[Frontiers in Bioscience, Scholar, 5, 478-495, January 1, 2013]

A statistical mechanics handbook for protein-ligand binding simulation

Walter Rocchia[1], Sara Bonella[2]

[1]Drug Discovery and Development, Istituto Italiano di Tecnologia, via Morego 30, 16163 Genova, Italy, [2]Department of Physics, University of Rome "La Sapienza" and CNISM Unit 1, Ple A. Moro 5, 00185 Rome, Italy

TABLE OF CONTENTS

1. ABSTRACT

In this work, the fundamental elements of statistical mechanics underlying the simulation of the protein-ligand binding process, such as statistical ensembles and the concept of microscopic estimators of macroscopic observables and free energy, are summarized in a self consistent fashion. Particular attention is then devoted to the introduction of some mathematical tools that are used in atomistic simulations aimed at estimating binding affinities and free energy profiles, and to the illustration of the origins of the difficulties encountered in this endeavor.

2. INTRODUCTION

Statistical mechanics provides a useful framework to describe and simulate, among many other systems, protein-ligand interactions. Estimating the binding energy of the process, in particular, is a task often addressed within this context. Indeed, the topic is relevant in many fields including Drug Design, and a large number of works on this topic can be found in the literature. In this paper, we wish to summarize the fundamentals needed to formalize the problem of protein-ligand binding, assuming a basic knowledge of classical Physics. We will emphasize the aspects introductory to the numerical simulation of the problem (the specific techniques will instead be discussed

in other contributions to the issue), point out the difficulties in obtaining an accurate estimate of the binding free energy, and indicate several potential pitfalls that can be encountered in these simulations. While some information is presented in complementary material at the end of the sections, a formal, and complete, derivation of the results presented is out of the scope of this work. Many exhaustive books on the topic exist and the interested reader will be referred to those for more in depth presentation and discussion. Our intention here is to present an introductory, self consistent framework where the researcher lacking a specific expertise in statistical mechanics can find a reasonably precise idea and form an intuitive picture of the theory underlying some of the most commonly used techniques for the simulation of protein-ligand binding. Occasionally, we hope to also suggest some deeper insights.

The paper is organized as follows. We begin by introducing the basic concepts of the dynamics and the statistical mechanics of a system. We will stress, in particular, the connection between microscopic and macroscopic observables and the concept of a statistical ensemble. The laws controlling the evolution of an isolated system will be summarized first, followed by the extension to the case of a closed system in equilibrium with a thermal reservoir. Then, the key quantity for our problem, i.e. the free energy, is introduced. We shall briefly discuss its role as the generator of microscopic estimators for thermodynamic quantities such as the pressure of the system, and then move to a more detailed description of its part in characterizing the thermodynamics, mechanism and kinetics of activated processes such as protein-ligand binding. We have tried to introduce a reasonably general definition of the free energy and its properties but, given the specific interest of this issue, the illustration of these properties is restricted to the specific case at hand (and to some of its peculiarities). To set the stage for the computational approaches discussed in other contributions, we also point out some of the difficulties in calculating the free energy numerically. Finally, to highlight differences and similarities among the two quantities, we present a brief comparison of the free energy with the potential energy of a physical system. We illustrate, in particular, the differences using a simple didactical example that, however, contains some of the ingredients of more complex systems.

3. BASICS

We will model our system, i.e. the protein, the ligand and the solvent, as a set of N interacting classical particles in a fixed volume V. The state of the system is completely specified by the knowledge of the set of the coordinates $\{\mathbf{r}_i, i=1,..,N\}$ and momenta $\{\mathbf{p}_i^r, i=1,..,N\}$ of the particles, $\mathbf{r}_i = \{r_{i,\alpha}, \alpha=1,2,3\}$ is the three dimensional vector identifying the position of particle i in Cartesian space, whereas in the following $\mathbf{r}$ indicates the 3N dimensional vector spanning both subscripts i and α, with similar notation for $\mathbf{p}_i^r$ and $\mathbf{p}^r$. Together, positions and momenta define a point in the so-called phase space, that we shall indicate as $\mathbf{\Gamma}=(\mathbf{r}, \mathbf{p}^r)$, of the system and that we shall consider as a classical microstate. The evolution of a system can be described by the dynamical trajectory $\{\mathbf{\Gamma}(t)\} = \{\mathbf{r}(t), \mathbf{p}^r(t)\}$, i.e. the set of points in phase space that the system visits in time (see the following for how to determine this sequence). Since the trajectory carries, for each value of the time, the complete information about the state of the system, it must be possible to express all experimentally measurable quantities (such as temperature, pressure, diffusion coefficients, spectra, etc.) as functions of the points in phase space visited by $\mathbf{\Gamma}(t)$. While this is in principle correct, experimental quantities are more conveniently characterized as time averages of such functions. To see why this may be so, consider that the time scale of displacement of a point in the phase space for a typical atomic system is of the order of the femtosecond (i.e. 10^{-15}s), much smaller than the typical human scale observation time (millisecond or even second). Thus, it is reasonable to postulate that the result of an observation, the macroscopic observable $\bar{O}$, is given by the average over time of the corresponding microscopic observable $O(\mathbf{\Gamma}(t))$. Thus, for an observation started at time t_0:

$$\bar{O} \triangleq \lim_{\tau\to\infty} \frac{1}{\tau - t_0} \int_{t_0}^{\tau} O(\Gamma(t))dt \tag{3.1}$$

The equation above is the cornerstone of classical statistical mechanics as first introduced by Boltzmann: it connects the microscopic dynamics of the system to the macroscopic observation via the operation of time average. In this context, the trajectory is not used to describe the evolution of the system "*per se*". Rather, it is a tool to generate, as a function of time, a sequence of points in phase space where the value of the observable is computed and then averaged. In the following, we are interested in so-called *equilibrium* properties of a system, where macroscopic observables can be taken as well-defined constants and the system has lost memory of the initial state at $t=t_0$. For systems at equilibrium, it is possible to restate the calculation of the averages in a language in which the time-dependent trajectory is no longer present. To see how this can be done, we present an argument that, under appropriate hypotheses (such as ergodicity[1]) can be turned into a rigorous statement. Let us consider the discretized version of the integral (3.1). Dividing the total time $\tau - t_0$ into *n* steps of duration h, we can write:

$$\bar{O} \approx \lim_{n\to\infty} \frac{1}{nh} \sum_{j=1}^{n} O_j h, \tag{3.2}$$

where $O_j = O(jh)$. Let us now divide the phase space of the system in L (hyper-)volume elements of the same "size" $d\mathbf{\Gamma}$ and assume that they are small enough, and the observable smooth enough, that the value of $O(\mathbf{\Gamma})$ in each element is a constant O_l. As the trajectory evolves, it visits any given cell more than once. The terms in the sum can then be rearranged based on the following observation: the equation above counts the contribution of each visit to a given cell to the total average by adding up the value of the observable every time that the trajectory enters the cell. If there have been n_l incursions of the trajectory in cell Γ_l, these visits can also be accounted for by a single term of

the form $n_l O_l$. The average can then be computed by summing over the different cells as:

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n} O_j = \lim_{n\to\infty}\frac{1}{n}\sum_{l=1}^{L} n_l O_l \tag{3.3}$$

Now, note that the ratio n_l/n is in fact the frequency with which the trajectory visits the cell Γ_l (in other words, the number of steps at which it is in that cell over the total number of steps). In the limit $n\to\infty$ this ratio tends to the probability to find the system in the cell according to that specific trajectory, i.e. $\rho_l d\boldsymbol{\Gamma}$ if ρ_l is the probability density. Thus, the average along the trajectory can also be expressed as:

$$\lim_{n\to\infty}\frac{1}{n}\sum_{l=1}^{L} n_l O_l = \sum_{l=1}^{L}\rho_l O_l\, d\boldsymbol{\Gamma} \cong \cong \int_{\cup\Gamma_l}\rho(\Gamma)O(\Gamma)d\boldsymbol{\Gamma} \triangleq \langle O\rangle \tag{3.4}$$

where the last relation becomes an equality for infinitesimal $d\boldsymbol{\Gamma} = d\mathbf{r}d\mathbf{p}^r$ and $\cup\Gamma_l$ is the whole phase space. The angle brackets <•> indicate the so-called ensemble average[2]. If the ergodic hypothesis is fulfilled, time and ensemble averages coincide, i.e. $\bar{O} = \langle O\rangle$, for any possible initial condition of the, infinitely long, trajectory. This implies that $\rho(\boldsymbol{\Gamma})$ is the stationary probability distribution that characterizes the system. In this case, then, macroscopic observables can also be computed as averages over the probability to find the system in a volume between $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}+d\boldsymbol{\Gamma}$ in the phase space. In this picture, time and the trajectory no longer appear and they are substituted by the idea of a weighted average over all the possible states accessible to the system. This formulation of statistical mechanics is originally due to Gibbs (1).

3.1. Fixed energy systems

Let us assume that our system is isolated so that its energy is constant. The microscopic quantity corresponding to the total energy is the Hamiltonian $H(\mathbf{r}, \mathbf{p}^r)$. In the following we will assume that it is given by the sum of coordinate-independent kinetic energy, K, and momentum-independent potential energy, $\mathfrak{V}$, so that:

$$H(\mathbf{r},\mathbf{p}^r) = K(\mathbf{p}^r) + \mathfrak{V}(\mathbf{r}) = \sum_{i=1}^{N}\frac{\mathbf{p}_i^r\cdot\mathbf{p}_i^r}{2m_i} + \mathfrak{V}(\mathbf{r}) \tag{3.5}$$

Note that, while the kinetic energy depends on the detailed nature of the system only via the atomic masses, the form of the potential energy function, which describes the interactions among the particles, is specific to a given system. Although this is not influential for the results we are presenting here, in the molecular systems of our interest $\mathfrak{V}$ often depends on mutual particle positions and not on the absolute position with respect to a spatial reference system. Once an initial condition, i.e. a set of coordinates and momenta at a given time t_0, is specified, the dynamical trajectory $\boldsymbol{\Gamma}(t)= \{\mathbf{r}(t; \mathbf{r}(t_0), \mathbf{p}^r(t_0)), \mathbf{p}^r(t; \mathbf{r}(t_0), \mathbf{p}^r(t_0))\}$, can be obtained, in principle, by solving the following set of 3N, coupled, first order differential equations:

$$\begin{cases}\frac{d\mathbf{r}_i}{dt} = \nabla_{\mathbf{p}_i^r}H(\mathbf{r},\mathbf{p}^r) = \nabla_{\mathbf{p}_i^r}K(\mathbf{p}^r) = \frac{\mathbf{p}_i^r}{m_i}\\ \frac{d\mathbf{p}_i^r}{dt} = -\nabla_{\mathbf{r}_i}H(\mathbf{r},\mathbf{p}^r) = -\nabla_{\mathbf{r}_i}\mathfrak{V}(\mathbf{r})\end{cases} \quad i=1\ldots N. \tag{3.6}$$

The equations above are Hamilton's evolution equations and they guarantee energy conservation. It can in fact be easily verified by direct substitution of the Hamiltonian in the time derivative that, if the phase space variables obey these equations, $dH(\boldsymbol{\Gamma})/dt=0$ so $H(\boldsymbol{\Gamma})$=const.=E, where E is the total energy of the system. The explicit form of the probability density $\rho(\boldsymbol{\Gamma})$ associated to this dynamics is obtained by assuming that, at equilibrium, all microscopic states compatible with the fixed energy constraint are equally probable. Thus, the probability to find the system in an infinitesimal volume $d\boldsymbol{\Gamma}$ around the point $\boldsymbol{\Gamma}$ must be zero if the phase space point is such that $H(\boldsymbol{\Gamma})\neq E$ and constant otherwise. This requirement is satisfied by choosing:

$$\rho(\boldsymbol{\Gamma})d\boldsymbol{\Gamma} = \lambda\,\delta(H(\boldsymbol{\Gamma}) - E)d\boldsymbol{\Gamma} \tag{3.7}$$

where $\delta(\cdot)$ is Dirac's delta[3]. The constant λ can be exploited to ensure that $\rho(\boldsymbol{\Gamma})$ satisfies a set of additional requirements: i) being normalized to one, ii) agreeing with the corresponding quantum probability density in the appropriate limit, iii) accounting for particle's indistinguishability (i.e. reflecting the invariance of the Hamiltonian upon relabeling of particles of the same chemical species[4]). It can be shown (2), that these requirements are satisfied by $\lambda=\delta E_0/(\Pi_\eta N_\eta! h^{3N}\Omega(N,V,E))$, with:

$$\Omega(N,V,E) = \frac{\delta E_0}{\prod_\eta N_\eta! h^{3N}}\prod_{i=1}^{N}\int_{\mathbb{R}^3} d\mathbf{p}_i^r\int_V \delta(H(\mathbf{r},\mathbf{p}^r)-E)d\mathbf{r}_i = = \frac{\delta E_0}{\prod_\eta N_\eta! h^{3N}}\int_{\mathbb{R}^{3N}}\int_{D(N,V)}\delta(H(\boldsymbol{\Gamma})-E)d\boldsymbol{\Gamma} \tag{3.8}$$

where h is Planck's constant, δE_0 is the uncertainty in the measurement of energy, η counts over the different chemical species composing the system and $\Sigma\eta\ N\eta=N$.[5] $\Omega(N,V,E)$, known as the microcanonical partition function, can be interpreted as a counter of the microscopic states of the system compatible with the macroscopic NVE constraints. The condition on E resides in the $\delta(\cdot)$ function, those on N and V in the domain D(N,V), the set of all the possible values of $\{\mathbf{r}\}$ compatible with the NV constraints, which is named configurational space and measures V^N.

