

N° d'ordre : 021

SOUTENANCE
en vue d'obtenir

L'HABILITATION À DIRIGER DES RECHERCHES
délivrée par l'École Normale Supérieure de Lyon

présentée et soutenue publiquement le Vendredi 2 Juillet 2010

par **Monsieur Paul FLEURAT-LESSARD**

**Mechanisms in chemistry:
Finding and Understanding reaction paths**

Après avis de:

Monsieur P. HIBERTY

Monsieur S. HUMBEL

Madame A. MILET

Devant la Commission d'examen formée de:

Dr. P. HIBERTY (Université de Paris-Sud) Membre/Rapporteur

Pr. S. HUMBEL (Université P. Cézanne) Membre/Rapporteur

Pr. A. MILET (Université J. Fourier) Membre/Rapporteur

Pr. F. VOLATRON (Université Paris 6) Membre

Dr. P. SAUTET (ENS de Lyon) Membre

Contents

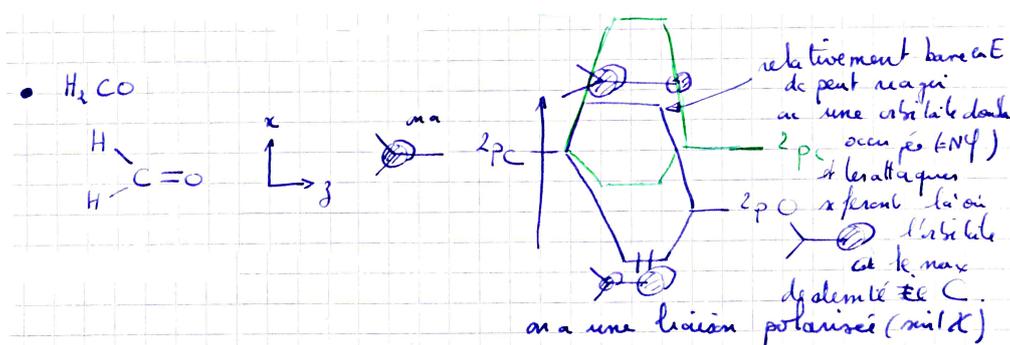
Introduction	1
I The one with the souvenir	3
1 Summary of my previous work	4
1.1 Geometrical and Electronic Structures	4
1.2 Chemical reactivity	5
2 CV	7
II The one with some methodology	14
3 The one with the Minimum Free Energy Path	15
4 The one with the Adaptive Biasing Density of States	33
III The one with some applications	42
5 The one with the unusual bond	43
6 The one with biochemistry	50
7 The one with organic reactions	62
IV The one with the future plans	67
1 Tools development	68
1.1 Describing the solvent	68
2 Applications	69
2.1 Biochemistry	69
2.2 Organic reactions with isocyanides	70

Introduction

What is chemistry ?

Clearly, the answer will depend on who you ask ! For me, chemistry is the science of molecules, and how they react onto each other. My first approach with “real” chemistry was in the second year of the *Classes Préparatoires* when I learnt all these organic reactions, with their mechanisms. At that time, I wondered what was the driving force behind these microscopic acts.

F. Volatron gave me the beginning of the answer during his lecture on quantum chemistry based on molecular orbitals:

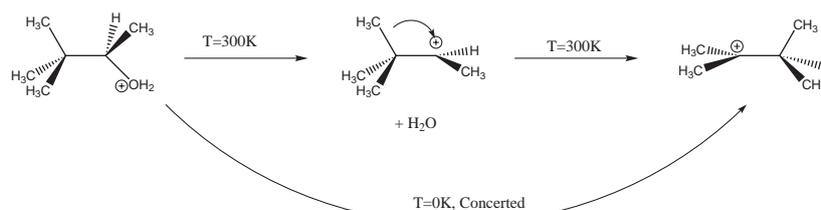


As indicated on the right, when one considers the anti-bonding π orbital of a carbonyl group, the largest coefficient is localized on the carbon. Therefore a nucleophilic specie will attack on this site. I had found my path !

During my master and Ph.D., both directed by F. Volatron, we studied the modifications induced by substituting a hydrogen atom by a π donor in organic or inorganic complexes.

After my Ph. D., I decided to orient my research themes towards the theoretical study of the chemical reactivity. The first step in such studies is to characterize the stationary points on the potential energy surface (PES) that correspond to the reactants, the products, the reaction intermediates and the transition states. The path connecting these points is called the Minimum Energy Path (MEP).

However, it might even happen that the MEP does not correspond at all to the path followed by the molecules at finite temperature because entropic effects are important for the reaction. For example, the pinacolic rearrangement is concerted at 0K but occurs in two steps at 300K:



In this case, it is important to be able to construct the most probable path (in a statistical way, that is the path that corresponds to the average of all trajectories) while taking the

temperature explicitly into account. Using molecular dynamic simulations, one can estimate directly the change in free energy associated to the transformation of the reactants to the products. As the free energy is a state function, it does not depend on a specific path. However, the transition state position will depend on the choice of the reaction coordinates and thus on the chosen path.

During my post-doctoral stay with T. Ziegler (Calgary), I have started working on the construction of the most probable path leading from the reactant to the product. We have first developed tools to optimize stationary points directly on the free energy surface.[P20, P21]* After my recruitment at the ENS Lyon in 2003, in collaboration with T. Ziegler, I have extended this project to the construction of the Minimum Free Energy Path: a path linking the reactants to the products on the free energy surface.[P22]

Since then, my research axis are always focused on the chemical reactivity: either for the development of new exploration tools, or for the application of these tools to complex biochemical or chemical reactions.

The rest of this memoir is organized as follows: in part I, I summarize the work that I have done during my Ph.D. and my first two post-doc. This part is concluded with my CV. Part II focuses on my methodological activity: it recalls the work done with T. Ziegler before exposing the method that we recently proposed for efficient free energy sampling. Part III presents some of the applications I have considered.

*Reference prefixed by *P* are given in my CV, pages 9 and following.

Part I

The one with the souvenir

1) Summary of my previous work

The work I have done before entering the École Normale Supérieure de Lyon can be divided into two parts: (i) electronic and geometrical studies of organic and inorganic molecules, and (ii) chemical reactivity.

1.1 Geometrical and Electronic Structures

During my Master and my PhD, both directed by F. Volatron in Orsay, we focused in understanding the chemical properties such as the equilibrium geometries or the redox potentials of organic and inorganic molecules. More specifically, we have focused on the replacement of hydrogen atoms by π -donor substituents such as NR_2 groups, that may lead to substantial changes in chemical properties. Such a situation is found in tetrakis(dimethylamino)-ethylene (TDAE): the structure characterized by Bock et al.[1] indicate that the four amino groups partially deconjugate in a conrotatory motion. In addition, the two $(\text{Me}_2\text{N})_2\text{C}$ halves partially deconjugate, the twist angle between them being equal to 28° . During my Master, we have shown that this resulted from a subtle balance between electronic and steric effects.[P2]

During my Ph. D., we have focused on amino ligands in inorganic complexes in which a single σ metal-nitrogen bond exists. Two properties of these ligands are worth noticing: on the one hand, when the alkyl group R is sterically demanding, they act as 'steric' shields to stabilize unsaturated species. On the other hand, the electronic structure of the complex may be dramatically affected by the interaction between the nitrogen lone pair and the metal orbitals. We have first studied triamido aluminum complexes, in which very different structures are found for different amido substituents. We found that the electronic structure results from a balance between the interaction of the aluminum with the lone pairs and the lone pair repulsion, leading to a shallow minimum. Finally, the $\text{Al}[\text{N}(\text{SiMe}_3)_2]$ and $\text{Al}(\text{N}^i\text{Pr}_2)_3$ have been studied through the hybrid IMOMM method. It is found that their different structures result from different way of minimizing the steric effects.[P6]

We have then studied some MH_4 and MCl_4 complexes. We showed that the molecular shape of these complexes can be predicted by considering the bielectronic terms in the d bloc. These terms are minimum when the degenerate orbital have the same occupancy. The energy evolution of the d bloc orbitals is shown Figure 1.1 when going from a tetrahedral geometry to a square planar one.

For instance, we found that d^0 , High Spin (HS) d^2 , Low Spin (LS) d^4 and HS d^5 favor tetrahedral geometries whereas d^4 LS and d^8 HS favor square planar structure.

In $\text{M}(\text{NH}_2)_4$, d_{z^2} and d_{xy} orbitals are no longer degenerate, which leads to a diamagnetic $\text{Mo}(\text{NMe}_2)_4$ complex. However, the tetrahedral geometry is still favored for the $\text{Ti}(\text{NH}_2)_4$, $\text{Cr}(\text{NH}_2)_4$ and $\text{Co}(\text{NH}_2)_4$ complexes.[P24] For $\text{Zn}(\text{NH}_2)_4$, π interactions are destabilizing and the square planar structure is unstable, leading to a bisamido complex $\text{Zn}(\text{NH}_2)_2$.

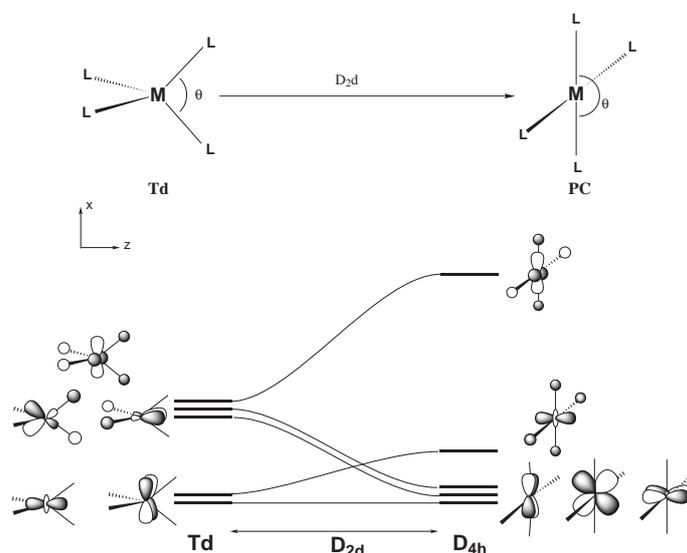
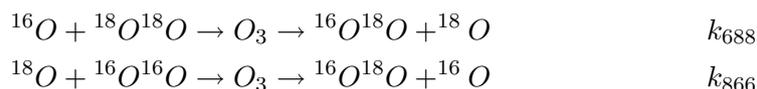


Figure 1.1: Evolution of d bloc orbitals for the tetrahedral to square planar deformation.

