
Modeling Soft Supramolecular Nanostructures by Molecular Simulations

Tânia F. Cova, Sandra C. Nunes, Bruce F. Milne,
Andreia F. Jorge and Alberto C. Pais

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.74939>

Abstract

The design and assembly of soft supramolecular structures based on small building blocks are governed by non-covalent interactions, selective host-guest interactions, or a combination of different interaction types. There is a surprising number of studies supporting the use of computational models for mimicking supramolecular nanosystems and studying the underlying patterns of molecular recognition and binding, in multi-dimensional approaches. Based on physical properties and mathematical concepts, these models are able to provide rationales for the conformation, solvation and thermodynamic characterization of this type of systems. Molecular dynamics (MD), including free-energy calculations, yield a direct coupling between experimental and computational investigation. This chapter provides an overview of the available MD-based methods, including path-based and alchemical free-energy calculations. The theoretical background is briefly reviewed and practical instructions are introduced on the selection of methods and post-treatment procedures. Relevant examples in which non-covalent interactions dominate are presented.

Keywords: molecular dynamics, free-energy, non-covalent interactions, host-guest systems, supramolecular structures

1. Introduction

In the past 3 decades, atomistic simulations and free-energy calculations have emerged as indispensable tools to tackle deep chemical and biological questions that experiment has left unresolved. To fully understand the vast majority of chemical and biological processes, it is often necessary to examine their underlying free-energy behavior [1]. This is the case, for

instance, of phase diagrams, protein-ligand and drug binding affinities, drug partitioning, reaction rates, equilibrium constants, acid-base equilibria, solvation contributions, or confinement energies. The prediction of these quantities, with relevance in the field of computer-aided drug design requires knowledge of the related free-energy changes [2]. Most of the research topics involving molecular modeling and simulation, with free-energy calculations, stem from pharmaceutical applications [1, 3, 4]. Those methods have provided insight into the time evolution of host-guest systems and of protein-ligand interactions [2, 4] and have revealed the respective dynamic behavior in water, thus establishing relevant recognition and affinity patterns. However, the formation of inclusion complexes is frequently used as a drug solubilizing approach, aiming at better therapeutic outcomes. This makes results sparse and focused on individual systems, lacking a comprehensive characterization for assessing the factors that govern, for example, the inclusion process. Despite the continuous progress and development of improved algorithms, the estimation of binding free energies by molecular dynamics (MD) simulations still requires significant computational efforts due to the mathematical complexity imposed by the solvated systems, often composed by myriads of atoms [3]. Recently, Pais and co-workers [5] have proposed an automated procedure based on umbrella sampling and the “flexible molecule” approximation for the calculation of binding constants in complexes formed between β -cyclodextrin and several naphthalene derivatives. In this type of complexes, the guest molecule may alter the structure of the host, leading to relevant cooperative effects, when compared to the free molecules. Inclusion of guests into host molecules is essentially governed by (i) relatively weak non-covalent interactions (NCI), including hydrophobic, van der Waals and electrostatic interactions, π - π stacking, hydrogen bonding, or a combination of these interactions, and also by (ii) stronger ionic and dipolar interactions. In addition to several reviews (see e.g. [6, 7]), including many dealing with specific host-guest systems [4, 7–12], a large number of research papers are available on this subject, including some focused on mechanistic aspects of the inclusion phenomena (see e.g. Refs. [5, 13–15].)

In fact, both the efficiency and accuracy of free-energy calculations have been greatly improved by a wide variety of methods [16]. However, it is increasingly difficult for researchers to find their way through the maze of available computational techniques. Why are there so many methods? Are they conceptually related? Do they differ in efficiency and accuracy? Why do methods that appear to be very similar carry different names? Which method is the best for a specific problem? How to choose the most relevant method to tackle the system at hand? These questions may leave researchers in the field confused and looking for clear guidelines. However, answers to these questions are not straightforward. A distinction has been made between two classes of free-energy transformations, namely those of alchemical and geometrical nature [17]. While the former exploits the malleability of the potential energy function and the virtually infinite possibilities of computer simulations to transform between chemically distinct states, the latter includes positional, orientational, and conformational changes in macromolecules and complexes thereof. At the practical level, such transformations are achieved using a variety of approaches, which can be separated into four main groups, (i) probability distributions and histograms, (ii) perturbation theory, (iii) non-equilibrium work and (iv) gradient-based methods [18]. One recurrent question concerns the selection of the most relevant approach to tackle the problem at hand, which can be reworded in terms of the best-suited, cost-effective method to obtain a reliable answer. To a large extent, this question

can be addressed by considering the actual nature of the transformation, alchemical, or geometrical, and subsequently by deciding which one, among the aforementioned methods, is more likely to yield the desired free-energy change at a minimal computational effort.

A detailed overview of the concepts and fundamentals of the available methods for estimating free energies is given in Ref. [16]. The guiding principle is that most of these are based on a few fundamental assumptions, devised by some pioneers in the field [19–23]. With some exceptions, more recent contributions are not based on innovative fundamental principles, but are “astute and ingenious ways of applying those already known” [16].

2. Context and relevant aspects

In the context of non-covalent binding, host-guest systems have emerged as useful models for assessing the accuracy of simulation methods. This can be explained by the fact that these systems display interesting features comparable to protein-ligand binding, including hydrogen bonds, conformational restriction and desolvation, and also possess an adequate size for conducting more precise thermodynamic calculations [24]. This is the case of cyclodextrin host-guest systems, which mimic several characteristics of protein-ligand binding with the advantage of being much more accessible due to their small size. These characteristics also facilitate the selection of the force field (FF) based on the attribution of the level of accuracy.

Specifically, the combination of MD and free-energy calculations with experimental results have revealed the relevant interaction patterns in the formation of inclusion complexes between several host (e.g. cyclodextrins [4, 5], cucurbiturils [25] and bambusurils [8, 26]) and guest molecules (e.g. antitumor and antiviral drugs [27], steroids [28], flavonoids [29] and small ionic species [26]). These small aggregates, comprising typically two (in 1:1 host-guest complexes) or three molecules (in 1:2 complexes) have been the basis of fundamental developments for establishing what governs associations in soft-matter and stability in larger supramolecular nanostructures, such as nanogels and targeted nanoparticles.

Despite significant advances in molecular modeling techniques and the comparative simplicity of host-guest systems, there is still a need for a tractable and theoretically sound computational method to interpret experimental data and help with the design of new hosts for targeted molecular guests. Similarly to other macromolecular systems, host-guest systems exhibit marked entropy-enthalpy compensation. Such property has been observed for NCI in both water and organic solvents. Among the wide applications of alchemical transformations, host-guest chemistry occupies the most prominent position [30].

A fundamental aspect to be understood is the precise manner in which the guest molecule binds to the host. MD simulations have greatly contributed to understanding molecular binding phenomena as a fully dynamical process. However, these simulations are limited by the time scales that can be routinely sampled. The introduction of special-purpose machines and the evolution of parallel codes, has increased enormously the time scales accessible by fully atomistic MD. However, guest molecules (e.g. drugs) with long residence times are common, and the respective association/dissociation from a host or receptor cannot be observed by

conventional MD calculations even when specialized hardware is employed. This well-recognized limitation of MD has led to the development of various algorithms to enhance the sampling of the high free-energy states and rare events, associated to high free-energy barriers.

Enhanced sampling methods can speed up conformational sampling by various means and represent an effective alternative to access with high accuracy the thermodynamics and possibly the kinetics that underlie these processes. Steered MD and umbrella sampling [31], with potential of mean force estimation (PMF) [32] allows inspecting the free-energy profile and the mechanistic aspects involved in formation of host-guest complexes [5].

In recent years, other strategies aimed at the same goal have been proposed. Replica exchange [33] metadynamics [34, 35], accelerated MD [36], milestoning [37], transition path sampling [38], and combinations of these are among the most widely used methods to enhance conformational sampling. For instance, the replica exchange methodology has been used in the atomistic simulations, as well as in a number of coarse-grained simulations (see e.g. [34]). This method has been applied to study free-energy landscape and folding mechanisms of several peptides and proteins [33], through several variants of the traditional temperature dependent scheme, available in some of the most popular MD packages, such as AMBER [39], GROMACS [40] and NAMD [41].

Among the enhanced sampling methods that fully explore the binding mechanism, metadynamics, especially in the well-tempered formulation, has emerged as a powerful approach for accelerating rare events [35]. This has been applied in drug docking to protein and enzymes, systems involving big conformational changes and relevant solvation effects. Directed dynamics such as the adaptive biasing force (ABF) and hyperdynamics were also derived from the same principles as adopted by metadynamics [35]. The implementation of metadynamics in MD codes, such as NAMD and GROMACS, have promoted a broad range of applications of the method, ranging from solid state physics to biological systems.

The adaptive biasing force algorithm (ABF) has also emerged as a promising strategy for mapping complex free-energy landscapes [42, 43], as it combines both constrained and unconstrained simulations into a highly efficient scheme, providing an uniform sampling of the order parameter. Briefly, as a simulation progresses, a continuously updated biasing force is added to the equations of motion, such that in the long-time limit it produces a Hamiltonian devoid of an average force acting along the transition coordinate of interest. In contrast to umbrella sampling schemes, based on probability distribution functions, ABF uses forces, which can be readily estimated without the need to sample broad ranges of the order parameter.

Another relevant problem in the early estimations of free-energy is related to the strong dependence on system size, in the presence of significant electrostatic interactions [44]. Once long-range corrections using Ewald lattice summation or the reaction field are included in molecular simulations, size effects in neutral systems decrease markedly. The problem, however, persists in charged systems, for example in determining the free energy of charging a neutral species in solution. In this context, it has been demonstrated that system-size dependence can be largely eliminated in these cases by careful treatment of the self-interaction term, which is associated with interactions of charged particles with their periodic images and a uniform neutralizing charge background [44].

