Supporting Information to "Towards quantitative estimates of binding affinities for protein-ligand systems involving large inhibitor compounds: a steered molecular dynamics simulation route"

Paolo Nicolini[♮], Diego Frezzato[‡], Cristina Gellini[†], Marco Bizzarri[†], and Riccardo Chelli[†*]
[♮]*Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord B4-B5, E-08034 Barcelona, Spain*
[‡]*Dipartimento di Scienze Chimiche, Università di Padova, Via Marzolo 1, I-35131 Padova, Italy*
[†]*Dipartimento di Chimica, Università di Firenze, Via della Lastruccia 3, I-50019 Sesto Fiorentino, Italy and European Laboratory for Non-linear Spectroscopy (LENS), Via Nello Carrara 1, I-50019 Sesto Fiorentino, Italy*
(Dated: February 28, 2013)

Notes

(i) In this report, any reference to Equations, Figures, Tables and Sections is intended to be made to those of the report itself, unless noted otherwise.

(ii) The references cited here are reported at the end of the report.

I. FORCE FIELD

A. Missing parameters

The protein is modeled using the AMBER03 force field[1, 2], gathering the missing parameters from the available data by similarity. The correspondences for the various types of potential are reported in Tables I, II, III and IV. The notation used for denoting the atom types is from AMBER03[1, 2].

Stretchings	AMBER03
CA - OS	CM - OS
N2 - CQ	N2 - CA
CK - NC	CK - NB
CB - CK	CB - CM
C* - CV	C* - CA
C* - HA	CM - HA
CV - N*	CA - N*

TABLE I: Stretching types that are missing in the AMBER03 force field (left column). In the right column, we report the stretching types of the AMBER03 force field[3] that we assign to the missing ones.

Bendings	AMBER03
CA - OS - CT	CM - OS - CT$^{(1)}$
CA - CA - OS	CM - CM - OS$^{(1)}$
N2 - CA - CA	N2 - CA - CM$^{(1)}$
CQ - N2 - CA	CM - N2 - CA$^{(1)}$
N2 - CQ - NC	NC - CQ - NC$^{(1)}$
H - N2 - CQ	H - N2 - CM$^{(1)}$
CV - N* - CB	CK - N* - CB$^{(1)}$
CB - N* - CA	CB - N* - C$^{(1)}$
CV - N* - CA	CB - N* - C$^{(1)}$
N* - CK - NC	N* - CA - NC$^{(1)}$
N* - CK - CB	N* - CA - CB$^{(1)}$
N* - CV - C*	N* - CA - C*$^{(1)}$
N* - CV - H4	N* - CA - H4$^{(1)}$
N* - CA - CA	N* - CA - CC$^{(1)}$
CV - C* - CB	CA - C* - CB$^{(1)}$
CQ - NC - CK	CA - NC - CA$^{(1)}$
CB - CK - NC	CB - CA - NC$^{(1)}$
NC - CA - H4	N* - CA - H4$^{(1)}$
C* - CV - H4	C* - CA - H4$^{(1)}$
CV - C* - HA	CA - CM - HA$^{(1)}$
NC - CA - CA	NC - CA - CM$^{(1)}$
CB - C* - HA	CA - CM - HA$^{(1)}$
CA - CB - CK	CA - CB - CN$^{(2)}$
C* - CB - CB	C* - CB - CA$^{(2)}$
CA - C - O2	CT - C - O2$^{(2)}$
CA - CT - H1	CA - CT - HC$^{(2)}$
C - CT - CA	C - CT - CT$^{(2)}$
NC - CA - CT	C - CT - CT$^{(2)}$

TABLE II: Bending types that are missing in the AMBER03 force field (left column). In the right column, we report the bending types of the AMBER03 force field that we assign to the missing ones. (1): see Ref. [3]; (2): see Ref. [4].

B. Atomic charges

In Tables V, VI and VII, we report the atomic charges and the atomic type assignments (on the basis of the AMBER03 force field) for the *16i*, *17g* and *32* ligands whose structural formulas are reported in Figs. 1, 2 and 3, respectively.

