JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

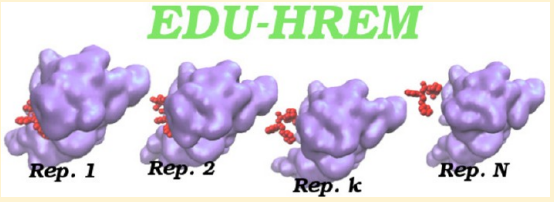
Energy-Driven Undocking (EDU-HREM) in Solute Tempering Replica Exchange Simulations

Piero Procacci,*[,†] Marco Bizzarri,[†] and Simone Marsili[‡]

[†]Dipartimento di Chimica, Università di Firenze, Via della Lastruccia 3, I-50019 Sesto Fiorentino, Italy

[‡]Centro Nacional de Investigaciones Oncologicas, Calle de Melchor Fernández Almagro, 3, E-28029 Madrid, Spain

ABSTRACT: We present a new computational strategy for calculating the absolute binding free energy for protein ligand association in the context of atomistic simulation in explicit solvent. The method is based on an appropriate definition of a solute tempering scheme enforced via Hamilton replica exchange method (HREM). The definition of "solute" includes both the ligand and the active site, with the remainder of the systems defined as "solvent". The hydrophilicity of the solute and the solute torsional plus nonbonded intrasolute interactions are increased and decreased, respectively, along the replica progression, thus favoring the extrusion of the drug form the active site in the scaled states of the generalized ensemble. The proposed technique, named "Energy Driven Undocking" (EDU-HREM), completely bypasses the need for defining and/or identifying the relevant reaction coordinates in a ligand receptor interactions and allows the calculation of the absolute binding free energy in one single generalized simulation of the drug-receptor system. The methodology is applied, with encouraging results, to the calculation of the absolute binding free energy of some FK506-related ligands of the peptidyl prolyl *cis*−*trans* isomerase protein (FKBP12) with known dissociation constants. Aspects of the binding/inhibition mechanism in FKBP12 are also analyzed and discussed.



1. INTRODUCTION

The determination of the absolute binding affinities for drug-receptor complexes from molecular modeling is probably one of the major ongoing challenges in pharmaceutical and medicinal chemistry. The importance of the in silico reliable evaluation of the drug-receptor binding free energies in the drug discovery process can be grasped by considering the high economic and time-consuming burden for the total synthesis of simple organic compounds, a process that has virtually no cost on a computer. The prediction of drug binding affinities of trial compounds is traditionally tackled in the pharmaceutical industry by means of docking posing/scoring technologies based on simplifying assumptions such as implicit solvent, restrained flexibility, and simplified force fields. These assumptions generally have a strong negative impact on the predictive power of docking methods whose pathologically poor specificity is typically prone to produce a significant share of false positives or false negatives. Molecular dynamics (MD) with explicit solvent and accurate force fields at the atomistic level are, in principle, a powerful tool for determining absolute binding free energy in drug receptor complexes with great accuracy.[1] In this context, modern approaches, such as the so-called alchemical decoupling scheme,[2,3] proceed by constructing ad hoc thermodynamic cycles whereby the binding free energy is broken up in several contributions (typically, the reversible alchemical decoupling of the ligand with its surroundings) to be computed individually by expensive simulations.[2−4] Other modern and very promising MD techniques, such as metadynamics[5−8] or non equilibrium steered MD,[9−12] relies on the definition of appropriately selected docking collective variables for evaluating the associated potential of mean force. Identification of low dissipation drug escaping paths and/or the relevant drug-receptor collective coordinates in these pioneering studies is, however, a nontrivial, highly system-dependent task.

In this contribution, we present an alternative route to the calculation of the absolute binding constant for the drug-receptor system in the context atomistic molecular dynamics simulations with explicit solvent based on an appropriate definition of a generalized ensemble (GE) enforced via the Hamilton replica exchange method (HREM) technology. The proposed technique completely bypasses the need to define and/or identify the relevant reaction coordinates in a ligand−receptor interaction and allows the calculation of the absolute binding affinity in one single generalized simulation of the drug-receptor system. The method starts from the observation that the equilibrium of a ligand−receptor binding reaction, and more in general of molecular recognition processes, is the result of the balance between ligand−receptor, ligand-solvent receptor-solvent and solvent−solvent enthalpic and entropic terms.[13] A specific HREM solute tempering protocol is then implemented whereby the solute in the system is defined, including the active site of the receptor and the drug; the hydrophilicity of the "solute" and the torsional and intrasolute nonbonded interactions are increased and decreased, respectively, along the replica progression while the explicit solvent is held in the thermodynamic conditions of the target state. The

Received: September 14, 2013

 © XXXX American Chemical Society dx.doi.org/10.1021/ct400809n | *J. Chem. Theory Comput.* XXXX, XXX, XXX−XXX

hydrophilicity increase, achieved by counterscaling the solute−solvent potential term in the GE, favors the extrusion of the drug from the active site, as a more hydrophilic solute, made of the binding pocket *and* the drug, tends to get solvated more easily. On the other hand, the selective heating of the solute internal degrees of freedom (while most of the system remains cold) helps to effectively span the *relevant* conformational space (i.e., that including the drug and the binding pocket) using an affordable number of replicas. The proposed protocol has general applicability and allows the reliable determination of the absolute binding constant with a single HREM simulation on the drug-receptor system.

This new methodology, named "Energy Driven Undocking" (EDU-HREM), is here applied to the determination of the binding affinities of the proline mimetic pipecolic ligands vs the peptidyl prolyl isomerase FKBP12 protein, a member of the family of the immunophilins, including cyclophilin, FKBP, and Parvulin.[14−16] The so-called peptidyl prolyl isomerase domain (PPI) in FKBP enzymes catalyzes the *cis*−*trans* isomerization of the peptide bond in proline. FKBPs are hence thought to have an important role in misfolding disorders, such as Alzheimer's disease[17,18] and Parkinson's disease.[19] Furthermore, immunophilins, when complexed with immunosuppressant drugs such as Rapamycin, FK506, and Cyclosporin, are important in the allosteric inhibition of Calcineurin and mTor proteins.[20,21] The functional mechanism of the immunophilin enzymes, i.e. the isomerization of the proline peptide bond, is still under intense debate with opposite views favoring the unassisted twisting mechanism with lowering of the *cis*−*trans* barrier due to desolvation effects of the peptide bond at the active site[22−24] or other hypothesis based on the existence of a specific epitope defined by highly conserved residues in the PPI binding pockets, such as Tyr82, Ile56, Trp59, Asp37[23] in FKBP12 or Arg55, Asp102, His126 in Cyclophilin A.[24,25] The EDU-HREM based techniques developed in the present study allows to reliably isolate, using consolidated reweighting techniques,[26] the contribution of selected portions of the FKBP12 to the computed binding free energy, thereby clarifying at the molecular level the energetic role in the binding process of each of these interactions.

The paper is organized as follows. In section 2, we present a theoretical introduction to the statistical mechanics of the binding process and we describe the theoretical basis for the EDU-HREM technique. In section 3, we present results for a series of FKBP12 pipecolic ligands with disparate binding constant, illustrating the merit and the shortcomings of the method. In section 44, we analyze and discuss, along the ligand series, the molecular details of the binding mechanism, examining the role of each site-specific interaction between the FKBP binding pocket and the ligand proline mimetic moiety. Finally conclusive remarks are presented in section 5.

2. THEORY AND COMPUTATIONAL METHODS

The standard free energy for *intermolecular* binding is given by

$$\Delta G = -k_B T \ln(C_0 K_a) \tag{1}$$

where K_a, the equilibrium constant, is defined as

$$K_a = \frac{V p_b}{p_u p_l} \equiv \frac{1}{K_d} \tag{2}$$

and has the dimension of a volume or an inverse concentration. In eq 2, p_b and p_u are the probability of observing the system in the bound or unbound state, respectively, while $p_l = p_u$ is the probability of observing the free ligand in solution. K_d, the inverse of the equilibrium constant, is the quantity normally measured in experiments. C_0 is the concentration of the standard state and is to be assigned a numerical value of 1 M, corresponding to a molecular average volume of 1661 Å^3. In a standard atomistic simulation, a receptor−drug pair such as FKBP12 and its ligands, with a dissociation constant of nanomolar order, the event of the observing an unbound complex is indeed a rare one. Such probability can be estimated[27] from the definition of the equilibrium constant,

$$K_a = \frac{p V_{box}}{(1-p)^2}$$

expressed in volume per molecule, where V_{box} is the volume of the MD cell (typically of the order of 10^6 Å^3) and p is the probability of observing the complex. For a dissociation constant of nanomolar order, the volume per molecule corresponds to $K_a \approx 10^{12}-10^{13}$ Å^3. Inserting this value in the expression of K_a and solving for p, one finds that the probability of observing the unbound complex in the simulation is given by $(1-p) \simeq (V_{box}/K_a)^{1/2} \simeq 10^{-3}-10^{-4}$. Given a typical diffusion coefficient of the free ligand in water of 0.01 Å^2 ps^{-1}, a reasonable sampling configuration time would be, at least, on the order of a few tens of picoseconds, yielding, on average, just *one* unbound event each tens of nanoseconds. In reality, in a simulation of a single drug−receptor pair lasting a few tens of nanoseconds, the system, were the starting configuration that of bound complex, has a good chance indeed to be observed systematically in the bound state. On the other hand, in the rare event of a dissociation, the diffusing ligand may spend a long time prior to redock the receptor in the active site, leading to severe underestimation of the binding affinity. A reliable brute-force determination of the typical drug-receptor dissociation constants via conventional MD is hence impractical as the combined effect of the strength of ligand−receptor pair and of the slow diffusion in water of the unbound ligand makes the statistics of binding/unbinding events highly unstable and erratic, even for long-lasting simulations with dedicated hardware.[28]

