Themis: a software to assess association free energies via direct estimative of partition functions

Felippe Mariano Colombari,[†] Asdrubal Lozada-Blanco,[‡] Kalil Bernardino,[¶]

Weverson Rodrigues Gomes,[‡] and André Farias de Moura^{*,‡}

[†]*Brazilian Biorenewables National Laboratory, Brazilian Center for Research in Energy and Materials, Campinas, SP, Brazil*

[‡]*Department of Chemistry, Federal University of São Carlos, São Carlos, SP, Brazil*

[¶]*Institute of Chemistry, University of São Paulo, São Paulo, SP, Brazil*

E-mail: moura@ufscar.br

Abstract

We present the program *Themis* – a computer implementation of a standard statistical mechanics framework to compute free energies, average energies and entropic contributions for association processes of two atom-based structures. The partition functions are computed analytically using a discrete grid in the phase space, whose size and degree of coarseness can be controlled to allow efficient calculations and to achieve the desired level of accuracy. With this strategy, applications ranging from molecular recognition, chiral discrimination, surface adsorption and even the interactions involving molecules in electronic excited states can be handled.

Introduction

Noncovalent interactions play a central role in diverse processes, both naturally occurring and artificially designed. For instance, molecular binding is of primary importance in all biological systems, being responsible for molecular recognition between specific molecules in different signaling pathways involved in metabolic and regulatory processes of living cells.¹⁻⁴ Intermolecular interactions also plays a fundamental role in chemical processes, whether single atoms or molecules are involved, as in detection and separation techniques, or supramolecular structures, as in self-assembling and self-organization processes.^{1,2,5} The ever increasing advances of nanotechnology and other applied fields have also boosted the interest in understanding and tuning intermolecular interactions to improve the efficiency of functional materials.⁶ In all these fields of application as well as in many others which we cannot discuss here, the understanding of specific interactions at the atomic level is mandatory to properly tailor the macroscopic properties of pure substances and more complex materials, and the connection between both levels lies in the realm of statistical thermodynamics.²

In principle, performing the ensembles averages over the position and momentum degrees of freedom should lead straightforwardly to the properties of interest, especially the free energy differences, which are the driving forces for processes as varied as solvation, molecular recognition, chiral discrimination, molecular binding and conformational changes, among others.^{7,8} But in practice both the computation and the measurement of these thermodynamic potentials remain as challenging tasks because they usually have a large entropic contribution, whose calculation and measurement would both require the sampling of the whole phase space of the system. While longer simulations and slower experiments might improve such sampling, there is a considerable fraction of the space phase with energies lying significantly higher above the global and local minima, either because repulsive contacts dominate at too short distances or because of vanishingly small attractive interactions at larger distances. These higher-energy, low-probability regions are usually not thermally accessible and thus they are not sufficiently sampled, affecting the accurate estimation of any entropic potential.^{7,8}

The intrinsic difficulties to obtain reliable free energy estimates are certainly among the reasons why most of the simulations done to date have relied on energy and enthalpy calculations to discuss

the driving forces behind chemical and biological processes, but neglecting of entropy can be misleading due to energy-entropy compensations. When compared to energy and enthalpy, free energy profiles can present their minima *i*) located at farther distances; *ii*) with more shallow well depths or *iii*) completely canceled out by entropic penalties.

Over the last few decades, several methods have been devised to improve the ability of simulation techniques to sample less probable regions of phase space, with variable degree of success depending on the type of problems being addressed. In a nutshell, these biased methods are designed to sample the phase space along paths in which the relevant transformations take place. To name a few of the most common techniques, the thermodynamic perturbation theory (TPT)⁹ and the thermodynamic integration (TI)¹⁰ perform the sampling between two thermodynamic states with different chemical composition, yielding a free energy corresponding to an alchemical transformation of one system into another. The potential of mean force (PMF), which can be computed by a number of different methods,^{7,11} performs the sampling between two thermodynamic states with different conformation and/or separation distances, corresponding to a free energy to perform a Cartesian transformation of one system into another. While the two former are frequently used to compute solvation free energies, the latter is very useful to compute free energies associated with intramolecular conformational changes and also binding free energies between different species. Metadynamics simulations¹² is probably the most flexible technique regarding the types of different phenomena that can be addressed, using the so-called collective variables to sample regions other than energy minima.

While these methods have been proven useful for many different applications, they still cannot reach states with energies arbitrarily higher than the global energy minimum. This limitation becomes more severe for complex systems having a large number of local free energy minima separated by large barriers, which would require an essentially infinite simulation to guarantee proper sampling, which is obviously unfeasible.⁷

We propose a methodology which is complementary to other computational techniques regarding the possibility of assessing states within arbitrary energy ranges while calculating the free energy of interaction between two structures. In the next sections we describe a computationally efficient implementation of standard statistical thermodynamic formalisms to obtain thermodynamics prop-

erties analytically from the configurational partition function. We named our computer code as *Themis*, the Ancient Greek titaness who was the personification of the natural law. Obviously, we also face the need to narrow the phase space sampling, so we discuss in the Methods section the simplifications that have been made, while in the Results and Discussions section we provide several examples of applications, demonstrating that our approach is not oversimplified and yields accurate results. *Themis* is a free software written in Fortran 2003 language, being available at <https://github.com/colombarifm/themis> under the GPLv3+ License. It runs under Linux environment with gfortran/gcc 5.4+ compilers.

