PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

II. Dissociation free energies in drug-receptor systems via non equilibrium alchemical simulations: application to the FK506-related immunophilin ligands.†

Francesca Nerattini,[b] Riccardo Chelli,[a] and Piero Procacci,*[a]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

The recently proposed fast switching double annihilation (FS-DAM), [Cardelli *et al.*, J. Chem. Theory Comput., 2015, **11**, 423] is aimed at computing the absolute standard dissociation free energies for the chemical equilibrium $\mathrm{RL} \rightleftharpoons \mathrm{R}+\mathrm{L}$ occurring in solution through molecular dynamics (MD) simulations at the atomistic level. The technique is based on the production of fast nonequilibrium annihilation trajectories of one of the species (the ligand) in the solvated RL complex and in the bulk solvent. As detailed in the companion theoretical paper, the free energies of these two nonequilibrium annihilation processes are recovered by using an unbiased unidirectional estimate derived from the Crooks's theorem exploiting the inherent Gaussian nature of the annihilation work. The FS-DAM technique was successfully applied to the evaluation of the dissociation free energy of the complexes of Zn(II) cation with an inhibitor of the Tumor Necrosis Factor α converting enzyme. Here we apply the technique to a real drug-receptor system, by satisfactorily reproducing the experimental dissociation free energies of FK506-related bulky ligands towards the native FKBP12 enzyme and by predicting the dissociation constants for the same ligands towards the mutant I56D. The effect of such mutation on the binding affinity of FK506-related ligands is relevant for assessing the thermodynamic forces regulating molecular recognition in FKBP12 inhibition.

1 Introduction

The determination of the binding free energy in ligand-receptor systems is the cornerstone of drug discovery. From a computational standpoint, the identification of small molecules as effective inhibitors for pharmaceutical relevant biological targets in the industrial environment is generally tackled using molecular docking techniques.[1,2] These methods do not actually evaluate directly the standard binding free energy of a ligand-receptor pair. Rather, they compute related quantities, the so-called scoring functions, relying on concepts such as minimal energy conformations, solvent accessible surface, knowledge-based potentials etc.. Scoring functions in molecular docking are computationally inexpensive but they can be quite involved as they need to surrogate in an empirical way many important effects in non covalent bonding that are not directly included in the calculations, like ligand and protein flexibility, reorganization energy upon binding, microsolvation effects, etc.. The computational efficiency of modern docking approaches, allowing to score thousands of rigid ligand poses on a typical protein target in few hours on a single CPU[3], is counterbalanced by the chronic lack of predictive power of the associated scoring functions that are often unable to discriminate between ligands of nanomolar, micromolar or millimolar affinity.[4,5] The pathologically large production of *false positive* hits in virtual screening techniques based on molecular docking, by letting many unfit candidates to proceed further in the drug discovery pipeline, greatly expands the cost of failures, with a direct negative economic impact in the overall drug discovery processes.[6,7]

In the last decades, in the framework of atomistic molecular dynamics simulations, several computational methods have been devised for rigorously determining the absolute binding free energy in drug receptor systems. Most of these methodologies are based on the so-called alchemical route[8] whereby, the interactions of the ligand with the environment, bulk (explicit) solvent or solvated bound state, are progressively switched off in the Hamiltonian of the system, using a λ coupling parameter varying in the interval [0,1] such that $H_{\lambda=1}$ and $H_{\lambda=0}$ correspond to

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

† Electronic Supplementary Information (ESI) available: Dissociation Free energy in systems with two competing binding poses. Force field parameters for the FK506 ligand. Preparation of the starting structures of the FKBP12 complexes. EDU-HREM scaling protocol for the FKBP12 bound states. Binding pattern for the FKBP12 (native and I56D) bound states. The FK506-FKBP12(I56D) system. See DOI: 10.1039/b000000x/

the Hamiltonians of the fully coupled and fully decoupled states, respectively. The free energy of ligand decoupling, ΔG_d, is obtained by discretizing the alchemical path [0,1] in a number of intermediate unphysical states, for which $1 < \lambda < 0$, running for each of these states *equilibrium* simulations. $\Delta G_d = G_{\lambda=0} - G_{\lambda=1}$ is finally recovered as a sum of the contributions from each of the λ-windows by applying the free energy perturbation[9] (FEP) or thermodynamic integration[10] (TI) techniques. The absolute dissociation free energy for the ligand-receptor pair corresponds to the difference between the two decoupling free energies of the bound and free ligand, except for a standard state correction (SSC).[11–13] There are many clever variants[14–24] of the reversible alchemical route to determine the binding free energies, most of them implementing λ-hopping schemes in the context of Generalized Ensemble (GE) simulations so as to favor conformational mixing at any intermediate λ state. It follows that in reversible techniques, a substantial part of the CPU time must be invested in equilibration.[14,25–27] Finally, all these alchemical methodologies, being rooted on the equilibrium assumption for all nonphysical intermediates connecting the bound and unbound thermodynamic states, must tackle the challenge of minimizing the free energy variance, that is, of choosing the λ alchemical protocol so that the total uncertainty for the transformation is the one which has an equal contribution across every point along the alchemical path.[28]

Recently Sandberg *et al.*[29] presented a new efficient nonequilibrium (NE) technique, the Fast Switching Double Annihilation Method (FS-DAM), aimed at computing the binding free energy in drug-receptor systems. FS-DAM is based on the production of hundreds of fast switching alchemical simulations[30–32] with a continuous dynamical evolution of the externally driven alchemical coordinate λ, completing the decoupling of the ligand in a matter of *few tens of picoseconds* rather than nanoseconds as in reversible alchemical simulations. The theory of NE in non covalent binding have been rigorously formalized in Ref.[33]. In that contribution, it is shown that the absolute binding free energy can be obtained from the annihilation work distributions resulting from the NE trajectories, by applying a unidirectional free energy estimate on the assumption that any observed work distribution is given by a mixture of Gaussian distributions,[13] whose normal components are identical in either direction of the nonequilibrium process, with weights regulated by the Crooks theorem.[34] The accuracy in FS-DAM free energy computation relies on the correct canonical sampling of the *initial fully coupled state alone* (or, more precisely, of the free energy basin referring to the principal pose of the ligand) and on the resolution of the work distribution, both ultimately depending on the number of independent NE trajectories and on the spread of the distribution. In FS-DAM, the sampling issue at intermediate λ state is therefore eliminated altogether allowing very precise decoupling free energy evaluations due to the small spread observed in the fast annihilation work of small molecules in condensed phases in standard conditions.[31,32]

In this study, we apply FS-DAM to the calculation of absolute binding free energies of FK506-related ligands of the peptidyl-prolyl-isomerase protein (FKBP12 or FK506 binding protein)[35], a system that has been investigated thoroughly and extensively in the past years.[27,36–38] The FS-DAM approach is combined with HREM simulations[15] aimed at identifying either the relevant ligand poses in the bound state or the main rotameric state of the free ligand in bulk solvent at the starting equilibrium configurations of fully coupled states. FS-DAM is shown to reproduce satisfactorily (within less that 1 kcal mol^{-1}) the experimental binding free energies between FKBP12 and three bulky and conformationally complex FK506-related ligands (including Tacrolimus). The calculations have been performed with parallel runs on 512 cores lasting few wall clock time hours, thus outperforming the reversible route approach on the same system (massively parallel computation of absolute binding free energy with well-equilibrated states[39]). We also investigated the effect of I56D mutation on the binding free energies of the same ligands in order to shed further light on the thermodynamic forces driving molecular recognition in immunophilins. The I56D replacement is in fact one of the most destabilizing tested mutations for the FKBP12 protein, yielding a decrease in the denaturation free energy of 3.2 kcal mol^{-1} [40] while the backbone NH moiety of Ile56 residue is consistently involved in the binding in all known co-crystal structures of FK506-related ligands-FKBP12 complexes.[35,41,42] Moreover, the binding affinities of FK506-related ligands for the I56D mutant are, to our knowledge, not yet available, exposing our FS-DAM predictions to experimental verification.

The paper is organized as follows. In Section "Theory", we briefly recall the theory of the NE alchemical approach. In Section "Methods", we first describe the implementation of an *ad-hoc* Hamiltonian replica exchange scheme[15] for identifying the relevant poses in the bound states of the complexes. We then detail the methodology for producing the FS-DAM trajectories on a parallel platform and for extracting unbiased unidirectional estimates from the corresponding annihilation work distributions. In section "Results", we present and discuss the absolute binding free energies of the FK506, N-Elte378[36] and SB3[35] ligands for the native FKBP12 protein and for the I56D mutant. In Section "Conclusions", we envisage perspectives and future directions of non equilibrium techniques focusing on a possible role for massively parallel FS-DAM in automated industrial virtual screening and ADME-Tox assessment in the drug discovery process.

2 Theory

The NE theory of alchemical transformations has been discussed thoroughly in Ref.[33]. Here we recall the essential concepts of the NE approach, referring to the paper[33] for a more in-depth explanation.