1.2 Chemical reactivity

At the end of my Ph. D., I decided to learn how to describe the dynamical properties of a system. During my first post-doc under the supervision of J.-C. Rayez in Bordeaux, we studied the addition of the CH radical to methane to form the C_2H_5 radical, as well as the dissociation of $C_2H_5^\bullet$. In agreement with the experiments, we found that all allowed path lead to the loss of hydrogen atoms, while loss of a H_2 molecule was not energetically possible.[P7]

I then moved to the group of R. Schinke in the Max Planck Institut für Strömungsforschung in Göttingen. My project dealt with the unusual mass independent isotopic enrichment experimentally observed for the ozone O_3 formation. After improving the global potential energy surface,[P8, P11] we used classical trajectories to compute the kinetic constants for the exchange reactions:



It is experimentally known that the ratio $Q = k_{866}/k_{688}$ is larger than one, and decrease from 1.27 at 300K to 1.2 at 400K.[2] More, by studying different combinations of oxygen isotopes, Mauesberger et al. showed that this ratio depends on the zero point energy (ZPE) of the ${}^X O^Y O$ molecule ($X=16,17$ or 18 ; $Y=16,17$ or 18).[3] We included the ZPE in our classical trajectories by either modifying the PES (method C) or by selecting only trajectories in which the O_2 vibrational energy was larger than its ZPE (method D). The results are show on Figure 1.2, along with the experimental data and the classical trajectories (method A).

Both methods C and D are in good agreement with the experiments, thus confirming the importance of the ZPE for the isotopic enrichment.[P14, P19, P23, P25]

Bibliography

- [1] a) Bock, H.; Borrmann, H.; Havlas, Z.; Oberhammer, H.; Ruppert, K.; Simon, A. *Angew. Chem. Int. Ed. Engl.* **1991**, *30*, 1678-1681. b) Bruckmann, J.; Krüger, C.; Borrmann, H.; Simon, A.; Bock, H. *Zeitschrift für Kristallographie* **1995**, *210*, 521.

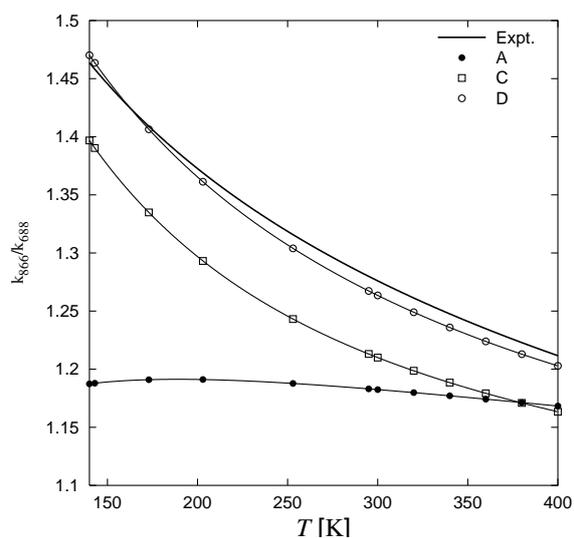


Figure 1.2: Influence on the temperature on the ratio $Q = k_{866}/k_{688}$. A: Classical trajectories on original PES, C: Classical trajectories on modified PES, D: Selected classical trajectories

[2] Wiegell, M. R.; Larsen, N. W.; Pedersen, T.; Egsgaard, H. *Int. J. Chem. Kinetics* **1997**, *29*, 745.

[3] a) Janssen, C.; Guenther, J.; Mauersberger, K.; Krankowsky, D. *Physical Chemistry Chemical Physics* **2001**, *3*, 4718-4721. b) Janssen, C.; Guenther, J.; Krankowsky, D.; Mauersberger, K. *Chemical Physics Letters* **2003**, *367*, 34 - 38.

2) CV

Paul FLEURAT-LESSARD

Né le: 20 Avril 1973, à Bordeaux (33)
 Nationalité: Française
 Adresse professionnelle: Laboratoire de Chimie, UMR 5182
 École Normale Supérieure de Lyon
 46 allée d'Italie
 69384 LYON Cedex 07.
 Grade: MCF
 Section CNU: 31
 Téléphone: 04 72 72 81 54
 Fax: 04 72 72 88 60
 E-mail: Paul.Fleurat-Lessard@ens-lyon.fr
 Web: <http://perso.ens-lyon.fr/paul.fleurat-lessard/>

Formation Professionnelle

- 09/03 -** **Maître de Conférence** au Laboratoire de Chimie de l'École Normale Supérieure (ENS) de Lyon.
- 01/03 - 08/03** **Stage Post-Doctoral** dirigé par T. ZIEGLER à l'Université de Calgary :
Chemins d'énergie libre de réactions organiques simples par Dynamique Moléculaire Ab Initio.
- 2001-2002** **Stage Post-Doctoral** dirigé par R. SCHINKE au Max Planck Institut für Strömungsforschung à Göttingen :
Étude théorique de la dissociation et de la recombinaison de l'ozone : $O + O_2 \rightleftharpoons O_3$.
- 01/01-07/01** **Stage post-doctoral** dirigé par J.-C. RAYEZ au Laboratoire de Physico-Chimie Moléculaire à Bordeaux :
Détermination des rapports de branchements pour $CH + CH_4$.
- 1997-2000** **Thèse de Chimie Théorique**, Université Paris XI
Étude théorique du rôle des ligands donneurs π dans les composés des métaux de transition et du groupe principal
 Direction : F. VOLATRON au Laboratoire de Chimie Physique.
 soutenue le 22 Décembre 2000 devant : B. LEVY, J.-Y. SAILLARD, F. MASERAS, T. MALLAH, H. BOLVIN, F. VOLATRON.
- 1996-1997** **Scientifique du Contingent**, École Polytechnique.
- 1995-1996** **Agrégation de Chimie** (rang 14)
- 1994-1995** 2^e année du *Magistère de Chimie*
D.E.A. de Physico-Chimie Moléculaire, Mention Très Bien.
Maîtrise, Mention Bien (Septembre 94).
- 1993-1994** 1^e année du *Magistère de Chimie*
Licence de Chimie, Mention Assez Bien (Février 94).
- 1991-1993** Classes Préparatoires (P'), **Entrée à l'ENS Ulm.**

Principales expériences d'enseignement

1998-2000	Moniteur à l'Université d'Orsay Encadrement de TD et TP en DEUG Sciences de la Matière <i>T. Mallah / F. Volatron</i> 50 heures par an TD de <i>Liaison Chimique</i> au Magistère Interuniversitaire de Chimie 15 heures par an
2003-2010	Cours de spectroscopie en M1, ENS Lyon 15 h/an
2003-2007	TP d'électrochimie en M1 et préparation à l'agrégation, ENS Lyon 20+20 h/an
2003-2010	TP de Chimie Orbitale M1, ENS Lyon 20 h/an
2003-2010	TD de Chimie Quantique (Hartree-Fock et au delà) M1, ENS Lyon 10 h/an
2003-2010	Cours de Chimie Orbitale et de Chimie du Solide Préparation à l'agrégation de Physique et de Chimie de l'ENS Lyon 50 h/an
2003-2010	Correction de devoirs, leçons et montages (Préparation à l'agrégation) 50 h/an
2006-2010	Cours de Dynamique Moléculaire avancée (CPMD, Thermostats, profil d'énergie libre), avec TP sur machines. 8 h/an
2007-2010	Responsable de la préparation à l'agrégation de Chimie 64 h/an

Encadrement d'étudiants

2004-2005	Stage de M2 de H. Berthoumieux, <i>Étude théorique d'une liaison faible peu usuelle : l'interaction N...CO</i>
2005-2006	Stage ATER de J. Pilmé, <i>Influence de la solvation sur l'interaction N...CO</i>
06-07 2007	Stage de L3 de F. Touti, <i>Influence des substituants sur la nature de l'interaction N...CO</i>
06-07 2007	Stage de L3 de V. Penin, <i>Utilisation de la théorie pour isoler un dihydroxyphosphorane</i>
2005-2008	Doctorat de J. Garrec, <i>Une liaison faible peu usuelle pour de nouveaux inhibiteurs pour les protéases aspartiques</i>
2007-2008	Stage ATER de D. Kozlowski <i>Calcul du profil d'énergie libre pour le repliement de molécules présentant l'interaction N...CO</i>
2007-2008	Stage de M2 de N. Cheron <i>Étude théorique du mécanisme du couplage de Ugi-Smiles</i>
2007-2008	Stage de M2 de E. Brunk <i>Influence du solvant sur la pré-organisation de molécules présentant l'interaction N...CO</i>
2008-2009	Stage de M2 de A. Adeniyi <i>Free Energy Profiles for the Folding of HIV Model Inhibitors: a QM/MM Approach</i>
2008-2009	Stage de M2 de D. De Olivera Gaspareto <i>Theoretical study of the CO2 capture by ethanolamine: influence of solvation.</i>
2008-2011	Doctorat de N. Cheron <i>Étude Théorique et Expérimentale du Mécanisme du Couplage Ugi-Smiles et Recherche d'Inducteurs Chiraux pour les Réactions Multicomposant</i>
2009-2010	Co-encadrement du stage de M2 de R. Ramozzi <i>Étude expérimentale et théorique du réarrangement de Mumm</i>

