership weights and the data are used to compute a new estimate of the parameters of the mixture θ_K. These steps are repeated until maximum log-likelihood is reached. The EM algorithm for a sample containing some hundreds of value is remarkably fast (few fraction of seconds on a laptop PC).[60] On the other hand, the error goes as $[1/(N/K)]^{1/2}$, i.e. increases with the number of components. In order to keep the confidence interval small when using mixtures, one can either increase the number N of work values by a factor K or one can increase the annihilation time τ at constant N, as this will correspondingly decreases the overall width of the distribution, thereby increasing accuracy (see e.g. Figure 7).

The determination of the binding site volume V_{site} is an other significant source of methodological error in the FSDAM technology. As discussed in the previous section, V_{site} in Table 1 was determined by the oscillation of the COM-COM ligand-host vector distance in the HREM stage and by computing the volume enclosed in in a Van Der Waals-like surface defined by spheres with centers at the sampled COM-COM vector points and radius 0.1 Å. This rather complicated approach tends to overestimate the binding site volumes producing values between a minimum of 9 Å^3 to a maximum of 30 Å^3. In the EM updated prediction, made assuming mixtures *in lieu* of a single Gaussian, following a suggestion given in Refs. [27,21], we have re-determined the binding site volumes simply as $V_{\text{site}} = \frac{4}{3}\pi(2\sigma)^3$, where σ^2 is the variance of the HREM determined COM-COM distributions reported in Figure 3.

In Table 2, we show the results obtained using the mixture method with the EM algorithm and the updated binding site volume correction on the same work data sets used in Table 1. We first note that according to the test statistic based on higher moments, Eq. 11, all the distributions that failed the AD tests exhibit, expectedly, a mixture character with $K > 1$. However, some of the work distributions that passed the AD tests, were also found with $K > 1$. For example, two unbound work distributions, i.e. those involving G4 and G7, were found with K=2, according to Eq. 11. This contradictory outcome is due to the slow convergence of the test statistic (e.g. sample bootstrap) of the γ and κ skewness and kurtosis, which makes the test behave erratically over under-sized or even in a reasonably large sample. Nonetheless, the resulting EM approach did not change significantly the estimates for the four ΔG_u annihilation free energies for which $K = 2$, while it had a negligible to small impact on the ΔG_b data that passed the AD tests (data not shown). As it can be seen from Table 2, the approach based on the mixture estimates Eqs. 10, **??** and 11 somewhat improved the agreement with the experimental dissociation free energies in both OA and TEMOA systems. Using the estimates Eq. 2 and Eq. 4, the correlation coefficients R^2 passed from 0.91 to 0.93 and from 0.31 to 0.46 for OA and TEMOA pairs, respectively and the RMSD went from 1.99 to 0.94 kcal mol^{-1} in OA and from 2.90 to 2.78 kcal mol^{-1} in TEMOA. Predictions of G2 and G4 in OA and of G7 in TEMOA improved due to a more negative contribution from the binding site volume.

In the Table 3 we show the quality of the FSDAM predictions as measured by the Pearson correlation coefficient R^2, the slope a and the intercept b, using the mixture approach for the dissociation free energy estimates as a function of the imposed restraint volume $V_r = (2\pi RT/K_h)^{3/2}$ in HREM and FNEA and of the ligand annihilation times $\tau_{b/u}$ in the bound and unbound state. Most of the EM-based FSDAM unsubmitted FSDAM predictions are only slightly better

System	ΔG_{exp}	$\Delta G_{\text{calc}}^{\text{EM}}$	$RT\ln(V_{\text{site}}/V_0)$	ΔG_{fs}	K_b	K_u
OA-G0	5.70	5.89 ± 0.21	-3.88	1.26	1	1
OA-G1	4.60	4.52 ± 0.19	-4.19	1.26	1	1
OA-G2	8.40	9.22 ± 0.41	-4.79	1.26	2	1
OA-G3	5.20	3.98 ± 0.62	-4.01	1.26	2	1
OA-G4	7.10	9.22 ± 0.25	-4.50	1.26	1	2
OA-G5	4.60	4.64 ± 0.29	-3.70	1.26	2	1
OA-G6	5.00	5.29 ± 0.25	-3.01	1.26	3	1
OA-G7	6.20	5.88 ± 0.21	-3.25	1.26	2	2
TEMOA-G0	6.06	3.29 ± 0.68	-3.67	1.26	2	1
TEMOA-G1	5.97	6.24 ± 0.57	-4.19	1.26	2	1
TEMOA-G2	6.81	0.44 ± 1.18	-3.79	1.26	2	1
TEMOA-G3	5.60	4.97 ± 0.22	-3.51	1.26	1	1
TEMOA-G4	7.80	11.06 ± 0.27	-4.55	1.26	1	2
TEMOA-G5	4.20	3.88 ± 0.26	-3.63	1.26	1	1
TEMOA-G6	5.40	5.75 ± 0.38	-3.78	1.26	1	1
TEMOA-G7	4.10	2.55 ± 0.52	-1.52	1.26	2	2