2.1 nonequilibrium alchemical transformations

Given two thermodynamic states - the ligand fully coupled and decoupled states, $\lambda = 1$, $\lambda = 0$, respectively - and an alchemical path connecting these states, one can generate many fast NE decoupling trajectories starting from a canonical sampling in $\lambda = 1$ and rapidly varying the λ parameter from 1 to 0 with a time protocol $\mathscr{T}_\tau$. From the work done in the NE decoupling trajectories, a forward work distribution $P(W_{1\to0})$ is constructed. The com-

mon switching off protocol $\mathscr{T}_\tau$ can be chosen such that, in the final NE states, the decoupled ligand can be located anywhere in the MD box with a random orientation.[33] The reverse process starts from a canonical sampling of the decoupled (gas-phase) ligand and $\lambda = 0$, producing NE trajectories done with an inverted time protocol $\overline{\mathscr{T}_\tau}$. These reverse trajectories end up in the final fully coupled NE state $\lambda = 1$, thus producing a reverse work distribution $P(-W_{0\to1})$. The forward and reverse work distribution obeys the Crooks NE work theorem[34]

$$\frac{P(W_{1\to0})}{P(-W_{0\to1})} = e^{\beta(W_{1\to0}-\Delta G)} \tag{1}$$

where ΔG is the decoupling free energy. The annihilation process of the ligand in the bulk solvent can be easily inverted[32] yielding a solvation free energy ΔG_s that can be straightforwardly obtained with bidirectional experiments using the Bennett acceptance ratio.[43,44] Alternatively, by exploiting the Markovian nature of the fast annihilation process[30–32], one can apply a Gaussian unidirectional estimate $\Delta G_s = \langle W^{(s)}_{1\to0}\rangle - \frac{1}{2}\beta\sigma_s^2$ to the forward/annihilation distribution.[13,29,31,32] When dealing with the annihilation of the unrestrained ligand in the complex in a MD box of volume $V_{\rm box} \gg V_{\rm site}$ (where $V_{\rm site}$ is the binding site volume in the receptor), the inversion of the process becomes more elusive. For tight binding ligands, if one uses few hundreds canonically sampled equilibrium starting phase-space points, these equilibrium microstates must basically all correspond to the bound state. A "tight binding ligand" has in fact a dissociation free energy ΔG_0 that exceeds 10 kcal mol^{-1}, yielding a sub-micromolar dissociation constant. In a MD box for a typical ligand receptor pair with volume of the order of $100 \div 500$ times V_0 (the standard state molecular volume), the probability ratio between unbound and bound state P_u/P_b at equilibrium is of the order of 10^{-3} or less.[45] Hence, for a ligand with $\Delta G_0 > 10$ kcal mol^{-1}, a canonical sampling of hundreds microstates should include only configurations belonging to the *bound state*. We assume, for the moment, that only one possible binding pose/basin exists in the binding site of volume $V_{\rm site}$. If we start from the bound state, we obtain an approximately Gaussian distribution for the annihilation work of the NE trajectories such that we may define a decoupling free energy $\Delta G_b = \langle W^{(b)}_{1\to0}\rangle - \frac{1}{2}\beta\sigma_b^2$. This quantity, provided that $V_{\rm box} \gg V_{\rm site}$, is clearly independent on $V_{\rm box}$. In other words, if we use the same sampling configurations for the fully coupled bound state, then we should always obtain the same $\langle W^{(b)}_{1\to0}\rangle$ and variance σ_b^2 and hence the same ΔG_b no matter how large is made the MD box. In the reverse process, however, the decoupled ligand should be brought to life *in presence of the solvated receptor* starting from a gas-phase random position/orientation in the MD box, thus ending up either in the bound state or in an unbound state. The probability of sampling the ligand in the bound state is given by $V_{\rm site}/V_{\rm box}$. So the dependence of the decoupling free energy on the volume $V_{\rm box}$ can be appreciated only in the reverse process, where the low free energy state corresponding to the unbound ligand in presence of the receptor gets exponentially amplified due to the mathematical structure of the Crooks NE work theorem. If the MD box is made large at will, the fast growth of the ligand in presence of the receptor will tend to generate a volume independent reverse distribution Gaussian $P(-W_{0\to1})$ of half maximum width $2\sigma_s$ and mean $-\langle W^{(s)}_{0\to1}\rangle\rangle$ as if the receptor were not present, hence basically identical to that obtained in the fast growth of the ligand in the pure solvent, corresponding to a hydration free energy of $-\Delta G_s = \langle W^s_{0\to1}\rangle - \frac{1}{2}\beta\sigma_s^2$. Based on the above, we assume that, for standard MD box volumes, the forward distribution is given by two volume independent components, a principal one related to $\Delta G_b = \langle W^{(b)}_{1\to0}\rangle - \frac{1}{2}\beta\sigma_b^2$ and a second shadow component[13] related to $\Delta G_s = \langle W^{(s)}_{1\to0}\rangle - \frac{1}{2}\beta\sigma_s^2$, i.e.

$$P(W_{1\to0}) = c_b N_b(W) + c_s N_s(W) \tag{2}$$

where $N_b(W)$ and $N_s(W)$ are the normal distributions of half maximum width $2\sigma_{b/s}$ centered at the volume independent mean work $\langle W^{(b)}_{0\to1}\rangle\rangle$ and $\langle W^{(s)}_{0\to1}\rangle\rangle$, respectively. Because of Eq. 1 applied to Gaussian mixtures, the forward distribution Eq. 2 must cross[13] the reverse distribution at the volume dependent free energy

$$\begin{aligned}\Delta G_{\rm box} &= -k_BT\ln(c_b e^{-\beta\Delta G_b} + c_s e^{-\beta\Delta G_s}) \\ &= -k_BT\ln\left[c_b e^{-\beta\Delta G_b}\left(1+\frac{c_s}{c_b}e^{-\beta(\Delta G_s-\Delta G_b)}\right)\right]\end{aligned} \tag{3}$$

According to the equilibrium theory of non covalent bonding[12,26,29,33,45], the volume dependent dissociation free energy is given by

$$\Delta G_{\rm sim} = \Delta G_0 - k_BT\ln(V_{\rm box}/V_0) \tag{4}$$

where V_0 and ΔG_0 are the standard state volume and dissociation free energy and where $\Delta G_0 = \Delta G_{0b} - \Delta G_s$ with ΔG_{0b} being the annihilation free energy of the ligand in the complex, referred to the standard state volume. As shown in Ref.[33], ΔG_b, referred to a volume $V_{\rm site}$, is connected to ΔG_{0b}, referred to the standard state volume of $V_0 = 1661$ Å^3, as $\Delta G_{0b} = \Delta G_b - k_BT\ln(V_0/V_{\rm site})$ so that Eq. 4 can be rewritten as

$$\Delta G_{\rm sim} = \Delta G_b - \Delta G_s - k_BT\ln(V_{\rm box}/V_{\rm site}) \tag{5}$$

Note that, according to Eq. 5, for infinite dilution the volume dependent dissociation free energy $\Delta G_{\rm sim}$ gets infinitely negative, meaning that the probability for a bound state goes to zero in this limit.[46,47] In NE theory of non covalent bonding, the Crooks theorem, as previously seen, imposes a different limit at infinity dilution, namely $\lim_{V_{\rm box}\to\infty}\Delta G_{\rm box} = \Delta G_s$. Let us hence see how the Crooks-based Eq. 3 and the equilibrium relation Eq. 5 can be reconciled. In Eq. 5, the volume dependent quantity $\Delta G_{\rm sim}$ may be viewed[45] as $-k_BT\ln(P_u/P_b)$ where P_b and P_u are the probability of finding the ligand-receptor pair in the bound and unbound states, respectively, at the volume $V_{\rm box}$ at equilibrium. So the equilibrium ratio between unbound and bound states in the MD volume $V_{\rm box}$ must be equal to the ratio of the weights c_s/c_b of the two normal components in Eq. 2

$$\frac{c_s}{c_b} = e^{-\beta[(\Delta G_b-\Delta G_s)-k_BT\ln(V_{\rm box}/V_{\rm site})]} \tag{6}$$

Eq. 6 is valid for $V_{\rm box}$ large at will. If $V_{\rm box}$ is chosen such that $k_BT\ln(V_{\rm box}/V_{\rm site}) = \Delta G_b - \Delta G_s$, then the probability for the un-

bound state P_u at equilibrium equals that of the bound state P_b and, correspondingly, $c_s/c_b = 1$. For a ligand with $\Delta G_0 = 10$ kcal mol^{-1}, such volume would be approximately $V_{\rm box} = V_0 \times 10^7$. Given that $c_s + c_b = 1$ due to the normalization condition for the probability $P(W_{1\to 0})$, we then have that

$$c_b = \frac{1}{1 + e^{-\beta[(\Delta G_b - \Delta G_s) - k_B T \ln(V_{\rm box}/V_{\rm site})]}} \tag{7}$$

Using Eq. 7 and 6, Eq. 3, after some trivial algebra, is transformed as

$$\Delta G_{\rm box} = \Delta G_s + k_B T \ln\left(\frac{e^{\beta(\Delta G_b - \Delta G_s)} + \frac{V_{\rm box}}{V_{\rm site}}}{1 + \frac{V_{\rm box}}{V_{\rm site}}}\right) \tag{8}$$

Eq. 8 is a nonequilibrium reformulation, based on the Crooks NE work theorem Eqs. 1 and 3, of the equilibrium Eq. 5. According to Eq. 8, when $V_{\rm box} \to \infty$, the volume dependent crossing point of the forward and reverse distribution for the decoupling of the ligand in the bound state is equal to ΔG_s. In MD simulations in explicit solvent, usually one has $V_{\rm box}/V_{\rm site} = 100 \div 1000$ so that we can neglect the 1 in the denominator on the rhs of Eq. 8. For tight binding nanomolar ligand, we have that $e^{\beta(\Delta G_b - \Delta G_s)} \simeq 10^8 \div 10^9$ (where we have tacitly assumed that $V_{\rm site} \simeq V_0$). Thus, we can neglect the term $V_{\rm box}/V_{\rm site}$ in the numerator on the rhs of Eq. 8, finally obtaining

$$\Delta G_{\rm box} = \Delta G_b - k_B T \ln \frac{V_{\rm box}}{V_{\rm site}} \tag{9}$$

Subtracting ΔG_s on both sides of Eq. 9 and considering that $\Delta G_{\rm sim} = \Delta G_{\rm box} - \Delta G_s$, we recover Eq. 5 . For what we have been saying, the quantity

$$\begin{aligned} \Delta G_0 &= \Delta G_b - k_B T \ln \frac{V_0}{V_{\rm site}} - \Delta G_s \\ &= \langle W_{1\to 0}^{(b)} \rangle - \langle W_{1\to 0}^{(s)} \rangle - \frac{1}{2}\beta(\sigma_b^2 - \sigma_s^2) - k_B T \ln \frac{V_0}{V_{\rm site}} \end{aligned} \tag{10}$$

provides an unbiased unidirectional estimates of the dissociation free energy in NE experiments. The variance in $\langle W \rangle$ and σ for normally distributed samples follows the ancillary t-statistics[48] and is in both cases proportional to $\sigma(\tau)/(N_\tau)^{1/2}$, where $\sigma(\tau)$ is the τ-dependent spread of the distribution and N_τ the number of NE trajectories all done with the same time protocol $\mathscr{T}_\tau$. So, if σ is of the order of few kcal mol $^{-1}$, *only few hundreds trajectories are needed to get an error on the free energy below 1 kcal mol*$^{-1}$. Unlike in reversible alchemical transformations, in the NE approach the overall error can be very naturally and reliably computed via standard block-bootstrapping from the collection of N_τ works. Moreover, reducing N_τ by a factor of G amplifies the error only by $G^{1/2}$ making Gaussian based estimates robust and reliable even with a very small number of sampling trajectories.[13]

The unidirectional estimate Eq. 10 for the dissociation free energy (with $V_{\rm site}$ as reference volume) is based on the Gaussian assumption of the underlying distributions. It may occur that the apparent annihilation work distribution $N_b(W)$ (or $N_s(W)$ when annihilating the free ligand in the bulk) is not Gaussian as verified by the standard Kolmogorov-Smirnov tests or by the evaluation of cumulants of order higher than two.[29] In that case, we shall assume that the distribution is given by a mixture of normal distributions due to, e.g., competing binding poses whose relative weights are again regulated by the Crooks NE work theorem.[13] A derivation of the corresponding equation for the two-binding poses analog of Eq. 9 is given in the Supporting Information.