Methodology

Systems of practical interest are beyond any possibility of having their actual partition functions known and exactly solved, so a number of simplifications has to be chosen to render the problems computationally tractable and yet accurate enough to yield useful results. The first step is the choice of a range of problems we want the method to be suitable for, which amounts to choose specific regions of the phase space to be studied in more detail, combined with regions with coarser detail or no detail at all. These choices may be summarized for the present method as follows:

- i*) systems should have constant chemical composition;
- ii*) the target property is the free energy of association of two structural unities (atoms, molecules, nanoparticles, etc.), but other properties can be computed from the knowledge of the probabilities of all the microstates;
- iii*) internal degrees of freedom of the structural unities will not be taken into account explicitly, meaning that all structural unities are rigid and remain at well-defined quantum states throughout the calculation;
- iv*) solvent effects (if any) are treated by means of continuum approaches.

The *Themis* program, as described below, will write down a detailed description of the most stable dimers within an ensemble of structures. Further refinement may be achieved by relaxing the internal degrees of freedom of these lowest-lying structures and/or including explicit solvent either at the molecular mechanics (MM) or at the quantum chemistry (QM) levels. In the current

implementation, one can handle molecules that have multiple relevant conformations by running *Themis* for these different structures and then build a global partition function.

Within the classical statistical thermodynamics formalism, the partition function is a multidimensional integral over positions (q) and momenta (p) of all N particles of the system, as presented in Eq. 1, for the NVT ensemble,

$$Z(N, V, T) = \frac{1}{h^{3N}} \int e^{-\beta H(q,p)} d^3q d^3p \quad (1)$$

where h is the Planck constant, $\beta = \frac{1}{k_B T}$, k_B is the Boltzmann constant and T is the absolute temperature.

Besides the intrinsic difficulty of handling a $6N$ -dimensional phase space integral, there is no analytic form of the Hamiltonian $H(p, q)$ for all phase space from an arbitrarily complex system, rendering the problem intractable without simplifications. The first approximation is the usual assumption that momenta contribution may be factored out from the integral and treated according to the Maxwell-Boltzmann distribution. The remaining configurational integral is still $3N$ -dimensional and as such remains intractable, so we replace the continuous and uncountable number of configurations of the integral by a finite summation over a discrete grid (eq. 2). This equation contains a further necessary approximation, the replacement of the unknown Hamiltonian by the interaction energy E_i calculated by any convenient methodology for each configuration comprising the discretized phase space.

$$Z_{\text{conf}} = \sum_i e^{-\beta E_i} \quad (2)$$

From this point on, the reference structure MOL1 is assumed to be fixed at the origin while the second structure MOL2 samples the accessible space around MOL1. The *Themis* program will read a user-defined grid with the relative distances between MOL1 and MOL2, which we named as the translation grid, consisting of a set of points distributed around MOL1. Examples of grids with different sizes and shapes that have been tested to match the peculiarities of either the system shape or the properties of interest or both include (Figure 1):

- i*) points distributed on a spherical surface;

ii) points distributed along a crystalline plane;¹³

iii) points distributed along a solvent accessible surface.¹⁴

Figure 1: Reference structures (MOL1) used for different types of calculation, depicting the translation grid points (pink spheres). The second structure (MOL2) is placed at each point in turn during the calculation. *i*) spherical grid around water molecule; *ii*) crystalline face grid on the surface of a TiO_2 nanoparticle;¹³ *iii*) solvent-accessible grid around CdTe/cysteine nanoparticle.¹⁴

The configurational ensemble corresponds to dimer conformations obtained by the application of three linear transformations on MOL2 coordinates while keeping MOL1 fixed at the origin. As a first example, we consider the interaction of two water molecules and the sampling will be performing by using user-defined atoms or virtual sites of both MOL1 and MOL2 as reference points for translation ($T_{\text{ref},1}$ and $T_{\text{ref},2}$, respectively) and rotation ($R_{\text{ref},1}$ and $R_{\text{ref},2}$, respectively) moves.

A **translation move** correspond to displace all atoms of the structure equally to a given point in space. In this example (detailed description of input files given in the supporting information) we defined both oxygen atoms as the references for translation moves. While $T_{\text{ref},1}$ is placed at the origin, $T_{\text{ref},2}$ will be placed on every point of the translation grid in turn (Figure 2).

Figure 2: Translation moves performed by MOL2 around MOL1 along a 42 point spherical grid.

Rotation moves, as stated by Euler’s rotation theorem, are rigid body motions in three-dimensional space that preserves at least one fixed point. Although similar to the classification of motions that objects may undergo in mechanics depending on the Euler angles involved, our grids do not depict actual motions and as such we prefer instead to classify these changes in the relative orientation of MOL2 with respect to MOL1 as two types of rotation that, applied consecutively to MOL2, create a set of orientations for each point of the translation grid. In order to build the rotation axis for both molecules ($T_{\text{ref},1} - R_{\text{ref},1}$ and $T_{\text{ref},2} - R_{\text{ref},2}$, respectively), a reference site (for instance, the center of mass in the example of water dimer in Figure 3c) was chosen for MOL1 whereas another one was chosen for MOL2 (one of its hydrogen atoms). Thus, **rot_{point}** is defined as the rotation of the axis around the translation point, in order to sample its orientations around a sphere (Fig. 3) for each point comprising the translation grid shown in Figure 2.