The considerations above lead to the normalized probability density for an isolated system:

$$\rho_{NVE}(\boldsymbol{\Gamma}) = \frac{\delta E_0\delta(H(\boldsymbol{\Gamma})-E)}{\prod_\eta N_\eta! h^{3N}\Omega(N,V,E)} \tag{3.9}$$

The space of events associated to this density is known as the *microcanonical ensemble*, or NVE.

3.2. Microscopic versus thermodynamic description

Based on the discussion above, if we start from a microscopic point of view, macroscopic observables such as temperature and pressure, can be calculated as time or ensemble averages of appropriate functions of the phase

space of the system. On the other hand, if we start from a macroscopic point of view, there is a well-established phenomenological theory, thermodynamics, that expresses the same observables as partial derivatives of a suitable state function, the thermodynamic potential, with respect to the so-called control variables. For an isolated system, the control variables are N, V, E and the thermodynamic potential is the entropy S(N,V,E) (3). Since both approaches have been validated experimentally, it must be possible to reconcile them. In the case of isolated systems, the bridge is provided by Boltzmann's relationship among the entropy and the partition function (note that these quantities depend on the same set of variables):

$$S(N,V,E) = k_B \ln\Omega(N,V,E) \tag{3.10}$$

where k_B is Boltzmann's constant. The relationship above is a fundamental *ansatz*, a prescription that identifies the macroscopic thermodynamic potential with an appropriate function of the microscopic ensemble. Note that, while we have previously identified observables with averages of suitable microscopic functions of the phase space, entropy in the NVE ensemble is related directly to the measure of the volume in phase space accessible to the system. The ansatz above, and in particular the presence of the logarithm, is validated by the consistency among the predictions based on the macroscopic and microscopic approaches. One immediate example is given by considering a situation in which we have two isolated subsystems, characterized by (N_1,V_1,E_1) and (N_2,V_2,E_2) that are brought in contact. Thermodynamics tells us that, if the systems do not interact, the entropy of the overall system is given by the sum of the two entropies; the same result must hold if we use the definition of the entropy based on the microcanonical partition function. This can be shown as follows: since the two systems are non interacting, the overall Hamiltonian is given by the sum of the Hamiltonians of the subsystems and it can be easily shown that these Hamiltonians are separately conserved by the dynamics (in this case the system is said to be separable). Indicating with $H_i(\mathbf{\Gamma}_i)$ the Hamiltonian of each subsystem (i =1, 2) and noting that the volume element for the combined system, $d\mathbf{\Gamma}$, is equal to the product of the individual volume elements $d\mathbf{\Gamma}_i$, we can write the partition function as:

$$\begin{aligned}&\Omega(N_1+N_2,V_1+V_2,E_1+E_2)=\\&=\frac{\delta E_0{}^2}{\prod_{\eta_1} N_{\eta_1}!\prod_{\eta_2} N_{\eta_2}!\, h^{3(N_1+N_2)}}\prod_{i=1}^{N_1}\int_{\mathbb{R}^3} d\mathbf{p}_i^r \int_V \delta(H_1(\mathbf{\Gamma_1})-E_1)d\mathbf{r_i}\cdot\\&\cdot\prod_{j=1}^{N_2}\int_{\mathbb{R}^3} d\mathbf{p}_j^r \int_V \delta(H_2(\mathbf{\Gamma_2})-E_2)d\mathbf{r_j}=\\&=\Omega(N_1,V_1,E_1)\Omega(N_2,V_2,E_2)\end{aligned} \tag{3.11}$$

This is somehow intuitive, since the possible states for the union of two non interacting systems are obtained by selecting one state for the first and then, independently, one for the other. Substituting in equation (3.10) we have:

$$\begin{aligned}S(N_1+N_2,V_1+V_2,E_1+E_2)&=k_B\ln\Omega(N_1+N_2,V_1+V_2,E_1+E_2)=\\&=k_B\ln[\Omega(N_1,V_1,E_1)\Omega(N_2,V_2,E_2)]=\\&=k_B[\ln\Omega(N_1,V_1,E_1)+\ln\Omega(N_2,V_2,E_2)]=\\&=S(N_1,V_1,E_1)+S(N_2,V_2,E_2)\end{aligned} \tag{3.12}$$

which indeed is consistent with the thermodynamic result.

Having established the framework of statistical mechanics and shown its connection to thermodynamics, we are, in principle, in the position to use either Eq. (3.1) or Eq. (3.4) to obtain measurable quantities. In practice, however, to do so we must still solve two problems. The first one is the definition of the function of phase space that corresponds to a given observable. This can be done either by microscopic considerations or by means of the thermodynamic relationships involving the thermodynamic potential. We will show some examples of these definitions in the next section after introducing an ensemble, the canonical ensemble, which is more closely related to the typical conditions in an experiment. The second problem is that to compute the averages we must either be able to solve the evolution equations of the system (Boltzmann) or to sample configurations[6] according to the ensemble probability density (Gibbs). Neither of these tasks can be performed analytically except for systems described by very simple Hamiltonians (free particle, harmonic oscillator and a few others). Both can however be tackled quite effectively by numerical methods that will be described in the next paper in this issue. In particular, the "Boltzmann" approach is realized using a set of techniques that go under the general name of molecular dynamics (MD), while the "Gibbs" approach is implemented via the so-called Monte Carlo methods (MC) (4-6).

3.3. Fixed temperature systems

We will now focus on the description of a closed system at fixed temperature, T. This situation is more interesting from the point of view of the connection with experiments since temperature is easier to control than overall energy. In statistical mechanics the ensemble corresponding to the new macroscopic constraints of fixed N,V,T is called canonical. This ensemble is modeled by considering a system σ that interacts with a much larger one, the so-called thermal reservoir ℜ, which is at fixed temperature. The characteristics of the interaction can be summarized as follows (3):

- the overall system Σ composed by σ and ℜ, is isolated;
- the number of degrees of freedom of ℜ is much larger than that of σ;
- there is only a weak and non specific coupling between σ and ℜ;
- the volume and the number of particles in each subsystem are kept constant as in the microcanonical ensemble.

With these assumptions, the equilibrium probability density for the microstate represented by the point $\mathbf{\Gamma}=(\mathbf{r}, \mathbf{p}^r)$ in the phase space of σ is the Boltzmann distribution (see details in Sect. 3.5.1.):

$$\rho_{NVT}(\mathbf{\Gamma}) = \frac{\exp\{-\beta\, H(\mathbf{\Gamma})\}}{\prod_\eta N_\eta! h^{3N} Q(N,V,T)} \tag{3.13}$$

Where $\beta \triangleq (k_B T)^{-1}$ and Q is the *canonical partition function*, which, similarly to the NVE case, is a weighted counter of all the microstates:

$$Q(N,V,T) = \frac{1}{\prod_\eta N_\eta! h^{3N}} \int_{D(N,V)} d\mathbf{r} \int_{R^{3N}} \exp\left(-\beta H(\mathbf{r},\mathbf{p}^r)\right) d\mathbf{p}^r \quad (3.14)$$

When, as we are assuming in our systems, the kinetic energy depends only on momenta and the potential energy only on coordinates Q can be expressed as the product:

$$Q = \left\{ \frac{1}{\prod_\eta N_\eta! h^{3N}} \int_{\mathbb{R}^{3N}} \exp\left(- \sum_{i=1}^{N} \frac{\boldsymbol{p}_i^r \cdot \boldsymbol{p}_i^r}{2 m_i k_B T} \right) d\boldsymbol{p}^r \right\} \cdot \left\{ \int_{D(N,V)} exp\left(-\beta U(\boldsymbol{r})\right) d\boldsymbol{r} \right\} \quad (3.15)$$

The first term in curly braces can be calculated analytically since it is the product of 3N integrals of Gaussian form over the whole real axis. When the potential energy is independent on **r**, this integral is the only non trivial contribution to Q. This is the case for a system of non interacting particles (i.e. an ideal gas), so this term is often referred to as the "*ideal*" part of the partition function. The second term, in contrast, depends on the specific form of the interactions among the particles and cannot be computed analytically for a generic potential. It is called *configurational integral* and often indicated as Z. It accounts for the contribution in excess to the ideal one (due to this fact, the configurational contribution to any thermodynamic property is called the "*excess*" contribution). Performing the Gaussian integrals and using the definition of Z, the partition function can be written as:

$$Q = \frac{\prod_{i=1}^{N} (2\pi k_B T m_i)^{\frac{3}{2}}}{\prod_\eta N_\eta!\, h^{3N}} Z \quad (3.16)$$

Excess quantities are of high interest and the relevant function for their description is the reduced probability density over the configurational space:

$$\rho_{NVT}^r(\mathbf{r}) = \int_{R^{3N}} \rho_{NVT}(\mathbf{r},\mathbf{p}^r) d\mathbf{p}^r = \frac{\exp\{-\beta U(\mathbf{r})\}}{Z(N,V,T)} \quad (3.17)$$

3.3.1. Consequences of the Boltzmann distribution

From the Boltzmann distribution, all statistical properties of any function of the coordinates can be derived as well as many important relationships among them. An important example descends from the expression of average value and variance of the energy of a canonical system. The *internal energy* U is the total energy of the system averaged in the canonical ensemble:

$$U \triangleq \langle E \rangle_{NVT} = \iint H(\Gamma)\rho_{NVT}(\Gamma) d\Gamma = \frac{-1}{\prod_\eta N_\eta! h^{3N} Q} \iint \frac{\partial \exp(-\beta H)}{\partial \beta} d\Gamma = -\frac{\partial \ln(Q)}{\partial \beta} \quad (3.18)$$

The energy variance is given by:

$$Var(E) = \langle (E-U)^2 \rangle_{NVT} = \langle E^2 \rangle_{NVT} - U^2 = \iint H^2(\Gamma)\rho_{NVT}(\Gamma) d\Gamma - U^2 = \frac{1}{\prod_\eta N_\eta! h^{3N} Q} \iint \frac{\partial^2 \exp(-\beta H)}{\partial \beta^2} d\Gamma = \frac{\partial^2 Q}{Q \partial \beta^2} - \left(\frac{\partial \ln(Q)}{\partial \beta}\right)^2 = \frac{\partial^2 \ln(Q)}{\partial \beta^2} = -\frac{\partial U}{\partial \beta} \quad (3.19)$$

If we consider the definition of heat capacity, $C_v \triangleq \left(\frac{\partial U}{\partial T}\right)_{N,V} = \frac{-1}{k_B T^2}\left(\frac{\partial U}{\partial \beta}\right)_{N,V}$, and identify the thermodynamic internal energy with $< E >_{NVT}$, we get $Var(E) = k_B T^2 C_v$. This expression relates the size of spontaneous energy fluctuations to a constitutive property of the physical system under consideration, such as the rate at which energy changes due to alterations in the temperature. Since both energy and heat capacity are extensive quantities, they are proportional to N and therefore the size of fluctuations relative to the average energy turns out to be inversely proportional to the square root of the number of particles: $\sqrt{Var(E)}/\langle E \rangle \propto 1/\sqrt{N}$. Further analysis proves that, around the most probable energy value U of the energy E, the energy has the following Gaussian probability density: $\frac{1}{\sqrt{2\pi N k_B T}} \exp\left(-\frac{(E-U)^2}{2(k_B T)^2 N}\right)$ (7).

Based on the result above, and considering that for a macroscopic system $N \sim 10^{23}$ (i.e. proportional to Avogadro's number), one can see that these fluctuations are negligible with respect to the mean energy of the system. Consequently, when macroscopic systems are considered, the distinction between fixed energy and fixed temperature ensembles becomes irrelevant, and so it is for the other statistical ensembles. In simulations, the affordable computational power sets a limit on the size of the system, represented by N. The size of the energy fluctuation is one of the (many) tests to validate the hypothesis that the simulated system is representative of its thermodynamic limit.

