

# Highly accurate computation of free energies in complex systems through *horsetail* QM/MM molecular dynamics combined with free-energy perturbation theory

Marilia T. C. Martins-Costa<sup>1</sup> · Manuel F. Ruiz-López<sup>1</sup>

Received: 10 January 2017 / Accepted: 1 March 2017 / Published online: 17 March 2017  
© Springer-Verlag Berlin Heidelberg 2017

**Abstract** We report a method to calculate very accurate free-energy values in complex systems using hybrid QM/MM approaches. The method combines the recently developed *horsetail* sampling technique in molecular dynamics and the *dual-level* technique derived from free-energy perturbation theory. The former is a particular type of multiple molecular dynamics that has been specifically designed for an efficient parallelization of QM/MM simulations. The latter allows estimating free-energy corrections at a high QM/MM level from a reference sampling obtained at a lower QM/MM level. The methodology is illustrated through the study of hydrogen peroxide at the vapor–liquid water interface, a system of considerable atmospheric and environmental relevance. We focus on the calculation of the free-energy profile for the torsional motion of the solute at the CCSD(T)/aug-cc-pVTZ level of theory. It is shown that the equilibrium angle is 87.7°, and that the free-energy barriers for *cisoid* and *transoid* transition states are 4.3 and 1.4 kcal/mol, respectively. These values are significantly different from the gas phase potential energy surface at the equivalent computational level (112.8°, 7.3 and 1.1 kcal/mol, respectively) suggesting that adsorption of hydrogen peroxide on water droplets could have major implications on its atmospheric chemistry.

**Keywords** Molecular dynamics simulations · QM/MM · Perturbation theory · Free energy · Hydrogen peroxide

## 1 Introduction

In recent years, the combined quantum mechanics/molecular mechanics (QM/MM) approach has become a standard tool in theoretical chemistry. First developed by Warshel and Levitt [1] and Karplus et al. [2] at the semiempirical level to study chemical processes of biological relevance, the method was rapidly extended to *ab initio* and density functional methods [3–6]. It opened an avenue to carry out molecular dynamics (MD) simulations of chemical events in solution and to improve our understanding of reaction mechanisms and activation to the transition state [7, 8].

In QM/MM MD methods, only a small part of the whole system is treated quantum mechanically and this represents a main advantage in terms of computational cost with respect to other methods based on full *ab initio* treatments. Nevertheless, most applications usually require long to very long CPU times. The bottleneck in such simulations is the need to get the wave function, the energy and the energy derivatives at each time step, which in the Born–Oppenheimer approximation requires the diagonalization of the Fock or Kohn–Sham matrices, making the whole procedure poorly parallelizable. In other words, not only the CPU time but also the wall-clock time in QM/MM MD simulations may rapidly become exceedingly long. For this reason, QM/MM simulations suffer in general from two main shortcomings: the use of low-level QM methods (typically semiempirical methods, or at best DFT-based methods with small basis sets), and a limited statistical sampling (typically a few tens or a few hundreds of picoseconds, depending on QM level and system size). The

---

Published as part of the special collection of articles derived from the 10th Congress on Electronic Structure: Principles and Applications (ESPA-2016).

---

✉ Manuel F. Ruiz-López  
Manuel.Ruiz@univ-lorraine.fr

<sup>1</sup> SRSMC, CNRS, University of Lorraine, BP 70239, 54506 Vandoeuvre-lès-Nancy, France

results are seldom accurate, therefore, and the situation is even worst when free energies are considered, as standard algorithms like umbrella sampling [9] are very computationally demanding.

In this paper, we describe a methodology to improve the accuracy of free-energy calculations in combined QM/MM MD simulations. It is based on the combination of two recently developed algorithms in our group. On the one hand, the *horsetail* sampling algorithm [10], which has allowed us for the first time to reach the nanosecond time scale in ab initio QM/MM simulations through the implementation of an efficient parallelization procedure. And on the other hand, the *dual-level* technique [11, 12], in which free-energy perturbation theory is used to estimate high QM/MM-level corrections to free energy obtained at a lower QM/MM level. The methodology is outlined in the next section, and it is then illustrated through the study of hydrogen peroxide dynamics at the vapor–liquid water interface. Specifically, we aim at determining the torsional free-energy surface of  $\text{H}_2\text{O}_2$  in this aqueous environment at the CCSD(T)/aug-cc-pVTZ level.