The ability to decouple relevant energy contributions from individual components is also far from a simple solution, being the basis of alchemical free-energy methods [19, 20], such as the thermodynamic integration (TI) [17, 44, 45], free-energy perturbation (FEP) [20] and molecular mechanics-Poisson-Boltzmann surface area (MM-PBSA) [46].

The feasibility of the estimated binding free-energy depends on the correct estimation of thermodynamic quantities and the concerted interaction components (enthalpy, entropy and solvent contributions) [15, 47]. For instance, the quantification of these interaction components have been directed at inclusion complexes between cyclodextrins and its derivatives and different model drugs (see Refs. [13, 15, 48]). However, only a few studies (e.g. [5, 13]) have introduced technical and practical directions to investigate the energy components underlying the stability of these complexes. Spoel and co-workers [13] have emphasized the role of solvent contribution resorting to steered molecular dynamics and PMF calculations. The latter techniques have also been used by Pais and co-workers [5] for assessing the relevance of non-included guest moieties in the binding constants of cyclodextrin inclusion complexes, estimating the weight of enthalpic and entropic contributions to the free-energy. In this context, MD and PMF calculations allow the modulation of the formation processes in inclusion complexes, along which the free-energy profiles are derived from an umbrella sampling procedure [5]. This approach has been employed in the characterization of the inclusion of a wide range of drugs on cyclodextrins, and also in the aggregation of host units either in dimerization [15, 48] or in rotaxanes formation [49, 50]. Other hybrid approaches [51], including quantum mechanics/molecular mechanics (QM/MM) [29] methods have also provided insight on the formation of inclusion complexes at a molecular level.

3. Using potentials of mean force to address host-guest recognition and association

Several research groups have opened the way for future progress through innovative applications of free-energy methods to physical and organic chemistry, as well as structural biology. An exhaustive account of the wide range of works published in the early years of free-energy calculations falls beyond the scope of this section. The reader is referred to Refs. [16, 52] for a full description of these efforts.

A complete thermodynamic characterization of the binding process implies the knowledge of the enthalpy and entropy of association. This is one of the key elements in identifying the stabilizing factors and in understanding how is the guest and host assemble. Standard molecular simulation methods have reproduced the absolute thermodynamic properties of binding (standard free-energy, enthalpy and entropy) between host-guest systems [53]. The absolute binding free-energy can be expressed as the sum of separate free-energy contributions corresponding to a step-by-step process describing the association process between the host and guest [5]. This can be accomplished by calculating the PMF profile along a specific reaction coordinate characterizing the association process (this reaction coordinate can be the distance between the center of mass of the host and that of the guest).

Different methodologies have been successfully applied for the calculation of the PMF profile. Among them, free-energy perturbation (FEP) [47, 50, 54, 55], thermodynamic integration (TI) [45, 47], umbrella sampling, with the weighted histogram analysis method (WHAM) [5, 56] and force constraint methods [16] are the most commonly used in the literature for chemically and biologically relevant systems. PMF calculations play a considerable role in understanding how chemical species recognize and associate. One might argue that they actually provide a more satisfactory picture of these phenomena than alchemical free-energy calculations, because they follow the molecules of interest from their free, unbound state, to the bound state of the complex. This point of view is obviously misleading, as the model reaction coordinate followed to describe association/dissociation bears certain arbitrariness. Furthermore, PMF have helped to decode complex recognition and association phenomena in thermodynamic signatures with a detailed atomic description of the underlying processes. A detailed description of the theoretical basis of PMF calculation is given in ref. [16]

The PMF, $W(\xi)$ may be interpreted as the potential that is produced by the average forces over all the configurations of a given system, in which a set of molecules is kept fixed (at a certain value of a reaction coordinate, ξ). This quantity was devised by Kirkwood [19, 32] and can be defined based on the average distribution function, $\rho(\xi)$, along the coordinate ξ , obtained from a Boltzmann weighted average,

$$\langle \rho(\xi) \rangle = \frac{\int dr \delta(\xi(r) - \xi) e^{-\frac{U(r)}{k_B T}}}{\int dr e^{-\frac{U(r)}{k_B T}}} \quad (1)$$

$$W(\xi) = W(\xi^*) - k_B \ln \left[\frac{\rho(\xi)}{\rho(\xi^*)} \right] \quad (2)$$

where $W(\xi^*)$ and ξ^* are arbitrary reference values, $\delta(\xi(r) - \xi)$ represents the Dirac delta function for the coordinate ξ , $U(r)$ corresponds the total energy of the system as a function of the coordinates r , and $\xi(r)$ is a function depending on the number of degrees of freedom, such as an angle, a distance or a more complex function of the cartesian coordinates of the system.

The distribution function $\rho(\xi)$ cannot be computed by standard MC or MD simulations, due to low sampling frequency in higher-energy configurations. Specific sampling techniques (non-Boltzmann sampling) have been designed for calculating the PMF along a coordinate ξ , from a MD trajectory. The umbrella sampling technique, proposed in the 1970s by Torrie and Valleau [22], is one of these approaches, and is commonly used to overcome the difficulty of sampling rare events, by modifying the potential function so that the unfavorable states are sampled appropriately.

Umbrella sampling is used to describe simulations in which an order parameter connecting the initial and final ensembles is divided into mutually overlapping regions, or windows, which are sampled using non-Boltzmann weights. To explain briefly, several independent simulations are performed in each of the imposed regions of the reaction coordinate (**Figure 1**, bottom), using a bias potential to constrain the simulations to adjacent windows. Such initial configurations are generated, each corresponding to a location wherein the guest molecule is restrained and centered at subsequent values of ξ_i , corresponding to decreasing or increasing

center of mass (COM) distances from the center of the host cavity, using an umbrella biasing potential, $w(\xi)$, often of quadratic form, $w_i(\xi) = 0.5 K(\xi - \xi_i)^2$. This restraint allows the guest molecule to sample the configurational space in a defined region along the association/dissociation coordinate. Then the biased system (**Figure 1**, middle) will sample configurations close to a defined position, z_i , even when these would not be sampled in the unbiased system. In general, the potential energy ($U(r) + w(\xi)$) is used to generate the biased simulations. The mean force, as a function of position, is calculated in each window, and the respective potentials, $W(\xi)$, are derived using an unbiasing procedure.

Statistical techniques, such as WHAM [56] are used to remove the umbrella bias and combine the local distributions, allowing free-energy to be computed (**Figure 1**, top). WHAM is the basic tool for constructing free-energy profiles from distributions derived through stratification, for which the path connecting the reference and the target states is separated into

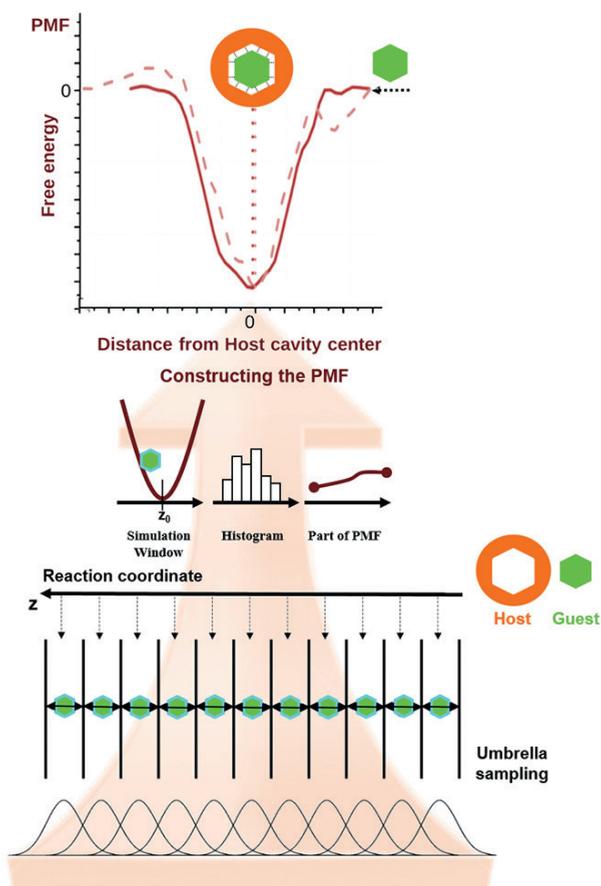


Figure 1. Schematic representation of the umbrella sampling procedure applied to a system along the pathway for the association/dissociation of a model guest with a host cavity.

intermediate states. The respective equations have also been used as the core of adaptive umbrella sampling approaches [42, 43], in which the efficiency of free-energy calculations are improved through refinement of the biasing potentials as the simulation progressed.

3.1. Thermodynamics of host-guest binding

The determination of binding affinities for host-guest complexes arising from the non-covalent association of two molecules is of paramount importance in different fields, including drug discovery. The host-guest binding affinity is related to the difference in Gibbs free energies of binding (ΔG_{bind}) corresponding to the associated and dissociated states of the system under study (see **Figure 2**).

ΔG_{bind} is composed of enthalpic and entropic terms. While the enthalpic contribution reflects changes in the inner energy of the systems, i.e. energies related to atomic motions and interactions, entropic contributions are also related with conformational changes upon binding [16].