Table 2 EM-based FSDAM prediction set assuming a mixture of normal components for the null hypothesis on the work distributions. All energy data are in kcal mol^{-1}. ΔG_{exp}: Experimental dissociation free energy[1]; $\Delta G_{\text{calc}}^{\text{EM}}$: EM-based FSDAM prediction; $RT\ln(V_{\text{site}}/V_0)$: standard volume correction with $V_{\text{site}} = \frac{4}{3}\pi(2\sigma)^3$; ΔG_{fs}: finite size correction; K_b, K_u: number of components found in the bound state and unbound state work distributions according to Eq. 11. The annihilation times for the bound and unbound states were $\tau_b = 360$ ps and $\tau_u = 360$ ps and $K_h = 0.003$ kcal mol^{-1}Å^{-2}

with respect to the SAMPL6 submitted blind prediction using the single Gaussian assumption based on Eq. 2. All the EM-predictions are strongly correlated one to another, demonstrating the expected insensitivity of the FSDAM technique (see Eq. 5) with respect to the choice of the reference volume V_r. It must be pointed out, however, a significant exception to this rule occurring in the TEMOA-G7 complex with the strongest restraint corresponding to a volume of 1.4 Å^3. In this case, the guest molecule was found in a pose set on the exterior of the cavitand, outside the funnel-shaped binding pocket in the majority of the sampled configurations in HREM, yielding a low annihilation energy component in Eq. 10, eventually producing a strongly negative dissociation free energy. This appears to be an artifact induced by the restraint, when imposing a ligand allowance volume V_r of the same order or less of V_{site}. Strong COM-COM restraints where at least one of the partner is a flexible molecule like in the TEMOA-G7 can hence occasionally induce high energetic barriers that make difficult for the ligand, once random fluctuations of the surrounding have brought it into a secondary pose, to re-enter the binding site.

3.3 Force field issues?

We finally conclude this section by commenting on the difference quality for the FSDAM prediction of the OA and TEMOA systems, irrespectively of the used approach, either using a single Gaussian (see dat on Figure 9) or a mixture of normal components (see Table 3) as the null hypothesis for the work distributions. We first note that in general, genuine blind predictions from all methodologies

adopted in the SAMPL6 challenge performed less satisfactorily for G2 and G4 in the OA systems and G2 G4 and G7 in the TEMOA complexes,[1] i.e. exactly the same host-guest pairs yielding the largest deviations in our blind predictions. As a matter of fact, if we eliminate these five host-guest pairs from our submitted blind prediction, the overall RMSD drops from 2.49 to only 1.14 kcal mol^{-1}. Similar results are obtained for the other EM-based unsubmitted predictions.