## 2 Accurate QM/MM calculations of thermodynamic properties

### 2.1 Horsetail sampling

The *horsetail* molecular dynamics (HMD) sampling method [10] is a particular version of the multiple molecular dynamics approach (MMD) [13]. As in MMD, in HMD multiple short trajectories are carried out in parallel seeking to obtain long-time dynamics behavior. The main characteristic of HMD is its multibranching structure, similar to that found in a horsetail. Along a main MD trajectory called the stem, many branched trajectories are launched at regular time intervals. The branching trajectories are started at selected stem configurations (the nodes) after redefining randomly the atom velocities from a Maxwell–Boltzmann distribution. This strategy is related to the rare event approach developed by Anderson [14, 15]. Details on the equilibrium fulfillment in such kind of simulations can be found elsewhere [16]. In order to achieve the highest parallel efficiency, the internode separation in the stem and the branching trajectories length should be equal. In this way, the segment in the stem separating two nodes and a whole set of branching trajectories are computed in parallel in a multi-core run, and the calculation proceeds node by node. Note that in HMD (in contrast to standard MMD) only one configuration is required to restart the calculation. In our original paper [10], this technique was applied to study the structure of hydrogen peroxide at the water liquid–vapor interface. The total simulation time was slightly more than

6 ns representing 5.1 years of CPU time but only 20 days of wall-clock time [10].

### 2.2 Free-energy perturbation theory

Let us now consider how the quality of the QM calculation can be increased while keeping the computational time within affordable limits. The basic scheme is inspired from the “double-slash” dual-level approach in quantum chemical calculations [17] for medium-size molecules in gas phase, where a high-level single-point computation is done on the geometry optimized at a lower level. Such computations are usually denoted HL/LL, where HL and LL stand for high-level and low-level methods, respectively. In our approach, LL calculations are used to generate the QM/MM sampling, while HL calculations are used afterward to obtain accurate energies on a selected set of snapshots from that sampling. Let us assume that the free-energy profile at LL along a reaction coordinate  $\xi$  has been obtained. If we further assume that the change from LL to HL Hamiltonian can be handled by means of perturbation theory [18], the free energy at HL can be estimated through the equation: [11]

$$W_{\text{HL}}(\xi) - W_{\text{LL}}(\xi) = -\frac{1}{\beta} \ln \left\langle e^{-\beta \Delta U(\xi)} \right\rangle_{\text{LL}} \quad (1)$$

where

$$\Delta U^i(\xi) = U_{\text{HL}}^i(\xi) - U_{\text{LL}}^i(\xi) \quad (2)$$

represents the potential energy difference between the high and low levels for configuration  $i$ , and  $\beta$  is the inverse temperature  $(k_{\text{B}}T)^{-1}$ . The average is calculated using a set of snapshots from the LL sampling selected at regular time intervals in the simulation and displaying a particular value of the reaction coordinate  $\xi$ . This is a major advantage of the method because one does not need to carry out high-level calculations along the whole reaction coordinate but only on selected  $\xi$  points, which limits significantly the computational effort to be done. It can be useful to rewrite the above expression using the fluctuations of the potential energy difference with respect to the average:

$$\delta \Delta U^i(\xi) = U_{\text{HL}}^i(\xi) - U_{\text{LL}}^i(\xi) - \langle \Delta U(\xi) \rangle_{\text{LL}} \quad (3)$$

which leads to:

$$W_{\text{HL}}(\xi) - W_{\text{LL}}(\xi) = \langle \Delta U(\xi) \rangle_{\text{LL}} - \frac{1}{\beta} \ln \left\langle e^{-\beta \delta \Delta U(\xi)} \right\rangle_{\text{LL}} \quad (4)$$

Here, the first term represents a free-energy correction due to differences on the potential energy average, which in general is expected to provide the largest contribution. The second term, which contains the fluctuations with respect to

the average and is connected to thermal corrections, is difficult to evaluate quantitatively. The distribution of  $\delta\Delta U^i(\xi)$  is approximately Gaussian so that the average of the exponential strongly depends on the low-energy tail, which corresponds to regions that are rarely sampled. To avoid large errors that can be introduced by these tails, several approximations can be done, which we briefly describe hereafter.