Figure 3: **rot_{point}** moves performed by MOL2 around MOL1 for the first translation point of the spherical grid.

Also, to further explore the configurational space of MOL2 around MOL1, **rot_{axis}** is defined as the rotation of the structure around its rotation own axis in order to samples its orientations around a circle (Fig. 4) for each **rot_{point}** move shown in Figure 3.

After moving $T_{\text{ref},1}$ (the oxygen atom of MOL1) to the origin of Cartesian system, the rotation axis $T_{\text{ref},1} - R_{\text{ref},1}$ will be aligned to z -axis. MOL1 orientation will be kept fixed along the whole procedure and the structural sampling of MOL2 around MOL1 will be performed in a triple nested loop (as presented in Algorithm 1) in order to obtain all possible combinations.

As regards MOL2, $T_{\text{ref},2}$ will be placed on every point of the translation grid in turn and, for each translation point, the relative orientation of MOL2 rotation axis will be sampled by applying

Figure 4: **rot_{axis}** moves performed by MOL2 around MOL1 at a given **rot_{point}** move.

a rotation matrix that keeps $T_{\text{ref},2}$ fixed while $R_{\text{ref},2}$ is translated to each point of a spherical grid centered at such translation point 3. For each one of these rotations of MOL2 rotation axis around the spherical grid, MOL2 will further undergo a number of rotations around its rotation axis (Fig. 4).

Algorithm 1: Energy calculation loop

```

Translate  $T_{\text{ref},1}$  to origin
Orient  $T_{\text{ref},1} - R_{\text{ref},1}$  along  $z$ -axis
do  $t = 1, nt$ 
    Translate  $T_{\text{ref},2}$  to  $t^{\text{th}}$  translation point
    do  $r_1 = 1, nr_1$ 
        Move  $R_{\text{ref},2}$  to  $r_1^{\text{th}}$  rotpoint point
        do  $r_2 = 1, nr_2$ 
            Perform  $r_2^{\text{th}}$  rotaxis rotation around  $T_{\text{ref},2} - R_{\text{ref},2}$  axis
            Calculate  $E(t, r_1, r_2)$ 
        end
    end
    Calculate  $Z_{\text{rot}}(t)$ 
end
Calculate  $Z_{\text{conf}}$ 
Calculate Boltzmann probabilities for all states
Order/search for most probable structures

```

The interaction energy of each microstate of the configurational ensemble is calculated in turn and stored in a three-dimension array for further calculation procedures. The number of microstates ($\mathcal{N} = nt \times nr_1 \times nr_2$) to be considered will depend on how finely/coarsely the phase space will be divided. The degree of coarsening of each type of move over the grid should be optimized by the user to achieve a compromise between accuracy and efficiency (more details presented in the supporting

information).

Once all energy values are calculated, it is straightforward to obtain the discrete configurational partition function for the system (eq. 3),

$$Z_{\text{conf}} = \sum_{t=1}^{nt} \sum_{r_1=1}^{nr_1} \sum_{r_2=1}^{nr_2} \exp(-E(t, r_1, r_2)/k_B T) \quad (3)$$

and using standard statistical mechanics formalism, thermodynamic properties such as free energy (eq. 4), ensemble energy (eq. 5) and entropic contribution (eq. 6) can be obtained:

$$A = -k_B T \ln Z_{\text{conf}} + k_B T \ln \mathcal{N} \quad (4)$$

$$\langle E \rangle = \frac{1}{Z_{\text{conf}}} \sum_i^{\mathcal{N}} E_i \exp(-E_i/k_B T) \quad (5)$$

$$-TS = A - \langle E \rangle \quad (6)$$

where $k_B T \ln \mathcal{N}$ corresponds to the contribution due to ideal gas configurational entropy, since all calculations have the ideal gas reference state defined for infinite intermolecular separation. Furthermore, the Boltzmann probability of each configuration (eq. 7) is also generated from the configurational partition function:

$$P_i = \frac{\exp(-E_i/k_B T)}{Z_{\text{conf}}} \quad (7)$$

While the triple summation described in eq. 3 led to a partition function and associated thermodynamic values for the whole configurational ensemble, such summation can be decomposed into its reorientational component for each translation point (eq. 8)

$$Z_{\text{conf}} = \sum_{t=1}^{nt} \left[\sum_{r_1=1}^{nr_1} \sum_{r_2=1}^{nr_2} \exp(-E(t, r_1, r_2)/k_B T) \right] = \sum_{t=1}^{nt} Z_{\text{rot}}(t) \quad (8)$$

Considering all $nr_1 \times nr_2$ moves at each translation point, thermodynamic functions can also be obtained for each one of them, rendering $A(t)$, $\langle E(t) \rangle$, $-TS(t)$ and $P(t)$ 3D landscapes along the surface of the reference structure (MOL1), which provide a convenient graphical representation of hot spots where the interaction between both structures is maximal as well as regions which behave as kinetic barriers between local minima. These surfaces can be visualized by a set of VMD¹⁵

scripts, written at the end of the run (users may adapt them to other visualization software that can handle surfaces with user defined coloring).

Themis also searches for the most probable structures in the whole ensemble, which can be overlaid on these 3D surfaces to provide insight on the most stable structural patterns. These high-probability structures may be used as input for any higher-level molecular mechanics or quantum chemistry calculations.