*Electronic address: riccardo.chelli@unifi.it

Proper Torsions	AMBER03
CB - N* - CA - NC	CQ - NC - CA - N*
CV - N* - CA - NC	CQ - NC - CA - N*
CB - N* - CA - CA	C - NA - CA - CM
CV - N* - CA - CA	C - NA - CA - CM
CA - N* - CV - H4	CA - N* - CM - H4
C* - CV - N* - CA	C* - CA - N* - CB
CB - C* - CV - H4	CB - C* - CA - H4
CB - N* - CV - H4	CA - N* - CM - H4
H4 - CV - C* - HA	H4 - CM - CM - HA
H - N2 - CQ - NC	H - N2 - CA - NC
N* - CK - NC - CQ	N* - CA - NC - CQ
C* - CB - CK - NC	C* - CB - CB - NC
CA - CB - CK - NC	CA - CB - CB - NC
CA - N2 - CQ - NC	CA - NC - CQ - NC
CB - CK - NC - CQ	CB - CB - NC - CQ
C* - CV - N* - CB	C* - CA - N* - CB
N* - CV - C* - CB	N* - CA - C* - CB
N* - CK - CB - C*	N* - CB - CB - C*
N* - CK - CB - CA	N* - CB - CB - CA
N* - CV - C* - HA	N* - CM - CM - HA

TABLE III: Proper-torsion types that are missing in the AMBER03 force field (left column). In the right column, we report the proper-torsion types of the AMBER03 force field[3] that we assign to the missing ones.

Improper Torsions	AMBER03
CA - CA - CA - OS	C - CM - CM - OS
CA - CA - CA - N2	CT - CZ - CM - N2
N2 - NC - CQ - NC	CT - NC - CQ - NC
CB - NC - CB - N	CT - NC - CQ - NC
CB - CV - C* - HA	CA - CM - CM - HA
CV - CB - N* - CA	CA - CB - N* - CT
CA - NC - CA - N	CT - NC - CQ - NC
CA - CB - CB - C*	CA - CA - CA - CT
CA - CK - CB - C*	CA - CA - CA - CT
CB - NC - CB - N*	CB - N2 - CA - NC
CA - NC - CA - N*	CB - N2 - CA - NC
CA - CA - CA - C	CA - CA - CA - CT
NC - CA - CA - CT	CA - CA - CA - CT

TABLE IV: Improper-torsion types that are missing in the AMBER03 force field (left column). In the right column, we report the improper-torsion types of the AMBER03 force field[3] that we assign to the missing ones.

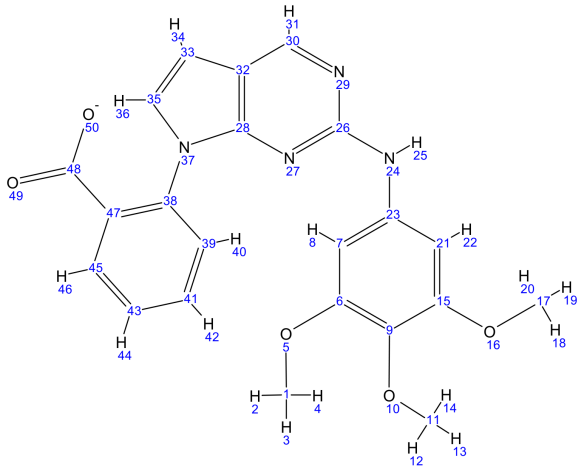

FIG. 1: Structural formula of the *16i* ligand[5].

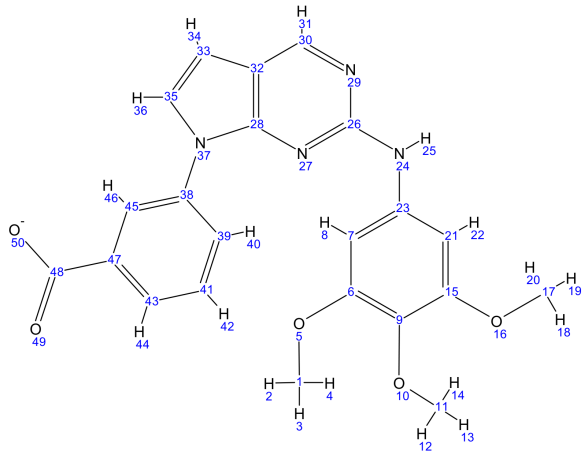

FIG. 2: Structural formula of the *17g* ligand[5].

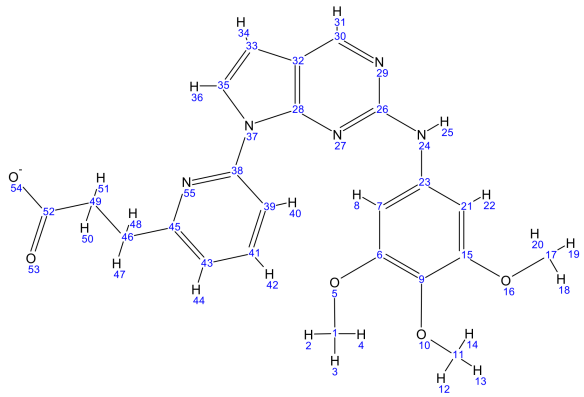

FIG. 3: Structural formula of the *32* ligand[5].