The calculation of (4) can be sufficiently accurate provided the distribution of  $\delta\Delta U(\xi)$  is well known up to two standard deviations [18]. In such case, it is possible to limit the numerical calculation to values in the range  $|\delta\Delta U(\xi)| \leq 2\sigma$ , where  $\sigma$  holds for the standard deviation of the distribution. A more rigorous approximation, however, consists in using a cumulant expansion [18, 19] limited to the second order:

$$\ln \langle e^{-\beta x} \rangle \approx -\beta \langle x \rangle + \frac{\beta^2}{2} (\langle x^2 \rangle - \langle x \rangle^2) \quad (5)$$

where the last term represents the variance of the variable  $x$ . Applying this approximation in our case, one obtains:

$$W_{\text{HL}}(\xi) - W_{\text{LL}}(\xi) \approx \langle \Delta U(\xi) \rangle_{\text{LL}} - \frac{\beta}{2} \langle \delta\Delta U(\xi)^2 \rangle_{\text{LL}} \quad (6)$$

since by definition  $\langle \delta\Delta U(\xi) \rangle_{\text{LL}} = 0$ . A similar relationship, although not totally equivalent, can be deduced if the distribution of  $\delta\Delta U(\xi)$  is fitted by a normalized Gaussian function having the general form:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right) \quad (7)$$

where for simplicity we use  $x = \delta\Delta U(\xi)$ ,  $\sigma^2$  and  $x_0$  are the variance and the position of the center of the Gaussian, respectively. Formally, the fitted Gaussian is not necessarily centered at 0 because the original distribution is not strictly symmetric in general, but if one wants to preserve the condition  $\langle \delta\Delta U(\xi) \rangle_{\text{LL}} = 0$ , it seems preferable to force  $x_0 = 0$ . Now, the second term in the right-hand side of Eq. (4) is calculated as:

$$-\frac{1}{\beta} \ln \langle e^{-\beta x} \rangle = -\frac{1}{\beta} \ln \int_{-\infty}^{+\infty} e^{-\beta x} P(x) dx \quad (8)$$

Integration of the previous equation leads finally to:

$$W_{\text{HL}}(\xi) - W_{\text{LL}}(\xi) = \langle \Delta U(\xi) \rangle_{\text{LL}} - \frac{\beta}{2} \sigma^2 \quad (9)$$

This expression has the same form as Eq. (6) using the cumulant expansion, the difference being the fact that the variance is calculated now from the fitted Gaussian rather than from the original distribution.

### 3 Computational details

As a case study, we report free-energy calculations for the torsional motion of hydrogen peroxide interacting with the vapor–liquid water interface. A *horsetail* QM/MM simulation for this system has been carried out in our recent work [10]. Details on the simulation can be found in that paper, so that only a brief description will be presented here. The simulation box contains a molecule of hydrogen peroxide described at the B3LYP/6-311+G(d) level and 499 TIP3P [20] water molecules (box size is  $24.662 \times 24.662 \times 130$  Å; we used periodic boundary conditions along the  $X$  and  $Y$  directions and a cutoff radius of 12.331 Å). Simulations were done in the NVT ensemble ( $T = 298$  K). A 100 ps MD trajectory (time step of 0.25 fs) was carried out to be used as the stem for the *horsetail* sampling. The latter consisted of 96 independent trajectories of 2.5 ps each launched at 26 internodes along the stem separated by 4 ps, for a total simulation time of 6.24 ns. Snapshots were saved every 25 fs for further analysis. The simulations were done using Gaussian 09 [21] for the QM calculations, Tinker 4.2 for the MD simulations [22] and the program developed by us [23].