The common methods for binding affinity estimation include very fast but less accurate similarity-based regression models and scoring functions and accurate but computationally demanding physical methods, which require a large number of MD (or Monte Carlo) simulations [57]. This suggests that the most convenient methods and algorithms are those that offer a good balance between computational costs and accuracy. Most accurate strategies provide estimates for thermodynamic quantities, with intrinsic difficulties associated to fundamental aspects. The quantification of atomic or ionic interactions, associated to repulsive and attractive forces, requires physical models assigned to different force fields (FF). The latter can be detailed using quantum mechanical (QM) *ab-initio* methods. However, a quantum mechanical representation of solvated biological systems, such as membranes and proteins consisting of large amounts of atoms is useless, even with approximations, such as the density functional theory (DFT) [58, 59]. Relevant considerations have been made [24] on the accuracy of FF for the calculation of binding thermodynamics. Some FF combinations may provide the most accurate binding enthalpies but the least accurate binding free energies, with implications in the development of new FF. These have been evaluated for a wide range of host-guest complexes and water models, using different partial charge assignment methods and host FF parameters, and resorting to the attach-pull-release (APR) method [24], which allows computing the absolute binding free energies, from a series of umbrella sampling simulations. In the



Figure 2. Schematic representations of a reversible non-covalent association/dissociation process of a guest molecule (green) and a host (orange) molecule, possessing a certain binding energy difference (ΔG_{bind}) between the associated and dissociated states.

latter, restraints controlling the host-guest complex are activated cumulatively for separating host and guest molecules, and then released to leave the guest at standard concentration.

In fact, the selection of FF and setup parameters (e.g. protonation states and slow motions) can be a difficult task, as new variants primarily concerned to the treatment of proteins, are constantly being produced. There are standard options such as OPLS [60], AMBER [61], CHARMM [62], and GROMOS [63] and other more specific FF, such as the Kirkwood-Buff [64] FF, directed at aromatic amino acids, the CHARMM polarizable FF, based on the classical Drude oscillator [65, 66], the induced dipole [67], for modeling polarization in proteins and protein-ligand complexes, the Jorgensen's approach [68] for parameterization by atoms-in-molecule electron density partitioning, AMOEBA [69] and GEM* [70], among many others [71–73]. Additionally, several water models such as TIP3P [74], TIP4Pew [75], and SPC/E [76], and the recently introduced OPC [77], TIP3P-FB and TIP4P-FB [78], TIP4P-D [79], and iAMOEBA [80] are also available. It is worth mentioning that the assessment of the ability of these FF to reproduce experimental data are very sparse, and based on the free-energy of hydration of small molecules [81], and on the structure of nucleic acids [82] and proteins [83]. With regard to explicit solvent methods, only a few studies have focused on thermodynamic data from non-covalent binding for validating the FF. Recently [84, 85], the binding free energies of cyclodextrin host-guest complexes were also used, but in the context of implicit solvent models.

It is worth to note that there are other issues in this type of systems that transcend the accuracy of the selected FF, and affect the simulation results. These are related to restrictions for the sampling and estimation of reliable thermodynamic quantities, including the respective binding constants. The latter can be determined from experimental techniques such as ¹H NMR, isothermal titration calorimetry and solubility experiments. The calculations of binding constants that provide a convenient counterpart to the experimental observations in inclusion complexes is sometimes addressed using a “flexible molecule” approximation [5], instead of the conventional rigid rotor harmonic oscillator [86–88]. The available volume for a guest molecule is described in terms of a cylindrical, PMF weighted region, in which the respective motion is measured by the positioning of the center of mass. The “flexible molecule” approximation is not as widespread as might be expected, and relatively recent publications [86, 89] both divulge the approximation and highlight controversies on the topic. In simple terms, the binding constant, K_{bind} can be estimated by integrating the PMF values ($\Delta G_{PMF}(\xi)$) along the association/dissociation coordinate, ξ , from

$$K_{bind} = \pi N_A \int r(\xi)^2 e^{\left(\frac{-\Delta G_{PMF}(\xi)}{RT}\right)} d\xi \quad (3)$$

where N_A and R corresponds to the Avogadro and the ideal gas constants, respectively and r is the average radius of the cross section of the cylinder in each sampling window, which is also a function of ξ .

The key factors affecting the thermodynamic signatures in host-guest systems can also be assessed along with an accurate description of the NCI including their spatial features. This can be carried out resorting to a recently developed Independent Gradient Method (IGM) of Lefebvre and co-workers [90] (see Section 4 for details), which allows the visualization

of regions of low charge density corresponding to stabilizing/destabilizing NCI, based on the analysis of the electronic charge density of the interacting molecules and the respective gradients. Although the original NCI method of Johnson and co-workers [91] provide a very similar qualitative analysis of NCI, the reduced density gradient, s , used to identify the interaction types is a dimensionless quantity and therefore difficult the evaluation of the respective strengths. The new IGM allows for a quantitative comparison of the strength of NCI through the calculation of the IGM descriptor, δ_g , which corresponds directly to the charge density gradient(s) in real-space.

4. Visualizing NCI using a charge density topology

The topological features of the electronic charge density have played an important part in several schemes aiming to provide a route to understanding molecular structure from first principles, that is, without resorting to any form of empirical or approximate models (other than the physically justified approximations underlying the electronic structure methods employed). Methods such as the theory of Atoms in Molecules (AIM) [92] and the Electron Localisation Function (ELF) [93–95] can provide great insight into features of the density. Both AIM and the ELF are useful for studying strong interactions such as covalent or ionic bonding but are not as useful in the analysis of very weak interactions of the type that are important for subjects as diverse as protein structure, drug design, catalysis, materials self-assembly, and many more. To address this, Johnson et al. [91, 96] introduced the NCI analysis method. This approach is based on the electronic charge density, ρ , and its derivatives. The first derivative $\nabla\rho$ enters into the expression for the reduced density gradient, s ,

$$s = \frac{1}{(2(3\pi^2))^{1/3} \rho^{4/3}} |\nabla\rho| \quad (4)$$

In regions of low ρ such as the density tails far from a molecule the gradient remains large and so s displays high values. However, in regions of low ρ between atoms where weak NCI occur the gradient (and therefore s) drop to zero.

In order to differentiate between stabilizing attractive interactions and those that are unfavorable and destabilize the system it is necessary to analyze the second derivative (Laplacian) of the charge density $\nabla^2\rho$. Decomposition of $\nabla^2\rho$ into the three eigenvalues representing the axes of maximal variation

$$\nabla^2\rho = \lambda_1 + \lambda_2 + \lambda_3, \quad (\lambda_1 \leq \lambda_2 \leq \lambda_3) \quad (5)$$

provides information on the nature of the interactions at a given point in space. λ_2 displays negative values in stabilizing/bonding regions (charge is flowing into this region indicating a local build-up of ρ) while in destabilizing/repulsive regions it is positive (charge flowing out, indicating a local depletion of ρ). Plotting s against $sign(\lambda_2)\rho$ will therefore permit the identification of NCI in regions where s and $sign(\lambda_2)\rho \rightarrow 0$. Although these quantities can be

readily obtained with first principles electronic structure methods, such calculations remain too computationally expensive for the large systems of interest in biological or materials science applications. In such cases it is possible to approximate the charge density with a sum of atomic densities providing a pro-molecular density, ρ^{pro} . This approximate density does not include relaxation of the atomic densities as would happen in self-consistent calculations for bonded systems. Fortunately, the largest deviation of ρ^{pro} from the fully relaxed density occurs in covalent bonding regions and has a minimal effect on the (low-density) NCI regions.

Substituting ρ^{pro} into the equations above has been found to have very little impact on the resulting NCI analysis. For purposes of interfacing with the results of molecular dynamics simulations, pro-molecular densities have the additional advantages that they are readily computed for both finite and extended/periodic systems and the fact that this approach requires orders of magnitude less computational time than the method based on first principles electronic structure calculations.

A recent development of the NCI method that makes use of the pro-molecular route to building the charge density of the system under study is the Independent Gradient Method (IGM) of Lefebvre et al. [90] As with the NCI method, IGM employs in addition to ρ quantities related to the first and second derivatives of the density. However, in IGM the reduced density gradient, s , is replaced with the descriptor δg^{inter} , defined as the difference between the first derivatives of the charge densities for the total system and the fragments

$$\delta g^{inter} = |\nabla \rho^{IGM, inter}| - |\nabla \rho| \quad (6)$$

$\delta g^{inter} > 0$ indicates the presence of NCI and the magnitude of the descriptor at a point in space gives an indication of the strength of the interaction. $\nabla \rho^{IGM, inter}$ is obtained from sums over all the atomic densities in the different fragments (here for two fragments A and B in the x-direction)

$$\left(\frac{\delta \rho}{\delta x}\right)^{IGM,inter} = \left| \sum_{i=1}^{N_A} \frac{\delta \rho_i}{\delta x} \right| + \left| \sum_{i=1}^{N_B} \frac{\delta \rho_i}{\delta x} \right| \quad (7)$$

The IGM approach has the attractive feature that δg^{inter} allows for a more facile comparison of the strength of the weak interactions than the quantity s in the original NCI [90]. This is due to the fact that the presence of significant non-covalent interactions is indicated when $s \rightarrow 0$ making interpretation difficult whereas the magnitude of the IGM quantity δg^{inter} increases in proportion to the strength of the interaction.

An illustrative example of the use of the NCI and IGM methods is given in **Figures 3** and **4**. The host-guest complex shown is between curcurbit[7]uril and the dodecyl serine-based monomeric cationic surfactant 12SerTFAc [97] (to make the example more accessible to the reader only a single configuration is shown). Calculation of the NCI and IGM descriptors was performed using IGMPLOT version 1.0 (<http://kisthelp.univ-reims.fr/IGMPLOT/>).