Ligand *16i*								
Label	Type	Charge	Label	Type	Charge	Label	Type	Charge
C1	CT	-0.106089	H18	H1	0.052610	C35	CV	-0.161949
H2	H1	0.096965	H19	H1	0.052610	H36	H4	0.220536
H3	H1	0.096965	H20	H1	0.052610	N37	N*	0.033143
H4	H1	0.096965	C21	CA	-0.692283	C38	CA	0.038264
O5	OS	-0.274320	H22	HA	0.242234	C39	CA	-0.196439
C6	CA	0.466804	C23	CA	0.632976	H40	HA	0.125770
C7	CA	-0.679344	N24	N2	-0.828815	C41	CA	-0.128426
H8	HA	0.312342	H25	H	0.377867	H42	HA	0.119664
C9	CA	-0.190583	C26	CQ	1.135574	C43	CA	-0.181238
O10	OS	-0.290843	N27	NC	-0.804379	H44	HA	0.125463
C11	CT	-0.075365	C28	CB	0.579039	C45	CA	-0.175817
H12	H1	0.078926	N29	NC	-0.885301	H46	HA	0.151415
H13	H1	0.078926	C30	CA	0.377548	C47	CA	-0.006811
H14	H1	0.078926	H31	H4	0.091637	C48	C	0.815993
C15	CA	0.496820	C32	CB	-0.316161	O49	O2	-0.770320
O16	OS	-0.402726	C33	C*	-0.328029	O50	O2	-0.770320
C17	CT	0.048205	H34	HA	0.188756			

TABLE V: Atom labels (refer to Fig. 1 to view the position in the molecule), atom type assignment in AMBER03 notation and charges for the *16i* ligand. Charges are in units of e.

Ligand *17g*								
Label	Type	Charge	Label	Type	Charge	Label	Type	Charge
C1	CT	-0.143894	H18	H1	0.050941	C35	CV	-0.190810
H2	H1	0.118953	H19	H1	0.050941	H36	H4	0.196767
H3	H1	0.118953	H20	H1	0.050941	N37	N*	-0.082394
H4	H1	0.118953	C21	CA	-0.634114	C38	CA	0.242479
O5	OS	-0.273593	H22	HA	0.230186	C39	CA	-0.317663
C6	CA	0.428274	C23	CA	0.479985	H40	HA	0.151516
C7	CA	-0.562551	N24	N2	-0.708319	C41	CA	-0.082889
H8	HA	0.258288	H25	H	0.355212	H42	HA	0.113010
C9	CA	-0.155539	C26	CQ	1.109670	C43	CA	-0.242755
O10	OS	-0.294515	N27	NC	-0.804801	H44	HA	0.157500
C11	CT	-0.081753	C28	CB	0.607497	C45	CA	-0.285891
H12	H1	0.083096	N29	NC	-0.888569	H46	HA	0.169445
H13	H1	0.083096	C30	CA	0.412768	C47	CA	0.076947
H14	H1	0.083096	H31	H4	0.089167	C48	C	0.798864
C15	CA	0.476234	C32	CB	-0.337435	O49	O2	-0.781403
O16	OS	-0.407727	C33	C*	-0.298982	O50	O2	-0.781403
C17	CT	0.055747	H34	HA	0.188475			

TABLE VI: Atom labels (refer to Fig. 2 to view the position in the molecule), atom type assignment in AMBER03 notation and charges for the *17g* ligand. Charges are in units of e.

Ligand *32*								
Label	Type	Charge	Label	Type	Charge	Label	Type	Charge
C1	CT	0.016214	H20	H1	0.069991	C38	CA	0.727822
H2	H1	0.065552	C21	CA	-0.653225	C39	CA	-0.658566
H3	H1	0.065552	H22	HA	0.233147	H40	HA	0.254693
H4	H1	0.065552	C23	CA	0.502942	C41	CA	0.236557
O5	OS	-0.377755	N24	N2	-0.720111	H42	HA	0.085511
C6	CA	0.435929	H25	H	0.352866	C43	CA	-0.620952
C7	CA	-0.564671	C26	CQ	1.086713	H44	HA	0.192359
H8	HA	0.251322	N27	NC	-0.794018	C45	CA	0.652630
C9	CA	-0.175891	C28	CB	0.570332	C46	CT	-0.090296
O10	OS	-0.310052	N29	NC	-0.869760	H47	HC	0.013864
C11	CT	-0.031745	C30	CA	0.354160	H48	HC	0.013864
H12	H1	0.073994	H31	H4	0.098879	C49	CT	-0.106333
H13	H1	0.073994	C32	CB	-0.238192	H50	H1	-0.001927
H14	H1	0.073994	C33	C*	-0.374171	H51	H1	-0.001927
C15	CA	0.511114	H34	HA	0.207181	C52	C	0.859747
O16	OS	-0.322451	C35	CV	-0.076196	O53	O2	-0.824449
C17	CT	-0.030037	H36	H4	0.262461	O54	O2	-0.824449
H18	H1	0.069991	N37	N*	-0.214894	N55	NC	-0.666847
H19	H1	0.069991						