The free-energy profile for the torsional angle  $\phi$  at the B3LYP/6-311+G(d) level has been calculated using the probability distribution obtained in the *horsetail* sampling, and the equation:

$$G(\phi) = -\frac{1}{\beta} \ln(P(\phi)) \quad (10)$$

This calculation depends on the number of bins used for the computation of the probability distribution. We used here 35 bins of  $5^\circ$  for the torsional angle between  $5^\circ$  and  $180^\circ$  but increasing (up to 45) or decreasing (up to 25) this number did not modify our results significantly. In order to get more accurate thermodynamic properties, free-energy perturbation theory has been used in the present paper. High-level QM/MM calculations were done using the CCSD(T)/aug-cc-pVTZ method for all the saved configurations displaying some specific values of the torsional angle  $\phi \pm 1^\circ$ . The region close to the free-energy minimum was explored in deeper detail. A total number of about 50000 QM/MM computations at the CCSD(T)/aug-cc-pVTZ level were done. The number of configurations used for each dihedral angle varies depending on the angle probability recovered with the original DFT sampling. It ranges between 3600 and 8600 for torsional angles close to the free-energy minimum (region  $[70^\circ-100^\circ]$ ), then it decreases as the free energy increases. For  $\phi = 10^\circ$ , which is in the highest energy part of the curve, only five configurations were

obtained, limiting the accuracy of the computation in this case. Standard errors will be provided in the next section. The free-energy corrections were calculated using perturbation theory and the cumulant approximation (6).

In the figures below, open circles represent the calculated points. These points are then fitted by a Fourier series torsional potential function,

$$E_{\text{tors}} = k_0 + \frac{k_1}{2}(1 + \cos(\phi)) + \frac{k_2}{2}(1 - \cos(2\phi)) + \frac{k_3}{2}(1 + \cos(3\phi)) + \frac{k_4}{2}(1 - \cos(4\phi)) + \frac{k_5}{2}(1 + \cos(5\phi)) \quad (11)$$

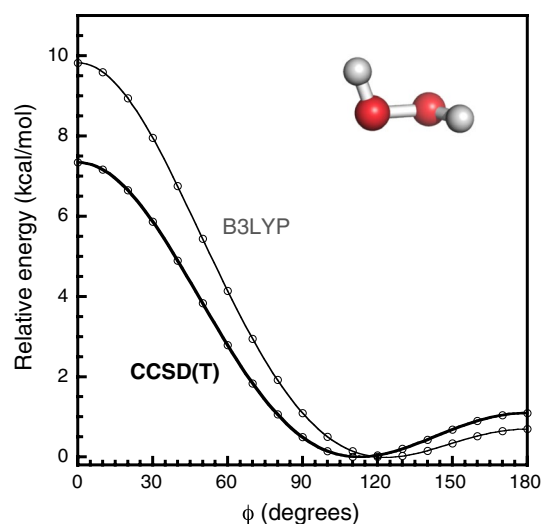
to estimate the free energy at  $\phi = 0^\circ$  and  $\phi = 180^\circ$  in the case of  $\text{H}_2\text{O}_2$  at the interface. The position of the energy minima  $\phi_0$  has been determined using a quadratic function that fits the points around the minimum (range  $[70^\circ\text{--}100^\circ]$ ).

## 4 Results and discussion

QM/MM statistical simulations of hydrogen peroxide in bulk water have been reported by several authors [24–26]. Simulations at the vapor–liquid water interface have also been carried out using both classical [27] and QM/MM [10] molecular dynamics. It has been shown that  $\text{H}_2\text{O}_2$  displays a significant affinity for the vapor–liquid water interface [10, 27] and this finding might have important implications on the reactivity of this chemical species in the low atmosphere. In particular, the UV–Vis absorption cross section is expected to be influenced by the solute–solvent interactions, which in turn may lead to a significant modification of the photolytic rate constant, as found in the case of ozone [28].

In order to achieve reliable predictions for the chemical reactivity of  $\text{H}_2\text{O}_2$  at the vapor–liquid water interface, it is critical to get an accurate description of its structure in this aqueous environment. The equilibrium value of the torsional angle and the height of the torsional energy barriers are the most problematic questions because, as discussed by other authors [29, 30], their calculation requires the use of elaborated correlation methods. Our focus in this work, therefore, has been to analyze this specific issue by means of the combined approach described above.

In Fig. 1, we compare the potential energy in gas phase at the B3LYP and CCSD(T) levels, as a function of the torsional angle  $\phi$ . The geometries were optimized at the B3LYP level for each  $\phi$  value (relaxed scan); then single-point CCSD(T) calculations were conducted on these geometries. Energy barriers and equilibrium angles are tabulated in Table 1.