3.3.2. Time evolution in the canonical ensemble

The evolution equations of a system belonging to the canonical ensemble are more complicated than those of a Hamiltonian system since they must take into account the average interaction with the reservoir $\Re$ under the same assumptions mentioned at the beginning of Sect. 3.3. Several models, called thermostats, have been developed to mimic the effect of this interaction in the evolution equations of a system and to generate trajectories that "sample" the distribution in Eq. (3.13). For example, in the Nosé-Hoover method the physical particles are coupled to a single extra degree of freedom whose evolution is tuned to influence the system's momenta so as to keep the temperature fluctuating around a preassigned value (4,6,8,9). Another possibility, which will be described more in detail in the following, is to use the Langevin model (10). Depending on the specific thermostat employed, the trajectories of the simulated system will be different and, of course, they will also be different from the ones induced by Hamilton's equations, which correspond to another ensemble. This is however not a problem since, to estimate the macroscopic observables, the only requirement on the trajectory is that it explores the phase space according to the correct statistics. Thus, as long as the microstates are visited with the frequency corresponding to the Boltzmann distribution, the details of the dynamics are irrelevant[7].

Let us now focus on the Langevin model. In this approach, each particle behaves as if it were undergoing a series of collisional events that can be modeled as a "viscous drag" force, opposite in sign and proportional in

intensity to the particle's velocity, summed to a random force. The (non Hamiltonian) evolution equations are:

$$\begin{cases} \frac{d\mathbf{r}_i}{dt} = \frac{\mathbf{p}_i^r}{m_i} \\ \frac{d\mathbf{p}_i^r}{dt} = -\nabla_{\mathbf{r}_i}\mathcal{U} - \gamma_i \mathbf{p}_i^r + \sqrt{D_i} m_i \mathbf{n}_i(t) \end{cases} \quad (3.20)$$

where $\boldsymbol{\Gamma}_i$ and D_i are the viscous friction coefficient and the diffusion constant, accounting for the number of collisions occurring in the unit time. $n_{i,\alpha}$ are white, i.e. delta-correlated, Gaussian noise terms, whose statistical properties are described by the expectation values:

$$E\{n_{i,\alpha}(t)\}=0 \;\forall\, t,\; E\{n_{i,\alpha}(t_1)\, n_{j,\beta}(t_2)\}=2\delta_{i,j}\,\delta_{\alpha,\beta}\,\delta(t_1 - t_2) \quad (3.21)$$

Due to the presence of the noise, Eq. (3.20) is a stochastic differential system, whose treatment needs suitable mathematical tools and whose solutions, i.e. the coordinates and momenta as a function of time, are a stochastic process. Many fundamental properties of this dynamical system can be derived by means of an important tool in the field of stochastic processes: the Fokker-Plank equation (11). In particular, see also Sect. 3.5.2., it can be shown that:

- the noise on average injects energy into the system σ whereas the frictional force, as it might be expected, has a dissipative effect in a way such that the average total energy of the system has a constant derivative:
$$\frac{d}{dt} < H >= \sum_{i=1}^{3N} D_i m_i - \frac{\gamma_i}{m_i} < \mathbf{p}_i^r \cdot \mathbf{p}_i^r > \quad (3.22)$$
- if the microscopic parameters are connected via the Einstein relationship:
$$k_B T = \frac{D_i m_i}{\gamma_i} \quad (3.23)$$
then the internal energy ($U = <K+\mathcal{U}>$) of the system is conserved, and the Boltzmann distribution is a stationary probability distribution for the system, proving the validity of this approach for the simulation within the canonical ensemble.

3.4. From microscopic quantities to macroscopic observables

Microscopically, a thermodynamic equilibrium state can be defined as a weighted ensemble of all the microstates that can be explored by the system when a consistent set of macroscopic constraints, involving control variables such as T, $\{N_k\}$, V, P etc.., are imposed. The specific combination of control variables defines the state itself, decides the value of the weights and identifies, as we have seen, the correct form of the "counter" of the accessible microstates, namely the *partition function*. This latter quantity contains all the statistical information concerning the system in that ensemble. The thermodynamic potential is proportional to the logarithm of the partition function and can be used, for instance, to obtain other thermodynamic observables by derivation with respect to the control variables. In the canonical ensemble:

$$F(N,V,T) = -k_B T \ln\left(Q(N,V,T)\right) \quad (3.24)$$

where F(N,V,T) is the Helmholtz free energy, i.e. the thermodynamic potential macroscopically associated to the constant temperature ensemble. Trusting this link, it becomes possible to obtain explicit microscopic estimators of macroscopic observables by exploiting the relationships among the free energy and energy, pressure etc., established by thermodynamic derivatives. It is also possible to derive these microscopic estimators starting from purely microscopic results (for example, via the generalized equipartition theorem) (2). The fact that the results obtained using these two approaches coincide is a further validation of the consistency of statistical mechanics. In the following, for convenience, we shall take the first route and derive a few relevant estimators.

We have already seen that the Hamiltonian is the microscopic estimator for the internal energy. If the kinetic energy term depends only on the momenta, we can derive another useful result:

$$\begin{aligned} \langle K \rangle &= \iint K(\boldsymbol{p}^r)\rho_{NVT}(\boldsymbol{r},\boldsymbol{p}^r)d\boldsymbol{r}d\boldsymbol{p}^r = \\ &= \frac{1}{\prod_\eta N_\eta! h^{3N} Q}\int_{D(N,V)} \exp\left(-\beta\mathcal{U}(\boldsymbol{r})\right)d\boldsymbol{r}\cdot \\ &\cdot \int_{\mathbb{R}^{3N}} K(\boldsymbol{p}^r) exp\left(-\beta K(\boldsymbol{p}^r)\right)d\boldsymbol{p}^r \end{aligned} \quad (3.25)$$

Now, recalling Eqs. (3.15) and (3.16), one obtains:

$$\begin{aligned} < K > &= \frac{\sum_{i=1}^{N}\frac{1}{2m_i}\sum_{\alpha=1}^{3}\int_{\mathbb{R}3N}(p_{i,\alpha}^r)^2 \exp\left(-\frac{K(\mathbf{p}^r)}{k_B T}\right)d\mathbf{p}^r}{\int_{\mathbb{R}^{3N}}\exp\left(-\frac{K(\mathbf{p}^r)}{k_B T}\right)d\mathbf{p}^r} = \\ &= \frac{\frac{3}{2}Nk_B T\left(\sqrt{(\pi k_B T)^3}\right)^N \prod_1^N m_i^{\frac{3}{2}}}{\left(\sqrt{(\pi k_B T)^3}\right)^N \prod_1^N m_i^{\frac{3}{2}}} = \frac{3}{2}Nk_B T \end{aligned} \quad (3.26)$$

This relationship is a particular case of the equipartition theorem and establishes the kinetic energy as the microscopic estimator of the temperature of the system (via the proportionality constant $3Nk_B/2$).

A further microscopic estimator, the one for the pressure, can be obtained from the thermodynamic derivative $P = -(\partial F/\partial V)_{N,T}$, again by substitution of the microscopic expression in Eq. (3.24) for the free energy, thus:

$$P = k_B T\frac{\partial \ln(Q)}{\partial V} = \frac{k_B T}{Q}\frac{\left(\prod_{i=1}^{N} m_i\right)^{\frac{3}{2}}}{\prod_\eta N_\eta!}\left(\frac{\sqrt{2\pi k_B T}}{h}\right)^{3N}\frac{\partial Z}{\partial V} \quad (3.27)$$

The algebra in this case is more involved because the configurational integral depends on the volume via its boundaries so that taking the derivative above is not trivial. For most geometries of interest, however, the dependence can be transferred from the boundaries to the argument of the potential energy via an appropriate rescaling of the coordinates (see for example ref. (6) and references therein)

and it is possible to show that the microscopic estimator for the pressure is:

$$P = \frac{1}{3V}\sum_{i=1}^{N}\sum_{\alpha=1}^{3}\left[\frac{(p_{i,\alpha}^{r})^2}{2m_i} + r_{i,\alpha}F_{i,\alpha}\right] \tag{3.28}$$

where $F_{i,\alpha}$ is the α-th Cartesian component of the total force acting on the i-th particle.

Thermodynamic quantities are equilibrium averages over all the phase space; however, it is sometimes useful to consider the microscopic definitions averaged over a reduced part of the phase space or over a limited time lapse to get a "local" or "instantaneous" version of these variables, e.g. an instantaneous temperature or a non equilibrium entropy. Numerical simulations give access to these variables as well as to the individual trajectories of any degree of freedom but it is important to remember that only equilibrium averages over the full phase space fulfill the laws of thermodynamics.

3.5. Complementary material to section 3

3.5.1. Derivation of Boltzmann distribution for the canonical ensemble

Let Σ be the union of a system σ, weakly and non-specifically coupled to a much larger thermal reservoir ℜ. Σ belongs to the microcanonical ensemble and has constant energy E_Σ. Let us now consider the α microstate for σ, it will have energy $E_\sigma=\varepsilon_\alpha$, with $\varepsilon_\alpha << E_\Re = E_\Sigma - \varepsilon_\alpha$, the corresponding energy of ℜ. We are interested in the equilibrium probability of σ to be in the state α. The weakness of the coupling implies that, apart from the overall energy conservation, the events in σ and ℜ are independent. Therefore, the desired probability, which is the probability that σ is in state α and that ℜ is in any compatible state, can be expressed as follows:

$$\wp_\alpha \triangleq \wp(\sigma \text{ in } \alpha \;/\; E_\sigma=\varepsilon_\alpha)\;\wp(E_\sigma = \varepsilon_\alpha;\, E_\Re = E_\Sigma - \varepsilon_\alpha) = \\ = \wp(\sigma \text{ in } \alpha \,/\, E_\sigma = \varepsilon_\alpha)\wp(E_\sigma = \varepsilon_\alpha)\wp(E_\Re = E_\Sigma - \varepsilon_\alpha) \tag{3.29}$$

Due to the *principle of equal a priori probabilities* each probability is proportional to the number of states that fulfill each energy constraint:

$$\wp_\alpha \propto \frac{1}{\Omega_\alpha(\varepsilon_\alpha)}\Omega_\sigma(\varepsilon_\alpha)\Omega_\Re(E_\Sigma - \varepsilon_\alpha) = \exp\{\ln\Omega_\Re(E_\Sigma - \varepsilon_\alpha)\} \tag{3.30}$$

By definition of microcanonical entropy and thermodynamic temperature, and again due to the assumption that the only effect of σ on ℜ is to exchange a, minimal, amount of energy, one can derive that:

$$\wp_\alpha \propto \exp\left\{\frac{1}{k_B}S_\Re(E_\Sigma - \varepsilon_\alpha)\right\} \approx \exp\left\{\frac{1}{k_B}\left[S_\Re(E_\Sigma) - \varepsilon_\alpha\frac{\partial S_\Re}{\partial E_\Re}\right]\right\} = \\ = \exp\left\{\frac{1}{k_B}S_\Re(E_\Sigma)\right\}\exp\left\{\frac{-\varepsilon_\alpha}{k_B T_\Re}\right\} \tag{3.31}$$

(The Taylor expansion of $S_\Re$ has been performed and truncated at the first order term). Considering that the first factor does not depend on α, that $\sum_\alpha \wp_\alpha = 1$, and that since ℜ and σ are in thermal contact and at equilibrium, the Zeroth Law of Thermodynamics guarantees that $T_\Re = T_\sigma = T$, one finally gets the Boltzmann distribution:

$$\wp_\alpha = \frac{\exp\left\{-\frac{\varepsilon_\alpha}{k_B T}\right\}}{\sum_\alpha \exp\left\{-\frac{\varepsilon_\alpha}{k_B T}\right\}} \tag{3.32}$$

The denominator in the formula embeds all the thermodynamic properties of σ and is named *canonical partition function* of σ :

$$Q = \sum_\alpha \exp\left\{-\frac{\varepsilon_\alpha}{k_B T}\right\} \tag{3.33}$$

This derivation is consistent with both a discrete energy spectrum, more typical of a quantum description, and the classical continuous energy spectrum. In this latter case, the microstate α can be associated to an elementary phase space volume $\mathbf{\Gamma}_\alpha$ where energy can always be assumed to be uniform. Consistency with Quantum Mechanics suggests a lower bound for this volume, namely h^{3N}. In this case, $\wp_\alpha = \iint_{\Gamma_\alpha}\rho(\boldsymbol{r},\boldsymbol{p}^r)\mathrm{d}\boldsymbol{r}\mathrm{d}\boldsymbol{p}^r = \rho\{(\boldsymbol{r},\boldsymbol{p}^r)\in\Gamma_\alpha\}h^{3N} \propto \exp\left\{-\frac{\varepsilon_\alpha}{k_B T}\right\}$ and the normalization of the probability takes the form: $\iint \rho(\boldsymbol{r},\boldsymbol{p}^r)\mathrm{d}\boldsymbol{r}\mathrm{d}\boldsymbol{p}^r = 1$. The final result is very similar to the discrete case, with the probability density for the continuum representation being:

$$\rho(\boldsymbol{r},\boldsymbol{p}^r) = \frac{\exp\left\{-\frac{H(\boldsymbol{r},\boldsymbol{p}^r)}{k_B T}\right\}}{\iint \exp\left\{-\frac{H(\boldsymbol{r},\boldsymbol{p}^r)}{k_B T}\right\}\mathrm{d}\boldsymbol{r}\mathrm{d}\boldsymbol{p}^r} \tag{3.34}$$

3.5.2. The Fokker-Plank equation

The Fokker-Plank equation is the equation that describes the evolution of the probability density $W(\mathbf{r},\mathbf{p}^r)$ for a stochastic dynamical system such as the one in Eq. (3.20) and can therefore also describe the evolution of average quantities starting from a given initial state. This equation takes the form:

$$\frac{\partial W(\mathbf{r},\mathbf{p}^r)}{\partial t} = \sum_{i=1}^{N}\sum_{\alpha=1}^{3}\left\{\frac{\partial U}{\partial r_{i,\alpha}}\frac{\partial W}{\partial p_{i,\alpha}} - \frac{p_{i,\alpha}}{m_i}\frac{\partial W}{\partial r_{i,\alpha}} + \right. \\ \left. + \gamma_i W + \gamma_i p_{i,\alpha}\frac{\partial W}{\partial p_{i,\alpha}} + D_i m_i^2\frac{\partial^2 W}{\partial p_{i,\alpha}^2}\right\} \tag{3.36}$$

The different contributions to the evolution of the system can easily be identified, the first two addends in the right hand side descend from the usual Hamiltonian dynamics, i.e. those described by the Liouville equation; the third and fourth terms are due to the viscous drag while the last one is caused by the noisy force. If we introduce the temperature T of the thermal bath, by means of the Einstein relationship:

$$k_B T = \frac{D_i m_i}{\gamma_i} \tag{3.37}$$

then, Eq. (3.36) admits the Boltzmann distribution as stationary solution.

