Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

Solvation free energies via alchemical simulations: let's get honest about sampling, once more

Piero Procacci,*[a]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

Free energy perturbation (FEP) approaches with stratification see widespread and increasing use in computational studies of biologically relevant molecules. However, when the molecular system are characterized by a complex conformational free energy landscape, the assessment of convergence remains a concern for many practitioners. The sampling problem in FEP has been authoritatively addressed in a recent perspective paper [D. Mobley, J. Comput. Aided Mol. Des., 26, 93 2012], incisively entitled "Let's get honest about sampling". Here, I return on the issue of sampling in the determination of the octanol-water partition coefficient for a synthetic precursor of Kinase inhibitors that has been included in the recent extension of the SAMPL6 blind challenge of LogP coefficients. I will show, that even for this simple compound, whose conformational space is essentially dictated by two sp_3 rotable bonds connecting rigid planar units, canonical sampling using standard techniques can be surprisingly hard to achieve. I will also show how the conformational sampling problem can be effectively bypassed using unidirectional and bidirectional non equilibrium work methods, reliably recovering the solvation energy with minimal methodological uncertainty.

Introduction

The determination of reliable solvation free energies using computational approaches is central in *in silico* drug discovery. Solvation free energies differences of drug size molecules in aqueous and hydrophobic phases (e.g. water/octanol partition coefficients, LogP) are useful for predicting the distribution of drugs within the body. Binding free energies themselves can be viewed[1,2] as the difference between the solvation free energy of the ligand in bulk solvent and in the composite environment of the protein binding pocket.

Solvation free energies and binding free energies of drug size molecules can nowadays be computed using seemingly well established computational techniques based on Free Energy Perturbation method (FEP)[3] with the system being represented by accurate atomistic force fields.[4–6] In FEP, the free energy between two end states, say 1 (gas-phase solute) and 0 (solvated solute), is given by

$$e^{-\beta\Delta G} = \langle e^{-\beta\Delta U}\rangle_0 \qquad (1)$$

where $\Delta U = U_1 - U_0$ is the potential energy difference between gas-phase and solvated solute and the canonical average $\langle . \rangle_0$ is done with the Hamiltonian of state 0. The dominant contribution to the exponential average in Eq. 1 comes from the configurations that are typical of state 1. When there is little overlap between the energy distributions at 0 and 1, these dominant microstates are generated very rarely using conventional molecular dynamics simulations (MD), leading to slow convergence.

To enhance accuracy, FEP approaches are normally implemented using the so-called stratification strategy or multistage sampling along a suitably defined "alchemical" coordinate λ, whereby the system is simulated in an appropriate number n of intermediate states corresponding to values of λ between 0 and 1, leading to the expression

$$e^{-\beta\Delta G} = \sum_{i=0}^{n-1} \langle e^{-\beta(U(\lambda_{i+1})-U(\lambda_i))}\rangle_{\lambda_i} \qquad (2)$$

where, typically, $U(\lambda) = U_0 + \lambda(U_1 - U_0)$. The solvation free energy is thus obtained as a sum of the free energy changes between consecutive λ states along the alchemical path with significant overlap of the corresponding potential energy distribution functions. The "curse"[7] of the exponential averages can be avoided altogether by storing just the $U(\lambda_i)$'s for each stratum[8] during the simulation and exploiting, in an inexpensive post-processing stage, powerful bidirectional maximum likelihood estimators like the so-called Bennett Acceptance Ratio (BAR).[9]

Stratification is costly since one has to perform *n equilibrium*

[a] *Department of Chemistry, University of Florence, Italy. E-mail: procacci@unifi.it*

† Electronic Supplementary Information (ESI) available: MBAR computed 2D FES as a function of the θ, ϕ dihedral angles. Replica exchange rates and GE states distributions in the replica walkers for the HREM simulations. Bidirectional PMF's in water and 1-octanol computed from the NEW data for SM02. See DOI: 10.1039/b000000x/

MD simulations to recover ΔG as a sum of BAR determined ΔG_i. On the other hand, these simulations are independent and can be performed in parallel with no communication overhead.

Although on paper the method may appear rather straightforward, the technicalities of modern FEP strategies are by all means non trivial.[8,10–14] The basic problem in implementing Eq. 2 is that 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",[15] a task that would require the prior knowledge of the dependence of ΔG on λ.[8] Besides, the rate of convergence of the canonical averages in the i-th stratum depends on the value of λ_i in an unknown way.[14] So, for example, barriers between conformational states in complex molecular solutes in polar solvents can increase dramatically when the screening between polar or charged moieties is gradually switched off along the alchemical path, hence making convergence at low values of λ extremely hard. A poor convergence in a single stratum may lead to a significant error in the final solvation free energy.

In the context of FEP with stratification, various workarounds to these difficulties have been devised. Among the most popular, we may cite here the so-called λ-hopping techniques[16,17] and the FEP/REST technology (FEP/Replica exchange with solute tempering).[18] Both these approach are based on the Hamiltonian Replica exchange method (HREM).[19] In λ-hopping simulations, the n independent simulations are lumped in a single replica exchange simulation on a generalized ensemble (GE) spanning the whole range [0,1] of the alchemical coordinate. Contiguous replicas attempt exchanges of their λ values at regular time intervals with a probability of acceptance regulated by a Metropolis criterion. In this manner one should have in principle n weakly communicating GE walkers crossing back and forth the stratification along the λ alchemical coordinate, possibly enhancing the sampling of conformational states separated by λ-dependent barriers with minimal computational overhead. It should be stressed, however, that λ-hopping, while mixing states from different Hamiltonians, has no effect whatsoever on the barrier heights between conformational states and hence cannot be considered, strictly speaking, as an *enhanced sampling* technique.

The FEP/REST approach was implemented as a modification of the λ-hopping technique precisely to overcome the sampling issue that are inherent in this technique. In FEP/REST some selected degrees of freedom of the solute are heated at intermediate λ values using a stratum-dependent scaling factor on an appropriate part of the potential energy function, smoothing conformational barriers and enhancing in this manner transitions between conformational minima. The extra scaling of the potential energy adds up to the overall ΔU between two contiguous strata so that this may require an increase of the total number of strata, n, in order to preserve the overlap between energy distributions and hence a smooth diffusion of the walkers in the GE.