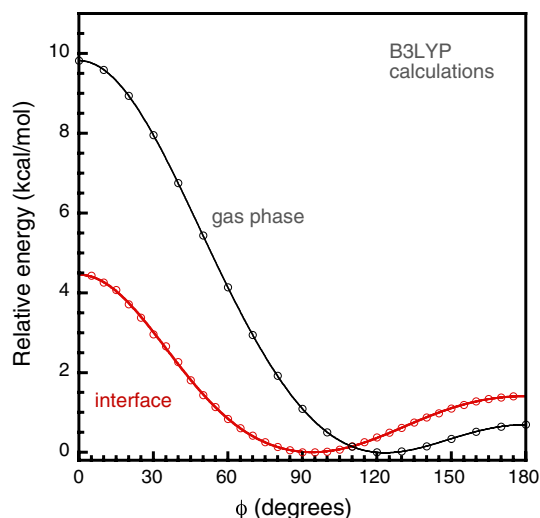


**Fig. 1** Relaxed potential energy surface for  $\text{H}_2\text{O}_2$  as a function of the HOOH torsional angle  $\phi$  in gas phase. B3LYP/6-311+G(d) vs CCSD(T)/aug-cc-pVTZ calculations. The CCSD(T) calculations have been carried out on the B3LYP optimized geometries with a fixed  $\phi$  angle. The lowest points are arbitrary taken as the zero energy in each case

**Table 1** Calculated energy barriers (kcal/mol) and torsional angle at equilibrium (degrees) of  $\text{H}_2\text{O}_2$  in the gas phase and at the vapor–liquid water interface at B3LYP/6-311+G(d) and CCSD(T)/aug-cc-pVTZ levels

	<i>Cisoid</i> barrier	<i>Transoid</i> barrier	$\phi_0$
Gas phase			
B3LYP	9.82	0.70	123.1
CCSD(T)	7.34	1.10	112.8
Interface			
B3LYP	4.45	1.41	92.3
CCSD(T)	4.28	1.38	87.7

As shown, both methods give similar qualitative results though the *cisoid* barrier is overestimated and the *transoid* barrier underestimated by the DFT approach. Moreover, there is a significant difference in the equilibrium angle  $\phi$ , which is predicted to be  $123.1^\circ$  and  $112.8^\circ$  by the B3LYP and CCSD(T) methods, respectively. The CCSD(T) value is in excellent agreement with the experimental estimate  $112^\circ$  reported by Koput [31] by fitting infrared and microwave transitions simultaneously to a large-amplitude Hamiltonian that accounts for vibration–torsion–rotation interaction, and also with other theoretical calculations at similar level [31, 32]. The CCSD(T) activation energies for the *transoid* and *cisoid* transition states, 1.1 and 7.3 kcal/mol, respectively, are also in excellent agreement with the experimental values:  $385/387\text{ cm}^{-1}$  (1.1 kcal/mol) and  $2488/2563\text{ cm}^{-1}$

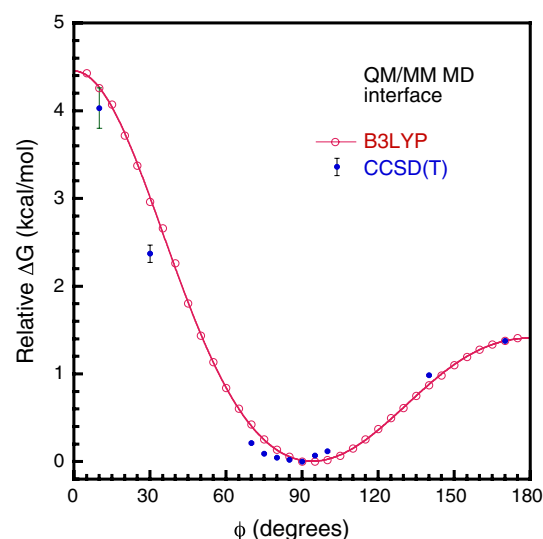


**Fig. 2** Relaxed potential energy surface in gas phase versus free-energy profile at the vapor–liquid water interface for  $\text{H}_2\text{O}_2$  as a function of the HOOH torsional angle  $\phi$ . Calculations at the B3LYP/6-311+G(d) level in both cases. Values at the interface are obtained from *horsetail* QM/MM MD simulations. The lowest points are arbitrary taken as the zero energy in each case

(7.1–7.3 kcal/mol), respectively (see Ref. [33] and references cited therein).