4. TOOLS AND CONCEPTS FOR THE DESCRIPTION OF THE BINDING PROCESS

4.1. Role of the free energy and of the internal constraint

In the previous section, we have used the free energy to generate, by differentiation, the microscopic estimator of the pressure (we could have done the same for the temperature). However, the importance of free energy is much broader. In the following, we will describe how it can be used to characterize the thermodynamics, mechanism, and kinetics of the protein-ligand binding process which is the focus of this contribution. Before doing so, we shall summarize the relationship between the free energy and the work necessary to bring a system from one state to another and point out some of the difficulties in a direct calculation of this (or any other) thermodynamic potential. Our considerations will refer to the NVT ensemble. In the field of chemical reactions and biological processes, the experimental conditions are usually NPT so we should focus on the Gibbs, rather than Helmholtz, free energy. However, in most cases of interest, the system can be approximated as incompressible, so the two thermodynamic potentials are essentially equivalent. Furthermore, simulations in the canonical ensemble are simpler to describe and the results presented in the next sections can be easily transposed to the NPT ensemble.

As a first step to appreciate the role and meaning of the free energy, let us consider an important consequence of the second law of thermodynamics: the *maximum entropy* or *minimum energy principle*. Let's imagine bringing a system to an out of equilibrium configuration, this can be done, for instance, by means of an *internal constraint* (12). An internal constraint limits the accessibility to a number, possibly large, of microstates without necessarily affecting the original control variables. A simple example is provided by a mechanical barrier dividing an adiabatic container full of gas in two parts at equilibrium. If we compare the situations with and without the barrier, we see that, although some variables, such as in this case T, V and N, may be unaffected, placing a constraint corresponds in general to forbidding some regions of the phase space, and this of course can result in possibly different values of some other quantities. From the microscopic point of view, this can be seen as the addition of a further control variable imposed by fixing the value of a function $\xi(\mathbf{r})$ of the spatial coordinates of the particles of the system. More complex control variables, involving the momenta, could be considered, but we will not do so in this work. The second law of thermodynamics tells us that, upon instantaneous removal of the constraint, the system relaxes to the equilibrium state of the unconstrained configuration, increasing its entropy. At the same time, both internal and free energies decrease. In other words, removing all internal constraints gives to the system the maximum freedom compatible with the remaining control variables and this leads to larger entropy and lower free and internal energies. This fact has interesting consequences for isothermal processes: let us imagine removing the internal constraint and observing the overall system until the new equilibrium state is reached. We know from the first law of thermodynamics that $\Delta U = \delta Q_R + \delta w$, where $\delta Q_R = T\Delta S_R$ is the heat entering the reservoir and δw is the work performed on the system. Due to the second law, $\Delta S_R + \Delta S \geq 0$, which implies: $\delta w \geq \Delta U - T\Delta S = \Delta F$. Thus, the free energy variation is always less or equal than the work performed on the system. The equality holds only for reversible processes, i.e. in absence of dissipation.

4.1.1. Obstacles to absolute free energy calculation

Let us now consider again, this time from a computational point of view, the microscopic expression for the free energy of a system in NVT:

$$F(N,V,T) = -k_B T \ln\left(\int_{D(N,V)} \exp\left(-\beta \mathcal{U}(\mathbf{r})\right) d\mathbf{r}\right) + -k_B T \ln\left\{\frac{\prod_{i=1}^{N}(2\pi k_B T m_i)^{\frac{3}{2}}}{\prod_{\eta} N_\eta! \, h^{3N}}\right\} \quad (4.1)$$

The direct calculation of this quantity poses several problems. First of all, Eq. (4.1) is not in the form of the average of an observable over the probability density (see Eq. (3.4)), but involves the evaluation of the configurational integral defined in Eqs. (3.15) and (3.16). While the methods mentioned in Sect. 3. are well suited to compute averages, they do not allow a direct and reliable calculation of Z(N,V,T). This difficulty can be circumvented to a certain extent using advanced methods, such as those mentioned later in this work or in other contributions to this issue, that take advantage of the fact that the relevant information is usually related to free energy differences (that can be expressed as averages) rather than absolute free energy values. Here, however, we will focus on the structure of the configurational integral to gain further insight on the free energy and to illustrate (some of the) difficulties in its evaluation.

The brute force numerical evaluation of Eq. (4.1) requires discretizing the configurational space volume in regions where the quantity $\exp(-\beta\mathcal{U}(\mathbf{r}))$can be assumed to be uniform[8]. In principle, the discretization process should involve the whole configurational space; this is a very challenging, if not impossible, task in high dimensional systems. Furthermore, the configurations that make contributions to the integral can be classified in three main categories:

- "low $\beta\mathcal{U}(\mathbf{r})$" regions, where the integrand contribution is large;
- "wide intermediate $\beta\mathcal{U}(\mathbf{r})$" regions, where the limited potential energy contribution is balanced by the entropic contribution provided by the large number of the corresponding microstates;
- high potential energy regions, that the exponential nature of the integrand makes negligible.

This qualitative classification underlines that the value of the potential energy alone is not sufficient to decide whether a configuration contributes appreciably to the integral or not. The entropic contribution, reflected by the numerosity of the microstates corresponding to a given potential energy, can have two opposite effects: making negligible low potential energy configurations corresponding to a vanishingly small part of the space, and,

conversely, making sets of configurations relevant not because of their "energetic convenience" but rather because of their numerousness. This is the main origin of the difference between potential energy surface and free energy surface. In Sect. 4.5., we will discuss a simple example highlighting this difference, that can have profound consequences extremely relevant in Drug Design.

4.1.2. Free energy differences calculation

As mentioned, the difficulties discussed above can be bypassed, to some extent, observing that the absolute value of the free energy is rarely needed since much information is contained in the free energy difference between two states of the system. To understand why this allows progress let us recall that, in statistical mechanics, a macroscopic state is a collection of microscopic configurations of the system (in general in phase space, but in this context we will assume that only coordinates are relevant). For example, in the protein-ligand binding process we could identify the unbound and bound states, *A* and *B* respectively, as the sets of microscopic configurations characterized by different mutual protein-ligand distances. Using the sole distance as a characterizing quantity is a very crude approximation and we will comment more on it in the following, it is however sufficient for introducing the following useful conceptual tools.

Let us consider the indicator function of the macrostate *A* (an analogous definition holds for *B*), $\chi_A(\mathbf{r})$, which equals 1 if **r** belongs to *A* and zero otherwise. The probability to find the system in state *A* can now be expressed as:

$$\wp(A) = \frac{\int_{D_A} \exp\{-\beta U(\mathbf{r})\}d\mathbf{r}}{\int_{D(N,V)} \exp\{-\beta U(\mathbf{r})\}d\mathbf{r}} = \frac{\int_{D(N,V)} \chi_A(\mathbf{r}) \exp\{-\beta U(\mathbf{r})\}d\mathbf{r}}{\int_{D(N,V)} \exp\{-\beta U(\mathbf{r})\}d\mathbf{r}} = \int_{D(N,V)} \chi_A(\mathbf{r}) \rho^{\Gamma}_{NVT}(\mathbf{r})\, d\mathbf{r} \quad (4.2)$$

where, in the first equality, D_A is the region of configurational space such that **r** belongs to *A*. The expression in Eq. (4.2) is quite natural, since it writes the probability of state *A* as the ratio of the number of microstates included in *A* over the total microstates number, each one weighted by its own probability. Moreover, importantly, it corresponds to the average value of a microscopic function, namely $\chi_A(\mathbf{r})$. If now we identify the partition function corresponding to a macrostate with the integral of the probability density over all its constituting microstates, and using Eq. (3.24), we can express the free energy difference between two macrostates as:

$$\Delta F_{AB} = -k_B T \ln\left[\frac{Q_B(N,V,T)}{Q_A(N,V,T)}\right] = -k_B T \ln\left[\frac{Z_B(N,V,T)}{Z_A(N,V,T)}\right] = -k_B T \ln\left[\frac{\wp(B)}{\wp(A)}\right] \quad (4.3)$$

where the second equality holds only if the definition of state *A* does not constrain the kinetics of the system differently from that of *B* so that the integrals over the momenta cancel out (see also endnote 12 after Eq. (4.10)).

The computational advantage of expression in Eq. (4.3) with respect to that in Eq. (4.1) is that it requires to evaluate the ratio of two average values that can be in principle computed, via Eq. (4.2), once a "configuration generator" compatible with the Boltzmann probability is available. This is the task of both MD and MC methods. Evaluating free energy differences with these techniques is still not trivial, due to the potentially very complex structure of the set D_A and to the nature of the microscopic observable to be averaged: the indicator function of a set is in fact a highly irregular function, abruptly changing from 0 to 1. A number of advanced methods have been developed to overcome these difficulties and are described in later contributions of this issue.

To conclude, we briefly mention that there are other possible approaches that estimate the free energy difference by separately estimating the free energy of the single states according to the following scheme:

$$\Delta F_{binding} = F_{complex} - F_{protein} - F_{ligand} \quad (4.4)$$

where each absolute free energy refers to a consistent set of control variables T, V, $\{N_k\}$. The difficulties in a full free energy calculation are in this expression often faced by means of implicit solvent and mean field approximations.

4.2. Free energy profiles and reaction paths

The ideas described in the previous subsection focus on evaluating the free energy difference between the initial and final states of a process, but provide no information on how the system evolves from one state to the other. A more detailed description is necessary if one is interested, for instance, in the mechanism or the rate of occurrence of the process. Considerable progress can be achieved by identifying the reaction coordinate, that is a variable whose values $\bar{\xi}$ characterize the state of advancement of the process. It is indeed possible to identify, by tackling the problem in the framework of the theory of stochastic processes, a "universal" and exact reaction coordinate (usually referred to as the *committor* function) that fulfills this task (13). The committor is however usually impossible to calculate (or to relate to an intuitive description of the process), so suitable approximations are commonly sought instead. In particular, in the following we shall use the fairly general assumption that the values of the reaction coordinate ξ are observable (and thus, in our scheme, "macroscopic") and can be obtained via a function $\boldsymbol{\xi}(\mathbf{r})$ of the coordinates of the system alone (no momenta). This function can map one, but usually more, microscopic configurations to a specific $\bar{\xi}$. Now, given a set of samples of the configurational space, generated by MC, MD or any other *ad hoc* tool, we can, through the $\boldsymbol{\xi}(\mathbf{r})$ mapping, obtain the corresponding samples in the $\bar{\xi}$ space, that can be interpreted as realizations of a putative random variable ξ. The importance of knowing the statistics of this random variable will be discussed in a moment.

A few more comments should be made on the nature and the choice of the mapping. Typically, the reaction coordinate is approximated by one or more quantities that are expected to be influential on the process

to be studied. In the protein-ligand case, the most intuitive is the distance between some atoms of the ligand and some of the binding site. This variable alone, however, may not be sufficient to univocally identify the binding process and is therefore complemented by other variables such as relative orientation, contact maps, water coordination values and so on. The reaction coordinate approximants usually undergo the name of collective variables (CVs), or order parameters or reaction variables, depending on the specific context; they may lead to a $\boldsymbol{\xi}(\mathbf{r})$ which is a vectorial, rather than scalar, function. CVs can in principle be either continuous or discrete variables, but we will assume in the following the former kind. The number of CVs is usually much smaller than that of the coordinates of the system, this dimensional reduction implies, in general, that many configurations correspond to the same $\hat{\xi}$ value.