Many of the MD-based methodology shared the same general force field approach that was used in our prediction, namely the GAFF or GAFF2 general parameterization set with atomic charges determined at the AM1-BCC level. In this regard, it should be noted[1] that the only blind MD-based prediction in SAMPL6 using the charmm CGenFF force field[61] performed well for G2, G4 and G7 in the TEMOA systems. This leads us to believe that there could be a problem in the GAFF parameterization of the torsional potential of the G2, G4 and G7 ligands, producing wrong conformational distributions in the guest molecule. For example, PrimaDORAC assigns the H type on methyl or methylene groups on the basis of the total withdrawn Mulliken charge found on the group rather than examining the character of the neighboring groups. In G7, all H atoms attached to aliphatic carbons were assigned the GAFF type `hc`, in spite of the proximity of the carboxylic and methylene moieties. The 1-4 interactions involving these hydrogen atoms could produce an excessive rigidity in the torsional potential, artificially favoring, in the unbound state, conformations with wrong hydration energies or that binds the hosts less effectively. Such an effect can possibly be amplified in a host with a more hindered pocket as in TEMOA. Moreover, the larger σ in the type `hc` with respect to a charge depleted H types like `h1` can make G7 binding to TEMOA more difficult than to the OA host. A similar argument could possibly apply also to G2 and G4: in G2 again all H atoms attached to aliphatic carbons are assigned by PrimaDORAC to the GAFF kind `hc`, while in G4 only those bound to the carbon linked to the carboxylate moiety are of the kind `h1`. On the other hand, for the other ligands for which the FSDAM prediction is satisfactory and that also contains methyl and methylene groups close to the carboxylate moiety, these considerations does not seem to apply as well, with troublesome implications on the PrimaDORAC assignment protocol and/or on the transferability of the GAFF2 atomic types across the SAMPL6 ligands. Further systematic computations are needed to assess these force field issues that appears to be shared by all GAFF-based SAMPL6 uncorrected prediction sets.

4 Conclusion

In this paper, we have computed, by means of a non equilibrium alchemical technique (termed FSDAM), the absolute dissociation free energy for the octa-acids host-guest systems provided by the SAMPL6 initiative. FSDAM is based on the production of canonical configurations of the bound and unbound states via replica exchange with solute tempering (HREM stage), followed by the generation of hundreds of fast non-equilibrium ligand annihilation trajectories (FNEA stage), eventually producing a collection of bound and unbound annihilation work values. The annihilation free energies of the ligand when bound to the receptor and in bulk solvent, are obtained from the collection of NE work values using an estimate based on the Crooks theorem for driven NE processes in the assumption that the NE

EM/Exp. correlation				OA			TEMOA		
K_h	V_r	τ_b	τ_u	R^2	a	b	R^2	a	b
0.003	40000	360	360	0.93	1.42	-2.00	0.46	1.19	-2.18
0.03	1661	360	360	0.88	1.34	-1.42	0.63	1.71	-4.11
0.15	125	360	360	0.93	1.43	-1.46	0.53	1.29	-2.12
0.30	45	360	360	0.95	1.46	-2.00	0.55	1.24	-1.69
3.0	1.4	360	360	0.96	1.14	-0.26	0.44	2.74	-13.9
0.03	1661	720	360	0.96	1.35	-0.20	0.59	1.38	-1.26
EM/EM(ref) correlation				OA			TEMOA		
K_h	V_r	τ_b	τ_u	R^2	a	b	R^2	a	b
0.003	40000	360	360	0.97	1.06	-0.65	0.97	1.07	-0.69
0.03	1661	360	360	0.95	1.02	-0.11	0.99	1.03	0.39
0.15	125	360	360	0.99	1.08	-0.42	0.99	1.03	-0.08
0.30	45	360	360	0.96	1.06	-0.18	0.99	1.08	0.20
3.0	1.4	360	360	0.90	0.76	1.36	0.80	1.78	6.20

Table 3 Quality of the EM-based FSDAM predictions for the OA and TEMOA SAMPL6 host-guest systems as measured by the R^2 Pearson coefficient and by the slope a and intercept b of the regression line using different values of the force constant in the COM-COM restraint potential (see Section 2.1) and of the annihilation time for the bound state τ_b.

work are normally distributed. The absolute binding free energies are recovered by the differences of the annihilation free energies of the ligand in bulk and in the bound state, corrected by a standard state and finite size terms.

For normal work distributions, the confidence interval of the FSDAM estimates can be assessed from the mean and variance following the t-statistics and chi-square distribution, respectively, or from standard bootstrapping techniques. The proposed technique is specifically tailored for modern multi-cores HPC platforms, easily engaging from hundreds (in the HREM stage) to thousands (in the FNEA stage) of parallel instances by means of an efficient hybrid OpenMP-MPI implementation, allowing job completion for a SAMPL6 host-guest pair in few wall-clock hours or minutes. FSDAM has been applied using a completely automatic procedure for the generation of the topology and parameter files of the host and guest molecules, of the input files and batch submission scripts on HPC Platforms, and of the data post-processing and methodological confidence level assessment. As such, the present calculation on the representative SAMPL6 set constitutes hence a prototypical test for industrial applications of FSDAM in ligand-receptor systems.