In spite of these advances, the FEP approach, especially when dealing with complex molecular solutes, may be still be plagued by typical HREM convergence problems, related to long round-trip times, bottlenecks, non communicating sub-ensemble ultimately leading to inhomogeneous, non converged stratum distribution of the GE walkers. H-REM-based FEP methodologies are still a matter of specialists and developers, with end users of MD packages mostly using the standard (non HREM-based) FEP technique described in the reference tutorials.[20] In all FEP-based submissions in the recent LogP/SAMPL6 challenge,[21] probably assuming a rather tame conformational landscape of the solutes, H-REM/FEP or FEP/REST were never used.

What is hardly grasped by the average FEP practitioner is that for each λ_i one has to provide a simulation time that guarantees stability (i.e. convergence) of all configurational averages, i.e. *statistical equilibrium must be attained on each point along the alchemical coordinate.* Best practice indications[8,22] cannot cover all possible sampling issues that could be faced in a complex molecular system. In case of highly flexible molecules such as many common drug-sized compounds, the emphasis on overlap and equal variance of potential energy distributions may be even misleading, leading to probe the free energy of a single unimportant conformational local minimum while all "best practice" diagnostic parameters (overlap, variance distribution) are apparently fine.

In this regard, few years ago David Mobley, one of the foremost organizers of the SAMPL project[23–25] and a leading scientist in computational biochemistry, wrote a seminal paper[26] entitled "Let's get honest about sampling", focused on the issue of adequate sampling in free energy calculations of complex molecular systems using classical force fields. According to Mobley, the discrepancies that are often observed in computed solvation or binding free energies "get [in most cases] blamed on force field deficiencies", rather than on specific sampling problems, eventually arguing that the latter and not the former "may be a leading cause of error. [..] With adequate sampling, we can quantitatively assess the accuracy of a particular force field, identify deficiencies, and improve it. Without adequate sampling, there is no such path forward." Unfortunately, this remarkable paper did not receive, by any means, the attention it deserved, and testing the adequacy of sampling in FEP studies remains rather scant in the specialist literature.

In this paper I return on the issue of sampling in the determination of the water octanol partition coefficient for a synthetic precursor of specific Kinase inhibitors, namely the N-[3-(trifluoromethyl)phenyl]quinazolin-4-amine (SM02), that has been included in the extension of the SAMPL6 blind challenge of LogP coefficients.[27] I will show that even for this simple molecule, whose conformational space is essentially dictated by two sp_3 rotable bonds connecting rigid planar units, canonical sampling in the standard FEP approach can be surprisingly hard to achieve, even using simulation times as long as 8 ns per λ state.

I will also show how the solvation energy can be reliably recovered, with minimal methodological uncertainty, using unidirectional and bidirectional non equilibrium work methods (NEW). The suggestive acronym "NEW" has been used in Ref.[8] very likely referring to the novelty of the technique, which is based on theorems in non equilibrium thermodynamics discovered at the end the past century.[28,29] Remarkably, NEW-based techniques requires adequate sampling of *just* the end thermodynamic states at $\lambda = 0,1$, which can be effectively implemented using a very efficient HREM on torsional tempering.[30] Starting from a rep-

resentative sample of the end state, the FEP stratification is replaced by launching several fast independent trajectories where the alchemical coordinate is rapidly and continuously driven to a final non equilibrium phase point of the other end state, eventually producing a non equilibrium work distribution. In this fashion, given that adequate sampling has been reached in the initial equilibrium states, the free energy can be computed straightforwardly using either bidirectional or unidirectional approaches by exploiting very general NE theorems. Most importantly, the confidence level of the free energy can in essence be related to a single parameter, i.e the variance of the non equilibrium work distributions.[7,8,31,32] As such, the non equilibrium stage in NEW can be easily implemented with ideal parallel scaling, thus allowing to reconstruct, *in a single parallel job*, the work distributions in a matter of few minutes (in the case of the LogP of SM02) on moderns multi-cores platforms.

In spite of the fact that NEW methods have been already applied quite successfully to solvation free energies[33,34] as well to drug-receptor[35–37], host-guest[38,39] and even protein-protein[32] binding free energies, the vast majority of practitioners still stick to FEP-based technology. In the recent LogP/SAMPL6 challenge, NEW was used only in my submissions. This reluctance in using NEW is probably due to incomplete or far-stretched analysis studies,[31,40] or to discouraging sentences in authoritative reviews such as[22] "we do not recommend that beginners use these [NEW] methods, as they add an extra degree of complication to both the simulation and the analysis".

In this paper, besides explicitly showing why in FEP-based studies we should get honest about sampling even in apparently simple cases, I shall try to convince beginners and practitioners that the NEW approach, contrary to the common belief, is indeed extremely simple, fast and can be far more reliable, for sound physical reasons, with respect to a "best practice" FEP computations in cases where a complex free energy conformational landscape is at stake.

Computational details

SM02 is depicted in Figure 1. The molecule is composed by a trifluoromethyl-phenyl and quinazolin moieties connected to a secondary amine. Neglecting the CF_3 free internal rotator, the conformational space of this molecule is dictated by the two amino rotable sp_3 bonds. The SM02 rotamers can be clustered into four main conformers which can be classified according to the gauche (G) or anti (A) setup of the two dihedral angles as indicated in the figure. "G" and "A" states are defined to have dihedral angle in the range $|\theta| \leq 60°$, and $120° \leq |\theta| \leq 180°$, respectively. The selected force field for SM02 was GAFF2[41] with AM1/BCC charges[42]. Atom types assignment and atomic charges calculation were done using the web interface PrimaDORAC.[43] The parametrization of the explicit water solvent in hydration free energy calculations is done using the recently developed OPC3[44] three-point site model. For 1-octanol as a solvent, we have adopted the PrimaDORAC generated GAFF2 force field parameters set. Such parametrization yields a density of 0.81 g/cm^3 and a dielectric constant of 6.1 in standard conditions, in satisfacory agreement with the experimental values[45] of 0.83 g/cm^3

