



UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

Scuola di  
Scienze Matematiche,  
Fisiche e Naturali

Corso di Laurea in Chimica  
Curriculum Scienze Chimiche

**Applicazione di trasformazioni alchemiche per il calcolo di energie libere di legame in sistemi host-guest:  $\beta$ -ciclodestrina con composti aromatici**

**Application of alchemical transformations for the calculation of binding free energies in host-guest systems:  $\beta$ -cyclodextrin with aromatic compounds**

**Relatore:**

Riccardo Chelli

**Correlatore:**

Piero Procacci

**Candidato:**

Matteo Cioni

Anno Accademico 2015-16



---

# Contents

---

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Theory</b>	<b>9</b>
2.1	Thermodynamics of the noncovalent binding . . . . .	9
2.2	Alchemical transformations: the double-decoupling method . . . . .	13
2.2.1	Decoupling the ligand from the receptor and solvent: $\Delta G_1^\circ$ calculation	14
2.2.2	Decoupling the ligand from the solvent: $\Delta G_2^\circ$ calculation . . . . .	21
2.3	Using the ligand-receptor distance as binding descriptor in the double-decoupling method . . . . .	23
2.4	Potential energy function and work in alchemical transformations . . . . .	26
<b>3</b>	<b>System, Simulations and Force Field</b>	<b>30</b>
3.1	Simulation technical details and force field . . . . .	30
3.2	System and operative simulation schemes . . . . .	33
<b>4</b>	<b>Results and Calculations</b>	<b>37</b>
4.1	Calculation of $\Delta G_2^\circ$ . . . . .	37
4.2	Calculation of $\Delta G_1^\circ$ . . . . .	42
4.2.1	Determination of $\Delta\phi$ . . . . .	42
4.2.2	Determination of the $\rho(r')$ contribution . . . . .	46
4.3	Calculation of the absolute binding free energy: $\Delta G_{\text{RL}}^\circ$ . . . . .	48

## CONTENTS

---

<b>5 Discussion and Conclusions</b>	<b>51</b>
<b>Bibliography</b>	<b>52</b>
<b>Acknowledgements</b>	<b>58</b>

# CHAPTER 1

---

## Introduction

---

A complete understanding of most chemical and biochemical processes of pharmacological interest, such as solvation and transport properties, protein-ligand binding, proton and electron transfer reactions, etc., requires a careful examination of the free energy difference between the initial and final states of the process under study. This thermodynamic quantity is indeed a fundamental measure of the relative stability of different states of a given system. In the case of a receptor-ligand complexation process, the absolute binding free energy (ABFE) is of paramount importance to assess the affinity between the two compounds. Strictly speaking, the ABFE can be expressed as the reversible thermodynamic work to separate the ligand from the receptor into solution.

In the last two decades, in the context of atomistic molecular dynamics (MD) simulations with explicit solvent, various computational techniques have been devised to compute the ABFE with unprecedented accuracy, such as the double decoupling method [1], potential of mean force [2], metadynamics [3–5] or generalized ensemble approaches like the binding energy distribution analysis [6], the adaptive integration method [7] or the energy driven undocking scheme [8]. All these methodologies bypass the sampling limitations that are inherent to classical MD simulations in ligand-receptor systems by appropriately modifying the interaction potential and/or by invoking geometrical restraints so as to force the binding/unbinding event in a simulation timescale typically in the order

of nanoseconds [9, 10].

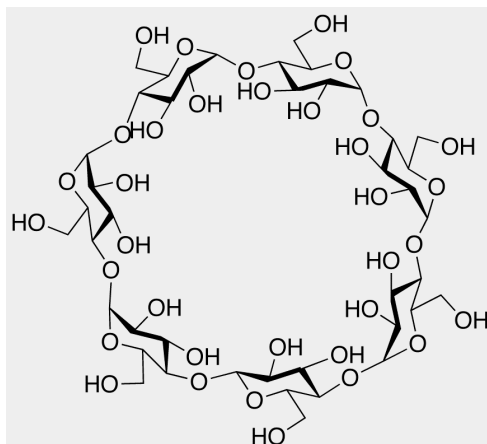
In the so-called alchemical transformations [1, 10–17], probably the most popular and widely used [12–17] of these methods, the ligand, in two distinct thermodynamic processes, is reversibly decoupled from the environment in the bulk solvent and in the binding site of the solvated receptor. Reversible decoupling is implemented by discretizing the non-physical alchemical path in a series of independent equilibrium simulations each with a different Hamiltonian  $H(\lambda_i)$  with the ligand-environment coupling  $\lambda_i$  parameter varying in small steps from  $\lambda = 1$  to  $\lambda = 0$  corresponding to the fully coupled and decoupled (gas-phase) state of the ligand, respectively. In most of the variants of the reversible alchemical route, a geometrical restraint, whose spurious contribution to the ABFE may be eliminated a posteriori, keeps the ligand in the binding site at intermediate values of the  $\lambda$  coupling parameter. The overall free energies for the two decoupling processes are computed by summing up the free energy differences relative to  $\lambda$ -neighboring Hamiltonians using either thermodynamic integration [18] or the free energy perturbation [19] scheme with the Bennett acceptance ratio [20]. The ABFE can be finally computed as the free energy difference between the two decoupling processes [11] using a correction [1, 21, 22] to account for the reversible work needed to bring the ligand volume from that imposed in the MD simulation to that of the standard state. The alchemical procedure can be merged with generalized ensemble approaches by letting  $\lambda$  hopping between neighboring  $\lambda$  states so as to favour conformational sampling of the ligand [6, 15, 23, 24].

A computer simulation requires a precisely defined model for the material of interest. The adopted molecular model describes the forces acting between the molecules, and defines the boundary conditions to be used. Actually, such a model describes interactions between molecules making up the system and their environment, as well as intra- and intermolecular interactions. A molecular potential energy function implicitly describes the geometric shapes of individual molecules or, more precisely, their electron clouds. Thus, when we specify the potential function, we establish the symmetry of molecules, whether they are rigid or flexible, how many interaction sites each molecule contains, where these sites are located, and so on. A detailed characterization of intermolecular potential functions can be given analytically or numerically. In any case, a quantitative form for the potential function defines the molecular model and hence it must be chosen before performing the simulation. Detailed information on MD simulation techniques can be found in the textbook [25].

In the present study, we apply MD simulation methods to compute ABFEs of sev-

eral ligand-receptor complexes involving  $\beta$ -cyclodextrin as the receptor and simple aromatic compounds (benzene, naphthalene and anthracene) as the ligands. The aim is twice. From one side, we investigate the capabilities of simulation approaches based on alchemical transformations and nonequilibrium work theorems to quantitatively predict the ABFEs, and hence the complexation equilibrium constants, of an important biochemical target involved in several biological processes. On the other side, we aim to test the simulation molecular model in general, and the GLYCAM [26] force field combined to our parametrization of the atomic charges in particular, for being employed in forthcoming computational studies addressed to ligands with pharmacological perspectives.

From the biomedical standpoint, complexes of cyclodextrins (CDs) have attracted a large interest, especially for their pharmaceutical applications. Since inclusion compounds of CDs with hydrophobic molecules are able to get into body tissues, they can be used to release biologically active compounds under specific conditions. The ability of CDs to alter physical, chemical and biological properties of guest molecules through the formation of inclusion complexes in solution is the basis for their use as pharmaceutical carriers of drugs, which are unstable at ambient conditions or poorly soluble into water. Indeed, a drug substance should have a certain level of solubility to be readily delivered to the cellular membrane, but it needs to be hydrophobic enough to cross the membrane: one of the unique properties of CDs is indeed their ability to enhance drug delivery through biological membranes. However, the inclusion complexes are important not only for solubilization and transport. In some cases, CDs catalyse the reaction of a guest molecule. Four species of CDs are known with rings including from 6 to 9 glucose units:  $\alpha$ -CD (6 units),  $\beta$ -CD (7 units),  $\lambda$ -CD (8 units) and  $\gamma$ -CD (9 units). In particular,  $\beta$ -CD consists of seven d-glucopyranose monomers covalently bound by  $\alpha$ -1,4-linkages. A ball-and-sticks representation of  $\beta$ -CD is shown in fig. 1.1. From a topological point of view, this macrocycle can be described as a truncated cone, in which the narrow rim (6.4 Å) bears the primary hydroxy groups, whereas the wide rim (15.4 Å) bears the secondary hydroxy groups. Since no hydroxy groups are present within the toroidal cavity of  $\beta$ -CD, this zone has a pronounced hydrophobic character. This feature, together with van der Waals forces and hydrogen bonding, allows  $\beta$ -CD to host small molecules efficiently in aqueous solutions. No covalent bonds are formed or broken during the complex formation, and drug molecules in the complex are in rapid equilibrium with free molecules in the solution. The main driving force for the complex formation is the release of enthalpy due to migration of water molecules outside the  $\beta$ -CD cavity. In fact, water molecules are



**Figure 1.1:**  $\beta$ -cyclodextrin

replaced by more hydrophobic guest molecules present into solution to achieve an apolar association and decrease of the  $\beta$ -CD ring strain resulting in a more stable lower energy state. The binding of guest molecules within the host  $\beta$ -CD is not fixed or permanent, but rather a dynamic equilibrium is established. Binding strength depends on how well the host-guest complex fits together and on specific local interactions between surface atoms. Complexes can be formed either in solution or in the crystalline state and water is typically the solvent of choice.

Thanks to their wide use in pharmaceutical applications, there is an enormous amount of studies about dynamical and thermodynamic properties of these kind of complexes with a drug as ligand. Whether we study systems theoretically or experimentally, the general procedure is the same: we manipulate and control certain observables (inputs), and then we measure the system response (outputs). Within this framework, MD simulations provide one of the most direct ways to theoretically investigate molecular behaviors that are not accessible to experimental approaches in complex systems, such as in the case of most host-guest interactions. Given a system composed of atoms interacting with empirical force fields, an MD simulation produces a dynamical trajectory by integrating Newton's equations of motion. Generally, a system of coupled second order non linear differential equations cannot be solved exactly; hence, an algorithm for numerical integration is necessary. Making use of MD simulations, it is possible to determine the system equilibrium distribution, as a function of specific structural parameters (interatomic distances, bending and dihedral angles, etc.). Considering the specific situation of a  $\beta$ -CD ligand complexation process, bounded and unbounded states can be defined in terms of a proper



coordinate. The evolution of an abstract coordinate may indeed represent the progress along the pathway of the reaction of interest. Under such assumptions, the Potential of Mean Force (PMF) represents a fundamental quantity in the calculation of ABFEs. The PMF is basically the free energy landscape in the space of selected collective coordinates. For example, if one defines the distance between two atoms as the collective coordinate, the PMF is monodimensional and corresponds to the free energy as a function of the separation between the atoms. In the present study, we choose the distance between the centers of mass of  $\beta$ -CD and ligand as the collective coordinate in which the PMF is represented. Because of the peculiar symmetry of the problem, such a distance appears as the most natural coordinate for representing the binding in  $\beta$ -CD complexes. By estimating the PMF profile, it is possible to evaluate the difference of energy between bounded and unbounded configurations. Unlikely, it is exactly this calculation, requiring an accurate exploration of the whole configurational space, that results prohibitive through a direct equilibrium sampling. This kind of numerical problem ultimately arises from the finite power of computers, which limits the simulation time available to get a complete sampling of the system.

A way to tackle this sampling problem is to resort to indirect strategies, based on a decomposition of the ABFE into a series of free energy contributions, related to each other through a thermodynamic cycle. In this respect, many physical processes, such as ligand binding in complex biological systems or transfer of a molecule from one phase to another, can be equivalently expressed as a sequence of intermediate transformations, even of “imaginary” physical meaning, associated with free energy changes which can be determined in an easier manner. The described strategy is the basis of the alchemical approach designed according to a thermodynamic cycle within the framework of the so-called Double Decoupling Method (DDM). DDM relies on “decoupling” the ligand from its surrounding, or, equivalently, switching off all the interactions of the ligand with the receptor and the solvent. So, in the first path of the cycle, the free energy change is associated with the turn-off of the intermolecular interactions involving the ligand located in the binding site. Intermolecular terms of the potential energy are progressively annihilated, nullifying the interactions of the ligand with receptor and solvent molecules. The intramolecular potential, instead, is unchanged. To complete the cycle, two other paths are necessary. The first consists into bringing the decoupled ligand from a configuration in the solvent to the binding site of the receptor. As it will be shown below, the associated free energy accounts for a geometrical change of the volume going from the solvent to

the binding site, which can be evaluated without performing simulations. The last path of the cycle is the activation of the ligand in the solvent, obtained by reintroducing the ligand-solvent intermolecular potential. Actually, in this work the opposite process, *i.e.* the ligand-solvent decoupling, will be simulated, whose free energy change differs only for the sign. For each alchemical transformation, the free energy change is computed by applying the Jarzynski equality, an almost recent relationship that relates free energy differences between two states to the work distribution along an ensemble of trajectories joining these states. Combining the free energy changes associated with these three paths allows to recover the standard ABFE of the complex. The described procedure is implemented on complexes that, owing to their relative simplicity, can be considered as benchmark compounds for binding of aromatic species, even if their practical impact in drug design appears to be modest. In particular, we have investigated complexes of  $\beta$ -CD with three ligands of different size, namely benzene, naphthalene and anthracene. This choice has been made taking into account the similar geometrical properties of these aromatic molecules, considering that they only differ for one phenyl ring. Since ligands show the same electronic structure and other features given by their aromatic nature, it will be of some interest to analyse the role of the ligand dimensions in the binding strength.

The computational procedure for anthracene and naphthalene is identical, whereas for benzene we have adopted a specific technique (called Umbrella Sampling) aimed at enforcing the sampling of the bounded state of the complex. Indeed, during the simulation of the benzene: $\beta$ -CD complex, we have observed that, after a certain time, the ligand leaves the binding site, resulting in a failure of sampling. For such a reason, we have introduced an additional external potential to restrain the benzene inside the binding cavity of  $\beta$ -CD, eventually ensuring an accurate sampling of the configurational space of the bounded state. Each simulation has been realized using the MD simulation program ORAC [27, 28].

In order to assess the validity of the overall computational approach, we have compared the results with experimental data reported in the literature. As a matter of fact, the agreement with experimental data is mandatory to validate both method and force field. Furthermore, it is worth noting that combining nonequilibrium alchemical transformations with a constraining scheme for the ligand in the binding site represents an original contribution in the examination of ligand-receptor systems of relevant biochemical interest.

## CHAPTER 2

---

### Theory

---

In the present chapter a general description of statistical mechanics principles is reported together with computational methods and procedures employed in this study. First, a brief introduction to classical and statistical thermodynamics is reported focusing on the milestone concepts of free energy and partition function. Then, we will enter in the specific detail of statistical mechanics theory of non-covalent binding and alchemical transformations. Other adopted methodologies, such as umbrella sampling [29] and nonequilibrium work theorems, will be discussed there where they are introduced. A rather general overview of molecular dynamics simulations will also be presented.

### 2.1 Thermodynamics of the noncovalent binding

The theory of noncovalent binding association and the basic relationships for the calculation of the ABFE through alchemical transformations has been reported by Gilson and coworkers in Ref. [1]. Here, we review the Gilson's results, preserving as much as possible his notation. During the discussion, we will outline some difference with respect to the Gilson's outcomes, especially arising from the possible geometries of the ligand related to its specific symmetry (spherical, linear, etc.). In Sec. 2.2.1, we go beyond the Gilson's treatment presenting two approaches to the alchemical transformations based on

constrained nonequilibrium simulations. In Sec. 2.3, we describe how the relationships for alchemical transformations change upon using the ligand-receptor distance as binding descriptor in the double-decoupling method.

The reaction we are interested to is the association of a ligand L with a receptor R to form a noncovalent complex RL,



At equilibrium, the chemical potentials of L, R, and RL into solution satisfy the following condition

$$\mu_{\text{sol},R} + \mu_{\text{sol},L} = \mu_{\text{sol},RL}. \quad (2.2)$$

The chemical potential of a species  $i$  at a given concentration  $C_i$  can be expressed as

$$\mu_{\text{sol},i} = \mu_{\text{sol},i}^{\circ} + RT \ln \frac{\gamma_i C_i}{C^{\circ}}, \quad (2.3)$$

where  $\mu_{\text{sol},i}^{\circ}$  is the standard chemical potential,  $\gamma_i$  is the activity coefficient,  $C^{\circ}$  is the standard concentration in the same units as  $C_i$  (1 M or 1 molecule/1661 Å<sup>3</sup>),  $R$  is the gas constant and  $T$  is the absolute temperature. As Gilson noted,  $\mu_{\text{sol},i}^{\circ}$  is the chemical potential in a hypothetical standard state in which each species is at standard concentration in the solvent, but does not interact with other solute molecules. It is worth noting that in the infinite dilution limit the activity coefficients of the solute species approach unity [30,31]. Recasting Eqs. 2.2 and 2.3, we obtain the relation between the standard free energy of binding and the binding constant  $K_{\text{RL}}$

$$\Delta G_{\text{RL}}^{\circ} \equiv \mu_{\text{sol},\text{RL}}^{\circ} - \mu_{\text{sol},R}^{\circ} - \mu_{\text{sol},L}^{\circ} = -RT \ln \left( \frac{\gamma_{\text{RL}}}{\gamma_R \gamma_L} \frac{C^{\circ} C_{\text{RL}}}{C_R C_L} \right) \equiv -RT \ln K_{\text{RL}}. \quad (2.4)$$

A fundamental relationship to link the ABFE,  $\Delta G_{\text{RL}}^{\circ}$ , and hence the equilibrium constant  $K_{\text{RL}}$ , to statistical thermodynamical quantities affordable computationally has been derived by Hill in Ref. [32] and modified by Gilson [1] to include explicitly the standard concentration,

$$\mu_{\text{sol},R}^{\circ} = -RT \ln \left( \frac{1}{V_{N,R} C^{\circ}} \frac{Q_{N,R}(V_{N,R})}{Q_{N,0}(V_{N,0})} \right) + P^{\circ} \bar{V}_R, \quad (2.5)$$

with analogous expressions for the ligand L and the complex RL. In the previous equation,  $Q_{N,R}(V_{N,R})$  is the canonical partition function for a solution consisting of  $N$  solvent molecules and one molecule R at volume  $V_{N,R}$ , which is the volume of this solution when it is at equilibrium at the standard pressure  $P^{\circ}$ . Analogously,  $Q_{N,0}(V_{N,0})$  is the canonical

partition function of  $N$  solvent molecules alone at the volume  $V_{N,0}$ , namely the equilibrium volume of the pure-solvent sample at standard pressure. Finally,  $\bar{V}_R = V_{N,R} - V_{N,0}$  is the volume change occurring at standard pressure when one molecule R is added to  $N$  molecules of solvent. It is worth noting that the term  $P^\circ \bar{V}_R$  into Eq. 2.5 is typically very small [33], because  $\bar{V}_R$  corresponds to about the molecular volume of R.