2.2 Canonical sampling of the $\lambda = 1$ fully coupled state

In FS-DAM, the issue of the equilibrium sampling of the starting $\lambda = 1$ fully coupled states is a indeed central one. It has been shown in recent studies[29,32] that hundreds of independent NE decoupling trajectories are sufficient for acquiring a reliable unidirectional estimate of the dissociation free energy ΔG_0 based on Eq. 10. These fast NE trajectories must be originated from phase space points at $H_{\lambda=1}$ that are representative of the Boltzmann corresponding distribution function $e^{-\beta H_{\lambda=1}}/Z_{\lambda=1}$ for the complex and for the ligand in bulk. When flexible ligands are simulated in bulk solvent, it might be hard, especially when dealing with ligand conformations separated by torsional barriers several $k_B T$ high, to *canonically* sample all accessible conformations using a conventional MD approach. In that case, it is more appropriate and effective to use Hamiltonian Replica Exchange approach with ligand torsional tempering.[49] According to a solute tempering scheme tested and refined in several past studies,[37,50,51] few nanoseconds of GE simulation are sufficient to acquire a complete picture of the conformational landscape of flexible drug size molecules (including small peptides[52]) in aqueous solvent in standard conditions. The few hundreds ligand conformations selected for the subsequent FS-DAM annihilation must be thus chosen according to their canonical weight as emerged from the H-REM simulation.

Let's now come to the canonical sampling of the bound state. We have seen in the preceding section that, for tight binding ligands, all the starting conformations of the system that includes the ligand *and* the receptor, must correspond to conformations of the complex. In order to sample canonically these bound states, that may include also multiple poses of the ligand in the binding site, we can use the recently developed Energy Driven Undocking scheme (EDU-HREM), where the drug-binding site interaction is progressively scaled along a replica progression in the context of a GE simulation, thus favoring unbinding events in the scaled replicas. In EDU-HREM,[15] a weak tethering potential between the drug and the receptor prevents the excessive wandering of the detached ligand into the bulk in the scaled (hot) replicas. EDU-HREM can be in principle be used directly to evaluate the dissociation free energy[15]

$$\Delta G_0 = -k_B T \ln(p_u/p_b) + k_b T \ln(V_{\rm eff}/V_0) \tag{11}$$

where $V_{\rm eff}$ is the effective volume available to the ligand imposed by the weak tethering potential and p_u and p_b are the probabilities of the unbound and bound state calculated using the whole GE statistics re-weighted on the target (unscaled) state via the multiple Bennett acceptance ratio (MBAR).[53,54] The EDU-HREM, unfortunately, has been proven inadequate for tight binding ligands,[55] as unbound states for strong binders occurs frequently

only at highly scaled replicas with negligibly small MBAR weights and very rarely at low replicas with higher MBAR factors. Therefore, a meaningful statistics for applying Eq. 11 can be collected only after unbearably long GE simulations. EDU-HREM, however, can be effectively used for a much simpler scope, i.e. that of sampling the conformations of the ligand in the bound state alone (a minor part of the accessible phase space for the system ligand and receptor), using the statistics collected in the target state only of the GE simulation and correcting, when necessary, for the biasing tethering potential.

3 Methods

We present dissociation free energy calculations for three tight binding FK506-related ligands of the peptidyl-prolyl (PPI) isomerase FKBP12 protein using the FS-DAM approach described in the preceding section and in Ref.[33] The selected ligands are the natural ligand FK506 (Tacrolimus), and two powerful synthetic ligands, SB3[35] and N-Elte378[51] (chemical structures are reported in in Figure 1). The dissociation constants of these three tight binding ligands are evaluated for the native FKBP12 protein and for the I56D mutant. Hence a total of six dug-receptor pairs are evaluated: SB3-FKBP12, SB3-FKBP12(I56D), N-Elte378-FKBP12, N-Elte378-FKBP12(I56D), FK506-FKBP12, FK506-FKBP12(I56D).

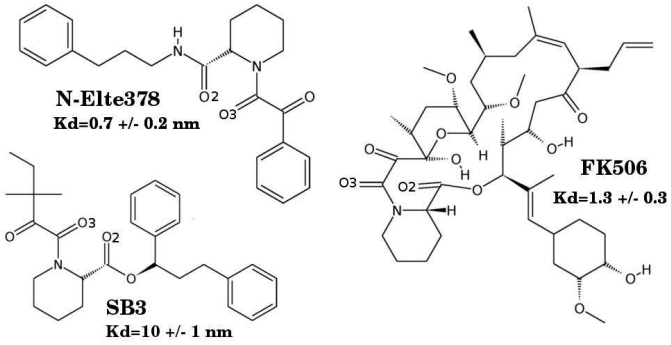

Fig. 1 Chemical structures of the two synthetic SB3 and N-Elte378 FKBP12 ligands and of the FKBP12 natural ligand, FK506. The dissociation constants of FK506 and N-Elte378 with respect to native FKBP12 are taken from Ref.[51]. The dissociation constant of SB3 is taken from Ref.[35].

3.1 Force field parameterization

The force field used for the protein system is based on the AMBER99sb parameter set.[56] For the SB3 ligand the force field parameters were taken from Ref.[50], and for N-Elte378 (neutral form) the potential parameters and atomic charges are those reported in Ref.[51]. For the FK506 ligand, the parameters were assigned according to the Generalized Amber Force Field (GAFF[57]). Atomic charges on FK506 atoms are computed on the AM1[58] optimized structure using the B3-LYP exchange-correlation functional[59,60] and the 6-31Gd split valence basis set, modeling solvation effects with a Polarizable Continuum Model[61] (PCM) with water as solvent, and evaluating a gridded electrostatic potential (ESP) according to the Merz-Singh-Kollman scheme[62] by single-point calculations. Equivalent ESP charges have been symmetrized. All *ab initio* calculations are done using the Gaussian program.[63] Atomic ESP charges and GAFF types for FK506 are reported in the Supporting Information. The TIP3P model[64] is adopted for the water solvent with standard Lorentz-Berthelot mixing rules for solute-solvent Lennard-Jones interactions.

3.2 Simulation details

All MD calculations were performed in the isothermal-isobaric ensemble (NPT). Constant pressure was enforced isotropically using a modification of the Parrinello-Rahman Lagrangian[65], and temperature control was achieved using the Nosé-Hoover thermostat[66,67]. Electrostatic interactions were computed using the smooth particle mesh Ewald algorithm with the convergence parameter set to 0.43 $Å^{-1}$ and grid spacing of $\simeq 1$ Å[68]. The equations of motion were integrated using a multiple time-step r-RESPA scheme[69] with a potential subdivision specifically tuned for bio-molecular systems in the NPT ensemble.[65,70]. Details of the integrator are provided in Ref.[70]. Distance constraints were enforced only for bonds involving a hydrogen atom (X-H bonds). All MD calculations, H-REM equilibrium, EDU-HREM and FS-DAM, were done with the ORAC code.[49,70]

Starting configurations for the ligands in bulk: the H-REM equilibrium simulations for the FK506, SB3 and N-Elte378 in bulk were done in cubic boxes containing the ligand and 678, 702 and 706 TIP3P water molecules, yielding a volume, at T=300 K and P=1 atm, of about 22.1 $\pm$ 0.4 nm^3 and lasted in all cases 5 ns. As done in a previous study[51], we scaled only the intraligand potential (torsional and non bonded interactions) using 8 replicas, with maximum and minimum scaling factors of 1.0 and 0.15 corresponding to a solute temperature ranging form 300 K to 2000 K. Configurations for the subsequent FS-DAM step were saved in the target replica at regular intervals every 9 ps, collecting a total of 512 initial fully coupled states. Further details on the H-REM protocol for the simulation of free ligands in bulk solvent are given in the Supporting Information of Ref.[51].