The resolution of the so-called "wandering ligand" problem[2] can be tackled using an appropriate restraint potential that prevents the ligand from "drifting away" in the solvent and redocks the protein in the correct location.[29] In this case, in order to compute the free energy for a given ligand receptor pair, because of the presence of the tethering potential, one must apply the formula for *intramolecular* binding,[27,30,31] where the probability ratio between bound and unbound states is not dependent on the ligand concentration, i.e.,

$$K_r = \frac{C_b}{C_u} = \frac{p_b}{p_u} \tag{3}$$

In the assumption of an ideal linker (no interactions between the linker and the tethered domains), the two equilibrium constants—K_r and K_a—are related[27] by the equation

$$K_r = \rho_l(\mathbf{r_0}) K_a \tag{4}$$

where $\rho_l(r_0)$ is the probability density of the end-to-end vector $\mathbf{r}$ of the linker evaluated at $\mathbf{r}_0$ (i.e., the ligand−receptor vector distance in the bound state). Intuitively, one may, in fact, assume that the role of a flexible tether is to fix the ligand at an

dx.doi.org/10.1021/ct400809n | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

effective concentration, say C_l. It can be then shown[30,32] that such effective concentration is given by

$$C_l = \rho_l(r_0) = \frac{1}{V_{\text{eff}}} \tag{5}$$

If the receptor and the ligand are linked via a simple harmonic potential, then the effective molecular volume, which is due to the linker, is a function of the equilibrium force constant alone and is given by

$$V_{\text{eff}} = 2\left(\frac{2\pi k_B T}{K}\right)^{3/2} \tag{6}$$

where K is the spring constant of the linker. Exploiting eqs 3, 4, and 5, the binding free energy (eq 1) may be rearranged as

$$\begin{aligned}\Delta G &= -k_B T \ln\left(\frac{p_b}{p_u}\right) - k_b T \ln\left(\frac{C_0}{\rho_l(r_0)}\right) \\ &= -k_B T \ln\left(\frac{p_b}{p_u}\right) - k_b T \ln\left(\frac{V_{\text{eff}}}{V_0}\right)\end{aligned} \tag{7}$$

with V_0 being the molecular volume of the standard state.

The Hamiltonian Replica Exchange Energy-Driven Undocking (EDU-HREM) Method for a Ligand Receptor Pair in Explicit Solvent. While the introduction of the tethering potential bypasses the problem of the slow diffusion of the ligand once detached from the receptor, still, for tight-binding ligands, the dissociation event leading to the unbound state is hardly sampled in conventional simulation of a single drug receptor pair lasting on the order of nanoseconds to microseconds. We shall therefore determine the binding affinity using eq 7 by favoring the *extrusion* of the drug from the binding site (i.e., enhancing p_u in eq 7) via modulation of the solute−solvent interaction in a GE simulation. To this end, we set up a Hamiltonian replica exchange (HREM) simulation according to the following protocol. In a standard MD box containing one ligand molecule, one receptor molecule, and the explicit solvent, we first partition the system in (i) a "solute" that includes the atoms of the binding pocket and its immediate surroundings and the atoms of the ligand and (ii) in the "solvent" comprising the remainder of the system, including the protein scaffold and the explicit solvent. Correspondingly, we rewrite the full interaction potential as

$$V(\mathbf{x}) = V_S(\mathbf{x}_S) + V_s(\mathbf{x}_s) + V_{Ss}(\mathbf{x}_s, \mathbf{x}_S) \tag{8}$$

where $\mathbf{x}$, $\mathbf{x}_S$, and $\mathbf{x}_s$ denote the coordinates of the entire system, of the solute and of the solvent and $V_S(\mathbf{x}_S)$, $V_s(\mathbf{x}_s)$, $V_{Ss}(\mathbf{x}_s,\mathbf{x}_S)$ are the intrasolute, intrasolvent, and solute−solvent interaction potential, respectively. Given n replicas in a GE simulation of the system, we then define the potential of the kth replica as

$$V_k = c_S^{(k)} V_S(\mathbf{x}_S) + V_s(\mathbf{x}_s) + c_{Ss}^{(k)} V_{Ss}(\mathbf{x}_s, \mathbf{x}_S) \tag{9}$$

where $c_S^{(k)}$ and $c_{Ss}^{(k)}$ are the intrasolute and solute−solvent scaling factor for the kth replica with $c_S^{(k)}$ varying between a maximum of 1 (for the target, unscaled replica) and a minimum of $c_S^{\min} < 1$, corresponding to a maximum intrasolute temperature of $1/(k_B\beta c_S^{\min})$ if the target replica is at $T = 1/(k_B\beta)$, and with the solute−solvent scaling factor $c_{Ss}^{(k)}$ scaled from a minimum of 1 in the target replica to a maximum of $c_{Ss}^{\max} > 1$ in the nth replica. If the simulation is started from the bound state, e.g., from the experimental PDB structure of the co-crystal, the counterscaling of the solute−solvent potential has the effect of gradually expelling the ligand from the binding site as the replica solute−solvent interactions are constantly increased according to the factors $c_{Ss}^{(k)}$ along the replica progression. This is so because in the replicas with large $c_{Ss}^{(k)}$ factors the solvent water molecules tend to interact more favorably with the "solute", i.e., with *both* the ligand and the binding pocket thereby intercalating between the two and favoring the separation of the ligand from the protein. The scaling of the intrasolute potential, on the other hand, helps to overcome the torsional barriers in the ligand and in the active site side chains, therefore enhancing the conformational sampling for a correct evaluation of the reorganization free-energy contribution[33] to the binding free energy.

Note that using the EDU-HREM protocol within the framework of the GE approach completely eliminates the need to define escaping coordinates for undocking the ligand from the active site. The optimal (least dissipation) escaping coordinates are automatically selected by the drug and the binding pocket in their effort to get in contact with the bulk solvent guided solely by the scaled solute−solvent potential.

In the proposed protocol, the intrasolvent interaction $V_s(\mathbf{x}_s)$ in eq 9 is not scaled and water mobility remains that which has been determined by the operating room temperature in all replicas. In the free ligand, on the other hand, the solvent cage stabilization induced by the scaling of the $V_{Ss}(\mathbf{x}_s,\mathbf{x}_S)$ term is partly compensated by the enhanced mobility of the ligand via the scaling of the intrasolute potential. As a consequence, the mobility of the free ligand in the scaled replicas does not deviate significantly from ordinary diffusivity in any of the replicas of the system. The presence of a replica-independent harmonic tethering potential prevents the drug from drifting away from the binding pocket. Such restraint potential opposes the expelling force due to the counterscaled solute−solvent interactions and tends to recollocate the ligand, as the counterscaling is decreased, in the binding pocket. In this manner, for a dissociation constant of nanomolar order and because of the tethering potential, in the target (unscaled) replica, the drug−receptor system is systematically found in the bound state and the binding free energy cannot be determined from the ratio between bound and unbound states using the statistics of the target replica only. However, because of the balance between the solute−solvent counterscaling and the weak tethering potential, unbound states for the ligand−receptor systems are now produced with a significant frequency in the scaled states in a nanosecond time span simulation. Therefore, the binding free energy for the drug−receptor pair can be evaluated using the entire generalized ensemble statistics by appropriately reweighting the ratio between bound and unbound states using techniques such as the multiple Bennett Acceptance Ratio (MBAR).[26] According to the MBAR method, the probability ratio between bound and unbound states is evaluated using the full generalized ensemble statistics as

$$\frac{p_b}{p_u} = \frac{\sum_i^N w_i(\mathbf{x}_i) H(\mathbf{x}_i)}{\sum_i^N w_i(\mathbf{x}_i)[1 - H(\mathbf{x}_i)]} \tag{10}$$

where the configuration index i runs on all the N ligand−receptor states sampled by the replicas in the GE simulation, $w_i(\mathbf{x}_i)$ is the MBAR normalized weight[26] for the ith configuration with coordinate $\boldsymbol{x}_i$ and the function $H(\mathbf{x}_i)$ is 1 for configurations corresponding to a bound state and is 0

dx.doi.org/10.1021/ct400809n | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

otherwise.[34] The MBAR weight for the target ensemble of the configuration x_i in the generalized ensemble are calculated as[26]

$$W_1(\mathbf{x}_i) = \frac{\exp[-\beta V_1(\mathbf{x}_i)]}{\sum_l N_l \exp[-\ln Z_l - \beta V_l(\mathbf{x}_i)]} \quad (11)$$

where $V_1(x_i)$ and $V_l(x_i)$ are the unscaled potential function of the target replica (replica 1) and of the lth replica (see eq 9). The partition functions Z_l are determined solving, by iteration, the equation

$$Z_n = \sum_{i=1}^{N} \frac{\exp[-\beta V_n(x_i)]}{\sum_{l=1}^{M} Z_l^{-1} N_l \exp[-\beta V_l(x_i)]} \quad (12)$$

where the index l runs on the number of replicas, N_l is the number of points sampled in the ensemble with potential functionV_l, and $N = \sum_l N_l$. The MBAR technique is fully described in ref 26. Such simulation protocol is schematically illustrated in Figure 1.

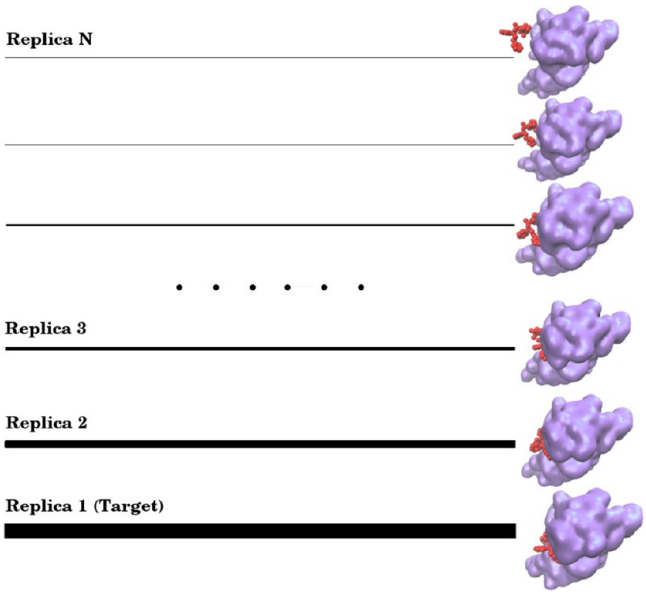


Figure 1. Schematic representation of HREM simulation of a tight binding ligand (in red) with solute−solvent counterscaling. Each horizontal line represents a replica simulation in the generalized ensemble with potential function given by eq 9. The thickness of the line for each replica is proportional to the MBAR weight of the corresponding GE configurations. Unbound states are found only in the last replicas with a small MBAR weight.