In general our FSDAM SAMPL6 blind prediction, with a RMSD of less than 2.5 kcal mol^{-1}, has a satisfactory ranking among all submitted genuine blind predictions (i.e without adjusted free energies based on retrospective data). We show that results can be improved, affecting only the data post-processing stage, by using a generalization of the FSDAM approach based on the assumption that the annihilation work samples are distributed according to a mixture of normal components rather than being represented by a single Gaussian distribution. For some host guest pairs, notably those involving the G2 and G4 guest molecules in both OA and TEMOA hosts, and the G7 species in the TEMOA molecule, the FSDAM dissociation free energies differ, on the average, by more than 3.5 kcal mol^{-1} with respect to the experimental values. Such an outcome is shared by many of the MD-based SAMPL6 predictions that used the same GAFF/TIP3P force

field approach adopted by us, leading to believe that the GAFF parameterization could be defective for the G2, G4 ligands and for the TIP3P mediated G7-TEMOA interaction with possible implications on the assignment protocol and/or on the transferability of the GAFF2 atomic types across the SAMPL6 ligands. Further systematic computations are needed to assess this force field issue that appears to be shared by all GAFF-based SAMPL6 uncorrected prediction sets.

Possible future improvements of the FSDAM technique could involve the setting up of bidirectional approaches based on ligand annihilation and growth processes. In this regard, as already successfully done in Ref. [10] for ligands in bulk, a still unexplored possibility would be that of computing the binding free energy of the reverse path for the bound state as well, using a suitably restrained ligand-receptor system for the HREM sampling and subsequent fast growth of the initially gas-phase ligand, [27,21] hence paving the way for the use of the reliable bidirectional Bennett Acceptance Ratio estimator.[62] Furthermore, the bound state distributions of the fast-growth processes would allow the direct calculation of the volume V_{site}, as the weight of the principal component corresponding to V_{site}/V_r. Standard implementation of bidirectional approaches would at least double the computational cost in FSDAM. However, as the majority of the fast-growth trajectories in the computationally demanding bound state would be highly dissipative, this could be accomplished using a coarse-grained path-breaking approach as suggested in Ref. [63]

References

1. Drug Design Data Resource (D3R), SAMPL6: Host-guest binding and physical property prediction. Information available at https://drugdesigndata.org/about/sampl (accessed 15 March 2018); Full free energy data available at https://drugdesigndata.org/upload/community-components/d3r/workshop2018/presentations/DavidMobley2018_SAMPL_overview.pdf (accessed 15 April 2018).

2. Jian Yin, Niel M. Henriksen, David R. Slochower, Michael R. Shirts, Michael W. Chiu, David L. Mobley, and Michael K. Gilson. Overview of the sampl5 host–guest challenge: Are we doing better? *J. of Comput. Aided Mol. Des.*, pages 1–19, 2016.

3. Hari S. Muddana, Andrew T. Fenley, David L. Mobley, and Michael K. Gilson. The sampl4 host–guest blind prediction challenge: an overview. *J. Comput Aided Mol. Des.*, 28(4):305–317, 2014.

4. Corinne L. D. Gibb and Bruce C. Gibb. Binding of cyclic carboxylates to octa-acid deep-cavity cavitand. *J Comput Aided Mol Des*, 28(4):319–325, Apr 2014. 24218290[pmid].

5. Haiying Gan and Bruce C. Gibb. Guest-mediated switching of the assembly state of a water-soluble deep-cavity cavitand. *Chem. Commun.*, 49:1395–1397, 2013.

6. Jacobs H. Jordan and Bruce C. Gibb. Molecular containers assembled through the hydrophobic effect. *Chem. Soc. Rev.*, 44:547–585, 2015.

7. G. M. Torrie and J. P. Valleau . Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *J. Comp. Phys.*, 23:187–199, 1977.

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. Piero Procacci and Chiara Cardelli. Fast switching alchemical transformations in molecular dynamics simulations. *J. Chem. Theory Comput.*, 10:2813–2823, 2014.

11. Robert B. Sandberg, Martina Banchelli, Carlo Guardiani, Stefano Menichetti, Gabriella Caminati, and Piero Procacci. Efficient nonequilibrium method for binding free energy calculations in molecular dynamics simulations. *J. Chem. Theory Comput.*, 11(2):423–435, 2015.

12. 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.

13. 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.