As can be seen from the molecular structure images, both the s and δg^{inter} surfaces provide qualitatively similar information and display the general distribution of the most significant

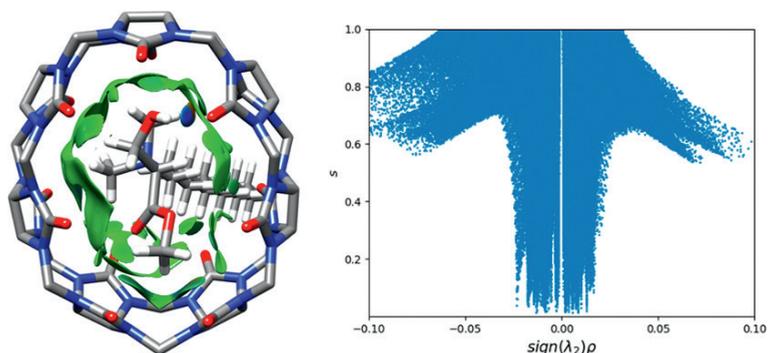


Figure 3. Host-guest complexation between curcubit[7]uril and the dodecyl serine-based monomeric cationic surfactant 12SerTFAC. (left) reduced density gradient, $s = 0.4$, isosurface colored by value of $\text{sign}(\lambda_2)\rho$ ($-0.03 \leq \text{sign}(\lambda_2)\rho \leq 0.03$ a.u.). Blue = stabilizing, red = destabilizing and green = weak interactions. (right) scatter plot of s and $\text{sign}(\lambda_2)\rho$ values.

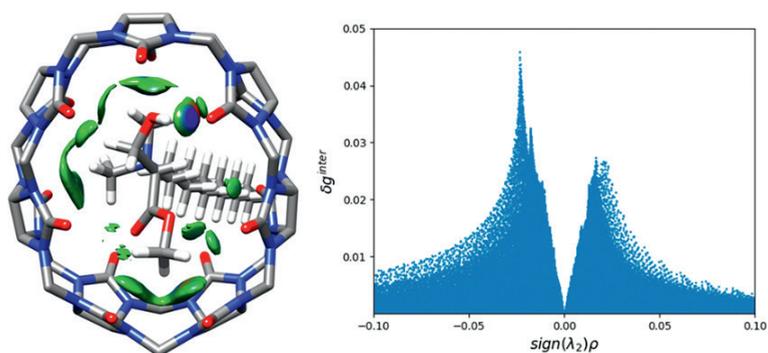


Figure 4. Host-guest complexation between curcubit[7]uril and 12SerTFAC surfactant. (left) IGM $\delta g^{\text{inter}} = 0.01$ a.u. Isosurface colored by value of $\text{sign}(\lambda_2)\rho$ ($-0.03 \leq \text{sign}(\lambda_2)\rho \leq 0.03$ a.u.). Blue = stabilizing, red = destabilizing and green = weak interactions. (right) scatter plot of δg^{inter} and $\text{sign}(\lambda_2)\rho$ values.

host-guest interactions. Similarly, the color-coding provided by the value of $\text{sign}(\lambda_2)\rho$ can be seen to predict that the interactions are largely of the weak van der Waals type with the exception of a single blue region denoting the hydrogen bond between the serine hydroxyl group and the host. A significant difference between the s and δg^{inter} surfaces is that the latter also contains information on the strength of the interactions through the volumes of the isosurface regions at a given point. The IGM approach therefore provides an important extra source of information on the nature of the interactions. The intrinsic difference between the two methods is even more clearly seen in the adjacent scatter plots. In the IGM plot (**Figure 4**, right) peaks of different magnitude corresponding to different strengths of interaction can be seen which give rise to the differing isosurface volumes. The NCI plot (**Figure 3**, right), however, displays all spikes in the non-covalent interaction region $-0.025 \leq \text{sign}(\lambda_2)\rho \leq 0.025$ in a very similar manner and makes interpretation much less easy. For this reason, despite very similar computational requirements for both methods making them equally suitable for analysis of large (ensembles

of) systems such as those arising from MD simulations it is likely that the future will see the newer IGM method being used much more extensively than the original NCI one.

5. Applications

This section presents some examples in which MD or MC simulations are used for establishing the contributions of NCI (e.g. electrostatic, hydrophobic and hydrogen-bond interactions) to the free energy, in systems in which these are the main source of recognition, assembly, and stability. These include, for instance, host-guest complexes, polyelectrolytes and proteins.

MD and free-energy calculations have provided insight on relevant aspects including the effect of substituents, the role of solvation, and rationales for the conformation and thermodynamic characterization of the systems under investigation. Among several studies on the topic, available in the literature, only a few are briefly reviewed. For instance, MD simulations in water, with PMF estimation based on both TI and constraint force methods, have been used for inspecting the thermodynamic properties of binding of some inorganic ions, such as ClO_4^- and SO_4^{2-} (**Figure 5**), and ferrocenyl alkanethiols with both free and gold-surface confined β -cyclodextrins [14, 98]. In general, the association process between the host and guest molecules is analyzed along the reaction coordinate defined by the distance between the centers of mass of both host and guest, along a reference coordinate (e.g. z-axis). In these particular systems, electrostatic interactions and hydrogen bonding have played a major role on the binding process. For instance, simulations have elucidated the association mode of sulfate anion with free and grafted β -cyclodextrin (see **Figure 5**). For the latter system, a small minimum in the PMF profile (**Figure 5**, top), positioned at a larger distance relative to the cavity of surface-grafted cyclodextrin, suggested the formation of a noninclusion complex. The ion binds to the host by forming hydrogen bonds with the free-cyclodextrin portal. Also relevant is the individual energy contributions to the cyclodextrin-ion interaction, which is governed by electrostatic interactions. However, this favorable electrostatic contribution is not sufficient to compensate desolvation of the anion, considering the respective hydration energy [14].

The release and transport of drugs mediated by cyclodextrin-based carriers, have also been investigated [99] systematically using MD and free-energy calculations, showing the preferred inclusion modes of such drugs for cyclodextrins. One example (see **Figure 6**), refers to the binding of amphotericin B (AmB), which possesses two sites, within the respective prolonged macrolide, with higher binding affinity to γ -cyclodextrin. The decomposition of the PMFs into free-energy contributions have suggested that van der Waals and electrostatic interactions are the main driving forces responsible for the formation of this type of complexes [99].

Recently, some of the authors [5] have demonstrated the relevance of non-included moieties in the stability constants of several cyclodextrin-based complexes. The binding constants for naphthalene derivatives forming complexes with β -cyclodextrin were calculated (ranging from 128 to 2.1×10^4 $\text{kJ}\cdot\text{mol}^{-1}$), pointing out the important effects of the substituents. Substitution of naphthalene promoted an increase in the binding constant (up to 100-fold), irrespective of the

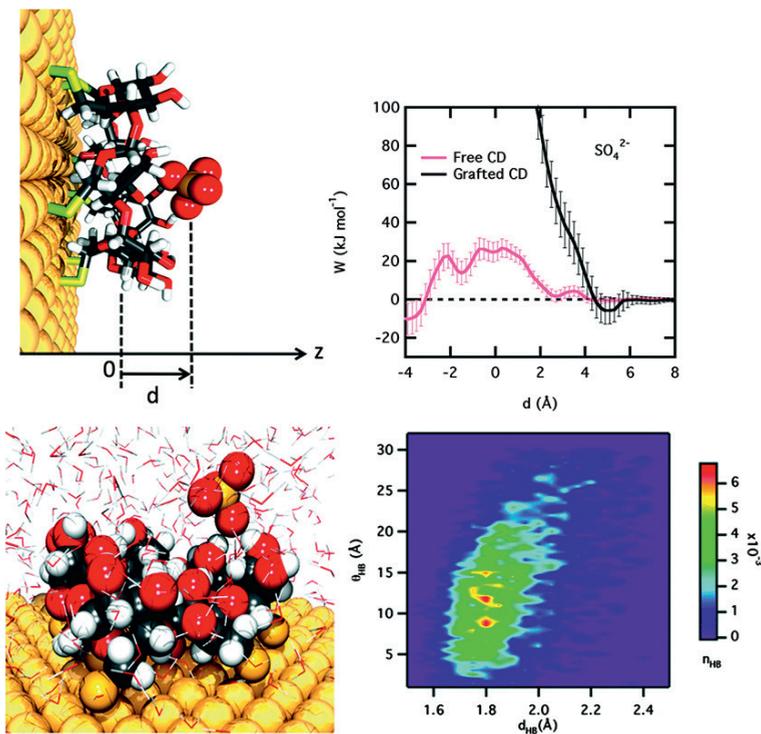


Figure 5. (Top) schematic representation of the association pathway (left) between SO_4^{2-} and a β -cyclodextrin grafted to a gold-surface, used for the construction of the PMF profile (right). The free-energy profile characterizing the association process, in water, of the ion into the free β -cyclodextrin is also shown. (Bottom) illustrative snapshot showing the presence of hydrogen bonding (right) as the ion approximates the cavity of the host, and (left) graphical distribution of distances and angles corresponding to the hydrogen bonds. Reproduced from Ref. [14] with permission from the Royal Society of Chemistry.

nature of the substituent, the latter comprising small hydrophobic and hydrophilic, including charged, groups. Enthalpic and entropic contributions were separated in order to estimate their weight in the free-energy. Also, the preferred orientation of the guest molecules within the cyclodextrin was established.