TABLE VII: Atom labels (refer to Fig. 3 to view the position in the molecule), atom type assignment in AMBER03 notation and charges for the *32* ligand. Charges are in units of e.

II. EXTRUSION PATHWAYS OF THE LIGANDS FROM THE FAK BINDING SITE

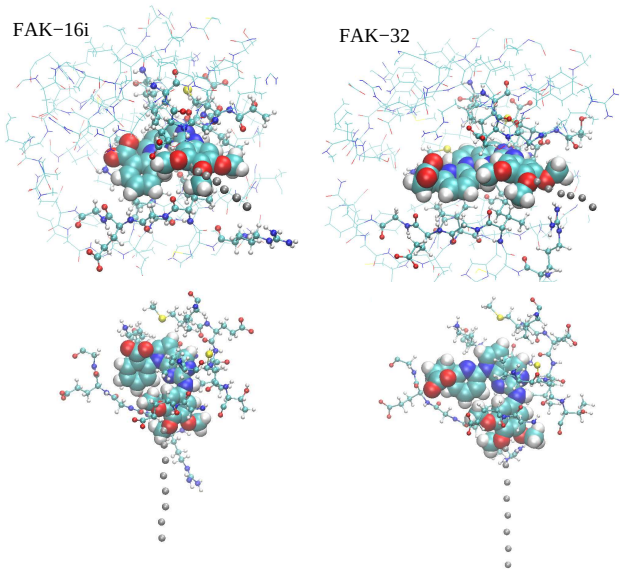

FIG. 4: Representation of the path of extrusion of the ligands *16i* and *32* from the FAK binding site (silver 3D balls). The ligands are represented in space-fill mode. The residues forming the binding sites are represented in ball and stick mode. For the *16i* ligand, the following residues are shown: Arg426, Ile428, Gly429, Glu430, Val436, Ala452, Lys454, Val484, Met499, Glu500, Leu501, Cys502, Thr503, Gly505, Glu506, Leu553. For the *32* ligand, the same residues plus Gly431 are shown. The residues forming the binding site are determined as explained in the main text. The protein atoms farther than 12 Å from the ligand are reported in thin-line mode. Water molecules are not shown. Two different perspectives are displayed.

III. RELATION BETWEEN BINDING CONSTANTS AND POTENTIAL OF MEAN FORCE

The purpose is to formally relate the potential of mean force (PMF), constructed via SMD simulations, to the binding equilibrium constants which are experimentally determinable. Our derivation is inspired by the work of Moro and Severin[6], aimed to express kinetic constants of bimolecular elementary reactions from the stochastic dynamics of the reactant molecules. A side product of that research was the derivation of the equilibrium constant for the forward/backward global process. Although many of the analytical issues, there developed for a model of particles interacting through Morse potential, are here lost due to chemical complexity and anisotropy of both ligand and protein molecules, the main features are kept and elaborated in the following treatment.

Be "A" the FAK protein, "B" the ligand, and "AB" the complex (all species are meant to be solvated in the medium at the equilibrium composition). The binding equilibrium is

$$\mathrm{A} + \mathrm{B} \rightleftharpoons \mathrm{AB}$$

for which $K_{\mathrm{eq}} = a_{\mathrm{AB}}\, a_{\mathrm{A}}^{-1}\, a_{\mathrm{B}}^{-1}$ is the thermodynamic equilibrium constant where $a_i = \gamma_i m_i / m^{\ominus}$ are the concentration-dependent activities of the species expressed according to the "model of dilute solutions" for the chemical potentials, m_i the molalities (with $m^{\ominus} = 1$ mol kg^{-1} the standard molality), and γ_i the activity coefficients. For dilute solutions, the relation between K_{eq} and the commonly adopted constant K_{c} based on molar concentrations $[i]$ is

