JCTC Journal of Chemical Theory and Computation

Tracing the Entropy along a Reactive Pathway: The Energy As a Generalized Reaction Coordinate

Carine Michel,[†] Alessandro Laio,*,[‡] and Anne Milet*,[†,#]

Département de Chimie Moléculaire, Chimie Théorique, UMR-5250, ICMG FR-2607, CNRS, Université Joseph Fourier Grenoble I, DU BP 53 38041 Grenoble Cedex 09 France, SISSA, ISAS, Via Beirut 2-4, I-34014 Trieste, Italy, and Institut Universitaire de France, France

Abstract: By using metadynamics at a temperature T_0 we reconstruct the free energy $F_{T_0}(E,s)$ as a function of the potential energy E and of a geometrical variable s. We show here that from $F_{T_0}(E,s)$ one can estimate the free energy also *at a different temperature*. This allows tracing the entropy and characterizing the properties of molecular systems at all temperatures by a single simulation. We validate this approach on the water dimer dissociation.

Metadynamics[1] is a novel technique that can be used for computing the free energy barriers and exploring new reactions pathways in a wide range of contexts, from protein folding[2] to mineral phase transitions[3] and organometallic reactivity.[4] It is based on biasing the dynamics with a history-dependent potential $V_G(s,t)$ defined in the space of a set of collective variables (CV).[5] In the limit of a long metadynamics run, the free energy surface can be reconstructed as a function of the CV s: $F(s) \sim -V_G(s,t)$. We show here that the potential energy E can also be used as a generalized coordinate for studying chemical reactions. E has already been used as a CV in metadynamics for reconstructing the density of states of an Ising model.[6] It has also been used as a helpful auxiliary collective variable to better explore the configuration space in some nucleation studies.[7] During a chemical process, the potential energy E varies as the system explores new intermediate/transition states. Thus, it is a relevant collective variable. We propose to use E in combination with ordinary geometrical CVs to conduct a chemical study. The advantages of such an extended set of CVs will be shown:

tracing the entropy along a reactive pathway and extrapolating the thermodynamic quantities at different temperatures.

Using the potential energy E as a CV together with ordinary geometrical variables s in a metadynamics scheme allows reconstructing, at a temperature T_0, the free energy F_{T_0} as a function of E and s simultaneously. Remarkably $F_{T_0}(E,s)$ contains the relevant information for characterizing the thermodynamic properties of the system at *all the temperatures T*, including the probability to observe the reactants/products and the activation entropy. In fact, for a system of potential energy $E(r)$ we have, within the canonical ensemble

$$F_{T_0}(s, E) = -k_B T_0 \ln \int dr \delta(E - E(r))\delta(s - s(r)) \exp\left(-\frac{1}{T_0}E(r)\right)$$
$$= E - k_B T_0 \ln \int dr \delta(E - E(r))\delta(s - s(r))$$
$$= E - k_B T_0 \ln (\Omega(E, s))$$

where, in the second passage, we used the properties of the Dirac delta. The density of states

$$\Omega(E, s) = \int dr \delta(E - E(r))\delta(s - s(r))$$

appearing in the last passage is a measure of the number of configurations r that exist at an energy $E(r)$ and for which $s(r) = s$. By definition this quantity does not depend on T_0. Thus, the free energy at a different temperature T can be directly computed from $F_{T_0}(E,s)$

$$F_T(E, s) = E - Tk_B \ln \Omega(E, s) = E + \frac{T}{T_0}(F_{T_0}(E, s) - E) \quad (1)$$

The internal energy profile $U_{T_0}(s)$ and the free energy profile $F_{T_0}(s)$ are also directly derived from $F_{T_0}(E,s)$

$$U_{T_0}(s) = \frac{\int dE E \exp(-F_{T_0}(E, s)/k_B T_0)}{\int dE \exp(-F_{T_0}(E, s)/k_B T_0)} \quad (2)$$

$$F_{T_0}(s) = -k_B T_0 \log(\int dE \exp(-F_{T_0}(E, s)/k_B T_0)) \quad (3)$$

Finally, the entropy $S_{T_0}(s)$ as a function of the reaction coordinate s is given by $S_{T_0}(s) = (U_{T_0}(s) - F_{T_0}(s))/T_0$. Thus, using eqs 1−3 allows computing the U, F, and S profiles at any temperature T by a run performed at a single temperature T_0.