Our EDU-HREM method for drug−receptor systems in explicit solvent is very similar in spirit to the so-called Binding Energy Distribution Analysis Method (BEDAM) approach[35] that is used to determine the binding free energy in an implicit solvent. The tethered state here is akin to the reference state in BEDAM of the constrained ligand in the binding site, in the absence of ligand−complex interaction. The ideal (entropy volumetric) term, $-k_bT \ln(V_{eff}/V_0)$, in the two method is identical, while the definition of the binding pocket for intrasolute and solvent−solute scaling (i.e., the part of the protein that is involved in the binding) in EDU-HREM is paralleled by the definition of the binding volume in BEDAM. Also in the BEDAM scheme, the sampling of the important states in the tail of the energy distributions of the uncoupled complex in the binding site is enhanced via a GE simulation approach, whereby the ligand−receptor interaction potential is gradually switched on along the replica progression. While in BEDAM, the rarely sampled states are the bound states, in the EDU-HREM protocol unbound states are the rare configurations whose sampling is greatly accelerated via GE-driven drug extrusion in the bulk.

3. IMPLEMENTATION OF EDU-HREM FOR THE CASE OF THE LIGANDS OF FKBP12

We here apply the EDU-HREM method to the case of FK506-related ligands of the immunophilin FKBP12,[14] which is a cytosolic enzyme that catalyzes the *cis*−*trans* isomerization (rotamerization) of prolyl amide bonds and that is involved in immunosuppression and neuronal functioning. FK506-related ligands are all characterized by the presence of an α-keto amide moiety. In all known co-crystals of the FKBP12 complexes,[36−41] these ligands invariably bind the protein through two hydrogen bonds between the donors OH of the highly conserved residues Tyr82 and NH of Ile56 and the acceptor carbonyl group in the pipecolyl α-keto amide region of the ligand (see Figure 8, presented later in this paper). Nevertheless, the presence of such chemical group does not guarantee by itself a large affinity,[37] implying that the binding mechanism in immunophilins is far more complex, with respect to the simple picture emerging form the static crystallographic structures. It has been recently shown[42,43] that the affinity of FK506-related ligands is actually a complex combination of enthalpic and entropic contributions, with a fundamental role played by the ligand reorganization energy.[33] Effective FKBP12 ligands are in fact those that mimic, in bulk solution, the rigid structure of the FK506 natural drug with respect to the orientation of the two carbonyl groups units of the ligand that will form, upon binding, hydrogen bonds with specific residues in the protein binding pocket. Such rigid structures in potent FK506-related ligands are favored by intraligand hydrophobic interactions that reduce the conformational entropy of the unbound state, hence enhancing the binding affinity. In the present paper, we test the EDU-HREM method to determine the *absolute* binding free energy via eqs 7 and 10 of three pipecolic ligands ligands (see Figure 2), namely, compounds **3**,

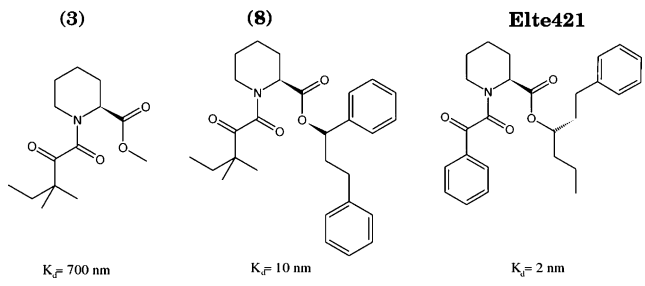


Figure 2. FKBP12 ligands. The experimental dissociation constants for compounds **3** and **8** are reported in ref 37. The dissociation constant for Elte421 is reported in ref 43.

8, and Elte421,[42] with known experimental binding affinities.[37,42] For each ligand−FKBP12 pair, the system was prepared as follows: the protein FKBP12 was built using the crystallographic structure of the complex **8**−FKBP12[37] (Protein Databank (PDB) ID: 1FKG). The force field for the protein was AMBER99sb.[44] For compounds **3** and **8**, the force field was taken from ref 45, while, for the Elte421, the potential parameters and atomic charges are those reported in ref 42.

dx.doi.org/10.1021/ct400809n | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

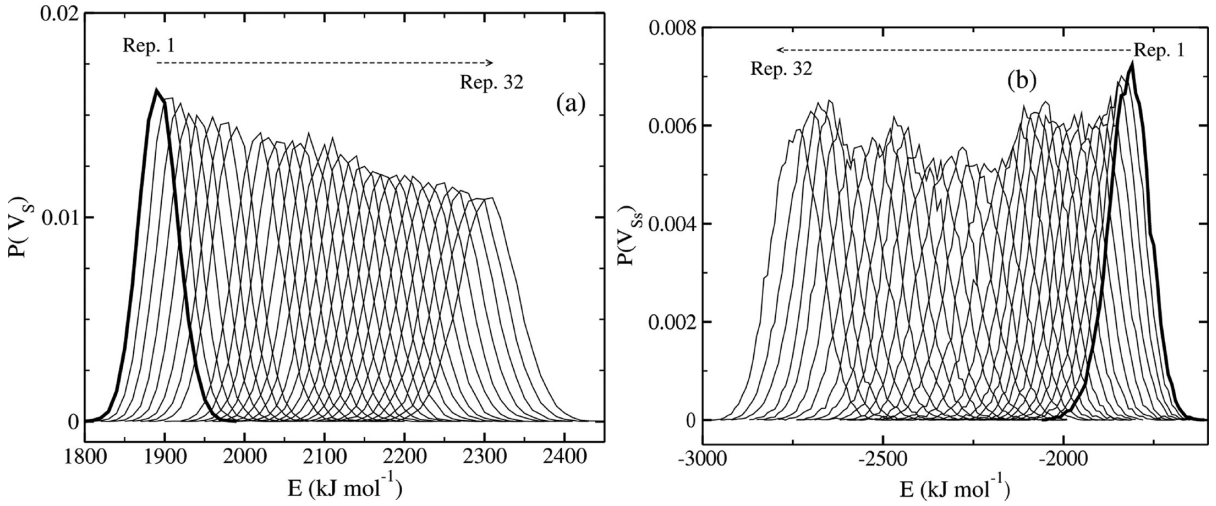


Figure 3. Energy distribution functions for (a) the solute−solute energy $V_S(\mathbf{x})$ and (b) the solute−solvent energy $V_{Ss}(\mathbf{x})$ in a 32 HREM simulation of the Elte421−FKBP12 complex. Thick lines refer to the target replica.

The ligand was initially collocated from a far distance into the FKBP12 binding pocket using a short (10 ps) steered molecular dynamics in vacuo, by means of a time-dependent harmonic external potential connecting the protein and the ligand and involving the distances O3-HO(Tyr93) and O2-HN(Ile56). (See Figure 8, presented later in this work, for the definition of O2 and O3 in the pipecolate α-keto amide core.) During the preparatory steered molecular dynamics (MD), the backbone atoms of FKBP12 were kept fixed at their experimental position. Once the ligand was in place, the bound state of the FKBP12-ligand complex was inserted in a triclinic simulation box with a volume of ∼125 000 Å^3. The box was then filled with TIP3[46] water molecules in random position at the density of ∼1 g/cm^3. For the **3**−FKBP12, **8**−FKBP12, and Elte421−FKBP12 systems, 3559, 3522, and 3515 solvent molecules, respectively, survived after excluding the water molecules overlapping with the protein−ligand complex. The system underwent a preliminary equilibration of 0.5 ns simulation in the isothermal−isobaric ensemble at T = 300 K and P = 1 atm. Isotropic stress constant pressure was enforced using a modification of the Parrinello−Rahman Lagrangian,[47] and temperature control was achieved using the Nosé thermostat.[48] Electrostatic interactions were computed using the smooth particle mesh Ewald algorithm with convergence parameter set to 0.43 Å^{-1} and grid spacing of 1.6 Å.[49] The equations of motion were integrated using a multiple time-step r-RESPA scheme[50] with a potential subdivision specifically tuned for biomolecular systems in the NPT ensemble.[47,51]

Thirty two (32) replica simulations of the ligand−receptor in the bound state were started from the last structure generated in the preliminary equilibration run. In the GE simulation, the drug was tethered to the protein with the restraint potential

$$V = \frac{1}{s}K\left[(r_{\mathrm{O3-HO_{Tyr82}}} - r_0)^2 + (r_{\mathrm{O2-HN_{Ile56}}} - r_0)^2\right]$$

involving the distances between the carbonyl oxygen atoms O2 and O3 of the ligand and the HO(Tyr82) and HN(Ile56) of FKBP12, with K = 0.01 kcal mol^{-1} Å^{-2} and r_0 = 1.9 Å. Such a tethering potential imposes a ligand concentration of 0.33 M or, equivalently, a ligand-accessible volume of ∼5000 Å^3 (see eq 7), which is ∼3 times the standard accessible volume. The force constant of the restraint potential is weak enough to allow the escape of the ligand in the "hot" replicas (see Figure 1) and sufficiently strong to avoid the diffusion of the ligand in the bulk solvent. Temperature and pressure were kept constant at T = 300 K and P = 1 atm throughout the GE, while the potential function was dependent on the state, according to eq 9. The constancy of the pressure in the GE is important, allowing for molar volume readjustment as the drug−receptor conformation changes from the bound state to the unbound state.

We now come to the definition of the other fundamental ingredient of our method, i.e., the definition of the "solute". Ideally, one should define the entire protein and the ligand as "solute", to allow all of the primary and secondary protein binding sites to contribute to the overall binding free energy. Clearly, the larger is the number of atoms of the protein involved in the scaling, the higher is the number of replica needed to span the selected range of the $c_{Ss}^{(k)}$ and $c_S^{(k)}$ factors. Moreover, by expanding the possible "binding site" to the entire receptor, the restraint potential should also be partially weakened, to let the drug move around the protein and sample all secondary binding site. In this manner, of course, convergence of the method becomes much slower, making the approach impractical. Fortunately, receptors generally have only one important binding site and FKBP12 makes no exception. This immunophilin resembles a cornucopia (see Figure 4, presented somewhat later in this work) with important (i.e., with significant population) binding sites at the edge of the concave region of the horn and secondary binding sites in the mostly hydrophobic and mobile loop Tyr80-His94.[52,53] Based on the above, the solute definition for the FKBP12 ligand complexes includes all residues in the β-sheet structure of the horn entrance and the mobile flexible loop Tyr80-His94, as specified in Figure 4. In total, besides the ligand, 1052 out of 1662 atoms of the protein were involved with the potential scaling of eq 9.