Bibliographie

- [P1] Pointet, K.; Milliet, A.; Renou-Gonnord, M.-F.; Fleurat-Lessard, P. *Mass spectrometric differentiation of isomeric polycyclic aromatic hydrocarbons by chemical ionization with diethylether, tetrahydrofuran and dimethylcarbonate* Eur. Mass. Spectrom. **1997**, *3*, 281-290.
- [P2] Fleurat-Lessard, P.; Volatron, F. *Structural consequences of π -Donation by NR₂ Groups: Ab initio Study of Tetrakis(dimethylamino)ethylene (TDAE) and Its Unsubstituted Analog* J. Phys. Chem. A **1998**, *102*, 10151-10158.
- [P3] Fleurat-Lessard, P.; Pointet, K.; Renou-Gonnord, M.-F. *Quantitative Determination of PAHs in Diesel Engine Exhausts by GC-MS* J. Chem. Ed. **1999**, *76*, 962-965.
- [P4] Aukauloo, A.; Ottenwaelder, X.; Ruiz, R.; Poussereau, S.; Pei, Y.; Journaux, Y.; Fleurat, P.; Volatron, F.; Cervera, B.; Muñoz, M. C. *A Square-Planar Dinickel(II) Complex with a Noninnocent Dinucleating Oxamate Ligand: Evidence for a Ligand Radical Species* Eur. J. Inorg. Chem. **1999**, 1067-1071.
- [P5] Fleurat-Lessard, P.; Pointet, K.; Milliet, A. *Contrasting Behavior of Tetracene and Perylene in Collision-induced Dissociation: a Theoretical Interpretation* J. Mass. Spectrom. **1999**, *34*, 768-773.
- [P6] Fleurat-Lessard, P.; Volatron, F. *Structure of Triamidoaluminum Complexes: A Theoretical ab initio/IMOMM Study* Inorg. Chem. **2000**, *39*, 1849-1854.
- [P7] Fleurat-Lessard, P.; Rayez, J.-C.; Bergeat, A.; Loison, J.-C. *Reaction of methylidine CH (X ²P) with CH₄ and H₂S: overall rate constant and absolute atomic hydrogen production.* Chem. Phys. **2002**, *279*, 87-89.
- [P8] Siebert, R.; Fleurat-Lessard, P.; Schinke, R.; Bittererová, M.; Farantos, S. C. *The vibrational energies of ozone up to the dissociation threshold: Dynamics calculations on an accurate potential energy surface* J. Chem. Phys. **2002**, *116*, 9749-9767.
- [P9] Fleurat-Lessard, P.; Durupthy, O.; Volatron, F. *Theoretical ab initio study of Xenon pentafluoride anion. Mechanism of Xenon pseudorotation* Chem. Phys. Letters **2002**, *363*, 505-508.
- [P10] Sivignon, G.; Fleurat-Lessard, P.; Onno, J. M.; Volatron, F. *Does back-bonding involve bonding orbitals in boryl complexes? A theoretical DFT study.* Inorg. Chem. **2002**, *41*, 6656-6661.
- [P11] Fleurat-Lessard, P.; Grebenshikov, S. Y.; Siebert, R.; Schinke, R.; Halberstadt, N. *Theoretical investigation of the temperature dependance of the O + O₂ exchange reaction* J. Chem. Phys. **2003**, *118*, 610-621.
- [P12] Babikov, D.; Kendrick, B. K.; Walker, R. B.; Pack, R. T.; Fleurat-Lessard, P.; Schinke, R. *Metastable states of ozone calculated on an accurate potential energy surface* J. Chem. Phys. **2003**, *118*, 6298-6308.
- [P13] Babikov, D.; Kendrick, B. K.; Walker, R. B.; Pack, R. T.; Fleurat-Lessard, P.; Schinke, R. *Formation of ozone: Metastable states and anomalous isotope effect* J. Chem. Phys. **2003**, *119*, 2577-2589.

- [P14] Schinke, R.; Fleurat-Lessard, P.; Grebenschikov, S. Y. *Isotope Dependence of the Lifetime of Ozone Complexes formed in O + O₂ collisions*. Phys. Chem. Chem. Phys. **2003**, *5*, 1966-1969.
- [P15] Kurkal, V.; Fleurat-Lessard, P.; Schinke, R. NO₂: *Global potential energy surfaces of the ground (1²A₁) and the first excited (1²B₂) electronic states* J. Chem. Phys. **2003**, *119*, 1489-1501.
- [P16] Fleurat-Lessard, P.; Grebenschikov, S. Y.; Schinke, R.; Janssen, C.; Krankowsky, D.; Mauesberger, K. *Isotope dependance of the O + O₂ exchange reaction: Experiment and theory*. J. Chem. Phys. **2003**, *119*, 4700-4712 Erratum in J. Chem. Phys. 2004, vol 120, 4993.
- [P17] Grebenschikov, S. Y.; Fleurat-Lessard, P.; Schinke, R.; Joyeux, M. *van der Waals states in ozone and their influence on the threshold spectrum of O₃ (X¹A₁). I. Bound states* J. Chem. Phys. **2003**, *119*, 6512-6523.
- [P18] Knight, L. K.; Piers, W. E.; Fleurat-Lessard, P.; Parvez, M.; McDonald, R. *β-Diketiminato Scandium Chemistry: Synthesis, Characterization, and Thermal Behavior of Primary Amido Alkyl Derivatives* Organometallics **2004**, *23*, 2087-2094.
- [P19] Schinke, R.; Fleurat-Lessard, P. *The transition-state region of the O(³P)+O₂(³Σ_g⁻) potential energy surface* J. Chem. Phys. **2004**, *121*, 5789-5793.
- [P20] Yang, S.; Fleurat-Lessard, P.; Hristov, I.; Ziegler, T. *Free Energy Profiles for the Identity S_N2 Reactions Cl⁻ + CH₃Cl and NH₃ + H₃BNH₃: A Constraint Ab Initio Molecular Dynamics Study* J. Phys. Chem. A **2004**, *108*, 9461-9468.
- [P21] Yang, S.; Hristov, I.; Fleurat-Lessard, P.; Ziegler, T. *Optimizing the Structures of Minimum and Transition State on the Free Energy Surface* J. Phys. Chem. A **2005**, *109*, 197-204.
- [P22] Fleurat-Lessard, P.; Ziegler, T. *Tracing the Minimum Energy Path on the Free Energy Surface* J. Chem. Phys. **2005**, *123*, 084101.
- [P23] Schinke, R.; Fleurat-Lessard, P. *The effect of zero-point energy differences on the isotope dependence of the formation of ozone: A classical trajectory study*. J. Chem. Phys. **2005**, *122*, 094317.
- [P24] Fleurat-Lessard, P.; Volatron, F. *Theoretical study of the distortion from regular tetrahedral structure of M(NH₂)₄ complexes* Theor. Chem. Acc. **2006**, *116*, 718-725.
- [P25] Schinke, R.; Grebenschikov, S. Y.; Ivanov, M.; Fleurat-Lessard, P. *Dynamical Studies of the Ozone Isotope Effect: A Status Report* Ann. Rev. Phys. Chem. **2006**, *57*, 625-661.
- [P26] Joubert, J.; Fleurat-Lessard, P.; Delbecq, F.; Sautet, P. *Simulating Temperature Programmed Desorption of Water on Hydrated γ-Alumina from First-Principles Calculations* J. Phys. Chem. B **2006**, *114*, 7392-7395.
- [P27] Zein, S.; Borshch, S. A.; Fleurat-Lessard, P.; Casida, M. E.; Chermette, H. *Assessment of the exchange-correlation functionals for the physical description of spin transition phenomena by DFT methods: All the same?* J. Chem. Phys. **2007**, *126*, 014105.

- [P28] Pilmé, J.; Berthoumieux, H.; Robert, V.; Fleurat-Lessard, P. *New insights into the ternary amino-carbonyl bond* Chem. Eur. J. **2007**, *13*, 5388-5393.
- [P29] Aronica, C.; Fleurat-Lessard, P.; Breteau, S.; Vigier, S. *Étude d'un système redox de référence pour les mesures électrochimiques : le ferrocène* B.U.P. **2007**, *101*, 23-31.
- [P30] Saaïdi, P.-L.; Guyonnet, M.; Jeanneau, E.; Fleurat-Lessard, P.; Hasserodt, J. *Trimerization Products of Trifluoroacetone: Critical Solvent Effect on Position and Kinetics of Anomeric Equilibria* J. Org. Chem. **2008**, *73*, 1209-1216.
- [P31] Fleurat-Lessard, P.; Volatron, F. *Structure of hexakis-methylthiobenzene: A theoretical study* Chem. Phys. Letters **2009**, *477*, 32-36.
- [P32] Geronimo, I.; Chéron, N.; Fleurat-Lessard, P.; Dumont, E. *How does microhydration impact on structure, spectroscopy and formation of disulfide radical anions ? An ab initio investigation on dimethyldisulfide*. Chem. Phys. Letters **2009**, *481*, 173 - 179.
- [P33] Garrec, J.; Cascella, M.; Rothlisberger, U.; Fleurat-Lessard, P. *Low inhibiting power of N...CO based peptidomimetic compounds against HIV-1 protease: Insights from a QM/MM study* J. Chem. Theory Comput. **2010**, *6*, 1369-1379.
- [P34] Dickson, B. M.; Legoll, F.; Lelièvre, T.; Stoltz, G.; Fleurat-Lessard, P. *Free energy calculations: An efficient adaptive biasing potential method* J. Phys. Chem. B **2010**, *114*, 5823-5830.
- [P35] Garrec, J.; Sautet, P.; Fleurat-Lessard, P. *A new QM model for the HIV protease active site: new insights for the formation of the intermediate*, en préparation.
- [P36] Chéron, N.; Grimaud, L.; El Kaïm, L.; Fleurat-Lessard, P. *A small step in understanding the Ugi-Smiles reaction: role of a Hydrogen bond*, en préparation.

Contrats et financements divers

Cluster 5	Allocation doctorale de la région Rhône-Alpes, 2005-2008. Projet <i>Une liaison inhabituelle pour la conception d'inhibiteurs de la protéase HIV</i> .
Cluster 5	Aide financière de la région Rhône-Alpes, 2007-2008. Porteur du projet <i>Une approche QM/MM-PBSA pour la conception d'inhibiteurs de la protéase HIV</i> .
ANR Sire	ANR <i>Calculs Intensifs et Simulations, CIS</i> , n° ANR-06-CIS-014 de Janvier 2006 à Juin 2010. Porteur ENS de la tâche <i>Carte</i> (Chemins Automatisés pour la Réactivité incorporant les effets de la Température, pression et de l'Environnement).
ANR Muse	ANR <i>Chimie et Procédés pour le Développement Durable, CP2D</i> , n° ANR-08-CP2D-15-02 de décembre. Porteur du projet à l'ENS.
PHC VanGogh	Partenariats Hubert Curien (PHC) VanGogh entre la France et les Pays-Bas. <i>Chemistry in Solution: A Multi-scale Molecular Dynamics Approach</i> , Participant. Durée 1 an (2010).

Organisation d'événements scientifiques

- CARTE Day Le 25 mai 2008, ENS Lyon.
Journée de présentation du logiciel CARTE développé dans le cadre de l'ANR SIRE. La présentation théorique a été suivie d'exercices sur machines.
- DFT09 31/08-4/09 2009, Lyon
Membre du comité local (Organising Committee).

Communications

ORALES

- Mai 2000 **Journées Francophones des Jeunes Physico-chimistes**, Paris.
Complexes tétracoordinés des métaux de transition : préférence géométrique en fonction du décompte électronique,
Fleurat-Lessard, P.; Volatron, F.
- Oct. 2000 **7^e Rencontres des Chimistes Théoriciens Francophones**, Gruissan.
Complexes tétracoordinés des métaux de transition : préférence géométrique en fonction du décompte électronique,
Fleurat-Lessard, P.; Volatron, F.
- Mai 2008 **HIM workshop on "Numerical methods in molecular simulation"**, Bonn.
Reaction path: a chemist view,
Fleurat-Lessard, P.
- Sep. 2009 **Workshop LN3M**, Lyon.
Low inhibiting power of N...CO based peptidomimetic compounds against HIV-1 protease: Insights from a QM/MM study ,
J.Garrec; Fleurat-Lessard, P.