$$K_{\mathrm{c}} = \frac{[\mathrm{AB}]_{\mathrm{eq}}}{[\mathrm{A}]_{\mathrm{eq}}[\mathrm{B}]_{\mathrm{eq}}} \simeq K_{\mathrm{eq}} \frac{\gamma_{\mathrm{A}}\gamma_{\mathrm{B}}}{\gamma_{\mathrm{AB}}} \left(\rho_{\mathrm{solv}} m^{\ominus}\right)^{-1} \tag{1}$$

where we have used $[i] \simeq \rho_{\mathrm{solv}} m_i$ with ρ_{solv} the density of the solvent. Contrary to K_{eq}, K_{c} bears a concentration dependence through the γ_i. The reference situation is that of "infinite dilution" where all activity coefficients tend to unity and

$$K_{\mathrm{c}}^{\infty} = K_{\mathrm{eq}} \left(\rho_{\mathrm{solv}} m^{\ominus}\right)^{-1} \tag{2}$$

Let us consider a sample portion of volume V containing N_{A} and N_{B} molecules of protein and ligand, respectively. The actual composition be specified by the collective variable $\Xi \equiv (v_{\mathrm{B}}, \epsilon_{\mathrm{AB}})$ where $v_{\mathrm{B}} = V/N_{\mathrm{B}}$ is the volume per ligand molecule and $\epsilon_{\mathrm{AB}} = N_{\mathrm{A}}/N_{\mathrm{B}}$. The specific volume v_{B} is used to specify the degree of dilution (v_{B} grows as the solution is more diluted at fixed number of molecules). A single protein molecule can be present either in free (unbound state, label "u" in the following) and complexed (bound state, label "b") forms. It has to be stressed that *bound state* terms nothing but the whole collection of internal/mutual configurations of A and B at the molecular level to be *associated* to what we conceive as the AB complex in the specific case; the same for the *unbound state*, for which the two molecules are considered to fall apart and to be separately solvated. Below, we shall give some more formal insight. Now consider the equilibrium population of the bound state, $P_{\mathrm{b,eq}}(\Xi)$ (in practice, the probability to pick, at equilibrium, a generic protein A and its closest ligand B configured in a way to be seen as complexed), and the counterpart $P_{\mathrm{u,eq}}(\Xi) = 1 - P_{\mathrm{b,eq}}(\Xi)$ for the unbound state. Notice that the populations depend on Ξ since the mean-field (thermally averaged) A-B interactions are in principle affected by the composition of the medium. Molar concentrations are then expressed as

$$[\mathrm{AB}]_{\mathrm{eq}} = \frac{N_{\mathrm{A}} P_{\mathrm{b,eq}}(\Xi)}{N_{\mathrm{Av}} V}, \tag{3}$$

$$[\mathrm{A}]_{\mathrm{eq}} = \frac{N_{\mathrm{A}} P_{\mathrm{u,eq}}(\Xi)}{N_{\mathrm{Av}} V}, \tag{4}$$

$$[\mathrm{B}]_{\mathrm{eq}} = \frac{N_{\mathrm{B}} - N_{\mathrm{A}} P_{\mathrm{b,eq}}(\Xi)}{N_{\mathrm{Av}} V} \tag{5}$$

where N_{Av} is the Avogadro number. Substitution of Eqs. 3, 4 and 5 into Eq. 1 yields

$$K_{\mathrm{c}}(\Xi) = N_{\mathrm{Av}} v_{\mathrm{B}} \, \frac{P_{\mathrm{b,eq}}(\Xi)}{P_{\mathrm{u,eq}}(\Xi)[1 - P_{\mathrm{b,eq}}(\Xi)\epsilon_{\mathrm{AB}}]} \tag{6}$$

The evaluation of $K_{\mathrm{c}}(\Xi)$, and then of the relevant limit $K_{\mathrm{c}}^{\infty} = \lim_{v_{\mathrm{B}} \to \infty} K_{\mathrm{c}}(\Xi)$, requires a statistical approach to make explicit the populations of the bound/unbound states. To simplify the notation, in the following the symbol "Ξ" will be omitted by leaving implicit that most quantities depend on composition up to take the "infinite dilution" limit at the very final stage.