Having said that the statistics describing the equilibrium distribution in the configurational space can be reduced, as far as the description of the reaction is concerned, to the statistics of the $\boldsymbol{\xi}$ random variable, we can use this latter to derive the probability of different states. We can also extend the ideas of Sect. 4.1.2. to characterize not only the free energy difference between initial and final states but also the free energy profile along the reaction's progress, provided our mapping $\boldsymbol{\xi}(\mathbf{r})$ provides a good approximation of the reaction coordinate. As it is customary in probability theory, the probability density of the $\boldsymbol{\xi}$ variable can be obtained by differentiation with respect to $\hat{\xi}$ of the corresponding cumulative distribution function represented by the probability of all of the microstates compatible with the $\xi(\mathbf{r}) \leq \hat{\xi}$ condition. The indicator function in this case takes the form of a step function:

$$\chi_{\xi(\mathbf{r})\leq\hat{\xi}}(\mathbf{r}) = \Theta\left(\hat{\xi} - \xi(\mathbf{r})\right) \tag{4.5}$$

where $\Theta(\cdot)$ is the Heaviside function, so that[3]:

$$\rho^{\mathrm{r}}_{\mathrm{NVT}\xi}(\hat{\xi}) = \frac{\mathrm{d}}{\mathrm{d}\hat{\xi}}\int_{\mathrm{D(N,V)}} \Theta\left(\hat{\xi} - \xi(\mathbf{r})\right)\rho^{\mathrm{r}}_{\mathrm{NVT}}(\mathbf{r})\,\mathrm{d}\mathbf{r} = \\ = \int_{\mathrm{D(N,V)}} \delta(\xi(\mathbf{r}) - \hat{\xi})\rho^{\mathrm{r}}_{\mathrm{NVT}}(\mathbf{r})\,\mathrm{d}\mathbf{r} \tag{4.6}$$

Here, the Dirac's delta can be interpreted as a counter of the microstates compatible with the condition $\xi(\mathbf{r}) = \hat{\xi}$. Consequently, the probability of a state can be expressed as the integral of (4.6) over the values that correspond to that state and Eq. (4.3) can be used to derive the free energy of any state with respect to a given reference[9]. With this definition, the free energy becomes a parametric function of the states described by $\hat{\xi}$, whose plot is referred to as the free energy profile (or surface, FES, for multidimensional CVs). From the practical point of view, if we can assume a smooth behavior of the probability density in Eq. (4.6), and if we have a suitable set of configurational samples of our system, we can partition the samples in bins of size Δ with respect to $\hat{\xi}$ and use the relative frequency of samples per bin as an estimator P_{est} of the probability to find the system in a state characterized by a given $\hat{\xi}$ value[10]:

$$\mathrm{F}_{\mathrm{NVT}\xi}(\hat{\xi}) - \mathrm{F}_{\mathrm{ref}} = -\mathrm{k_B T}\ln\left[\frac{\int_{\hat{\xi}-\frac{\Delta}{2}}^{\hat{\xi}+\frac{\Delta}{2}}\rho^{\mathrm{r}}_{\mathrm{NVT}\xi}(\xi)\mathrm{d}\xi}{\int_{\xi_{\mathrm{ref}}-\frac{\Delta}{2}}^{\xi_{\mathrm{ref}}+\frac{\Delta}{2}}\rho^{\mathrm{r}}_{\mathrm{NVT}\xi}(\xi)\mathrm{d}\xi}\right] \cong \\ \cong -\mathrm{k_B T}\ln\frac{\mathrm{P_{est}}(\hat{\xi})}{\mathrm{P_{est}}(\xi_{\mathrm{ref}})} \tag{4.7}$$

A possible alternative way to get to the expression above for the free energy relies on the concept of internal constraint introduced in Sect. 4.1. Adopting this thermodynamic point of view, the collective variable is seen as a further control variable that can be applied to the system, i.e. a constraint. In this interpretation, $F_{NVT\xi}$ is the actual free energy of the system where we imposed the $\xi(\mathbf{r}) = \hat{\xi}$ constraint. Evidently, while the conceptual or computational imposition of such a constraint is not particularly problematic, its experimental counterpart could be even impossible to realize in practice. One interesting consequence of this observation and of the *minimum energy principle* is that the system exerts a force on the constraint pushing it in a way that lowers the free energy. In partial analogy to the classical mechanical case, where the force exerted by a conservative field is the opposite of the gradient of the potential energy, we can observe what happens when $\boldsymbol{\xi}(\mathbf{r})=[\mathbf{r}_1;\ldots;\mathbf{r}_{n<<N}]$. In this case $F_{NVT\xi}$ coincides with the so-called Potential of Mean Force and one can write (14):

$$-\nabla_{\hat{\xi}}\,\mathrm{F}_{\mathrm{NVT}\xi}(\hat{\xi}) = \langle\vec{\mathrm{F}}_{\xi}\rangle_{\mathrm{NVT}\hat{\xi}} \tag{4.8}$$

where $\vec{F}_{\xi}$ is the force acting on the *n* constrained particles[11].

Another useful consequence of the internal constraint perspective can be seen if we multiply both sides of Eq. (4.6) by Z(N,V,T), and obtain the following expression:

$$\mathrm{Z(N,V,T)}\rho^{\mathrm{r}}_{\mathrm{NVT}\xi}(\hat{\xi}) = \int_{\mathrm{D(N,V)}}\delta(\xi(\mathbf{r}) - \hat{\xi})\exp(-\beta\mathcal{U}(\mathbf{r}))\,\mathrm{d}\mathbf{r} \triangleq \\ \triangleq \mathrm{Z(N,V,T},\hat{\xi}) \tag{4.9}$$

In this context $Z(N,V,T,\hat{\xi})$ can be interpreted as the configurational integral of the system where the function $\boldsymbol{\xi}(\mathbf{r})$ is constrained to assume the value $\hat{\xi}$. Analogously, an expression very similar and substantially equivalent to that of Eq. (4.7) can be used to indicate the free energy of the constrained state:

$$\mathrm{F}_{\mathrm{NVT}\xi}(\hat{\xi}) = -\mathrm{k_B T}\ln\left[\frac{\rho^{\mathrm{r}}_{\mathrm{NVT}\xi}(\hat{\xi})}{\mathrm{C}}\right] \tag{4.10}$$

where C is a constant which accounts for the reference state and dimensional consistency[12].

4.2.1. The case of an harmonic restraint

The application of a holonomic, i.e. of the form $\xi(\mathbf{r}) = \hat{\xi}$, constraint is sometimes tricky to be imposed in a simulation (15). A commonly adopted alternative is to restrain the value of the CV via a biasing potential of the form:

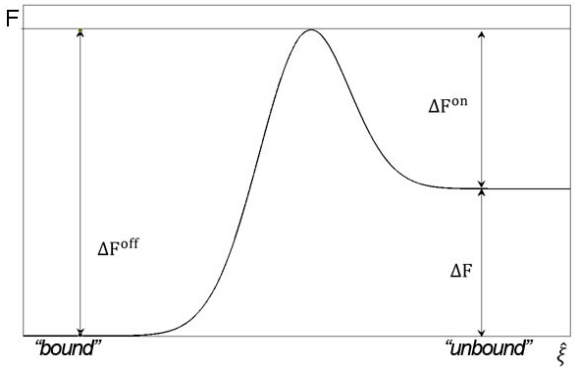

Figure 1. Schematic representation of the free energy profile in a binding process as a function of the reaction coordinate value $\tilde{\xi}$. ΔF is the free energy difference between the unbound and the bound states.

$$\mathcal{U}_{\text{bias}}(\mathbf{r}) = \tfrac{1}{2}k\left(\xi(\mathbf{r}) - \tilde{\xi}\right)^2 \tag{4.11}$$

where k is the spring constant, the higher its value, the stiffer is the spring and the stronger is the restraint. The probability density corresponding to this situation is:

$$\rho^{r}_{\text{NVTbiased}}(\mathbf{r},\tilde{\xi}) = \frac{\exp\{-\beta[\mathcal{U}(\mathbf{r})+\mathcal{U}_{\text{bias}}(\mathbf{r})]\}}{\int_{D(N,V)}\exp\{-\beta[\mathcal{U}(\mathbf{r})+\mathcal{U}_{\text{bias}}(\mathbf{r})]\}d\mathbf{r}} = \\ = \frac{\exp\left\{-\frac{\beta k}{2}\left(\xi(\mathbf{r})-\tilde{\xi}\right)^2\right\}\exp(-\beta\mathcal{U}(\mathbf{r}))}{\int_{D(N,V)}\exp\left\{-\frac{\beta k}{2}\left(\xi(\mathbf{r})-\tilde{\xi}\right)^2\right\}\exp(-\beta\mathcal{U}(\mathbf{r}))d\mathbf{r}} \tag{4.12}$$

By multiplying both numerator and denominator by Z(N,V,T) and after some algebraic manipulation, the expression above becomes:

$$\rho^{r}_{\text{NVTbiased}}(\mathbf{r},\tilde{\xi}) = \frac{\exp\left\{-\frac{\beta k}{2}\left(\xi(\mathbf{r})-\tilde{\xi}\right)^2\right\}\rho^{r}_{\text{NVT}}(\mathbf{r})}{\langle\exp\left\{-\frac{\beta k}{2}\left(\xi(\mathbf{r})-\tilde{\xi}\right)^2\right\}\rangle_{\text{NVT}}} = \\ = \frac{\sqrt{\frac{\beta k}{2\pi}}\exp\left\{-\frac{\beta k}{2}\left(\xi(\mathbf{r})-\tilde{\xi}\right)^2\right\}\rho^{r}_{\text{NVT}}(\mathbf{r})}{\langle\sqrt{\frac{\beta k}{2\pi}}\exp\left\{-\frac{\beta k}{2}\left(\xi(\mathbf{r})-\tilde{\xi}\right)^2\right\}\rangle_{\text{NVT}}} \tag{4.13}$$

Due to the fact that the Dirac delta can be approximated by a Gaussian function, as the width of this latter tends to zero, we see that:

$$\rho^{r}_{\text{NVTbiased}}(\mathbf{r},\tilde{\xi}) \xrightarrow{k\to\infty} \delta\left(\xi(\mathbf{r})-\tilde{\xi}\right)\frac{\rho^{r}_{\text{NVT}}(\mathbf{r})}{\rho^{r}_{\text{NVT}\xi}(\tilde{\xi})} \tag{4.14}$$

which is, a we will see later, the probability density of the system in the constrained configurational space obtained by fixing $\tilde{\xi}$ as well as NVT.

The practical usefulness of the harmonic restraint stems from the fact that it can be imposed by simple addition of a biasing potential. Importantly, it also provides an estimator of the free energy gradient with respect to the CV. Adding the biasing to the physical potential in Eq. (4.1) and differentiating with respect to $\tilde{\xi}$ one obtains:

$$-\nabla_{\tilde{\xi}} F_{\text{NVTbiased}}(\tilde{\xi}) = \frac{k_B T\int_{D(N,V)}\nabla_{\tilde{\xi}}\exp\left\{-\frac{\beta k}{2}\left(\xi(\mathbf{r})-\tilde{\xi}\right)^2\right\}\exp(-\beta\mathcal{U}(\mathbf{r}))d\mathbf{r}}{\int_{D(N,V)}\exp\left\{-\frac{\beta k}{2}\left(\xi(\mathbf{r})-\tilde{\xi}\right)^2\right\}\exp(-\beta\mathcal{U}(\mathbf{r}))d\mathbf{r}} = \\ = \frac{\int_{D(N,V)}k\left(\xi(\mathbf{r})-\tilde{\xi}\right)\exp\{-\beta[\mathcal{U}(\mathbf{r})+\mathcal{U}_{\text{bias}}(\mathbf{r})]\}d\mathbf{r}}{\int_{D(N,V)}\exp\{-\beta[\mathcal{U}(\mathbf{r})+\mathcal{U}_{\text{bias}}(\mathbf{r})]\}d\mathbf{r}} = \\ = \langle k\left(\xi(\mathbf{r})-\tilde{\xi}\right)\rangle_{\text{NVTbiased}} \tag{4.15}$$

Note that when CVs are not just linear combinations of the Cartesian coordinates of the system, the equation above is a generalization of the definition of mean force. Expression (4.15) has an intuitive interpretation that relates the free energy gradient to the average elastic force exerted on the harmonic restraint by the system.