Finally, the "bound" state (i.e., the state for which $H(\mathbf{x}) = 1$) is defined by introducing the contact function

$$F_c(\mathbf{x}) = \frac{1}{N_L}\sum_i^{N_L}\sum_j^{N_P}\mathcal{H}(r_d - r_{ij}(\mathbf{x})) \qquad (13)$$

dx.doi.org/10.1021/ct400809n | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

where N_L and N_p are the number of atoms (including hydrogen atoms) in the ligand and in the protein, respectively; r_{ij} is the distance between the atom i on the ligand and the atoms j on the protein; and $\mathcal{H}(x)$ is the Heaviside step function, such that $\mathcal{H}(x) = 1$ if $x > 0$ and $\mathcal{H}(x) = 0$ if $x < 0$. The atom−atom contact threshold (r_d) is taken to be 4.5 Å. According to eq 13, the function F_c goes from zero (unbound ligand) to the average number of protein contacts per atom of the ligand in the bound state. The function $H(\mathbf{x})$ (see discussion near eq 10) for a given configuration $\mathbf{x}$ in the GE simulation can be defined as

$$H(\mathbf{x}) = \mathcal{H}(F_c(\mathbf{x}) - F_0) \tag{14}$$

where F_0 is a threshold contact function value such as $H(\mathbf{x}) = 0$, for $F_c(\mathbf{x}) < F_0$, i.e., for the unbound state.

The HREM simulations were run for ∼20 ns for the tight-binding ligands (**8** and Elte421) and for 12 ns for the weaker ligand (**3**). All computation were done using the program ORAC.[54,55] In Figure 3, we show the energy distribution along the replica progression for the solute−solvent energy and for the intrasolute energy, as obtained in the last 9 ns HREM simulation of the Elte421-FKBP12 complex with the scaling protocol defined in Figure 4. The overlap of the distributions is significant for each contiguous pair in the replica progression for both energy terms, yielding an average acceptance ratio of 0.6 ± 0.1. The mean state intrasolute energies $V_S(\mathbf{x}_k)$, as shown in Figure 3, increase steadily as k is increased and, correspondingly, with decreasing scaling factor $c_S^{(k)}$. Such steady intrasolute energy increase along the replica progression indicates that the hot (large k) replicas effectively sample solute configurations with high proper and improper torsional energy hence helping in overcoming barriers toward secondary conformational minima. These states can eventually be transmitted, through replicas exchange, to the low-lying replicas with the appropriate canonical weight, hence allowing for a correct evaluation of the contribution of secondary binding sites and of the ligand and binding site reorganization energy to the overall binding free energy through eq 10. The counter-scaled $V_{Ss}(\mathbf{x})$ energy, on the other side, decreases with growing replica index. Therefore, in the states with large k the solute−solvent interaction is stronger (more negative), with an actual temperature felt by these interactions that is *cooler* with respect to the temperature of the target replica. In the 32nd replica, the effective temperature, as far as the solute−solvent interactions are concerned, is 250 K. As discussed before the $V_{Ss}(\mathbf{x})$ scaling, by enhancing the hydrophilicity of the active site and of the ligand, helps to push the drug outside the binding region in the replicas with high k-index, allowing for a more-efficient sampling of the unbound states in the GE simulation.

Residues	C_{SS}^{min}	C_{Ss}^{max}
Lys17-Thr21	0.5	1.2
Val24-Glu60	0.5	1.2
Leu74	0.5	1.2
Ile76	0.5	1.2
Tyr80-Val101	0.5	1.2
Ligand(108)	0.5	1.2

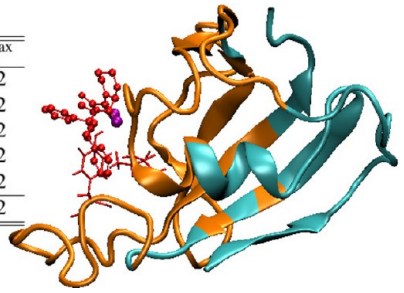


Figure 4. EDU-HREM scaling protocol for the FKBP12−ligand system. On the right, we report the starting structure of the **8**−FKBP12 complex. The protein structure in orange color is scaled. The ligand (also scaled) is in a ball-and-stick representation, while the hydrogen-bonded Tyr86 and Ile56 residues are depicted as bond models. The purple atoms on the ligand are the two hydrogen-bonded oxygen atoms.

In Figure 5, we report the cumulative running time-average, $F_{GE}(t) = \langle F_c(\mathbf{x}(t))\rangle_{GE}$, of the function defined in eq 13 for the FKBP12-Elte421 complex. As it can be seen, in the first half, $F_{GE}(t)$ is steadily decreasing. In this nonequilibration stage of the GE ensemble, the ligand migrates, in the hot replicas, from the binding site of the starting configuration (identical for all states, as shown in the Figure) toward the bulk phase. After an ∼10 ns run, a multicanonical stationary configuration is reached with the drug being gradually more and more bent torward the bulk with increasing replica index (as depicted in Figure 5, right). The FKBP12−Elte421 complex was the system characterized by the longest equilibration time. As we shall see in the next section, this is indeed an early indication of a large binding affinity. The MBAR weight for all points in the GE simulation (655 360 for **3** and 819 200 for **8** and Elte421), have been calculated according to eqs 11 and 12, using the ancillary program MBAR provided in the distribution of the

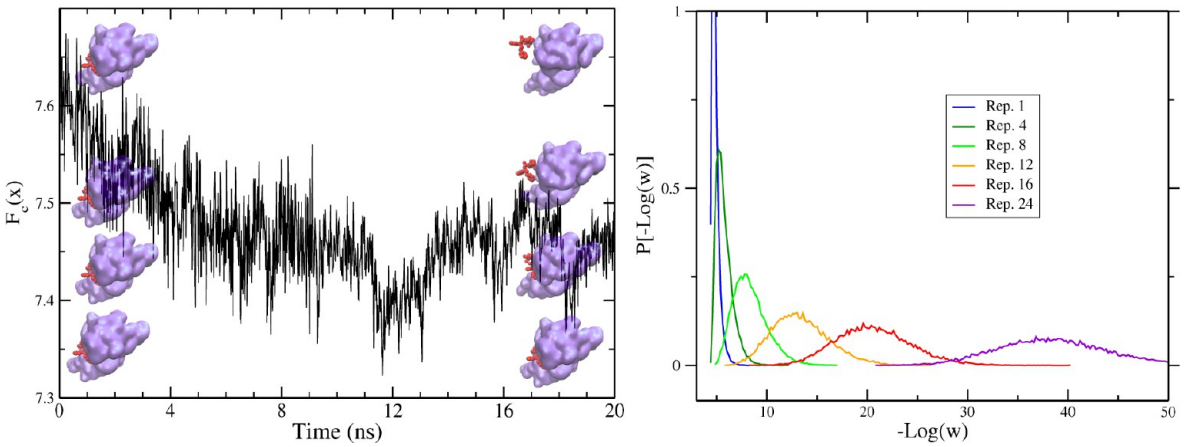


Figure 5. (Left) Time record of the function $F_{GE}(t) = (1/N_{ref})\sum_k F_c(\mathbf{x}_k(t))$ for the EDU-HREM simulation of the Elte421−FKBP12 complex. (Right) Probability distribution of $-\log[W_1(\mathbf{x}_k)]$ (see eq 11), with $W_1(\mathbf{x}_k)$ being the MBAR weights of the *bf**x*** point belonging to state k for the target replica ($k = 1$).

dx.doi.org/10.1021/ct400809n | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

ORAC program.[54] As an example, in Figure 5 (right), we report the distribution of the logarithm of the MBAR weights (see eq 11) for the some representative states in the GE of the Elte421-FKBP12 system, calculated in the last 9 ns run (i.e., when the cumulative running average $F_{GE}(t)$ is stationary; see Figure 5, left). The contribution of a point $\mathbf{x}_k$ in the GE ensemble to the mean of eq 10 is driven by its MBAR weight. The latter is non-negligible if the k-distribution of the logarithm of the MBAR weight overlap significantly with the distribution corresponding to the target replica. As shown in the right-hand panel of Figure 5, for the Elte421−FKBP12 complex, the overlap between the target (k = 1) weight distribution and the kth replica weight distribution becomes negligible for k > 12.

Finally, in Figure 6, we report the histograms of the replica contact functions $F_c(\mathbf{x}_k)$ for the **3**−FKBP12, **8**−FKBP12, and Elte421−FKBP12 complexes, evaluated in the last 9 ns of simulation for a set of representative replicas uniformly sampled from the GE simulation. As it can be seen, for all three tested ligands, the maxima in the distribution of the contact functions are shifted more and more toward smaller values as the replica index k is increased, indicating that, on average, the ligand spends more time outside the binding pocket in the "hot" (large k) replicas, i.e., the states characterized by small and large values of the $c_S^{(k)}$ and $c_{sS}^{(k)}$ scaling constants, respectively. Unbound or nearly unbound states (i.e., state with small F_c) are found in the left tail of the distributions. Remarkably, compound **3** is the ligand that is more easily expelled from the binding site as k is increased, in accordance with the large measured dissociation constant (see Figure 2). Elte421, once more, exhibits a much harder resistance to the solvating force of the scaled potential, with respect to the other ligands, in accordance with its strong affinity (see Figure 2).