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

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

16. Piero Procacci. 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, 2017.

17. Piero Procacci. 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, 2016.

18. Cresco: Centro computazionale di ricerca sui sistemi complessi. Italian National Agency for New Technologies, Energy (ENEA). See `https:/http://www.cresco.enea.it` (accessed date 24/06/2015).

19. Consorzio Interuniversitario del Nord est Italiano Per il Calcolo Automatico (Interuniversity Consortium High Performance Systems) http://www.cineca.it (accessed 22 January 2018).

20. GAFF and GAFF2 are public domain force fields and are part of the AmberTools16 distribution, available for download at http://amber.org internet address (accessed March 2017). According to the AMBER development team, the improved version of GAFF, GAFF2, is an ongoing poject aimed at "reproducing both the high quality interaction energies and key liquid properties such as density, heat of vaporization and hydration free energy". GAFF2 is expected "to be an even more successful general purpose force field and that GAFF2-based scoring functions will significantly improve the successful rate of virtual screenings.".

21. Edoardo Giovannelli, Piero Procacci, Gianni Cardini, Marco Pagliai, Victor Volkov, and Riccardo Chelli. Binding free energies of host-guest systems by nonequilibrium alchemical simulations with constrained dynamics: Theoretical framework. *J. Chem. Theory Comput.*, 12:5874–5886, 2017.

22. Edoardo Giovannelli, Matteo Cioni, Piero Procacci, Gianni Cardini, Marco Pagliai, Victor Volkov, and Riccardo Chelli. Binding free energies of host-guest systems by nonequilibrium alchemical simulations with constrained dynamics: Illustrative calculations and numerical validation. *J. Chem. Theory Comput.*, 13:5887–5899, 2017.

23. 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.

24. S. Marsili, G. F. Signorini, R. Chelli, M. Marchi, and P. Procacci. Orac: A molecular dynamics simulation program to explore free energy surfaces in biomolecular systems at the atomistic level. *J. Comput. Chem.*, 31:1106–1116, 2010.

25. M. K. Gilson, J. A. Given, B. L. Bush, and J. A. McCammon. The statistical-thermodynamic basis for computation of binding affinities: A critical review. *Biophys. J.*, 72:1047–1069, 1997.

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

27. 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.
28. Jan Hermans and Lu Wang. 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, 1997.
29. 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.
30. Y. Deng and B. Roux. Computations of standard binding free energies with molecular dynamics simulations. *J. Phys. Chem. B*, 113:2234–2246, 2009.
31. Hengbin Luo and Kim Sharp. On the calculation of absolute macromolecular binding free energies. *Proc. Natnl. Acad. Sci. USA*, 99(16):10399–10404, 2002.
32. P. Procacci, S. Marsili, A. Barducci, G. F. Signorini, and R. Chelli. Crooks equation for steered molecular dynamics using a nosé-hoover thermostat. *J. Chem. Phys.*, 125:164101, 2006.
33. Sanghyun Park and Klaus Schulten. Calculating potentials of mean force from steered molecular dynamics simulations. *J. Chem. Phys.*, 120(13):5946–5961, 2004.
34. Kalimuthu Krishnamoorthy. *Handbook of Statistical Distributions with Applications*. Chapman and Hall/CRC, London (UK), 2006.
35. C.R. Sanders. Biomolecular ligand-receptor binding studies: theory, practice, and analysis. Nashville: Vanderbilt University, 2010. from: https:structbio.vanderbilt.edu/sanders/Binding_Principles_2010.pdf as of 22 April 2018.
36. D.A. Lauffenburger and J.J. Linderman. *Receptors: Models for Binding, Trafficking, and Signaling,*. Oxford University Press, 1993.
37. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. A smooth particle mesh ewald method. *J. Chem. Phys.*, 103:8577–8593, 1995.
38. Tom Darden, David Pearlman, and Lee G. Pedersen. Ionic charging free energies: Spherical versus periodic boundary conditions. *J. Chem. Phys.*, 109(24):10921–10935, 1998.
39. Araz Jakalian, David B. Jack, and Christopher I. Bayly. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.*, 23(16):1623–1641, 2002.
40. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart. Am1: A new general purpose quantum mechanical model. *J. Am. Chem. Soc.*, 107, 1985.
41. W. L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926–935, 1983.
42. M. Parrinello and A. Rahman. Crystal structure and pair potentials: A molecular-dynamics study. *Phys. Rev. Lett.*, 45:1196–1199, Oct 1980.
43. M. Marchi and P. Procacci. Coordinates scaling and multiple time step algorithms for simulation of solvated proteins in the npt ensemble. *J. Chem. Phys.*, 109:5194–520g2, 1998.
44. M. Tuckerman and B. J. Berne. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.*, 97:1990–2001, 1992.
45. P. Procacci, E. Paci, T. Darden, and M. Marchi. Orac: A molecular dynamics program to simulate complex molecular systems with realistic electrostatic interactions. *J. Comput. Chem.*, 18:1848–1862, 1997.
46. R. M. Levy E. Gallicchio. Agbnp: An analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling. *J. Comput. Chem.*, 25:479–499, 2004.
47. Y. Sugita and Y. Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, 314:141–151, 1999.
48. G. Ponti, F. Palombi, D. Abate, F. Ambrosino, G. Aprea, T. Bastianelli, F. Beone, R. Bertini, G. Bracco, M. Caporicci, B. Calosso, M. Chinnici, A. Colavincenzo, A. Cucurullo, P. Dangelo, M. De Rosa, P. De Michele, A. Funel, G. Furini, D. Giammattei, S. Giusepponi, R. Guadagni, G. Guarnieri, A. Italiano, S. Magagnino, A. Mariano, G. Mencuccini, C. Mercuri, S. Migliori, P. Ornelli, S. Pecoraro, A. Perozziello, S. Pierattini, S. Podda, F. Poggi, A. Quintiliani, A. Rocchi, C. Scio, F. Simoni, and A. Vita. The role of medium size facilities in the hpc ecosystem: the case of the new cresco4 cluster integrated in the eneagrid infrastructure. In *Proceeding of the International Conference on High Performance Computing & Simulation*, pages 1030–1033. Institute of Electrical and Electronics Engineers (IEEE), 2014.