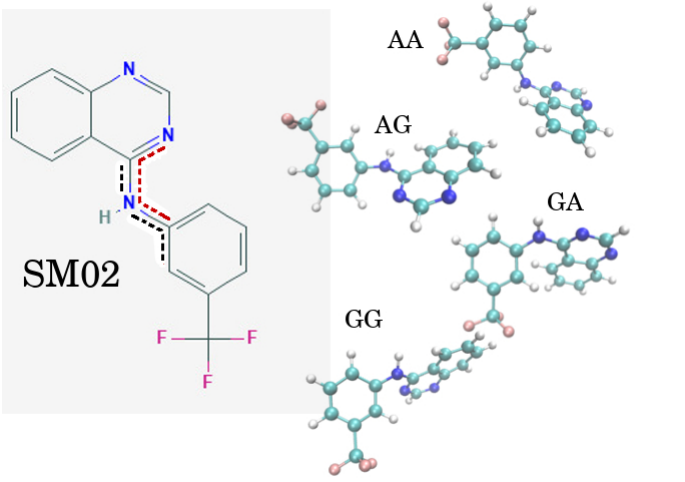

Fig. 1 SM02 molecule (left) and main conformational clusters (right) classified (see text) according to the two dihedral angles traced with a dashed line on the left.

and 10.3, respectively. The SM02 molecule was considered in the neutral form (no protonated amino group).

Solvation free energies were evaluated by dissolving the solute in about 1240 water molecules or 125 molecules of octanol in a cubic MD box. All simulations were done in the NPT isothermal-isobaric ensemble, yielding a mean side-length around 32-33 Å in both water and 1-octanol. The external pressure was set to 1 atm using a Parrinello-Rahman Lagrangian[46] with isotropic stress tensor while temperature was held constant at 300 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[47] with a potential subdivision specifically tuned for bio-molecular systems in the NPT ensemble.[46,48]. The long range cut-off for Lennard-Jones interactions was set to 13 Å in all cases. Long range electrostatic were treated using the Smooth Particle Mesh Ewald method,[49] with an α parameter of 0.38 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.

The Hamiltonian Replica exchange simulations in water or 1-octanol solution and gas-phase SM02 are done using torsional tempering. Torsional tempering, a specialized solute tempering[50] scheme described for the first time in Ref.[30], allows to surgically enhance the sampling on the relevant degrees of freedom of the system keeping the replica number to a minimum. For SM02, the scaling involves just the torsional potentials (including 1-4 non bonded interactions) around the two amino rotable bonds shown in Figure 1 with a minimum scaling factor of $c = 0.1$, corresponding to a "torsional temperature" of 3000 K. Only the scaling factors are communicated among replicas, minimizing interprocessor communication and hence allowing for frequent exchange attempts (15 fs). As discussed in Ref.[51], where the effect of exchange attempts frequency was analyzed in a series of HREM simulations of peptides in explicit solvent, "exchanges should be attempted extremely often, providing gains in efficiency and no

undesired effects."

The torsional GE space is covered using four replicas only, with the scale factors[30] $c_m = c^{(m-1)/3}$.

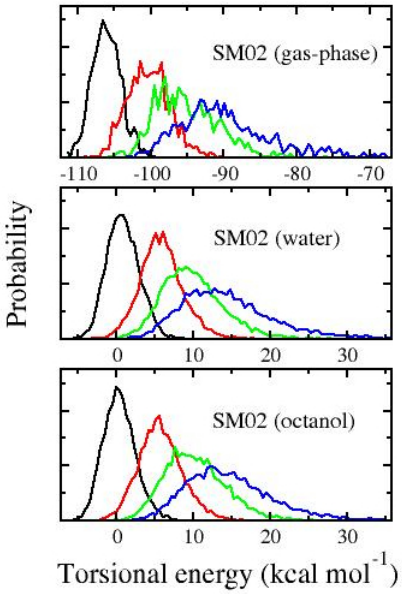

Fig. 2 HREM torsional energy (see text) distributions (from left to right $c_1 = 1,\ c_2 = 0.464,\ c_3 = 0.215,\ c_4 = 0.1$) of SM02 in gas-phase ($\lambda = 1$), water and 1-octanol ($\lambda = 0.0$)

In Figure 2, I report the torsional energy distribution obtained in a 8 ns H-REM simulation in the four replicas for SM02 in the gas-phase, in water and in octanol. As it can be seen, the overlap of contiguous energy replica distributions is significant in all media. In all media, acceptance ratios are between 20% and 60% while round-trip times are of the order of 1 ps or less. Further details on HREM convergence (Replica exchange rates and GE states probability distributions for each replica walker) are provided in the Supporting Information. All calculations were done using the program ORAC.[37] The ORAC code, including the source, can be freely downloaded from the site http://www.chim.unifi.it/orac or from the mirror http://lx03.sm.chim.unifi.it/orac.

Results

A pre-analysis for a FEP implementation on SM02

FEP calculations of solvation or binding free energies are normally done using standardized protocols prepared for the most popular simulation packages.[4,52,53] In most of these specialized tutorials,[20] simple examples are usually provided with simulation times per λ states that rarely exceed few hundreds of picoseconds. The issue of convergence is either not addressed at all (gromacs tutorial), or only vaguely alluded such as "a more serious simulation project [with respect to the tutorial example] would need to run much longer to reach convergence and also may need additional λ-points (NAMD tutorial)". The hydration of ethanol, according to the gromacs tutorial, can be achieved simulating with conventional MD for 200 ps per stratum in six λ strata. This indication is consistent with the slow growth/annihilation simulation of ethanol in water, yielding a reversible path in less than 2 ns.[33] SM02 is made of mostly rigid units and should have a relatively tame conformational landscape. Hence one would expect to use a similar setup for FEP calculations of solvation energy.

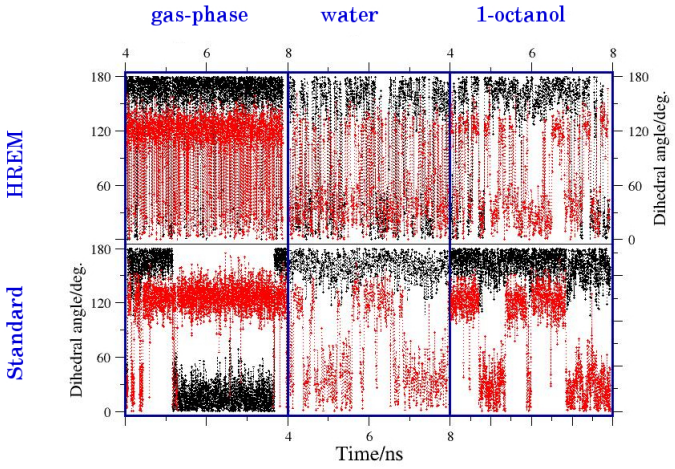

Fig. 3 Time record of the dihedral angles of SM02 (see Figure 1) in the gas-phase, in water and in 1-octanol at 300K, obtained on the target state ($c = 1$) of HREM (upper panels) and with standard MD (lower panels).