Results and discussion

Free energy of association of the water dimer

The water dimer is a very simple and yet extremely important model system whose energetic and structural properties were widely discussed elsewhere from both computational and experimental viewpoint.^{16,17} Therefore, it is a convenient choice for the validation of the methods implemented in *Themis*.

In order to study the association free energy of the water dimer, a potential of mean force (PMF)-like calculation was performed by considering spherical translation grids (Figure 1) around the first water molecule (MOL1). The same grid may be scaled either up or down by multiplication of its Cartesian coordinates by suitable scaling factors (ξ). Considering the oxygen atoms of both MOL1 and MOL2 water molecules as $T_{\text{ref},1}$ and $T_{\text{ref},2}$, ξ will then correspond to the separation distance between them. The association free energy ($\Delta A_{\text{bind}} = A_{\xi_{\text{eq}}} - A_{\xi_{\infty}}$) can be evaluated by sampling the ξ coordinate from a large separation distance (a 15.0 Å intermolecular distance suffices since water molecules are small and at this distance, the interaction energy already vanished) to a close molecular contact (2.0 Å, in which strong repulsive interactions builds up) at 0.1 Å intervals, resulting in 131 $O - O$ separation distances. For each ξ value, a conformational ensemble was generated in an independent run by using a spherical grid containing 42 points for both translation and rot_{point} moves. The rot_{axis} moves around one $O - H$ bond from MOL2 were performed in 3.0 degree intervals, yielding an ensemble containing $\mathcal{N} = 42 \times 42 \times 120 = 211680$ microstates for each $O - O$ separation distance. Intermolecular energy for each configuration was calculated by the

sum of Lennard-Jones and Coulomb contributions obtained with the TIP3P water parameters.¹⁸

For validation purposes, the thermodynamic potentials along the $O - O$ separation coordinate were compared to PMF calculation carried out using the umbrella sampling (US) methodology¹⁹ as implemented in the GROMACS 5.1.2 molecular dynamics package^{20–22} (detailed description presented in the supporting information), as shown in Figure 5.

Figure 5: Thermodynamic properties for water-water interactions as a function of $O - O$ distance at $T = 300$ K obtained from a 500 ns molecular dynamics/US simulation (left) and from *Themis* program (right).

As one can notice, *Themis* calculation has the advantage of delivering all thermodynamic properties straightforwardly while the US technique requires that the same calculation should be done at different temperatures in order to obtain the entropic contribution by finite differences. This additional numerical treatment, besides increasing the computational cost of such calculation, introduces undesired noise to both entropy and energy curves. Although both methods agree qualitatively, thermodynamic functions differ quantitatively at the equilibrium distance (Table 1). Since the partition function calculation can sample phase space more effectively, one observes both free energy and energy values more negative and a slightly more positive value for the entropic term. Another important difference that should be accounted relies in the temperature role in both methods: while *Themis* accounts the exact temperature value to build the Boltzmann factor and therefore, the partition function, MD simulations relies in algorithms to control such quantities. For such small system (6 atoms in vacuum), considerably oscillations were observed, which can directly affect both

the sampling and the energy calculations.

Table 1: Thermodynamic data for water dimer in vacuum at the equilibrium $O-O$ distance obtained for both umbrella sampling methods and the *Themis* program.

Calculation method	ξ_{eq} (Å)	A (kJ/mol)	E (kJ/mol)	-TS (kJ/mol)
umbrella sampling	2.8	-14.3	-20.6	6.3
<i>Themis</i>	2.8	-16.0	-23.7	7.7

The direct sampling of the phase space yielded smoother curves along the separation coordinate at a reduced computational cost: the full thermodynamic profile ($211680 \times 131 \approx 2.8 \times 10^7$ microstates) took less than 5 minutes of wall-time. On the other hand, although different MD simulation lengths – ranging from 100 ns to 500 ns – rendered almost the same free energy profiles, resulting ΔE_{bind} and $-T\Delta S_{\text{bind}}$ profiles were very noisy, indicating that a better sampling was required (Fig. S1). Thus, even for this simple system, considerably long simulations were required to appropriately deliver converged thermodynamic profiles, which took more than 5 days of wall-time to complete. It is important to highlight that although the biased umbrella potential guarantees a proper sampling along ξ , it still fails to enforce the sampling of the whole spherical shell for all combinations of the polar angle θ and the azimuthal angle ϕ around the reference structure at a given separation distance, resulting, for example, in the smaller entropy at the PMF minimum. .

After sampling the whole ensemble, Cartesian coordinates of the most stable structures are written along with their probabilities. Here, as expected, the lowest-lying and most probable structure found at the free energy minimum was a hydrogen-bonded water dimer (Figure 6). As a matter of fact, at ξ_{eq} , *Themis* reported more than 100 structures with interaction energies lying within 1.0 kJ/mol from the most stable structure intermolecular energy (-28.4 kJ/mol).

Further, thermodynamic potentials mapped onto the 2.8 Å sphere around MOL1 (Figure 7) presented well-defined regions where the water-water interaction was more favorable, corresponding to MOL1 being either an H-bond donor or an H-bond acceptor.

Although there are some obvious limitations in *Themis* approach, it is noteworthy to highlight that, for particular cases – such as rigid molecules in vacuum – the methodology implemented in

Figure 6: Structure of the lowest interaction energy dimer taken from the $O - O$ separation distance corresponding to the free energy minimum.