AFFICHES

- Juil. 1998 **Journées Francophones des Jeunes Physico-chimistes**, Montpellier.
Étude théorique des distorsions structurales pour le Tetrakis(diméthylamino)éthène,
Fleurat-Lessard, P.; Volatron, F.
- Juil. 1998 **Journées Francophones des Jeunes Physico-chimistes**, Montpellier.
Analyse théorique de la structure des dérivés triamino de l'aluminium $Al(NR_2)_3$,
Fleurat-Lessard, P.; Volatron, F.
- Sep. 1999 **Cinquième école d'été de physicochimie théorique**, Marly-le-Roi.
Analyse théorique de la structure des dérivés triamino de l'aluminium $Al(NR_2)_3$,
Fleurat-Lessard, P.; Volatron, F.
- Mai 2000 **Journées Francophones des Jeunes Physico-chimistes**, Paris.
Structure électronique du pentafluorure du Xénon. Pseudorotation de l'atome central dans les composés AX_5E_2 ?,
Fleurat-Lessard, P.; Volatron, F.; Durupthy, O.
- Mai 2000 **Dynam-2000** Arcachon.
Statistical analysis of the photodissociation of tungsten hexacarbonyl $W(CO)_6^+$ complex,
Fleurat-Lessard, P.; Ohanessian, G.; Dutuit, O.; Alcaraz, C.; Vervloet, M.

- Juin 2000 **Xth International Congress of Quantum Chemistry**, Menton.
Tetracoordinated transition metal complexes: geometrical dependence upon electron counting,
Fleurat-Lessard, P.; Volatron, F.
- Juin 2000 **Xth International Congress of Quantum Chemistry**, Menton.
Xenon pentafluoride anion XeF_5^- : electronic structure and pseudo rotation mechanism in AX_5E_2 species,
Fleurat-Lessard, P.; Volatron, F.; Durupthy, O.
- Oct. 2000 **7^e Rencontres des Chimistes Théoriciens Francophones**, Gruissan.
Structure électronique du pentafluorure du Xénon. Pseudorotation de l'atome central dans les composés AX_5E_2 ?,
Fleurat-Lessard, P.; Volatron, F.; Durupthy, O.
- Juil. 2004 **Canadian Symposium on Theoretical Chemistry 2004**, Quebec (Canada).
Tracing the Minimum Energy Path on the Free Energy Surface,
Fleurat-Lessard, P.; Ziegler, T.
- Sep. 2004 **Conference on Computational Physics 2004**, Gênes (Italie).
Tracing the Minimum Energy Path on the Free Energy Surface,
Fleurat-Lessard, P.; Ziegler, T.
- Avr. 2006 **Chemical Reactivity Symposium 2006**, Bruxelles (Belgique).
Tracing the Minimum Energy Path on the Free Energy Surface,
Fleurat-Lessard, P.; Ziegler, T.

Participation à des congrès

- Juin 2001 **Journées de Cinétique et Photochimie en Phase Gazeuse**, Lille (France).
- Fév. 2002 **Ecole d'hiver sur la théorie de la Fonctionnelle de la Densité (DFT)**, Mariapfarr (Autriche).
- Mai 2002 **Workshop: Quantum Dynamical Concepts: From Diatomics to Biomolecules**, Dresde (Allemagne).
- Fév. 2009 **Tutorial CP2K: enabling the power of imagination in MD simulations**, Zürich (Suisse).

Responsabilité administratives

- 2004-2005 Membre élu du Conseil de Laboratoire du LCH.
- 2005-2008 Membre élu du Conseil Scientifique de l'ENS Lyon.
- 2005-2008 Membre de la Commission de Spécialistes 31 de l'UCBL.
- 2007-2008 Membre de la Commission de Spécialistes de Chimie de l'ENS Lyon.

Part II

The one with some methodology

3) The one with the Minimum Free Energy Path

The Minimum Energy Path (MEP) is a path linking the reactants to the products via one or more transition states on the PES. However, in many cases, the influence of the temperature (via the entropy term especially) is such that the observed path at finite temperature differs qualitatively from the MEP. With this in mind, it becomes logical to design some tools to find a similar path on the Free Energy Surface (FES).

This was the goal of my project with T. Ziegler in Calgary, and this work is presented in the following article. Our main conclusion is that one can easily adapt the usual tools used for finding MEP to the construction of the Minimum *Free* Energy Path (MFEP) via a molecular dynamic simulation. In particular, in this article, we propose to construct the MFEP as the mass weighted steepest descent path starting from the Transition State (TS) directly optimized at finite temperature on the FES.¹ To do so, one only needs to compute the free energy gradient:

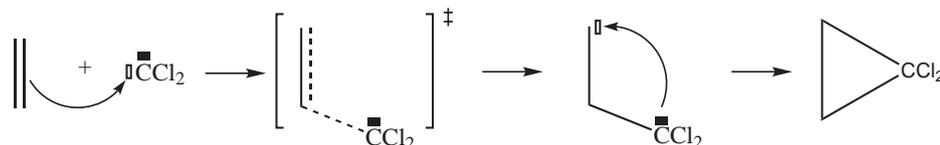
$$\frac{\partial A}{\partial \xi} = \frac{1}{\langle Z_{\xi}^{-1/2} \rangle_{\xi^*}} \left\langle Z_{\xi}^{-1/2} \left(\lambda_{\xi} + \frac{1}{2\beta Z_{\xi}^2} \left\{ \sum_{i=1}^{i=3N} \frac{\partial \xi}{\partial x'_i} \frac{\partial Z_{\xi}}{\partial x'_i} \right\} \right) \right\rangle_{\xi^*},$$

where ξ is the reaction coordinate, Z_{ξ} is the inverse mass matrix element related to ξ and λ_{ξ} is the Lagrange multiplier given by the Shake algorithm.

As this equation involves second derivatives of the reaction coordinate ξ (through the derivative of Z_{ξ}), its evaluation is not always straightforward. Therefore, we also showed that for many usual constraints (such as fixing a bond distance or the center of mass), the equations become quite simple. Indeed, in such cases, the free energy gradient is just the average of the potential energy gradient:

$$\left. \frac{\partial A(\xi)}{\partial \xi} \right)_{\xi^*} = \left\langle \frac{\partial V}{\partial \xi} \right\rangle_{\xi^*}$$

We then applied this method to the construction of the MFEP for the addition of CCl_2 to ethylene:



¹a) Yang, S.; Fleurat-Lessard, P.; Hristov, I.; Ziegler, T. *J. Phys. Chem. A* **2004**, *108*, 9461–9468. b) Yang, S.; Hristov, I.; Fleurat-Lessard, P.; Ziegler, T. *J. Phys. Chem. A* **2005**, *109*, 197–204.

Tracing the minimum-energy path on the free-energy surface

Paul Fleurat-Lessard^{a)}

Laboratoire de Chimie, UMR CNRS 5182, Ecole Normale Supérieure de Lyon, 46 allée d'Italie, 69364 Lyon Cedex 07, France

Tom Ziegler^{b)}

Department of Chemistry, University of Calgary, University Drive 2500 Calgary, Alberta, Canada T2N 1N4

(Received 5 October 2004; accepted 13 May 2005; published online 26 August 2005)

The free-energy profile of a reaction can be estimated in a molecular-dynamics approach by imposing a mechanical constraint along a reaction coordinate (RC). Many recent studies have shown that the temperature can greatly influence the path followed by the reactants. Here, we propose a practical way to construct the minimum-energy path directly on the free-energy surface at a given temperature. First, we follow the blue-moon ensemble method to derive the expression of the free-energy gradient for a given RC. These derivatives are then used to find the actual minimum-energy reaction path at finite temperature, in a way similar to the intrinsic reaction path of Fukui on the potential-energy surface. [K. Fukui, *J. Phys. Chem.* **74**, 4161 (1970)]. Once the path is known, one can calculate the free-energy profile using thermodynamic integration. We also show that the mass-metric correction cancels for many types of constraints, making the procedure easy to use. Finally, the minimum-free-energy path at 300 K for the addition of CCl_2 to ethylene is compared with a path based on a simple one-dimensional reaction coordinate. A comparison is also given with the reaction path at 0 K.

I. INTRODUCTION

Being able to understand, or better to predict, the evolution of a complex system is of critical importance in all areas of chemistry and biology. In turn, this understanding requires the knowledge of not only the mechanism at a microscopic level but also of the free-energy change associated with the reaction under investigation. In principle, molecular-dynamics simulation can give access to the free-energy profile of chemical processes, and indeed, free-energy simulations have become a key tool in the study of many chemical and biochemical problems.¹

However, chemical reactions are usually rare events and they would require a much too long simulation time for a process to occur without any bias. Thus, finding an efficient way to accurately compute the free-energy difference for a given reaction is still a very active field of research.²⁻⁶ Several methods have been proposed to evaluate the change of the free energy along a given path: free-energy perturbation,⁷ umbrella sampling,⁸ thermodynamic integration,⁹ and, more recently, the Jarzynski equality.^{5,6} (See also Ref. 10 for a review.)

Of particular importance is the understanding of the link between constrained and unconstrained simulations, which was put on a firm basis a decade ago by Carter *et al.* who introduced the blue-moon relation.¹¹ During the past few

years, this relation has been refined so that many formulas are now at hand to evaluate the derivative of the free energy along a given reaction coordinate.¹²⁻¹⁴

The access to analytical gradients on the potential-energy surface was a very important step forward in standard quantum chemistry; it became possible to find the optimum geometry of complex systems, optimizing transition states became easier, and frequencies could be obtained much faster with a higher accuracy. On the other hand, despite the fact that the exact equations for the evaluation of gradients of the free-energy surface have been available for many years, their use has mainly been restricted to the evaluation of the free-energy change along a predefined path. Evaluating such a free-energy profile is common in both chemistry and biology. However, it seems that the potential of free-energy gradients has not been fully appreciated.

In this study, we propose to use the available equations for the gradient to explore the free-energy surface in much the same way as gradients have been used to explore the potential-energy surface. Special attention will be given to finding the minimum-free-energy path.

The account of this investigation is organized as follows. In Sec. II we first review the main equations employed to evaluate the derivative of the free energy along a reaction coordinate in a uniform way. In Sec. III, the scope of these expressions is extended, and their actual use in a simulation is discussed. In Sec. IV, they are then used to find a minimum-energy path on the free-energy surface for the addition of dihalide carbene to ethylene. Section V offers the concluding remarks.