The entropy obtained in this manner could alternatively be derived computing U as the average of E for fixed s and the free energy F as a function of s alone by thermodynamic integration or one-dimensional metadynamics. However, as it is clear from eqs 2 and 3, internal energy, free energy, and

* Corresponding author e-mail: Anne.milet@ujf-grenoble.fr (A.M.), laio@sissa.it (A.L.).

† Université Joseph Fourier Grenoble-I.

‡ Statistical and Biological Physics, SISSA.

Institut Universitaire de France.

entropy depend implicitly on the temperature T_0. Thus, if one wants to characterize the behavior of the system at several temperatures T, one would be forced to repeat the calculation at all T. One can also obtain the entropy as a function of a geometrical parameter within the harmonic and the rigid rotator approximations,[8] but those approximations break down in solvated systems and at high temperature where anharmonicity effects become important. Instead, our approach allows computing the entropy profile at a minimal computational cost without those approximations. Furthermore, it provides F, U, and S profiles in a wide range of temperatures by a single calculation.

In practical application, $F_{T_0}(E,s)$ is estimated by metadynamics that, by construction, explores only a finite region of the CV space. Thus, even if eq 3 is in principle exact, the larger the difference between T and T_0 is, the less the extrapolation will be meaningful. Indeed, the CV regions that are relevant at T and T_0 might be different. We here perform the extrapolation if two criteria are satisfied:

1. The free energy surface $F_{T_0}(E,s)$ has to be sufficiently explored. For a given (E,s), the accumulation of Gaussians must be larger than $2k_BT_0$ to extrapolate $F_{T_0}(E,s)$ into $F_T(E,s)$.

2. In order to compute the profiles $F_T(s)$ and $U_T(s)$, one has to integrate $F_T(E,s)$ along E according to eq 2 and eq 3. Thus, for a given value of s, the extrapolated free energy surface $F_T(E,s)$ has to be sufficiently explored as a function of E in the region around the minimum of F, as this region dominates the integrals. Denoting by $E_0(s)$ the value of E minimizing $F_T(E,s)$ at fixed s, we require that all the values of E for which $F_T(E,s) - F_T(E_0(s),s) < 2k_BT$ are explored. This is imposed requiring the history-dependent potential to be larger than $2k_BT_0$ in this region.

This procedure has been validated on a simple system, the water dimer treated at the DFT level. At $T = 0$, the only stable state is the bound dimer, while at high T, a dissociated state becomes gradually stabilized by configuration entropy. In order to retain the physical meaning of the total energy of the system, we perform our calculations using the Born–Oppenheimer *ab initio* molecular dynamics approach, using the CP2K-QuickStep program.[11] QuickStep is an implementation of the Gaussian Plane Waves (GPW) method based on the Kohn–Sham formulation of density functional theory (DFT). It is a hybrid method using a linear combination of Gaussian-type orbitals to describe the Kohn–Sham orbitals, whereas an auxiliary plane waves basis set is employed to expand the electronic charge density. The basis set used is a quadruple-valence set of Gaussian orbitals with a set of three polarization functions added for all atoms in conjunction with the Goedecker-Teter-Hutter pseudopotentials. The auxiliary PW basis set was defined by a cubic box of $12A^3$ and by a density cutoff of 500 Ry for the larger grid. The use of the BLYP functional for this system has been validated through comparison with experimental results[9] and high-level calculations[10] The four hydrogen atoms have been replaced by deuterium in order to increase the dynamics step to 0.5 fs. Velocity rescaling has been used to enforce a constant temperature T_0.