Figure 7: Free energy, energy and entropic penalty representation around the spherical grid corresponding to the global minima ($\xi = 2.8 \text{ \AA}$) of water potential of mean force. Red corresponds to lower values and blue corresponds to higher values. Top row corresponds to MOL1 frontal view; Bottom row corresponds to MOL1 lateral view.

Themis correspond to a fair approximation. Thus, a better convergence at a lower computational cost should be expected as compared to US.

Free energies including quantum mechanical degrees of freedom

Still during the development of the *Themis* code we were faced with an actual problem which could not be handled by classical forcefields and would require energies to be computed by quantum chemical methods, so we have adapted the code to have an interface to external quantum chem-

istry codes and a **RERUN** functionality, that would parse the output files of the individual quantum chemical calculations to get the energies and then compute the partition function and the thermodynamic properties of the systems. The application of that development version of the *Themis* program proved effective to sample the surface of a CdTe nanoparticle functionalized with cysteine molecules (Figure 1) and find the preferential binding sites of the four DNA nucleobases.¹⁴ Over 2 millions structures comprised the phase space and their energies were computed using the PM7 Hamiltonian²³ as implemented in the MOPAC software.²⁴

More recently, it was shown that free energy surfaces of first coordination shell of ion pairs of ionic liquids obtained at DFT level produces results in excellent agreement with the ones obtained from liquid phase MD simulations,^{25,26} demonstrating that even in some complex cases, *Themis* is able to probe the structure of condensed phases. This enables the comparison of forcefields with higher theory methods such as DFT, which are too costly to be applied even for small clusters. It was also possible to study the interaction of carbon nitride sheets in order to build a multilayer structure and its further interaction with different cations²⁷ to build a single-cation catalyst using quantum chemistry energies from the GFN1-xTB Hamiltonian.²⁸ These very successful applications of our method within a quantum chemical potential energy surface might seem enough to prove the effectiveness of our proposal, but we will now present the proof of concept for a more complex application which raises a considerable interest in applied problems ranging from chemistry to biology and materials science: the thermodynamics of interactions involving electronic excited states.

Unlike forcefields, which treat the potential energy as a combination of additive terms and thus allow the direct assessment of intermolecular interactions, quantum chemical calculations provide only the total energy of the dimer for the different distances and relative orientations, and so the dimer energies have to be computed for distances large enough to reach a plateau where we define the dimer energy as E_∞ and then compute the intermolecular interaction for all the remaining points of their phase space using Equation 9:

$$E_{\text{inter},i} = E_i - E_\infty \tag{9}$$

where E_i corresponds to the energy of the i th dimer configuration. Energy convergence issues

that could arise for structures with very repulsive contacts between MOL1 and MOL2 may be avoided by the specification of a minimum interatomic cutoff distance below which *Themis* should not generate the input file for the external quantum chemical program. If this option is used, a large positive energy will be ascribed to those structures that had any atomic clashes and their contribution the partition function will be essentially null.

We have computed the PMF-like profile for the association of two biphenyl molecules, using the middle of the bond between the two rings as the reference site for translation. For sake of comparison, we have computed the profile with three different conditions:

- i*) using the classical OPLS-AA forcefield;
- ii*) using the PM7 Hamiltonian for the two biphenyl molecules in the ground state;
- iii*) using the PM7 Hamiltonian with one molecule in the first singlet excited state.

In all cases, the molecules were optimized at the specified level of theory prior to the calculation, which means that in case *iii*) one molecule was optimized at the S_0 state while the other was optimized at the S_1 state (which has the two rings at coplanar orientation). Unlike the water molecules discussed in the previous section, biphenyl molecules are flexible as the rings can rotate with respect to each other around the central single bond, so we have to choose a conformation here and the less arbitrary choice is always the conformation arising from a full geometry optimization at the level of theory of choice. *

First thing to note is the large differences between the classical forcefield and the PM7 results (Figure 8), which demonstrate the flawed description of simple forcefield for $\pi-\pi$ interactions, which are properly captured by the PM7 Hamiltonian both at the ground state and at the first excited state (Figure 9). Regarding the excited state calculation, its minimum free energy was considerably lower than the minimum obtained for the ground state simulation. Interestingly, the free energy decrease of the excited state has a considerable entropic contribution arising from the lower steric hindrance that arises when one of the molecules becomes planar.

Although basis set superposition errors (BSSE) may be large with minimal basis set methods

*Only the lowest energy conformation was considered here because this is a proof of concept of the general aspects of the methodology, but a more accurate result could be obtained by computing the same free energy surfaces for different conformers and combining the results including their internal energy differences in a global partition function.

Figure 8: Free energy, average energy and entropic penalty components for the association of two biphenyl molecules using either the classical OPLS-AA forcefield or the semiempirical PM7 Hamiltonian, which was applied to the ground state dimer and for the $S_0 - S_1$ dimer.

Figure 9: Most stable dimers obtained at free energy minima of each biphenyl-biphenyl interaction. The pink spheres on the bond between the two rings are the virtual sites used as the T_{ref} reference sites during the calculations.

such as PM7,²³ structures are very similar along the reaction coordinate and errors should nearly cancel out when we compare the same region of phase space at two different quantum states. Besides, the error may be mitigated by choosing a quantum chemical method that handles BSSE properly. Also the quality of the methods may be improved as necessary. In either case, a larger computational cost should be expected, but the protocol we demonstrate here should hold as a proof of concept.

