



FEG optimizations using dual-level approach

Carlos BISTAFA

Graduate School of Informatics, Nagoya University, Japan

2nd CREST Workshop FY2018
Nagoya, September 18th, 2018

Today's Talk

Cost-Effective Method for Free-Energy Minimization in Complex Systems with Elaborated Ab Initio Potentials

Carlos Bistafa,[†] Yukichi Kitamura,[†] Marilia T. C. Martins-Costa,[‡] Masataka Nagaoka,^{*,†,§,||} and Manuel F. Ruiz-López^{*,‡,⊥}

Cost-Effective Method for Free-Energy Minimization in Complex Systems with Elaborated Ab Initio Potentials

Carlos Bistafa,[†] Yukichi Kitamura,[†] Marilia T. C. Martins-Costa,[‡] Masataka Nagaoka,^{*,†,§,||} and Manuel F. Ruiz-López^{*,‡,⊥}

[†]Department of Complex Systems Science, Graduate School of Informatics, Nagoya University, Chikusa Ku, Furo Cho, Nagoya, Aichi 4648601, Japan

[‡]Laboratoire de Physique et Chimie Théoriques, UMR CNRS 7019, Faculté des Sciences et Technologies, Université de Lorraine, CNRS, BP 70239, 54506 Vandœuvre-lès-Nancy Cedex, France

[§]ESICB, Kyoto University, Kyodai Katsura, Nishikyo-ku, Kyoto 6158520, Japan

^{||}Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Honmachi, Kawaguchi 3320012, Japan

[⊥]Future Value Creation Research Center, Graduate School of Informatics, Nagoya University, Chikusa Ku, Furo Cho, Nagoya, Aichi 4648601, Japan

Supporting Information

ABSTRACT: We describe a method to locate stationary points in the free-energy hypersurface of complex molecular systems using high-level correlated *ab initio* potentials. In this work, we assume a combined QM/MM description of the system although generalization to full *ab initio* potentials or other theoretical schemes is straightforward. The free-energy gradient (FEG) is obtained as the mean force acting on relevant nuclei using a dual level strategy. First, a statistical simulation is carried out using an appropriate, low-level quantum mechanical force-field. Free-energy perturbation (FEP) theory is then used to obtain the free-energy derivatives for the target, high-level quantum mechanical force-field. We show that this composite FEG-FEP approach is able to reproduce the results of a standard free-energy minimization procedure with high accuracy, while simultaneously allowing for a drastic reduction of both computational and wall-clock time. The method has been applied to study the structure of the water molecule in liquid water at the QM/MM level of theory, using the sampling from QM/MM molecular dynamics simulations at the B3LYP/6-311+G(d,p) level. The obtained values for the geometrical parameters and for the dipole moment of the water molecule are within the experimental error, and they also display an excellent agreement when compared to other theoretical estimations. The developed methodology represents therefore an important step toward the accurate determination of the mechanism, kinetics, and thermodynamic properties of processes in solution, in enzymes, and in other disordered chemical systems using state-of-the-art *ab initio* potentials.

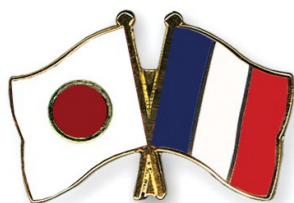
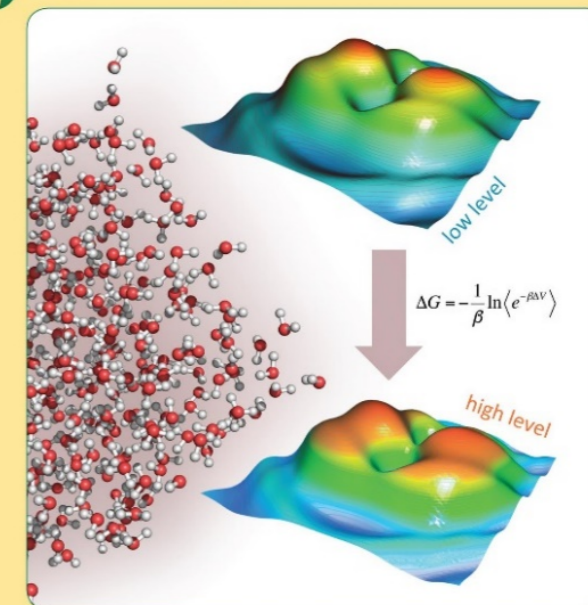
1. INTRODUCTION