In a completely different system [26], the same strategy was used for exploring the ion caging ability of bambusurils in aqueous media (**Figure 7**). Specifically, new insights on the conformation, hydration and energy changes involved in the complexation process between the water-soluble benzoyl-substituted bambus[6]uril and chloride ion were provided. The structural features of three single bambusuril derivatives, and the relative energy contributions to the formation of the benzoyl-substituted bambus[6]uril-chloride complex were computed. The estimated standard free-energy of binding and the respective binding constant were found to be $-11.7 \text{ kJ}\cdot\text{mol}^{-1}$ and 112, respectively. Binding occurred with complete desolvation

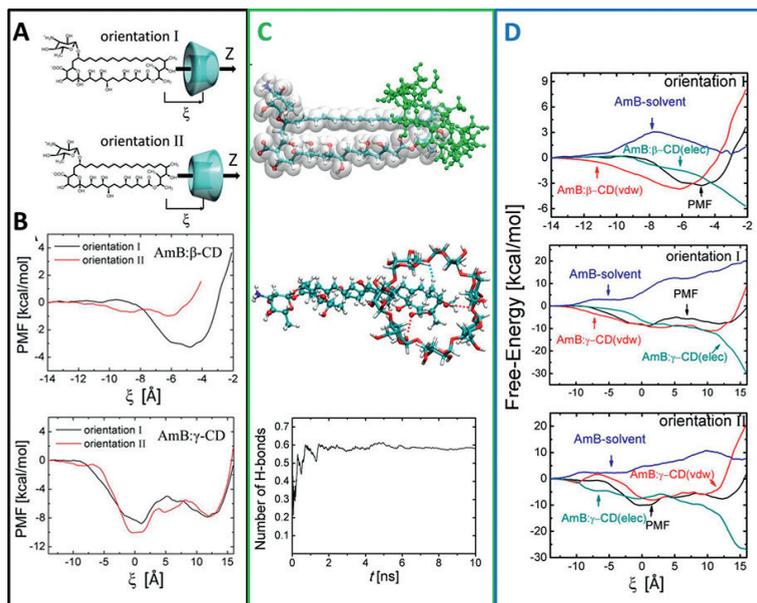


Figure 6. (A) Schematic representation of the two inclusion pathways along which AmB approximates to β - or γ -cyclodextrins. (B) PMF profiles for the inclusion of AmB with β - and γ -cyclodextrins considering the two orientations. (C) Equilibrium configuration of the complex between AmB and β -cyclodextrin (orientation I) followed by a representation of the hydrogen bonds formed between host and guest molecules and the respective averaged number ($O\cdots H\cdots O$ angle $>135^\circ$ and $O\cdots O$ distance <3.5 Å), as a function of the simulation time. (D) Decomposition of the PMFs into van der Waals and electrostatic host-guest and host-solvent contributions, for the complexes between AmB and β - and γ -cyclodextrin in orientation I, and AmB and γ -cyclodextrin in orientation II. Reproduced from Ref. [99] with permission from the American Chemical Society.

of both guest and host cavity. One of the most interesting observations was that chloride was hermetically sealed inside the host cavity, as a result of a concerted action of both conformation change and desolvation (see **Figure 7**) [26].

Solvent contributions to the thermodynamics of binding have been scrutinized [13], emphasizing the importance of using explicit models for the accurate calculation of binding thermodynamics. A detailed analysis on the energetics of the host-guest complexation between β -cyclodextrin and several model drugs (e.g. puerarin, daidzin, daidzein, and nabumetone) have demonstrated that the flexibility of the binding partners and solvation-related enthalpy and entropy changes must be included explicitly for the precise estimation of thermodynamic parameters involved in molecular association. In this study, it was also demonstrated that implicit models are not suitable to provide detailed information on how free-energy is decomposed into enthalpy and entropy [13].

The dimerization of cyclodextrins has also been recognized as a relevant step in the construction of nanostructured materials. Only a few studies (see e.g. refs [15, 48].) involving umbrella

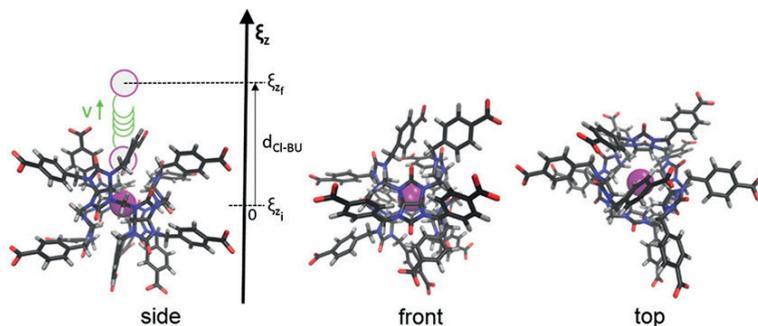


Figure 7. Typical conformation of the inclusion complex between one chloride ion (purple sphere) and the dodeca (4-carboxybenzyl)bambus[6]juril, in water, sampled during the MD run, at 298 K. (Right) Scheme of the dissociation pathway, along the z -component (ϵ_z), between host and guest. The steering force is applied on the chloride ion for generating the initial configurations for the umbrella sampling procedure. Reproduced from Ref [26] with permission from Elsevier.

sampling simulations and PMF calculations have been carried out for investigating the binding affinity of dimers in the presence/absence of a guest molecule, and the preferred orientation of interglucopyranose hydrogen bonds, at the cyclodextrin portals. These include β -cyclodextrin monomers and dimers in aqueous and nonaqueous media [48]. Polar solvents with stronger hydrogen bond accepting abilities can easily disrupt intermolecular hydrogen bonds, decreasing the stability of the dimers. In the same environment, higher binding affinities are achieved if guest molecules are included in the channel-type cavity of such dimers. These results allowed concluding that the formation of dimers is solvent-dependent and guest-modulated [48]. In another related study [15], it was shown that the cooperative binding of cyclodextrin cavities to guest molecules can facilitate the dimerization process, favoring the overall stability and assembly of nanostructures. Different dimerization modes yielded different binding strengths. This has been demonstrated in systems consisting of isoflavone drug analogs used as drug templates, and cyclodextrins (**Figure 8**). A detailed quantification of host, guest and solvent contributions to the thermodynamics of binding has revealed that head-to-head dimerization promotes the most stable complexes, which can be used as building blocks for template-stabilized nanostructures. Desolvation of cyclodextrin dimers and entropy changes upon complexation also affect significantly the cooperative binding [15].

Other strategies have been adopted resorting to both MD and MC simulations. As an example of application [100], free-energy calculations and lattice chain MC simulations have been used for understanding the formation of polyrotaxanes (**Figure 9**), including the identification of the most favorable conformations of cyclodextrin molecules in a polyrotaxane and the quantification of dimerization free energies, related to different spatial arrangements of two consecutive cyclodextrin molecules. It has been suggested, that the binary association is controlled by the formation of hydrogen bonds between two adjacent molecules, promoting an overall structural stabilization [100].

The thermodynamic cost of confining polyelectrolytes spherical cells of different radii can also be quantified by free-energy calculations, resorting to TI [101]. Simulation studies of

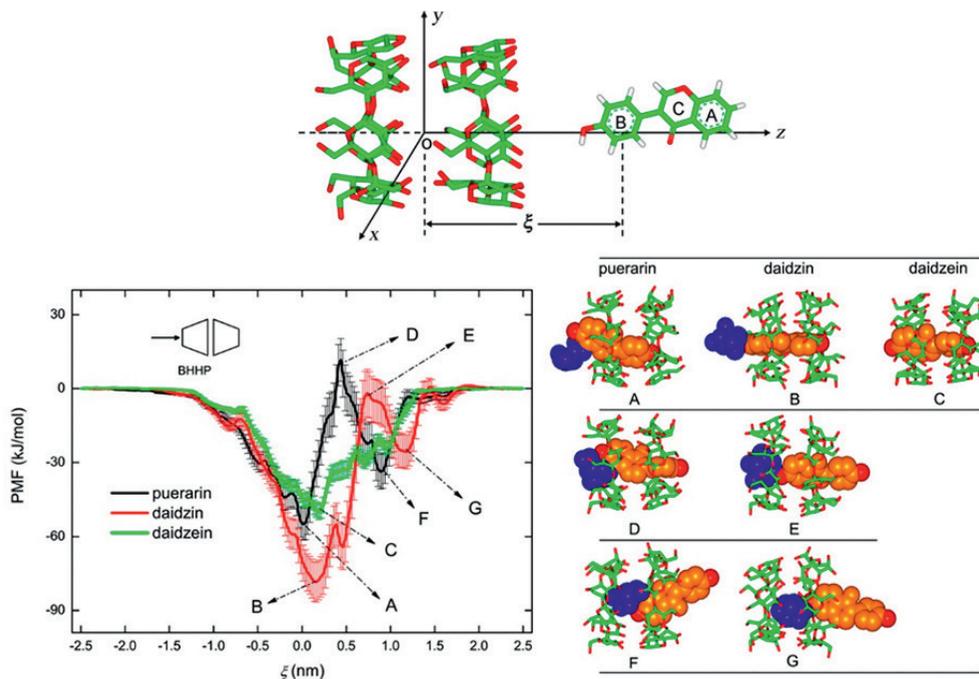


Figure 8. (Top) schematic representation of the association coordinate for the 2:1 complex between a β - cyclodextrin dimer and daidzein. (Bottom) PMF profiles and representative configurations of the 2:1 complexes between the specific BHHP arrangement of β - cyclodextrin dimer and three different model drugs. Reproduced from Ref. [15] with permission from the American Chemical Society.

polyelectrolytes, based on MC and MD, have been precluded by the respective chain lengths and the need of large concentrations of compacting agents. The conformational and energetic changes in DNA delivery systems have been studied computationally by some of the authors [101, 102]. For instance, a coarse-grained model was proposed [101] for exploring structural and thermodynamics aspects of confining polyions in spherical cavities, with different linear charge densities and with counterions of different valences. The free-energy of the confined polyion and counterions were estimated as a function of the sphere radius, from a dilute solution corresponding to the largest sphere. A positive free-energy difference was found for all systems and was associated to the compression. This means that the system containing the polyion with the largest linear charge density and monovalent counterions displays the largest resistance to being compressed. The penalty is thus largest for the polyion with the largest linear charge density and monovalent counterions. The replacement of the monovalent counterions with trivalent ones induced a compactation of the polyion, as a consequence of the stronger electrostatic polyion-counterion attraction. This leads to a nearly full compensation of the ideal and electrostatic contributions to the free-energy of their confinement.