We now provide a more detailed expression of the standard chemical potentials  $\mu_{\text{sol,R}}^\circ$  and  $\mu_{\text{sol,L}}^\circ$ , by exploiting the representation of the canonical partition functions in terms of the classical statistical thermodynamics [34,35]. In this framework, the partition function  $Q_{N,R}(V_{N,R})$  can be written as a phase-space integral separable as the product of an integral over the positional variables, *i.e.* the atomic coordinates, and two integrals over the dynamical variables, *i.e.* the conjugate momenta related to the solute and solvent atoms:

$$Q_{N,R}(V_{N,R}) = \frac{1}{\sigma_{\text{sol,R}} \sigma_{\text{S}}^N} \int e^{-\beta U(\mathbf{r}'_R, \mathbf{r}_S)} d\mathbf{r}'_R d\mathbf{r}_S \int \exp\left(-\beta \sum_{i=1}^{M_R} \frac{p_i^2}{2m_i}\right) d\mathbf{p}_R \int \exp\left(-\beta \sum_{i=M_R+1}^{M_R+M_S} \frac{p_i^2}{2m_i}\right) d\mathbf{p}_S, \quad (2.6)$$

where  $\beta = (RT)^{-1}$ , the symbols  $\mathbf{r}'_R$ ,  $\mathbf{p}_R$  and  $M_R$  denote the atomic coordinates, the conjugate momenta and the number of atoms of the R molecule, respectively, and  $\mathbf{r}_S$ ,  $\mathbf{p}_S$  and  $M_S$  are the analogous quantities for the  $N$  solvent molecules. We note that, at variance with the integral over the conjugate momenta, the integral over  $\mathbf{r}'_R$  and  $\mathbf{r}_S$  cannot be split, because the coordinates of solute and solvent are intimately connected through mixed terms in the energy  $U(\mathbf{r}'_R, \mathbf{r}_S)$  of the system. In Eq. 2.6,  $\sigma_{\text{sol,R}}$  and  $\sigma_S$  are the symmetry numbers of R into solution and of the solvent molecule. To specify that the symmetry number of R is related to the solution environment is necessary because we will show that analogous factors will be introduced for the gas phase and the complex RL. It is worth considering that the prefactor arising from the quantum-mechanical correction is not included in the expression of  $Q_{N,R}(V_{N,R})$ , because it does not contribute to the calculation of the ABFE. Following Gilson, we introduce a molecular axis system to separate the lab-frame coordinates  $\mathbf{r}'_R$  of R into internal and external, even if other definitions are also possible. This molecular axis system is built taking as reference three atoms of R. Atom 1 becomes the origin of the molecular coordinates, denoted as  $\mathbf{R}_R$ . The vector joining atom 1 with atom 2 defines the  $x$ -axis. The direction of the  $y$ -axis is given by the direction of the vector joining atoms 2 and 3, minus the  $x$ -component of this vector. The  $z$ -axis is constructed as the cross-product of the  $x$ - and  $y$ -axes. The six coordinates thus fixed,

namely  $\mathbf{R}_R$  plus the Eulerian angles  $\xi_{R,1}$ ,  $\xi_{R,2}$  and  $\xi_{R,3}$  that specify the orientation of the molecular frame relative to the lab frame, correspond to the external coordinates of R. The set of  $3M_R - 6$  internal coordinates of R will be indicated with  $\mathbf{r}_R$ . Noting that the integrals over  $\mathbf{r}_R$  and over  $\mathbf{r}_S$  do not depend upon the position and orientation of R, the integrals over  $\mathbf{R}_R$ ,  $\xi_{R,1}$ ,  $\xi_{R,2}$  and  $\xi_{R,3}$  can be done at once. Considering that R is typically a nonlinear molecule, the integrals yield  $8\pi^2 V_{N,R}$ . Moreover, considering that the integral over the momenta components of an atom  $i$  yields a factor  $(2\pi m_i RT)^{3/2}$ , the partition function of Eq. 2.6 can be written as

$$Q_{N,R}(V_{N,R}) = \frac{8\pi^2 V_{N,R} Z_{N,R}}{\sigma_{\text{sol},R} \sigma_S^N} \prod_{i=1}^{M_R+M_S} (2\pi m_i RT)^{3/2}, \quad (2.7)$$

where

$$Z_{N,R} = \int e^{-\beta U(\mathbf{r}_R, \mathbf{r}_S)} d\mathbf{r}_R d\mathbf{r}_S \quad (2.8)$$

is the configuration integral for a system consisting of one R molecule into  $N$  solvent molecules. In a similar way, we may express the partition function of  $N$  solvent molecules as

$$\begin{aligned} Q_{N,0}(V_{N,0}) &= \frac{1}{\sigma_S^N} \int e^{-\beta U(\mathbf{r}_S)} d\mathbf{r}_S \int \exp\left(-\beta \sum_{i=M_R+1}^{M_R+M_S} \frac{p_i^2}{2m_i}\right) d\mathbf{p}_S \\ &= \frac{Z_{N,0}}{\sigma_S^N} \prod_{i=M_R+1}^{M_R+M_S} (2\pi m_i RT)^{3/2}, \end{aligned} \quad (2.9)$$

where  $Z_{N,0}$  is the configuration integral for the solvent sample

$$Z_{N,0} = \int e^{-\beta U(\mathbf{r}_S)} d\mathbf{r}_S. \quad (2.10)$$

Substituting Eqs. 2.7 and 2.9 into Eq. 2.5, we obtain

$$\mu_{\text{sol},R}^\circ = -RT \ln \left( \frac{8\pi^2}{C^\circ \sigma_{\text{sol},R}} \prod_{i=1}^{M_R} (2\pi m_i RT)^{\frac{3}{2}} \frac{Z_{N,R}}{Z_{N,0}} \right) + P^\circ \bar{V}_R. \quad (2.11)$$

Similar arguments lead to a relationship for  $\mu_{\text{sol},L}^\circ$ . However considering that the ligand can be also linear in shape and even a single atom, integration over the orientational degrees of freedom can give  $8\pi^2$ ,  $4\pi$  and 1, respectively (from now on, this geometry factor will be denoted as  $\mathcal{V}_{\xi_L}$ ). Therefore, the generic expression for  $\mu_{\text{sol},L}^\circ$  is

$$\mu_{\text{sol},L}^\circ = -RT \ln \left( \frac{\mathcal{V}_{\xi_L}}{C^\circ \sigma_{\text{sol},L}} \prod_{i=1}^{M_L} (2\pi m_i RT)^{\frac{3}{2}} \frac{Z_{N,L}}{Z_{N,0}} \right) + P^\circ \bar{V}_L. \quad (2.12)$$

Here, the product is extended to the  $M_L$  components of the conjugate momenta of L.

The calculation of the standard chemical potential of the complex  $\mu_{\text{sol,RL}}^\circ$  requires a specific treatment of the external and internal coordinates of RL. The former are assumed to be the external coordinates of R, while the external coordinates of L, indicated as  $\zeta_L \equiv (\mathbf{R}_L, \xi_{L,1}, \xi_{L,2}, \xi_{L,3})$ , are taken to be defined relative to R, so that they become internal coordinates of the complex. The arguments adopted to determine  $\mu_{\text{sol,R}}^\circ$  and  $\mu_{\text{sol,L}}^\circ$  may also be used here with the difference that the configuration integral of the complex must be restricted to the configurations for which R and L are complexed [35, 36]. This can be realized introducing a step function  $I(\zeta_L)$  that holds 1 for complexed configurations and 0 otherwise. We can thus obtain the following expression:

$$\mu_{\text{sol,RL}}^\circ = -RT \ln \left( \frac{8\pi^2}{C^\circ \sigma_{\text{cp,L}} \sigma_{\text{cp,R}}} \prod_{i=1}^{M_L+M_R} (2\pi m_i RT)^{\frac{3}{2}} \frac{Z_{N,\text{RL}}}{Z_{N,0}} \right) + P^\circ \bar{V}_{\text{RL}}, \quad (2.13)$$

where the product is extended to the  $M_L + M_R$  components of the conjugate momenta of L and R and  $\sigma_{\text{cp,L}}$  and  $\sigma_{\text{cp,R}}$  are the symmetry numbers associated with L and R when the complex is formed (we notice that Gilson indicated the symmetry number of the complex as  $\sigma_{\text{AB}}$ ). These numbers may indeed differ from  $\sigma_{\text{sol,L}}$  and  $\sigma_{\text{sol,R}}$ , as long the molecular symmetry is not preserved upon complexation. In Eq. 2.13, the configuration integral of RL into solution is

$$Z_{N,\text{RL}} = \int I(\zeta_L) J_{\zeta_L} e^{-\beta U(\zeta_L, \mathbf{r}_L, \mathbf{r}_R, \mathbf{r}_S)} d\zeta_L d\mathbf{r}_L d\mathbf{r}_R d\mathbf{r}_S, \quad (2.14)$$

where  $J_{\zeta_L}$  is the absolute value of the Jacobian determinant for the rotation and translation of L relative to R. We remark that, for purposes of generality, we keep the full dependence of  $J_{\zeta_L}$  on the translational and rotational (external) coordinates of L. Instead, Gilson and coworkers take a Jacobian determinant dependent on the only rotation of L, implicitly assuming that the position of L is relative to a Cartesian reference frame on R.

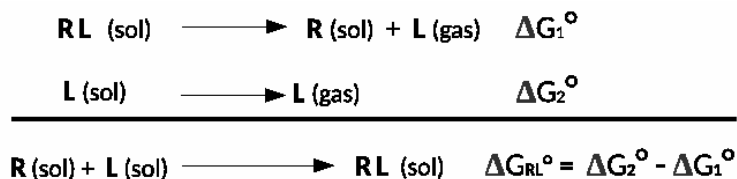
Recasting Eqs. 2.11, 2.12 and 2.13 into Eq. 2.4, we get the expression for the ABFE

$$\Delta G_{\text{RL}}^\circ = -RT \ln \left( \frac{C^\circ}{\mathcal{V}_{\zeta_L}} \frac{\sigma_{\text{sol,L}} \sigma_{\text{sol,R}}}{\sigma_{\text{cp,L}} \sigma_{\text{cp,R}}} \frac{Z_{N,\text{RL}} Z_{N,0}}{Z_{N,\text{R}} Z_{N,\text{L}}} \right) + P^\circ (\bar{V}_{\text{RL}} - \bar{V}_{\text{R}} - \bar{V}_{\text{L}}). \quad (2.15)$$

## 2.2 Alchemical transformations: the double-decoupling method

The DDM is a route to the estimate of  $\Delta G_{\text{RL}}^\circ$  and is based on the calculation of the free energy differences between two independent processes, represented in the scheme of

Fig. 2.1: the annihilation of L from the solvated complex RL and the annihilation of L from the solvent. In the former case, annihilation is accomplished by turning off the interactions of L with solvent and receptor R in a solution of RL, while in the latter case annihilation is performed by turning off the interactions of L with the solvent in a solution of L. It is worth noting that, in both situations, we do not deal with a total annihilation of L.



**Figure 2.1:** Thermodynamic analysis of double-decoupling method

L, but rather with a decoupling of L from its environment. Its intramolecular interactions are left in place and hence it is virtually “transformed” in a ideal-gas molecule. Before discussing the annihilation processes and in particular the details of our approach, it is convenient to see how  $\Delta G_{\text{RL}}^\circ$  is correlated to the free energies  $\Delta G_1^\circ$  and  $\Delta G_2^\circ$  of the processes represented in Fig. 2.1 and how these quantities can be expressed in terms of configuration integrals. According to Gilson [1],  $\Delta G_1^\circ$  and  $\Delta G_2^\circ$  can be written as

$$\Delta G_1^\circ = \mu_{\text{sol,R}}^\circ + \mu_{\text{gas,L}}^\circ - \mu_{\text{sol,RL}}^\circ, \quad (2.16)$$

$$\Delta G_2^\circ = \mu_{\text{gas,L}}^\circ - \mu_{\text{sol,L}}^\circ, \quad (2.17)$$

where  $\mu_{\text{gas,L}}^\circ$  is the standard chemical potential of L in the ideal gas phase and the other standard chemical potentials are defined in Eqs. 2.11, 2.12 and 2.13. Considering Eq. 2.4 together with Eqs. 2.16 and 2.17, it is immediate to show that

$$\Delta G_{\text{RL}}^\circ = \Delta G_2^\circ - \Delta G_1^\circ. \quad (2.18)$$

### 2.2.1 Decoupling the ligand from the receptor and solvent: $\Delta G_1^\circ$ calculation

The standard chemical potential of L in the ideal gas phase,  $\mu_{\text{gas,L}}^\circ$ , is related to the natural logarithm of the molecular partition function as

$$\mu_{\text{gas,L}}^\circ = -RT \ln Q_{0,\text{L}}(V^\circ), \quad (2.19)$$



where it is explicitly reported that the partition function must be evaluated in the phase space limited to the standard volume  $V^\circ = 1/C^\circ$ . Following the arguments leading to Eq. 2.12, we get

$$\mu_{\text{gas,L}}^\circ = -RT \ln \left( \frac{\mathcal{V}_{\xi_L}}{C^\circ \sigma_{\text{gas,L}}} \prod_{i=1}^{M_L} (2\pi m_i RT)^{\frac{3}{2}} Z_{0,L} \right). \quad (2.20)$$

In the previous equation,  $\mathcal{V}_{\xi_L}$  is from the integral over the orientation of L ( $\mathcal{V}_{\xi_L} = 8\pi^2, 4\pi, 1$  for non linear, linear and atomic ligands, respectively),  $\sigma_{\text{gas,L}}$  is the symmetry number of L in the ideal gas phase and  $Z_{0,L}$  is the configuration integral in the internal coordinates:

$$Z_{0,L} = \int e^{-\beta U(\mathbf{r}_L)} d\mathbf{r}_L. \quad (2.21)$$

The external coordinates of L are integrated in Eq. 2.20, giving the contribution  $\mathcal{V}_{\xi_L}/(C^\circ \sigma_{\text{gas,L}})$ . Substituting Eqs. 2.11, 2.13 and 2.20 into Eq. 2.16, we obtain

$$\Delta G_1^\circ = -RT \ln \left( \frac{\mathcal{V}_{\xi_L}}{C^\circ} \frac{\sigma_{\text{cp,L}} \sigma_{\text{cp,R}}}{\sigma_{\text{gas,L}} \sigma_{\text{sol,R}}} \frac{Z_{N,R} Z_{0,L}}{Z_{N,RL}} \right) + P^\circ (\bar{V}_R - \bar{V}_{RL}). \quad (2.22)$$

In the DDM, an artificial energy function  $U(\lambda, \zeta_L, \mathbf{r}_L, \mathbf{r}_R, \mathbf{r}_S)$  dependent on a control parameter  $\lambda \in [0, 1]$  is introduced, whose functional form is rather arbitrary. The only requirements are that for  $\lambda = 0$  and  $\lambda = 1$  the artificial energy function must correspond to the energy functions of the coupled and uncoupled states of the ligand in the complex, respectively:

$$U(0, \zeta_L, \mathbf{r}_L, \mathbf{r}_R, \mathbf{r}_S) = U(\zeta_L, \mathbf{r}_L, \mathbf{r}_R, \mathbf{r}_S), \quad (2.23)$$

$$U(1, \zeta_L, \mathbf{r}_L, \mathbf{r}_R, \mathbf{r}_S) = U(\mathbf{r}_R, \mathbf{r}_S) + U(\mathbf{r}_L). \quad (2.24)$$

Exploiting the artificial energy function, a free energy function dependent parametrically on  $\lambda$  can be built as

$$g(\lambda) = -RT \ln \int I(\zeta_L) J_{\zeta_L} e^{-\beta U(\lambda, \zeta_L, \mathbf{r}_L, \mathbf{r}_R, \mathbf{r}_S)} d\zeta_L d\mathbf{r}_L d\mathbf{r}_R d\mathbf{r}_S. \quad (2.25)$$

According to  $g(\lambda)$  and to the requirements of Eqs. 2.23 and 2.24, the free energy difference between the final and initial states is

$$\begin{aligned} g(1) - g(0) &= -RT \ln \frac{\int I(\zeta_L) J_{\zeta_L} e^{-\beta U(\mathbf{r}_R, \mathbf{r}_S)} e^{-\beta U(\mathbf{r}_L)} d\zeta_L d\mathbf{r}_L d\mathbf{r}_R d\mathbf{r}_S}{\int I(\zeta_L) J_{\zeta_L} e^{-\beta U(\zeta_L, \mathbf{r}_L, \mathbf{r}_R, \mathbf{r}_S)} d\zeta_L d\mathbf{r}_L d\mathbf{r}_R d\mathbf{r}_S} \\ &= -RT \ln \frac{V_I V_{\xi_L} Z_{N,R} Z_{0,L}}{Z_{N,RL}}. \end{aligned} \quad (2.26)$$

This equation results from assuming  $\int I(\zeta_L) J_{\zeta_L} d\zeta_L = V_I V_{\xi_L}$ . The  $V_I$  term arises from the integral over the position, while the  $V_{\xi_L}$  term arises from the integral over the orientation. These two quantities represent, together, the binding site volume. This implies that both translational and rotational degrees of freedom of L relative to R (external coordinates of L) within the binding site contribute to the ABFE, and Eq. 2.26 quantifies such a contribution. The definitions of the configuration integrals, Eqs. 2.8, 2.14 and 2.21, have also been used in deriving the second line of Eq. 2.26. In the method of thermodynamic integration, the quantity  $g(1) - g(0)$  is directly evaluated via equilibrium simulations, so that one must also determine  $V_I$  and  $V_{\xi_L}$  to gain an estimate of  $\Delta G_1^\circ$  (see Eq. 28 of Ref. [1]); note that  $V_{\xi_L}$  corresponds to  $\xi_I$  of the Gilson’s article). The calculation of  $V_I$  and  $V_{\xi_L}$  may not be straightforward and requires a separate simulation.