The theoretical study of chemical reactions in solution and in complex molecular environments often requires the knowledge of the solute's geometry at the stationary points of the free-energy landscape of the system (reactants, products, intermediates, transition structures). Since the pioneer applications of implicit continuum models to describe reaction paths in solution,¹ thousands of new studies have been carried out using these models, providing valuable insights into solvation effects on the kinetics, thermodynamics, and/or reactivity of chemical processes. In continuum models (for further details see reviews reported before^{2–5}), it is straightforward to compute the derivatives of the free energy of solvation with respect to the solute's parameters^{6,7} and therefore to locate the relevant stationary points. However, due to crude approximations, implicit models are not always

sufficiently accurate, and large efforts have been made in the last years to develop more elaborated models.

In this line, Molecular Dynamics (MD) using QM/MM (Quantum Mechanics/Molecular Mechanics) force-fields has become one of the most popular and successful techniques to study complex molecular systems.⁸ The price to be paid is a much higher computational cost, and hence most applications assume semiempirical or density functional theory QM methods and MD sampling limited to a few tens or hundreds of picoseconds (though recent algorithms⁹ have been allowed to reach multianosecond time scales). In MD simulations, the calculation of free-energy reaction profiles is usually accomplished by the use of *ad hoc* techniques, which require the prior

Received: March 16, 2018
Published: May 9, 2018



Free Energy Gradient (FEG) Method

From a classical simulation, we can obtain the forces acting in each atom of the solute molecule in all the steps. By time-averaging these force values, the force on the Free Energy Surface can be obtained as:

$$F(q) = -\frac{\partial G(q)}{\partial q} = -\left\langle \frac{\partial V(q)}{\partial q} \right\rangle$$

where $G(q)$ is the free energy and V is the sum of the solute intramolecular potential energy and the solute-solvent interaction energy at coordinate q .

Advantage: We can optimize atomistically and free-energetically the solute molecule in solution;

Limitation: Gradient calculations are performed in all steps of the QM/MM simulation, so the computational cost can be very demanding.

Sequential-QM/MM^{a,b} and ASEC-FEG^c

1: Classical Simulation

2: Select Relevant Configurations

3: QM Solute and Classical Solvent

4: ASEC^d

Steps

QM

$$F = \left\langle -\frac{\partial V(q)}{\partial q} \right\rangle \cong -\frac{\partial \langle V(q) \rangle}{\partial q}$$

(Ref. e)

a) Canuto & Coutinho, *Int. J. Quantum Chem.* **77** (2000)192;

b) Solvent Effects on Molecules and Biomolecules, S. Canuto (ed.), Springer 2008 (chap. 6);

c) Georg & Canuto, *J. Chem. Phys. B.* **116** (2012) 11247;

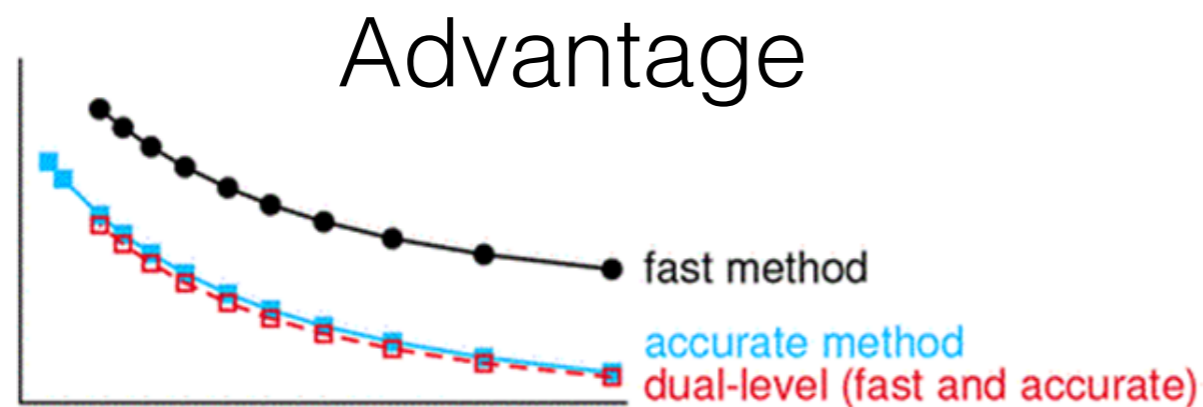
d) Coutinho et. al., *Chem. Phys. Lett.* **437** (2007) 148;