^{a)}Electronic mail: paul.fleurat-lessard@ens-lyon.fr

^{b)}Electronic mail: ziegler@ucalgary.ca

II. THEORY

A. Intrinsic reaction path

We consider a chemical system composed of N atoms of mass m_i described by $3N$ Cartesian coordinates x_i with $i = 1, \dots, 3N$. We want to construct the minimum-energy path connecting the reactants to the products on the free-energy surface (FES). On the potential-energy surface (PES), Fukui^{15,16} has defined such a path: the intrinsic reaction path (IRP). On each point of this path, the atomic Cartesian coordinates satisfy

$$\dots = \frac{m_i dx_i}{\partial V / \partial x_i} = \frac{m_j dx_j}{\partial V / \partial x_j} = \dots, \quad (1)$$

where V is the potential energy and m_i is the mass of the atom with coordinate x_i . This equation can be simplified by using mass-weighted Cartesian coordinates,

$$x'_i = \sqrt{m_i} x_i, \quad i = 1, \dots, 3N. \quad (2)$$

In its simplified form, Eq. (1) reads

$$\dots = \frac{dx'_i}{\partial V / \partial x'_i} = \frac{dx'_j}{\partial V / \partial x'_j} = \dots. \quad (3)$$

Thus, the IRP corresponds to a steepest-descent path using the mass-weighted coordinates.

In this work, we want to construct a similar path on the FES. Therefore, we have to find a path satisfying

$$\dots = \frac{dx'_i}{\partial A / \partial x'_i} = \frac{dx'_j}{\partial A / \partial x'_j} = \dots, \quad (4)$$

where A stands for the Helmholtz free energy.

In order to find this path, we have to calculate the gradient of the free energy for each point of this path. In practice, this path will be discretized by a set of k points in the configurational space, i.e., by the set of k molecular geometries,

$$\mathbf{x}^j = \{x_i^j; i = 1, \dots, 3N\}, \quad j = 1, \dots, k. \quad (5)$$

B. Generalized coordinates

Even though one can use the $3N$ Cartesian coordinates to describe the system and its evolution, it is usually easier to employ a set of $3N-6$ generalized internal coordinates, as well as the overall rotation and translation of the molecule.

Further, chemical intuition tells us that most of the time only a few degrees of freedom (reaction coordinates) are sufficient to describe the reaction path. Thus, it appears natural to split the generalized coordinates into two subsets corresponding to the active coordinates, denoted by $\xi = (\xi_1, \dots, \xi_r)$, and the inactive coordinates, denoted by $\mathbf{q} = (q_1, \dots, q_n)$. A more quantitative criterion for this separation will be discussed later. These two sets are associated with two groups of generalized momenta $\mathbf{p}_\mathbf{q}$ and \mathbf{p}_ξ , and a velocity vector,

$$\mathbf{v}_{\mathbf{q}\xi} = \begin{pmatrix} \dot{\mathbf{q}} \\ \dot{\xi} \end{pmatrix}.$$

The IRP will then be constructed in the subset of the active coordinates ξ .

This separation is similar to the adiabatic separation used in quantum dynamics between the slow modes (corresponding to our active set) and the fast modes (corresponding to our inactive set). One common point is that the inactive coordinates can vary along the reaction path. In other words, being inactive does not mean being frozen; and inactive coordinate is characterized by the fact that it does *not* contribute to the direction of the minimum-free-energy path as the thermal motions along these coordinates are nearly harmonic. However, motion along the inactive coordinates might contribute to the changes in the free energy as the curvature of the harmonic potential changes along the free-energy path. This point will be illustrated in the Application section.

It is worth discussing here the meaning of a structure on the potential and on the free-energy surfaces. On the potential-energy surface, a structure corresponds to a stationary point; for such a point, the derivatives of the potential are zero for all coordinates. On the free-energy surface, we can use a similar description: a structure corresponds to a point in which the derivatives of the free energy are zero for all coordinates. By definition of the inactive coordinate, they do not contribute to a change in the direction of the minimum-free-energy path. Therefore, the derivative of the free energy along an inactive coordinate is zero: $\partial A / \partial q_i = 0$ for all inactive coordinates. As a consequence, a structure can be defined as a point in which the derivatives of the free energy are zero for all *active coordinates*. To complete the geometrical description, we will use the thermal average of the inactive coordinates during the molecular-dynamics simulation.

C. Notations

Many expressions have already been proposed to evaluate the derivative of the free energy in a predefined direction.^{12-14,17-19} In this section, we will recall the expressions that will be used throughout this work. For the sake of simplicity, the expression will be given in the case where the active set contains only one coordinate ξ . It is worth noting that once the reaction path has been constructed, only one coordinate is needed because the reaction coordinate can be uniquely defined as the mass-weighted curvilinear distance from the reactants to the point of interest along the path.

We denote by \mathbf{J} the Jacobian matrix for the transformation from the Cartesian coordinates to the generalized coordinates, $\mathbf{x} \rightarrow (\mathbf{q}, \xi)$,

$$\mathbf{J} = \begin{pmatrix} \frac{\partial \mathbf{x}}{\partial \mathbf{q}} & \frac{\partial \mathbf{x}}{\partial \xi} \end{pmatrix},$$

and by $|\mathbf{J}|$ its determinant (that is the Jacobian). We also define \mathbf{J}' as the Jacobian matrix for the transformation from the mass-weighted Cartesian coordinates to the generalized coordinates.

We introduced the mass matrix $\mathbf{A}_{\mathbf{q}\xi}$ defined by $\mathbf{A}_{\mathbf{q}\xi} = \mathbf{J}' \mathbf{M} \mathbf{J}$, where \mathbf{M} is a $3N \times 3N$ diagonal matrix containing all atomic masses.

The generalized momenta are related to the velocity vector by

$$\mathbf{p}_{q\xi} = \begin{pmatrix} \mathbf{p}_q \\ p_\xi \end{pmatrix} = A_{q\xi} \begin{pmatrix} \dot{\mathbf{q}} \\ \dot{\xi} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_q \dot{\mathbf{q}} + \mathbf{B}_\xi \dot{\xi} \\ \mathbf{B}'_\xi \dot{\mathbf{q}} + C_\xi \dot{\xi} \end{pmatrix}. \quad (6)$$

For further convenience, we decompose $\mathbf{A}_{q\xi}$ and its inverse $\mathbf{A}_{q\xi}^{-1}$ into the following blocks:

$$\mathbf{A}_{q\xi} = \begin{pmatrix} \mathbf{A}_q & \mathbf{B}_\xi \\ \mathbf{B}'_\xi & C_\xi \end{pmatrix}, \quad \mathbf{A}_{q\xi}^{-1} = \begin{pmatrix} \mathbf{X}_q & \mathbf{Y}_\xi \\ \mathbf{Y}'_\xi & Z_\xi \end{pmatrix}. \quad (7)$$

Some properties of these matrices are given in Appendix A.

Last, let us write explicitly the form of Z_ξ and \mathbf{Y}_ξ ,

$$Z_\xi = \sum_{i=1}^{3N} \frac{1}{m_i} \frac{\partial \xi}{\partial x_i} \frac{\partial \xi}{\partial x_i} = \sum_{i=1}^{3N} \frac{\partial \xi}{\partial x'_i} \frac{\partial \xi}{\partial x'_i}, \quad (8)$$

$$(\mathbf{Y}_\xi)_j = \sum_{i=1}^{3N} \frac{1}{m_i} \frac{\partial q_j}{\partial x_i} \frac{\partial \xi}{\partial x_i} = \sum_{i=1}^{3N} \frac{\partial q_j}{\partial x'_i} \frac{\partial \xi}{\partial x'_i}. \quad (9)$$

D. Gradient of the free energy

The free energy $A(\xi^*)$ is related to the partition function $Q(\xi^*)$ by $A(\xi^*) = -k_B T \ln(Q(\xi^*))$. Therefore, a prerequisite to the determination of $\partial A / \partial \xi$ is the evaluation of $\partial Q / \partial \xi$.

The partition function is defined by

$$Q(\xi^*) = \int d\mathbf{q} \int d\mathbf{p}_q dp_\xi \exp(-\beta \mathbf{H}), \quad (10)$$

where \mathbf{H} is the Hamiltonian associated with our system,

$$\mathbf{H}(\mathbf{q}, \xi, \mathbf{p}_q) = \frac{1}{2} \mathbf{p}_q^t \mathbf{A}_{q\xi}^{-1} \mathbf{p}_q + V(\mathbf{q}, \xi). \quad (11)$$

However, in a molecular-dynamics simulation, constraining the reaction coordinate ξ to remain constant and equal to ξ^* implies imposing the additional constraint $\dot{\xi}=0$. Therefore, the ensemble average during the molecular-dynamics (MD) simulation is not the one needed in Eq. (10) because p_ξ is *not* sampled. When the reaction coordinate ξ is constrained to a specific value, $\xi = \xi^*$, the Hamiltonian associated with the system becomes

$$\mathbf{H}_{\xi^*}^c(\mathbf{q}, \mathbf{p}_q) = \frac{1}{2} \mathbf{p}_q^t \mathbf{A}_{q\xi^*}^{-1} \mathbf{p}_q + V(\mathbf{q}, \xi^*). \quad (12)$$

In the following, we will denote by

$$\langle \mathcal{O} \rangle_{\xi^*} = \frac{\int d\mathbf{q} \int d\mathbf{p}_q \mathcal{O} \exp(-\beta \mathbf{H}_{\xi^*}^c)}{\int d\mathbf{q} \int d\mathbf{p}_q \exp(-\beta \mathbf{H}_{\xi^*}^c)} \quad (13)$$

the average of a function \mathcal{O} over the constrained ensemble. The notation $\langle \rangle_{\xi^*}$ indicates that the sampling is done along \mathbf{p}_q and \mathbf{q} while ξ remains constant and equal to ξ^* . The average for an unconstrained simulation is

$$\langle \mathcal{O} \rangle = \frac{\int d\mathbf{q} d\xi \int d\mathbf{p}_{q\xi} \mathcal{O} \exp(-\beta \mathbf{H})}{\int d\mathbf{q} d\xi \int d\mathbf{p}_{q\xi} \exp(-\beta \mathbf{H})}. \quad (14)$$

The first step in evaluating the derivative of the free energy is to relate these two averages.