Other interesting example of application relates to protein-ligand interactions [103–107], which are typically pursued by significant conformational and energetic changes. These

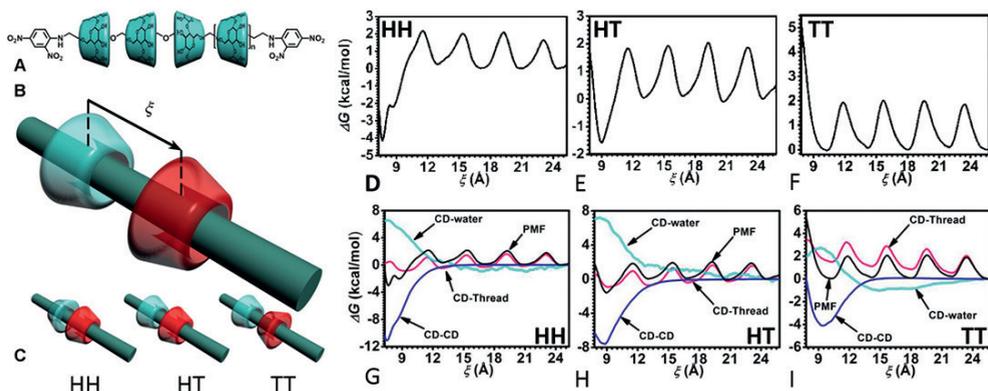


Figure 9. (A) Structural arrangement of a polyrotaxane formed by a poly(ethylene glycol chain included in multiple α -cyclodextrins, (B) definition of the reaction coordinate, ξ , in which both the geometrically restrained (cyan) and free (red) cyclodextrins are presented. (C) Three possible conformations (HH, HT and TT) showing different spatial arrangements of two consecutive α -cyclodextrin molecules. (D-F) Free-energy profiles corresponding to the dimerization of α -cyclodextrins on the poly(ethylene glycol chain, for the HH, HT and TT conformations. (G-I) Individual cyclodextrin-cyclodextrin, cyclodextrin-water, and cyclodextrin-thread energy contributions for the total free-energy, in the three conformations. Reproduced from Ref. [100] with permission from the American Chemical Society.

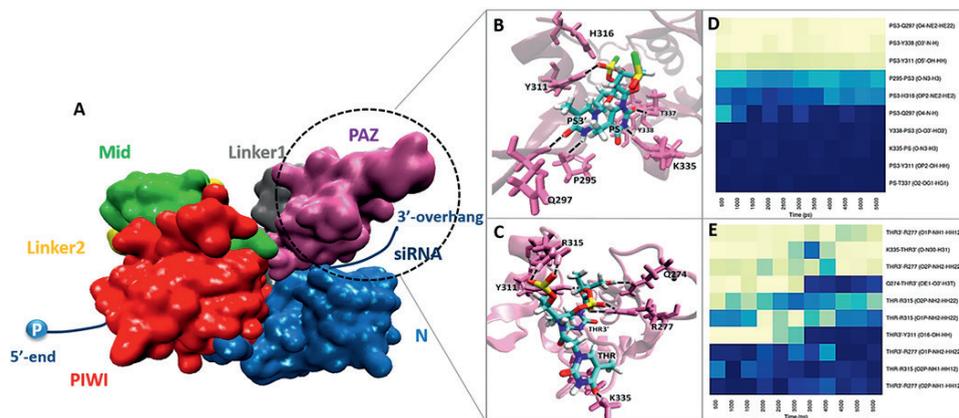


Figure 10. Representative conformation of the complex between PAZ domain of the human argonaute 2 (hArgo2) and the chemically modified siRNAs at 3' overhang, obtained from the last nanoseconds of the MD production runs. Panel A illustrates the relative positioning of domains and interdomain linker of the hArgo2 (using the crystal coordinates of PDB ID: 4F3T) and the positioning of the siRNA antisense strand. Panels B and C show the most representative structures of the complexes formed between PAZ domain and siRNAs, modified with phosphorothioate thymidine (PS) and L-threoninol-thymine (THR), respectively. PS and THR correspond to the second-last residues of the modified siRNAs while PS3' and THR3' correspond to the last one. Hydrogen-bonding interactions between the siRNA nucleotides and the PAZ amino acid residues are represented in black dash lines. The phosphorus, oxygen, nitrogen, hydrogen and sulfur atoms are colored in yellow, red, blue, white, and green respectively. Panels D and E correspond to the occupancies of the most prominent hydrogen bonds formed between the PAZ binding pocket and 2-nt modified nucleotides with PS and THR, respectively. The analysis of the instantaneous hydrogen bond formation was obtained using an in-house algorithm. Adapted from Ref. [108].

changes often occur on time scales that make direct atomic-level simulations useless. One possible alternative relies on the assumption that the change in the binding free-energy resulting, for e.g. from a mutation of a ligand bounded to a protein, complies a linear response scheme [104], with parameters estimated from training sets of protein-ligand complexes, and used for predicting binding affinities of new ligands. Another strategy is the computational design of the ligand in free and bound states. In this type of systems, the emergence of some reliable implicit solvation models and classical statistical-mechanical approaches have been part of the solution. These include, for instance, the use of the molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) model [46]. A theoretical study, recently published [108], encompasses MD and free-energy calculations for evaluating the structural and thermodynamic signatures involved in the interaction of siRNA molecules, bearing chemical modifications at 3'-overhang, with the PAZ domain of human Argonaute (**Figure 10**). In these systems, a reduction of the complex binding affinity has been observed upon siRNA modification, being explained by the hampering of hydrogen bond formation in the active site of PAZ. Analyses of free-energy, achieved from simulations based on the Generalized Born and surface area continuum solvation (MM-GBSA) method, and of hydrogen bonds, have provided a complete identification of the most relevant residues for PAZ-siRNA interaction (see **Figure 10**). This data will contribute to the improved design of synthetic nucleotide analogues, circumventing some of the intrinsic siRNA drawbacks.

6. Concluding remarks

Molecular simulations, including free-energy calculations is still a fertile ground for research, from both theoretical and applied perspectives. Among the various challenges in the field of free-energy calculations, the integration of thermodynamics and kinetics through the concomitant determination of potentials of mean force and diffusivities in biased simulations, will benefit from substantial efforts. Although there is a general consensus that free-energy calculations are an important tool of computational chemistry and structural biology, thriving beyond academic walls, namely in industrial environments, some concerns still remain, dealing with the question to whether or not those methods provide convincing and reliable answers.

Acknowledgements

The authors acknowledge the Fundação para a Ciência e Tecnologia (FCT), Portuguese Agency for Scientific Research, through the Projects n. 016648 POCI-01-0145-FEDER-016648 and COMPETE POCI-01-0145-FEDER-007440. The Coimbra Chemistry Centre is also supported by FCT through the Projects PEst-OE/QUI/UI0313/2014 and POCI-01-0145-FEDER-007630. Tânia F.G.G. Cova, Sandra C. C. Nunes and Andreia. F. Jorge acknowledge, respectively, the PhD and post-doctoral research Grants SFRH/BD/95459/2013, SFRH/BPD/71683/2010 and SFRH/BPD/104544/2014, assigned by FCT.

Author details

Tânia F. Cova, Sandra C. Nunes, Bruce F. Milne, Andreia F. Jorge and Alberto C. Pais*

*Address all correspondence to: pais@qui.uc.pt

Department of Chemistry, Faculty of Science and Technology, Coimbra Chemistry Centre, CQC, University of Coimbra, Portugal

References

- [1] Abel R, Wang L, Harder ED, Berne BJ, Friesner RA. Advancing drug discovery through enhanced free energy calculations. *Accounts of Chemical Research*. 2017;**50**:1625-1632
- [2] Mobley DL, Gilson MK. Predicting binding free energies: Frontiers and benchmarks. *Annual Review of Biophysics*. 2017;**46**:531-558
- [3] Ganesan A, Coote ML, Barakat K. Molecular dynamics-driven drug discovery: Leaping forward with confidence. *Drug Discovery Today*. 2017;**22**:249-269
- [4] Qianqian Z, Weixiang Z, Runmiao W, Yitao W, Defang O. Research advances in molecular modeling in cyclodextrins. *Current Pharmaceutical Design*. 2017;**23**:522-531
- [5] Cova TFGG, Nunes SCC, Pais AACC. Free-energy patterns in inclusion complexes: The relevance of non-included moieties in the stability constants. *Physical Chemistry Chemical Physics*. 2017;**19**:5209-5221
- [6] Schneider H-J. Binding mechanisms in supramolecular complexes. *Angewandte Chemie International Edition*. 2009;**48**:3924-3977
- [7] Biedermann F, Nau WM, Schneider H-J. The hydrophobic effect revisited—Studies with supramolecular complexes imply high-energy water as a noncovalent driving force. *Angewandte Chemie International Edition*. 2014;**53**:11158-11171
- [8] Cova TFGG, Nunes SCC, Valente AJM, Pinho e Melo TMVD, Pais AACC. Properties and patterns in anion-receptors: A closer look at bambusurils. *Journal of Molecular Liquids*. 2017;**242**:640-652
- [9] Abdolmaleki A, Ghasemi F, Ghasemi JB. Computer-aided drug design to explore cyclodextrin therapeutics and biomedical applications. *Chemical Biology & Drug Design*. 2017;**89**:257-268
- [10] Schmidt BVKJ, Barner-Kowollik C. Dynamic macromolecular material design—The versatility of cyclodextrin-based host–guest chemistry. *Angewandte Chemie International Edition*. 2017;**56**:8350-8369
- [11] Ryzhakov A, Do Thi T, Stappaerts J, Bertoletti L, Kimpe K, Sá Couto AR, Saokham P, Van den Mooter G, Augustijns P, Somsen GW, Kurkov S, Inghelbrecht S, Arien A, Jimidar MI, Schrijnemakers K, Loftsson T. Self-assembly of cyclodextrins and their complexes in aqueous solutions. *Journal of Pharmaceutical Sciences*. 2016;**105**:2556-2569