In this study, we propose a change of paradigm for the ligand-receptor binding, adopting a criterion based on the only position of L relative to R, say  $\mathbf{R}_L$ . From a mathematical standpoint, this corresponds to turn from a binding function expressed in terms of position and orientation of L, *i.e.*  $I(\zeta_L) \equiv I(\mathbf{R}_L, \xi_{L,1}, \xi_{L,2}, \xi_{L,3})$ , to a binding function expressed in terms of the position of L alone, *i.e.*  $I(\mathbf{R}_L)$ . This assumption is consistent with the common idea that binding occurs basically when ligand and receptor approach to each other, regardless the mutual orientation defined here by the variables  $\xi_{L,1}$ ,  $\xi_{L,2}$  and  $\xi_{L,3}$ . Of course, for a generic position  $\mathbf{R}_L$  satisfying the binding condition  $I(\mathbf{R}_L) = 1$ , most orientations of L relative to R will have a negligible probability of being observed in the coupled state, because of strong atomic overlaps between R and L. As a consequence, these configurations will contribute negligibly to the denominator of Eq. 2.26. In the numerator of Eq. 2.26 (uncoupled state), the integral over the orientations can instead be done at once, giving a factor  $\mathcal{V}_{\xi_L}$ . This allows us to rewrite the free energy function of Eq. 2.25 as follows

$$g(\lambda) = -RT \ln \int I(\mathbf{R}_L) J_{\mathbf{R}_L} J_{\xi_L} e^{-\beta U(\lambda, \mathbf{R}_L, \xi_L, \mathbf{r}_L, \mathbf{r}_R, \mathbf{r}_S)} d\mathbf{R}_L d\xi_L d\mathbf{r}_L d\mathbf{r}_R d\mathbf{r}_S, \quad (2.27)$$

where  $\xi_L$  is a shorthand for  $(\xi_{L,1}, \xi_{L,2}, \xi_{L,3})$ ,  $J_{\xi_L}$  and  $J_{\mathbf{R}_L}$  are the Jacobian determinants for the (external) rotational and translational coordinates of L, respectively, and  $d\xi_L \equiv d\xi_{L,1} d\xi_{L,2} d\xi_{L,3}$ . As noted below Eq. 2.14, the Jacobian determinant  $J_{\mathbf{R}_L}$  is in general different from 1, being 1 only when  $\mathbf{R}_L$  is expressed in a Cartesian reference system. The

free energy difference  $g(1) - g(0)$  of Eq. 2.26 is then restated straightforwardly,

$$\begin{aligned}
 g(1) - g(0) &= -RT \ln \frac{\int I(\mathbf{R}_L) J_{\mathbf{R}_L} J_{\xi_L} e^{-\beta U(\mathbf{r}_R, \mathbf{r}_S)} e^{-\beta U(\mathbf{r}_L)} d\mathbf{R}_L d\xi_L d\mathbf{r}_L d\mathbf{r}_R d\mathbf{r}_S}{\int I(\mathbf{R}_L) J_{\mathbf{R}_L} J_{\xi_L} e^{-\beta U(\mathbf{R}_L, \xi_L, \mathbf{r}_L, \mathbf{r}_R, \mathbf{r}_S)} d\mathbf{R}_L d\xi_L d\mathbf{r}_L d\mathbf{r}_R d\mathbf{r}_S} \\
 &= -RT \ln \frac{V_I \mathcal{V}_{\xi_L} Z_{N,R} Z_{0,L}}{Z_{N,RL}}. \tag{2.28}
 \end{aligned}$$

In the second line of Eq. 2.28, we have carried out the integrals over  $\mathbf{R}_L$  and  $\xi_L$  in the numerator, obtaining  $V_I$  and  $\mathcal{V}_{\xi_L}$ . We point out that, owing to the adopted binding criterion, the integral over  $\xi_L$  is analytical, while numerical into Gilson's expression Eq. 2.26 ( $\mathcal{V}_{\xi_L}$  vs.  $V_{\xi_L}$ ).

Since the paradigm of binding is based on  $\mathbf{R}_L$ , it is convenient to introduce a potential of mean force as a function of  $\lambda$  including the internal coordinates of R and L, the coordinates of the solvent and the orientational coordinates of L. This potential thus results to be a function of both  $\lambda$  and  $\mathbf{R}_L$ :

$$e^{-\beta\phi(\lambda, \mathbf{R}_L)} = \int J_{\xi_L} e^{-\beta U(\lambda, \mathbf{R}_L, \xi_L, \mathbf{r}_L, \mathbf{r}_R, \mathbf{r}_S)} d\xi_L d\mathbf{r}_L d\mathbf{r}_R d\mathbf{r}_S. \tag{2.29}$$

According to the above definition of potential of mean force, the free energy function  $g(\lambda)$  (Eq. 2.27) takes the following simplified form

$$g(\lambda) = -RT \ln \int I(\mathbf{R}_L) J_{\mathbf{R}_L} e^{-\beta\phi(\lambda, \mathbf{R}_L)} d\mathbf{R}_L. \tag{2.30}$$

Using the definition 2.30 of  $g(\lambda)$  into Eq. 2.28, we obtain

$$g(1) - g(0) = -RT \ln \frac{\int I(\mathbf{R}_L) J_{\mathbf{R}_L} e^{-\beta\phi(1, \mathbf{R}_L)} d\mathbf{R}_L}{\int I(\mathbf{R}_L) J_{\mathbf{R}_L} e^{-\beta\phi(0, \mathbf{R}_L)} d\mathbf{R}_L}. \tag{2.31}$$

Here, we report on two alternative algorithms to compute  $\Delta G_1^\circ$ . One is analogous to the method by Gilson and co-workers, being based on the direct estimate of  $g(1) - g(0)$ . In particular, we compute  $g(1) - g(0)$  from nonequilibrium molecular dynamics simulations instead of equilibrium simulations via thermodynamic integration. First, we sample the binding configurations for the various positions  $\mathbf{R}_L$  using an equilibrium simulation. The correct sampling weight  $e^{-\beta\phi(0, \mathbf{R}_L)}$  of the initial microstates (coupled RL complex) is thus guaranteed implicitly by the simulation. This provides an amount of isothermally and isobarically sampled microstates, say  $N_{\text{traj}}$ , to be taken as initial phase-space points for the nonequilibrium alchemical trajectories. Equation 2.31 establishes that L be in the same binding site in both the initial and final states. This can be accomplished by

creating a bijective mapping between these states, with the aim of preventing the ligand from leaving the binding site. Recently, some of us developed a nonequilibrium approach able to guarantee such a mapping [37], allowing the estimate of free energy differences between two configurational domains by means of steered molecular dynamics associated with nonequilibrium work theorems. The method is based on the creation of a phase-space mapping applied during the nonequilibrium trajectories, whether to the control parameter employed to switch the system from the initial to the final state or to some phase-space variable (not directly correlated to the control parameter) taken to define the two configurational domains. The latter is just the situation that we may apply to the alchemical transformations. While the  $\lambda$  control parameter is evolved in time from 0 (coupled ligand) to 1 (uncoupled ligand) according to some established time schedule, the coordinate  $\mathbf{R}_L$  of the ligand relative to the receptor is mapped to bring the system from a coupled to an uncoupled configuration within the binding site. This is accomplished by fixing the  $\mathbf{R}_L$  coordinate to the initial value (obtained from the equilibrium sampling) during the alchemical transformation, thus preventing the ligand from leaving the binding site. A constraint to  $\mathbf{R}_L$  can be applied whether using some constraining method, such as RATTLE [38] or SHAKE [39], or more simply by enforcing stiff (harmonic) potentials to the three components of  $\mathbf{R}_L$ . The latter is the algorithm employed in this work. Using this simulation scheme, we thus produce  $N_{\text{traj}}$  alchemical trajectories that allow to compute the free energy difference  $g(1) - g(0)$  by using the Jarzynski equality [40]:

$$g(1) - g(0) = -RT \ln \langle e^{-\beta W} \rangle, \quad (2.32)$$

where the average is performed over the  $N_{\text{traj}}$  work values  $W$  associated to the alchemical trajectories. For a generic trajectory, the work is computed via the standard formula [41]

$$W = \int_0^\tau \frac{\partial U(\lambda, \mathbf{R}_L, \xi_L, \mathbf{r}_L, \mathbf{r}_R, \mathbf{r}_S)}{\partial t} dt, \quad (2.33)$$

where  $\mathbf{R}_L$  is fixed to the value of the initial microstate and  $\tau$  is the (simulation) time of the alchemical trajectory. We outline that the explicit dependence on time lies only on the  $\lambda$  parameter, while the other variables are uncontrolled degrees of freedom.

Once the quantity  $g(1) - g(0)$  is estimated, the contribution  $\Delta G_1^\circ$  to the ABFE can be computed through the following relationship (use the second line of Eq. 2.28 into Eq. 2.22)

$$\Delta G_1^\circ = g(1) - g(0) - RT \ln \left( \frac{1}{V_I C^\circ} \frac{\sigma_{\text{cp,L}} \sigma_{\text{cp,R}}}{\sigma_{\text{gas,L}} \sigma_{\text{sol,R}}} \right) + P^\circ (\bar{V}_R - \bar{V}_{\text{RL}}). \quad (2.34)$$

Here, it is important to note that, as in thermodynamic integration method, a calculation of the binding site volume  $V_I$  needs to be carried out.

In order to avoid the calculation of  $V_I$ , which implies to know a way of evaluating the function  $I(\mathbf{R}_L)$ , we may resort to an indirect method based on the calculation of the denominator of the first line of Eq. 2.31. This is the second approach that we propose, outlining that other similar schemes based on equilibrium simulations have been published [42]. Noting that  $e^{-\beta\phi(1,\mathbf{R}_L)}$  does not depend on  $\mathbf{R}_L$  and that  $\int I(\mathbf{R}_L)J_{\mathbf{R}_L}d\mathbf{R}_L = V_I$ , we can rewrite Eq. 2.31 as

$$g(1) - g(0) = -RT \ln \frac{V_I e^{-\beta\phi(1,\mathbf{R}_L)}}{\int I(\mathbf{R}_L)J_{\mathbf{R}_L}e^{-\beta\phi(0,\mathbf{R}_L)}d\mathbf{R}_L}. \quad (2.35)$$

In the previous equation, the quantity  $\mathcal{V}_{\xi_L}$  does not appear in the numerator because the integral over the orientational coordinates of L is included into  $e^{-\beta\phi(1,\mathbf{R}_L)}$  (see Eq. 2.29). Substituting Eq. 2.35 into Eq. 2.34 yields

$$\Delta G_1^\circ = -RT \ln \left( \frac{\sigma_{\text{cp,L}} \sigma_{\text{cp,R}}}{C^\circ \sigma_{\text{gas,L}} \sigma_{\text{sol,R}}} \frac{e^{-\beta\phi(1,\mathbf{R}_L)}}{\int I(\mathbf{R}_L)J_{\mathbf{R}_L}e^{-\beta\phi(0,\mathbf{R}_L)}d\mathbf{R}_L} \right) + P^\circ(\bar{V}_R - \bar{V}_{\text{RL}}). \quad (2.36)$$

With respect to the first alchemical scheme represented by Eq. 2.34, in this second approach the knowledge of  $V_I$  is not necessary. On the other side, here we need to compute the integral over  $\mathbf{R}_L$ , and especially to determine the difference between the potentials of mean force for the coupled and uncoupled systems as a function of  $\mathbf{R}_L$ , *i.e.*  $\phi(0, \mathbf{R}_L) - \phi(1, \mathbf{R}_L)$ . Indeed, this may not be a simple task. To tackle the problem, we resort to an intermediate configuration of the complex RL featured by an established position  $\mathbf{R}'_L$  of L, the corresponding potential of mean force in the coupled state being  $\phi(0, \mathbf{R}'_L)$ . The definition of this configurational state allows us to write

$$\frac{e^{-\beta\phi(1,\mathbf{R}_L)}}{\int I(\mathbf{R}_L)J_{\mathbf{R}_L}e^{-\beta\phi(0,\mathbf{R}_L)}d\mathbf{R}_L} = \frac{e^{\beta[\phi(0,\mathbf{R}'_L)-\phi(1,\mathbf{R}'_L)]}}{\int I(\mathbf{R}_L)J_{\mathbf{R}_L}e^{\beta[\phi(0,\mathbf{R}'_L)-\phi(0,\mathbf{R}_L)]}d\mathbf{R}_L}, \quad (2.37)$$

where, considering that  $\phi(1, \mathbf{R}_L)$  is independent upon  $\mathbf{R}_L$ , the equality  $\phi(1, \mathbf{R}_L) = \phi(1, \mathbf{R}'_L)$  has been used. Numerator and denominator of the right side of Eq. 2.37 can be computed separately. The denominator can be computed from an equilibrium simulation of the bounded RL complex (for tight binding) or using a method to sample preferentially bounded complex configurations, such as the umbrella sampling method [29] (for soft binding). In any case, configurations featured by  $\mathbf{R}_L = \mathbf{R}'_L$  must be sampled during the equilibrium simulation, for being defined the function  $\phi(0, \mathbf{R}'_L)$ . Therefore, even if  $\mathbf{R}'_L$  can

in principle be chosen arbitrarily, it is statistically convenient that  $I(\mathbf{R}'_L) = 1$ , or better that  $\mathbf{R}'_L$  falls in a binding site region with small value of the potential of mean force (high probability region).

From the operational point of view, it is important to note that the denominator of Eq. 2.37 corresponds to the probability density of finding the ligand at the position  $\mathbf{R}'_L$  once the complex RL is formed, namely such that  $I(\mathbf{R}_L) = 1$ . This can be realized writing the denominator as follows

$$\rho(\mathbf{R}'_L) \equiv \frac{e^{-\beta\phi(0,\mathbf{R}'_L)}}{\int I(\mathbf{R}_L) J_{\mathbf{R}_L} e^{-\beta\phi(0,\mathbf{R}_L)} d\mathbf{R}_L} = \frac{\delta p(\mathbf{R}'_L)}{J_{\mathbf{R}'_L} \delta \mathbf{R}_L}, \quad (2.38)$$

where  $\delta p(\mathbf{R}'_L)$  is the infinitesimal probability that the system is found in the volume element  $J_{\mathbf{R}'_L} \delta \mathbf{R}_L$  centered into  $\mathbf{R}'_L$  during an equilibrium sampling of the *only* bounded state of the complex. Note that the Jacobian determinant  $J_{\mathbf{R}'_L}$  is computed at the position  $\mathbf{R}'_L$ . Let suppose that the bounded state of the complex is sampled through an equilibrium simulation, or, more generally, through a simulation adopting some biasing potential (*e.g.*, using umbrella sampling), in which we are able to sample the bounded state of the complex. In such a situation, we can define a position  $\mathbf{R}'_L$  of L and a resolution  $\delta \mathbf{R}_L$  for establishing when the system takes that position. Denoting the number of times the system visits the volume element  $J_{\mathbf{R}'_L} \delta \mathbf{R}_L$  centered into  $\mathbf{R}'_L$  as  $\delta M_{\mathbf{R}'_L}$  and the total number of bounded configurations sampled during the simulation as  $M_{\text{tot}}$ , the probability of interest is simply computed as

$$\delta p(\mathbf{R}'_L) = \frac{\delta M_{\mathbf{R}'_L}}{M_{\text{tot}}}. \quad (2.39)$$

The numerator of Eq. 2.37 is estimated through an alchemical transformation. Analogously to the first alchemical scheme, a number of initial microstates are sampled by enforcing a fixed position  $\mathbf{R}'_L$  of the ligand. Starting from these microstates, nonequilibrium trajectories are performed with an established time schedule for  $\lambda$ , from  $\lambda = 0$  to  $\lambda = 1$ . The works computed from these trajectories via Eq. 2.33 are thus employed in the Jarzynski equality [40] to evaluate the free energy difference between the initial and final states, which corresponds to  $\phi(1, \mathbf{R}'_L) - \phi(0, \mathbf{R}'_L)$ .

In summary, considering the introduction of an intermediate configuration (Eq. 2.37) and the definition of probability density (Eq. 2.38),  $\Delta G_1^\circ$  can be rewritten as

$$\Delta G_1^\circ = \phi(1, \mathbf{R}'_L) - \phi(0, \mathbf{R}'_L) - RT \ln \left( \rho(\mathbf{R}'_L) \frac{\sigma_{\text{cp,L}} \sigma_{\text{cp,R}}}{C^\circ \sigma_{\text{gas,L}} \sigma_{\text{sol,R}}} \right) + P^\circ (\bar{V}_R - \bar{V}_{\text{RL}}), \quad (2.40)$$

where  $\phi(1, \mathbf{R}'_L) - \phi(0, \mathbf{R}'_L)$  and  $\rho(\mathbf{R}'_L)$  are computed as described above.

## 2.2.2 Decoupling the ligand from the solvent: $\Delta G_2^\circ$ calculation

The contribution  $\Delta G_2^\circ$  to  $\Delta G_{\text{RL}}^\circ$  is obtained considering Eq. 2.17. Substituting Eqs. 2.12 and 2.20 into Eq. 2.17 yields

$$\Delta G_2^\circ = -RT \ln \left( \frac{\sigma_{\text{sol,L}}}{\sigma_{\text{gas,L}}} \frac{Z_{N,0} Z_{0,L}}{Z_{N,L}} \right) - P^\circ \bar{V}_L, \quad (2.41)$$

where  $Z_{N,L}$ ,  $Z_{N,0}$  and  $Z_{0,L}$  are the known configurational integrals. It is interesting to note that, at variance with  $\Delta G_1^\circ$ , the contribution  $\Delta G_2^\circ$  does not depend upon the choice of the standard concentration.