Let us introduce a freely chosen Cartesian reference frame attached to a ligand molecule (the "B-frame", BF), and an analogous frame tethered to the protein (the "A-Frame", AF). Be $\mathbf{r}$ the positional vector which specifies the location of the ligand (center of BF) with respect to the AF. Then we introduce, in all generality, a collection $\mathbf{a}$ of structural degrees of freedom of the protein aimed to specifies its internal conformation in the AF. For the ligand we introduce an array $\mathbf{b}$ which gives both its conformation in the BF and the orientation of the BF with reference to AF axes. For completeness, we could introduce also a collection of solvent degrees of freedom, here ignored for the sake of notation. Consider now the *generalized first-neighbor distribution function*, $G_1(\mathbf{r}, \mathbf{a}, \mathbf{b})$, such that $G_1(\mathbf{r}, \mathbf{a}, \mathbf{b})$ $\mathrm{d}\mathbf{r}$ $\mathrm{d}\mathbf{a}$ $\mathrm{d}\mathbf{b}$ is the probability to find the nearest ligand molecule within the elemental volume of configurational space at location $\mathbf{r}$ from a protein, with protein and ligand in configurations $\mathbf{a}$ and $\mathbf{b}$ respectively. Normalization is $\int G_1(\mathbf{r}, \mathbf{a}, \mathbf{b})$ $\mathrm{d}\mathbf{r}$ $\mathrm{d}\mathbf{a}$ $\mathrm{d}\mathbf{b} = 1$, where the spatial integration is extended over the bulk of volume V. Notice that, due to the intrinsic angular anisotropy of the system, specially close to the binding site, Cartesian coordinates need to be used here in place of the single radial variable usually adopted for most common cases where spherical symmetry (of B *vs.* central A) holds with good approximation. Under the realistic assumption that multiple binding of ligands to the same protein site is avoided, the probability of finding a protein bound to its neighbor ligand is given by

$$P_{\mathrm{b,eq}} = \int G_1(\mathbf{r}, \mathbf{a}, \mathbf{b}) \; S_{\mathrm{b}}(\mathbf{r}, \mathbf{a}, \mathbf{b}) \; \mathrm{d}\mathbf{r} \; \mathrm{d}\mathbf{a} \; \mathrm{d}\mathbf{b} \tag{7}$$

where $S_{\mathrm{b}}(\cdot)$ is a selector function whose value is 1 if the specified pair-configuration belongs to the bound state, 0 elsewhere.

The crucial problems to be faced now arise: *i)* to model the first-neighbor distribution for the specific system, and *ii)* to establish an *objective* criterion to specify the selector function. The first problem has no workable solutions for so complex systems (explicit forms can be obtained only for rigid molecules of simple geometries). This prevents to tackle point *ii)*. In fact, following Moro and Severin, a natural solution of the second problem could be achieved by looking, in principle, at the multidimensional landscape of the effective potential $-\ln G_1(\mathbf{r}, \mathbf{a}, \mathbf{b})$: there might appear a bistability with two well-separated compact basins of configurations (microstates). In this case, the bound and unbound (macro)states correspond to these collections of microstates, and the boundary of the bound state is naturally identified with the "separatix hypersurface" (in the full $(\mathbf{r}, \mathbf{a}, \mathbf{b})$ space) between the two basins. Such a kind of analysis was done explicitly in Ref. [6] for spherical particles interacting via Morse potential, while here the same approach is clearly infeasible and only chemical intuition may reasonably identify these domains via concepts like "binding pocket" or similar representations. We shall proceed by adopting the crudest approach which shall lead, however, to a rationale of a commonly adopted expression for the equilibrium constant in terms of PMF (Eq. 4 of the main text).

First assume that $S_{\mathrm{b}}(\mathbf{r}, \mathbf{a}, \mathbf{b})$ has a weak dependence on the conformational/orientational degrees of freedom of protein and ligand. In this limit one can work with the reduced distribution

$$\overline{G}_1(\mathbf{r}) = \int G_1(\mathbf{r}, \mathbf{a}, \mathbf{b}) \; \mathrm{d}\mathbf{a} \; \mathrm{d}\mathbf{b} \tag{8}$$

normalized as $\int \overline{G}_1(\mathbf{r}) \; \mathrm{d}\mathbf{r} = 1$. Thus,

$$P_{\mathrm{b,eq}} = \int_{I_{\mathrm{b}}} \overline{G}_1(\mathbf{r}) \; \mathrm{d}\mathbf{r} \tag{9}$$

where I_{b} stands for the spatial domain, surrounding the protein, which corresponds to bound configurations (the "binding pocket", to say). The purpose is to establish a link between $\overline{G}_1(\mathbf{r})$ and the usual dimensionless pair-correlation function $g(\mathbf{r})$ related to the probability to find a ligand molecule at displacement $\mathbf{r}$ from the protein (with $\int g(\mathbf{r}) \; \mathrm{d}\mathbf{r} = V$). The $g(\mathbf{r})$ is then associated to the PMF, $\Phi(\mathbf{r})$, through

$$\mathrm{e}^{[\phi - \Phi(\mathbf{r})]/(RT)} = g(\mathbf{r}) \tag{10}$$

where ϕ denotes the value of the PMF at infinite distance of the ligand from the protein, that is, in the "free" state, R is the molar gas constant and T is the temperature. Clearly, any offset in the PMF values deletes in the difference $\phi - \Phi(\mathbf{r})$. In particular, as stressed in the main text concerning the SMD practice, the PMF values can be implicitly meant as shifts from the reference value associated to *any* location $\mathbf{r}_0$ freely chosen.