e) Galván et. al., *J. Chem. Phys.* **118** (2003) 255.

Dual-Level Method

FEP:
$$W^B(\xi) - W^A(\xi) = -k_B T \left\langle e^{-\frac{U^B(\xi) - U^A(\xi)}{k_B T}} \right\rangle$$

In particular:
$$A_{HL}(q) = A_{LL}(q) - \frac{1}{\beta} \left\langle e^{-\beta \Delta V} \right\rangle$$



Limitation: High-Level (HL) and Low-Level (LL) have to generate close enough samplings.

Motivation

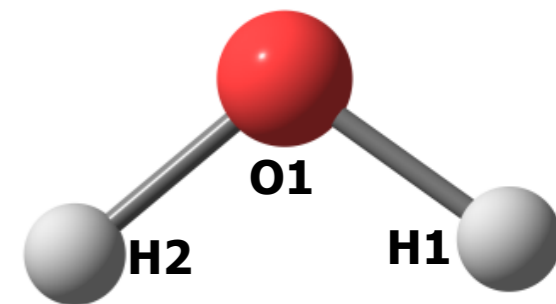
Perform high-level FEG calculations without the limitations of mean-field approaches.

In order to achieve this we use a dual-level approach, performing the QM/MM simulation and FEG calculations with an inexpensive low-level method, and then correcting the values to a higher level using the dual level approach.

Computational Details

- **System: 1 QM water molecule + 592 TIP3P water molecules.**
(small, interesting and data is available for comparison)
- **Two cases were studied**
 - 1) B3LYP::HF -> QM/MM simulations performed at the HF/6-31G(d) level, with FEG calculations corrected to B3LYP/6-311++G(d,p).
In order to validate our approach, usual QM/MM simulation and FEG calculations at the B3LYP/6-311++G(d,p) level were also performed.
 - 2) QCISD::B3LYP -> QM/MM simulations performed at the B3LYP/6-311++G(d) level, with FEG calculations corrected to QCISD/aug-cc-pVTZ.

In both cases: NVT, 300K, 40 ps, time-step of 1fs.



Dual level calculations

Using FEP equations, we can prove that the derivatives of the free energy at HL are related to the LL by:

$$\frac{\partial G_{HL}(q)}{\partial q} = \left\langle \frac{\partial V_{HL}(q;s)}{\partial q} w(q;s) \right\rangle_{LL}$$

Where w is a weighting factor expressed by:

$$w(q;s) = \frac{e^{-\beta\Delta V(q;s)}}{\left\langle e^{-\beta\Delta V(q;s)} \right\rangle_{LL}}$$

We use 10% of the configurations sampled at LL to calculate the values at HL.

Results: “close enough samplings”