In this case, the artificial energy function  $U(\lambda, \zeta_L, \mathbf{r}_L, \mathbf{r}_S)$  does not depend upon the internal coordinates of R (indeed, we deal with a molecule of L in the solvent). The requirements on  $U$  are that for  $\lambda = 0$  and  $\lambda = 1$  the artificial energy function must correspond to the energy functions of the coupled and uncoupled states of the ligand in the solvent, respectively:

$$U(0, \zeta_L, \mathbf{r}_L, \mathbf{r}_S) = U(\zeta_L, \mathbf{r}_L, \mathbf{r}_S), \quad (2.42)$$

$$U(1, \zeta_L, \mathbf{r}_L, \mathbf{r}_S) = U(\mathbf{r}_S) + U(\mathbf{r}_L). \quad (2.43)$$

We outline that the external coordinates of L are relative to the lab frame rather than to R. A free energy function dependent parametrically on  $\lambda$  can be built exploiting the artificial energy function,

$$g(\lambda) = -RT \ln \int J_{\zeta_L} e^{-\beta U(\lambda, \zeta_L, \mathbf{r}_L, \mathbf{r}_S)} d\zeta_L d\mathbf{r}_L d\mathbf{r}_S. \quad (2.44)$$

According to  $g(\lambda)$  and to the requirements of Eqs. 2.42 and 2.43, the free energy difference between the final and initial states is

$$\begin{aligned} g(1) - g(0) &= -RT \ln \frac{\int J_{\zeta_L} e^{-\beta U(\mathbf{r}_S)} e^{-\beta U(\zeta_L, \mathbf{r}_L)} d\zeta_L d\mathbf{r}_L d\mathbf{r}_S}{\int J_{\zeta_L} e^{-\beta U(\zeta_L, \mathbf{r}_L, \mathbf{r}_S)} d\zeta_L d\mathbf{r}_L d\mathbf{r}_S} \\ &= -RT \ln \frac{V V_{\xi_L} Z_{N,0} Z_{0,L}}{V V_{\xi_L} Z_{N,L}} = -RT \ln \frac{Z_{N,0} Z_{0,L}}{Z_{N,L}}. \end{aligned} \quad (2.45)$$

In the numerator,  $V$  is the volume of the container (simulation box) and arises from the integral over the position  $\mathbf{R}_L$ , while  $V_{\xi_L}$  arises from the integral over the orientation  $(\xi_{L,1}, \xi_{L,2}, \xi_{L,3})$ . In the denominator, the integrals over the internal coordinates of the solute,  $\mathbf{r}_L$ , and over the coordinates of the solvent,  $\mathbf{r}_S$ , do not depend upon the position or orientation of the solute,  $\zeta_L$ . Therefore, the integrals over  $\zeta_L$  may be carried out at once yielding the factor  $V V_{\xi_L}$ .

We may now define the free energy function  $g(\lambda)$  in terms of the potential of mean force as a function of position and orientation of the ligand:

$$g(\lambda) = -RT \ln \int J_{\zeta_L} e^{-\beta\phi(\lambda, \zeta_L)} d\zeta_L, \quad (2.46)$$

where

$$e^{-\beta\phi(\lambda, \zeta_L)} = \int e^{-\beta U(\lambda, \zeta_L, \mathbf{r}_L, \mathbf{r}_S)} d\mathbf{r}_L d\mathbf{r}_S. \quad (2.47)$$

As observed above, the integrals over  $\mathbf{r}_L$  and  $\mathbf{r}_S$  into Eq. 2.47 do not depend upon  $\zeta_L$  and hence the potential of mean force  $\phi(\lambda, \zeta_L)$  is independent on  $\zeta_L$ . This allows to write Eq. 2.46 as

$$g(\lambda) = \phi(\lambda, \zeta_L) - RT \ln(VV_{\zeta_L}). \quad (2.48)$$

Substituting Eq. 2.48 into Eq. 2.45 and the resulting equation into Eq. 2.41, we obtain

$$\Delta G_2^\circ = \phi(1, \zeta_L) - \phi(0, \zeta_L) - RT \ln \left( \frac{\sigma_{\text{sol,L}}}{\sigma_{\text{gas,L}}} \right) - P^\circ \bar{V}_L. \quad (2.49)$$

It is worth mentioning here that the knowledge of  $\sigma_{\text{gas,L}}$  is actually unnecessary because it drops out when Eqs. 2.40 and 2.49 are recasted into Eq. 2.18.

Operatively, the free energy contribution  $\Delta G_2^\circ$  to the ABFE can be computed using nonequilibrium molecular dynamics simulations in the usual way. First, a set of initial microstates is produced through an equilibrium simulation of a molecule L into  $N$  molecules of solvent (without any constraint). Starting from these microstates, nonequilibrium trajectories are performed with an established time schedule for  $\lambda$ , starting from  $\lambda = 0$  (coupled ligand) and ending to  $\lambda = 1$  (uncoupled ligand). The works computed from these alchemical trajectories via Eq. 2.33, are then employed in the Jarzynski equality [40] (Eq. 2.32) to evaluate the free energy difference  $\phi(1, \zeta_L) - \phi(0, \zeta_L)$  between the initial and final states, to be finally used into Eq. 2.49.

In the present study, we do assume that the symmetry of R and L does not change in going from the bulk phase to the complexed state, namely  $\sigma_{\text{sol,L}} = \sigma_{\text{cp,L}}$  and  $\sigma_{\text{sol,R}} = \sigma_{\text{cp,R}}$ . The symmetry number of L in the gas phase,  $\sigma_{\text{sol,L}}$ , does not matter because it drops out when  $\Delta G_1^\circ$  and  $\Delta G_2^\circ$  are recasted into Eq. 2.18. Moreover, we make the approximation that the expansion/compression work is negligible in the reaction (Eq. 2.1), so that  $P^\circ \bar{V}_L = 0$ . According to the above approximation and assumptions, the expression for  $\Delta G_2^\circ$  employed in this work simplifies

$$\Delta G_2^\circ = \phi(1, \zeta_L) - \phi(0, \zeta_L). \quad (2.50)$$



## 2.3 Using the ligand-receptor distance as binding descriptor in the double-decoupling method

First of all, we outline that, among the two DDM schemes described in Sec. 2.2, in this study we employ to the one associated with Eq. 2.40 (for  $\Delta G_1^\circ$  calculation), along with a binding descriptor based on the distance between ligand and receptor. Therefore, in the present section, we discuss how the basic relationship of the DDM, namely Eqs. 2.40 together with the “applicative” Eqs. 2.38 and 2.39, are modified upon using the only distance between ligand and receptor as the binding descriptor. A discussion on how Eq. 2.49 changes according to assumptions on the symmetry numbers and on the work of compression/expansion is reported at the end of Sec. 2.2.2.

Specifically, we adopt the distance  $r$  between the centers of mass of the  $\beta$ -CD and the aromatic compound to establish when the complex is in place, *i.e.*, when  $I(r) = 1$ . In principle, in order to apply this criterion, we need to define two threshold distances, say  $r_1^*$  and  $r_2^*$ , such that  $I(r) = 1$  if  $r_1^* < r < r_2^*$  and  $I(r) = 0$  otherwise. However, we note that the binding function  $I(\cdot)$  enters the DDM indirectly via Eq. 2.39. This suggests that one does not need to define  $r_1^*$  and  $r_2^*$ , but rather to find a way of sampling most of the binding-site region during an equilibrium simulation. This is indeed the procedure followed for the complexes formed by  $\beta$ -CD with naphthalene and anthracene. Instead, for the  $\beta$ -CD benzene complex, the bounded configurations have been sampled by means of an umbrella sampling simulation supplemented with a soft harmonic restraint. Of course, this strategy introduces an error, which can be relevant for soft-binding complexes, because the ABFE basin can be ill defined. This is why it is necessary to use a restraining potential (umbrella sampling), with a prior idea of its softness. In tight binding complexes, such a drawback is less dramatic because the sampling is naturally limited around the binding pocket during an equilibrium simulation.

We start considering that the distance  $r$  we are talking about corresponds to the distance between the origin of the R reference system and the origin of the L reference system, namely the module of  $\mathbf{R}_L$ . This suggests of using spherical coordinates to represent  $\mathbf{R}_L$ , *i.e.*  $\mathbf{R}_L \equiv (r, \theta, \phi)$ , where  $\theta$  is the angle between  $\mathbf{R}_L$  and the  $z$ -axis of the R-frame and  $\phi$  is the angle formed by the projection of  $\mathbf{R}_L$  on the  $xy$ -plane of the R-frame and the  $x$ -axis of the same frame. Then, we make explicit the coordinates  $r$ ,  $\theta$  and  $\phi$  into Eq.

2.27, expressing the binding function  $I(\cdot)$  in terms of the only  $r$  coordinate:

$$g(\lambda) = -RT \ln \int I(r) r^2 \sin \theta J_{\xi_L} e^{-\beta U(\lambda, \mathbf{R}_L, \xi_L, \mathbf{r}_L, \mathbf{r}_R, \mathbf{r}_S)} d\mathbf{R}_L d\xi_L d\mathbf{r}_L d\mathbf{r}_R d\mathbf{r}_S, \quad (2.51)$$

where for the sake of compactness  $\mathbf{R}_L \equiv (r, \theta, \phi)$ ,  $d\mathbf{R}_L \equiv dr d\theta d\phi$  and the quantity  $r^2 \sin \theta$  is the Jacobian determinant  $J_{\mathbf{R}_L}$ . The other symbols in Eq. 2.51 have preserved their original meaning. Thus, the free energy difference  $g(1) - g(0)$  of Eq. 2.28 becomes

$$\begin{aligned} g(1) - g(0) &= -RT \ln \frac{\int I(r) r^2 \sin \theta J_{\xi_L} e^{-\beta U(\mathbf{r}_R, \mathbf{r}_S)} e^{-\beta U(\mathbf{r}_L)} d\mathbf{R}_L d\xi_L d\mathbf{r}_L d\mathbf{r}_R d\mathbf{r}_S}{\int I(r) r^2 \sin \theta J_{\xi_L} e^{-\beta U(\mathbf{R}_L, \xi_L, \mathbf{r}_L, \mathbf{r}_R, \mathbf{r}_S)} d\mathbf{R}_L d\xi_L d\mathbf{r}_L d\mathbf{r}_R d\mathbf{r}_S} \\ &= -RT \ln \frac{4\pi V_I \mathcal{V}_{\xi_L} Z_{N,R} Z_{0,L}}{Z_{N,RL}}. \end{aligned} \quad (2.52)$$

In the second line of the previous equation,  $4\pi$  arises from integration over  $\theta$  and  $\phi$ ,  $\mathcal{V}_{\xi_L}$  arises from integration over the orientational coordinates of L (*i.e.*,  $\xi_L$ ) and  $V_I$  is now the integral  $\int I(r) r^2 dr$ . The second line of Eq. 2.52 allows us to write  $\Delta G_1^\circ$  of Eq. 2.22 as (*viz.* Eq. 2.34)

$$\Delta G_1^\circ = g(1) - g(0) - RT \ln \left( \frac{1}{4\pi V_I C^\circ} \frac{\sigma_{\text{cp,L}} \sigma_{\text{cp,R}}}{\sigma_{\text{gas,L}} \sigma_{\text{sol,R}}} \right) + P^\circ (\bar{V}_R - \bar{V}_{RL}). \quad (2.53)$$

The unnormalized average probability of finding the ligand in a generic point at a distance  $r$  from the origin of the R-frame (for a given  $\lambda$ ) corresponds, up to a multiplication factor, to the radial distribution function, which, in turn, equals the exponential of the potential of mean force,  $e^{-\beta\phi(\lambda,r)}$ :

$$e^{-\beta\phi(\lambda,r)} = \frac{1}{4\pi} \int \sin \theta J_{\xi_L} e^{-\beta U(\lambda, r, \theta, \phi, \xi_L, \mathbf{r}_L, \mathbf{r}_R, \mathbf{r}_S)} d\theta d\phi d\xi_L d\mathbf{r}_L d\mathbf{r}_R d\mathbf{r}_S. \quad (2.54)$$

The quantity  $4\pi r^2 e^{-\beta\phi(\lambda,r)} dr$  is therefore proportional to the probability of finding L into a spherical shell of radius  $r$  and thickness  $dr$  centered at the origin of the R-frame. According to the above definition of potential of mean force, the free energy function  $g(\lambda)$  (Eq. 2.51) becomes

$$g(\lambda) = -RT \ln \int I(r) 4\pi r^2 e^{-\beta\phi(\lambda,r)} dr. \quad (2.55)$$

Used into the first line of Eq. 2.52, the previous relationship gives the free energy difference  $g(1) - g(0)$

$$g(1) - g(0) = -RT \ln \frac{4\pi V_I e^{-\beta\phi(1,r)}}{\int I(r) 4\pi r^2 e^{-\beta\phi(0,r)} dr}. \quad (2.56)$$

In this equation, integration over  $r$  is carried out because  $\phi(1, r)$  does not depend on  $r$ . Using Eq. 2.56 into Eq. 2.53 yields

$$\Delta G_1^\circ = -RT \ln \left( \frac{\sigma_{\text{cp,L}} \sigma_{\text{cp,R}}}{C^\circ \sigma_{\text{gas,L}} \sigma_{\text{sol,R}}} \frac{e^{-\beta\phi(1,r)}}{\int I(r) 4\pi r^2 e^{-\beta\phi(0,r)} dr} \right) + P^\circ (\bar{V}_R - \bar{V}_{\text{RL}}). \quad (2.57)$$

As in Sec. 2.2.1, we introduce an intermediate configuration corresponding to  $r = r'$ . This allows to write

$$\frac{e^{-\beta\phi(1,r)}}{\int I(r) 4\pi r^2 e^{-\beta\phi(0,r)} dr} = \frac{e^{\beta[\phi(0,r') - \phi(1,r')]}}{\int I(r) 4\pi r^2 e^{-\beta\phi(0,r)} dr} e^{-\beta\phi(0,r')} \quad (2.58)$$

where, the equality  $\phi(1, r) = \phi(1, r')$  has been used because  $\phi(1, r)$  is independent upon  $r$ .

As already noted (Sec. 2.2.1), the free energy difference  $\phi(1, r') - \phi(0, r')$  appearing in the numerator of Eq. 2.58 is estimated through alchemical transformations. A number of initial microstates of the coupled system ( $\lambda = 0$ ) are sampled at the fixed distance  $r'$  of L from the origin of the R-frame. Starting from these microstates, nonequilibrium trajectories are performed with an established time schedule for  $\lambda$ , from  $\lambda = 0$  to  $\lambda = 1$ . The works computed from these trajectories via Eq. 2.33 are thus employed in the Jarzynski equality [40] (Eq. 2.32).

The remaining part of the numerator of Eq. 2.58 is computed noting that it corresponds to the probability density of finding L in a generic point at the distance  $r'$  from the origin of the R-frame, once the complex is in a bounded configuration, *i.e.*,  $I(r) = 1$ :

$$\rho(r') \equiv \frac{e^{-\beta\phi(0,r')}}{\int I(r) 4\pi r^2 e^{-\beta\phi(0,r)} dr} = \frac{\delta p(r')}{4\pi r'^2 \delta r}, \quad (2.59)$$

where  $\delta p(r')$  is the infinitesimal probability that the system is in a spherical shell of volume  $4\pi r'^2 \delta r$  around the radius  $r'$  during an equilibrium simulation of the *only* bounded state of the complex. As explained in Sec. 2.2.1 (see discussion about Eq. 2.39), computing  $\delta p(r')$  from an equilibrium simulation of the bounded RL complex is straightforward.

In summary, considering the introduction of an intermediate configuration (Eq. 2.58) and the definition of probability density (Eq. 2.59),  $\Delta G_1^\circ$  of Eq. 2.57 can be rewritten as

$$\Delta G_1^\circ = \phi(1, r') - \phi(0, r') - RT \ln \left( \rho(r') \frac{\sigma_{\text{cp,L}} \sigma_{\text{cp,R}}}{C^\circ \sigma_{\text{gas,L}} \sigma_{\text{sol,R}}} \right) + P^\circ (\bar{V}_R - \bar{V}_{\text{RL}}), \quad (2.60)$$

where  $\phi(1, r') - \phi(0, r')$  and  $\rho(r')$  are computed as described above.

The same assumptions applied to  $\Delta G_2^\circ$  (see discussion at the end of Sec. 2.2.2) lead to the final expression for  $\Delta G_1^\circ$ :

$$\Delta G_1^\circ = \phi(1, r') - \phi(0, r') - RT \ln \frac{\rho(r')}{C^\circ}. \quad (2.61)$$

## 2.4 Potential energy function and work in alchemical transformations

In a system subject to a continuous alchemical transformation, only the non-bonded potential energy function associated with the interactions between alchemical and non-alchemical species undergoes a switch-off or switch-on. All intramolecular interactions and the interactions between nonalchemical species are unchanged during the transformation. The general theory underlying alchemical transformations is reported in Sec. 2.2, where a generic potential energy function dependent on the alchemical parameter  $\lambda(t)$  is introduced (see Eq. 2.25).