The free energy has been reconstructed at $T_0 = 100$ K by metadynamics[11] as a function of the potential energy E and a geometrical collective variable, namely the coordination number between the two oxygen atoms, $s = cn(O,O) = [1-(r/r_0)^3]/[1-$

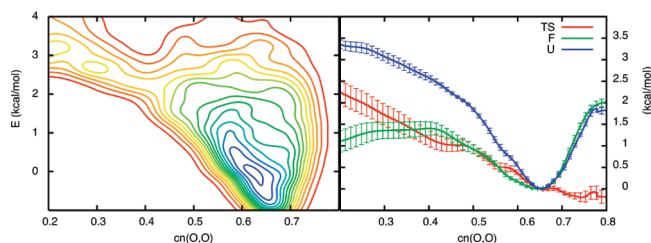**Figure 1.** Left panel: Free energy in function of the potential energy E and $s = cn(O,O)$ at $T_0 = 100$ K reconstructed from a metadynamics run. Right panel: Free energy F, internal energy U, and temperature × entropy T_0S profiles along the reaction coordinate $s = cn(O,O)$, resulting from the mean value of 5 independent metadynamics runs at $T_0 = 100$ K. Error bars correspond to the standard deviation. High values of cn(O,O) correspond to the bound water dimer; low values of cn(O,O) correspond to the dissociated dimer.

$(r/r_0)^6]$ where r is the O–O distance and $r_0 = 3.5$ Å. The shape of the added Gaussians is defined by the height (0.063 kcal/mol ≈ 0.3 k_BT_0), the width along s (0.03), and the width along E (0.188 kcal/mol). The Gaussians are added every 20 fs.

The metadynamics runs have been stopped after 1500 Gaussians have been added: after this time, a diffusive behavior in CV space and several recrossings of the transition state region are observed. The free energy surface $F_{T_0}(E,s)$ is then reconstructed on a 500 × 500 grid in the (0.2 < cn(O, O) < 0.8) × (0 kcal/mol< E < 10 kcal/mol) region. The cn(O, O) limits derive from the size of the box, which restrains the accessible range for the lower region, and on repulsion between nuclei, which forbids the system visiting the upper region. All the energies in the following are expressed as a difference with respect to the global total energy minimum. The two-dimensional free energy $F_{T_0 = 100K}(E,s)$ is shown in Figure 1. Then, following the general procedure detailed above, we extracted from $F_{T_0 = 100K}(E,s)$ the internal energy $U_{T_0 = 100K}(s)$, the free energy $F_{T_0 = 100K}(s)$, and the entropy $S_{T_0 = 100K}(s)$ as a function of the geometrical variable s. The result is shown in Figure 1, right panel. The error bars are estimated repeating the procedure in 5 independent but equivalent simulations.[12] All the free energy profiles fall in a range of 0.5 kcal/mol around the mean. Similarly, all the entropy profiles fall in a range of 5 cal/mol/K × 100 K = 0.5 kcal/mol around the mean. From now, the discussion will be based on the more accurate mean profiles.

The free energy profile $F_{T_0 = 100K}(s)$ and the entropy profile $S_{T_0 = 100K}(s)$ obtained at $T_0 = 100$ K are plotted together with the internal energy profile $U_{T_0 = 100K}(s)$ in Figure 1. Both F and U present a minimum at $s = 0.64$ corresponding to the bound water dimer, with an O–O distance $r = 2.89$ Å. For smaller s, along the dissociation path, those two profiles start becoming different: the free energy profile $F(s)$ presents a maximum around $s = 0.40$, whereas the mean energy profile $U(s)$ increases monotonically. It is well-known that the origin of those different behaviors lies in the entropic effect: the higher the temperature is, the more the dissociated dimer is stabilized by configurational entropy. Indeed, the entropy monotonically increases during the dissociation (see Figure 1): from $s = 0.64$ to $s = 0.2$, the dissociation of the water dimer leads to a net entropy increase of $T_0 \times S = 2.2$ kcal/mol at $T_0 = 100$ K.[13] Thus, the entropic contribution (2.2 kcal/mol)