Starting configurations for the bound states: the starting configuration for the FKBP12 complexes were prepared as described in the Supporting Information. In the Figure **??**, we show as an example, a representation of the FKBP12-FK506 complex. Each equilibrium GE simulation involves 16 replicas where only the ligand-binding site interactions are scaled according to a solute tempering approach[49,72]. The "solute" involves the ligand and 15 surrounding protein residues. The scaling factors along the replica progression apply to the intrasolute torsional and non bonded potential and follow a standard geometrical progression.[73] Further details on the EDU-HREM scaling protocol for the FKBP12 complexes are given in the Supporting Information. EDU-HREM GE simulations, lasting in all cases 5 nanoseconds, are performed in the NPT ensemble (T=300 K and P=1 atm) on the native and mutated forms of the FKBP12 complexes. In each EDU-HREM simulation, a weak tethering potential (see Supporting Information for details) was imposed to prevent excessive drifting of the ligand off the binding site in the "hot" replicas. A total of 512 bound states configurations were acquired at regu-

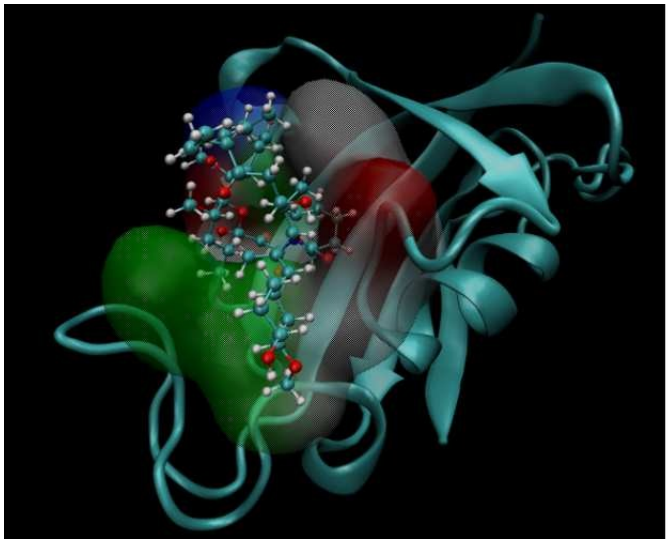

Fig. 2 Ribbon representation of the native FKBP12 protein (PDB code 1FKJ)[71] in complex with the FK506 ligand (ball and stick representation). The binding pocket is shown as a surface representation with a color coded scheme for residue type (grey: hydrophobic, green:polar, red: acidic, blue:basic)

lar time intervals using the last 3.5 ns of the target (unscaled) replica.

FS-DAM simulations: the driven alchemical annihilation simulations were performed starting from the $\lambda = 1$ (fully coupled) equilibrium configurations collected in the preceding HREM or EDU-HREM step, according to the scheme described in Ref.[32] The annihilation protocol $\mathscr{T}_\tau$ was common for the ligand in bulk and in the bound state. $\mathscr{T}_\tau$ lasted in all cases a total of 270 ps per NE-trajectory. In the first 120 ps, the electrostatic interactions were linearly switched off. In the following 30 ps, 2/3 the Lennard-Jones potential was turned off and in the last 120 ps, the residual 1/3 was finally switched off with a soft-core Beutler potential[74] regularization as λ is approaching to zero. The cut-off for the Lennard-Jones interaction was set to 13 Å.

4 Results

4.1 Canonical Sampling of the fully coupled states

The EDU-HREM simulations for collecting the starting canonical configurations of the fully coupled complex indicate that in all studied cases a single pose appears to contribute to the binding at equilibrium. As a matter of fact, the 512 initial configurations for the ligand in the bound state, refer in all cases to a very tight single pose involving persistent ligand-residue contacts as it can be inferred from Figure 5 of the Supporting Information. It should be said that, for the mutated system, the equilibrium binding pattern at $\lambda = 1$ is sensibly smoothed off for N-Elte378 and SB3, with contributions coming from various hydrophobic contacts. For the simulation of the free ligands in bulk water, N-Elte378, as discussed in Ref.[51], exhibits a competition between a compact structure, stabilized by stacking interactions between the terminal aromatic moieties of the molecule, and a less stable extended structure (see Figure 7 in ref.[51]). These transient ligand conformations differ by less than 2 kcal mol^{-1}, easily inter-converting in the target state of the H-REM equilibrium simulations. As we shall see later on, the presence of two principal conformational states in free N-Elte378 in bulk will be reflected in the bimodality of the corresponding annihilation work distribution.

4.2 The Gaussian nature of the nonequilibrium annihilation work distribution

Concerning the subsequent FS-DAM stage, we previously stated that, when the annihilation free energy is due to a single primary pose/conformation of the ligand, the annihilation works for the bound state and for the ligand in bulk should be normally distributed (with the exception of N-Elte378 in bulk) at any τ or, equivalently, at any λ alchemical state. The dissociation free energy can be thus recovered from the distribution of the final works ($\tau = 270$ ps) using the unbiased estimator of Eq. 10. In Figure 3 we show, as an example, the time record of the work calculated

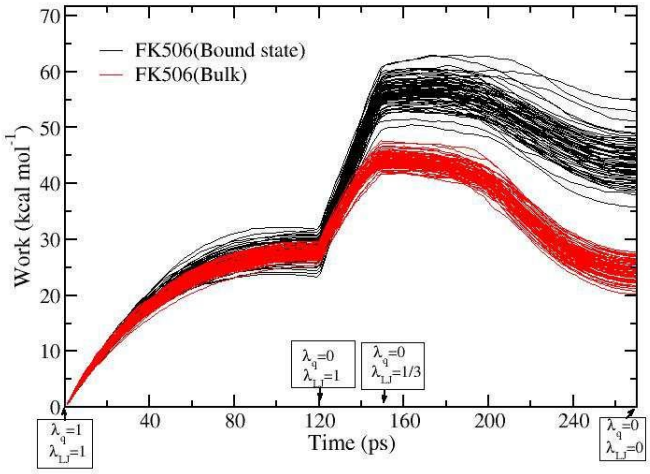

Fig. 3 Work computed in a representative sample of the decoupling trajectories for the FK506 ligand in the bound state (black lines) and in bulk solvent (red lines). The bound state involve native FKBP12. In the boxes, the stage of the annihilation protocol is shown at the indicated simulation time.

during the driven annihilation trajectories of the FK506 ligand in the complex [FKBP12(native)] and in bulk solvent. For clarity, only 72 trajectories out of 512 are shown. During the linear discharging of the ligand, occurring in the first 120 ps, the work monotonically increases for both the ligand in the bound state and the ligand in bulk solvent. In the following 30 ps, while the Lennard-Jones potential is linearly scaled up to a factor of 1/3 of its original value, the energy loss is more pronounced for the bound state. In the last stage, when the Lennard-Jones potential is finally brought to zero and the ligand is fully transferred to the gas-phase, we observe in both cases an energy recovery due to the enthalpic gain of the water progressively refilling the ligand cavity. The recovery is on the average less pronounced for the ligand in the bound state since in this case, water cannot refill optimally the entire cavity due to the interference of the pocket residues.

The patter shown in Figure 3 is common to all ligand-receptor pairs examined in this study.

In order to be able to use the unbiased dissociation free energy estimator of Eq. 10, the final annihilation work values should be normally distributed. In Figure 4, we report the probability density distributions of the final annihilation works along with the corresponding cumulative distribution functions (solid lines) compared to the reference normal counterparts (dashed lines) obtained from the first two moments, $\langle W_{1\to0}\rangle$, σ of the work samples. In order to quantify the distance between the sample work distributions (computed using all 512 works with a bin size of 0.5 kcal mol^{-1}) and the reference normal distributions,[30] we use the Kolmogorov-Smirnov (KS) statistic $D = \sup_x |P_n(W_{1\to0}) - N(\langle W_{1\to0}\rangle, \sigma, W)|$ where $\sup_x$ is the supremum of the set of differences. The *goodness-of-fit* test or the Kolmogorov-Smirnov test is constructed by using the critical values of the Kolmogorov distribution $D_c = c(\alpha)/n^{1/2} =$ at the 1 % of the significant level ($\alpha = 0.01$). The null hypothesis is rejected at this level if $D > D_c = 0.101$. We can see from Figure 4 that the sample appears to be drawn from the reference normal distribution in all cases except for the annihilation of N-Elte378 in bulk solvent. In this case the work distribution function has a marked bimodal character exhibiting two broad peaks centered at $W_{1\to0} \simeq 11$ and 17 kcal mol^{-1}. Such an outcome is a direct consequence of the partitioning of the N-Elte378 512 starting equilibrium conformational configurations as belonging to the extended and compact structural basins, each characterized by its own hydration free energy. We must also remark that for N-Elte378 in bulk, the heights, widths and distances between the two main peaks in the work distribution do not satisfy the "no mixing" hypothesis (see also Section "Dissociation free energy in systems with two competing binding poses" in the Supporting Information) indicating that a significant mutual contamination occurs between the two conformational free energy basins during the ligand annihilation process. As the sampled work values for the N-Elte378 annihilation in bulk did not pass the KS-test for a normal distribution, in this single case out of the nine reported in Figure 4 we use a mixture of two normal components as reference distribution, i.e. $P_{\rm ref}(W_{1\to0}) = c_1 N(W, W_1, \sigma_1) + c_2 N(W, W_2, \sigma_2)$, where the weights c_1, $c_2 \equiv 1 - c_1$ and the moments W_1, W_2, σ_1, σ_2 are fitted according to the procedure described in Ref.[13]. The resulting hydration free energy is

$$\Delta G_S = -k_B T \ln\left[c_1 e^{-(\beta(W_1 - \beta\sigma_1^2/2)} + c_2 e^{-(\beta(W_2 - \beta\sigma_2^2/2)}\right] \qquad (12)$$

4.3 Free energy contributions to binding

In Figure 5, we report the potential of mean force (PMF), $\mathscr{W}_{b/s}(\tau)$, as a function of the decoupling time τ for the annihilation process of the three ligands in FKBP12(native), FKBP12(I56D) and in bulk, calculated assuming, at each τ, a normal work distribution such that $\mathscr{W}_{b/s}(\tau) = \langle W_{0\to1}^{(b/s)}(\tau)\rangle - \frac{1}{2}\beta\sigma^2_{b/s}(\tau))$, with the exception of N-Elte378 in bulk where the PMF is given by

$$\mathscr{W}_s(\tau) = -k_B T \ln\left[c_1(\tau) e^{-\beta(W_1(\tau) - \beta\sigma_1^2(\tau)/2)} + c_2(\tau) e^{-\beta(W_2(\tau) - \beta\sigma_2^2(\tau)/2)}\right]. \qquad (13)$$