In the program ORAC [27], employed to carry out all the MD simulations, the non-bonded contribution to the total energy of a system subject to an alchemical transformation is expressed as

$$\begin{aligned} U(\mathbf{r}_1, \dots, \mathbf{r}_N, \lambda, \eta) = & \sum_i \sum_{j>i} [1 - \lambda_{ij}(t)] \frac{Q_i Q_j}{r_{ij}} \operatorname{erfc}(\alpha r_{ij}) \\ & - \frac{\alpha}{\pi^{1/2}} \sum_i [1 - \lambda_i(t)]^2 Q_i^2 + U_{\text{rl}} + U_{\text{alch}} - U_{\text{intra}} \\ & + \sum_i \sum_{j>i} 4\epsilon_{ij} [1 - \eta_{ij}(t)] \left( \frac{1}{[\alpha \eta_{ij}(t) + (r_{ij}/\sigma_{ij})^6]^2} - \frac{1}{\alpha \eta_{ij}(t) + (r_{ij}/\sigma_{ij})^6} \right) \end{aligned} \quad (2.62)$$

where  $\alpha$  is the Ewald convergence parameter related to the width of the Gaussian spherical charge distribution,  $\operatorname{erfc}(x)$  is the complementary error function,  $r_{ij}$  is the distance between the atoms  $i$  and  $j$  and  $Q_i$  is the net charge on the atom  $i$ . The first term in the non-bonded energy corresponds to the electrostatic interactions in the direct lattice. The second term refers to the self-interactions of the Gaussian charge distributions. The  $U_{\text{rl}}$  term corresponds to the interactions between Gaussian distributions in the zero cell as well as in the infinite direct lattice, reformulated as an absolutely convergent summation in the reciprocal lattice. Its expression is the following

$$U_{\text{rl}} = \frac{1}{2\pi U} \sum_{\mathbf{m} \neq \mathbf{0}} \frac{\exp(-\pi^2 |\mathbf{m}|^2 / \alpha^2)}{|\mathbf{m}|^2} S(\mathbf{m}, \lambda) S(-\mathbf{m}, \lambda), \quad (2.63)$$

where  $V$  the unit cell volume,  $\mathbf{m}$  a reciprocal lattice vector and the quantity  $S(\mathbf{m}, \lambda)$  is

$$S(\mathbf{m}, \lambda) = \sum_i (1 - \lambda_i(t)) Q_i \exp(-2\pi i \mathbf{m} \cdot \mathbf{r}_i). \quad (2.64)$$

The term  $U_{\text{alch}}$  in Eq. 2.62 is an alchemical correction to the electrostatic energy having the following expression

$$U_{\text{alch}} = \sum_{ij>14} Q_i Q_j [1 - (1 - \lambda_i(t))(1 - \lambda_j(t))] \frac{\text{erf}(\alpha r_{ij})}{r_{ij}}, \quad (2.65)$$

where the sum is extended to the atoms of the alchemical species separated by more than 4 covalent bonds and  $\text{erf}(x)$  is the error function. The  $U_{\text{intra}}$  term in Eq. 2.62 is added to remove the 1-2, 1-3 and 1-4 intramolecular interactions between the alchemical species introduced spuriously in the  $U_{\text{rl}}$  contribution (Eq. 2.63):

$$U_{\text{intra}} = \sum_{ij\text{-excl.}} (1 - \lambda_i(t))(1 - \lambda_j(t)) Q_i Q_j \frac{\text{erf}(\alpha r_{ij})}{r_{ij}}, \quad (2.66)$$

where the sum is extended to all 1-2, 1-3 and 1-4 interactions. Finally, the last term in Eq. 2.62 corresponds to the modified atom-atom van der Waals interactions introduced in Ref. [43] incorporating a soft-core parametrization, where the infinity in the Lennard-Jones interaction is smoothed to zero as a function of  $\eta_{ij}(t)$ . The parameter  $\gamma$  is a positive constant (usually set to 0.5; see Ref. [44]) that controls the smoothing to zero of the derivatives of the Lennard-Jones function as  $r_{ij}$  tends to zero [45].

As we can see from Eqs. 2.62, 2.64, 2.65 and 2.66, two kinds of alchemical parameters actually inter into play. The  $\lambda_i(t)$  parameter is introduced to scale the charges of the alchemical species. Note that  $\lambda_{ij}(t)$  is not an additional parameter because it depends on  $\lambda_i(t)$  and  $\lambda_j(t)$  according to the combination rules reported in Table 2.1. The  $\eta_i(t)$  parameter is associated with the Lennard-Jones potential through  $\eta_{ij}(t)$  (also reported in Table 2.1). The time dependence of the  $\lambda_i(t)$  and  $\eta_i(t)$  parameters used to switch a part of the non-bonded potential energy is externally imposed using an appropriately selected time protocol. The non-bonded potential energy  $U(\mathbf{r}_1, \dots, \mathbf{r}_N, \lambda, \eta)$  (Eq. 2.62) coincides with the standard potential energy of a system with no alchemical species when all the alchemical parameters,  $\lambda_i(t)$  and  $\eta_i(t)$  are constant and equal to zero. At the other extreme, when  $\lambda_i(t) = \eta_i(t) = 1$ , the alchemical species “disappear” according to the combination rules for  $\lambda_{ij}(t)$  and  $\eta_{ij}(t)$  specified in Table 2.1. These rules are such that the modified alchemical potential is enforced only when one of the two interacting

---

$i$	$j$	$\lambda_{ij}(t)$	$\eta_{ij}(t)$
Alchemical	Nonalchemical	$\lambda_i(t)$	$\eta_i(t)$
Nonalchemical	Nonalchemical	$\lambda_j(t)$	$\eta_j(t)$
Nonalchemical	Nonalchemical	0	0
Alchemical	Alchemical	0	0

---

**Table 2.1:** Atom-atom combination rules for alchemical and nonalchemical species.

atoms is alchemical while atom-atom interactions within a given alchemical species are accounted for with the standard potential or simply set to zero when they do refer to atoms on different alchemical species. In general, the time protocol for the van der Waals and electrostatic control parameters may differ from each other, which gives us some flexibility in devising the protocol of the alchemical transformation. For example, in a decoupling process (switch off of the non-bonded potential), in order to prevent interatomic overlap during the simulated trajectory, the first parameter undergoing increase (from 0 to 1) is  $\lambda_i(t)$ . Once the electrostatic interactions of the alchemical species are switched off, the parameters  $\eta_i(t)$  are increased (from 0 to 1) according to some other time schedule. Indeed, this is the time protocol adopted in our alchemical calculations, as it will be detailed in Sec. 3.1.

In the context of nonequilibrium alchemical transformations, the work performed on the system to drive the alchemical parameters from 1 to 0 during a simulation of length  $\tau$  can be written as

$$W = \int_0^\tau \frac{\partial H(x, \lambda, \eta)}{\partial \lambda} \dot{\lambda} dt + \int_0^\tau \frac{\partial H(x, \lambda, \eta)}{\partial \eta} \dot{\eta} dt. \quad (2.67)$$

In a NVT or NPT simulation with an ongoing alchemical process, the alchemical work, Eq. 2.67, could be computed simply by monitoring the changes in the total energy of the systems, that includes the real potential and kinetic energy of system and the potential and kinetic energies of the barostat and the thermostats. This energy is a constant of the motion and hence any variation of it must correspond to the work done on the system. Alternatively, the work can be computed by analytically evaluating the  $\lambda$  and  $\eta$  derivatives of the non-bonded energy of Eq. 2.62.

The work can also be computed numerically observing that the differential work due

to a  $\delta\lambda$  or  $\delta\eta$  increment of the alchemical factors is given by

$$dW = \frac{1}{2}(U(\lambda + \delta\lambda, \eta) - U(\lambda - \delta\lambda, \eta) + U(\lambda, \eta + \delta\eta) - U(\lambda, \eta - \delta\eta)), \quad (2.68)$$

where  $U(\lambda, \eta)$  is a shorthand for  $U(\mathbf{r}_1, \dots, \mathbf{r}_N, \lambda, \eta)$ . Equation 2.68 is correct to order  $o(\delta\lambda^2)$  and  $o(\delta\eta^2)$ . The quantity  $dW$  can be easily accumulated during the simulated trajectory, providing, at the end of the trajectory, the total work to be used in nonequilibrium work relationships (see Sec. 3.2). This last approach is the one employed in our calculations, as it is computationally cheaper.

More technical details about the potential energy function of Eq. 2.62 and the calculation of the alchemical work can be found in Ref. [46].

---

## System, Simulations and Force Field

---

### 3.1 Simulation technical details and force field

In the present study, all MD simulations were performed in the isothermal-isobaric (NPT) ensemble by using cubic periodic boundary conditions. Constant pressure of 1 atm was enforced isotropically using a modification of the Parrinello-Rahman Lagrangian [47] and temperature control (298 K) was achieved by using the Nosé-Hoover thermostat [48]. Electrostatic interactions were computed using the smooth particle mesh Ewald method with the convergence parameter set to  $0.43 \text{ \AA}^{-1}$  and a grid spacing of  $\cong 1 \text{ \AA}$  for the fast Fourier transform calculation. Equations of motion were integrated using a multiple time-step r-RESPA scheme [49], with the potential subdivision described in Ref. [49]. In molecules, the fastest motions are the stretching vibrations, especially those involving hydrogen atoms. These degrees of freedom, however, have relatively little influence on many properties, including the noncovalent binding affinity we are interested in. Therefore, from the computational standpoint it is advantageous to constrain all covalent bonds involving hydrogen atoms, which allows for longer simulation time-steps and ultimately faster simulations.

The force field is based on the GLYCAM parameter set [26]. The interactions among atoms are described by a potential energy based on intramolecular and intermolecular



functions:

$$U_{tot} = U_{nb} + U_b. \quad (3.1)$$

The first term of Eq. 3.1 is the non-bonded potential energy, which is partitioned into electrostatic and van der Waals interactions:

$$U_{nb} = U_{vdw} + U_{el}. \quad (3.2)$$

The potential energy  $U_{nb}$  is described in detail in Section 3.1. Here, we remark that van der Waals interactions are modelled by using Lennard-Jones type potentials with parameters  $\sigma_{ij}$  and  $\epsilon_{ij}$  (see Eq. 2.62), for which Lorentz-Berthelot mixing rules are adopted:

$$\sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j) \quad (3.3)$$

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}. \quad (3.4)$$

The second term in Eq. 3.1 is the bonded potential energy, which is partitioned into several contributions

$$U_b = U_{stretch} + U_{bend} + U_{p-tors} + U_{i-tors} + U_{14}, \quad (3.5)$$

where  $U_{stretch}$  and  $U_{bend}$  are the stretching and bending energies, respectively:

$$U_{stretch} = \sum_{bonds} K_r (r - r_0)^2, \quad (3.6)$$

$$U_{bend} = \sum_{angles} K_\theta (\theta - \theta_0)^2, \quad (3.7)$$

where  $K_r$  and  $K_\theta$  are the bonded force constants associated with bond stretching and angle bending and  $r_0$  and  $\theta_0$  are the equilibrium values. The torsional potential includes two contributions: the improper torsional potential,  $U_{i-tors}$  and the proper torsional potential,  $U_{p-tors}$ . The latter has the following form:

$$U_{p-tors} = \sum_{proper} K_\phi [1 + \cos(n\phi - \gamma)] \quad (3.8)$$

where  $\phi$  is the dihedral angle formed by four atoms linked through three adjacent covalent bonds,  $K_\phi$  is the force constant,  $n$  accounts for the periodicity of the torsional potential

and  $\gamma$  is a phase factor. The improper torsions typically account for out-of-plane bending motions and may have a harmonic functional form

$$U_{i-tors} = \sum_{improper} K_{\zeta}(\zeta - \zeta_0)^2, \quad (3.9)$$

where  $\zeta$  is the angle indicating the deviation of a covalent bond from the plane defined by three atoms (*e.g.*, in benzene, the angle formed by the CH bond with respect to the plane formed by the carbon atoms in the frame C-CH-C).  $K_{\zeta}$  and  $\zeta_0$  are the force constant and the equilibrium value of the angle, respectively.

In Eq. 3.5, the term  $U_{14}$  represents the interaction between two atoms in positions 1-4, namely separated by three consecutive covalent bonds. This interaction consists of an electrostatic and a Lennard-Jones term properly scaled down to account for the overlap of the electronic densities of the two atoms. In general, the scaling factors depend on the force field. In present force field,  $U_{14}$  is

$$U_{14} = 0.8333 U_{14}^{el} + 0.5 U_{14}^{lj}, \quad (3.10)$$

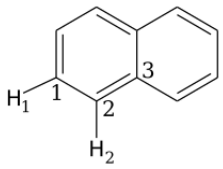
where  $U_{14}^{el}$  and  $U_{14}^{lj}$  are the standard electrostatic and Lennard-Jones energies.

The electrostatic energy depends critically on the net charges assigned to the atoms. As prescribed by the GLYCAM force field [26], the charges on the  $\beta$ -CD are determined by means of an ElectroStatic Potential (ESP) fit. In short, the ESP approach is based on the fit of the potential, computed *ab initio* on a grid around the molecule, through a distribution of point charges located on the atomic nuclei. The *ab initio* calculation of the potential has been performed using the density functional theory at the B3LYP/6-31G\* level, after a structural optimization carried out with the same level of theory. The same procedure has been employed to compute the atomic charges of benzene, naphthalene and anthracene. The TIP3P model [50] has been adopted for the water molecules. The atomic charges of benzene, naphthalene and anthracene are reported in Tables 3.1, 3.2 and 3.3, respectively.

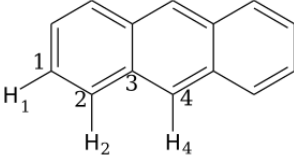
**Table 3.1:** Charges for benzene.

<b>C</b>	-0.130521
<b>H</b>	0.130521

**Table 3.2:** Charges for naphthalene

<b>C1</b>	-0.300257	<b>C2</b>	-0.137388	<b>C3</b>	0.229439	
<b>H1</b>	0.174399	<b>H2</b>	0.148526			

**Table 3.3:** Charges for anthracene

<b>C1</b>	-0.266416	<b>C2</b>	-0.123133	<b>C4</b>	-0.343480	
<b>H1</b>	0.1635645	<b>H2</b>	0.142433	<b>H4</b>	0.181745	
<b>C3</b>	0.1628695					

## 3.2 System and operative simulation schemes

The initial structure of the complex formed by benzene and  $\beta$ -CD has been taken from the PDB file 2Y4S [51]. The initial structures of the complexes having naphthalene and anthracene as ligands have been built by means of a molecular modelling program (Avogadro [52]), taking  $\beta$ -CD from the PDB file and the ligands from the database of the Avogadro program. Specifically, the ligands were in turn dragged manually into the  $\beta$ -CD cavity and then the structure of the complex has been partially optimized (through an empirical force field) to avoid large stress in the initial geometries. After obtaining the structures of the complexes, water molecules were introduced in a cubic box of 40 Å side-length in random positions, taking the complex in the center of the box. The number of water molecules for the various simulations is reported below. As stated above

thermodynamic conditions for all simulations correspond to a pressure of 1 atm and a temperature of 298 K.

As previously discussed (see Sec. 2.3), the paradigm adopted here to evaluate the binding affinity between  $\beta$ -CD and aromatic compounds (benzene, naphthalene and anthracene) is based on the distance  $r$  between the centers of mass of ligand and receptor. The basic relationships are Eqs. 2.50 and 2.61. In these equations, quantities appear that require to carry out different types of simulations. These simulations are described below.

1. *Equilibrium MD simulations of the solvated (fully coupled) ligand-receptor complex without using constraints aimed at computing  $\rho(r')$  (see Eq. 2.61).* These simulations are also used to determine  $r'$  through the evaluation of the PMF of the system as a function of the distance  $r$  (details are given in Sec. 4.2.2). The initial structures of the solvated complexes are determined as described in Sec. 3.1. The systems consist of a ligand molecule, a  $\beta$ -CD molecule and 922 water molecules. The simulation setups are very similar for the systems containing naphthalene and anthracene. To an equilibration run of about 6 ns, a production run of 54 ns follows. In the case of the complex formed by benzene, after a few hundred ps of an equilibrium simulation, we observed an unbinding event. Thus, in order to keep the  $\beta$ -CD-benzene complex in the bounded state, as required for the calculation of  $\rho(r')$ , an umbrella sampling simulation has been performed, enforcing a restraining potential  $U_{us}(r) = k_{us}r^2$  on the coordinate  $r$ . In the potential energy function  $U_{us}(r)$ , the force constant  $k_{us}$  is  $0.2 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ . Canonical averages have been computed using the proper reweighting procedure [53]. Also in this case, equilibration takes about 6 ns, while the production run is 54 ns long. In all simulations, the atomic coordinates needed for the analysis have been saved every 120 fs.
2. *Equilibrium MD simulations of the solvated (fully coupled) ligand-receptor complex with constraint  $r = r'$ , to produce the initial microstates for the nonequilibrium alchemical simulations (see point 3).* The setup for these simulations is similar to that described at the point 1, with the basic difference that, in this case, a constraint is enforced to the distance  $r$  between the centers of mass of ligand and receptor. This is accomplished by applying a stiff harmonic potential of the type  $U_{constr}(r) = k(r - r')^2$ , with force constant  $k = 1000 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  and  $r' = 1 \text{ \AA}$ . The choice of  $r'$  will be justified in Sec. 4.2.2. Clearly, in these simulations we do not need to employ umbrella sampling, because the bounded state of the complex is

preserved via  $U_{constr}(r)$ . An equilibration run of 600 ps and a production run of 600 ps have been carried out, by saving all the dynamical variables (atomic coordinates, atomic velocities, thermostat and barostat variables) every 600 fs, for a total of 1000 microstates to be used in the nonequilibrium simulations described at the point 3.