1. Blue-moon correction

Following the work of Carter *et al.*,¹¹ we can rewrite the kinetic part of the unconstrained Hamiltonian as

$$\mathbf{p}'_{q\xi} \mathbf{A}_{q\xi}^{-1} \mathbf{p}_{q\xi} = \mathbf{p}'_q \mathbf{A}_q^{-1} \mathbf{p}_q + (p_\xi + Z_\xi^{-1} \mathbf{Y}'_\xi \mathbf{p}_q)^t Z_\xi (p_\xi + Z_\xi^{-1} \mathbf{Y}'_\xi \mathbf{p}_q). \quad (15)$$

We can then rewrite the average value of an operator \mathcal{O} independent of p_ξ ,

$$\begin{aligned} & \int d\mathbf{q} \int d\mathbf{p}_{q\xi} \mathcal{O} \exp(-\beta \mathbf{H}) \\ &= \int d\mathbf{q} \int d\mathbf{p}_q \mathcal{O} \exp\left(-\beta \left[\frac{1}{2} \mathbf{p}'_q \mathbf{A}_q^{-1} \mathbf{p}_q + V(\mathbf{q}, \xi^*) \right]\right) \\ & \quad \times \int dp_\xi \exp\left[-\frac{1}{2} \beta (p_\xi + Z_\xi^{-1} \mathbf{Y}'_\xi \mathbf{p}_q)^t Z_\xi (p_\xi + Z_\xi^{-1} \mathbf{Y}'_\xi \mathbf{p}_q)\right]. \end{aligned} \quad (16)$$

Using the definition of $\mathbf{H}_{\xi^*}^c$ and the properties of Gaussian integrals [see Eq. (B1) of Appendix B], it comes

$$\begin{aligned} & \int d\mathbf{q} \int d\mathbf{p}_{q\xi} \mathcal{O} \exp(-\beta \mathbf{H}) \\ &= \int d\mathbf{q} \int d\mathbf{p}_q \mathcal{O} \exp(-\beta \mathbf{H}_{\xi^*}^c) Z_\xi^{1/2}. \end{aligned} \quad (17)$$

Thus, the relation between the constrained and unconstrained simulations reads

$$\langle \mathcal{O} \rangle = \frac{\int d\mathbf{q} \int d\mathbf{p}_q \mathcal{O} \exp(-\beta \mathbf{H}_{\xi^*}^c) \mathcal{O} Z_\xi^{1/2}}{\int d\mathbf{q} \int d\mathbf{p}_q \exp(-\beta \mathbf{H}_{\xi^*}^c) Z_\xi^{1/2}} = \frac{\langle \mathcal{O} Z_\xi^{1/2} \rangle_{\xi^*}}{\langle Z_\xi^{1/2} \rangle_{\xi^*}}, \quad (18)$$

which corresponds to the standard blue-moon correction.¹¹

Using Eq. (17), the partition functions can now be written as

$$Q(\xi^*) = \int d\mathbf{q} \int d\mathbf{p}_q \exp(-\beta \mathbf{H}_{\xi^*}^c) Z_\xi^{1/2}, \quad (19)$$

which leads to

$$\begin{aligned} & \left(\frac{\partial Q(\xi)}{\partial \xi} \right)_{\xi^*} \\ &= \int d\mathbf{q} \int d\mathbf{p}_q e^{-\beta \mathbf{H}_{\xi^*}^c} Z_\xi^{-1/2} \left(-\beta \frac{\partial \mathbf{H}_{\xi^*}^c}{\partial \xi} - \frac{1}{2} Z_\xi^{-1} \frac{\partial Z_\xi}{\partial \xi} \right). \end{aligned} \quad (20)$$

Starting from this equation, two different procedures have been proposed. In the first one, this equation is expanded using the properties of the mass matrix and of the Gaussian integrals. The second one follows more closely the philosophy of a MD simulation and tries to relate the derivative of the partition function to the force λ_ξ acting on the reaction coordinate ξ . These two procedures will be detailed in the next sections.

2. Expanding the Hamiltonian

From Eq. (12), we have

$$\frac{\partial \mathbf{H}_{\xi^*}^c}{\partial \xi} = \frac{1}{2} \mathbf{p}_q^t \frac{\partial \mathbf{A}_q^{-1}}{\partial \xi} \mathbf{p}_q + \frac{\partial V}{\partial \xi}. \quad (21)$$

Plugging Eq. (B2) of Appendix B and Eq. (21) into Eq. (20), one finds

$$\left(\frac{\partial Q(\xi)}{\partial \xi} \right)_{\xi^*} = \int d\mathbf{q} \int d\mathbf{p}_q e^{-\beta \mathbf{H}_{\xi^*}^c} Z_{\xi^*}^{-1/2} \left(-\frac{1}{2} \text{Tr} \left(\mathbf{A}_q \frac{\partial \mathbf{A}_q^{-1}}{\partial \xi} \right) - \beta \frac{\partial V}{\partial \xi} - \frac{1}{2} Z_{\xi^*}^{-1} \frac{\partial Z_{\xi^*}}{\partial \xi} \right). \quad (22)$$

Using $\text{Tr}[\mathbf{A}_q(\partial \mathbf{A}_q^{-1}/\partial \xi)] = -(\partial \ln|\mathbf{A}_q|/\partial \xi)$ and Eq. (A7), one finally gets

$$\left(\frac{\partial A(\xi)}{\partial \xi} \right)_{\xi^*} = \frac{\langle |Z_{\xi^*}|^{-1/2} ((\partial V/\partial \xi) - kT(\partial \ln|\mathbf{J}|/\partial \xi)) \rangle_{\xi^*}}{\langle |Z_{\xi^*}|^{-1/2} \rangle_{\xi^*}}. \quad (23)$$

In the case of multiple constraints, the derivative along the constraint ξ_i reads^{12,17}

$$\left(\frac{\partial A(\xi)}{\partial \xi_i} \right)_{\xi^*} = \frac{\langle |Z_{\xi^*}|^{-1/2} [(\partial V/\partial \xi_i) - kT(\partial \ln|\mathbf{J}|/\partial \xi_i)] \rangle_{\xi^*}}{\langle |Z_{\xi^*}|^{-1/2} \rangle_{\xi^*}}. \quad (24)$$

Although these formulas are exact, they are not as useful as one might have expected because they require the knowledge of the full Jacobian matrix. This in turn implies that the full set of generalized coordinates is known, which is something usually not desirable for a big molecule.

3. Relation with the Lagrange multiplier

In cases where constructing the full set of generalized coordinates \mathbf{q} is not desirable, one has to get rid of the explicit dependence of the previous equations on the Jacobian. This is what is done in the second approach; instead of relating the derivative of the constrained Hamiltonian $\mathbf{H}_{\xi^*}^c$ to the Jacobian, we will relate it to the force acting on the reaction coordinate during the simulation.

Indeed, during a MD simulation, one makes use of the ergodicity principle, and the averages are actually performed over time and not directly over the phase space. Then, in order to ensure that the reaction coordinate ξ is constant and equal to ξ^* , a modified Lagrangian is used,

$$\mathcal{L}^* = \frac{1}{2} \mathbf{v}_{\mathbf{q}\xi}^t \mathbf{A}_{\mathbf{q}\xi} \mathbf{v}_{\mathbf{q}\xi} - V(\mathbf{q}, \xi) + \lambda_\xi (\xi - \xi^*), \quad (25)$$

where \mathcal{L} is the Lagrangian of the unconstrained system and $\mathbf{v}_{\mathbf{q}\xi}$ is the velocity vector. λ_ξ is the Lagrange multiplier asso-

ciated with the reaction coordinate ξ . Its value is adjusted at each step of the simulation so that $\dot{\xi} = \dot{\xi}^*$. In practice, the SHAKE algorithm was used for our simulations.²⁰ In this algorithm, λ_ξ is adjusted to ensure that $\ddot{\xi} = 0$.

For an unconstrained simulation, the equation of motion of the reaction coordinate is

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{\xi}} \right) = \frac{\partial \mathcal{L}}{\partial \xi}, \quad (26)$$

which gives

$$C_{\xi \ddot{\xi}} = \frac{\partial \mathcal{L}}{\partial \xi} - \frac{d}{dt} (\mathbf{B}'_\xi \dot{q}) - \frac{dC_{\xi \dot{\xi}}}{dt}. \quad (27)$$

Taking into account that $\ddot{\xi} = 0$ in a constrained simulation, one has

$$C_{\xi \ddot{\xi}} = \frac{\partial \mathcal{L}^*}{\partial \xi} - \frac{d}{dt} (\mathbf{B}'_\xi \dot{q}) \quad (28)$$

$$= \frac{\partial \mathcal{L}}{\partial \xi} - \frac{d}{dt} (\mathbf{B}'_\xi \dot{q}) + \lambda_\xi. \quad (29)$$

Demanding that $\ddot{\xi} = 0$ in a constrained simulation, we have

$$\lambda_\xi = - \left\{ \frac{\partial \mathcal{L}}{\partial \xi} - \frac{d}{dt} (\mathbf{B}'_\xi \dot{q}) \right\}. \quad (30)$$

Using the definition of the Lagrangian, it comes

$$\lambda_\xi = - \left\{ \frac{1}{2} \dot{\mathbf{q}}^t \frac{\partial \mathbf{A}_q}{\partial \xi} \dot{\mathbf{q}} - \frac{\partial V}{\partial \xi} - \frac{d}{dt} (\mathbf{B}'_\xi \dot{q}) \right\} \quad (31)$$

$$= \frac{1}{2} \mathbf{p}_q^t \frac{\partial \mathbf{A}_q^{-1}}{\partial \xi} \mathbf{p}_q + \frac{\partial V}{\partial \xi} + \frac{d}{dt} (\mathbf{B}'_\xi \dot{q}) \quad (32)$$

$$= \frac{\partial \mathbf{H}_{\xi^*}^c}{\partial \xi} + \frac{d}{dt} (\mathbf{B}'_\xi \dot{q}). \quad (33)$$

Inserting Eq. (33) in the expression for the derivative of the partition function gives

$$\frac{\partial Q}{\partial \xi} = -\beta \int dp_q dq e^{-\beta \mathbf{H}_{\xi^*}^c} Z_{\xi^*}^{-1/2} \times \left(\lambda_\xi - \frac{d}{dt} (\mathbf{B}'_\xi \dot{q}) + \frac{1}{2\beta} Z_{\xi^*}^{-1} \frac{\partial Z_{\xi^*}}{\partial \xi} \right). \quad (34)$$