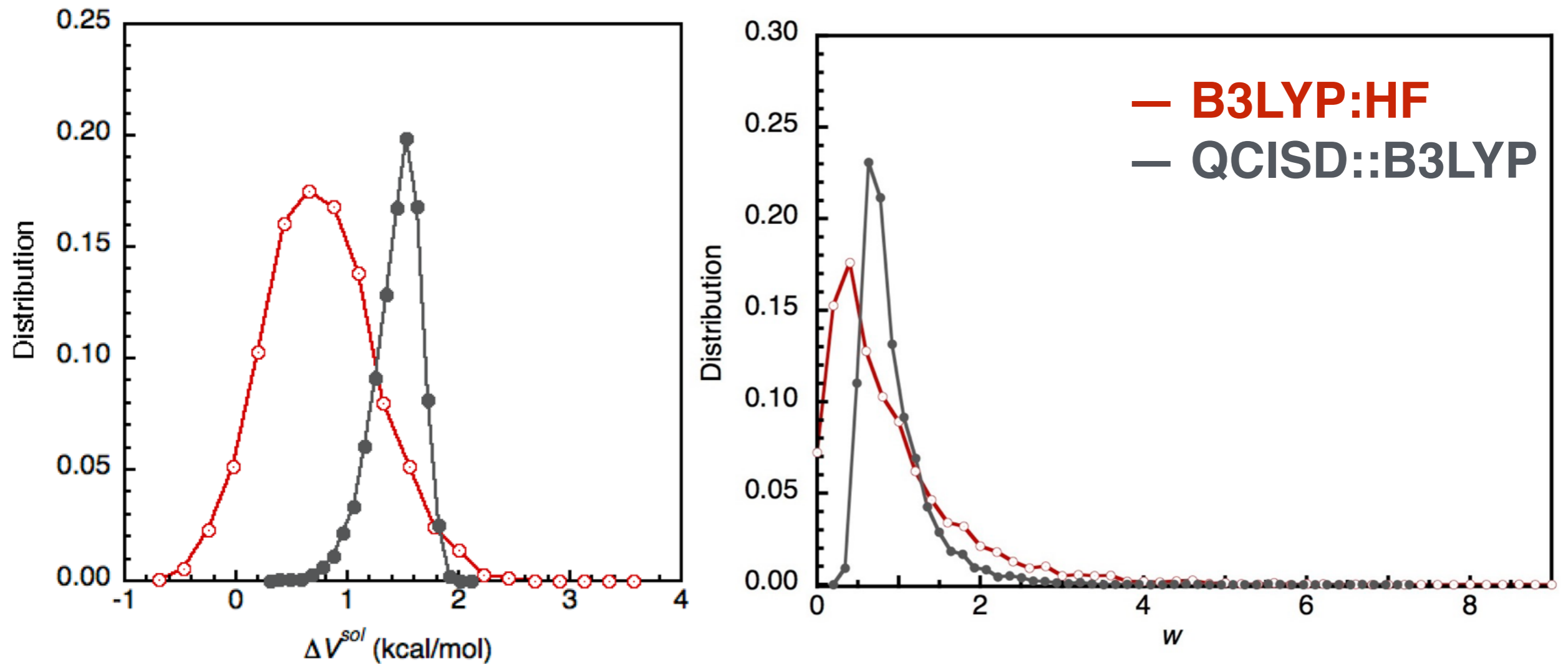


Figure 1: Normalized distributions of ΔV^{sol} (left) and w (right) in the two studied cases for the 10% saved configurations (in total, 4000) in the first iteration of the QM/MM MD simulation at LL of each optimization procedure. ΔV^{sol} is the difference between the HL and LL QM/MM solvation energies of the water molecule in liquid water defined by $\Delta V^{sol} = \Delta V(q;s) - (E_{HL}^0 - E_{LL}^0)$.

Results: “close enough samplings”