3. *Nonequilibrium alchemical simulations of the solvated ligand-receptor complex with constraint  $r = r'$  to compute the quantity  $\phi(1, r') - \phi(0, r')$  entering Eq. 2.61.* Starting from the microstates produced as described at the point 2, alchemical trajectories (for a total of 1000) have been performed by varying the parameters  $\eta$  and  $\lambda$  (see Eq. 2.62) to progressively decouple the ligand from its environment, by keeping the ligand bounded to the receptor through the potential energy  $U_{constr}(r) = k(r - r')^2$  (see point 2). All nonequilibrium alchemical trajectories have been realized by switching off the electrostatic interactions in the first half of the run, *i.e.* from 0 to  $\tau/2$ , while the Lennard-Jones interactions have been switched off in the second half of the run, *i.e.* from  $\tau/2$  to  $\tau$ . Both alchemical parameters  $\eta$  and  $\lambda$  have been varied linearly in time. At the end of the alchemical trajectories, the system consists of the solvated receptor, with the ligand in gas phase and constrained to the receptor via  $U_{constr}(r)$ . Four time schedules have been tested, differing to each other in the simulation length:  $\tau = 30, 60, 120, 240$  ps. During each nonequilibrium alchemical trajectory, the work performed on the system is computed by means of Eq. 2.68. The 1000 work values are then exploited into Eq. 2.32 to evaluate  $g(1) - g(0) \equiv \phi(1, r') - \phi(0, r')$ .
4. *Equilibrium simulations of the solvated (fully coupled) ligand to produce the initial microstates for the nonequilibrium alchemical simulations* (see point 5). Also in this case, the simulation setup is similar to that described at the point 1. Here, the system consists of a ligand molecule in a solvent consisting of 922 (benzene) or 920 (naphthalene and anthracene) water molecules. In all systems, equilibration and production runs lasting 6 and 0.6 ns, respectively, have been carried out. During the production run, the dynamical variables (atomic coordinates, atomic velocities, thermostat and barostat variables) have been saved every 600 fs, for a total of 1000 microstates to be used in the nonequilibrium simulations described at the point 5.
5. *Nonequilibrium alchemical simulations of the solvated ligand to compute the quantity  $\phi(1, \zeta_L) - \phi(0, \zeta_L)$  entering Eq. 2.50.* Starting from the microstates produced

as described at the point 4 (ligand into solvent), alchemical trajectories (for a total of 1000) have been performed by varying the parameters  $\eta$  and  $\lambda$  (see Eq. 2.62) to progressively decouple the ligand from the solvent. All nonequilibrium alchemical trajectories have been realized by switching off the electrostatic interactions in the first half of the run, *i.e.* from 0 to  $\tau/2$ , while the Lennard-Jones interactions have been switched off in the second half of the run, *i.e.* from  $\tau/2$  to  $\tau$ . Both alchemical parameters  $\eta$  and  $\lambda$  have been varied linearly in time. At the end of the alchemical trajectories, the system consists of neat solvent with the gas-phase ligand in the simulation box volume. Two time schedules, corresponding to different simulation time lengths, have been tested:  $\tau = 30, 240$  ps. During each nonequilibrium trajectory, the work performed on the system is computed by means of Eq. 2.68. The 1000 work values are then exploited into Eq. 2.32 to evaluate  $g(1) - g(0) \equiv \phi(1, \zeta_L) - \phi(0, \zeta_L)$ . It is worth noting that the opposite of the quantity  $\phi(1, \zeta_L) - \phi(0, \zeta_L)$  corresponds to the solvation free energy of the ligand.

## CHAPTER 4

---

### Results and Calculations

---

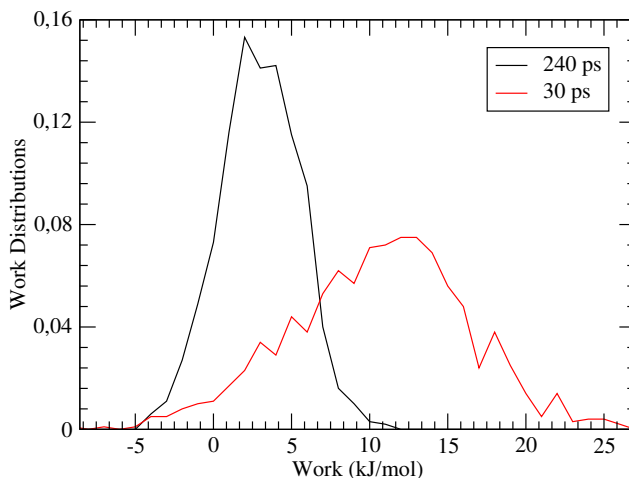
Calculations of  $\Delta G_{\text{RL}}^{\circ}$  for each complex have been split in the evaluation of two fundamental quantities,  $\Delta G_1^{\circ}$  (Eq. 2.61) and  $\Delta G_2^{\circ}$  (Eq. 2.50). The former contribution represents the decoupling free energy of the ligand from receptor and solvent in the bound-state configuration of the complex, while the latter is the solvation free energy of the ligand. The relation between these two quantities and  $\Delta G_{\text{RL}}^{\circ}$  is expressed in Eq. 2.18. In Sec. 4.1, we discuss the calculations of  $\Delta G_2^{\circ}$ , while  $\Delta G_1^{\circ}$  will be introduced in Sec. 4.2. The latter quantity is in turn split in two contributions: the decoupling free energy of the constrained bounded-state complex and the free energy contribution arising from  $\rho(r')$  (see Eq. 2.61 and related discussion).

### 4.1 Calculation of $\Delta G_2^{\circ}$

In the present section, we report on the ligand solvation free energy, which corresponds to  $\Delta G_2^{\circ}$  (Eq. 2.50). In order to rationalize the results, we also analyse the works derived from the alchemical trajectories.

## Benzene

To compute  $\Delta G_2^\circ$ , we have realized two series of nonequilibrium alchemical trajectories, according to the point 5 of Sec. 3.2. These two series of simulations are featured by different time schedules (for the evolution of the alchemical parameters  $\eta$  and  $\lambda$ ), differing in the simulation time-length, which is 30 and 240 ps. From now on, we will identify the calculations arising from the two series of simulations as *fast* and *slow*, respectively. This choice is common to all the ligands under study. The basic quantity determined during the nonequilibrium alchemical trajectories is the work performed on the system, which is employed to compute  $\Delta G_2^\circ$  as described in Sec. 2.2.2. Interesting informations about the reversibility and the accuracy of the calculation are gained from the work distribution functions, which depend not only on the nature of the process, but also on the time schedule adopted for the alchemical parameters and especially on the rate of the alchemical trajectories. The work distribution functions obtained for the decoupling process of benzene into water, realized with *fast* and *slow* simulations, are shown in Fig. 4.1. From the figure, it can be observed that the duration of the simulation affects the work



**Figure 4.1:** Works distribution functions obtained from nonequilibrium simulations of benzene into water. The simulation time corresponding to the curves is reported in the legend.

distributions significantly. These differences are typical of nonequilibrium simulations, independent on the type of process under investigation. On one side, when the rate of change of the control parameter is infinitely slow, the process is quasi-static, and hence reversible, and the total work  $W$  performed on the system equals the free energy



difference between the initial and final states, which is  $\Delta G_2^\circ$  in the present case. On the other side, when the control parameters are switched at a finite rate, the work depends on the microscopic initial conditions of the system, as well as on the fluctuations of the uncontrolled degrees of freedom during the nonequilibrium simulation trajectory. Thus, an amount of nonequilibrium trajectories can be carried out, each featured by a different work. In any case, in agreement with the second law of thermodynamics, the average work will exceed  $\Delta G_2^\circ$ , namely  $\langle W \rangle > \Delta G_2^\circ$ , even if single antidissipative trajectories (those for which  $W < \Delta G_2^\circ$ ) can be observed with some probability [40]. The difference  $\langle W \rangle - \Delta G_2^\circ$  is the average dissipated work:

$$\langle W_{diss} \rangle = \langle W \rangle - \Delta G_2^\circ. \quad (4.1)$$

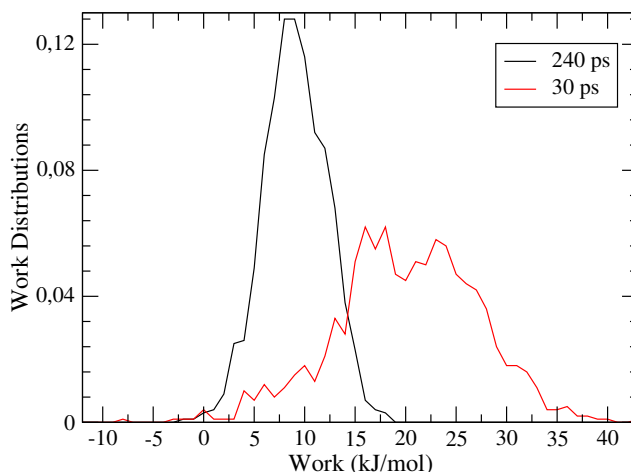
In summary, increasing the simulation time of the nonequilibrium trajectories, the process becomes more reversible. In this way, the dispersion of the work obtained in a series of trajectories decreases, and its distribution becomes narrower around an average value progressively smaller, globally lowering the dissipation. In the limit of  $\tau \rightarrow \infty$ , the work distribution  $P(W)$  takes the shape of a Dirac delta function,  $P(W) = \delta(W - \Delta G_2^\circ)$ , centred into  $\Delta G_2^\circ$ . This represents the reversible limit, according to which it is possible to consider the transformation slow enough that the system has the possibility of sampling canonically all the configurations. This is the reason why the work distribution function arising from *slow* simulations has a trend more similar to a Gaussian function. Considering the average work  $\langle W \rangle$ , we find  $\langle W_{slow} \rangle = 2.9$  kJ/mol and  $\langle W_{fast} \rangle = 10.5$  kJ/mol. The *slow* distribution appears narrower, with values approximately in the range -5 and 12 kJ/mol, while the *fast* distribution extends over a wider range of values, from  $\sim -7$  to  $\sim 27$  kJ/mol. The reason for such a marked difference is consistent with the average dissipated work  $\langle W_{diss} \rangle$ . To estimate this quantity it is necessary to evaluate the solvation free energy gained from the two simulations, according to the JE (Eq. 2.32). Applying the JE to the two sets of data of alchemical work, we obtain  $\Delta G_2^\circ = 2.7$  kJ/mol for the *fast* simulations and  $\Delta G_2^\circ = 1.6$  kJ/mol for the *slow* simulations. According to Eq 4.1, we find  $W_{diss} = 7.9$  kJ/mol and  $W_{diss} = 1.3$  kJ/mol for the *fast* and *slow* simulations, respectively.

Concerning the solvation free energy, it is important to consider that trajectories associated with low work values are those that mostly contribute to the exponential average appearing in the JE (Eq. 2.32). These critical work values are located in the left tail of the work distribution. If they are not sampled adequately, an overestimate of  $\Delta G_2^\circ$

may arise. This is a known biasing problem in calculations based on the JE [54]. The problem can be tackled in two ways: increasing the number of trajectories entering the exponential average in the JE, or increasing the simulation time of the nonequilibrium trajectories. It has been recognized that the latter approach is more efficient to improve the accuracy of the calculation. Therefore, we may consider the  $\Delta G_2^\circ$  value obtained from the *slow* simulations more accurate than that recovered from the *fast* simulations. This choice will be extended to all systems and nonequilibrium simulations considered in this study.

## Naphthalene

The second investigated complex is  $\beta$ -CD and naphthalene. The work distribution functions obtained from the nonequilibrium alchemical trajectories for the ligand-solvent decoupling process are displayed in Fig. 4.2. Although both work distributions are local-



**Figure 4.2:** Works distribution functions obtained from nonequilibrium simulations of naphthalene into water. The simulation time corresponding to the curves is reported in the legend.

ized in different work intervals, they maintain the fundamental characteristics described for the benzene system. In this case, the *fast* work distribution falls in the range  $(-4, 40)$  kJ/mol, while the *slow* one in the interval  $(-2, 17)$  kJ/mol. Comparing these results with those obtained for benzene, it emerges that, in the present case, the work distributions are wider, while the average work is significantly larger. In fact, we find  $\langle W_{fast} \rangle = 20.1$

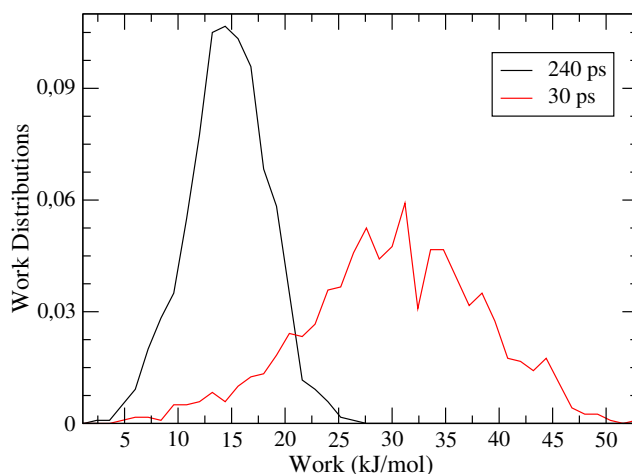
kJ/mol and  $\langle W_{slow} \rangle = 9.1$  kJ/mol. The work increase is pronounced especially for the *fast* simulation.

By applying the JE to the two series of work values, we obtain estimates of the solvation free energy:  $\Delta G_2^\circ = 8.0$  kJ/mol and  $\Delta G_2^\circ = 7.0$  kJ/mol for the *fast* and *slow* simulations, respectively. The difference between the two values is in line with our expectations and with the data obtained on benzene, but quite moderate.

Observing the larger data dispersion in the present work distributions, we can expect a greater dissipated work. This is confirmed by introducing the solvation free energies reported above in the expression for the dissipated work (Eq. 4.1):  $\langle W_{diss} \rangle = 12,1$  kJ/mol for the *fast* simulations and  $\langle W_{diss} \rangle = 2.1$  kJ/mol for the *slow* simulations.

### Anthracene

The last complex is formed by anthracene and  $\beta$ -CD. The work distribution functions obtained from the nonequilibrium alchemical trajectories for the ligand-solvent decoupling process are displayed in Fig. 4.3. It is possible to observe a first difference with respect to



**Figure 4.3:** Works distribution functions obtained from nonequilibrium simulations of anthracene into water. The simulation time corresponding to the curves is reported in the legend.

the previous functions. For anthracene, *fast* and *slow* work distributions fall in domains of positive work values. Indeed, *slow* work distribution covers a range from 1 to 26 kJ/mol, while the *fast* work distribution from 5 to 50 kJ/mol. Anyway, apart from this difference, the overall scenario does not change. The average work increases further to

30.1 kJ/mol for *fast* simulations and to 14.6 kJ/mol for *slow* simulations. The solvation free energies for *fast* and *slow* simulations are  $\Delta G_2^\circ = 15.1$  kJ/mol and  $\Delta G_2^\circ = 11.8$  kJ/mol, respectively.

The regular trend of  $\Delta G_2^\circ$  with growing the ligand dimensions can be ascribed to a lowering of intermolecular energy. In fact, with increasing the size of the ligand, water-water intermolecular energy contributions are lost without being recovered from the weaker interactions between water and ligand (this is indeed a hydrophobic molecule). In terms of dissipated work, we get:  $W_{diss} = 2.9$  kJ/mol (*slow* simulations) and  $W_{diss} = 15.0$  kJ/mol (*fast* simulations).

In conclusion, the solvation free energies of the three ligands employed in the calculations of the ABFEs reported below are summarized in Table 4.1.

**Table 4.1:** Solvation free energies in kJ/mol.

	benzene	naphthalene	anthracene
$\Delta G_2^\circ$	1.6	7.0	11.8

## 4.2 Calculation of $\Delta G_1^\circ$

In the present section, we will discuss the results obtained for  $\Delta G_1^\circ$ , *i.e.* the decoupling free energy of the ligand from receptor and solvent, expressed in Eq. 2.61. According to this equation, the calculation is split in two parts. A part dealing with the estimation of  $\Delta\phi = \phi(1, r') - \phi(0, r')$  through nonequilibrium alchemical simulations and a part concerning the determination of  $\rho(r')$  via equilibrium simulations.

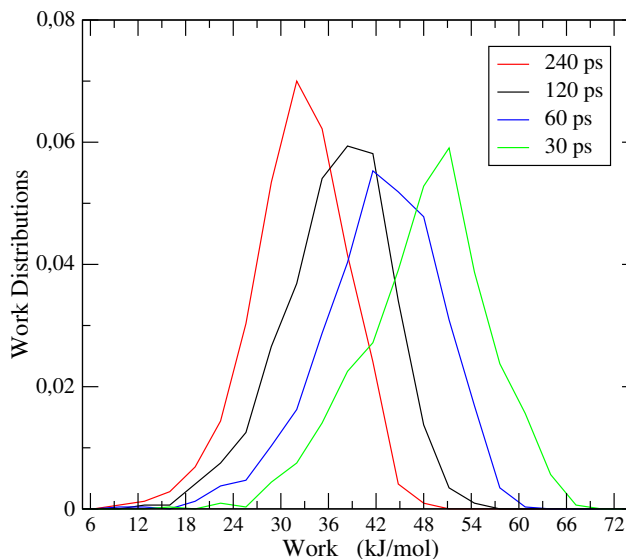
### 4.2.1 Determination of $\Delta\phi$

#### Benzene

The contribution of  $\Delta\phi$  to  $\Delta G_1^\circ$  is computed according to the points 2 and 3 of Sec. 3.2. Four different time-lengths have been considered to evaluate the degree of convergence of the free energy estimates:  $\tau = 30, 60, 120$  and  $240$  ps. The quantity  $\Delta\phi$  has been computed from the JE, obtaining the following values:  $\Delta\phi_{30} = 32.2$  kJ/mol,  $\Delta\phi_{60} = 26.9$  kJ/mol,  $\Delta\phi_{120} = 26.3$  kJ/mol and  $\Delta\phi_{240} = 22.9$  kJ/mol. A clear decrease of  $\Delta\phi$  is observed from  $\tau = 30$  ps to  $\tau = 240$  ps. As in the case of  $\Delta G_2^\circ$ , we will assume that

simulations of 240 ps provide more accurate results, and for this reason, the final quantity  $\Delta G_{\text{RL}}^{\circ}$  will be related to values obtained with that simulation time.