- [12] Loh XJ. Supramolecular host-guest polymeric materials for biomedical applications. *Materials Horizons*. 2014;**1**:185-195
- [13] Zhang H, Tan T, Hetényi C, van der Spoel D. Quantification of solvent contribution to the stability of noncovalent complexes. *Journal of Chemical Theory and Computation*. 2013;**9**:4542-4551
- [14] Filippini G, Bonal C, Malfreyt P. How does the dehydration change the host-guest association under homogeneous and heterogeneous conditions? *Physical Chemistry Chemical Physics*. 2014;**16**:8667-8674
- [15] Zhang H, Tan T, Hetényi C, Lv Y, van der Spoel D. Cooperative binding of cyclodextrin dimers to isoflavone analogues elucidated by free energy calculations. *The Journal of Physical Chemistry C*. 2014;**118**:7163-7173
- [16] Chipot C, Pohorille A. *Free Energy Calculations*. Berlin, Heidelberg, New York: Springer; 2007
- [17] Robert A, Lingle W, David LM, Richard AF. A critical review of validation, blind testing, and real- world use of alchemical protein-ligand binding free energy calculations. *Current Topics in Medicinal Chemistry*. 2017;**17**:2577-2585
- [18] Chipot C. *Frontiers in free-energy calculations of biological systems*. Wiley Interdisciplinary Reviews: Computational Molecular Science. 2014;**4**:71-89
- [19] Kirkwood JG. Statistical mechanics of fluid mixtures. *The Journal of Chemical Physics*. 1935;**3**:300-313
- [20] Zwanzig RW. High-temperature equation of state by a perturbation method. I. Nonpolar gases. *The Journal of Chemical Physics*. 1954;**22**:1420-1426
- [21] Widom B. Some topics in the theory of fluids. *The Journal of Chemical Physics*. 1963;**39**:2808-2812
- [22] Torrie GM, Valleau JP. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*. 1977;**23**:187-199
- [23] Bennett CH. Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics*. 1976;**22**:245-268
- [24] Henriksen NM, Gilson MK. Evaluating force field performance in thermodynamic calculations of cyclodextrin host-guest binding: Water models, partial charges, and host force field parameters. *Journal of Chemical Theory and Computation*. 2017;**13**:4253-4269
- [25] Malhis LD, Bodoor K, Assaf KI, Al-Sakhen NA, El-Barghouthi MI. Molecular dynamics simulation of a cucurbituril based molecular switch triggered by pH changes. *Computational and Theoretical Chemistry*. 2015;**1066**:104-112
- [26] Cova TFGG, Nunes SCC, Pinho e Melo TMVD, Pais AACC. Bambusurils as effective ion caging agents: Does desolvation guide conformation? *Chemical Physics Letters*. 2017;**672**:89-96

- [27] Cao R, Wu S. In silico properties characterization of water-soluble γ -cyclodextrin bi-capped C60 complex: Free energy and geometrical insights for stability and solubility. *Carbohydrate Polymers*. 2015;**124**:188-195
- [28] Cai W, Sun T, Liu P, Chipot C, Shao X. Inclusion mechanism of steroid drugs into β -cyclodextrins. Insights from free energy calculations. *The Journal of Physical Chemistry B*. 2009;**113**:7836-7843
- [29] Sancho MI, Andujar S, Porasso RD, Enriz RD. Theoretical and experimental study of inclusion complexes of β -cyclodextrins with chalcone and 2',4'-dihydroxychalcone. *The Journal of Physical Chemistry B*. 2016;**120**:3000-3011
- [30] Giovannelli E, Procacci P, Cardini G, Pagliai M, Volkov V, Chelli R. Binding free energies of host-guest systems by nonequilibrium alchemical simulations with constrained dynamics: Theoretical framework. *Journal of Chemical Theory and Computation*. 2017;**13**:5874-5886
- [31] Vijayaraj R, Van Damme S, Bultinck P, Subramanian V. Molecular dynamics and umbrella sampling study of stabilizing factors in cyclic peptide-based nanotubes. *The Journal of Physical Chemistry B*. 2012;**116**:9922-9933
- [32] Roux B. The calculation of the potential of mean force using computer simulations. *Computer Physics Communications*. 1995;**91**:275-282
- [33] Bernardi RC, Melo MCR, Schulten K. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochimica et Biophysica Acta (BBA) - General Subjects*. 2015;**1850**:872-877
- [34] Abrams C, Bussi G. Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration. *Entropy*. 2014;**16**:163
- [35] Cavalli A, Spitaleri A, Saladino G, Gervasio FL. Investigating drug-target association and dissociation mechanisms using metadynamics-based algorithms. *Accounts of Chemical Research*. 2015;**48**:277-285
- [36] Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *The Journal of Chemical Physics*. 2004;**120**:11919-11929
- [37] Faradjian AK, Elber R. Computing time scales from reaction coordinates by milestoning. *The Journal of Chemical Physics*. 2004;**120**:10880-10889
- [38] Bolhuis PG, Chandler D, Dellago C, Geissler PL. Transition path sampling: Throwing ropes over Rough Mountain passes, in the dark. *Annual Review of Physical Chemistry*. 2002;**53**:291-318
- [39] Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, DeBolt S, Ferguson D, Seibel G, Kollman P. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications*. 1995;**91**:1-41

- [40] Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015;**1-2**:19-25
- [41] Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*. 2005;**26**:1781-1802
- [42] Comer J, Gumbart JC, Hénin J, Lelièvre T, Pohorille A, Chipot C. The adaptive biasing force method: Everything you always wanted to know but were afraid to ask. *The Journal of Physical Chemistry B*. 2015;**119**:1129-1151
- [43] Zhao T, Fu H, Lelièvre T, Shao X, Chipot C, Cai W. The extended generalized adaptive biasing force algorithm for multidimensional free-energy calculations. *Journal of Chemical Theory and Computation*. 2017;**13**:1566-1576
- [44] Straatsma TP, Berendsen HJC. Free energy of ionic hydration: Analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations. *The Journal of Chemical Physics*. 1988;**89**:5876-5886
- [45] Martins SA, Sousa SF, Ramos MJ, Fernandes PA. Prediction of solvation free energies with thermodynamic integration using the general amber force field. *Journal of Chemical Theory and Computation*. 2014;**10**:3570-3577
- [46] Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery*. 2015;**10**:449-461
- [47] He J, Chipot C, Shao X, Cai W. Cooperative recruitment of amphotericin B mediated by a cyclodextrin dimer. *The Journal of Physical Chemistry C*. 2014;**118**:24173-24180
- [48] Zhang H, Tan T, Feng W, van der Spoel D. Molecular recognition in different environments: β -cyclodextrin dimer formation in organic solvents. *The Journal of Physical Chemistry B*. 2012;**116**:12684-12693
- [49] Loethen S, Kim JM, Thompson DH. Biomedical applications of cyclodextrin based polyrotaxanes. *Polymer Reviews*. 2007;**47**:383-418
- [50] Yu Y, Cai W, Chipot C, Sun T, Shao X. Spatial arrangement of α -cyclodextrins in a rotaxane. Insights from free-energy calculations. *The Journal of Physical Chemistry B*. 2008;**112**:5268-5271
- [51] Masatake S, Fumio H. Predicting the binding free energy of the inclusion process of 2-hydroxypropyl- β -cyclodextrin and small molecules by means of the MM/3D-RISM method. *Journal of Physics: Condensed Matter*. 2016;**28**:384002
- [52] Kollman P. Free energy calculations: Applications to chemical and biochemical phenomena. *Chemical Reviews*. 1993;**93**:2395-2417
- [53] Ghoufi A, Malfreyt P. Calculation of the absolute thermodynamic properties of association of host-guest systems from the intermolecular potential of mean force. *The Journal of Chemical Physics*. 2006;**125**:224503

- [54] Lamb ML, Jorgensen WL. Computational approaches to molecular recognition. *Current Opinion in Chemical Biology*. 1997;1:449-457
- [55] Cai W, Sun T, Shao X, Chipot C. Can the anomalous aqueous solubility of [small beta]-cyclodextrin be explained by its hydration free energy alone? *Physical Chemistry Chemical Physics*. 2008;10:3236-3243
- [56] Hub JS, de Groot BL, van der Spoel D. g_wham—A free weighted histogram analysis implementation including robust error and autocorrelation estimates. *Journal of Chemical Theory and Computation*. 2010;6:3713-3720
- [57] Brandsdal BO, Österberg F, Almlöf M, Feierberg I, Luzhkov VB, Åqvist J. Free Energy Calculations and Ligand Binding, *Advances in Protein Chemistry*. Uppsala, Sweden: Academic Press; 2003. pp. 123-158
- [58] DiLabio GA, Otero-de-la-Roza A. Noncovalent Interactions in Density Functional Theory, *Reviews in Computational Chemistry*. Hoboken, New Jersey: John Wiley & Sons, Inc; 2016. pp. 1-97
- [59] Simons J. An experimental chemist's guide to ab initio quantum chemistry. *The Journal of Physical Chemistry*. 1991;95:1017-1029
- [60] Robertson MJ, Tirado-Rives J, Jorgensen WL. Improved peptide and protein torsional energetics with the OPLS-AA force field. *Journal of Chemical Theory and Computation*. 2015;11:3499-3509
- [61] ITADREDTJDA Case, Betz RM, Botello-Smith W, Cerutti DS, Cheatham TE, AMBER. 16, University of California; 2016
- [62] Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, Mackerell AD. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry*. 2010;31:671-690
- [63] Reif MM, Hünenberger PH, Oostenbrink C. New interaction parameters for charged amino acid side chains in the GROMOS force field. *Journal of Chemical Theory and Computation*. 2012;8:3705-3723
- [64] Ploetz EA, Smith PE. A Kirkwood-buff force field for the aromatic amino acids. *Physical Chemistry Chemical Physics*. 2011;13:18154-18167
- [65] Baker CM, Anisimov VM, MacKerell AD. Development of CHARMM polarizable force field for nucleic acid bases based on the classical Drude oscillator model. *The Journal of Physical Chemistry B*. 2011;115:580-596
- [66] Lemkul JA, MacKerell AD. Balancing the interactions of Mg²⁺ in aqueous solution and with nucleic acid moieties for a polarizable force field based on the classical Drude oscillator model. *The Journal of Physical Chemistry B*. 2016;120:11436-11448
- [67] Wang Z-X, Zhang W, Wu C, Lei H, Cieplak P, Duan Y. Strike a balance: Optimization of backbone torsion parameters of AMBER polarizable force field for simulations of proteins and peptides. *Journal of Computational Chemistry*. 2006;27:781-790