In Fig. 2, we report the free-energy curve obtained at the vapor–liquid water interface using the *horsetail* MD simulations at the B3LYP level. For comparison, the potential energy curve in the gas phase is also reported. The solvation effect leads, as expected [24], to a strong decrease in the *cisoid* barrier (by 5.4 kcal/mol) and a slight increase in the *transoid* one (by 0.4 kcal/mol), which is primarily due to the variation of the  $\text{H}_2\text{O}_2$  dipole moment with the torsional angle (the *cisoid* form being highly polar, while the *transoid* form has zero dipole moment). For the same reason, the solvation effect does also modify the equilibrium value  $\phi_0$ . It changes from  $123.1^\circ$  in the gas phase to  $92.3^\circ$  at the vapor–liquid water interface. This change represents a huge modification of the peroxide structure, which may have significant consequences in terms of the electronic properties and in particular of the electronic absorption spectrum [10].

In order to get more accurate data at the interface, we have calculated free-energy corrections at the CCSD(T) level using the *dual-level* method. The energy profiles are plotted in Fig. 3 (see also data in Table 1), and in Table 2 we have collected the contributions to the free-energy corrections in the cumulant expansion (6). Error bars for the calculated energies are represented in Fig. 3, although they are only visible for  $\phi = 10^\circ, 30^\circ$  (in the other cases, standard errors are too small to be displayed, see Table 2 and the discussion below). For comparison, we also include in Table 2 the contribution of the variance when a fitted



**Fig. 3** Free energy profile at the vapor–liquid water interface for  $\text{H}_2\text{O}_2$  as a function of the HOOH torsional angle  $\phi$ . *Horsetail* B3LYP/6-311+G(d) calculations vs *dual-level* CCSD(T)/aug-cc-pVTZ values. The lowest points are arbitrary taken as the zero energy in each case

**Table 2** Calculated contributions to the cumulant expansion of the free-energy correction (kcal/mol) for different values of the torsional angle ( $^\circ$ ) of  $\text{H}_2\text{O}_2$  at the vapor–liquid water interface

$\phi$	$\langle \Delta U(\phi) \rangle_{LL}$	$-\frac{\beta}{2} \langle \delta \Delta U(\phi)^2 \rangle_{LL}$	$-\frac{\beta}{2} \sigma^2$
10	2.03 (0.23)	−0.23	
30	1.84 (0.10)	−0.40	
70	2.16 (0.01)	−0.34	−0.34
75	2.21 (0.01)	−0.34	−0.35
80	2.26 (0.01)	−0.31	−0.33
85	2.32 (0.01)	−0.32	−0.32
90	2.35 (0.01)	−0.32	−0.32
95	2.41 (0.01)	−0.30	−0.31
100	2.42 (0.01)	−0.29	−0.29
140	2.36 (0.01)	−0.21	−0.21
170	2.22 (0.02)	−0.18	−0.17

For comparison, we also display (right column) the contribution when a fitted Gaussian is used instead of the cumulant expansion (Eq. 9)

SE for  $\langle \Delta U(\phi) \rangle_{LL}$  are indicated in parenthesis. Here, we use relative potential energies at B3LYP/6-311+G(d) and CCSD(T)/aug-cc-pVTZ levels, the reference being the calculated energies of  $\text{H}_2\text{O}_2$  at its B3LYP optimized geometry

Gaussian is used instead of the cumulant expansion (see Eq. 9; note that we do not provide values for torsional angles of  $10^\circ$  and  $30^\circ$  because fitting a Gaussian to the corresponding  $\delta \Delta U(\phi)$  distributions would not be realistic due to too limited samplings). As shown, the values are very close to those obtained with the cumulant expansion in the

present case, and therefore, they will not be discussed further here.