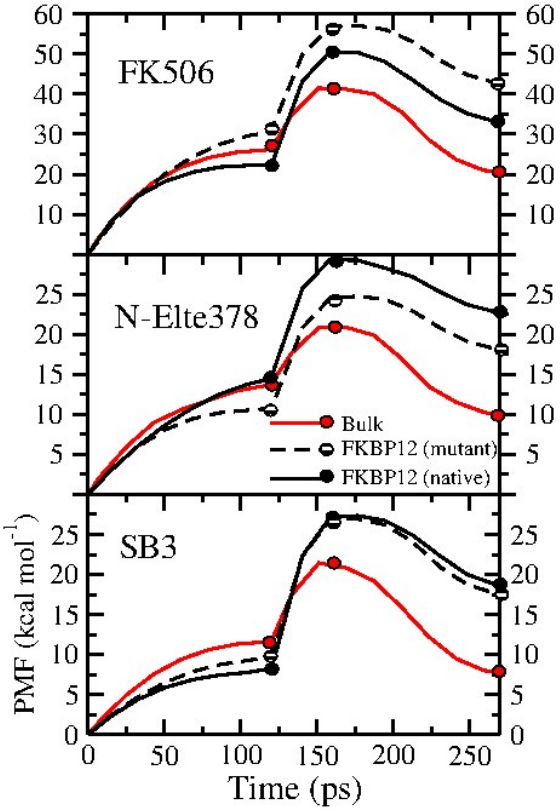

Fig. 5 PMF for the annihilation of the ligand in the bound state (native and I56D mutant) and in bulk. The symbols mark the contributions of the electrostatic, dispersion and cavity free energies to the PMF (see Table 1)

From the final values of the PMF's, the dissociation free energies of the ligands with respect to native FKBP12, $\Delta G^{(n)}$, and the I56D mutant $\Delta G^{(m)}$ can be obtained as $\Delta G^{(n/m)} = [\mathscr{W}_b^{(n/m)}(\tau) - \mathscr{W}_s(\tau))]_{\tau=270\ \rm ps}$, not including a small volume correction discussed further below. By inspection of Figure 5 showing a common trend for the annihilation PMF of the three ligands in bulk or in the protein, the dissociation free energies of each FKBP12 binder can be roughly separated into three contributions, namely i) that due to the electrostatic free energy, $\Delta G_{\rm el} = [\mathscr{W}_b(\tau) - \mathscr{W}_s(\tau)]_{\tau=120\ \rm ps}$ obtained by discharging the ligand, ii) the contribution due to the Lennard-Jones dispersive interactions, $\Delta G_{\rm disp} = [\mathscr{W}_b(\tau) - \mathscr{W}_s(\tau)]_{\tau=150\ \rm ps} - \Delta G_{\rm el}$, and finally the contribution related to the annihilation of the cavity controlled by the repulsive part of the Lennard-Jones potential $\Delta G_{\rm cav} = [\mathscr{W}_b(\tau) - \mathscr{W}_s(\tau)]_{\tau=270\ \rm ps} - \Delta G_{\rm disp} - \Delta G_{\rm el}$. In Table 1, we report these contributions for the three FK506-related ligands-FKBP12 (native and I56D mutant) complexes. We note that for the complexes formed by the native FKBP12 protein, the electrostatic contribution is strongly destabilizing for the SB3 and FK506 ligands, while it moderately favors binding in the case of N-Elte378. For the mutant, the electrostatic term, while being still destabilizing the complex SB3-FKBP12(I56D), gets in general more favorable even becoming an important stabilizing contribution in the case of FK506-FKBP12(I56D) complex. The electrostatic energy gain in the FK506-FKBP12(I56D) complex is very likely corre-

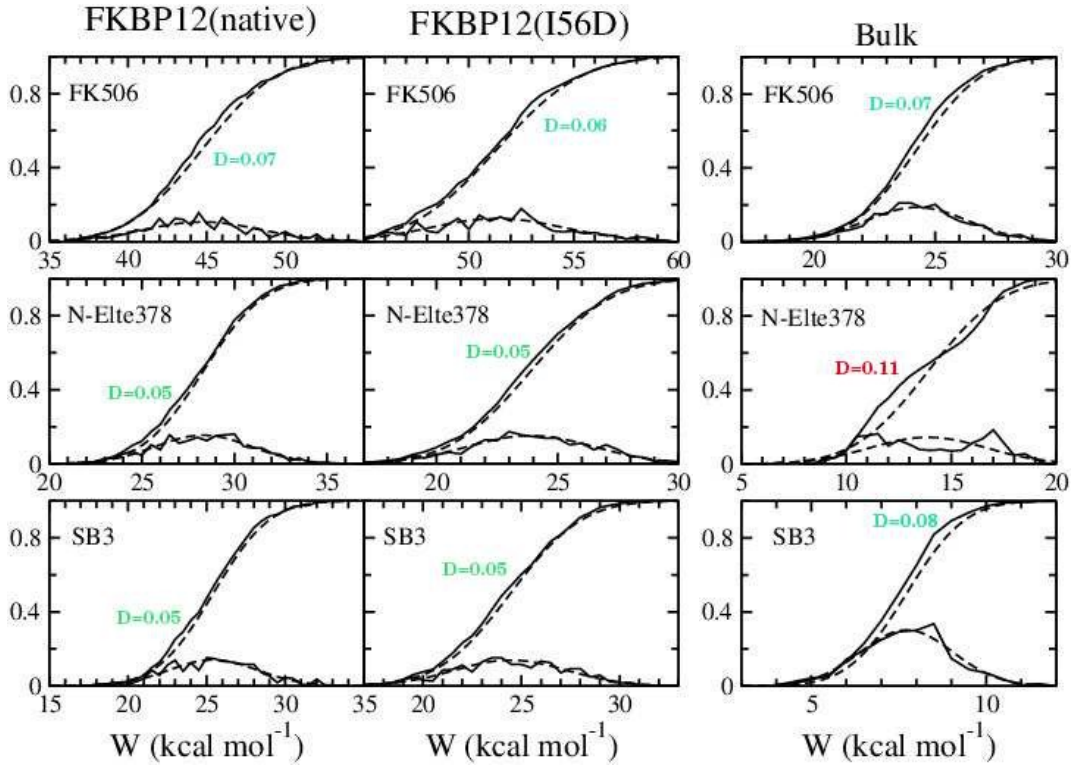

Fig. 4 Kolmogorov-Smirnov (KS) test on the cumulative distribution of the annihilation work, assuming only one Gaussian component (null hypothesis). The plots in the left panel refers to KS-test for the annihilation of the ligand in the complexes. The plots in the right panel refers to the KS-test for the annihilation of the ligand in bulk. The critical value for rejecting the null hypothesis is set at $D_c = 0.101$ at the $\alpha = 0.01$ level.

lated to the enhanced contribution due to $\Delta G_{\rm cav}$ that increase by 3.5 kcal mol^{-1} upon mutation. The dispersive and cavity contributions, $\Delta G_{\rm disp}$ and $\Delta G_{\rm cav}$, strongly stabilize the complex in all cases. As noted in previous studies,[75] the dispersive free energy is the largest contribution to the overall binding affinity in all cases, with only a moderate and consistent attenuation upon mutation. This result appears to be in substantial agreement with recent extensive site directed mutagenesis studies[76], where the activity of various FKBP12 mutants (not including Ile56, unfortunately), involving aminoacids in the neighborhood of the binding pocket, was measured using a peptide analog by the standard chymotrypsin digestion assay[77] and then compared with wild-type FKBP12. In that study, it was found that site-specific interactions by the side chains of amino acid residues constituting the substrate-binding cavity were not essential for the PPIase activity, hence suggesting that the binding site for PPIase activity has a marked hydrophobic character (see Figure 2).

At variance with the other two ligands SB3 and FK506, in the case of the N-Elte378 complex, as far as $\Delta G_{\rm el}$ is concerned, the I56D mutation appears to strongly destabilize the binding free energy by nearly 4 kcal mol^{-1}. This was somewhat expected given that N-Elte378 was engineered[51] for an optimal mean exposure of the O2 and O3 oxo moieties in order to accept two simultaneous H-bonds involving the HO in Tyr82 and the HN of the backbone of Ile56 (See Figure 2 of the Supporting Information). When the latter residue is mutated to Asp, the excess negative charge on Asp repels the two N-Elte378 H-bond oxo groups acceptors so that the O2-HN(Asp56) H-bond is basically suppressed and the O3-HO(tyr82) significantly weakened as shown by the corresponding distance distribution function reported in Figure 5 of the SI, with consequent loss of ligand-protein electrostatic energy. Such electrostatic energy loss, as Table 1 shows, is responsible for most of the decline of the N-Elte378 affinity upon mutation.

Ligand	$\Delta G_{\rm el}$		$\Delta G_{\rm disp}$		$\Delta G_{\rm cav}$	
	Native	I56D	Native	I56D	Native	I56D
FK506	-4.1	4.3	13.2	11.0	3.5	7.0
N-Elte378	0.7	-3.0	7.7	6.6	4.5	4.7
SB3	-3.2	-1.7	9.5	7.8	6.1	5.2

Table 1 Electrostatic, dispersive and cavity(repulsive) contributions to the dissociation free energy of FK506-related ligands based on the PMF reported in Figure 5. For the definition of $\Delta G_{\rm el}$, $\Delta G_{\rm disp}$ and $\Delta G_{\rm cav}$ see text. All units are in kcal mol^{-1}. The dissociation free energy, except for a small volume term, corresponds to the sum of $\Delta G_{\rm el}$, $\Delta G_{\rm disp}$ and $\Delta G_{\rm cav}$. A discussion of the errors on the free energies in the context of FS-DAM can be found in the discussion of Table 2.