Chiral discrimination

Molecular recognition between chiral species is a subtle and very important class of phenomena in biological systems and understanding its working principles at the molecular level is becoming increasingly important in nanotechnology.^{14,29} Since chirality is a property of the shape and not the size, the chiral discrimination phenomena are expected to depend on the way chiral species are arranged around each other and how the spatial configuration changes for different combinations of the enantiomers.^{30,31}

Taking cysteine amino acid as a test case, two types of dimers are expected:

- i*) homochiral (L-L and D-D pairs);
- ii*) heterochiral (L-D and D-L pairs).

Although all the L-L, L-D, D-L and D-D dimers may be formed, they are related by mirror inversion: L-L and D-D pairs are thermodynamically identical in the absence of any external chiral fields, as well as the L-D and D-L enantiomers. So we need to consider only the L-L homochiral and the L-D heterochiral pairs, which are expected to behave as diastereomers and consequently to have different association thermodynamics.

Taking L-cysteine as the reference structure MOL1 in *Themis* calculations and using spherical grids to compute the PMF for either another L-cysteine or a D-cysteine as the second structure MOL2, we obtained a complex thermodynamic profiles for both systems (Figure 10). Calculations were performed using OPLS-AA parameters and an ensemble of $\mathcal{N} = 362 \times 362 \times 120$ configurations for each one of the 167 separation distances from 3.4 Å to 20.0 Å, rendering more than 2.6×10^9 microstates. See detailed discussion in the supporting information.

Experimentally, separation of amino acids into each pure enantiomer usually requires very long chromatographic columns due to the very small differences in the interactions of each enantiomer with the chiral stationary phase.³² Likewise, the difference in the interaction for the L-L and the L-D pairs is also expected to be small throughout the PMF profile. Specifically at the global minima ($\xi = 5.5$ Å) the difference amounted to only $\Delta\Delta A = 0.05$ kJ/mol and at their second minima ($\xi = 6.9$ Å) to only $\Delta\Delta A = 0.08$ kJ/mol, both much lower than $k_B T$ at $T = 300$ K.

The most stable structures for these low-lying minima correspond to hydrogen-bonded pairs, whose energy is large enough to remain stable in spite of the chiral distortions, so only minor energy

Figure 10: Thermodynamic properties for cysteine-cysteine interactions as a function of chiral carbon atoms (C_α) separation at $T = 300$ K. Top left: Helmholtz free energy; Middle left: Energy; Bottom left: Entropic penalty. Inset plots correspond to differences between homochiral and heterochiral curves. Right panels correspond to the most stable structure found at each mentioned separation distance.

changes were observed, as might be expected.

Interestingly, the energy curves had a further shallow minima at $\xi = 3.9$ Å that was not observed in the free energy curves due to energy-entropy compensation effects, arising from the severe spatial restrictions at closer contact. This finding exemplifies the risks of relying on energy scans instead of free energy scans for systems that may form multiple energy minima.

Further, if one calculates the *chirodiastaltic* curves, i.e., the curves that result from the differences between the homochiral profiles and the heterochiral profiles (Equations 10, 11, 12), the preference

for either homochiral or heterochiral dimers as a function of the separation distance become clearer. For instance, ΔA_{chir} curve shows an homochiral preference for values < 0 whereas the opposite behavior is found for values > 0 . Thus, there is a major chiral discrimination in favour of the LL dimer at $\xi = 4.9$ Å.

$$\Delta A_{\text{chir}}(\xi) = A_{\text{LL}}(\xi) - A_{\text{LD}}(\xi) \quad (10)$$

$$\Delta E_{\text{chir}}(\xi) = E_{\text{LL}}(\xi) - E_{\text{LD}}(\xi) \quad (11)$$

$$-T\Delta S_{\text{chir}}(\xi) = -T\Delta S_{\text{LL}}(\xi) + T\Delta S_{\text{LD}}(\xi) \quad (12)$$

Free energy of adsorption

Adsorption phenomena play a central role in materials science and most theoretical investigations still rely solely on energy calculations and, even worse, they usually sample very few positions and relative orientations of the adsorbing molecules on the surfaces.

As a case study we have chosen to model the differential adsorption profiles of citrate and water molecules on the surface of gold nanoparticles (Au NPs), since citrate is a very common stabilizer for these nanostructures and as such the adsorption free energy is expected to be more negative for citrate than for water molecules. Instead of computing a PMF profile as we did previously we have chosen a more complex scenario: the adsorption of either citrate or water on the surface of a dimer of Au NPs in order to evaluate not only the adsorption free energy, but also to determine if the adsorption is thermodynamically more favorable on the surfaces between the two NPs, enabling the adsorbate to interact with both NPs simultaneously, or on the opposite surfaces.