49. Piero Procacci, Marco Bizzarri, and Simone Marsili. Energy-driven undocking (edu-hrem) in solute tempering replica exchange simulations. *J Chem. Theory Comp.*, 10:439–450, 2014.

50. Note that the PMF route to K_d still needs the definition of a binding site volume, that is of an "indicator function"[**?**,31,25] depending on the relative guest-host coordinates that can distinguish between "bound" state and "unbound" state.

51. T.C. Beutler, A.E. Mark, R.C. van Schaik, P.R. Gerber, and W.F. van Gunsteren. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.*, 222:5229–539, 1994.

52. Daniel K. Shenfeld, Huafeng Xu, Michael P. Eastwood, Ron O. Dror, and David E. Shaw. Minimizing thermodynamic length to select intermediate states for free-energy calculations and replica-exchange simulations. *Phys. Rev. E*, 80:046705, 2009.

53. 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.*, 11:2536–2549, 2015.

54. T. W. Anderson and D. A. Darling. A test of goodness of fit. *J. Am. Stat. Ass.*, 49:765–769, 1954.

55. M.A. Stephens. Test of fit for the logistic distribution based on the empirical distribution function. *Biometrika*, 66:591–595, 1979.

56. N. M. Razali and Y. B. Wah. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *J. Stat. Model. Anal.*, 2:21–33, 2011.

57. Tanweer Ul Islam. Stringency-based ranking of normality tests. *Commun. Stat. Simul. Comput.*, 46(1):655–668, 2017.

58. A.P. Dempster, N.M. Laird, and D.B Rubin. Maximum likelihood from incomplete data via the em algorithm. *J. Royal Stat. Soc. B*, 39:1–38, 1977.

59. Maya R. Gupta and Yihua Chen. Theory and use of the em algorithm. *Foundations and Trends in Signal Processing*, 4(3):223–296, 2011.

60. An EM Fortran90 code is available with the ORAC6 distribution tarball at the ORAC site `www.chim.unifi.it/orac` .

61. A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wirkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B*, 102(18):3586–3616, 1998. PMID: 24889800.

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

63. Edoardo Giovannelli, Cristina Gellini, Giangaetano Pietraperzia, Gianni Cardini, and Riccardo Chelli. Combining path-breaking with bidirectional nonequilibrium simulations to improve efficiency in free energy calculations. *J. Chem. Phys.*, 140(6):064104, 2014.