In Figure 2, I report the time record of the last (supposedly equilibrated) 4 ns of the two dihedral angles of SM02 (indicated in Figure 1) obtained in the gas-phase, in water and in octanol during a conventional 8 ns MD simulation and in a HREM 8 ns simulation(target state) with torsional tempering. In the HREM simulations, sampling of the all four conformational states is uniform with distribution reflecting the canonical probability at the target state $c = 1$. In the standard MD simulations, sampling of the dihedral angles in all media proceeds by jumps, related to the rare events of the interchange between A and G conformations. The G state in one of the two dihedral angles (the black trait) is *never* visited in both water and 1-octanol while it appears to be over-sampled in the gas-phase. It should also be noted that conformational sampling (expressed by the frequency of the GA swaps) in conventional simulation is much harder in the decoupled solute (gas-phase). This should come with no surprise due to the fact that, in the solvent, GA jumps can be triggered by molecular collisions. Thus, we may reasonably expect that conformational sampling efficiency in standard MD will decrease in going from a fully coupled ($\lambda = 0$) to fully decoupled ($\lambda = 1$) state.

The rational of using the apparently "expensive" torsional HREM for SM02 is that the expected quality of sampling in a enhanced sampling simulation outperforms by far that of conventional run investing the same amount of computational resources, as the collected results neatly show.

In Figure 4, the 2D probability of the dihedral angles, $P(\theta,\phi)$ is translated into a corresponding 2D free energy surface (FES), $-RT\log P(\theta,\phi)$. The difference in the FES between conventional and HREM is indeed striking. Apparently, conventional MD was unable to sample correctly the GA and GG configurations of SM02

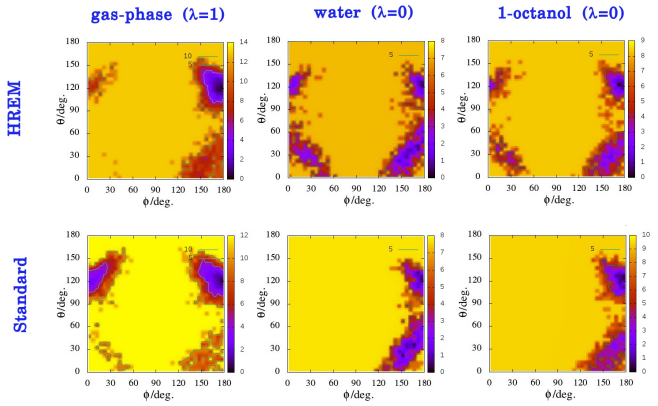

Fig. 4 Free energy surface at T=300 K with respect to the two dihedral angles (ϕ and θ dihedral angles are indicated in black and red color in in Figure 1) of SM02 in the gas-phase, water and 1-octanol with HREM (upper panels) and with standard MD (lower panels). The z-energy scale is in kJ mol^{-1}. In each plot, the zero level of the FES corresponds to the maximum probability state.

in both water and 1-octanol, while it definitely oversamples the GA states in the gas-phase. The free energy differences between the various conformational states at the end states $\lambda = 1$ (fully coupled solute) and $\lambda = 0$ (gas-phase) are of the order of 1 kcal mol^{-1} and do affect the final solvation energies in a complex manner. If these differences are not correct, because of the existence of high energetic barriers separating conformational free energy basins, one cannot expect the final FEP value to be correct either. FES barriers heights between rotameric states on the target state can be easily afforded in HREM simulations by computing the weight of each of the sampled point in the GE using the so-called multi-state Bennett acceptance ratio estimator[54,55] (MBAR). 2D FES plots computed with MBAR are shown in the Supporting Information along with a complete error analysis on relative free energy differences of the four rotameric states (GG, GA, AG, AA, see discussion near Figure 1) in water, octanol and gas-phase computed from the HREM and standard simulations.

Given these results, would it be reasonable to expect reproducible solvation free energies of SM02 from a conventional FEP simulation of few hundreds picoseconds per stratum on few strata? Of course not. In the supporting information, we show that, in order to recover the FES landscape obtained with an 8 ns of HREM simulation, *at least* 100 ns of simulation are needed for the decoupled (gas-phase) SM02 molecule. Standard FEP, in this apparently simple case, should be simply and honestly discouraged, as a basically unreliable methodology unless one is willing to witlessly waste microseconds of conventional simulation.

The λ-hopping techniques are of little help in this case, since conformational transitions are rare at full coupling and ever more at the other extreme of the alchemical coordinate.

FEP/REST could be of some help, but the computational time would definitely increase and the λ protocol (number and spacing of strata) should be carefully tuned with the REST scaling. In this regard, we point out that although the so-called FEP/REST method was originally engineered as a mono-dimensional HREM,[18] two independent scaling parameters are used in the technique and therefore the cost should be reasonably comparable to that of a bi-dimensional HREM.[16] As a matter of fact, neither λ-hopping nor FEP/REST were used in the recent LogP/SAMPL6 challenge. In all FEP submssions[21] standard MD simulations on 10 to 40 λ states were used with simulation times per state ranging from a minumum of 5 ns to a maximum of 20 ns.

We can make a conservative estimate of the computational cost for FEP/REST for this system based on the results reported so far. Using 20 λ states[40] including the end states,[56] each simulated for 4 ns with a four replica exchange solute tempering scheme, we would need (optimistically) a total of 0.64 microsecond of simulation for a *reproducible* LogP coefficient.