Likewise B3LYP, compared to gas phase, CCSD(T) calculations at the interface predict a large torsional angle decrease (by approximately  $25^\circ$ ), a significant decrease in the *cisoid* energy barrier (by roughly 3 kcal/mol), and a slight increase in the *transoid* one (by roughly 0.3 kcal/mol). Interestingly, at the interface, the CCSD(T) energy barriers are close to B3LYP calculations (see Table 1 and Fig. 3), suggesting that intermolecular interaction corrections compensate  $\text{H}_2\text{O}_2$  intramolecular energy corrections. This fact can be explained by considering that B3LYP dipole moments are overestimated with respect to the more accurate CCSD(T) calculations. For instance, for the structure with  $\phi = 10^\circ$  in gas phase, the low-level dipole moment is 3.66 D while the high-level value for the same structure is 3.27 D. We then expect intermolecular interactions to be slightly overestimated at the low level, as indeed found. Thus, taking the set of structures with  $\phi = 10^\circ \pm 1$  in the QM/MM simulation, the solute–solvent interactions are overestimated by roughly 4 kcal/mol (on average) at B3LYP level. Note finally that, quantitatively, the predicted decrease for the *cisoid* energy barrier is affected by the large error bars of the high-level corrections at  $\phi = 10^\circ, 30^\circ$ . However, the energy decrease with respect to gas phase is substantially larger than the standard errors at the interface gathered in Table 2.

As shown in Table 2, when one compares the high-level free-energy corrections attributed to the change in the potential energy and to its variance, one notes that the first term undergoes the largest variations with the torsional angle. It ranges between 1.84 and 2.42 kcal/mol, whereas the contributions due to the variance ranges between  $-0.18$  and  $-0.40$  kcal/mol. The latter, however, depends on  $\beta$  and should play a slightly larger role at lower temperatures.

## 5 Conclusions

The calculation of free-energy profiles in complex molecular systems remains an important challenge for theoretical chemistry, especially in the context of *ab initio* molecular dynamics and related simulation techniques. The reason for that is twofold. On the one hand, the convergence of the free-energy calculation is generally slow because it critically depends on the effectiveness of the phase space sampling. On the other hand, elaborated quantum chemical calculations have a high computational cost and in addition they are not efficiently parallelizable, limiting therefore the quality of the potential energy function that can be used in the simulations. In this work, we have proposed a possible strategy to address this challenge. It is based on

the combination of the *horsetail* molecular dynamics and *dual-level* approaches previously described for calculations using QM/MM partitions. The central idea is to obtain a realistic sampling of the system using an appropriate cost-effective statistical method, then to use perturbation theory to ameliorate the probability distribution in some selected points of the reaction coordinate. In this paper, we have used the *horsetail* sampling to get the probability distributions and free energies from direct simulations but it would be possible to use a similar approach in connection with techniques based on biased molecular dynamics. For instance, in umbrella sampling simulations, each simulation window could be carried out within the *horsetail* sampling scheme, which in principle should improve the accuracy of the distributions and accelerate the convergence of the free-energy calculation.

Using the above computational scheme, we have succeeded in obtaining the free-energy profile associated with the torsional motion of the hydrogen peroxide molecule adsorbed at the vapor–liquid water interface at the CCSD(T)/aug-cc-pVTZ level. We have shown, in particular, that the equilibrium angle of  $\text{H}_2\text{O}_2$  is significantly changed with respect to the gas phase value, which may have important implications on the photochemistry of this system interacting with water droplets in the low atmosphere [10].

Overall, the proposed methodology opens new opportunities for the calculation of very accurate thermodynamic properties in large disordered systems, and accordingly, it can be compared to standard composite methods for the study of isolated molecules in the gas phase. Further extensions in this direction will be considered in forthcoming work.

**Acknowledgements** The authors are grateful to the French CINES (project lct2550) for providing computational resources and to the French–Japan CNRS–JSPS PRC program (project CoSyDy).