In Fig. 4.4, the work distributions of the decoupling process of benzene are displayed. Also in this case, it is possible to get information about accuracy of the calculation and



**Figure 4.4:** Works distribution functions of the complex with benzene obtained through nonequilibrium alchemical simulation. The simulation time related to each distribution is reported in the legend.

reversibility of the process from the shape and position of the work distributions. A clear trend can be observed with increasing  $\tau$ . *Faster* simulations are located at greater work values, as it can be inferred quantitatively from the average work:  $\langle W \rangle_{30} = 48.1$  kJ/mol,  $\langle W \rangle_{60} = 42.4$  kJ/mol,  $\langle W \rangle_{120} = 36.9$  kJ/mol and  $\langle W \rangle_{240} = 32.3$  kJ/mol. To the light of the considerations reported in the previous section, such a trend is not surprising. However, in this case, *slower* work distributions do not shrink significantly, even if an asymmetric broadening of the left-side tail seems to become more pronounced for faster simulations.

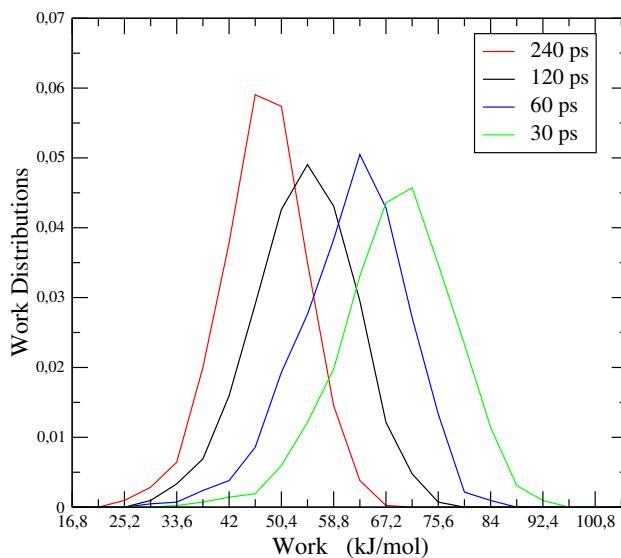
In particular, the ranges of covered works is: (25.6, 70.4) kJ/mol for  $\tau = 30$  ps, (16, 60.8) kJ/mol for  $\tau = 60$  ps, (16, 57.6) kJ/mol for  $\tau = 120$  ps and (9.6, 51.2) kJ/mol for  $\tau = 240$  ps. These distributions appear wider than those related to  $\Delta G_2^{\circ}$ , and for this reason we do expect larger values of the dissipated work  $W_{\text{diss}}$ . In fact, from  $\Delta\phi$  and the average work  $\langle W \rangle$ , we obtain  $W_{\text{diss}}^{30} = 15.9$  kJ/mol,  $W_{\text{diss}}^{60} = 15.5$  kJ/mol,  $W_{\text{diss}}^{120} = 10.6$

kJ/mol and  $W_{diss}^{240} = 9.4$  kJ/mol, where superscripts refer to  $\tau$ .

## Naphthalene

Also in the case of naphthalene four simulation times are considered. The quantity  $\Delta\phi$  has been computed from the JE, obtaining the following values:  $\Delta\phi_{30} = 48.5$  kJ/mol,  $\Delta\phi_{60} = 44.1$  kJ/mol,  $\Delta\phi_{120} = 41.1$  kJ/mol,  $\Delta\phi_{240} = 37.0$  kJ/mol. As usual, a decreasing trend of  $\Delta\phi$  is observed with increasing  $\tau$ .

The work distributions are shown in Fig. 4.5. These functions fall at greater work



**Figure 4.5:** Works distribution functions of the complex with naphthalene obtained through nonequilibrium alchemical simulation. The simulation time related to each distribution is reported in the legend.

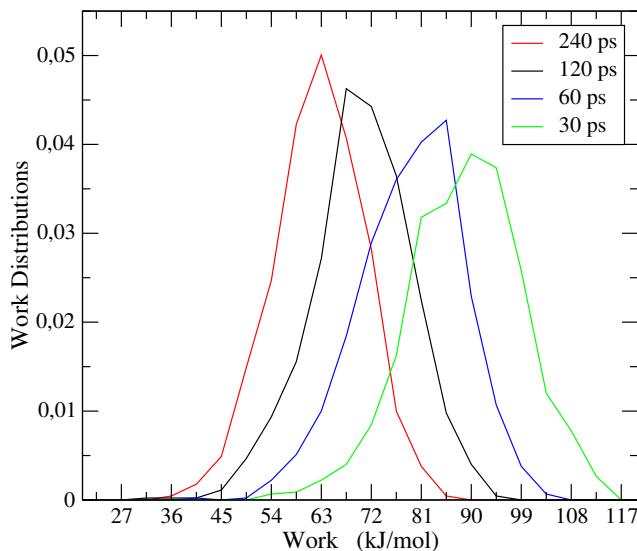
values with respect to those of benzene, and this fact is in agreement with the ligand dimensions, as discussed in the previous section. In fact,  $\langle W \rangle_{30} = 68.9$  kJ/mol,  $\langle W \rangle_{60} = 61.6$  kJ/mol,  $\langle W \rangle_{120} = 53.9$  kJ/mol,  $\langle W \rangle_{240} = 47.5$  kJ/mol. For  $\tau = 30$  ps, the work distribution ranges from 33.6 to 96.6 kJ/mol, for  $\tau = 60$  ps from 25.2 to 88.2 kJ/mol, for  $\tau = 120$  ps from 25.2 to 79.8 kJ/mol and for  $\tau = 240$  ps from 21 to 67.2 kJ/mol. These ranges show once again that the *slower* distributions are narrower, and the width decreases with the increase of  $\tau$ . Compared to the benzene case, work distributions of naphthalene cover larger ranges, resulting broader; accordingly, we do expect larger values

of  $W_{diss}$ . From Eq. 4.1, we obtain  $W_{diss}^{30} = 20.4$  kJ/mol,  $W_{diss}^{60} = 17.5$  kJ/mol,  $W_{diss}^{120} = 12.8$  kJ/mol,  $W_{diss}^{240} = 10.4$  kJ/mol. These dissipations are slightly larger than those of benzene.

### Anthracene

In the case of anthracene, no significant departures from the behavior observed for benzene and naphthalene are observable. The values of  $\Delta\phi$  obtained for this system are  $\Delta\phi_{30} = 67.1$  kJ/mol,  $\Delta\phi_{60} = 56.1$  kJ/mol,  $\Delta\phi_{120} = 49.7$  kJ/mol and  $\Delta\phi_{240} = 48.7$  kJ/mol.

In Fig. 4.6 we report the work distributions for the complex composed by anthracene and  $\beta$ -CD. Due to the ligand size, we expect irregularities as in the case of  $\Delta G_2^\circ$ . For the



**Figure 4.6:** Works distribution functions of the complex with anthracene obtained through nonequilibrium alchemical simulation. The simulation time related to each distribution is reported in the legend.

work distribution related to  $\tau = 30$  ps, the shape is very different from a Gaussian, but also in the other distributions the symmetry is absent, and each distribution present a sharp peak; so, the complex with anthracene results as the system that differs mostly from the ideal behavior. Averages work values are the largest and their decrease is regular:  $\langle W \rangle_{30} = 88.9$  kJ/mol,  $\langle W \rangle_{60} = 79.1$  kJ/mol,  $\langle W \rangle_{120} = 70.2$  kJ/mol and  $\langle W \rangle_{240} = 62.5$  kJ/mol. In this case, the average dissipated work is  $W_{diss}^{30} = 21.8$  kJ/mol,  $W_{diss}^{60} = 23.0$  kJ/mol,  $W_{diss}^{120} = 20.5$  kJ/mol and  $W_{diss}^{240} = 13.8$  kJ/mol.

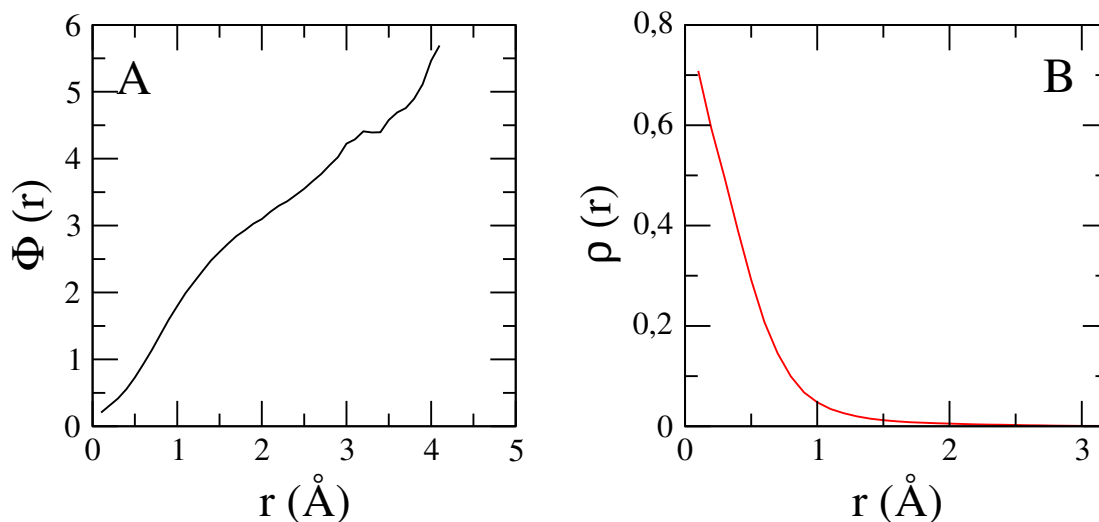
In conclusion, in Table 4.2 we report the  $\Delta\phi$  values for the three systems.

**Table 4.2:**  $\Delta\phi$  in kJ/mol for  $\tau=240$  ps

	benzene	naphthalene	anthracene
$\Delta\phi$	22.9	37.0	48.7

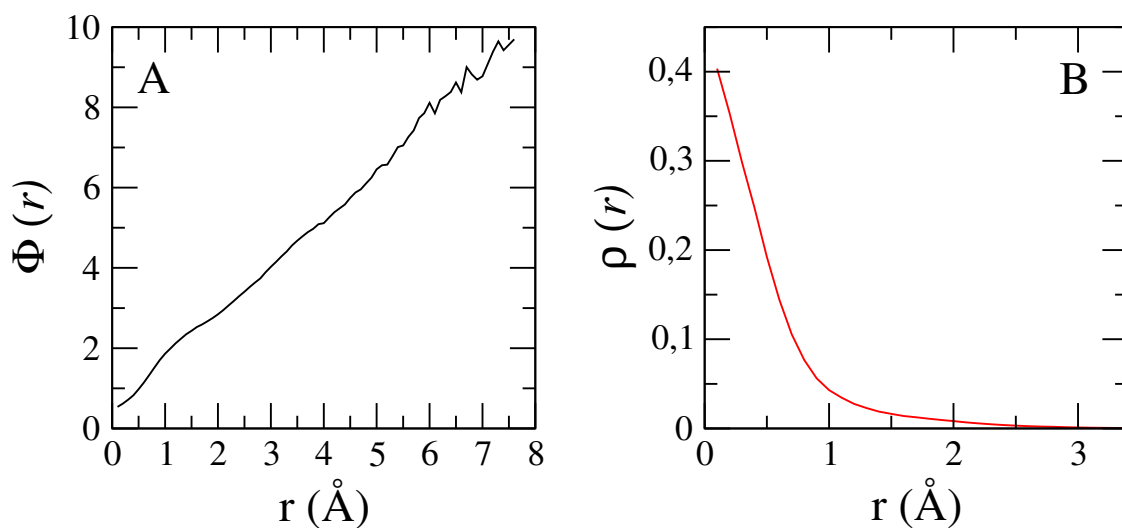
### 4.2.2 Determination of the $\rho(r')$ contribution

Numerically, the quantity  $\rho(r')$  (Eq. 2.59) is evaluated from equilibrium simulations of the ligand-receptor complex in the bounded state, as described at the point 1 of Sec. 3.2. The choice of  $r'$  for a given complex, albeit arbitrary, depends upon the PMF computed from the equilibrium simulation. In order to get a good statistical sampling,  $r'$  should correspond to highly probable configurations, namely it should fall in a domain where the value of the PMF is low. Figures 4.7, 4.8 and 4.9 display the PMF of the three complexes as a function of the distance between the centers of mass of ligand and receptor. The corresponding radial distribution functions  $\rho(r)$ , from which the PMF  $\Phi(r)$  is evaluated (through the relationship  $\Phi(r) = -RT \ln \rho(r)$ ), are also reported in the figures.

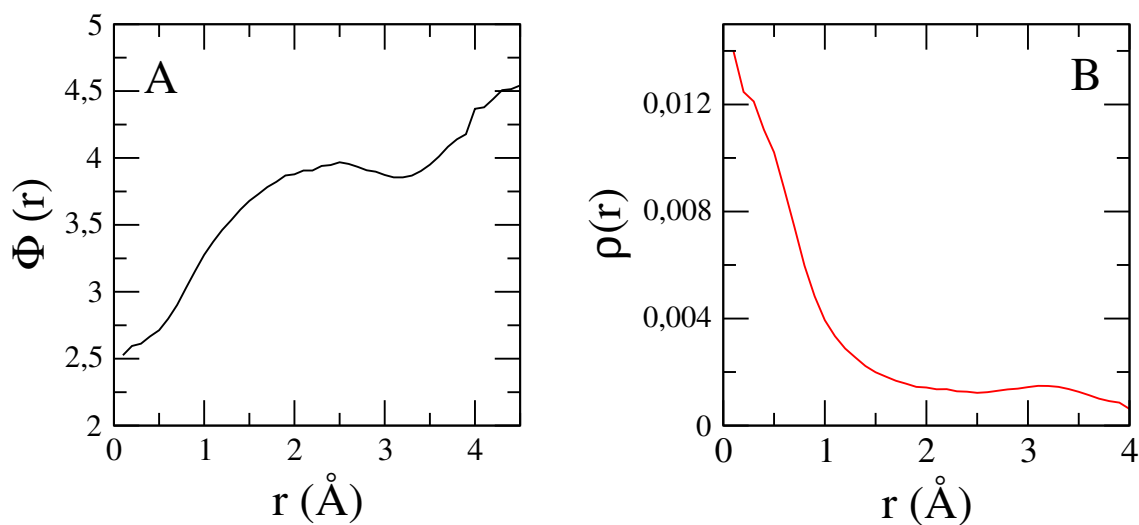


**Figure 4.7:** Panel A: PMF for the complex with benzene, evaluated through an equilibrium simulation. Panel B: radial distribution function,  $\rho(r)$ , used to compute the PMF.





**Figure 4.8:** Panel A: PMF for the complex with naphthalene, evaluated through an equilibrium simulation. Panel B: radial distribution function,  $\rho(r)$ , used to compute the PMF.



**Figure 4.9:** Panel A: PMF for the complex with anthracene, evaluated through an equilibrium simulation. Panel B: radial distribution function,  $\rho(r)$ , used to compute the PMF.

On the basis of the PMF trends, we opted for  $r' = 1 \text{ \AA}$ . According to the previous observations, it would be possible to choose a lower value for  $r'$ , corresponding to a region of higher probability. However, values around  $0 \text{ \AA}$  must be avoided to prevent numerical

errors arising from the (too small) spherical shell volume employed in the calculation of  $\rho(r)$ . In this respect, it is worth noting that, according to the theory, the specific value of  $r'$  does not affect the estimate of  $\Delta G_1^\circ$ .

Furthermore, from Eq. 2.59, we can see that  $\rho(r')$  depends apparently on  $\delta r$ . Actually, as also  $\delta p(r')$  depends (linearly) on  $\delta r$ , the probability density  $\rho(r')$  does not. In spite of this theoretical outcome, from the numerical standpoint we may observe a certain dependence of  $\rho(r')$  upon the choice of  $\delta r$ . This has been verified for all complexes, estimating  $\rho(r')$  for various values of  $\delta r$ , specifically 0.01, 0.02, 0.05, 0.1 and 0.2 Å.

In Fig. 4.10, we report the quantity  $RT \ln \frac{\rho(r')}{C^\circ}$  as a function of  $\delta r$  for the three complexes. This energy term contributes to  $\Delta G_1^\circ$  as established from Eq. 2.61. The nearly constant trend of  $RT \ln \frac{\rho(r')}{C^\circ}$  is very satisfactory, confirming the substantial independence of the free energy estimates on  $\delta r$ . The value that we have chosen to estimate  $\Delta G_1^\circ$  is  $\delta r = 0.1$  Å. With such a value, we have obtained  $RT \ln \frac{\rho(r')}{C^\circ} = 9.4, 10.0, 10.5$  kJ/mol for benzene, naphthalene and anthracene, respectively. These results are summarized in Table 4.3. According to Eq. 2.61, we can recast the data of Tables 4.2 and 4.3 into  $\Delta G_1^\circ$ ,

	benzene	naphthalene	anthracene
$RT \ln \frac{\rho(r')}{C^\circ}$	9.4	10.0	10.5

**Table 4.3:**  $RT \ln \frac{\rho(r')}{C^\circ}$  in kJ/mol computed with  $\delta r = 0.1$  Å

whose estimates are reported in Table 4.4 for the three complexes.