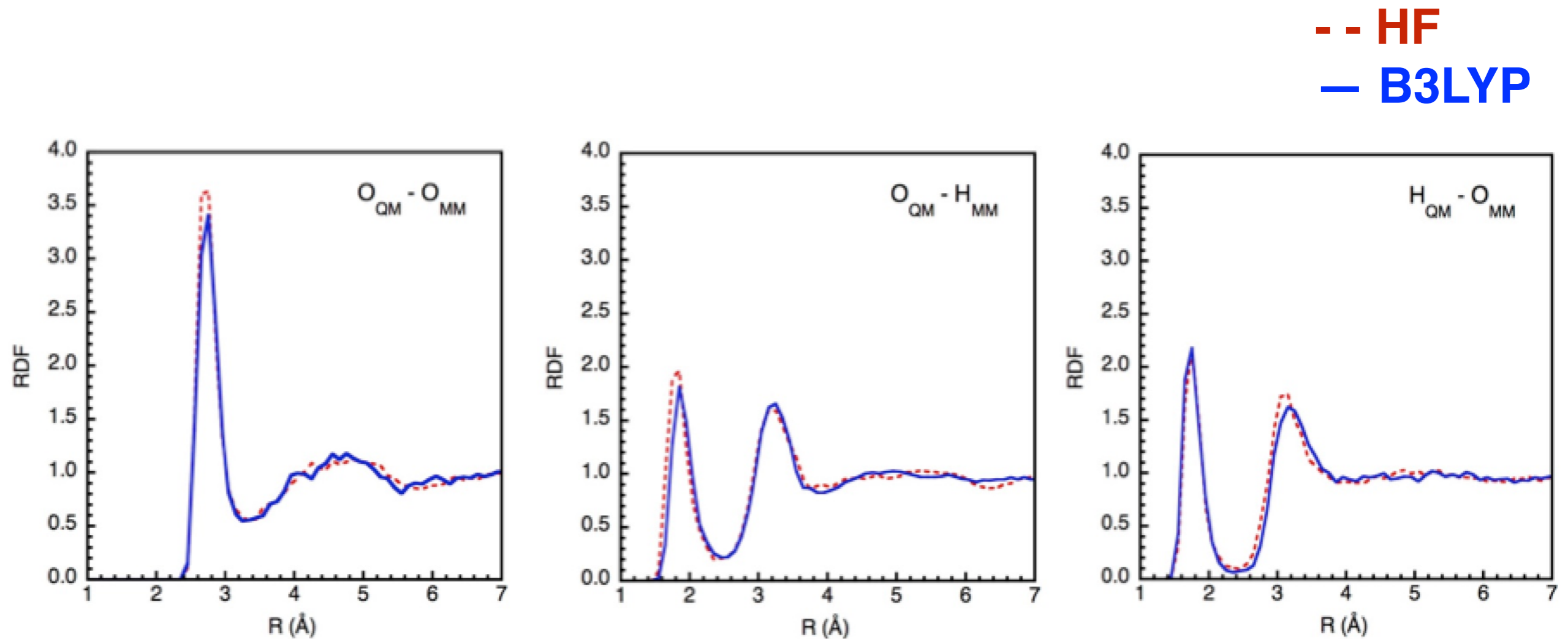


Figure 2: Radial distribution functions of a water molecule (QM) in liquid water (MM) from QM/MM simulations at HF (red, dashed) and B3LYP (blue, plain) levels in case study 1, B3LYP::HF, at the initial point of the optimization procedure.

Results: FEG and FEG-FEP comparison

-o- B3LYP::HF (FEG-FEP)
-> B3LYP (FEG)