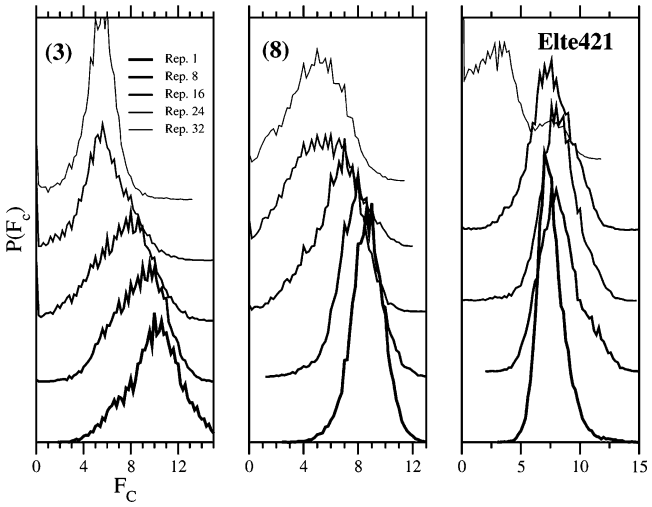


Figure 6. Probability contact function $\delta(F - F_c(\mathbf{x}_k))$ distributions for various replica in the GE simulation for the **3**−FKBP12, **8**−FKBP12, and Elte421−FKBP12 complexes.

4. RESULTS AND DISCUSSION

Ideally in the EDU-HREM method, the binding free energy should be calculated from eq 7, using the GE-averaged p_b/p_u ratio with the threshold contact function F_0 set to zero in the definition described by eq 14. For tight-binding ligands, we have seen that unbound states are extremely rare in conventional simulations, with one or none of these events sampled within tens of nanoseconds. These events are more and more favored by the EDU-HREM protocol with increasing k replica index. Unbound states with $F_c(\mathbf{x}) = 0$ for ligand **3** start to appear already in the k = 4 state and are in the hundreds near replica with k = 12, i.e., ∼0.5% of the sampled configuration per state (see Figure 7, left). Clearly, rare states with $F_c(\mathbf{x}) = 0$ occurring at low replica index k can strongly affect the GE averaged ratio eq 10, especially for tight-binding ligands. A single unbound state sampled at low k may have a MBAR weight exceeding the sum of the MBAR weights of all other unbound sampled states in the GE simulation, leading to strong overestimation of the binding affinity. To overcome this statistical problem, we can compute eq 10 and, hence, the free

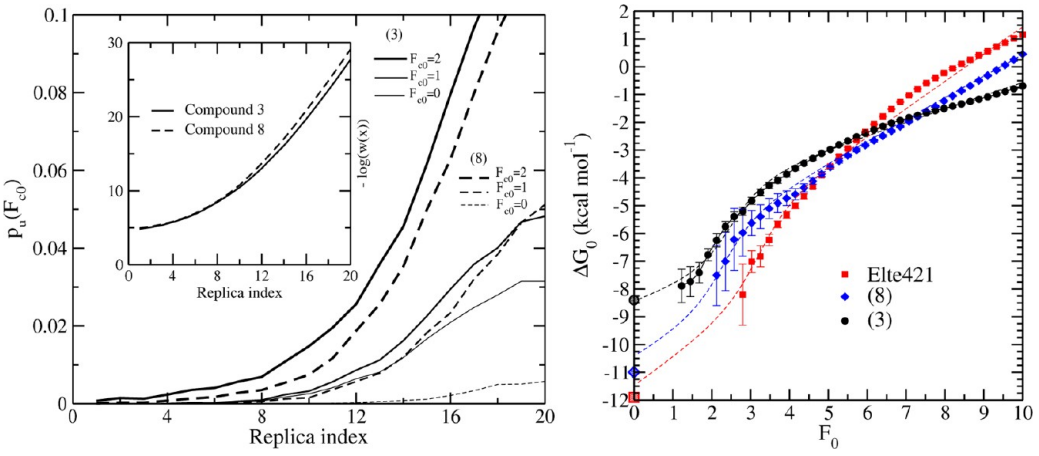


Figure 7. (Left) Probability for the unbound states ($H(\mathbf{x} = 0)$ for compounds **3** and **8**, as a function of the replica index for various threshold values F_c (see text). In the inset, the logarithm of inverse weights are reported as function of the replica index. (Right) Binding free energy for FKBP12 ligands, as a function of the threshold F_0 for counting unbound states (see eq 14)). The lines are the result of a fit using the sigmoid function of eq 15. For compound **3**, we obtain a = 0.47 kcal mol^{-1}, d = 1.65 kcal mol^{-1}, n_c = 2.5, G_0 = −6.9 kcal mol^{-1}. For compound **8**, we obtain a = 0.77 kcal mol^{-1}, d = 1.65 kcal mol^{-1}, n_c = 2.3, and G_0 = −8.9 kcal mol^{-1}. For Elte421, we obtain a = 0.97 kcal mol^{-1}, d = 1.65 kcal mol^{-1}, n_c = 3.2, and G_0 = −9.9 kcal mol^{-1}. Error bars have been calculated using a bootstrap Monte Carlo procedure. Crosshatched symbols at F_0 = 0 correspond to the experimental values of the standard binding free energy.[37,43]

dx.doi.org/10.1021/ct400809n | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

energy via eqs 10 and 7, as a function of the threshold constant F_0 in the definition of $H(\mathbf{x})$ (eq 14). Nearly unbound states or weakly bound states with, e.g., $F_c(\mathbf{x}) = 3$, are, in fact, rather frequent, even at low replica index, allowing for a reliable statistical determination of the ratio presented in eq 10. In Figure 7 (left), we report the probability of an unbound state as a function of the replica index for various values of the threshold F_0. Of course, nonzero values of the threshold F_0 lead to a overestimation of the incidence of the unbound state and, hence, an underestimation of the binding affinity. However, the correct binding free energy can be recovered by extrapolating the function $\Delta G(F_0)$ at $F_0 = 0$. In Figure 7 (right panel), we show the binding free energy $\Delta G(F_0)$ as a function of the threshold F_0 for the three ligands. Points for which $\Delta G(F_0)$ was determined by less than 1000 unbound configurations are not shown. These points are not statistically reliable, as the probability for the unbound states p_u was contributed by very few dominating events with large MBAR weights.

In ligand **3**, the statistical error for all shown points is generally small, although it is noticeably growing as F_0 approaches zero. As above outlined, the reason for such behavior lies in the fact that the unbound states at low F_0 becomes more and more scarce for low replica indexes, hence producing a larger error. This does not appear to be a serious problem for **3** and our EDU-HREM setup for FKBP12 drugs of the strength of that of **3** appears to be, indeed, quite robust. However, the growth of the error as F_0 goes to zero can be a problem for the tighter binders Elte and **8**, as is manifestly shown by the error bars reported in Figure 7 (right). It is reasonable to expect that, for a given ligand, the function $\Delta G(F_0)$, superimposed to the trivial linear trend, should exhibit a progression of the slope up to climax (the detaching event), depending on the shape and size of the ligand and of the binding site and, of course, of the scaling protocol. Indeed, such sigmoidal behavior of the function $\Delta G(F_0)$ is clearly observed for compound **3** possibly implicating that the molecular recognition in the docking process may occur via a learning curve modality. Hence, we may tentatively use the fitting function

$$\Delta G(F_0) = aF_0 + d\,\tanh(F_0 - n_c) - G_0 \qquad (15)$$

to extrapolate the value of ΔG at $F_0 = 0$. For ligand **3**, the fitting of the EDU-HREM data yields a binding free energy of −8.5 to be compared to the experimental value[37] of −8.4 kcal mol^{-1}.

While the agreement between EDU-HREM computation and experimental binding affinity obtained for the relatively weak binder **3** is indeed embarrassingly good, things appear to be more complicated indeed for the potent **8** and Elte421 ligands. As shown in Figure 7, the behavior of the function $\Delta G(F_0)$ discernibly indicates that both **8** and Elte421 are much stronger binders than **3** but for these two ligands the states with $F_0 < 2{:}3$ are very scarce, preventing a clear identification of the inflection point of the sigmoid (see Figure 7, right).

We now make the ansatz that the parameter d, i.e., the sigmoid spread, is a *ligand-independent term* related to the active site shape and volume and to the adopted scaling protocol. Therefore, in tentatively fitting the function $\Delta G(F_c)$ for **8** and Elte421, we use the fixed value of d = 1.65 kcal mol^{-1} obtained for **3**, optimizing the other free parameters a, n_c, and k in eq 15. The extrapolated binding free energy for **8** and Elte421 are −10.5 and −11.5 kcal mol^{-1}, respectively, to be compared with the experimental values[37,42] of −11.0 and −11.9 kcal mol^{-1}, respectively.

Given the quantitative agreement obtained by fitting the EDU-HREM computed $\Delta G(F_c)$ data with the function defined in eq 15, one may wonder whether eq 15 is of general applicability or if it is peculiar for the FK506-related ligands and the FKBP12 system and whether the parameters in the arbitrarily chosen fitting function, eq 15, have some physical meaning. We have seen that the spread of the sigmoid function, d (which is the same for the three ligands), could be related to the nature of the FKBP12 active site. The a parameter, for example, may be connected to $(\partial\Delta G(n)/\partial n)$, i.e., the rate of change of the free energy with increasing ligand−protein atom−atom contact. Then, n_c is the average number of surrounding protein atom (per atom of the ligand) for which the rate $(\partial\Delta G(n)/\partial n)$ is maximal. The parameter G_0, finally, could be controlled by the dimension of the ligand, the latter being linearly related to the binding free energy. Only an extensive analysis on drug−receptor pairs could help elucidate these issues, an analysis that is clearly beyond the scope and aim of the present study.

We now focus on the binding mechanism of FK506-related ligands. As suggested in previous studies, a crucial ingredient in the binding affinity of pipecolic-based ligands is the stabilization, *in bulk solution*, of the pipecolate and α-keto-amide moiety[42,43,45] (see Figure 8) in its correct binding conformation, akin to that found in the macrolide FK506 or Rapamycin.[37] In such "correct" orientation, the ligand experiences a minimal loss of reorganization free energy in passing from the bulk solvent to the bound state in FKBP12, thus significantly enhancing the binding affinity. In Figure 9, we report the free energy surface (FES), with respect to the ω and ψ dihedral angles.