Letter

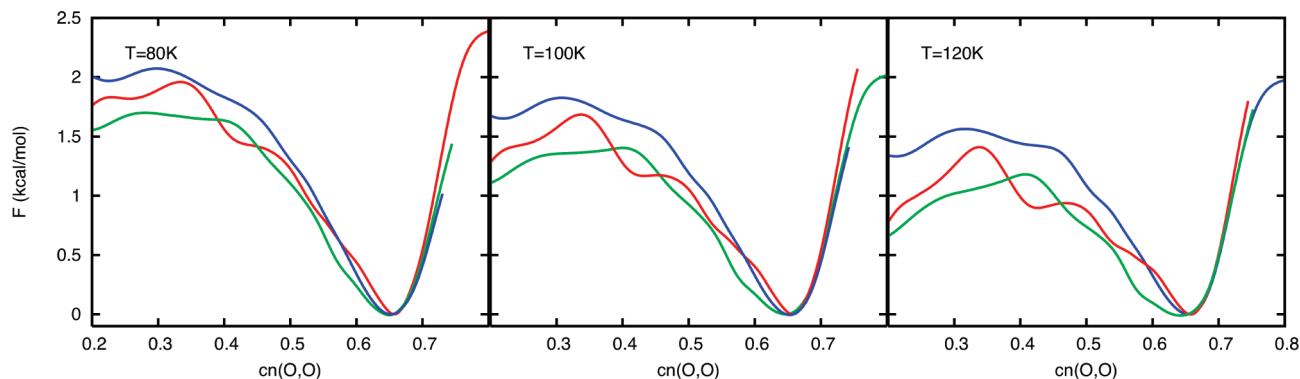*J. Chem. Theory Comput., Vol. 5, No. 9, 2009* **2195**

Figure 2. Free energy profiles extrapolated at various temperatures T: 80 K (left panel), 100 K (middle panel), 120 K (right panel). In red, the free energy profile extrapolated from metadynamics runs performed at $T_0 = 80$ K, in green, the free energy profile extrapolated from metadynamics runs performed at $T_0 = 100$ K, in blue, the free energy profile extrapolated from metadynamics runs performed at $T_0 = 120$ K.
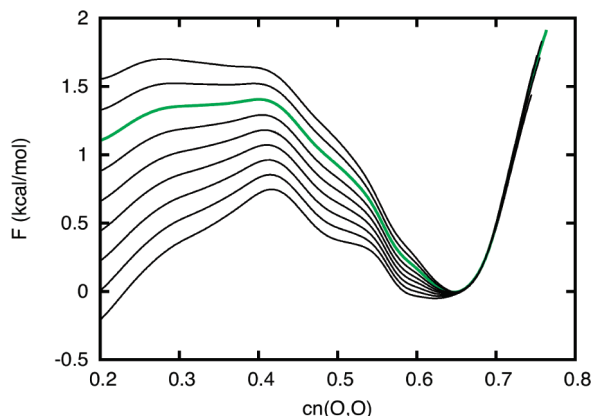

Figure 3. Free energy profiles extrapolated at various temperatures T (from 80 K to 160 K, every 10 K) from the metadynamics runs at $T_0 = 100$ K. In green, the free energy profile at 100 K. Notice the lower profile ($T = 160$ K): the dissociated dimer is more stable than the associated one.

compared with the net energy increase (3.4 kcal/mol) is far from being negligible.

Any extrapolation of a given property at another temperature T based on the entropy profile $S_{T_0}(s)$ has to be taken with caution: this procedure cannot be valid for large changes in temperature, as $S_{T_0}(s)$ depends on T_0. On the other hand, computing the free energy F as a simultaneous function of a geometrical parameter s and the potential energy E allows extrapolating the results over a much broader temperature range. We have performed the same calculations at lower ($T_0 = 80$ K) and higher ($T_0 = 120$ K) temperature. In Figure 2, the free energy profile F_{T_0} at temperature T_0 obtained from a metadynamics at this same temperature T_0 is plotted together with the free energy profiles F_T extrapolated from $F_{T_0'}(E,s)$ obtained at another temperature $T_0' = T_0 \pm \Delta T$ ($\Delta T = 20$ K and/or 40 K). Those free energy profiles perfectly illustrate the stabilization the dissociated dimer gains at higher temperature. Indeed, this behavior is observed for both the original and the extrapolated profiles. At the beginning of the dissociation process (cn(O,O) > 0.40), the agreement between simulated and extrapolated profiles is particularly good. Once the dimer is dissociated (cn(O,O) < 0.40), the difference between simulated and extrapolated profile remains lower than 0.7 kcal/mol. Following this idea, we have extrapolated the free energy at much higher temperature from

the runs at $T_0 = 100$ K: the dissociated dimer is found to be more stable for $T > 150$ K (see Figure 3).