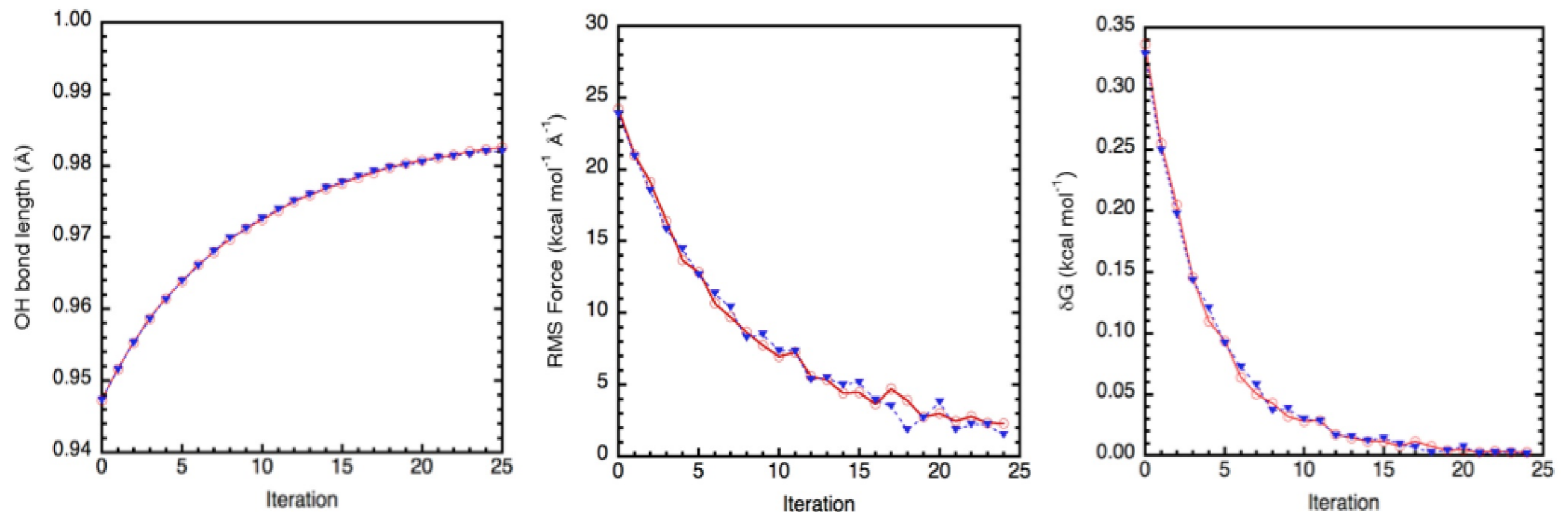


Figure 3: Convergence of average OH bond length, RMS force, and predicted change in free energy (δG) for the water molecule in aqueous solution along the approximate FEG-FEP (red, circles, plain line) and standard FEG (blue, triangles, dashed line) optimization methods in B3LYP::HF case.

Results: Optimized Structure of Water Molecule in Aqueous Solution

Method	Property	Gas	Liquid	Δ
B3LYP/6-311+G(d,p) FEG^a	d_{OH} (Å)	0.962	0.982	+0.020
	α (°)	105.1	105.2	+0.1
	μ (D)	2.16	2.91	+0.75
B3LYP::HF	d_{OH} (Å)	0.962	0.982	+0.020
	α (°)	105.1	105.2	+0.1
	μ (D)	2.16	2.95	+0.79
QCISD::B3LYP	d_{OH} (Å)	0.959	0.979	+0.020
	α (°)	104.4	105.5	+1.1
	μ (D)	1.86	2.69	+0.83
Experimental values	d_{OH} (Å)	0.957 ^b	0.970±0.005 ^c	+0.013
	α (°)	104.5 ^b	~106±1.7 ^c	+1.5
	μ (D)	1.85 ^d	2.6 ^e	+0.75

^a Standard FEG optimization;

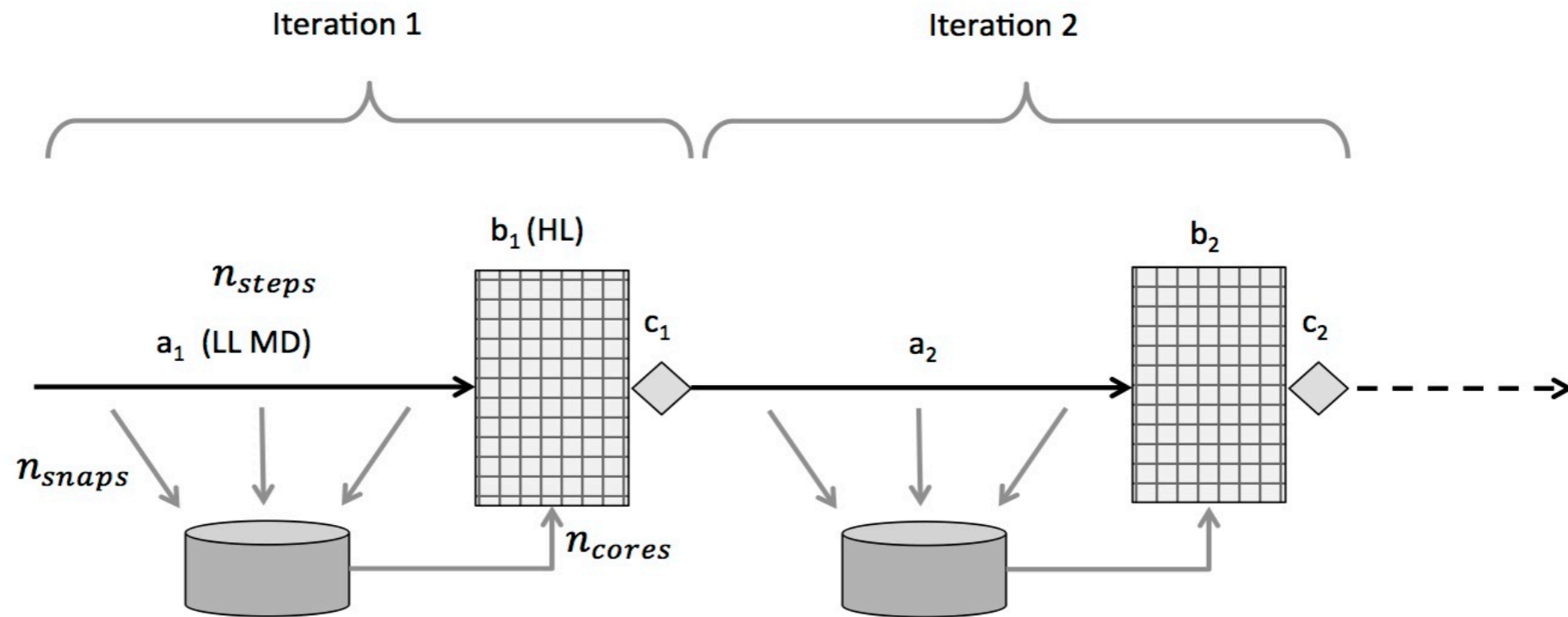
^b Benedict et. al, *J. Chem. Phys.*, **24**, 1139 (1956);