4.2.2. Characteristics of the free energy profile

Let us now consider a simplified description in the case of protein-ligand binding; the function $\xi(\mathbf{r})$:D(N,V)→[0;+∞], that could be, for instance, the distance from the ligand to the binding site, maps a set of coordinates {**r**} on the scalar parameter $\tilde{\xi}$. In the mentioned context of bound and unbound states, the "unbound" region A can be mapped to the $[\tilde{\xi}_A;+\infty]$ interval while the "bound" region B is mapped to $[0;\tilde{\xi}_B]$, with $\tilde{\xi}_B < \tilde{\xi}_A$13. Intermediate values of the parameter $\tilde{\xi}$ correspond to a continuum of intermediate states that can be visited during the process and that depend on the specific mapping induced by ξ(r). The presence of states which are neither bound nor unbound is physically sound since it accounts for a gradual "switching on" of the interaction and for the possible existence of states where protein and ligand interact in a way which is non functional to the sought biological activity. A typical free energy profile in protein-ligand binding is shown in Figure 1. It is characterized by regions of low free energy separated by a barrier. These regions (which correspond to high probability values for ξ) are associated to the $\tilde{\xi}$ values that identify the reactants (unbound state "*A*") and products (bound state "*B*") of the binding process.

The free energy profile contains more information than just the equilibrium population of some states. In fact, free energy, through its gradient, indirectly rules the average force acting over the system as the reaction progresses, and its complete profile accounts not only for thermodynamic but also for average kinetic properties. When the reaction can be schematized as the transition between two basins separated by a barrier, the overall free energy difference affects the equilibrium population of the two states, but the energy profile encountered by the system during the reaction decides the rates at which the transitions occur. More specifically, the inward barrier height determines the rate of binding, namely k_{on}, while the outward barrier height determines of the unbinding rate, k_{off}. The shape of the profile indicates that there are low (free) energy "long lived" states, namely bound and unbound, where the system spends most of the time but there are also unlikely configurations, which in the $\tilde{\xi}$ representation are positioned around the top of the barrier, that, although energetically unfavorable, must be visited when the system moves from A to B and vice versa, i.e. during the (un)binding events. The state corresponding to these configurations is named *transition state* (TS). When

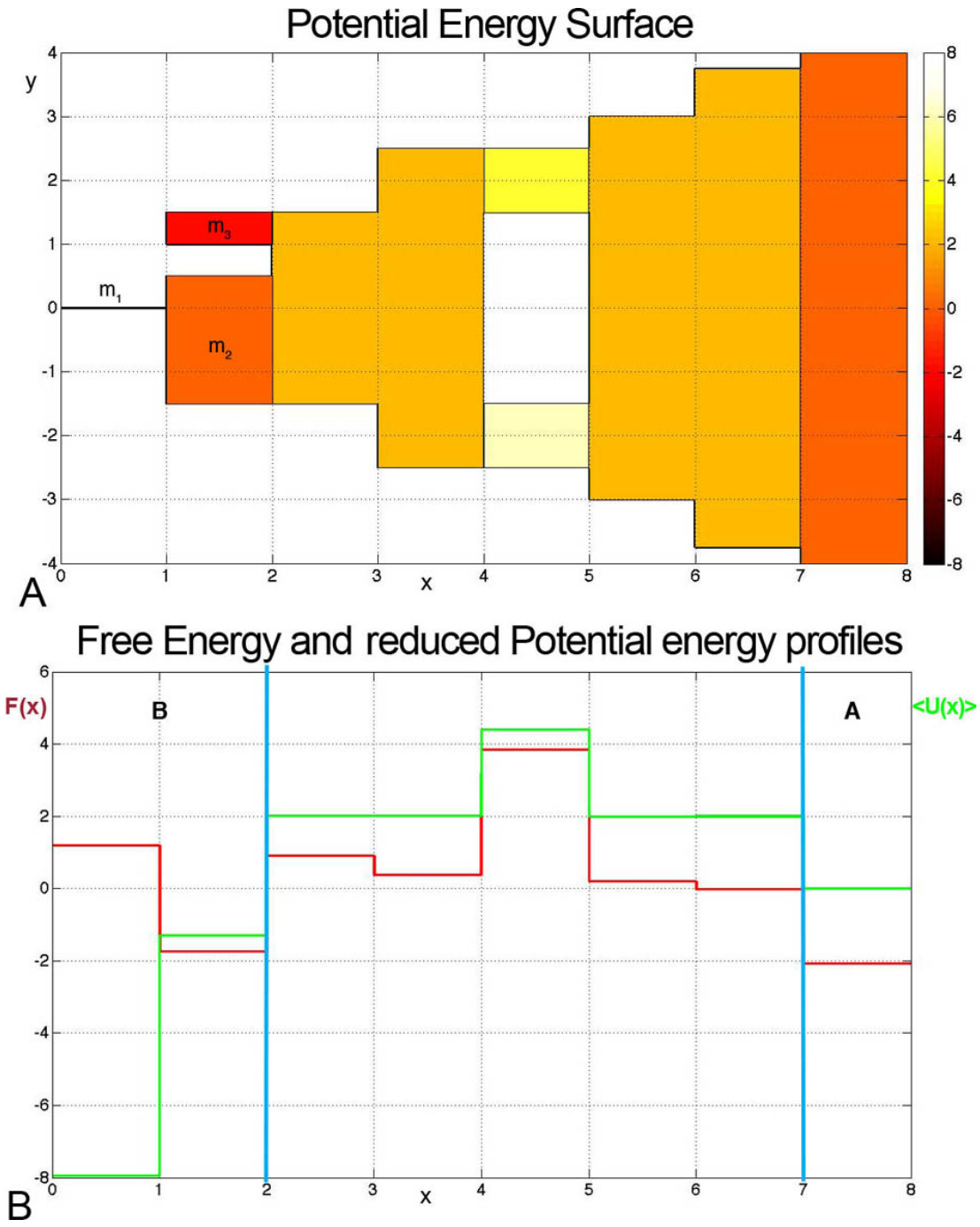

Figure 2. a) potential energy landscape of a 2D paradigmatic example; the depth is color coded as shown in the right bar, from brown to white. b) free energy profile (red) and potential energy (green) of a (NVT$\tilde{\xi}$) constrained system, where $\tilde{\xi}$=x. Vertical sky-blue bars indicate the choice of unbound and bound states, A and B, respectively.

the barrier is considerably larger than kBT the computational evaluation of the free energy profile becomes very expensive. This is qualitatively explained by the fact that a high barrier implies a low probability of sampling the TS. For instance, if the sampling is done via a MD simulation it will be unlikely that the trajectory, mapped on the ξ space, crosses the barrier. If, in contrast, the configurational samples are obtained through MC, the probability that a MC move pointing to the TS is accepted is extremely low, leading to a bad characterization of the TS. So called *enhanced sampling techniques*, described in other contributions of this issue, aim at improving the sampling in low probability regions,

such as the TS, which are however important for the description of the reaction.

4.3. The definition of bound and unbound states and the reaction coordinate

A thorough characterization of bound and unbound states is far from trivial, see Gilson et al. for reference, especially if the protein and the ligand form a loosely bound complex or if there are peripheral interaction sites (16). Basically, the region of the configurational space that must be counted in the bound (*B*) state should include all the configurations where protein binding site and the ligand make a strong interaction while all non interacting microstates contribute to the unbound state (*A*). Differences in momenta space are also plausible, with lower instantaneous kinetic energy values for bound than for unbound states, but a clear distinction is made unlikely by the large fluctuations of kinetic energy of the solvated ligand. It is not expected, however, a strong dependence of the bound/unbound free energy difference on the precise definition of the states, this aspect could be seen as a defining characteristic of a complex (16). It is important to note, however, that the definition of *A* and *B* might depend on many factors, such as the experimental protocol one aims at describing. For instance, considering the signal provided by a FRET experiment, one obtains basically a two-state characterization of the process, where all the microstates involving some ligand-protein interaction occurring out of the binding site are to be assigned to the unbound state *A* (17). This latter is a different definition from the one given previously and may well lead to numerical discrepancies in the free energy estimates. From the energetic point of view, we should imagine the unbound state as a very large free energy basin, substantially flat, whereas the bound state is more limited and likely presents a rougher landscape and several different minima. There are basically no rules concerning the landscape of the remaining part of the configurational space, although a rough profile with plenty of peaks and a few valleys is to be expected.

An accurate identification of the reaction coordinate is even more complicated than the definition of the states, as already mentioned. We would like to stress again that the free energy profile relies crucially on the choice of the CVs and that there exists a rigorous and "universal" definition of a reaction coordinate, called the *committor*, derived in the context of Transition State Theory (TST), whose description is out of the scope of the present contribution and whose calculation in practice is prohibitive (13). Testing whether a particular choice of CVs is good or not is again a topic pertaining to TST and usually heuristic approaches are adopted to select putative CVs to be tested later. We should also stress that such a test should be always performed since a bad choice of the CVs can, and in several cases does, lead to misinterpreting the process. As an example, distance *per se* cannot exhaustively describe the binding process; for instance, it does not uniquely identify the relative position of the ligand with respect to the binding site (18). Most likely, it is flawed thinking to believe that the same set of collective variables can describe the protein-ligand process in all cases, since Biology got us used to seeing very different mechanisms adopted in different cases, according to the specific needs of the process.

As it has already been mentioned, following the reaction coordinate is not mandatory for the correct estimation of the free energy difference but only for reconstructing the intermediate phases; as a matter of fact, there are techniques that envision the transition occurring along unphysical, sometimes called "alchemical", paths and that provide nonetheless a reliable estimation of the free energy difference.

4.4. Potential versus free energy surface

The most intuitive energetic landscape where the system evolution can be represented is the potential energy surface (PES) in the 3N-dimensional configurational space. PES is also the natural landscape for constant energy (isolated) systems. However, we have seen that for systems in thermal equilibrium with the environment and when discussing the evolution between "macrostates" specified by collective variables, the free energy is the right potential to consider. The Free Energy Surface (FES), is defined in a space with the same dimension of $\hat{\xi}$. In order to better understand similarities and differences between the two quantities, the following considerations can be useful. First, let us imagine to employ as CVs the 3N coordinates of the system $\xi(\boldsymbol{r}) \triangleq \boldsymbol{r} = [\boldsymbol{r}_1; \dots; \boldsymbol{r}_N]$ so that $\hat{\xi} = \hat{\boldsymbol{r}} = [\hat{\boldsymbol{r}}_1; \dots; \hat{\boldsymbol{r}}_N]$. In this case, by means of Eqs. (4.6) and (4.10), the free energy coincides with the potential energy:

$$\mathrm{F}(\mathrm{N}, \mathrm{V}, \mathrm{T}, \hat{\mathbf{r}}) = \mathrm{F}(\mathrm{N}, \mathrm{V}, \mathrm{T}, \{\hat{\mathbf{r}}_1; \dots; \hat{\mathbf{r}}_\mathrm{N}\}) = \mathcal{U}(\hat{\mathbf{r}}) \quad (4.16)$$

This result leads to the following observations: i) if really all coordinates are necessary to characterize the advancement of the process (which is extremely hard to imagine), then the only relevant effects are those coming directly from the physical interactions within the system. In this case the free energy "looses" its thermodynamic aspect and the evolution of the system is described by mechanics alone. ii) If, on the other hand, $\xi(r)$ lives in a space of lower dimension than 3N, the behavior of free energy is richer. Eq. (4.6) and (4.10) state that any arrangement of the coordinates resulting in the same value of the collective variable (weighted by the corresponding Boltzmann factor) contributes to the free energy at a given $\hat{\xi}$. The Dirac delta acts as a selector of these configurations. This provides information about the multiplicity of the microscopic configurations that correspond to the same value of the "control variable" ξ and corresponds to the entropic contribution in the thermodynamic definition of the free energy. The argument above becomes

more evident in the classical low temperature limit, where it is possible to show that $Z(NVT) \xrightarrow{\beta\to\infty} \alpha \exp[-\beta \mathcal{U}(\boldsymbol{r}^*)]$ with r* being the configuration of minimum potential energy, see 4.7.1. Consistently with the intuitive interpretation that a fully classical system at low temperature is confined in the lowest potential energy basin, the free energy tends to the absolute minimum of $\mathcal{U}$ (see Eq. (4.1)).