The model system consisted of two spherical Au NPs with FCC crystalline structure and a radius of 12 Å (321 gold atoms in each particle), with a fixed distance of 30 Å between their centers of mass, and either one citrate ion or one water molecule. The interactions of gold atoms were described by a Lennard-Jones potential,³³ along with the citrate parameters from the OPLS-AA forcefield³⁴ and the SPC water model.¹⁸

As regards *Themis* INPUT, the two nanoparticles are considered as a single reference structure MOL1 and the grid for the translation a square placed parallel to the equatorial planes of both NPs, which was divided into equidistant points separated by 0.005 Å. Grid points inside the NPs were excluded resulting in 17678 translational points, along with a spherical grid with $nr_1 = 92$ points for the **rot_{point}** rotation around each point in the translation grid and $nr_2 = 45$ points for the **rot_{axis}** rotation around each one of the **rot_{point}**, yielding over 73 million structures for each molecule. 2D free energy surfaces show favorable adsorption of both citrate and water onto the surface of the Au NPs (Figure 11). In both cases, the adsorption in the region between the two NPs is more favorable than in regions where the molecules can interact with only one NP at a time, but for the citrate anion the midpoint between the particles is unfavorable due to its large size as compared to the separation left between both NPs, while this very spot is where the much smaller water molecule was more stable, because it fits in the narrow separation without incurring in large

entropic penalties as is the case for citrate. Obviously, the separation between the Au NPs was set arbitrarily here to demonstrate yet another capability of the *Themis* program, but the calculations could have been repeated for a range of separations between the Au NPs, allowing a more detailed investigation of adsorption patterns of water and citrate as the NPs move closer together.

Figure 11: Free energy surface for the translation of one water molecule (top) and one citrate molecule (bottom) over a midplane slice between two Au NPs. On the right, the most stable structures found in each case are displayed.

The energy change to bring the citrate from the farthest point on the grid to the most favorable point of the grid (in the region between the two NPs) amounted to -201 kJ/mol while the entropic penalty was only 18 kJ/mol, resulting in a free energy change of -183 kJ/mol. The free energy associated with the same process for the water amounted to only -24 kJ/mol, being the adsorption of the citrate much more favorable than that of water, as expected.

The major limitation of this kind of model is that in the vacuum there is no competition between molecules for an adsorption site, while in solution the free energy should be much smaller since a number of water molecules must be removed from the Au surface in order to make room for the adsorption of the citrate, and *vice-versa*. But we foresee our methodology as a powerful tool to treat the adsorption of molecules in gas phase onto solid surfaces, which is typical of many applications in heterogeneous catalysis.

Conclusions

We have developed the *Themis* program, an efficient implementation of a standard statistical mechanics framework that allows the calculation of the free energy and its components for the interaction between any two atom-based structures. The configurational partition function is analytically computed using a discrete grid of relative positions and orientations between the two structures. Energies entering into the partition function may be computed using the built-in forcefields of the program or may be calculated using any external program that can return total energy values. This possibility of using external programs allows the inclusion of solvent effects, the use of more sophisticated forcefields and even the use of quantum chemical energies. Regarding the latter, we have demonstrated the proof of concept of how the *Themis* program can be used to compute the free energy of association for molecules in electronic excited states, which should have impact in future investigations of photocatalysis. A word of caution is necessary here: unreliable results were obtained using three-body forcefields like the Tersoff parametrization for C, Si and Ge,³⁵ whose cutoff distance is so short that association PMF profiles were poorly captured. We have described a number of applications using the *Themis* program either as a standalone tool or combined with an external quantum chemistry program, but there are other possibilities still under testing that are worth mentioning here: *i*) as a tool for forcefield parametrization using force match procedures for structures in vacuum (the same grid is used for quantum chemical calculations and for the classical forcefield being tuned); and *ii*) the minimum free energy structures at each distance along a PMF can be exported to be used as the starting structures in a more detailed modeling, like an umbrella sampling calculation with explicit solvent, ions, etc. (it is expected some time saving if the structures are already aligned at a reasonable relative orientation).

Acknowledgement

The authors are grateful to the Brazilian funding agencies CAPES, CNPq and FAPESP (processes 2012/15147-4, 2013/07296-2 and 2017/12063-8) for financial support. AFM is grateful to MEC/PET for a fellowship. The authors acknowledge the National Laboratory for Scientific Computing (LNCC/MCTI, Brazil) for providing HPC resources of the SDumont supercomputer

(<http://sdumont.lncc.br>) and the Cloud@UFSCar (<http://portalcloud.ufscar.br/>), which have afforded the necessary computational resources to the research reported within this paper.

Supporting Information Available

The following files are available free of charge.

- manual.pdf: contains detailed information about *Themis* usage, input and output files.

References

- (1) Mihailescu, M.; Gilson, M. K. On the Theory of Noncovalent Binding. *Biophysical Journal* **2004**, *87*, 23–36.
- (2) Zhou, H.-X.; Gilson, M. K. Theory of Free Energy and Entropy in Noncovalent Binding. *Chemical Reviews* **2009**, *109*, 4092–4107.
- (3) Persch, E.; Dumele, O.; Diederich, F. Molecular Recognition in Chemical and Biological Systems. *Angewandte Chemie International Edition* **2015**, *54*, 3290–3327.
- (4) Caro, J. A.; Harpole, K. W.; Kasinath, V.; Lim, J.; Granja, J.; Valentine, K. G.; Sharp, K. A.; Wand, A. J. Entropy in molecular recognition by proteins. *Proceedings of the National Academy of Sciences* **2017**, *114*, 6563–6568.
- (5) Rebek, J. Introduction to the Molecular Recognition and Self-Assembly Special Feature. *Proceedings of the National Academy of Sciences* **2009**, *106*, 10423–10424.
- (6) Ariga, K.; Ito, H.; Hill, J. P.; Tsukube, H. Molecular recognition: from solution science to nano/materials technology. *Chemical Society Reviews* **2012**, *41*, 5800.
- (7) Berendsen, H. J. C. *Simulating the Physical World: Hierarchical Modeling from Quantum Mechanics to Fluid Dynamics*; Cambridge University Press: Cambridge, United Kingdom, 2007.