In principle, an approximate integral relation between $\overline{G}_1(\mathbf{r})$ and $g(\mathbf{r})$ could be worked out by adapting the strategy indicated in Ref. [7] and developed by Torquato *et al.* in Ref. [8]. The starting point is to apply the composed-probability rule to link the first-neighbor distribution to the "exclusion" probability (*i.e.*, the probability of finding no ligand molecules within the spherical cavity of radius r) through a "conditional" pair-correlation (see Eqs. (2.12) and (2.17) in Ref. [8] paying attention to the different notation). The crucial point is

to model this latter function in a way devoid as much as possible of subjective bias. This is implicitly the strategy pursued in Ref. [6], where a detailed form of the first-neighbor distribution on the large spatial scale was needed to display the bistability feature of the effective potential as stressed above. However, as pointed out in Ref. [7], subjective choices may lead to approximations whose reliability is hardly assessable. At any rate, for our purposes only a likely approximation of $\overline{G}_1(\mathbf{r})$ within the bound state is needed; still according to the assumption that a single ligand can populate the binding pocket, in the absence of detailed information an unbiased choice is to take

$$\overline{G}_1(\mathbf{r}) \simeq \rho_{\mathrm{B}}\ g(\mathbf{r}) \tag{11}$$

for locations $\mathbf{r}$ within I_{b}, where $\rho_{\mathrm{B}} = N_{\mathrm{B}}/V = v_{\mathrm{B}}^{-1}$ is the particle-density of ligand molecules (notice that Eq. 11 is also suggested in Ref. [7]). Clearly, such a relation holds only within the bound state (normalization, in fact, is not fulfilled on the whole volume), and it simply states that the probability of finding the reference-center of the (nearest) ligand in the volume element $\mathrm{d}\mathbf{r}$ at $\mathbf{r}$ (*i.e.*, $\overline{G}_1(\mathbf{r})\ \mathrm{d}\mathbf{r}$) is given by the probability of finding a ligand (*i.e.*, $g(\mathbf{r})/V\mathrm{d}\mathbf{r}$) multiplied by the number of available ligands (N_{B}). Substitution of Eq. 11 into Eq. 9, and use of Eq. 10, yields

$$P_{\mathrm{b,eq}} \simeq v_{\mathrm{B}}^{-1} \int_{I_{\mathrm{b}}} \mathrm{e}^{[\phi-\Phi(\mathbf{r})]/(RT)}\ \mathrm{d}\mathbf{r} \tag{12}$$

Now consider that the extension of the domain I_{b} is expected to be very weakly dependent on dilution. In other words, the counterpart I_u (spatial domain of the unbound state) depends sensibly on v_{B} while the separatrix between I_{b} and I_{u} is essentially independent of it. This has been demonstrated in Ref. [6] for spherical particles. Thus, as v_{B} increases at denominator in Eq. 12, $P_{\mathrm{b,eq}}$ takes values so small that the following approximation is acceptable for the complementary population of the unbound state:

$$P_{\mathrm{u,eq}} \simeq 1 \tag{13}$$

By inserting Eqs. 12 and 13 into Eq. 6 we get the final result

$$K_{\mathrm{c}} = N_{\mathrm{Av}} \int_{I_{\mathrm{b}}} \mathrm{e}^{[\phi-\Phi(\mathbf{r})]/(RT)}\ \mathrm{d}\mathbf{r} \tag{14}$$

which is Eq. 4 of the main text. Although Eq. 14 might sound familiar, we stress again that its validity depends on the quality of approximation Eq. 11 and, even more important, on the concrete possibility to discharge (average out) all orientational/conformational correlations between ligand and protein when adopting the reduced distribution $\overline{G}_1(\mathbf{r})$ through Eq. 8 for the positional variables only. These assumptions may be responsible for an unpredictable quality degradation of Eq. 14.

[1] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. M. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, et al., J. Comput. Chem. **24**, 1999 (2003).

[2] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, J. Comput. Chem. **25**, 1157 (2004).

[3] Amber Home Page (Last consultation: February 28, 2013), http://ambermd.org/dbase.html.