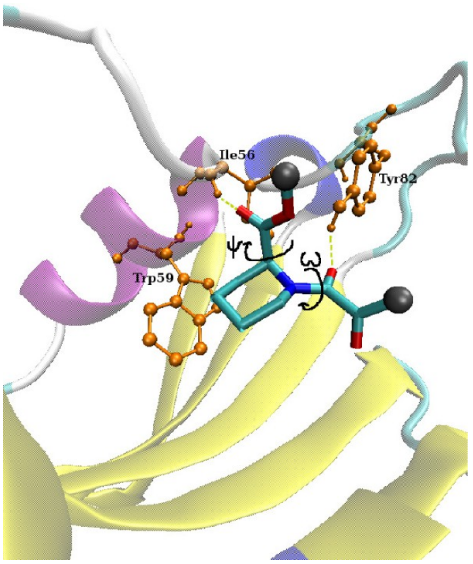


Figure 8. Binding pocket of the complex FKBP12 with a pipecolic FK506-related ligand as obtained from the PDB file 1FKG.[37] The ω and ψ dihedral angles in the ligand are highlighted. The hydrogen-bonded Tyr82 and Ile56 and the vicinal Trp59 are shown in orange color. The hydrogen bonds between HN(Ile56)−O2(sb3) and HO(Tyr82)−O3(sb3) are indicated with light green dashed lines.

The FES is calculated as $G(\omega,\psi) = -RT \ln P(\omega,\psi)$, with $P(\omega,\psi)$ being the corresponding distribution probability calculated in the GE simulation, using the MBAR weights (see eq 11). The maximum of the function $P(\omega,\theta)$ defines the

dx.doi.org/10.1021/ct400809n | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

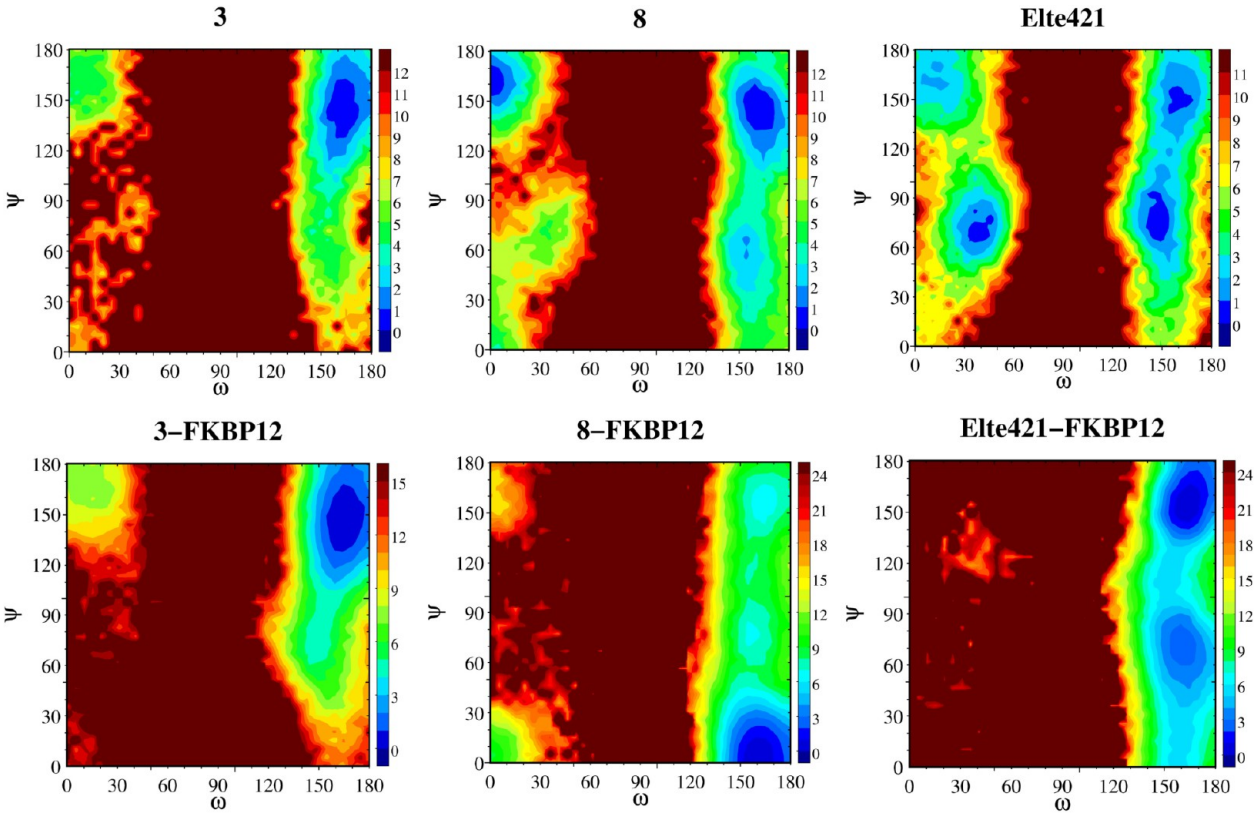


Figure 9. Free energy surface (FES) with respect to the dihedral angles ω and ψ (see Figure 8 for definitions) of compounds **3**, **8**, and Elte421 in bulk (top graphs) and in the complex (bottom graphs). The color-coded energy scale is shown in units of kJ mol^{-1}.

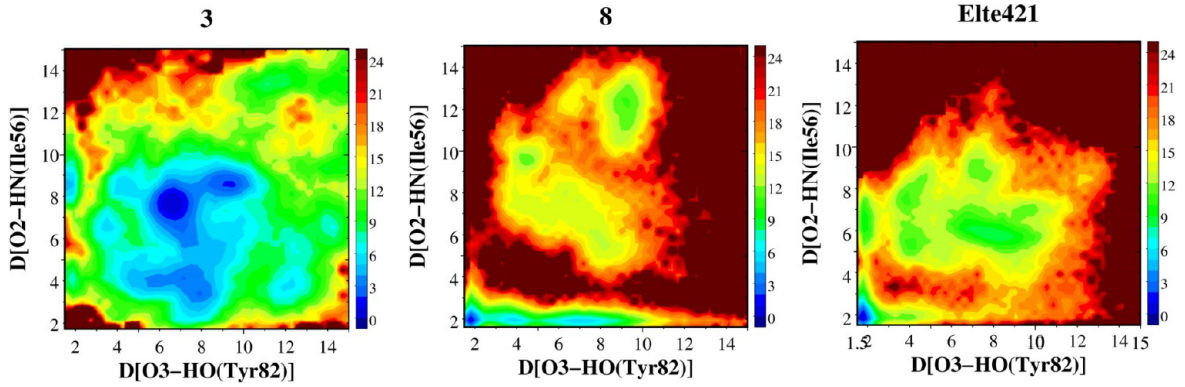


Figure 10. Free energy surface (FES) as a function of the two atom−atom distance HN(Ile56)−O2 and HO(Tyr82)−O3 in the FKBP12 complex with compounds **3**, **8**, and Elte421. The color-coded energy scale is shown in units of kJ mol^{-1}.

zero in the FES. The FES in the top diagrams have been calculated using the data taken from HREM simulations of the ligands in bulk solution.[42] Details of these simulations are given in the Supporting Information of ref 42. The bottom diagrams in Figure 9 have been computed from the present HREM simulations and refer to the ligands when interacting with the FKBP12 protein. The "binding orientation" in the FES lies in the bottom right quadrant of the contour plots, i.e., for ω and ψ in *trans* and *cis* conformations, respectively. With these values of the dihedral angles, the carbonyl moieties in the pipecolate and α-keto-amide core structure (see Figure 8) are optimally exposed for engaging in two hydrogen bonds with HO−Tyr82 and HN−Ile56, as detected in many FKBP12 co-crystals.[36−41] The comparison in Figure 9 of the FES in the bulk and in the complex exposes the importance of conformational preorganization for the binding affinity of FKBP12 ligands.[56] Ligand **3** has an orientation of the pipecolate and α-keto-amide core that is not suited for binding as the ligand in bulk mostly exhibits the ψ dihedral in the quasi-*trans*, rather than *cis* binding conformation. When **3** is bound to the protein, the ligand is surprisingly unable to adjust the conformation so as to establish the HN(Ile56)−O2 and HO(Tyr82)−O3 hydrogen bonds. The absence of the two simultaneous hydrogen bonds in the bound state of **3** is confirmed also in the FES, with respect to the HN(Ile56)−O2 and HO(Tyr82)−O3 distances reported in Figure 10. In the diagram corresponding to **3**, the main basin ($r_{HN(Ile56)-O2}$ = 6:8 Å and $r_{HO(Tyr82)-O3}$ = 6:8 Å) refers to conformations bearing no hydrogen bond at all, while the two HN(Ile56)−O2 and HO(Tyr82)−O3 hydrogen bonds appear to be rather weak and anticorrelated (i.e., when one is formed,

dx.doi.org/10.1021/ct400809n | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

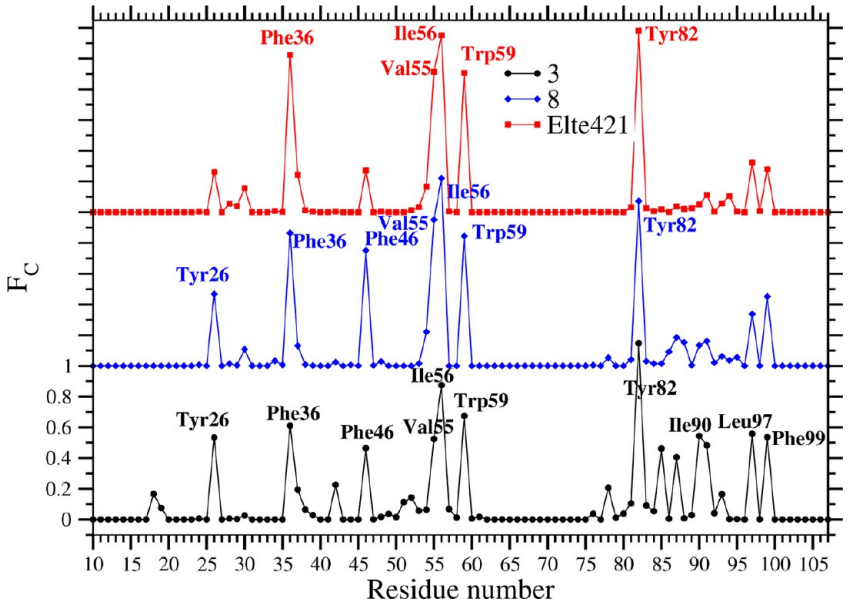


Figure 11. Contribution to the function $F_c(\mathbf{x})$ (see eq 13) from each residue for the various FKBP12 complexes.

the other is not present and vice versa), populating the secondary minima at $r_{HN(Ile56)-O2}$ = 2:4 Å and $r_{HO(Tyr82)-O3}$ = 6:8 Å and at $r_{HN(Ile56)-O2}$ = 7:10 Å and $r_{HO(Tyr82)-O3}$ = 2:3 Å. We have found that, in **3**, the hydrogen-bond acceptor in the ligand can also be the oxygen O4 (see Figure 8) with the donor HO(Tyr82). This "nonstandard" and rather weak ligand−protein hydrogen bond is moderately correlated with the ψ > 90° conformations (data not shown). The lack of the two concurrent HN(Ile56)−O2 and HO(Tyr82)−O3 hydrogen bonds has a negative impact on the binding enthalpy, thus reducing the affinity of **3** for FKBP12.