Two possible ways of comparing potential energy versus free energy differences follow:

- comparing two individual configurations, namely $\hat{\boldsymbol{r}}$ and $\tilde{\boldsymbol{r}}$. In this case the correct comparison is between $(\mathcal{U}(\hat{\boldsymbol{r}}) - \mathcal{U}(\tilde{\boldsymbol{r}}))$ and $\left(F(N,V,T,\xi(\hat{\boldsymbol{r}})) - F(N,V,T,\xi(\tilde{\boldsymbol{r}}))\right)$. This analysis is not very informative since in the free energy the individual contribution of the $\hat{\boldsymbol{r}}$ and $\tilde{\boldsymbol{r}}$ microstates is not very meaningful and it should not be distinguished from that of all the other microstates that are mapped by the function $\xi(\cdot)$ to the same value of the CV;
- comparing the collective states identified by two different values of the internal constraint in the NVT ensemble. In this case, the relationship to be considered is $F(N,V,T,\bar{\xi}) = U_{\bar{\xi}} - TS_{\bar{\xi}} = \langle K\rangle_{NVT\bar{\xi}} + \langle \mathcal{U}\rangle_{NVT\bar{\xi}} - TS_{\bar{\xi}}$. Since different values of the constraint do not affect the kinetic energy, the following expression holds for the energy differences:

$$F(N,V,T,\bar{\xi}_B) - F(N,V,T,\bar{\xi}_A) = \\ = \left(\langle \mathcal{U}\rangle_{NVT\bar{\xi}_B} - \langle \mathcal{U}\rangle_{NVT\bar{\xi}_A}\right) - T\left(S_{NVT\bar{\xi}_B} - S_{NVT\bar{\xi}_A}\right) \quad (4.17)$$

Changes in free energy, which drive processes, may therefore result from changes in average potential energy or in entropy, which may work in synergy or against each other. It can be interesting, in this context, to describe how the average of a microscopic function can be expressed in a system where also a CV is constrained. In the case of potential energy, from expression (4.9) of the configurational integral, we obtain:

$$\langle \mathcal{U}(\mathbf{r})\rangle_{NVT\bar{\xi}} = -d\ln\left(Z(N,V,T,\bar{\xi})\right)/d\beta = \\ = \int_{D(N,V)} \mathcal{U}(\mathbf{r})\delta(\xi(\mathbf{r}) - \bar{\xi}) \frac{\exp(-\beta\mathcal{U}(\mathbf{r}))}{Z(N,V,T,\bar{\xi})} d\mathbf{r} \quad (4.18)$$

This relationship outlines the role of $\delta(\xi(\boldsymbol{r}) - \bar{\xi}) \frac{\exp(-\beta\mathcal{U}(\boldsymbol{r}))}{Z(N,V,T,\bar{\xi})} = \delta(\xi(\boldsymbol{r}) - \bar{\xi}) \frac{\rho^{\boldsymbol{r}}_{NVT}(\boldsymbol{r})}{\rho^{\boldsymbol{r}}_{NVT\bar{\xi}}(\bar{\xi})}$ as the correct probability distribution of the system in the constrained configurational space obtained by fixing $\bar{\xi}$ as well as NVT.

The next section describes a very simple example where the different contributions to Eq. (4.17) emerge in a simple two-dimensional case. We will, for example, illustrate the fact that knowing the global potential energy minimum of a system is inessential when entropy accounts for a sizeable fraction of its free energy.

4.5. A didactic example

The example that follows highlights some of the peculiarities of the energetic characterization of a reaction. The key aspect we wish to stress is the role of entropic contribution, here a measure of the number of states having the same potential energy value, in the free energy behavior. Let us consider a system made of two one-dimensional particles moving in the piecewise constant potential energy represented in Figure 2a. The configurational space here is identified by {r = (x,y)}. In the figure, the PES, that is the graph of $\mathcal{U}$(r), is shown in a color coded fashion. The PES presents three pronounced minima: one, m1, is very deep and narrow around the point (0,0); the other two, m2 and m3, are less deep but much wider disjoint minima with x∈(1;2). Between x=2 and x=4 and between x=5 and x=7 we have two featureless regions where there are no significant interactions. Notice however that, as x increases, so does also the number of accessible states. In between these regions there are two passes of different height. Finally, for x greater than 7, there is a plateau of null potential energy and maximum number of accessible states. We use this landscape as an extremely simplified model for the binding of a ligand and a protein. Intuitively, one would expect that deep minima in the potential represent bound states of the system, while regions of no interactions model the free ligand and protein. Given the features of the potential in Figure 2a, a simple (and reasonable) choice of CV seems to be ξ(r)=x. The region where $x\epsilon(0;\bar{\xi}_B = 2)$ represents the bound state, whereas $x\epsilon(\bar{\xi}_A = 7;8)$ the unbound one. In Figure 2b the FES (in red) and the average potential energy of the constrained system (green) profiles are shown. For this simple system all the statistical mechanical quantities can be analytically expressed both in the unconstrained and in the constrained system. With the potential represented in panel *a*, one gets the following probabilities of observing states *B* and *A*, respectively: $\wp(B) = 0.35$ and $\wp(A) = 0.47$. This results in a free energy difference in kBT units, see Eq. (4.3), of $\Delta F_{AB}/k_BT = 0.29$, implying, counterintuitively, that the unbound state, *A* is the most stable between the two.

We would like to point to some very important facts that can be generalized to high dimensional spaces:

- the definition of bound and unbound states can be somehow ambiguous; the actual "bound state" could just correspond to *m1* rather than to the union of *m1*, *m2* and *m3*;
- the identification of the reaction coordinate is far from trivial and needs some knowledge of the phenomena. Even in our toy system, the constraint ξ(r)=x does not distinguish, for example, region *m2* from *m3*. If only one of these regions corresponded to the interaction with the

biologically relevant site, this choice would not correctly represent the process;

- while the average potential energy is remarkably sensitive to the presence of low potential energy regions and much less to their size, the free energy is largely influenced by the number of accessible states to the energy minima. This entropic phenomenon is very well illustrated by the fact that in our case the global FES minimum is between x=1 and x=2 whereas the minimum average potential energy is between x=0 and x=1;
- another entropic effect is visible in the FES that decreases as the number of accessible states increases (see intervals [2;4] and [5;7]), and vice versa (see interval [4;5]).

4.6. Volumetric effect on the unbound state

If we look at the behavior of the free energy in the previous example for x belonging to the [2;4] and [5;7] intervals, we can notice a decrement corresponding to the increase of the number of accessible microstates, without any other change in the potential energy. This often occurs when, as it is commonly done, the distance between ligand and protein is taken as the collective variable for representing the binding process. It is therefore worth considering it in more detail.

Let us separate the potential energy in two parts, namely $\mathcal{U}(\boldsymbol{r}) = \mathcal{U}_1(\boldsymbol{r}) + \mathcal{U}_{LP}(\boldsymbol{r})$, where the subscript LP indicates the direct interaction between the ligand and the protein; we assume, as it is usually done, that the dependence of $\mathcal{U}_1(\mathbf{r})$ on $\mathbf{r}_P$ and $\mathbf{r}_L$ is additive. Then, let us define our collective variable as the distance between $\mathbf{r}_P$ and $\mathbf{r}_L$, $\xi(\boldsymbol{r}) = \|\boldsymbol{r}_L - \boldsymbol{r}_P\|$, P and L being one atom of the binding site and one of the ligand, respectively. The constrained configurational integral can be written as follows[14]:

$$Z(\mathrm{NVT}\hat{\xi}) = \int_{D(N,V)} \delta(\xi(\mathbf{r}) - \hat{\xi}) \exp(-\beta \mathcal{U}_1(\mathbf{r})) \exp(-\beta \mathcal{U}_{LP}(\mathbf{r})) \, d\mathbf{r} \quad (4.19)$$

If we now make a change of variables where the vector ($\mathbf{r}_P - \mathbf{r}_L$) is expressed in spherical coordinates, so to move from $\mathbf{r}$ to $(\xi, \theta, \varphi, \tilde{\mathbf{r}})$[15] and observe that only $\mathcal{U}_{LP}$ depends on the displacement of the ligand with respect to the protein, we get the following relationship:

$$\begin{aligned} Z(\mathrm{N,V,T},\hat{\xi}) &= \int d\tilde{\mathbf{r}} \exp(-\beta \mathcal{U}_1(\tilde{\mathbf{r}})) \int_0^{\pi} d\varphi \int_0^{2\pi} d\theta \\ &\int_0^{\infty} \delta(\xi - \hat{\xi}) \exp(-\beta \mathcal{U}_{\mathrm{LP}}(\xi, \theta, \varphi, \tilde{\mathbf{r}})) \, \xi^2 \sin\varphi \, d\xi = \\ &= \int d\tilde{\mathbf{r}} \exp(-\beta \mathcal{U}_1(\tilde{\mathbf{r}})) \int_0^{\pi} d\varphi \\ &\int_0^{2\pi} \exp\left(-\beta \mathcal{U}_{\mathrm{LP}}(\xi, \theta, \varphi, \tilde{\mathbf{r}})\right) \xi^2 \sin\varphi \; d\theta \end{aligned} \quad (4.20)$$

If the ligand is constrained to stay far enough from the protein so that it can explore the whole solid angle without making any direct interaction with it, then $\mathcal{U}_{LP}$ can safely be assumed to be null, regardless of the values of θ and φ. In this region:

$$Z(NVT\hat{\xi}) = \int 4\pi\hat{\xi}^2 exp(-\beta \mathcal{U}_1(\tilde{\mathbf{r}})) \, d\tilde{\mathbf{r}} = 4\pi\hat{\xi}^2 Z_1 \quad (4.21)$$

If we define the unbound state as the one that corresponds to the interval $\xi \in [R_P; R]$, where R_p is the radius of a sphere centered in the origin so that out of it there is no direct interaction between protein and ligand, we obtain:

$$\begin{aligned} F_{unbound} - F_{ref} &= -k_B T ln \int_{R_P}^{R} Z(N,V,T,\hat{\xi}) d\hat{\xi} = \\ &= -k_B T ln \tfrac{4\pi}{3}(R^3 - R_P^3) - k_B T ln Z_1 \end{aligned} \quad (4.22)$$

One can see that the free energy of the unbound state contains a term, i.e. $4\pi[R^3-(R_P)^3]/3$, corresponding to the volume of the space region where the ligand is free to move; its independence from the potential energy indicates its entropic nature. A few remarks should to be made, to compare this derivation to a real case scenario; i) in a real case application, a more precise choice of the CV would be appropriate, considering the distance between an average position of atoms composing the binding site and that of the ligand, this complicates the calculations but does not affect the results. ii) The spherical shape for the unbound state is clearly an approximation made to simplify calculations. A more proper description would involve more complex geometries but the concept of available volume of no interaction would still remain valid. Finally, iii) we must recall that the external radius R cannot be taken arbitrarily large, since there is a global constraint on the volume of the system and this has to be chosen compatibly with the experimental concentration of reactants.

4.7. Complementary material to section 4

4.7.1. The zero temperature limit for free energy

In order to study the low temperature limit for the free energy, we will use the multivariate Laplace method of integration, which is a useful tool for approximating the integral of exponential functions when the exponent tends to minus infinity (19). We will adopt the so called harmonic approximation, in which the potential energy is expanded in Taylor expansion around its global minimum, so to retain the most relevant contribution to the integral:

$$\begin{aligned} \mathcal{U}(\mathbf{r}) = \; & \mathcal{U}(\mathbf{r}^*) + [\nabla_{\mathbf{r}} \mathcal{U}(\mathbf{r})]^{\mathrm{T}}{}_{\mathbf{r}=\mathbf{r}^*} (\mathbf{r} - \mathbf{r}^*) + \\ & + \tfrac{1}{2} (\mathbf{r} - \mathbf{r}^*)^{\mathrm{T}} \mathbf{H}(\mathbf{r}^*)(\mathbf{r} - \mathbf{r}^*) + O(\|\mathbf{r} - \mathbf{r}^*\|^3) \end{aligned} \quad (4.23)$$

where r* are the coordinates of the minimum, (assumed here to be non degenerate), $[\nabla_{\mathbf{r}} \mathcal{U}(\mathbf{r})]^{\mathrm{T}}{}_{\mathbf{r}=\mathbf{r}^*}$ is the gradient of the potential energy in r*, that in our case is null since we are in a minimum, and H(r*) is the Hessian matrix of the potential in r*, i.e. the symmetric matrix containing the second partial derivatives of the potential with respect to the coordinates. The fact that we are in a minimum guarantees that this matrix is positive definite. If we now substitute the second order, i.e. harmonic, approximation of the potential in the expression of the configurational integral, we obtain, ($\Delta\mathbf{r} = \mathbf{r}\text{-}\mathbf{r}^*$):

$$Z(N,V,T) \cong \int_{D(N,V)} \exp\{-\beta[U(\mathbf{r}^*) + \tfrac{1}{2}\Delta\mathbf{r}^T\mathbf{H}(\mathbf{r}^*)\Delta\mathbf{r}]\}\,d\mathbf{r} = \\ = \exp\{-\beta U(\mathbf{r}^*)\} \int_{D(N,V)} \exp\left\{-\frac{\beta}{2}\Delta\mathbf{r}^T\mathbf{H}(\mathbf{r}^*)\Delta\mathbf{r}\right\} d\mathbf{r} = \\ = \exp\{-\beta U(\mathbf{r}^*)\} \sqrt{\frac{1}{\det[\mathbf{H}(\mathbf{r}^*)]}} \sqrt{\left(\frac{2\pi}{\beta}\right)^{3N}} \quad (4.24)$$

With the same argument we can study the behavior of the NVT average of a generic well behaved microscopic function h(**r**), which is linearly Taylor-expanded around the **r*** point (details have been skipped):

$$\int_{D(N,V)} h(\mathbf{r})\rho^{r}_{NVT}(\mathbf{r})d\mathbf{r} = \int_{D(N,V)} \frac{h(\mathbf{r})e^{-\beta U(\mathbf{r})}}{Z(N,V,T)} d\mathbf{r} \cong \\ \cong \frac{h(\mathbf{r}^*)\exp\{-\beta U(\mathbf{r}^*)\}\sqrt{\frac{1}{\det[\mathbf{H}(\mathbf{r}^*)]}}\sqrt{\left(\frac{2\pi}{\beta}\right)^{3N}}}{\exp\{-\beta U(\mathbf{r}^*)\}\sqrt{\frac{1}{\det[\mathbf{H}(\mathbf{r}^*)]}}\sqrt{\left(\frac{2\pi}{\beta}\right)^{3N}}} = h(\mathbf{r}^*) \quad (4.25)$$

This is the definition of a Dirac delta[3] so that:

$$\rho^{r}_{NVT}(\mathbf{r}) \xrightarrow{\beta\to\infty} \delta(\mathbf{r} - \mathbf{r}^*) \quad (4.26)$$

The free energy limit can be derived by the use of (4.24) and the definition of free energy:

$$F(N,V,T) = -k_B T \ln Z(N,V,T) \cong \\ \cong U(\mathbf{r}^*) - k_B T \tfrac{1}{2} \ln \frac{(2\pi k_B T)^{3N}}{\det[\mathbf{H}(\mathbf{r}^*)]} \quad (4.27)$$

In this representation, the first term is the potential energy in the minimum and also the NVT average potential energy in this limit due to (4.26). The remaining term, which tends to zero as the temperature approaches the absolute zero, has therefore an entropic origin and depends on the shape, in particular the curvature, of the potential energy around its minimum, through its partial derivatives with respect to the coordinates.