[4] W. D. Cornell, P. Cieplak, I. Bayly, I. R. Gould, K. M. J. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).

[5] H.-S. Choi, Z. Wang, W. Richmond, X. He, K. Yang, T. Jiang, D. Karanewsky, X. Gu, V. Zhou, Y. Liu, et al., Bioorg. Med. Chem. Lett. **16**, 2689 (2006).

[6] G. J. Moro and G. Severin, J. Chem. Phys. **114**, 4565 (2001).

[7] Y. Zhang and J. A. McCammon, J. Chem. Phys. **118**, 1821 (2003).

[8] S. Torquato, B. Lu, and J. Rubinstein, Phys. Rev. A **41**, 2059 (1990).

	FAK-*16i*			FAK-*17g*			FAK-*32*		
$\mathbf{q}_i$	Met475	Thr532	Val605	Met475	Thr532	Val605	Met475	Thr532	Val605
$\mathbf{q}_0$	14.7	20.6	23.9	14.7	20.5	24.0	15.0	20.9	24.6
$\mathbf{q}_1$	15.6	21.1	24.3	15.6	21.0	24.2	15.9	21.4	25.1
$\mathbf{q}_2$	16.4	21.5	24.5	16.5	21.4	24.4	16.8	21.8	25.4
$\mathbf{q}_3$	17.1	21.8	24.6	17.3	21.8	24.7	17.7	22.2	25.7
$\mathbf{q}_4$	18.0	22.3	24.9	18.3	22.3	25.0	18.6	22.7	26.2
$\mathbf{q}_5$	18.9	22.8	25.3	19.2	22.9	25.5	19.5	23.2	26.7
$\mathbf{q}_6$	19.8	23.4	25.8	20.2	23.5	26.0	20.5	23.8	27.2
$\mathbf{q}_7$	20.7	24.0	26.3	21.1	24.2	26.6	21.4	24.4	27.8
$\mathbf{q}_8$	21.6	24.6	26.7	22.1	24.8	27.2	22.4	25.1	28.4
$\mathbf{q}_9$	22.5	25.1	27.1	23.0	25.5	27.7	23.4	25.9	29.1
$\mathbf{q}_{10}$	23.5	25.8	27.7	24.0	26.2	28.4	24.3	26.6	29.8
$\mathbf{q}_{11}$	24.4	26.5	28.4	25.0	27.0	29.1	25.3	27.4	30.5
$\mathbf{q}_{12}$	25.4	27.3	29.1	25.9	27.8	29.7	26.3	28.1	31.2
$\mathbf{q}_{13}$	26.3	28.1	29.8	26.9	28.5	30.4	27.2	28.8	31.8
$\mathbf{q}_{14}$	27.3	28.9	30.5	27.9	29.3	31.1	28.2	29.4	32.3
$\mathbf{q}_{15}$	28.2	29.8	31.4	28.8	30.1	31.8	29.1	30.1	33.0
$\mathbf{q}_{16}$	29.2	30.6	32.2	29.8	30.9	32.5	30.1	30.8	33.6
$\mathbf{q}_{17}$	30.2	31.5	33.0	30.7	31.7	33.3	31.0	31.7	34.4
$\mathbf{q}_{18}$	31.1	32.4	33.9	31.7	32.6	34.2	32.0	32.6	35.3
$\mathbf{q}_{19}$	32.1	33.3	34.7	32.7	33.5	35.0	32.9	33.5	36.2
$\mathbf{q}_{20}$	33.0	34.2	35.5	33.6	34.4	35.9	33.9	34.4	37.1
$\mathbf{q}_{21}$	34.0	35.1	36.3	34.6	35.3	36.8	34.8	35.4	38.0
$\mathbf{q}_{22}$	35.0	36.0	37.2	35.5	36.2	37.6	35.8	36.3	38.9
$\mathbf{q}_{23}$	35.9	36.9	38.0	36.5	37.2	38.5	36.7	37.2	39.8
$\mathbf{q}_{24}$	36.9	37.9	38.9	37.5	38.1	39.4	37.7	38.2	40.7
$\mathbf{q}_{25}$	37.8	38.8	39.7	38.4	39.0	40.3	38.7	39.1	41.7
$\mathbf{q}_{26}$	38.8	39.7	40.6	39.4	39.9	41.2	39.6	40.1	42.6

TABLE VIII: Pulling coordinate $\mathbf{q}_i$ for the FAK-*16i*, FAK-*17g* and FAK-*32* systems. The entries represent the distances (in Å) between the protein reference-atoms (α-carbons of Met475, Thr532 and Val605) and the ligand reference-atom.