- (8) Hansen, N.; van Gunsteren, W. F. Practical Aspects of Free-Energy Calculations: A Review. *Journal of Chemical Theory and Computation* **2014**, *10*, 2632–2647.
- (9) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *The Journal of Chemical Physics* **1954**, *22*, 1420–1426.
- (10) Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *The Journal of Chemical Physics* **1935**, *3*, 300–313.
- (11) Roux, B. The calculation of the potential of mean force using computer simulations. *Computer Physics Communications* **1995**, *91*, 275–282.
- (12) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proceedings of the National Academy of Sciences* **2002**, *99*, 12562–12566.
- (13) Santos, U. S. Computational study of the interactions between titanium dioxide nanoparticles involved in the oriented attachment process. Ph.D. thesis, Federal University of Sao Carlos, Sao Carlos, SP, Brazil, 2018.
- (14) Sun, M.; Xu, L.; Qu, A.; Zhao, P.; Hao, T.; Ma, W.; Hao, C.; Wen, X.; Colombari, F. M.; de Moura, A. F.; Kotov, N. A.; Xu, C.; Kuang, H. Site-selective photoinduced cleavage and profiling of DNA by chiral semiconductor nanoparticles. *Nature Chemistry* **2018**, *10*, 821–830.
- (15) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *Journal of Molecular Graphics* **1996**, *14*, 33–38, 27–28.
- (16) Mukhopadhyay, A.; Cole, W. T. S.; Saykally, R. J. The water dimer I: Experimental characterization. *Chemical Physics Letters* **2015**, *633*, 13–26.
- (17) Mukhopadhyay, A.; Xantheas, S. S.; Saykally, R. J. The water dimer II: Theoretical investigations. *Chemical Physics Letters* **2018**, *700*, 163–175.
- (18) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, *79*, 926–935.

- (19) Torrie, G.; Valleau, J. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **1977**, *23*, 187–199.
- (20) Berendsen, H.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **1995**, *91*, 43–56.
- (21) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **2005**, *26*, 1701–1718.
- (22) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.
- (23) Stewart, J. J. P. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *Journal of Molecular Modeling* **2012**, *19*, 1–32.
- (24) Stewart, J. J. P. MOPAC2012. Stewart Computational Chemistry: Colorado Springs, CO, USA.
- (25) Bernardino, K.; Lima, T. A.; Ribeiro, M. C. C. Low-Temperature Phase Transitions of the Ionic Liquid 1-Ethyl-3-methylimidazolium Dicyanamide. *The Journal of Physical Chemistry B* **2019**, *123*, 9418–9427.
- (26) Bernardino, K.; Goloviznina, K.; Gomes, M. C.; Pádua, A. A. H.; Ribeiro, M. C. C. Ion pair free energy surface as a probe of ionic liquid structure. *The Journal of Chemical Physics* **2020**, *152*, 014103.
- (27) Colombari, F. M.; da Silva, M. A. R.; Homsí, M.; Souza, B. R. L.; Araujo, M.; Francisco, J. L.; Silva, G. T. S. T.; Silva, I. F.; de Moura, A. F.; Teixeira, I. F. Graphitic Carbon Nitrides as Platforms for Single-Atom Photocatalysis. *Faraday Discussions* **2020**,
- (28) Grimme, S.; Schreiner, P. R. Computational Chemistry: The Fate of Current Methods and Future Challenges. *Angewandte Chemie International Edition* **2017**, *57*, 2–9.

- (29) Ma, W.; Xu, L.; de Moura, A. F.; Wu, X.; Kuang, H.; Xu, C.; Kotov, N. A. Chiral Inorganic Nanostructures. *Chemical Reviews* **2017**, *117*, 8041–8093.
- (30) Davankov, V. A. The nature of chiral recognition: Is it a three-point interaction? *Chirality* **1997**, *9*, 99–102.
- (31) Colombari, F. M.; de Moura, A. F.; Freitas, L. C. G. Chiral recognition of liquid phase dimers from gamma-valerolactone racemic mixture. *Journal of Molecular Modeling* **2018**, *24*, 215.
- (32) Deáková, Z.; Ďuračková, Z.; Armstrong, D. W.; Lehotay, J. Separation of Enantiomers of Selected Sulfur-Containing Amino Acids by Using Serially Coupled Achiral-Chiral Columns. *Journal of Liquid Chromatography & Related Technologies* **2015**, *38*, 789–794.
- (33) Heinz, H.; Vaia, R. A.; Farmer, B. L.; Naik, R. R. Accurate Simulation of Surfaces and Interfaces of Face-Centered Cubic Metals Using 12-6 and 9-6 Lennard-Jones Potentials. *The Journal of Physical Chemistry C* **2008**, *112*, 17281–17290.
- (34) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society* **1996**, *118*, 11225–11236.
- (35) Tersoff, J. Modeling solid-state chemistry: Interatomic potentials for multicomponent systems. *Phys. Rev. B* **1989**, *39*, 5566–5568.

Graphical TOC Entry