5. CONCLUSIONS

In this paper, we reviewed some of the basic ingredients for studying, using computer simulation, ligand protein binding. We begun by formulating a suitable microscopic model of the system, and by quickly reviewing its (classical) evolution equations. We stressed the connection among microscopic and macroscopic observables, and showed how some significant thermodynamic quantities, e.g. temperature, pressure, can be obtained as averages of suitable functions of the microscopic configurations of the system in phase space. We also illustrated the equivalence, under the ergodic hypothesis, of the concept of time and ensemble average which underlie molecular dynamics and Monte Carlo simulations, respectively. These ideas were presented both for an isolated system, microcanonical ensemble, and for a system in thermal equilibrium with the environment, canonical ensemble; the equivalence of these two ensembles was briefly discussed. We then introduced the key quantity of this contribution, the (Helmholtz) free energy, and argued how it can be used, once an appropriate set of collective variables has been identified, to characterize the progress of an activated event, such as the protein-ligand binding, from the reactants to the product and *vice versa* and to obtain insights on the mechanism and the rate of the process. We discussed some of the difficulties of a brute force calculation of the free energy and the similarities and differences among free and potential energy. These concepts were illustrated on a simple toy model. We concluded with an analysis of some specific issues related to the evaluation of the free energy for ligand protein binding.

The material contained in this introductory review is, inevitably, incomplete and, probably, biased by the experience and taste of the authors. Excellent in depth discussions of the topics presented here can be found in the scientific literature and in text books (some present in our bibliography). Our goal was to illustrate how the very general concepts and tools discussed are useful for addressing many different aspects of the ligand protein binding process. In this spirit, we attempted to present a coherent framework that may be useful as a first "guide" to the concepts discussed and that sets the stage for more specialized contributions.

6. ACKNOWLEDGEMENTS

The Authors gratefully acknowledge the IIT platform "Computation". Funding from the IIT SEED project SIMBEDD is also acknowledged.

7. REFERENCES

1. L D Landau, E M Lifshitz. Statistical Physics. Pergamon Press (1980)

2. K Huang. Statistical Mechanics. Wiley (1987)

3. J R Waldram. The theory of thermodynamics. Cambridge University Press (1985)

4. M P Allen, D Tildesley. Computer Simulation Of Liquids. Clarendon Press, Oxford, UK (1987)

5. D Frenkel, B Smit. Understanding Molecular Simulation. Academic Press (2001)

6. M E Tuckerman. Statistical Mechanics: Theory and Molecular Simulations. Oxford University Press, UK (2010)

7. W Greiner, L Neise, H Stöcker. Thermodynamics and Statistical Mechanics. Springer (1995)

8. S Nosé: A unified formulation of the constant temperature molecular dynamics methods. *J Chem Phys* 81, 511 (1984)

9. W G Hoover: Canonical dynamics: Equilibrium phase-space distributions, *Phys Rev A* 31, 1695–1697 (1985)

10. T Schlick. Molecular Modeling and Simulation. Springer (2002)

11. H Risken. The Fokker-Planck Equation: Methods of Solution, Applications. Springer-Verlag, Berlin, New York (1984)

12. D Chandler. Introduction to Modern Statistical Mechanics. Oxford University Press, USA (1987)

13. E Vanden-Eijnden. Transition Path Theory. In: Computer Simulations in Condensed Matter: From Materials to Chemical Biology. Eds. M. Ferrario, G. Ciccotti, Springer, (2006)

14. D A McQuarrie. Statistical Mechanics. Sausalito, California: University Science Books (2000)

15. M Sprik, G Ciccotti: Free energy from constrained molecular dynamics. *J Chem Phys*, 109, 7737 (1998)

16. M K Gilson, J A Given, B L Bush, J A McCammon: The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. *Biophys J,* 72, 1047-1069 (1997)

17. T M Jovin, D J Arndt-Jovin. FRET microscopy: Digital imaging of fluorescence resonance energy transfer. Application in cell biology. In: Cell Structure and Function by Microspectrofluometry. Eds. E. Kohen, J. G. Hirschberg and J. S. Ploem. Academic Press, London, UK, 99–117 (1989)

18. M Masetti, A Cavalli, M Recanatini, F L Gervasio: Exploring complex protein-ligand recognition mechanisms with coarse metadynamics. *J Phys Chem B*, 113, 4807-4816 (2009)

19. A Azevedo-Filho, R Shachter. Laplace's Method Approximations for Probabilistic Inference in Belief Networks with Continuous Variables. In: Mantaras, R.; Poole, D., Uncertainty in Artificial Intelligence. Morgan Kauffman, San Francisco, (1994)

Endnotes:

[1] During its exploration of the phase space, an ergodic system will go, in the long run, arbitrarily close to every conceivable microstate. Ergodicity is assumed in all of the systems studied in this kind of works, albeit for complex systems, such as biological ones, it is impossible to be proven.

[2] That is, the average over many statistically equivalent replicas of the system. This condition is resembling experimental situations where the same observable is measured over several equivalent samples.

[3] For the scope of the present work, the Dirac's delta has to be seen just as an integral kernel such that, for a suitable class of test functions f(x), it happens that $\int f(x)\delta(x-x_0)dx=f(x_0)$, provided that the domain of integration contains the point x_0. The physical dimension of a Dirac's delta is the reciprocal of that of its argument. It can be thought as the generalized derivative of the Heaviside step function. Its behavior can be approximated by the limit of a Gaussian function as its width goes to zero: $\frac{1}{\sqrt{2\pi}\sigma} exp\frac{-(x-x_0)^2}{2\sigma^2} \xrightarrow{\sigma\to 0} \delta(\mathbf{x}-\mathbf{x_0})$.

[4] The assignment of the volume element $d\mathbf{r}_i d\mathbf{p}^r_i$ to the particle "i" is somehow arbitrary since the Hamiltonian would remain unchanged upon re-labeling of particles of the same chemical species. To avoid over-counting of the (different) microscopic states accessible to the system, a prefactor containing the factorial of all the numbers of identical particles, must be included.

[5] In this concise description, the number N stands for the composition of the system and can be thought either as a synonym of N_η in the NVE expression or as $\Sigma_\eta N_\eta$ in expressions such as h^{3N} or $\Sigma^N_{i=1}(\cdot)$. The subscript η spans over all the chemical species composing the system. In the present formulation two entities are chemically identical if the Hamiltonian is not affected by their exchange; this should not engender any ambiguity in the chemical context where a whole molecule can be considered as a chemical species.

[6] Unlike in (bio)chemistry, where the terms *configuration* and *conformation* have two precise and different meanings, in this context we will identify a configuration with a point in phase space after integration of the momenta, characterized by the conformation(s) of the macromolecule(s) as well as by the positions and orientations of solvent molecules.

[7] This statement applies since we restrict our attention to time-independent equilibrium properties. More care would be necessary in discussing, for example, equilibrium time correlation functions.

[8] Throughout the work we distinguish between time-invariant, i.e. constant, quantities, and space-invariant, i.e. uniform, quantities.

[9] Here we imagine, having the protein-ligand case in mind, that, both bound and unbound states correspond to intervals of values of the CV. If we use their mutual distance *d* as CV, for example, we can associate the unbound state with the values $d > d_{th}$, where d_{th} is a given threshold value.

[10] It is possible to connect this result to the one in Eq. (4.4) by thinking that state *B* corresponds to the state of the complex and that in state *A* the protein and the ligand do not interact, therefore $F(N,V,T,\text{unbound state}) = F_{protein} + F_{ligand}$. To avoid any misinterpretation, we would like to stress that extensive quantities can be algebraically added only when the corresponding subsystems are non-interacting (or at least interacting weakly and non

specifically) so that their state distributions can be considered independent. This can be proven trivially in the following cases:

- Internal energy:

$U_{A\cup B}=\langle H_{A\cup B}\rangle = \langle H_A + H_B\rangle = \langle H_A\rangle + \langle H_B\rangle = \mathrm{U_A}+\mathrm{U_B}$, keeping in mind that the global Hamiltonian can be written as the sum of different individual Hamiltonians only if the subsystems are independent.

-Entropy:

$S_{A\cup B}=-k_B\sum_{a,b}\wp_{a,b}ln\wp_{a,b} = -k_B(\sum_a\wp_a ln\wp_a + \sum_b\wp_b ln\wp_b)$ $=\mathrm{S_A}+\mathrm{S_B}$, also this identity is true only if the two systems *A* and *B* are independent, so that the joint probability $\wp_{a,b}$ is equal to the product $\wp_a\wp_b$ (see Sect. 3.2.).

[11] If the CV is not a linear combination of the Cartesian coordinates of the system, the gradient with respect to ξ does NOT correspond to the force acting on the constraint but there is an additional term containing the Jacobian of the coordinate transformation from the Cartesian set to one that include the CV (see ref. (15)).

[12] With the internal constraint concept in mind, it is easier to understand that constraining configurations affects also momenta. Actually, the expression of the free energy contains also a term which depends on momenta:

$$\frac{F_{NVT\xi}(\hat{\xi})}{-k_BT} = ln\left[Q_{NVT\xi}(\hat{\xi})\right] = ln\left[Z_{NVT\xi}(\hat{\xi})\right] + \\ + ln\left[\int_{\mathbb{R}^{3N}}\prod_{j=1}^{d}\delta\left(\sum_{i=1}^{N}\sum_{\alpha=1}^{3}\left(\frac{\partial\xi_j(\boldsymbol{r})}{\partial r_{i,\alpha}}\frac{p^r_{i,\alpha}}{m_i}\right)\right)\frac{exp(-\beta K(\boldsymbol{p}^r))}{\prod_\eta N_\eta! h^{3N-d}}d\boldsymbol{p}^r\right],$$

where *d* is the dimensionality of the constraint $\boldsymbol{\xi}$ (if it is a scalar then *d* = 1). It is to be pointed out that the last term, the detailed form of which is out of the scope of this work, takes into account how the kinetics of the system is affected by the application of the configurational constraint and depends on the constraint itself via its structure, that is its dimensionality and its gradient with respect to the Cartesian coordinates, but it does not depend on the particular value of $\hat{\xi}$. This useful observation allows us to say that the kinetic term cancels out when free energy differences with respect to a given CV are calculated.

[13] The descriptor of binding inspiring us here is the distance between site and ligand, which is one of the most widely used. Other descriptors can be envisioned, however, so that $\xi(\boldsymbol{r}): D(N,V) \longmapsto [\hat{\xi}_\mathrm{B};\hat{\xi}_\mathrm{A}]$ and that $\xi(\boldsymbol{r}\in A)=\hat{\xi}_A$ and $\xi(\boldsymbol{r}\in B)=\hat{\xi}_\mathrm{B}$, this latter example, however, involves more complex mathematics and is left to more advanced discussions.

[14] The imposition of null global momentum subtracts 3 degrees of freedom to the configurational space D(N,V), whose measure now becomes V^{3N-3}.

[15] In polar coordinates, a vector is represented by means of its modulus ρ and two angles, θ and φ. Here, the modulus coincides with the collective variable that we chose to describe the binding.

Key Words: Statistical mechanics; free energy; activated processes; collective variables; protein-ligand binding.

Send correspondence to: Walter Rocchia, Drug Discovery and Development, Istituto Italiano di Tecnologia, via Morego 30, 16163 Genova, Italy, Tel:+39 01071781552, Fax:39 01071781228, E-mail: walter.rocchia@iit.it