## References

1. Warshel A, Levitt M (1976) *J Mol Biol* 103:227–249
2. Field MJ, Bash PA, Karplus M (1990) *J Comput Chem* 11:700–733
3. Stanton RV, Hartsough DS, Merz KM (1993) *J Phys Chem* 97:11868–11870
4. Gao J, Xia X (1992) *Science* 258:631–635
5. Tuñón I, Martins-Costa MTC, Millot C, Ruiz-López MF (1995) *J Mol Mod* 1:196–201
6. Tuñón I, Martins-Costa MTC, Millot C, Ruiz-Lopez MF (1995) *Chem Phys Lett* 241:450–456
7. Tuñón I, Martins-Costa MTC, Millot C, Ruiz-López MF (1997) *J Chem Phys* 106:3633–3642
8. Strnad M, Martins-Costa MTC, Millot C, Tuñón I, Ruiz-López MF, Rivail JL (1997) *J Chem Phys* 106:3643
9. Torrie GM, Valleau JP (1977) *J Comput Phys* 23:187

10. Martins-Costa MTC, Ruiz-López MF (2017) *J Comput Chem* 38:659
11. Retegan M, Martins-Costa M, Ruiz-López MF (2010) *J Chem Phys* 133:064103
12. Martins-Costa MTC, Ruiz-Lopez MF (2013) *J Phys Chem B* 117:12469–12474
13. Auffinger P, Westhof E (1996) *Biophys J* 71:940–954
14. Anderson JB (1973) *J Chem Phys* 58:4684–4692
15. Anderson JB (1995) *Adv Chem Phys* 91:381–431
16. Bennett CH (1977) Molecular dynamics and transition state theory: the simulation of infrequent events. In: Christoffersen RE (ed), *Algorithms for chemical computations*, chap 4. ACS symposium series, vol 46. American Chemical Society, Washington, DC, pp 63–97
17. Corchado JC, Truhlar DG (1998) Dual-level methods for electronic structure calculations of potential energy functions that use quantum mechanics as the lower level. In: Gao J, Thompson MA (eds), *Combined quantum mechanical and molecular mechanical methods*, chap 7. ACS Symposium Series, vol 712. American Chemical Society, Washington, DC, pp 106–127
18. Chipot C, Pohorille A (2007) *Free energy calculations*. Springer, Berlin
19. Park S, Khalili-Araghi F, Tajkhorshid E, Schulten K (2003) *J Chem Phys* 119:3559–3566
20. Jorgensen WL, Chandrasekar J, Madura JD, Impey WR, Klein ML (1983) *J Chem Phys* 79:926–935
21. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery JA Jr, Peralta JE, Ogliaro F, Bearpark MJ, Heyd J, Brothers EN, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell AP, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam NJ, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas Ó, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ (2009) *Gaussian 09*. Gaussian Inc, Wallingford
22. Ponder JW (2004) *TINKER: software tools for molecular design 4.2*. Washington University School of Medicine, Saint Louis, MO
23. Martins-Costa MTC (2014) A Gaussian 09/Tinker 42 interface for hybrid QM/MM applications. University of Lorraine—CNRS, Lorraine
24. Martins-Costa MTC, Ruiz-Lopez MF (2007) *Chem Phys* 332:341–347
25. Cristina Caputo M, Provasi PF, Benitez L, Georg HC, Canuto S, Coutinho K (2014) *J Phys Chem A* 118:6239–6247
26. Fedorov DG, Sugita Y, Choi CH (2013) *J Phys Chem B* 117:7996–8002
27. Vacha R, Slavicek P, Mucha M, Finlayson-Pitts BJ, Jungwirth P (2003) *J Phys Chem* 108:11573
28. Anglada JM, Martins-Costa M, Ruiz-Lopez MF, Francisco JS (2014) *Proc Natl Acad Sci USA* 111:11618–11623
29. Maciel GS, Bitencourt ACP, Ragni M, Aquilanti V (2006) *Chem Phys Lett* 432:383–390
30. Margules L, Demaison J, Boggs J (2000) *J Mol Struct THEOCHEM* 500:245–258
31. Koput J (1995) *Chem Phys Lett* 236:516–520
32. Maciel GS, Bitencourt ACP, Ragni M, Aquilanti V (2007) *J Phys Chem A* 111:12604–12610
33. Dorofeeva OV, Iorish VS, Novikov VP, Neumann DB (2003) *J Phys Chem Ref Data* 32:879–901