In order to obtain the final formula we need, we have to integrate analytically the term $(d/dt)(\mathbf{B}'_\xi \dot{q})$. This will be done by following the work of den Otter and Briels.¹³ First, we use

$$Z_{\xi^*}^{-1/2} \frac{d}{dt} (\mathbf{B}'_\xi \dot{q}) = \frac{d}{dt} (Z_{\xi^*}^{-1/2} \mathbf{B}'_\xi \dot{q}) + \frac{1}{2} \mathbf{B}'_\xi \dot{q} Z_{\xi^*}^{-1/2} Z_{\xi^*}^{-1} \frac{dZ_{\xi^*}}{dt}. \quad (35)$$

Using the ergodicity principle, we have²¹

$$\int dp_q dq e^{-\beta H_\xi^c} \frac{d}{dt} (Z_\xi^{-1/2} \mathbf{B}_\xi^t \dot{q})$$

$$= \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^\tau dt \frac{d}{dt} (Z_\xi^{-1/2} \mathbf{B}_\xi^t \dot{q}) = 0. \quad (36)$$

As $\dot{\xi}=0$, we have $dZ_\xi/dt = \dot{\mathbf{q}}^t \partial Z_\xi / \partial \mathbf{q}$, and thus

$$\int dp_q dq e^{-\beta H_\xi^c} \mathbf{B}_\xi^t \dot{q} \frac{dZ_\xi}{dt} = \int dp_q dq e^{-\beta H_\xi^c} \mathbf{B}_\xi^t \dot{\mathbf{q}} \dot{\mathbf{q}}^t \frac{\partial Z_\xi}{\partial \mathbf{q}}.$$

Using Eqs. (B2) and (A3), we get

$$\int dp_q dq e^{-\beta H_\xi^c} \mathbf{B}_\xi^t \dot{q} \frac{dZ_\xi}{dt}$$

$$= \int dp_q dq e^{-\beta H_\xi^c} \frac{1}{\beta} \mathbf{B}_\xi^t \mathbf{A}_q^{-1} \frac{\partial Z_\xi}{\partial \mathbf{q}}$$

$$= \int dp_q dq e^{-\beta H_\xi^c} \left(-\frac{1}{\beta} Z_\xi^{-1} \mathbf{Y}_\xi^t \frac{\partial Z_\xi}{\partial \mathbf{q}} \right). \quad (37)$$

Last,

$$\frac{\partial Z_\xi}{\partial x_i} = \sum_j \frac{\partial Z_\xi}{\partial q_j} \frac{\partial q_j}{\partial x_i} + \frac{\partial Z_\xi}{\partial \xi} \frac{\partial \xi}{\partial x_i}$$

and then

$$\mathbf{Y}_\xi^t \frac{\partial Z_\xi}{\partial \mathbf{q}} = -Z_\xi \frac{\partial Z_\xi}{\partial \xi} + \frac{\partial \xi}{\partial \mathbf{x}'} \frac{\partial Z_\xi}{\partial \mathbf{x}'}. \quad (38)$$

Collecting Eqs. (34)–(38), we have shown

$$\int dp_q dq e^{-\beta H_\xi^c} Z_\xi^{-1/2} \frac{d}{dt} (\mathbf{B}_\xi^t \dot{q})$$

$$= \int dp_q dq e^{-\beta H_\xi^c} Z_\xi^{-1/2} \left(\frac{1}{2\beta} Z_\xi^{-1} \frac{\partial Z_\xi}{\partial \xi} - \frac{1}{2\beta Z_\xi^2} \frac{\partial \xi}{\partial \mathbf{x}'} \frac{\partial Z_\xi}{\partial \mathbf{x}'} \right). \quad (39)$$

Using this last equation, we recover the general formula derived by Sprik and Ciccotti,¹² den Otter and Briels,¹³ and Darve and Pohorille,¹⁴

$$\frac{\partial A}{\partial \xi} = \frac{1}{\langle Z_\xi^{-1/2} \rangle_{\xi^*}} \left\langle Z_\xi^{-1/2} \left(\lambda_\xi + \frac{1}{2\beta Z_\xi^2} \left\{ \sum_{i=1}^{i=3N} \frac{\partial \xi}{\partial x'_i} \frac{\partial Z_\xi}{\partial x'_i} \right\} \right) \right\rangle_{\xi^*}. \quad (40)$$

This equation is readily evaluated during a simulation because all the terms depend only on known quantities such as ξ and Z_ξ .

When many coordinates are constrained, the derivative along ξ_k reads^{17,18}

$$\frac{\partial A}{\partial \xi_k} = \frac{1}{\langle |Z_\xi|^{-1/2} \rangle_{\xi^*}} \left\langle |Z_{\xi_k}|^{-1/2} \left(\lambda_{\xi_k} + \frac{kT}{2|Z_\xi|} \right. \right.$$

$$\left. \left. \times \sum_{j=1}^{j=r} (Z_\xi^{-1})_{kj} \left\{ \sum_{i=1}^{i=3N} \frac{\partial \xi_j}{\partial x'_i} \frac{\partial |Z_\xi|}{\partial x'_i} \right\} \right) \right\rangle_{\xi^*}. \quad (41)$$

4. Comparing the two procedures

At first glance Eqs. (23) and (40) seem very different and it is not obvious that they both refer to the same quantity. The connection between these formulas is more clearly seen if one rewrites their numerators,

$$\frac{\partial A}{\partial \xi} \propto \left\langle Z_\xi^{-1/2} \left(\left[\frac{1}{2} \mathbf{p}_q^t \frac{\partial \mathbf{A}_q^{-1}}{\partial \xi} \mathbf{p}_q + \frac{\partial V}{\partial \xi} \right] + \frac{1}{2\beta} Z_\xi^{-1} \frac{\partial Z_\xi}{\partial \xi} \right) \right\rangle_{\xi^*}, \quad (23')$$

$$\frac{\partial A}{\partial \xi} \propto \left\langle Z_\xi^{-1/2} \left(\left[\frac{1}{2} \mathbf{p}_q^t \frac{\partial \mathbf{A}_q^{-1}}{\partial \xi} \mathbf{p}_q + \frac{\partial V}{\partial \xi} \right] + \frac{d}{dt} (\mathbf{B}_\xi^t \dot{q}) \right. \right.$$

$$\left. \left. + \frac{1}{2\beta Z_\xi^2} \left\{ \sum_{i=1}^{i=3N} \frac{\partial \xi}{\partial x'_i} \frac{\partial Z_\xi}{\partial x'_i} \right\} \right) \right\rangle_{\xi^*}. \quad (40')$$

Then, inserting Eq. (39) into Eq. (40') leads to Eq. (23').

III. PRACTICAL CONSIDERATIONS

The exact formulas for evaluating the free-energy derivatives have been recalled in Sec. II. However, actually using them in a molecular-dynamics simulation requires to tackle two problems. The first one is related to the chemical system under study: one has to choose the coordinates that belong to the active space, so that all the relevant coordinates are considered. The second problem is more technical; the previous formulas look quite complicated to evaluate and one might wonder how to compute them efficiently. We will first propose a way to decide whether a coordinate should be included into the active space. Then, we will show that in many cases, the previous formulas can be greatly simplified.

A. Monitoring the active space

The first step of a molecular-dynamics simulation aiming at calculating the change in the free energy along a minimum-energy path is to divide the degrees of freedom into the active and inactive coordinates. Of course, it is possible to consider all coordinates as active and to calculate the complete set of derivatives using the previous formulas. However, that would require launching essentially one MD simulation for each degree of freedom; such a procedure would be quite expensive. Moreover, in contrast with the reaction path on the potential-energy surface, not all degrees of freedom are required; many degrees of freedom will move in a nearly harmonic well. As a consequence, the derivative of the free energy along these modes is small, and one can safely discard them. Such a case is observed in the forthcoming application for the CH distances of the ethylene molecule; the CH bond length slightly increases as the cyclopropane is formed, but at each point of the reaction path, they move in a quasiharmonic well. Hence, their contribution to the direction of the minimum-free-energy path can be neglected.

Therefore, one must select only a restricted set of coordinates. However, for complex systems undergoing a reaction, chemical intuition might not be sufficient. We propose here a way to construct the active and the inactive sets and to

monitor this separation along the construction of the path. The value of the derivative of the free energy will be used as a *quantitative* criterion to discriminate between active and inactive coordinates. Ideally, this derivative should be zero for an inactive coordinate. However, in practice, the actual reaction coordinate has nonzero component on all coordinates, including the inactive coordinates. The criterion will then be to compare the derivative of the free energy along an inactive coordinate to a given threshold. If it is bigger than this threshold then one must consider incorporating q_n into the active set. Moreover, we will show that one can use the data of a simulation to estimate the derivative of the free energy along an unconstrained coordinate.

Let us consider a system for which the beginning of the minimum-free-energy path has already been constructed as a set of k points (ξ^0, \dots, ξ^k). The purpose of this section is to detail the procedure to find the next point, $k+1$, of the path. By definition of the reaction path, this point can be obtained by following the gradient of the free energy along a small distance ds ,

$$\xi^{k+1} = \xi^k + \frac{\mathbf{grad} A}{|\mathbf{grad} A|} ds.$$

The previous formulas can be applied to the data of the MD simulation conducted at point ξ^k to calculate the derivatives of the free energy along the active coordinates: $\{\partial A / \partial \xi_i\}_{i=1, \dots, r}$. Before constructing the point $k+1$, we should update the active and inactive sets; if the derivative of the free energy along an active coordinate is zero, then this coordinate should be taken out of the active set. The reciprocal question is then: Shall we include any of the inactive q_i variable into the active set in order to better describe the rest of the path? Let us consider the coordinate q_n as an example. We will now give the expression of the derivative of the free energy along this inactive coordinate q_n . The previous formulas cannot be used directly because they necessitate that the coordinate under study is *constrained* during the simulation. Let us denote by q_n^k the particular value of q_n at which we want to calculate the derivative of the free energy along q_n : $(\partial A / \partial q_n)_{q_n^k, \xi}$. In order to evaluate this quantity, one can follow the same procedure as before, and finds

$$\begin{aligned} & \left(\frac{\partial A(q_n, \xi)}{\partial q_n} \right)_{q_n=q_n^k, \xi} \\ &= \frac{\langle |Z_\xi|^{-1/2} \delta(q_n - q_n^k) ((\partial V / \partial q_n) - kT(\partial \ln |\mathbf{J}| / \partial q_n)) \rangle_\xi}{\langle |Z_\xi|^{-1/2} \delta(q_n - q_n^k) \rangle_\xi}. \end{aligned} \quad (42)$$

In this expression, we have explicitly included the delta function $\delta(q_n - q_n^k)$ which ensures that the average corresponds to a conditional sampling of the phase space in which q_n is equal to q_n^k .