^c Ichikawa et. al., *Mol. Phys.*, **73**, 79 (1991);

^d Clough et. al., *J. Chem. Phys.*, **59**, 2254 (1973);

^e Coulson et. al., *Proc. R. Soc. London., Ser. A*, **291**, 445 (1966).

Conclusions and Perspectives



- Our calculations show that the results obtained using FEG and FEG-FEP are the same;
- FEG-FEP is computationally very efficient. For instance, the QCISD optimization took 20 days, while our estimation is that it would take 1.2 years using the usual FEG approach and same computational resources;
- Our results obtained at QCISD are the most accurate theoretical results for structure and dipole of the water molecule in aqueous solution available so far. They are consistent with previous calculations and with experimental data;
- **Perspectives:** Extend the equations for the Hessian case, in order to perform transition states optimizations and infrared spectrum calculations (very promising).

Conclusions and Perspectives

Table 1: The Hessian obtained with B3LYP::HF (FEG-FEP) method and with usual FEG method at B3LYP. The difference in the values obtained in both calculations is mild. Values in kcal/molÅ².

$$\frac{\partial^2 G_{HL}}{\partial q_{\alpha i} \partial q_{\beta j}} = \left\langle \frac{\partial^2 V}{\partial q_{\alpha i} \partial q_{\beta j}} w \right\rangle_{LL}$$

B3LYP	Ox	Oy	Oz	H1x	H1y	H1z	H2x	H2y	H2z
Ox	108.171								
Oy	-2.892	1501.454							
Oz	1.888	12.302	1049.192						
H1x	-41.541	0.704	-0.948	55.419					
H1y	-0.125	-678.513	380.373	2.317	718.138				
H1z	-0.152	489.880	-472.186	-1.587	-414.751	-41.579			
H2x	-0.305	0.049	3.267	-0.049	-0.213	-677.827	-379.338		
H2y	-0.117	-62.014	-0.002	-489.526	-473.188	0.251	56.840	439.414	
H2z	0.054	59.740	-56.387	-0.441	717.735	36.316	-0.096	407.253	435.127

B3LYP::HF	Ox	Oy	Oz	H1x	H1y	H1z	H2x	H2y	H2z
Ox	106.959								
Oy	-3.185	1515.630							
Oz	-0.348	-2.393	1048.908						
H1x	-43.025	0.787	-0.822	62.187					
H1y	0.173	-676.209	378.459	-0.099	711.517				
H1z	-0.244	487.451	-472.200	-0.520	-404.380	-42.404			
H2x	-0.273	0.148	3.485	-0.036	-0.392	-676.956	-379.000		
H2y	-0.094	-61.853	-0.199	-488.168	-472.276	0.134	55.959	436.734	
H2z	-0.081	61.824	-56.215	0.782	711.808	37.147	0.957	405.847	437.377