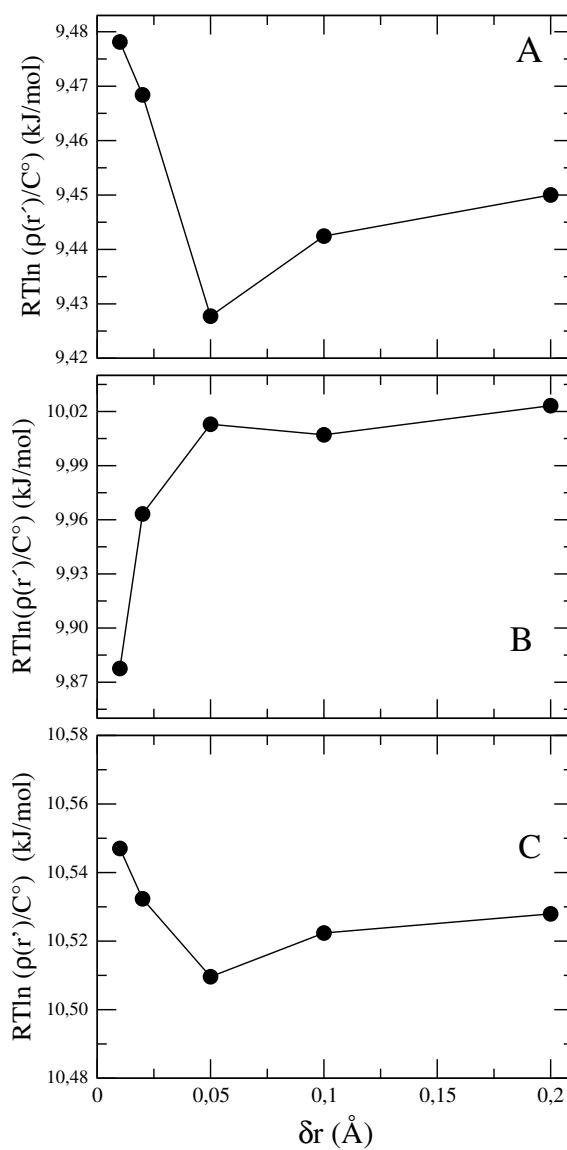
	benzene	naphthalene	anthracene
$\Delta G_1^\circ$	13.5	27.0	38.2

**Table 4.4:**  $\Delta G_1^\circ$  in kJ/mol for the three complexes.

### 4.3 Calculation of the absolute binding free energy:

$$\Delta G_{\text{RL}}^\circ$$

Estimation of  $\Delta G_2^\circ$  and  $\Delta G_1^\circ$  allows us to determine the values of  $\Delta G_{\text{RL}}^\circ$  related to the three complexes, as reported in Eq. 2.18. Thus, exploiting the data reported in the



**Figure 4.10:**  $RT \ln \frac{\rho(r')}{C^\circ}$  as a function of  $\delta r$  for the complexes with benzene, naphthalene and anthracene represented in panels A, B and C, respectively. Those behaviours confirm the limited dependence of  $RT \ln \frac{\rho(r')}{C^\circ}$  on the choice of the resolution.

Tables 4.4 and 4.1, we obtain:

$$\text{benzene} : \Delta G_{\text{RL}}^\circ = 1.6 \text{ kJ/mol} - 13.5 \text{ kJ/mol} = -11.9 \text{ kJ/mol} \quad (4.2)$$

$$\text{naphthalene} : \Delta G_{\text{RL}}^\circ = 7.0 \text{ kJ/mol} - 27.0 \text{ kJ/mol} = -20.0 \text{ kJ/mol} \quad (4.3)$$

$$\text{anthracene} : \Delta G_{\text{RL}}^{\circ} = 11.8 \text{ kJ/mol} - 38.2 \text{ kJ/mol} = -26.4 \text{ kJ/mol} \quad (4.4)$$

This is the central result of this study. To evaluate the accuracy of the calculations, we compare the computed ABFEs to the experimental counterparts, reported in Ref. [55]. The data are summarized in Table 4.5. We note that all the computed ABFEs overes-

Ligand	<b>benzene</b>	<b>naphthalene</b>	<b>anthracene</b>
Experiment	-12.7	-16.2	-18.9
Calculation	-11.9	-20.0	-26.4
$\delta\Delta G_{\text{RL}}^{\circ}$	0.8	3.8	7.5

**Table 4.5:** Absolute binding free energies for the complex ligand- $\beta$ -CD expressed in kJ/mol

timate the experimental data by an amount, which is decreasing in the order benzene, naphthalene, anthracene:  $\delta\Delta G_{\text{RL}}^{\circ} = 0.8, 3.8, 7.5$  kJ/mol. These data are also reported in Table 4.5. These features will be discussed in the next section.

## CHAPTER 5

---

### Discussion and Conclusions

---

The aim of the present study is the determination, via computer simulations, of the ABFE of complexes formed by  $\beta$ -CD with three aromatic compounds, specifically, benzene, naphthalene and anthracene. Among the variety of methods devised for ABFE calculations, we adopted MD simulations with full atomistic details adapted to realize alchemical transformations within the framework of the so-called “double decoupling method” [1].

Elements of interest and innovation are related to the theoretical strategy of supplementing alchemical transformations with a scheme based on the constrained dynamics of relevant coordinates of the system, devised to distinguish between the bound and unbound states of the complex. In our case, such a coordinate is the distance between the centers of mass of  $\beta$ -CD and ligand. Furthermore, the alchemical transformations are carried out by employing nonequilibrium simulations with alchemical parameters (regulating the intermolecular interactions between ligand and environment) as driving control parameters. Decoupling free energies are computed according to well-known nonequilibrium work theorems, such as the Jarzynski equality [40]. Our approach is alternative to computational schemes based on either equilibrium MD simulations [1, 56, 57] or nonequilibrium MD simulations without constrained dynamics [58, 59].

The results obtained using this method, summarized in Table 4.5, are very satisfac-

tory, especially for benzene. The reason for the growing overestimate (with respect to experiments) of the ABFE, with going from benzene to naphthalene and anthracene (see Table 4.5), can be ascribed to the specific nature of the calculation of the decoupling free energies, performed via Jarzynski equality. It is known that the Jarzynski equality suffers a statistical bias for a finite number of samples (work values), which can be very large [54]. Such a bias is more and more important when the nonequilibrium process becomes more irreversible, or equivalently, when the alchemical trajectories become more dissipative. This is in agreement with our analysis of the work distribution functions. For a given rate (*i.e.*, simulation time) of the nonequilibrium alchemical trajectories, dissipation increases, in average, when the molecule undergoing decoupling, which is the ligand in our case, grows in size. This is ultimately due to the fact that larger molecules introduce greater perturbations during the alchemical process.

The drawback arising from bias could be tackled making the trajectories more reversible, namely adopting longer simulation times. This is indeed confirmed by our analysis. Thus, we do expect that, increasing the simulation time beyond the longest one used in this study (240 ps), an improvement of the calculated estimates might occur. However, we must acknowledge that other sources of error, first of all the force field, may also affect the ABFE estimates.

Nonetheless, already with the current simulation time (240 ps), the differences between calculated and experimental ABFEs are quite low (less than 8 kJ/mol) and in line with the errors observed with other methodologies. Thus, in conclusion, the results of the present investigation are satisfactory for two reasons at least: the verification of the developed methodology and the validation of the force field employed for the molecules under study.

---

## Bibliography

---

- [1] M. K. Gilson, J. A. Given, B. L. Bush, and J. A. McCammon. *Biophys. J.*, 72:1047–1069, 1997.
- [2] H.-J. Woo and B. Roux. Calculation of absolute protein–ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. USA*, 102(19):6825–6830, 2005.
- [3] A. Laio and M. Parrinello. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA*, 99(20):12562–12566, 2002.
- [4] J. Fidelak, J. Juraszek, D. Branduardi, M. Bianciotto, and F. L. Gervasio. Free-energy-based methods for binding profile determination in a congeneric series of cdk2 inhibitors. *J. Phys. Chem-Us. B*, 114(29):9516–9524, 2010.
- [5] X. Biarnés, S. Bongarzone, A. V. Vargiu, P. Carloni, and P. Ruggerone. Molecular motions in drug design: the coming age of the metadynamics method. *J. Comput. Aided Mol. Des.*, 25(5):395–402, 2011.
- [6] E. Gallicchio, M. Lapelosa, and R. M. Levy. Binding energy distribution analysis method (bedam) for estimation of protein- ligand binding affinities. *J. Chem. Theory Comput.*, 6(9):2961–2977, 2010.
- [7] M. Fasnacht, R. H Swendsen, and J. M. Rosenberg. Adaptive integration method for monte carlo simulations. *Physical Review E*, 69(5):056704, 2004.

## BIBLIOGRAPHY

---

- [8] P. Procacci, M. Bizzarri, and S. Marsili. Energy-driven undocking (edu-hrem) in solute tempering replica exchange simulations. *J. Chem. Theory Comput.*, 10(1):439–450, 2013.
- [9] N. Deng, S. Forli, P. He, A. Perryman, L. Wickstrom, R.S.K. Vijayan, T. Tiefenbrunn, D. Stout, E. Gallicchio, A. J. Olson, et al. Distinguishing binders from false positives by free energy calculations: fragment screening against the flap site of hiv protease. *J. Phys. Chem-Us. B*, 119(3):976–988, 2014.
- [10] J. C. Gumbart, B. Roux, and C. Chipot. Standard binding free energies from computer simulations: What is the best strategy? *J. Chem. Theory Comput.*, 9(1):794–802, 2012.
- [11] W. L. Jorgensen and C Ravimohan. Monte carlo simulation of differences in free energies of hydration. *J. Chem. Phys.*, 83(6):3050–3054, 1985.
- [12] M. R. Shirts, D. L. Mobley, and J. D. Chodera. Alchemical free energy calculations: ready for prime time? *Annu. Rep. Comput. Chem.*, 3:41–59, 2007.
- [13] W. L. Jorgensen and Laura L. T. Perspective on free-energy perturbation calculations for chemical equilibria. *J. Chem. Theory Comput.*, 4(6):869–876, 2008.
- [14] Y. Deng and B. Roux. Computations of standard binding free energies with molecular dynamics simulations. *J. Phys. Chem-Us. B*, 113(8):2234–2246, 2009.
- [15] E. Gallicchio and R. Levy. Advances in all atom sampling methods for modeling protein–ligand binding affinities. *Curr. Opin. Struc. Biol.*, 21(2):161–166, 2011.
- [16] N. Hansen and W. F. Van Gunsteren. Practical aspects of free-energy calculations: A review. *J. Chem. Theory Comput.*, 10(7):2632–2647, 2014.
- [17] L. Wang, Y. Wu, Y. Deng, B. Kim, L. Pierce, G. Krilov, D. Lupyan, S. Robinson, M. K. Dahlgren, J. Greenwood, et al. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.*, 137(7):2695–2703, 2015.
- [18] J. G. Kirkwood. Statistical mechanics of fluid mixtures. *J. Chem. Phys.*, 3(5):300–313, 1935.



## BIBLIOGRAPHY

---

- [19] R. W. Zwanzig. High-temperature equation of state by a perturbation method. i. nonpolar gases. *J. Chem. Phys.*, 22(8):1420–1426, 1954.
- [20] C. H. Bennett. Efficient estimation of free energy differences from monte carlo data. *J. Comput. Phys.*, 22(2):245–268, 1976.
- [21] I. J. General. A note on the standard state’s binding free energy. *J. Chem. Theory Comput.*, 6(8):2520–2524, 2010.
- [22] P. Procacci. Unbiased free energy estimates in fast nonequilibrium transformations using gaussian mixtures. *J. Chem. Phys.*, 142(15):154117, 2015.
- [23] J. D. Chodera, D. L. Mobley, Mi. R. Shirts, R. W. Dixon, K. Branson, and V. S. Pande. Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struc. Biol.*, 21(2):150–160, 2011.
- [24] J. W Kaus and J. A. McCammon. Enhanced ligand sampling for relative protein–ligand binding free energy calculations. *J. Phys. Chem-Us. B*, 119(20):6190–6197, 2015.
- [25] J. Smit. *Introduction to quantum fields on a lattice*. Cambridge University Press, 2002.
- [26] R. J. Woods, R. A. Dwek, C. J. Edge, and B. Fraser-Reid. Molecular mechanical and molecular dynamic simulations of glycoproteins and oligosaccharides. 1. glycam\_93 parameter development. *J. Phys. Chem-Us.*, 99(11):3832–3846, 1995.
- [27] P. Procacci, T. A. Darden, E. Paci, and M. Marchi. Orac: A molecular dynamics program to simulate complex molecular systems with realistic electrostatic interactions. *J. Comput. Chem.*, 18(15):1848–1862, 1997.
- [28] S. Marsili, G. F. Signorini, R. Chelli, M. Marchi, and P. Procacci. Orac: A molecular dynamics simulation program to explore free energy surfaces in biomolecular systems at the atomistic level. *J. Comput. Chem.*, 31(5):1106–1116, 2010.
- [29] G.M. Torrie and J.P. Valleau. *J. Comput. Phys.*, 23:187–199, 1977.
- [30] S.Glasstone. *Thermodynamics for Chemists*. van Nostrand, New York, 1947.

## BIBLIOGRAPHY

---

- [31] M. Randall D. Lewis. *Thermodynamics*. McGraw-Hill, 1961.
- [32] K. S. Knaebel and F. B. Hill. Pressure swing adsorption: development of an equilibrium theory for gas separations. *Chem. Eng. Sci.*, 40(12):2351–2360, 1985.
- [33] A Ben Naim. *Statistical thermodynamics for chemists and biochemists*. Plenum Press, New York, 1992.
- [34] D. A. Mcquarrie. *Statistical Mechanics*. Harper and Row, New York, 1973.
- [35] D. Chandler and L. R. Pratt. Statistical mechanics of chemical equilibria and intramolecular structures of nonrigid molecules in condensed phases. *J. Chem. Phys.*, 65(8):2925–2940, 1976.
- [36] T.L. Hill. *Cooperativity Theory in Biochemistry*. Springer-Verlag, New York, 1985.
- [37] E. Giovannelli, G. Cardini, Volkov V., and R. Chelli. Nonequilibrium work theorems applied to transitions between configurational domains. *J. Stat. Mech*, in press, 2016.
- [38] H. C. Andersen. Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.*, 52(1):24–34, 1983.
- [39] J. Ryckaert, G. Ciccotti, and H. J.C. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.*, 23(3):327–341, 1977.
- [40] C. Jarzynski. *Phys. Rev. Lett.*, 78:2690, 1997.
- [41] C. Chipot and A. Pohorille. *Free energy calculations*. Springer, 2007.
- [42] Y. Deng and B. Roux. *J. Chem. Theory Comput.*, 2:1255–1273, 2006.
- [43] M. R Shirts and Vijay S Pande. Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. *J. Chem. Phys.*, 122(13):134508, 2005.
- [44] M. R Shirts, Jed W Pitera, W. C Swope, and Vijay S Pande. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *J. Chem. Phys.*, 119(11):5740–5761, 2003.

## BIBLIOGRAPHY

---

- [45] S. W. de Leeuw, J. W. Perram, and E. R. Smith. Simulation of electrostatic systems in periodic boundary conditions. i. lattice sums and dielectric constants. In *Proc. R. Soc. A*, volume 373, pages 27–56. The Royal Society, 1980.
- [46] P. Procacci and Chiara Cardelli. Fast switching alchemical transformations in molecular dynamics simulations. *J. Chem. Theory Comput.*, 10(7):2813–2823, 2014.
- [47] M. Parrinello and Aneesur Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.*, 52(12):7182–7190, 1981.
- [48] S. Nosé. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, 81(1):511–519, 1984.
- [49] D. D. Humphreys, R. A. Friesner, and B. J. Berne. A multiple-time-step molecular dynamics algorithm for macromolecules. *J. Phys. Chem-Us.*, 98(27):6885–6892, 1994.
- [50] W. L Jorgensen, J. Chandrasekhar, J. D Madura, R. W Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2):926–935, 1983.
- [51] M.B. Vester-Christensen, M.A. Hachem, B. Svensson, and A. Henriksen. Crystal structure of an essential enzyme in seed starch degradation: Barley limit dextrinase in complex with cyclodextrins. *J. Mol. Biol.*, 403(5):739–750, 2010.
- [52] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, and G. R. Hutchison. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.*, 4(1):1, 2012.
- [53] G. M Torrie and J. P. Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.*, 23(2):187–199, 1977.
- [54] J. Gore, F. Ritort, and C. Bustamante. Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proc. Natl. Acad. Sci. USA*, 100(22):12564–12569, 2003.
- [55] M. V. Rekharsky and Y. Inoue. Complexation thermodynamics of cyclodextrins. *Chemical reviews*, 98(5):1875–1918, 1998.

## BIBLIOGRAPHY

---

- [56] Y. Deng and B. Roux. Calculation of standard binding free energies: Aromatic molecules in the t4 lysozyme l99a mutant. *J. Chem. Theory Comput.*, 2(5):1255–1273, 2006.
- [57] S. Boresch, F. Tettinger, M. Leitgeb, and M. Karplus. Absolute binding free energies: a quantitative approach for their calculation. *J. Phys. Chem-Us. B*, 107(35):9535–9551, 2003.
- [58] P. Procacci. Dissociation free energies of drug–receptor systems via non-equilibrium alchemical simulations: a theoretical framework. *Phys. Chem. Chem. Phys.*, 18(22):14991–15004, 2016.
- [59] F. Nerattini, R. Chelli, and P. Procacci. Dissociation free energies in drug–receptor systems via nonequilibrium alchemical simulations: application to the fk506-related immunophilin ligands. *Phys. Chem. Chem. Phys.*, 18(22):15005–15018, 2016.

### Acknowledgements

Desidero ringraziare tutti coloro che mi hanno aiutato nella stesura della tesi con suggerimenti, critiche ed osservazioni: a loro va la mia gratitudine più sincera.

Il primo doveroso quanto spontaneo ringraziamento è rivolto al Ph. D. Edoardo Giovannelli per tutto l’aiuto fornitomi e per l’inesauribile pazienza con cui mi ha affiancato durante l’elaborazione della tesi; i suoi consigli sono stati per me fondamentali, e sono certo che farò tesoro di tutto ciò che mi ha trasmesso e insegnato.

Non posso far a meno di ringraziare il Prof. Gianni Cardini e il Dott. Marco Pagliai, per la loro costante presenza e collaborazione, per gli spunti che molto spesso mi hanno aiutato e motivato.

Grazie di cuore agli altri tesisti e tirocinanti, coloro che hanno contribuito a rendere nostra quella stanza, che mi hanno permesso di trasformare in piacevole e stimolante un periodo impegnativo come quello che abbiamo affrontato insieme.

Come potrei non nominare, infine, il Relatore Prof. Riccardo Chelli e il Correlatore Prof. Piero Procacci. Grazie alla loro determinazione e preparazione ho potuto vivere in prima persona un contesto fino ad’ora a me sconosciuto. L’impegno e l’entusiasmo con cui mi hanno guidato saranno per me fonte di ispirazione e punti di partenza.