4.4 FS-DAM dissociation free energies for FK506-related ligands

In Table 2, we report the overall dissociation free energies for the six FKBP12 complexes computed using the unbiased estimator

	ΔG_{b0}	ΔG_s	$\Delta G_0^{\mathrm{FS-DAM}}$	$\Delta G_0^{\mathrm{Exp.}}$
FK506(native)	33.0±0.7	20.4 ±0.1	**12.6** ±0.8	*12.2 ± 0.1*
FK506(I56D)	42.5±0.4	20.4 ±0.1	22.1 ±0.5	-
N-Elte378(native)	22.5±0.1	9.8 ±0.4	**12.7** ±0.5	*12.6* ± 0.1
N-Elte378(I56D)	17.9±0.1	9.8 ±0.4	8.1 ±0.5	-
SB3(native)	18.5±0.3	6.3 ±0.1	**12.2** ±0.4	*11.0* ± 0.1
SB3(I56D)	17.3±0.3	6.3 ±0.1	11.0 ±0.4	-

Table 2 Dissociation free energies of FKBP12 (native and I56D mutant) complexes for some FK506-related ligands. $\Delta G_{b0} = \Delta G_b + k_B T \ln(V_{\mathrm{site}}/V_0)$ is the annihilation free energy of the ligand in the bound state including a SSC correction of $k_B T(\ln V_{\mathrm{site}}/V_0) = -0.2 \pm 0.1$ kcal mol^{-1}. ΔG_s is the annihilation free energy of the free ligand in bulk. The experimental absolute dissociation free energies for FK506-FKBP12(native) and N-Elte378-FKBP12(native) are taken from Ref.[51], while that of SB3-FKBP12(native) is taken from Ref.[35]. All units are in kcal mol^{-1}.

of Eq. 10, with the only exception of N-Elte378 in bulk solvent (see Figure 4 and related discussion), for which ΔG_s was evaluated using a mixture of two normal components and applying the corresponding unidirectional estimator of Eq. 12. The free energies were all evaluated from the 512 final work values, using a bootstrap based error analysis on 40 random subsets containing 256 work samples. The volume of the cavity allocating the substrate, V_{site}, is unknown. The cavity, however, should be large enough to accommodate the natural FKBP12 binder FK506. The SSC term was therefore evaluated using, as binding site volume, the mean volume of the FK506 ligand in solution, i.e. $V_{\mathrm{site}} = V_{\mathrm{FK506}} = 1117 \pm 242$ Å^3. This volume was computed by means of a Voronoi polihedra analysis on 512 H-REM configurations in the target state of the solvated ligand.[78] The agreement between FS-DAM computed standard dissociation free energies and the corresponding experimental data, all referring to the native species, is excellent, especially for FK506 and N-Elte378. The experimental dissociation constants of these two ligands were measured in the same experimental conditions, using the intrinsic tryptophan quenching upon binding.[51] In the case of the SB3 ligand, the experimental data is actually derived from an *inhibition constant* measured[35] using the standard chymotrypsin digestion assay.[77] This fact may explain the slight overestimation (by 1.2 kcal mol^{-1}) of the calculated dissociation free energy of the SB3-FKBP12(native) complex with respect to the experimental data.

4.5 The impact of the I56D mutation in binding

In case of N-Elte378 and SB3, the I56D mutation weakens the binding affinity, especially for the former compound. This is consistent with the idea[37] that the binding strength and specificity in these FK506-related ligands is due to the formation of two simultaneous H-bonds involving the hydroxy hydrogen in Tyr82 and the backbone amide hydrogen in Ile56. When Ile is mutated to Asp, this H-bond network is disrupted and the electrostatic balance in the binding becomes negative as shown in Table 1. As far as the FK506 ligand is concerned, the effect of the I56D mutation on the dissociation free energy is remarkable indeed, with an unexpected increase of nearly 10 kcal mol^{-1} with respect to the native protein producing a dissociation constant in the femtomolar range. If experimentally confirmed, this result would reveal the FK506-FKBP12(I56D) system as one of the most stable non-covalent protein-ligand complexes in nature.[79] In the complex with the native FKBP12 form, FK506 engages in four H-bonds involving Asp37, Glu54, Ile56 and Tyr82 (see Figure 5 of the Supporting Information) and in a series of hydrophobic contacts. Upon I56D mutation, the H-bond network remains essentially unaffected including the H-bond involving Asp56 that now binds O3 rather than O2 of FK506 (see Figure 1), resulting again in tight anchoring of FK506 in the pocket, as for the native form. The two carboxylate oxygen atoms in Asp56 do not involve additional H-bonds; rather, the Asp56 side chain remains mainly solvent exposed during the EDU-HREM equilibrium simulation, while Ile56 in the native form is constantly in contact with several non polar groups of the ligand. So, apparently, there is no clear-cut reason, related to a specific ligand-protein interactions, for the observed increase in the binding affinity of FK506 in the I56D mutant and water molecules in the binding site could be possibly involved. The situation in FK506-FKBP12 complexes resemble strikingly to that of the well known tightly bound streptavidin-biotin system,[80] where the ligand is enclosed in a wide hydrophobic binding pocket inserted in an antiparallel β-barrel tertiary structure, similar to that of FKBP12, stabilized by several H-bonds. To the best of our knowledge, FK506 affinity for the FKBP12(I56D) mutant is not available experimentally. This is somewhat surprising given that I56D is one of the most destabilizing tested mutations for the FKBP12 protein, yielding a decrease of its denaturation free energy by 3.2 kcal mol^{-1}.[40] Such a destabilization of the unbound native structure upon mutation may in part explain the computed super-affinity of FK506 *vs* FKBP12(I56D). On the other hand, it is known[81] that the solvation of the protein active sites that are characterized by hydrophobic enclosure and correlated H-bonds induces atypical entropic and enthalpic penalties of hydration. These complex and concurrent penalties, like in the case of the streptavidin-biotin pair or, possibly, the FK506-FKBP12(I56D) system, stabilize the protein-ligand complex with respect to the independently solvated ligand and protein, which leads to enhanced binding affinities. The 10 kcal mol^{-1} difference in the FS-DAM computed binding affinity between FK506 and FKBP12 upon I56D mutation should in any case be taken with caution. The record dissociation free energy of the FK506-FKBP12(I56D) complex could be in fact partially debunked if one assumes an underlying two component reference work distributions with a manifold of alternative poses in the FKBP12(I56D) binding site. In fact, the FK506-FKBP12(I56D)

complex is the only system for which an underlying two normal components work distribution function yields an annihilation free energy that, while being still important, is significantly lower than that obtained by using a reference work distribution with a single normal component (16.5 *vs* 22.1 kcal mol^{-1}, see last section in the Supporting Information). Nonetheless, our results strongly call for an experimental verification.

4.6 FS-DAM precision and comparison with equilibrium based FEP studies

In FEP or TI *reversible* methodologies, a canonical sampling should in principle apply at any stage along the alchemical path. As the accessible conformational states may depend on the value of the alchemical parameter λ, in equilibrium techniques, with or without λ-hopping schemes, choosing the λ alchemical protocol so as to get an equal contribution to the overall uncertainty across every point along the alchemical path is a far from trivial task.[28] Such difficulties have an impact on the reliability of the computed free energies, whose values and uncertainties may vary considerably with mean simulation times on the λ states ranging from less than one to several nanoseconds.[25,82,83] For example, in Ref.[25], the FEP dissociation free energy of FK506-FKBP12(native) complex, using the AMBER/GAFF force field as in the present study, is found to be 10.1 ± 3.5 kcal mol^{-1} running simulations for a total of 3 ns time span (2 ns of equilibration) at each of the 31 λ points. Although not explicitly stated, the volume restraint used for ligand in Ref.[25] was very likely implemented via a relatively weak stretching potential (force constant of 0.24 kcal mol^{-1} Å^{-2}) between the pipecolic nitrogen and the binding site centroid.[75] The error on the dissociation free energy was evaluated repeating the 3 ns alchemical simulations on the 31 λ points for twelve times, starting, for each λ point, from the same configurational state with different initial momenta. In Ref.[75], Shirts reports for the SB3-FKBP12 complex a dissociation free energy of 7.3 $\pm$1 kcal mol^{-1} using simulation times of the order of 10 ns per λ point. Wang, Deng and Roux[82] repeated the calculations reported in Ref.[25] for the FK506-FKBP12 complex using FEP with translational/rotational restraints for the ligand in the binding pocket, finding a dissociation free energy of 10.8 ± 3.0 kcal mol^{-1}. In all cases, they used very tight translational/rotational/conformational restraint potentials (see Table 4 of Ref.[82]) allowing for relatively short alchemical simulations (1 to 2 ns). The error was determined repeating the complete FEP/MD calculation for 3-5 times. For the SB3-FKBP12 complex, they report a dissociation free energy of 10.3 ± 1.2 kcal mol^{-1}. The tight restraints make their approach equivalent, *de facto*, to a sophisticated single pose docking technology with analytical evaluation of the so-called cratic free energy[84] coming from the restraints. In this spirit, Sunhwan *et al.*[83] performed calculations on FKBP12 complexes with FK506-related ligands using the same FEP procedure with restraints described in Ref.[82], but employing much shorter simulations at the intermediate λ points (0.1 ns) and adopting the CHARMM PARAM22 parameter set.[85] They so obtained a dissociation free energy of 11.5 ± 2.4 kcal mol^{-1} for the FK506-FKBP12 complex and of 13.1 ± 3 kcal mol^{-1} for the SB3-FKBP12 complex.