In conclusion, we have here shown that by using the potential energy E as a collective variable in a metadynamics framework, it is possible to address two issues that are considered very challenging in computational chemistry: tracing the entropy along a reactive pathway and extrapolating thermodynamic quantities at different temperatures. If used in combination with other geometrical collective variables s, the potential energy E allows reconstructing the free energy surface $F_{T_0}(E,s)$. From this quantity one can directly obtain the entropy along the reactive pathway. Moreover, using simple thermodynamics identities, from $F_{T_0}(E,s)$ one can estimate the free energy at a different temperature T. This allows extracting from a single simulation the relevant thermodynamic quantities in a wide range of temperatures.

References

(1) Laio, A.; Parrinello, P. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562.

(2) (b) Bussi, G.; Gervasio, F.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2006**, *128*, 13435. (c) Bonomi, M.; Gervasio, F.; Tiana, G.; Provasi, D.; Broglia, R. A.; Parrinello, M. *Biophys. J.* **2007**, *93*, 2813.

(3) (a) Martonak, R.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2003**, *90*, 075503. (b) Organov, A. R.; Martonak, R.; Laio, A.; Raiteri, P.; Parrinello, M. *Nature* **2005**, *438*, 1142.

(4) (a) Stirling, A.; Iannuzzi, M.; Parrinello, M.; Molnar, F.; Bernhart, V.; Luinstra, G. A. *Organometallics* **2005**, *24*, 2533. (b) Michel, C.; Laio, A.; Mohamed, F.; Krack, M.; Parrinello, M.; Milet, A. *Organometallics* **2007**, *26*, 1241.

(5) Any differentiable function of the ionic coordinates can be used as a collective variable, such as distances, angles, etc.

(6) Micheletti, C.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2004**, *92*, 170601.

(7) (a) Donadio, D.; Raiteri, P.; Parrinello, M. *J. Phys. Chem. B* **2005**, *109*, 5421. (b) Trudu, F.; Donadio, D.; Parrinello, M. *Phys. Rev. Lett.* **2006**, *97*, 105701. (c) Quigley, D.; Rodger, P. M. *J. Chem. Phys.* **2008**, *128*, 221101.

(8) McQuarrie, D. A. *Statistical Thermodynamics*; Harper and Row: New York, 1976.

2196 *J. Chem. Theory Comput., Vol. 5, No. 9, 2009*

Letter

(9) Odutola, J. A.; Dyke, T. R. *J. Chem. Phys.* **1980**, *72*, 5062.

(10) Klopper, W.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Duijneveldt, F. B. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2227. Xu, X.; Goddard, W. A., III *J. Phys. Chem. A* **2004**, *108*, 2305.

(11) In order to keep the physical meaning of the total energy of the system, we perform metadynamics based on a Born-Oppenheimer ab initio dynamics using the CP2K program. (VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, *J. Comput. Phys. Commun.* **2005**, *167*, 103. The time step is 0.5 fs. The energy has been computed at the DFT level using the BLYP functional. The basis set used was a quadrupole-valence set of Gaussian orbitals with a set of three polarization functions added for all atoms in conjunction with the Goedecker-Teter-Hutter pseudopotentials. The auxiliary PW basis set was defined by a cubic box of 12 Å^3 and by a density cutoff of 500 Ry for the larger grid.

(12) For each simulation, the initial velocities have been randomly chosen according to the Boltzmann distribution at T_0.

(13) Within the usual approximations framework, a calculation at the BLYP/6-311+g(3df,3pd) level gives an entropy increase of 21 cal/mol/K at 100 K.