NEW techniques for the LogP of SM02

Before embarking in the FEP calculations, we should ask whether the LogP could be computed more efficiently and rapidly with NEW techniques, possibly also providing a decent confidence interval. In NEW, starting form the canonical sampling of one end states, the system is rapidly brought to the other end state in n concurrent and independent NE trajectories by continuously varying the alchemical coordinate λ from 0 to 1 (fast growth) or from 1 to 0 (fast annihilation) and computing the external work as $W = \int \frac{\partial U}{\partial \lambda} \dot{\lambda} dt$,[8] eventually producing a forward (growth) work distribution $P_G(W)$ and a reverse (annihilation) work distribution $P_A(-W)$. The solvation free energy can be recovered from the collection of NE works using either unidirectional or bidirectional estimates based on the Jarzynski and Crooks non equilibrium theorems:

$$\Delta G = \sum_{i=1}^{n} e^{-\beta W_i(G)} \tag{3}$$

$$\Delta G = -\sum_{i=1}^{n} e^{-\beta W_i(A)} \tag{4}$$

$$\sum_{i=1}^{n} \frac{1}{1+e^{\beta(W_i(G)-\Delta G)}} = \sum_{i=1}^{n} \frac{1}{1+e^{\beta(W_i(A)+\Delta G)}} \tag{5}$$

where Eqs. 3, 4, 5 are the unidirectional Jarzynski forward estimate, the undirectional Jarzynsky reverse estimate and the bidirectional Crooks estimates based on the maximum likelihood Bennett acceptance ratio.[57] Note that the fast switching protocol of the forward and reverse transformations must be related by time reversal in order to use the Crooks theorem in Eq. 5.

We have already collected the sampling of the end states, done with HREM (upper panels in Figure 2). What we need now is to produce the fast growth and annihilation of SM02 in a swarm of independent NE trajectories initiated by representative samples of the canonical end states at $\lambda = 0, 1$. In setting up the NE stage in NEW, one has to know that the accuracy of the method decreases with increasing "dissipation", namely the difference between the mean NE work $\langle W \rangle$ and the unknown underlying free energy. This is so since the accuracy of the estimate depends on the overlap of the forward and reverse distribution.[8,58]

From past experience on solvation energies of simple organic

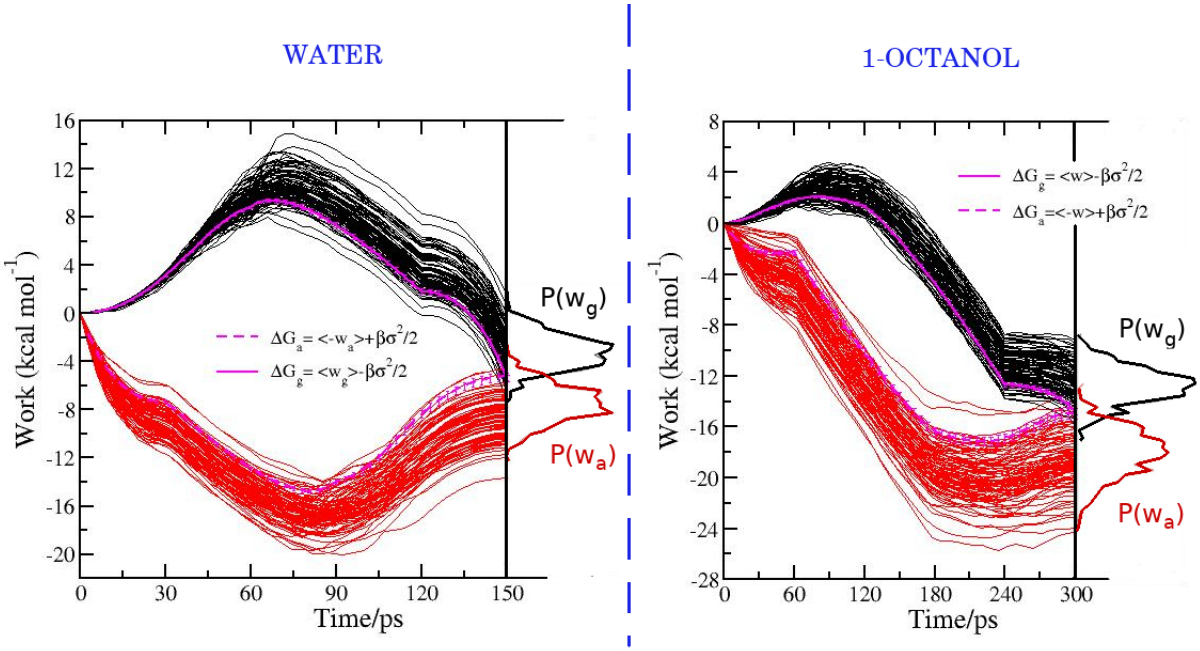

Fig. 5 Work functions during 100 representative NE fast switching alchemical trajectories for SM02 in water and 1-octanol. Growth and annihilation trajectories are in black and red color, respectively. In each solvent, the fast growth alchemical protocol involves sequentially the Lennard-Jones interaction ($1.0 \leq \lambda < 0.2$) using a soft-core Beutler potential[59] regularization as $\lambda \to 1$ and the electrostatic interactions $0.2 \leq \lambda \leq 0.0$. The fast annihilation process is done with inverted λ schedule. The forward (growth), $P(W_G)$, and backward (annihilation), $P(W_A)$, work distributions are shown on the right side of the plots.

molecules[33,34,40], we will conservatively use a fast switching time for SM02 of 150 ps in water and of 300 ps in 1-octanol and compute the NE work distribution from 240 NE trajectories in each direction. The λ NE protocol provides that electrostatic and Lennard-Jones solute-solvent interactions are switched on or off sequentially.[33] NE fast switching (growth or annihilation) computations were done with the hybrid OpenMP/MPI ORAC program in a single parallel job on the CRESCO6 HPC platform provided by ENEA[60] and based on the Intel(R) Xeon(R) Platinum 8160 CPU 2.10GHz 48-cores processor. Each of these four parallel jobs (growth and annihilation in water and 1-octanol) engaged a total of 1440 cores, namely 240 instances on the MPI layer for the non communicating NE trajectories each running with six OpenMP threads. The total wall clock time to get the LogP result was less than that half an hour. The total computational time invested in the calculation was less than 250 ns including the 32 ns of the initial equilibrium state. In the panels of Figure 5, the NEW methodology for the computation of LogP in SM02 is illustrated in Figure 5. The work done during the NE growth and minus the work done in the NE annihilation processes are represented in black and red color, respectively. Only 100 representative works (out of the 240 produced) in each direction are shown. The work distributions at the end of the transformations are depicted on the right of the plots.