For compound **8** in bulk solution, a significant amount of conformers populate the binding region if the FES has the following parameters: ω > 150, ψ < 50. As shown in Figure 9, this ability to frequently adopt the optimal binding conformation *in the solvent* induces a dramatic change in the FES(ω,ψ) when compound **8** is bound to the protein, with most of the ligand conformations populating the binding conformational state. Correspondingly, **8** appears to be tightly locked in the active site with the two strong and highly correlated hydrogen bonds HN(Ile56)−O2 and HO(Tyr82)−O3 (see deep minimum at $r_{HN(Ile56)-O2} \simeq 2$ Å and $r_{HO(Tyr82)-O3} \simeq 2$ Å in Figure 10). The conformational preorganization of compound **8**, along with its reduced configurational entropy due to persistent hydrophobic intraligand interactions,[42,43] allows the formation of two strong and highly correlated hydrogen bonds in the bound state, making compound **8** almost 2 orders of magnitude more potent than compound **3**.[37]

As shown in Figure 9, similar to compound **8**, Elte421, in bulk solution, also exhibits a wide and deep free energy conformational minimum in the binding region of pipecolate-α-keto-amide conformational space. This minimum in the Elte421 FES(ω,ψ) becomes even wider and deeper upon binding to FKBP12, although, at variance with compound **8**, states with ψ > 90° are still highly populated in the bound state. This occurrence does not seem to have a significant impact on the capability of Elte421 to establish the two H-bonds involving the pairs HN(Ile56)-O2 and HO(Tyr82)-O3 as shown in Figure 10. Exactly as in **8** and at variance with compound **3**, in the Elte421 FES for these two distances (Figure 10), a clear minimum is seen when *both* of the hydrogen bonds HN(Ile56)−O2 and HO(Tyr82)−O3 are formed. In the ψ > 90° conformations, Elte421 is still able to interact effectively with Tyr82, via stacking interactions with the phenyl ring in the phenyl−hexyl moiety (data not shown).

The strength of the two hydrogen bonds HN(Ile56)−O2 and HO(Tyr82)−O3 in the strong ligands **8** and Elte421 found in our EDU-HREM simulations strongly suggests that the inhibition of the PPIase activity by pipecolic FKBP ligands mainly occurs via *polar* enthalpy driven, rather than hydrophobic, protein−ligand interactions. This result is indeed in agreement with the observation that molecular recognition by FKBP12 FK506-related ligands is invariably based in all known PDB structures of ligand−FKBP complexes[36−40] on the formation two strong hydrogen bonds in the PPI domain involving two highly conserved residues across the FKBP family,[14] namely, Tyr82 and Ile56. However, the prominent role of specific electrostatic interactions in the binding of FKBP12 is seemingly in contrast with the common belief[22,23,57] that, being the binding pocket of the PPI domain mostly hydrophobic, PPIase activity is only driven by solvent-mediated hydrophobic interactions for capturing the Pro-containing peptide. In a site-directed mutagenesis study,[23] in particular, each of the aminoacids Asp37, Arg42, Phe46, Val55, Trp59, and Tyr82 was mutated with the 20 common amino acid residues, finding that "site-specific interactions by the side chains of amino acid residues constituting the substrate-binding cavity were not essential for the PPIase activity, although the 37th, 55th, and 82nd amino acid residues significantly contributed to the activity". Ile56 was not tested in that study. These results were in part contradicted by a recent microbiology assay on a novel FK-506-binding-like protein that lacks peptidyl-prolyl isomerase activity.[58] This inactive FKBP protein exhibits several mutations in residues that have been proposed to mediate PPIase activity in other FKBPs, most of them overlapping with those mutated by Ikura et al., *including* Tyr82 and Ile56 (FKBP12 numbering). Our EDU-HREM simulations provide an opportunity to shed some light in these debated issues. To this end, we can quantitatively assess the contribution of each

dx.doi.org/10.1021/ct400809n | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

protein residue in the binding of FK506-related ligands by decomposing the $F_c(\mathbf{x})$ function of eq 13, as

$$F_c^r(\mathbf{x}) = \frac{1}{N_L}\sum_i^{N_L}\sum_j^{N_p^{(r)}}\mathcal{H}(r_d - r_{ij}) \tag{16}$$

where the index r refers to the rth residue in FKBP12 and $N_p^{(r)}$ is the corresponding number of atoms. The weighted MBAR average on the GE of the function $F_c^{(r)}(\mathbf{x})$ represents the contribution from the rth residue to the overall averaged contact function $\langle F_c(\mathbf{x})\rangle_{GE}$. In Figure 11, we show $\langle F_c^{(r)}(\mathbf{x})\rangle_{GE}$ as a function of the residue number for the three ligands. The three ligands show indeed a very similar pattern. Remarkably, for all the ligands, the five most important amino acids for the binding are invariably Phe36, Val55, Ile56, Trp59, and Tyr82 with $\langle F_c(\mathbf{x})\rangle_{GE} > 0.6$, and with smaller contribution from Phe46 and Tyr26. While the interactions with Phe36, Val55, and Trp59 are certainly hydrophobic in nature, we have seen that this is not so for the other two binding residues, Tyr82 and Ile56, involved in the formation of highly specific hydrogen bonds with the pipecolate α-keto amide moiety. The strong contribution from Ile56 to binding stems, in fact, from the hydrogen bond involving the backbone NH donor and not from the hydrophobic side chain.

The collected data seems to suggest that *both* the polar and hydrophobic interactions are important in immunophilin binding and inhibition with a cooperative mechanism already noted in FKBP12[59,60] and in the parent cyclophilin protein.[24] In compound **8** and Elte421, in particular, the ligand anchors to the FKBP12 binding pocket at Tyr82 (via CO of the flanking the piperidinic ring) and at Ile56 (via the following carbonyl in the peptidomimetic chain) with simultaneous stacking of Trp59 on the piperidinic ring. The two anchoring hydrogen bonds prevent (see discussion of Figure 9), the *trans*−*cis* isomerization of the ω dihedral, locking the ligand in the binding pocket. This inhibition modality suggest that in the Pro *cis*−*trans* isomerization catalysis the substrate (bearing Pro and not the six-membered piperidine moiety) can anchor the concave PPI binding pocket through only one hydrogen bond (probably at Tyr82, via the CO of the flanking Pro residue) and through proline-Trp59 stacking.

The prominent role of the O3-OH(Tyr82) and O2-HN(Ile56) hydrogen bonds in modulating the *enthalpic* contribution in the docking process of FK506-related ligand could be experimentally confirmed by substituting the hydrogen bond donors Ile56 and Tyr82 with negatively charged residues, such as Asp or Glu, possibly detecting a significant drop in the affinity for the mutated protein.

5. CONCLUSIONS AND PERSPECTIVES

We have introduced and described a new computational tool, "Energy Driven Undocking" (EDU-HREM), to calculate the absolute binding free energy for protein ligand association in the context of atomistic simulation in explicit solvent. The methodology is based on solute tempering scheme, enforced via Hamiltonian REM simulation, that allows to compute the dissociation constant using the full generalized ensemble statistics. The solute tempering scaling protocol is engineered to facilitate the extrusion of the ligand into the bulk by enhancing the hydrophilicity of both ligand and active site along the replica progression. An appropriate restrain potential prevents the ligand to diffuse away once detached from the binding site. The proposed technique completely bypasses the need for defining and/or identifying the relevant reaction coordinates in a ligand receptor interactions and allows the calculation of the absolute binding affinity in one single generalized simulation of the drug-receptor system. The algorithm has been applied to the determination of the binding affinities of three FK506-related ligands with known potency. The results presented in this study show that, to the least, the EDU-HREM method is able to discriminate potent binders from weak binders. The computed standard binding free energies of the analyzed ligands are found to be in quantitative agreement with the experimental measurements. The results, on the overall, are very encouraging, but much more work must be invested to improve the scaling protocol with respect to the solute/solvent definition and to tethering potential (i.e., the fixed ligand concentration) and their effect on the convergence of the simulation. We have seen, in fact, that for tight (nanomolar) binders (especially for Elte421), unbound states are sampled too late in the replica progression, yielding very small MBAR weights for the target state. Also, the definition of one part of the system as "solute" and (correspondingly) a second complementary part of the system as "solvent" poses severe limitations on the scaling protocol. The "solvent" may in fact include portion of the protein that can act as spurious secondary binding sites, as the ligand feels these unscaled parts of the receptor as "solvent", strongly interacting with them. In order to prevent this unwanted effect, a large part of the protein must be included in the "solute" definition. Also, due care must be taken in the tuning of the restraint potential in order to avoid contacts with distal and unscaled portions of the protein, once the ligand leaves the binding pocket to join the "solvent". The large size of the solute in the dual solute−solvent system has a direct impact of the number of replicas that one must introduce for spanning a given range of scaled potentials.

A possible solution to these shortcomings of the EDU-HREM technique could be that of expanding the partitioning of the systems, abandoning the dual solute−solvent representation, to build a more complex and effective scaling protocol. For example, one can lump the atoms of the protein in an active site and in an inert (unscaled) part, thus avoiding the insurgence of spurious binding sites far away from the binding pocket. Work in this direction is underway in our laboratory.