In FS-DAM, given that the few hundreds of starting configurations collected during the EDU-HREM simulations at the fully coupled state have produced a correct Boltzmann sampling of the ligand in the binding site or in the bulk, then the overall error on the dissociation free energies can be very naturally and reliably computed via standard block-bootstrapping from the collection of N annihilation works and applying to each independent sample either Eq. 10 or Eq. 12 for estimating the dissociation free energy. Random block bootstrapping is fully justified in our case as the target state sampling in EDU-HREM (for the complex) or HREM (for the ligand in bulk) simulations is not time ordered and is composed by the uncorrelated contributions coming from eight independent replicas in the GE. In Table 3, the annihilation free energies and corresponding errors have been computed by block bootstrapping the 512 work values into 40 samples with 64, 128 and 256 elements. As expected, the annihilation free energy estimates based on Eq. 10 remain stable, while errors increase moderately even for very rough sampling. The above is true also when the estimate is based on a two-component distribution as for the case of N-Elte378 in bulk where Eq. 12 applies. Table 3 is important from a practical standpoint. An accuracy within 1 kcal/mol for the dissociation free energy, that would be pharmaceutically useful, can be granted by FS-DAM for drug size molecules such as N-Elte378 or SB3 using as little as 64 NE non communicating trajectories each lasting few hundreds picoseconds, hence investing a limited amount CPU resources. On per ligand basis, the EDU-HREM and subsequent FS-DAM steps are completed in about 12 wall clock time hours, using $\simeq$ 600 core hours (core unit Intel Nehalem E5530 2.4 GHz) on the ENEA-CRESCO2 HPC system.[86]

	$N = 256$	$N = 128$	$N = 64$
FK506(native)	33.2±0.5	33.5±1.8	33.6±2.9
FK506(I56D)	42.7±0.4	42.9±1.3	43.0±2.5
FF506(bulk)	20.4±0.1	20.3±0.3	20.3±0.6
N-Elte378(native)	22.7±0.1	22.9±0.3	23.0±0.6
N-Elte378(I56D)	18.1±0.1	18.0±0.3	18.1±0.9
N-Elte378(bulk)	9.8±0.4	9.9±0.7	9.9 ±0.8
SB3(native)	18.7±0.3	19.0±0.5	19.1±1.0
SB3(I56D)	17.5±0.3	17.5±1.0	17.7±1.8
SB3(bulk)	6.3±0.1	6.2±0.1	6.3±0.1

Table 3 Annihilation free energies computed for various number N of randomly sampled works out 512 available work values. The error has been evaluated using 40 boostrapped samples. The free energies are estimated using Eq. 10 except for N-Elte378 in bulk where for which Eq. 12 is applied. All units are in kcal mol^{-1}.

Regarding this crucial aspect for a possible application of FS-DAM technology for a second generation of High-Throughput Virtual Screening in drug discovery, quoting Shirts[75] the FEP equilibrium approach for the case of the FKBP12 ligands required "$\simeq$ 600 ns per λ value per ligand for the ligand decoupling simulations, with individual simulations averaging $\simeq$20 ns in length and $\simeq$150-200 ns per λ value per ligand for the complexes, with typical lengths between 8 and 10 ns" running on the `Folding@home`[87] distributed platform. Assuming 31 λ val-

ues,[25,75] this would make about 18 μs per ligand. On a per ligand basis, FS-DAM requires a total of 80 ns equilibrium EDU-HREM GE simulation (5 ns per replica on 16 states) for collecting the starting canonically sampled configuration of the bound states and a total of 138 ns (for 512 trajectories) for the annihilation of the ligand in the complex. The FS-DAM approach is by nature embarrassingly parallel.[29] These data are summarized in Table 4. They show that FS-DAM outperforms FEP approach,[25,27,75]

	N_τ	N_λ	Simulation time (ns per ligand)	Mean error on ΔG_0 (kcal mol^{-1})
FS-DAM	512	n/a	218	0.3
FS-DAM	256	n/a	149	0.7
FS-DAM	128	n/a	115	1.5
FEP[75]	n/a	31	18000	1.5
FEP[25]	n/a	33	400	4.5
FEP/BAR[27]	n/a	32	900	3.0
FEP-restraint[82]	n/a	25	250	1.5

Table 4 Performances of NE FS-DAM and equilibrium FEP. N_τ and N_λ indicate the number of independent NE trajectories (applicable in FS-DAM only) and the number of λ intermediate states (applicable in FEP only). All data refer to the FKBP12 receptor on per ligand basis.

both in terms of precision/reliability and of CPU time. FS-DAM appears at least twice as efficient and precise with respect to the FEP with tight restraints adopted in Ref.[82]. However, while FS-DAM, being based on an unrestricted equilibrium sampling of the bound states via EDU-HREM GE simulations, is a methodology aimed at identifying and handling multiple poses simultaneously, the FEP with restraints is by construction a post-docking *single pose* technique and[82] "is most useful when used in conjunction with accurate starting configurations, obtained either from X-ray crystallography or high-quality docking models."

5 Conclusions

In this study we have applied the Fast Switching Double Annihilation Method (FS-DAM)[29] to the determination of the standard dissociation free energy of FK506-related ligands, N-Elte378, SB3 and FK506 with respect to native FKBP12 and to the I56D mutant. FS-DAM is the nonequilibrium variant of the Double Annihilation Method invented in 1985 by Jorgensen and Ravimohan[8] and is based on the production of few hundreds of fast nonequilibrium (NE) trajectories where the ligand is annihilated in the bulk and in the complex. In both cases, the NE independent annihilation trajectories last few hundreds of picoseconds, each starting from a canonically sampled microstate of the fully coupled system at $\lambda = 1$. The equilibrium initial configurations at $\lambda = 1$ are produced using enhanced sampling techniques. For the annihilation of the ligand in bulk, we use Replica Exchange with Solute Tempering with scaling of the intraligand potential (torsional and non bonded contributions) allowing for a rapid canonical sampling of the ligand conformational states contributing to the hydration energy. For the ligand in the complex, we also use Replica Exchange simulations coupled with the Energy Driven Undocking technique[15] to canonically collect bound state configurations. Excessive wandering of the ligand into the bulk, possibly occurring in the hot, highly scaled, replicas is prevented by imposing a weak tethering potential with a minimal biasing effect in the target state corresponding to the tightly bound complex. Using EDU-HREM, we can get in few nanoseconds an essentially unrestrained and enhanced sampling of the bound state for subsequent fast annihilation step, a sampling that is far superior to the conventional sampling at $\lambda = 1$ using standard MD as in Free Energy Perturbation (FEP) or Thermodynamic Integration (TI) methods with restraints.

The fast switching stage of the coupled states is found to produce normally distributed work values allowing for a very precise and robust unbiased estimate of the annihilation free energies, based on the application of the Crooks NE work theorem to Gaussian work distributions. The dissociation free energies are recovered, as in the equilibrium variant, as a difference between the annihilation free energies of the ligand in the complex and in bulk plus a small standard state correction. They are characterized by a common trend for all FKBP12-related ligands, with most of the binding energy arising from the dispersive-repulsive contribution of the ligand-environment interaction potential, in agreement with previous FEP studies and with experimental indications. For the systems involving the native protein, we have obtained standard dissociation free energies that are in close agreement with the available experimental data. For the mutated protein, for which no experimental dissociation constants are available, we predict a slight weakening of the binding strength of SB3 and a significant decline of the affinity of the N-Elte378 ligand, elicited by an important loss of favorable and highly specific electrostatic interactions. For Tacrolimus (FK506), the I56D mutation induces a remarkable enhancement in the binding affinity, yielding a dissociation constant in the femtomolar range. This implies that the FK506-FKBP12(I56D) complex is predicted to be one of the most stable non-covalent protein-ligand complexes in nature. Such results strongly call for an experimental verification.

Assuming a correct sampling of the initial configurations of the fully coupled state, where usually one pose has overwhelming weight with respect to all others alternate ligand-protein interactions, we show that the errors in the FS-DAM determination of the standard dissociation free energies can be very reliably assessed using standard bootstrapping techniques. The errors in all analyzed cases are only moderately affected by substantial reduction of the number of independent NE trajectories, while the dissociation free energy values remain remarkably stable. When compared to the parent equilibrium variants such as FEP or TI, the FS-DAM outperforms, both in terms of precision/reliability and of CPU time investment, the equilibrium approaches. The efficiency, simplicity and inherent parallel nature of the FS-DAM algorithm, project the methodology as a possible effective tool for a second generation High Throughput Virtual Screening in drug discovery and design. In this context, FS-DAM can be easily adapted and used for a reliable bio-availability ADME-tox assessment of candidate drugs, evaluating hydration energy, water-octanol partition coefficients and membrane affinities.

6 Acknowledgements

The computing resources and the related technical support used for this work have been provided by CRESCO/ENEAGRID High Performance Computing infrastructure and its staff.[88] 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 http://www.cresco.enea.it/english for information

References

1 T. Ewing, S. Makino, A. Skillman and I. Kuntz, *Journal of Computer-Aided Molecular Design*, 2001, **15**, 411–428.

2 G. Morris and M. Lim-Wilby, *Molecular Modeling of Proteins*, Humana Press, 2008, vol. 443, pp. 365–382.

3 G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *J Comput Chem*, 2009, **30**, 2785–2791.

4 N. Deng, S. Forli, P. He, A. Perryman, L. Wickstrom, R. S. K. Vijayan, T. Tiefenbrunn, D. Stout, E. Gallicchio, A. J. Olson and R. M. Levy, *The Journal of Physical Chemistry B*, 2015, **119**, 976–988.

5 *Chemogenomics and Chemical Genetics.A User's Introduction for Biologists, Chemists and Informaticians*, ed. E. Marechal, S. Roy and L. Lafanechere, Springer-Verlag Berlin Heidelberg, 2011.

6 B. Munos, *Nat Rev Drug Discov*, 2009, **8**, 959–968.

7 J. W. Scannell, A. Blanckley, H. Boldon and B. Warrington, *Nat Rev Drug Discov*, 2012, **11**, 191–200.

8 W. Jorgensen and C. Ravimohan, *J. Chem. Phys.*, 1985, **83**, 3050–3054.

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

10 J. G. Kirkwood, *J. Chem. Phys.*, 1935, **3**, 300–313.

11 I. J. General, *Journal of Chemical Theory and Computation*, 2010, **6**, 2520–2524.

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

13 P. Procacci, *The Journal of Chemical Physics*, 2015, **142**, 154117.

14 E. Gallicchio, M. Lapelosa and R. M. Levy, *J. Chem. Theory Comput.*, 2010, **6**, 2961–2977.

15 P. Procacci, M. Bizzarri and S. Marsili, *J Chem. Theory Comp.*, 2014, **10**, 439–450.

16 J. Chodera, D. Mobley, M. Shirts, R. Dixon, K.Branson and V. Pande, *Curr. Opin Struct. Biol*, 2011, **21**, 150–160.

17 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.