The solvation free energy is computed straightaway from the work distributions using BAR.[8,9,61] In BAR, the error can be evaluated using standard bootstrap methods or even analytically, using the theoretical variance of the BAR estimate based on the Fischer information.[61] The BAR estimates for the solvation energy of SM02 in water and 1-octanol using the GAFF2 force field are found to be $\Delta G(w) = -5.52 \pm 0.14$ kcal mol^{-1} and $\Delta G(o) = -15.10 \pm 0.15$ kcal mol^{-1}, respectively, yielding a partition coefficient of LogP $= 7.05 \pm 0.21$. NEW computations using the OPLS-AA force field[6,62] (work data not shown) yields LogP $= 6.24 \pm 0.43$. The experimental LogP value for SM02 is 4.1 Full SAMPL6 data are availble in Ref.[21].

In Figure 5, the magenta lines refer to the potential of mean force (PMF) computed using the Gaussian assumption on every point of the λ alchemical coordinate, namely

$$\begin{aligned} \Delta G(\lambda) &= \langle W_G(\lambda) \rangle - \frac{1}{2}\beta\sigma_G(\lambda)^2 \\ \Delta G(\lambda) &= -\langle W_A(\lambda) \rangle + \frac{1}{2}\beta\sigma_A(\lambda)^2 \end{aligned} \qquad (6)$$

where $\langle W_{G/A} \rangle$ and $\sigma^2_{G/A}$ are the mean and variance of the work distribution computed in the growth or annihilation process. The two equations in Eq. 6 constitutes two *independent* estimates of the same quantity $\Delta G(\lambda)$. The normality of the distributions can be checked using well established normality tests, such as the Anderson-Darling test.[63] If the NE work is normally distributed, the Crooks theorem implies that the forward and reverse work distribution are symmetric with respect the crossing point $W = \Delta G$, with $\beta\sigma_A(\lambda)^2/2 = \beta\sigma_G(\lambda)^2/2 = W_{\text{diss.}}$ corresponding to the *dissipated work* along the alchemical coordinate. It should be stressed that Eqs. 6 are asymptotically (i.e. for $n_{\text{works}} \to \infty$) *exact*[64] if the distribution is Gaussian. Hence, each of the two equations in Eq. 6 provides a reliable *unidirectional* and *unbi-*

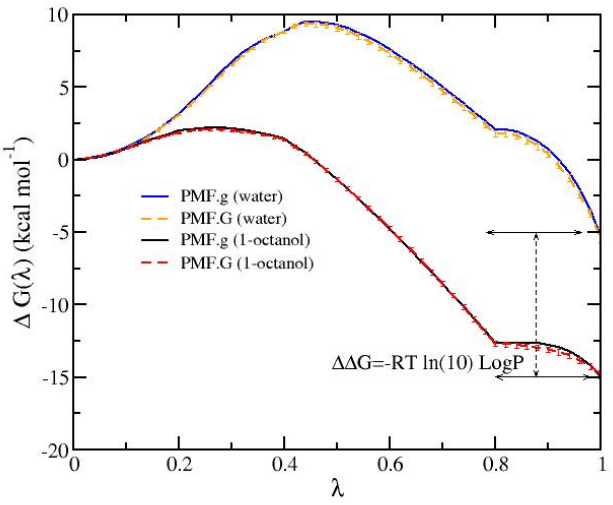

Fig. 6 pmf

ased estimate for finite work samples, with confidence interval depending only on the sample variance and given by[8,32]

$$\delta\Delta G(\lambda) = 1.96\left|\frac{\sigma(\lambda)}{n_{\text{works}}^{1/2}} + \frac{1}{2}\beta\left(\frac{2}{n_{\text{works}}}\right)^{1/2}\sigma(\lambda)^2\right| \qquad (7)$$

Note that halving the number NE trajectories, increases the error by a factor $\sqrt{2}$.

The correctness of the Gaussian assumption can be instantly checked using a single `tac` unix command on the PMF's obtained from the annihilation side (`PMF.A`) or the growth side (`PMF.G`), i.e.

```
$ tac PMF.A | awk 'BEGIN {tt=$1;dg=$2}\
 {print -($1-tt),-($2-dg)}' } > PMF.g
$ tac PMF.G | awk 'BEGIN {tt=$1;dg=$2}\
 {print -($1-tt),-($2-dg)}' } > PMF.a
```

If the Gaussian assumption is correct, then the reconstructed PMF's `PMF.g` and `PMF.a` must be identical to the `PMF.G` and `PMF.A` functions, respectively. In Figure 6, we report the actual and reconstructed PMF from the growth side, `PMF.G` and `PMF.g` for SM02 in water and in 1-octanol. The PMF with $\lambda > 0.8$ corresponds to the SM02 charging process in both water and 1-octanol. The Figure 6 represents a *stunning* demonstration of the beauty and power of the Crooks theorem: the function `PMF.g`, obtained by `tac` inversion of `PMF.A`, computed with a completely independent set of NE measurements with respect to `PMF.G`, follows the function `PMF.G` in the tiniest details along the λ coordinate, always remaining within the confidence interval provided by Eq. 7. This result is important, since, when we are dealing with normal distributions, we can reliably compute the LogP via Eqs 6 using unidirectional measurements instead of using the BAR approach on bidirectional experiments, hence drastically reducing the computational demand of the NEW method. The bidirectional PMF for SM02, calculated using the estimator proposed in Ref.[65], is reported in the Supporting Information.

I must stress here that the growth-based unidirectional estimate is computationally very convenient since the end states sampling is based on a H-REM simulation on a *isolated*, gas-phase molecule embedded as a ghost in a common equilibrated solvent for all solutes. Preliminary analysis of NEW SAMPL6/LogP submission[21] has shown that accuracy is only moderately downgraded in the unidirectional (growth) estimates which are on other hand, extremely favorable from a computational standpoint: a single LogP can be computed in a *matter of minutes* on a Tier-1 HPC system such as the CRESCO6 ENEA cluster equipped with Intel Skylake 48 cores CPU 2.4 GHz.[66]

Conclusion

In this paper I have shown why "we should get honest about sampling"[11] when implementing standard FEP technique with stratification, using, as a specific example, the determination of the solvation free energies of a simple organic compound that has been included in the forthcoming extension of the SAMPL6 blind challenge of LogP coefficients. I show that standard MD is unable to reproduce the HREM determined FES with respect to the dihedral angles connecting the planar moieties of SM02 in any media even when running for as long as 8 ns. These results call for much more careful assessment of the convergence issue in FEP approaches than those proposed in best practice guides of widespread MD suites. The issue of adequate sampling in FEP calculation of solvation and binding free energies is intimately connected to that of providing a *reliable* methodological (i.e. not force field related) confidence interval, a quantity that is of no less importance of the free energies themselves.