- [68] Cole DJ, Vilseck JZ, Tirado-Rives J, Payne MC, Jorgensen WL. Biomolecular force field parameterization via atoms-in-molecule electron density partitioning. *Journal of Chemical Theory and Computation*. 2016;**12**:2312-2323
- [69] Shi Y, Xia Z, Zhang J, Best R, Wu C, Ponder JW, Ren P. Polarizable atomic multipole-based AMOEBA force field for proteins. *Journal of Chemical Theory and Computation*. 2013;**9**:4046-4063
- [70] Duke RE, Starovoytov ON, Piquemal J-P, Cisneros GA. GEM*: A molecular electronic density-based force field for molecular dynamics simulations. *Journal of Chemical Theory and Computation*. 2014;**10**:1361-1365
- [71] Grimme S. A general quantum mechanically derived force field (QMDF) for molecules and condensed phase simulations. *Journal of Chemical Theory and Computation*. 2014;**10**:4497-4514
- [72] Monti S, Corozzi A, Fristrup P, Joshi KL, Shin YK, Oelschlaeger P, van Duin ACT, Barone V. Exploring the conformational and reactive dynamics of biomolecules in solution using an extended version of the glycine reactive force field. *Physical Chemistry Chemical Physics*. 2013;**15**:15062-15077
- [73] Gao J, Truhlar DG, Wang Y, Mazack MJM, Löffler P, Provorse MR, Rehak P. Explicit polarization: A quantum mechanical framework for developing next generation force fields. *Accounts of Chemical Research*. 2014;**47**:2837-2845
- [74] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*. 1983;**79**:926-935
- [75] Horn HW, Swope WC, Pitner JW, Madura JD, Dick TJ, Hura GL, Head-Gordon T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *The Journal of Chemical Physics*. 2004;**120**:9665-9678
- [76] Berendsen HJC, Grigera JR, Straatsma TP. The missing term in effective pair potentials. *The Journal of Physical Chemistry*. 1987;**91**:6269-6271
- [77] Izadi S, Anandakrishnan R, Onufriev AV. Building Water Models: A Different Approach. *The Journal of Physical Chemistry Letters*. 2014;**5**:3863-3871
- [78] Wang L-P, Martinez TJ, Pande VS. Building force fields: An automatic, systematic, and reproducible approach. *The Journal of Physical Chemistry Letters*. 2014;**5**:1885-1891
- [79] Piana S, Donchev AG, Robustelli P, Shaw DE. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *The Journal of Physical Chemistry B*. 2015;**119**:5113-5123
- [80] Wang L-P, Head-Gordon T, Ponder JW, Ren P, Chodera JD, Eastman PK, Martinez TJ, Pande VS. Systematic improvement of a classical molecular model of water. *The Journal of Physical Chemistry B*. 2013;**117**:9956-9972

- [81] Mobley DL, Bayly CI, Cooper MD, Shirts MR, Dill KA. Small molecule hydration free energies in explicit solvent: An extensive test of fixed-charge atomistic simulations. *Journal of Chemical Theory and Computation*. 2009;**5**:350-358
- [82] Galindo-Murillo R, Robertson JC, Zgarbová M, Šponer J, Otyepka M, Jurečka P, Cheatham TE. Assessing the current state of amber force field modifications for DNA. *Journal of Chemical Theory and Computation*. 2016;**12**:4114-4127
- [83] Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of Chemical Theory and Computation*. 2015;**11**:3696-3713
- [84] Wickstrom L, He P, Gallicchio E, Levy RM. Large scale affinity calculations of cyclodextrin host-guest complexes: Understanding the role of reorganization in the molecular recognition process. *Journal of Chemical Theory and Computation*. 2013;**9**:3136-3150
- [85] Zhang H, Yin C, Yan H, van der Spoel D. Evaluation of generalized born models for large scale affinity prediction of cyclodextrin host-guest complexes. *Journal of Chemical Information and Modeling*. 2016;**56**:2080-2092
- [86] Zhou H-X, Gilson MK. Theory of free energy and entropy in noncovalent binding. *Chemical Reviews*. 2009;**109**:4092-4107
- [87] Yang T, Wu JC, Yan C, Wang Y, Luo R, Gonzales MB, Dalby KN, Ren P. Virtual screening using molecular simulations. *Proteins*. 2011;**79**:1940-1951
- [88] Suarez D, Diaz N. Conformational and entropy analyses of extended molecular dynamics simulations of [small alpha]-, [small beta]- and [gamma]-cyclodextrins and of the [small beta]-cyclodextrin/nabumetone complex. *Physical Chemistry Chemical Physics*. 2017;**19**:1431-1440
- [89] De Jong DH, Schäfer LV, De Vries AH, Marrink SJ, Berendsen HJC, Grubmüller H. Determining equilibrium constants for dimerization reactions from molecular dynamics simulations. *Journal of Computational Chemistry*. 2011;**32**:1919-1928
- [90] Lefebvre C, Rubez G, Khartabil H, Boisson J-C, Contreras-Garcia J, Henon E. Accurately extracting the signature of intermolecular interactions present in the NCI plot of the reduced density gradient versus electron density. *Physical Chemistry Chemical Physics*. 2017;**19**:17928-17936
- [91] Contreras-García J, Johnson ER, Keinan S, Chaudret R, Piquemal J-P, Beratan DN, Yang W. NCIPLOT: A program for plotting noncovalent interaction regions. *Journal of Chemical Theory and Computation*. 2011;**7**:625-632
- [92] Bader RFW. *Atoms in Molecules: A Quantum Theory*. Oxford: Clarendon Press; 1994
- [93] Becke AD, Edgecombe KE. A simple measure of electron localization in atomic and molecular systems. *The Journal of Chemical Physics*. 1990;**92**:5397-5403
- [94] Silvi B, Savin A. Classification of chemical bonds based on topological analysis of electron localization functions. *Nature*. 1994;**371**:683

- [95] Burnus T, Marques MAL, Gross EKV. Time-dependent electron localization function. *Physical Review A*. 2005;**71**:010501
- [96] Johnson ER, Keinan S, Mori-Sánchez P, Contreras-García J, Cohen AJ, Yang W. Revealing noncovalent interactions. *Journal of the American Chemical Society*. 2010;**132**:6498-6506
- [97] Brito RO, Silva SG, Fernandes RMF, Marques EF, Enrique-Borges J, do Vale MLC. Enhanced interfacial properties of novel amino acid-derived surfactants: Effects of headgroup chemistry and of alkyl chain length and unsaturation. *Colloids and Surfaces B: Biointerfaces*. 2011;**86**:65-70
- [98] Filippini G, Goujon F, Bonal C, Malfreyt P. Host-guest complexation in the ferrocenyl alkanethiols-Thio β -cyclodextrin mixed self-assembled monolayers. *The Journal of Physical Chemistry C*. 2014;**118**:3102-3109
- [99] He J, Chipot C, Shao X, Cai W. Cyclodextrin-mediated recruitment and delivery of amphotericin B. *The Journal of Physical Chemistry C*. 2013;**117**:11750-11756
- [100] Liu P, Chipot C, Shao X, Cai W. How do α -cyclodextrins self-organize on a polymer chain? *The Journal of Physical Chemistry C*. 2012;**116**:17913-17918
- [101] Pais AACC, Miguel MG, Linse P, Lindman B. Polyelectrolytes confined to spherical cavities. *The Journal of Chemical Physics*. 2002;**117**:1385-1394
- [102] Dias R, Lindman B. DNA Interactions with Polymers and Surfactants. Hoboken, New Jersey: John Wiley & Sons; 2008
- [103] Vettoretti G, Moroni E, Sattin S, Tao J, Agard DA, Bernardi A, Colombo G. Molecular dynamics simulations reveal the mechanisms of allosteric activation of Hsp90 by designed ligands. *Scientific Reports*. 2016;**6**:23830
- [104] Khandelwal A, Balaz S. Improved estimation of ligand-macromolecule binding affinities by linear response approach using a combination of multi-mode MD simulation and QM/MM methods. *Journal of Computer-Aided Molecular Design*. 2007;**21**:131-137
- [105] Abriata LA, Dal Peraro M. Assessing the potential of atomistic molecular dynamics simulations to probe reversible protein-protein recognition and binding. *Scientific Reports*. 2015;**5**:10549
- [106] Childers MC, Daggett V. Insights from molecular dynamics simulations for computational protein design. *Molecular Systems Design & Engineering*. 2017;**2**:9-33
- [107] Kuhn B, Tichý M, Wang L, Robinson S, Martin RE, Kuglstatter A, Benz J, Giroud M, Schirmeister T, Abel R, Diederich F, Hert J. Prospective evaluation of free energy calculations for the prioritization of Cathepsin L inhibitors. *Journal of Medicinal Chemistry*. 2017;**60**:2485-2497
- [108] Alagia A, Jorge AF, Avino A, Cova TF, Crehuet R, Grijalvo S, Pais AC, Eritja R. Exploring PAZ/3'-overhang interaction to improve siRNA specificity. A combined experimental and modeling study. *Chemical Science*. 2018;**9**:2074-2086