18 M. R. Shirts and V. S. Pande, *J. Chem. Phys.*, 2005, **122**, 144107.

19 M. Fasnacht, R. H. Swendsen and J. M. Rosenberg, *Phys. Rev. E*, 2004, **69**, 056704.

20 D. L. Mobley, A. P. Graves, J. D. Chodera, A. C. McReynolds, B. K. Shoichet and K. A. Dill, *Journal of Molecular Biology*, 2007, **371**, 1118 – 1134.

21 A. Pohorille, C. Jarzynski and C. Chipot, *The Journal of Physical Chemistry B*, 2010, **114**, 10235–10253.

22 J. C. Gumbart, B. Roux and C. Chipot, *J. Chem. Theory Comput.*, 2013, **9**, 974–802.

23 N. Hansen and W. F. van Gunsteren, *Journal of Chemical Theory and Computation*, 2014, **10**, 2632–2647.

24 J. W. Kaus and J. A. McCammon, *J. Phys. Chem. B*, 2015, **119**, 6190–6197.

25 H. Fujitani, Y. Tanida, M. Ito, G. Jayachandran, C. D. Snow, M. R. Shirts, E. J. Sorin and V. S. Pande, *The Journal of Chemical Physics*, 2005, **123**, 084108.

26 Y. Deng and B. Roux, *J. Phys. Chem. B*, 2009, **113**, 2234–2246.

27 H. Fujitani, Y. Tanida and A. Matsuura, *Phys. Rev. E*, 2009, **79**, 021914.

28 L. N. Naden and M. R. Shirts, *Journal of Chemical Theory and Computation*, 2015, **11**, 2536–2549.

29 R. B. Sandberg, M. Banchelli, C. Guardiani, S. Menichetti, G. Caminati and P. Procacci, *Journal of Chemical Theory and Computation*, 2015, **11**, 423–435.

30 M. Goette and H. Grubmüller, *Journal of Computational Chemistry*, 2009, **30**, 447–456.

31 V. Gapsys, D. Seeliger and B. de Groot, *J. Chem. Teor. Comp.*, 2012, **8**, 2373–2382.

32 P. Procacci and C. Cardelli, *J. Chem. Theory Comput.*, 2014, **10**, 2813–2823.

33 P. Procacci, Unpublished.

34 G. E. Crooks, *J. Stat. Phys.*, 1998, **90**, 1481–1487.

35 D. A. Holt, J. J. I. Luengo, D. S. Yamashita, H. J. Oh, A. L. Konialian, H. K. Yen, L. W. Rozamus, M. Brandt, M. J. Bossard, M. A. Levy, D. S. Eggleston, J. Liang, L. W. Schultz, T. J. Stout and J. Clardyt, *J. Am. Chem. Soc.*, 1993, **115**, 9925–9938.

36 M. Banchelli, C. Guardiani, E. Tenori, S. Menichetti, G. Caminati and P. Procacci, *Phys. Chem. Chem. Phys.*, 2013, **15**, 18881–18893.

37 M. Bizzarri, E. Tenori, M. Martina, S. Marsili, G. Caminati, S. Menichetti and P. Procacci, *J. Phys. Chem. Letters*, 2011, **2**, 2834–2839.

38 M. Lawrenz, D. Shukla and V. S. Pande, *Sci. Rep.*, 2015, **5**, 7918.

39 T. Yamashita, A. Ueda, T. Mitsui, A. Tomonaga, S. Matsumoto, T. Kodama and H. Fujitani, *Chemical and Pharmaceutical Bulletin*, 2015, **63**, 147–155.

40 E. R. G. Main, K. F. Fulton and S. E. Jackson, *Biochemistry*, 1998, **37**, 6145–6153.

41 G. Van-Duyne, R.F.Standaert, S.L.Schreiber and J.Clardy, *J. Am. Chem. Soc.*, 1991, **113**, 7433–7434.

42 J. Griffith, J. Kim, E. Kim, M. Sintchak, J. Thomson, M. Fitzgibbon, M. Fleming, P. Caron, K. Hsiao and M. Navia, *Cell*, 1995, **82**, 507–522.

43 C. H. Bennett, *J. Comp. Phys.*, 1976, **22**, 245–268.

44 M. R. Shirts, E. Bair, G. Hooker and V. S. Pande, *Phys. Rev. Lett.*, 2003, **91**, 140601.

45 H.-X. Zhou and M. K. Gilson, *Chem. Rev.*, 2009, **109**, 4092–

4107.

46 The statistical average of the quantity $I(\mathbf{r},\Omega)$ defining the complex R-L tends to zero at infinite dilution, such that the product of the $< I(\mathbf{r},\Omega) >$ times V tends to the equilibrium constant as V tends to infinity:[47]

$$< H(\mathbf{r},\Omega) > V|_{\lim\to\infty} = V \frac{\int I(\mathbf{r},\Omega) e^{-\beta w(\mathbf{r},\Omega)} d\mathbf{r} d\Omega}{\int e^{-\beta w(\mathbf{r},\Omega)} d\mathbf{r} d\Omega}$$

$$= \frac{1}{8\pi^2} \int I(\mathbf{r},\Omega) e^{-\beta w(\mathbf{r},\Omega)} d\mathbf{r} d\Omega = K$$

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

47 H. Luo and K. Sharp, *Proc. Natnl. Acad. Sci. USA*, 2002, **99**, 10399–10404.

48 K. Krishnamoorthy, *Handbook of Statistical Distributions with Applications*, Chapman and Hall/CRC, London (UK), 2006.

49 S. Marsili, G. F. Signorini, R. Chelli, M. Marchi and P. Procacci, *J. Comp. Chem.*, 2010, **31**, 1106–1116.

50 M. Bizzarri, S. Marsili and P. Procacci, *J. Phys. Chem. B*, 2011, 6193–6201.

51 M. R. Martina, E. Tenori, M. Bizzarri, S. Menichetti, G. Caminati and P. Procacci, *J. Med. Chem.*, 2013, **56**, 1041–1051.

52 C. Guardiani, G. F. Signorini, R. Livi, A. M. Papini and P. Procacci, *J. Phys. Chem. B*, 2012, **116**, 5458–5467.

53 M. R. Shirts and J. D. Chodera, *J. Chem. Phys.*, 2008, **129**, 124105.

54 P. Procacci, *J. Chem. Phys.*, 2013, **139**, 124105.

55 F. Nerattini, *MSc thesis*, University of Florence, Italy, 2014.

56 V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins*, 2006, **65**, 712–725.

57 J. Wang, R. Wolf, J. Caldwell, P. Kollman and D. Case, *J. Comp. Chem.*, 2004, **25**, 1157–1174.

58 M. J. S. Dewar, E. G. Zoebisch, E. F. Healy and J. J. P. Stewart, *Journal of the American Chemical Society*, 1985, **107**, 3902–3909.

59 A. D. Becke, *Phys. Rev. A*, 1988, **33**, 3098.

60 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785–789.

61 J. Tomasi, B. Mennucci and E. Cancés, *J. Mol. Struct. (Theochem)*, 1999, **464**, 211–226.

62 U. Singh and P. Kollman, *J. Comput. Chem.*, 1984, **5**, 129–145.

63 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ã. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09 Revision A.1*, Gaussian Inc. Wallingford CT 2009.

64 W. L. Jorgensen, J. Chandrasekhar, J. Madura, R. Impey and M. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

65 M. Marchi and P. Procacci, *J. Chem. Phys.*, 1998, **109**, 5194–5202.

66 S. Nosé, *Mol. Phys.*, 1984, **52**, 255–268.

67 W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695–1697.

68 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.

69 M. Tuckerman and B. J. Berne, *J. Chem. Phys.*, 1992, **97**, 1990–2001.

70 P. Procacci, E. Paci, T. Darden and M. Marchi, *J. Comp. Chemistry*, 1997, **18**, 1848–1862.

71 K. P. Wilson, M. M. Yamashita, M. D. Sintchak, S. H. Rotstein, M. A. Murcko, J. Boger, J. A. Thomson, M. J. Fitzgibbon, J. R. Black and M. A. Navia, *Acta Crystallographica Section D*, 1995, **51**, 511–521.

72 P. Liu, B. Kim., R. A. Friesner and B. J. Berne, *Proc. Acad. Sci.*, 2005, **102**, 13749–13754.

73 H. Fukunishi, O. Watanabe and S. Takada, *J. Chem. Phys.*, 2002, **116**, 9058–9067.

74 T. Beutler, A. Mark, R. van Schaik, P. Gerber and W. van Gunsteren, *Chem. Phys. Lett.*, 1994, **222**, 5229–539.

75 M. R. Shirts, *PhD thesis*, Stanford University CA, Stanford, California 94305, 2005.

76 T. Ikura and N. Ito, *Protein Sci.*, 2007, **16**, 2618–2625.

77 G. Fischer, H. Bang and C. Mech, *Biomed. Biochim. Acta*, 1984, **43**, 1101–1111.

78 P. Procacci and R. Scateni, *International Journal of Quantum Chemistry*, 1992, **42**, 1515–1528.

79 I. D. Kuntz, K. Chen, K. A. Sharp and P. A. Kollman, *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 9997–10002.

80 N. M. Green, *Advances in Protein Chemistry*, Academic Press, 1975, vol. 29, pp. 85–133.

81 T. Young, R. Abel, B. Kim, B. J. Berne and R. A. Friesner, *Proceedings of the National Academy of Sciences*, 2007, **104**, 808–813.

82 J. Wang, Y. Deng, B. and Roux, *Biophys. J.*, 2006, **91**, 2798–2814.

83 S. Jo, W. Jiang, H. S. Lee, B. Roux and W. Im, *Journal of Chemical Information and Modeling*, 2013, **53**, 267–277.

84 J. Hermans and L. Wang, *Journal of the American Chemical Society*, 1997, **119**, 2707–2714.

85 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, *The Journal of Physical Chemistry B*, 1998, **102**, 3586–3616.

86 *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).

87 V. Pande, *Folding@home*, See *https://folding.stanford.edu* (accessed date 24/06/2015).

88 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.