The conformational sampling problem can be to a large extent bypassed at an acceptable computational cost using unidirectional and bidirectional non equilibrium work (NEW) methods, reliably recovering the solvation energy with minimal methodological uncertainty. NEW, at variance with FEP stratification technology, requires equilibrium sampling only of the end states sparing that of the inner alchemical states by exploiting powerful non equilibrium theorems. When dealing with system with hard-to-sample conformational landscapes, NEW is shown, in the specific case of SM02, to require less computational resources with respect to a well converged FEP calculation, providing at the same time a far more reliable methodological confidence interval for the computed LogP. In this regard uncertainty on the computed value should be considered as a key quantity in drug development, being strictly related to the investment risk in industrial projects based on theoretical predictions. Since the NE work histograms provides at the same time the free energy and the associated confidence interval, in the NEW approach, so long as the HREM sampling of the end states reflects their canonical distribution, by design there cannot be no "uncertainty on the uncertainty",[26] and hence there is no need for optimizing the alchemical protocol as, in e.g, adaptive[14] equilibrium FEP-based alchemical calculations.

NEW approaches are still rarely used in free energy calculations. In the recent SAMPL6 initiative for blind prediction of host-

guest binding free energies[25], the only NEW submission[39] turned out be one of the best performing methods in the challenge. These good performances of the NEW approach have been confirmed in the second round of the SAMPL6 challenge for the LogP.[21] In spite of the fact that the rare applications of NEW methods have been quite successfully for solvation and binding free energies of complex molecular systems,[32–39] the vast majority of practitioners still rely upon FEP-based technology, in most cases systematically and incautiously underrating the sampling issue.[26] This is probably due to incomplete or far-stretched analysis studies[31,40] where NEW methods have been shown, in specific examples, to be computationally comparable to or less efficient than equilibrium approaches. As our apparently simple case has showed, comparison done on the Widom insertion in a 32-particles hard spheres fluid[31] or on hydration free energies[40] of small rigid molecules may lead to incautious conclusions and cannot be, by any means, of general validity.

In the recently released SAMPL6 challenge results on LogP coefficient,[21] the six NEW submissions (Cgen, GAFF2 and OPLS-AA force fields with BAR or unidirectional fast-growth estimates) were consistently among the highest ranking in the context of the MD-based prediction sets. Most importantly, NEW produced similar and mutually correlated LogP estimates, independently on the force field used or on the the kind of estimate (BAR-based or unidirectional fast-growth). This is strinkigly at variance with the very disparate behaviour of FEP submissions in SAMPL6, with different variants yielding results on opposite side of the performance scale, revealing the sampling deficiencies that may occur in standard or uncareful implementation of this technique.

It is expected that the present paper, besides recalling once more that "only with adequate sampling [at all λ's] we can quantitatively assess the accuracy of a particular force field, identify deficiencies, and improve it" and that no data can be trusted if convergence has not accurately (and expensively) being tested, may be of some help in overcoming the reluctance of beginners and practitioners to explore new "paths forward"[26] in free energy calculations.

Acknowledgments

The computing resources and the related technical support used for this work have been provided by CRESCO/ENEAGRID High Performance Computing infrastructure and its staff.[60] 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).

References

1 M. K. Gilson, J. A. Given, B. L. Bush and J. A. McCammon, *Biophys. J.*, 1997, **72**, 1047–1069.

2 P. Procacci and R. Chelli, *J. Chem. Theory Comput.*, 2017, **13**, 1924–1933.

3 R. W. Zwanzig, *J. Chem. Phys.*, 1954, **22**, 1420–1426.

4 R. Salomon-Ferrer, D. A. Case and R. C. Walker, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2013, **3**, 198–210.

5 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. WiÃşrkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586–3616.

6 L. S. Dodda, J. Z. Vilseck, J. Tirado-Rives and W. L. Jorgensen, *J. Phys. Chem. B*, 2017, **121**, 3864–3870.

7 U. Ryde, *J. Chem. Theory Comput.*, 2017, **13**, 5745–5752.

8 A. Pohorille, C. Jarzynski and C. Chipot, *J. Phys. Chem, B*, 2010, **114**, 10235–10253.

9 C. H. Bennett, *J. Comput. Phys.*, 1976, **22**, 245–268.

10 W. Jorgensen and L. Thomas, *J. Chem. Theory Comput.*, 2008, **4**, 869–876.

11 D. L. Mobley and P. V. Klimovich, *J. Chem. Phys.*, 2012, **137**, 230901–230901.

12 N. Hansen and W. F. van Gunsteren, *J. Chem. Theory Comput.*, 2014, **10**, 2632–2647.

13 P. Procacci, *J. Mol. Graph. and Model.*, 2017, **71**, 233–241.

14 Z. X. Sun, X. H. Wang and J. Z. H. Zhang, *Phys. Chem. Chem. Phys.*, 2017, **19**, 15005–15020.

15 L. N. Naden and M. R. Shirts, *J. Chem. Theory Comput.*, 2015, **11**, 2536–2549.

16 W. Jiang and B. Roux, *J. Chem. Theory Comput.*, 2010, **6**, 2559–2565.

17 E. Gallicchio and R. M. Levy, *Current Opinion in Structural Biology*, 2011, **21**, 161 – 166.

18 L. Wang, Y. Deng, J. L. Knight, Y. Wu, B. Kim, W. Sherman, J. C. Shelley, T. Lin and R. Abel, *J. Chem. Theory Comput.*, 2013, **9**, 1282–1293.

19 Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.*, 1999, **314**, 141–151.

20 See for example the GROMACS manual and the tutorial for alchemical calculations: *Hands-on tutorial Solvation free energy of ethanol* available at http://www.gromacs.org (accessed date 04/04.2014). For NAMD, See the tutorial: *In silico alchemy: A tutorial for alchemical free-energy perturbation calculations with NAMD* available at http://www.ks.uiuc.edu (accessed date 04/04.2014).