AUTHOR INFORMATION

Corresponding Author

*E-mail: procacci@unifi.it.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Early tests and the tuning of the EDU-HREM algorithm for FKBP12 ligands were performed using the high-performance facilities provided by the Computational Research Center for Complex Systems (CRESCO) of the Italian Agency for New Technologies (ENEA). The production part of the calculations was done on the FERMI Blue Gene/Q architecture provided under the Italian Super-Computing Resource Allocation program (ISCRA) (Call B Project No. FKBP12-LHP10BDV7I7) released by the Italian Supercomputing Center of CINECA. We also thank Regione Toscana for financial support in the initial stage of the project.

REFERENCES

(1) Gilson, M.; Given, J.; Bush, B.; McCammon, J. *Biophys. J.* **1997**, *72*, 1047−1069.
(2) Gumbart, J. C.; Roux, B.; Chipot, C. *J. Chem. Theor. Comput.* **2013**, *9*, 974−802.
(3) Deng, Y.; Roux, B. *J. Phys. Chem. B* **2009**, *113*, 2234−2246.
(4) Woo, H.-J.; Roux, B. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6825−6830.
(5) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562−12566.
(6) Barducci, A.; Bussi, G.; Parrinello, M. *Phys. Rev. Lett.* **2008**, *100*, 020603.
(7) Fidelak, J.; Juraszek, J.; Branduardi, D.; Bianciotto, M.; Gervasio, F. L. *J. Phys. Chem. B* **2010**, *114*, 9516−9524.
(8) Vargiu, A. V.; Ruggerone, P.; Carloni, A. M. P. *Nucleic Acids Res.* **2008**, 5910−5921.
(9) Patel, J. S.; Branduardi, D.; Masetti, M.; Rocchia, W.; Cavalli, A. *J. Chem. Theory Comput.* **2011**, *7*, 3368−3378.
(10) Colizzi, F.; Perozzo, R.; Scapozza, L.; Recanatini, M.; Cavalli, A. *J. Am. Chem. Soc.* **2010**, *132*, 7361−7371.
(11) Favia, A. D.; Masetti, M.; Recanatini, M.; Cavalli, A. *PLos One* **2011**, *6*, e25375.
(12) Nicolini, P.; Frezzato, D.; Gellini, C.; Bizzarri, M.; Chelli, R. *J. Comput. Chem.* **2013**, *34*, 1561−1576.
(13) Hummer, G. *Nat. Chem.* **2010**, *2*, 906−906.
(14) Blackburn, E. A.; Walkinshaw, M. D. *Curr. Op. Pharm.* **2011**, *11*, 365−371.
(15) Chattopadhaya, S.; Harikishore, A.; Yoon, H. S. *Curr. Med. Chem.* **2011**, *18*, 5380−5397.
(16) Galat, A. *J. Chem. Inf. Model.* **2008**, *48*, 1118−1130.
(17) Sugata, H.; Matsuo, K.; Nakagawa, T.; Takahashi, M.; Mukai, H.; Ono, Y.; Maeda, K.; Akiyama, H.; Kawamata, T. *Neurosci. Lett.* **2006**, *350*, 472−477.
(18) Liu, F.; Liu, P.; Shao, H.; Kung, F. *Biochem. Biophys. Res. Commun.* **2006**, *350*, 472−477.
(19) Gerard, M.; Deleersnijder, A.; Daniels, V.; Schreurs, S.; Munck, S.; Reumers, V.; Pottel, H.; Engelborghs, Y.; Van den Haute, C.; Taymans, J.; Debyser, Z.; Baekelandt, V. *J. Neurosci.* **2010**, *30*, 2454−2463.
(20) Lisi, L.; Navarra, P.; Cirocchi, R.; Sharp, A.; Stigliano, E.; Feinstein, D. L.; Dello Russo, C. *J. Neuroimmun.* **2012**, *243*, 43−51.
(21) Lin, J. T.; Stein, E. A.; Wong, M. T.; Kalpathy, K. J.; Su, L. L.; Utz, P. J.; Robinson, W. H.; Fathnnan, C. G. *Clin. Immunol.* **2012**, *142*, 127−138.
(22) Park, S. T.; Aldape, R. A.; OlgaFuter; DeCenzo, M. T.; Livingston, D. J. *J. Biol. Chem.* **1992**, *267*, 3316−3324.
(23) Ikura, T.; Ito, N. *Protein Sci.* **2007**, *16*, 2618−2625.
(24) Hamelberg, D.; McCammon, J. A. *J. Am. Chem. Soc.* **2009**, *131*, 147−152.
(25) Leone, V.; Lattanzi, G.; Molteni, C.; Carloni, P. *PLos Comput. Biol.* **2009**, *5*, e1000309.
(26) Shirts, M. R.; Chodera, J. D. *J. Chem. Phys.* **2008**, *129*, 124105.
(27) Zhou, H.-X.; Gilson, M. K. *Chem. Rev.* **2009**, *109*, 4092−4107.
(28) Shaw, D. E.; et al. *Commun. ACM (ACM)* **2008**, *51*, 91−97.
(29) Hermans, J.; Shankar, S. *Isr. J. Chem.* **1986**, *27*, 225−227.
(30) Zhou, H.-X. *Biophys. J.* **2006**, *91*, 3170−3181.
(31) Zhou, H.-X. *J. Am. Chem. Soc.* **2001**, *123*, 6730−6731.
(32) Zhou, H.-X. *Biochemistry* **2001**, *40*, 15069−15073.
(33) Yang, C.-Y.; Sun, H.; Chen, J.; Nikolovska-Coleska, Z.; Wang, S. *J. Am. Chem. Soc.* **2009**, *131*, 13709−13721.
(34) Luo, H.; Sharp, K. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 10399−10404.
(35) Gallicchio, E.; Lapelosa, M.; Levy, R. M. *J. Chem. Theory Comput.* **2010**, *6*, 2961−2977.
(36) Van-Duyne, G.; Standaert, R. F.; Schreiber, S. L.; Clardy, J. *J. Am. Chem. Soc.* **1991**, *113*, 7433−7434.
(37) Holt, D. A.; Luengo, J. J. I.; Yamashita, D. S.; Oh, H. J.; Konialian, A. L.; Yen, H. K.; Rozamus, L. W.; Brandt, M.; Bossard, M. J.; Levy, M. A.; Eggleston, D. S.; Liang, J.; Schultz, L. W.; Stout, T. J.; Clardyt, J. *J. Am. Chem. Soc.* **1993**, *115*, 9925−9938.
(38) Griffith, J.; Kim, J.; Kim, E.; Sintchak, M.; Thomson, J.; Fitzgibbon, M.; Fleming, M.; Caron, P.; Hsiao, K.; Navia, M. *Cell* **1995**, *82*, 507−522.
(39) Kotaka, M.; Ye, H.; Alag, R.; Hu, G.; Bozdech, Z.; Preiser, P.; Yoon, H.; Lescar, J. *Biochemistry* **2008**, *47*, 5951−5961.
(40) Alag, R.; Qureshi, I.; Bharatham, N.; Shin, J.; Lescar, J.; Yoon, H. S. *Protein Sci.* **2010**, *19*, 1577−1586.
(41) Gopalakrishnan, R.; Kozany, C.; Gaali, S.; Kress, C.; Hoogeland, B.; Bracher, A.; Hausch, F. *J. Med. Chem.* **2012**, *55*, 4114−4122.
(42) Bizzarri, M.; Tenori, E.; Martina, M. R.; Marsili, S.; Caminati, G.; Menichetti, S.; Procacci, P. *J. Phys. Chem. Lett.* **2011**, *2*, 2834−2839.
(43) Martina, M. R.; Tenori, E.; Bizzarri, M.; Menichetti, S.; Caminati, G.; Procacci, P. *J. Med. Chem.* **2013**, *56*, 1041−1051.
(44) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct. Funct. Bioinf.* **2006**, *65*, 712−725.
(45) Bizzarri, M.; Marsili, S.; Procacci, P. *J. Phys. Chem. B* **2011**, *115*, 6193−6201.
(46) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926−935.
(47) Marchi, M.; Procacci, P. *J. Chem. Phys.* **1998**, *109*, 5194−5202.
(48) Nose, S. *J. Chem. Phys.* **1984**, *81*, 511−519.
(49) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *101*, 8577−8593.
(50) Tuckerman, M.; Berne, B. J.; Martyna, G. *J. Chem. Phys.* **1992**, *97*, 1990−2001.
(51) Procacci, P.; Paci, E.; Darden, T.; Marchi, M. *J. Comput. Chem.* **1997**, *18*, 1848−1862.
(52) Banaszynski, L.; C.W., L.; Wandless, T. *J. Am. Chem. Soc.* **2004**, *127*, 4715−4721.
(53) Liang, J.; Cho, i. J.; Clardy, J. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1999**, *55*, 736−744.
(54) Marsili, S.; Signorini, G. F.; Chelli, R.; Marchi, M.; Procacci, P. *J. Comput. Chem.* **2010**, *31*, 1106−11161.
(55) Procacci, P.; Marchi, M.; Marsili, S.; Signorini, G.; Chelli, R. Orac, release 5, Rev. 3.1. The orac suite can be freely downloaded at http://www.chim.unifi.it/orac as a compressed tar archive. The suite includes the MD parallel engine (including the source code), accessory analysis tools for examining the conformational behavior of biological molecules and related documentation.
(56) Wang, Y.; Kirschner, A.; Fabian, A.-K.; Gopalakrishnan, R.; Kress, C.; Hoogeland, B.; Koch, U.; Kozany, C.; Bracher, A.; Hausch, F. *J. Med. Chem.* **2013**, *56*, 3922−3935.
(57) Armistead, D. M.; Badia, M. C.; Deininger, D. D.; Duffy, J. P.; Saunders, J. O.; Tung, R. D.; Thomson, J. A.; Decenzo, M. T.; Futer, O.; Livingston, D. J.; Murcko, M. A.; Yamashitaand, M. M.; Navia, M. A. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1995**, *D51*, 522−528.
(58) Norville, I. H.; Breitbach, K.; Eske-Pogodda, K.; Harmer, N. J.; Sarkar-Tyson, M.; Titball, R. W.; Steinmetz, I. *Microbiology* **2011**, 2629−2638.
(59) Lamb, M. L.; W, L. J. *J. Med. Chem.* **1998**, *41*, 3928−3939.
(60) Lamb, M.; Tirado-Rives, J.; Jorgensen, W. L. *Bioorg. Med. Chem.* **1999**, *7*, 851−860.