21 Drug Design Data Resource (D3R), SAMPL6: Host-guest binding and physical property prediction. Information available at https://drugdesigndata.org/about/sampl (accessed 15 May 2019); A prelimary analysis on LogP data submissions is available at https://github.com/MobleyLab/SAMPL6 (accessed 15 May 2019).

22 M. R. Shirts and D. L. Mobley, *Methods in molecular biology*, 2013, **924**, 271–311.

23 H. S. Muddana, A. T. Fenley, D. L. Mobley and M. K. Gilson, *J. Comput Aided Mol. Des.*, 2014, **28**, 305–317.

24 J. Yin, N. M. Henriksen, D. R. Slochower, M. R. Shirts, M. W. Chiu, D. L. Mobley and M. K. Gilson, *J. Comput. Aided Mol.*

Des., 2016, **31**, 1–19.
25 A. Rizzi, S. Murkli, J. N. McNeill, W. Yao, M. Sullivan, M. K. Gilson, M. W. Chiu, L. Isaacs, B. C. Gibb, D. L. Mobley and J. D. Chodera, *J. Comput. Aided Mol. Des.*, 2018, **32**, 937–963.
26 D. L. Mobley, *J. Comput. Aided Mol. Des.*, 2012, **26**, 93–95.
27 Drug Design Data Resource (D3R), SAMPL6: Host-guest binding and physical property prediction. Information available at https://drugdesigndata.org/about/sampl (accessed 15 December 2018).
28 G. E. Crooks, *J. Stat. Phys.*, 1998, **90**, 1481–1487.
29 C. Jarzynski, *Phys. Rev. Lett.*, 1997, **78**, 2690–2693.
30 S. Marsili, G. F. Signorini, R. Chelli, M. Marchi and P. Procacci, *J. Comput. Chem.*, 2010, **31**, 1106–1116.
31 H. Oberhofer, C. Dellago and P. L. Geissler, *J. Phys. Chem. B*, 2005, **109**, 6902–6915.
32 P. Procacci, *J. Chem. Teory Comput.*, 2018, **14**, 3890–3902.
33 P. Procacci and C. Cardelli, *J. Chem. Theory Comput.*, 2014, **10**, 2813–2823.
34 V. Gapsys, D. Seeliger and B. de Groot, *J. Chem. Teor. Comput.*, 2012, **8**, 2373–2382.
35 P. Procacci, *Phys. Chem. Chem. Phys.*, 2016, **18**, 14991–15004.
36 F. Nerattini, R. Chelli and P. Procacci, *Phys. Chem. Chem. Phys.*, 2016, **18**, 15005–15018.
37 P. Procacci, *J. Chem. Inf. Model.*, 2016, **56**, 1117–1121.
38 R. B. Sandberg, M. Banchelli, C. Guardiani, S. Menichetti, G. Caminati and P. Procacci, *J. Chem. Theory Comput.*, 2015, **11**, 423–435.
39 P. Procacci, M. Guarrasi and G. Guarnieri, *J. Comput. Aided Mol. Des.*, 2018, **32**, 965–982.
40 A. Yildirim, T. A. Wassenaar and D. van der Spoel, *J. Chem. Phys.*, 2018, **149**, 144111.
41 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.".
42 A. Jakalian, D. B. Jack and C. I. Bayly, *J. Comput. Chem.*, 2002, **23**, 1623–1641.
43 P. Procacci, *J. Chem. Inf. Model.*, 2017, **57**, 1240–1245.
44 S. Izadi and A. V. Onufriev, *J. Chem. Phys.*, 2016, **145**, 074501.
45 S. Kim, P. A. Thiessen, E. E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B. A. Shoemaker, J. Wang, B. Yu, J. Zhang and S. H. Bryant, *Nucleic Acids Res.*, 2016, **44**, D1202–D1213.
46 M. Marchi and P. Procacci, *J. Chem. Phys.*, 1998, **109**, 5194–520g2.
47 M. Tuckerman and B. J. Berne, *J. Chem. Phys.*, 1992, **97**, 1990–2001.
48 P. Procacci, E. Paci, T. Darden and M. Marchi, *J. Comput. Chem.*, 1997, **18**, 1848–1862.
49 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.
50 P. Liu, B. Kim., R. A. Friesner and B. J. Berne, *Proc. Acad. Sci.*, 2005, **102**, 13749–13754.
51 D. J. Sindhikara, D. J. Emerson and A. E. Roitberg, *J. Chem. Theory Comput.6*, 2010, **6**, 2804–2808.
52 M. J. Abraham, T. Murtola, R. Schulz, S. PÃąll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1-2**, 19–25.
53 J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, L. Skeel and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781–1802.
54 M. R. Shirts and J. D. Chodera, *J. Chem. Phys.*, 2008, **129**, 124105.
55 P. Procacci, *J. Chem. Phys.*, 2013, **139**, 124105.
56 G. Duarte, G. Calabró and D. Mobley, *Infinite Dilution Activity Coefficients as Constraints for Force Field Parameterization and Method Development*, 2018, chemrxiv preprint, https://doi.org/10.26434/chemrxiv.7199786.v2.
57 M. R. Shirts, E. Bair, G. Hooker and V. S. Pande, *Phys. Rev. Lett.*, 2003, **91**, 140601.
58 P. Procacci, *J. Chem. Phys.*, 2015, **142**, 154117.
59 T. Beutler, A. Mark, R. van Schaik, P. Gerber and W. van Gunsteren, *Chem. Phys. Lett.*, 1994, **222**, 5229–539.
60 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, *Proceeding of the International Conference on High Performance Computing & Simulation*, Institute of Electrical and Electronics Engineers (IEEE), 2014, pp. 1030–1033.
61 M. R. Shirts, E. Bair, G. Hooker and V. S. Pande, *Phys. Rev. Lett.*, 2003, **91**, 140601.
62 L. S. Dodda, I. Cabeza de Vaca, J. Tirado-Rives and W. L. Jorgensen, *Nucleic Acids Res.*, 2017, **45**, W331–W336.
63 T. W. Anderson and D. A. Darling, *J. Am. Stat. Ass.*, 1954, **49**, 765–769.
64 P. Procacci, S. Marsili, A. Barducci, G. F. Signorini and R. Chelli, *J. Chem. Phys.*, 2006, **125**, 164101.
65 R. Chelli and P. Procacci, *Phys. Chem. Chem. Phys.*, 2009, **11**, 1152.
66 *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).