If we want to avoid calculating the full set of the generalized coordinates, we have to derive an expression similar to Eq. (40) for the coordinate q_n . An important point here is that, in contrast with ξ , q_n was not constrained during the simulation. Therefore, the simulation data contain the sampling over p_{q_n} that would have been missing in a simulation

where *both* ξ and q_n would have been constrained.

The expression for the derivative of the free energy along this unconstrained coordinate q_n will be done in two steps. In the first step, we will establish the expression for the derivative of the free energy along an unconstrained reaction coordinate, that is during a simulation without the constraint $\dot{\xi}=0$. Then, we will use the blue-moon relation to obtain the expression of the free energy along an inactive, unconstrained, coordinate q_n during a simulation with a constrained reaction coordinate ξ .

First, using Eq. (6), we have

$$\begin{pmatrix} \dot{\mathbf{q}} \\ \dot{\xi} \end{pmatrix} = \mathbf{A}_{q\xi}^{-1} \begin{pmatrix} \mathbf{p}_q \\ p_\xi \end{pmatrix} = \begin{cases} \mathbf{X}_q \mathbf{p}_q + \mathbf{Y}_\xi p_\xi \\ \mathbf{Y}_\xi^T \mathbf{p}_q + Z_\xi p_\xi \end{cases} \quad (43)$$

Using this equation and Eq. (A3), we can rewrite p_ξ , as

$$p_\xi = -Z_\xi^{-1} \mathbf{Y}_\xi^T \mathbf{p}_q + Z_\xi^{-1} \dot{\xi} = \mathbf{B}_\xi^T \mathbf{A}_q^{-1} \mathbf{p}_q + Z_\xi^{-1} \dot{\xi}. \quad (44)$$

Remembering $\partial \mathbf{H} / \partial \xi = -(d/dt)(p_\xi)$, one finds

$$\begin{aligned} \frac{\partial Q}{\partial \xi} &= \beta \left\{ \int dp_q dq dp_\xi e^{-\beta H} \frac{d}{dt} (\mathbf{B}_\xi^T \mathbf{A}_q^{-1} \mathbf{p}_q) \right. \\ &\quad \left. + \int dp_q dq dp_\xi e^{-\beta H} \frac{d}{dt} (Z_\xi^{-1} \dot{\xi}) \right\}. \end{aligned} \quad (45)$$

The ergodicity principle allows us to say that the first integral is zero. Therefore, we have

$$\begin{aligned} \frac{\partial Q}{\partial \xi} &= \beta \int dp_q dq dp_\xi e^{-\beta H} \frac{d}{dt} (Z_\xi^{-1} \dot{\xi}) \\ &= \beta \int dp_q dq dp_\xi e^{-\beta H} \left(Z_\xi^{-1} \ddot{\xi} - Z_\xi^{-2} \dot{\xi} \frac{dZ_\xi}{dt} \right). \end{aligned} \quad (46)$$

We shall now pursue the derivation following a procedure similar to that of Pohorille and co-workers.^{14,18}

The second term of Eq. (46) reads

$$\begin{aligned} & \int dp_q dq dp_\xi e^{-\beta H} Z_\xi^{-2} \dot{\xi} \frac{dZ_\xi}{dt} \\ &= \int dp_q dq dp_\xi e^{-\beta H} Z_\xi^{-2} \dot{\xi} \frac{\partial Z_\xi}{\partial \mathbf{x}'} \mathbf{p}_x' \\ &= \int dp_q dq dp_\xi e^{-\beta H} Z_\xi^{-2} \dot{\xi} \frac{\partial Z_\xi}{\partial \mathbf{x}'} \\ &\quad \times (\mathbf{J}_\xi'^{-1} p_\xi + \mathbf{J}_q'^{-1} \mathbf{p}_q). \end{aligned} \quad (47)$$

We now introduce a new set of generalized momenta $(\tilde{\mathbf{p}}_q, \tilde{p}_\xi)$ defined by

$$\begin{aligned} \tilde{\mathbf{p}}_q &= \mathbf{p}_q, \\ \tilde{p}_\xi &= Z_\xi^{-1} \dot{\xi} = p_\xi + Z_\xi^{-1} \mathbf{Y}_\xi^T \mathbf{p}_q. \end{aligned} \quad (48)$$

This transformation is valid as it is invertible. Moreover, the Jacobian $|\tilde{\mathbf{J}}|$ associated with it is equal to 1. With these new variables, Eq. (15) reads

$$\frac{1}{2} \mathbf{p}_q^T \mathbf{A}_{q\xi}^{-1} \mathbf{p}_q + \frac{1}{2} \tilde{p}_\xi^2 Z_\xi^{-2} \tilde{p}_\xi. \quad (49)$$

Equation (47) then reads

$$\begin{aligned}
& \int dp_q dq dp_\xi e^{-\beta H} Z_\xi^{-2} \xi \frac{dZ_\xi}{dt} \\
&= \int dq e^{-\beta V} \int d\tilde{p}_q e^{-1/2\beta\tilde{p}_q^T \mathbf{A}_q^{-1} \tilde{p}_q} \\
&\quad \times \int d\tilde{p}_\xi |\tilde{\mathbf{J}}| e^{-1/2\beta\tilde{p}_\xi^T Z_\xi^{-1} \tilde{p}_\xi} \frac{\partial Z_\xi}{\partial \mathbf{x}'} \\
&\quad \times [\mathbf{J}_\xi'^{-1} \tilde{p}_\xi + (\mathbf{J}_q'^{-1} - Z_\xi^{-1} \mathbf{Y}_\xi^T) \tilde{\mathbf{p}}_q]. \quad (50)
\end{aligned}$$

As the Hamiltonian is even in \tilde{p}_ξ , after integration over \tilde{p}_ξ , only even terms remain,

$$\begin{aligned}
& \int dq dp_q dp_\xi e^{-\beta H} Z_\xi^{-2} \xi \frac{dZ_\xi}{dt} \\
&= \int dq d\tilde{p}_q d\tilde{p}_\xi |\tilde{\mathbf{J}}| e^{-\beta H} Z_\xi^{-1} \tilde{p}_\xi^2 \frac{\partial Z_\xi}{\partial \mathbf{x}'} \mathbf{J}_\xi'^{-1}. \quad (51)
\end{aligned}$$

Making use of Eq. (B2) to integrate over \tilde{p}_ξ , it comes

$$\begin{aligned}
& \int dq dp_q dp_\xi e^{-\beta H} Z_\xi^{-2} \xi \frac{dZ_\xi}{dt} \\
&= \int dq d\tilde{q} d\tilde{p}_\xi |\tilde{\mathbf{J}}| e^{-\beta H} \frac{1}{\beta Z_\xi^2} \frac{\partial Z_\xi}{\partial \mathbf{x}'} \mathbf{J}_\xi'^{-1} \quad (52)
\end{aligned}$$

$$= \int dq dp_q dp_\xi e^{-\beta H} \frac{1}{\beta Z_\xi^2} \frac{\partial Z_\xi}{\partial \mathbf{x}'} \frac{\partial \xi}{\partial \mathbf{x}'}. \quad (53)$$

Collecting the previous equations leads to

$$\frac{\partial A}{\partial \xi} = \frac{1}{Q} \int dp_q dq dp_\xi e^{-\beta H} \left(-Z_\xi \ddot{\xi} + \frac{1}{\beta Z_\xi^2} \frac{\partial Z_\xi}{\partial \mathbf{x}'} \frac{\partial \xi}{\partial \mathbf{x}'} \right). \quad (54)$$

This equation is the same as the one obtained by Darve *et al.* [Eq. (24) of Ref. 18], but we arrived at it with different assumptions.

We now apply this equation to the case of a simulation where ξ is constrained but q_n is not. Applying the blue-moon relation [see Eq. (18)], one finds^{14,18}

$$\begin{aligned}
\left(\frac{\partial A(q_n, \xi)}{\partial q_n} \right)_{q_n, \xi} &= \frac{1}{\langle |Z_\xi|^{-1/2} \delta_{q_n} \rangle_\xi} \left\langle |Z_\xi|^{-1/2} \delta_{q_n} \left(-Z_{q_n}^{-1} \ddot{q}_n \right. \right. \\
&\quad \left. \left. + \frac{kT}{Z_{q_n}^2} \left\{ \sum_{i=1}^{i=3N} \frac{\partial q_n}{\partial x_i} \frac{\partial Z_{q_n}}{\partial x_i} \right\} \right) \right\rangle_\xi, \quad (55)
\end{aligned}$$

where δ_{q_n} stands for $\delta(q_n - q_n^k)$, and Z_{q_n} is the part of the inverse mass matrix corresponding to q_n : $Z_{q_n} = \sum_i (1/m_i) (\partial q_n / \partial x_i) (\partial q_n / \partial x_i)$. As already noted,¹⁸ this expression is slightly different from Eq. (40), despite the fact that they both relate the derivative of the free energy along q_n to the force acting on q_n during the MD simulation. The origin of this difference comes from the nature of the sampling used to estimate the two expressions: Eq. (55) corresponds to a conditional average performed during a simulation in which q_n was *not* constrained whereas Eq. (40) corresponds to a simulation in which both ξ and q_n were constrained. It is worth stressing here that both expressions are valid but correspond to different contexts. Moreover, as long as the sampling

along q_n is sufficient, evaluating the derivative of the free energy along q_n during a simulation in which only ξ is constrained using the conditional averaging of Eq. (55) will lead to the same numerical result as Eq. (40) using a MD simulation with both ξ and q_n constrained.

Practical use of this equation is described in Appendix D.

To conclude this section, we detail one possible use of Eqs. (40) and (55) for the construction of a reaction path on the free-energy surface. We suppose as before that the active set is known and denoted by ξ , and that the path is partially constructed, up to the point k corresponding to the value $\xi^k = \{\xi_i^k\}_{i=1, \dots, r}$ of the active coordinates. To find the next point ξ^{k+1} , one should

- (1) launch a simulation while constraining each active coordinates ξ_i ($i = 1, \dots, r$) to be constant, and equal to ξ_i^k ;
- (2) use formula (40) (or (23)) to compute the derivative of the free energy along the active coordinates, $\partial \mathbf{A} / \partial \xi$;
- (3) take out of the active set the coordinates corresponding to zero derivatives;
- (4) then, use the *simulation data* and Eq. (55) to evaluate the derivative of the free energy along the inactive coordinates $\partial \mathbf{A} / \partial \mathbf{q}$, then include into the active set the coordinates associated to non-zero derivatives;
- (5) recollect *all* derivatives to obtain the full gradient, $\mathbf{grad} A = \{\partial \mathbf{A}' / \partial \xi, \partial \mathbf{A}' / \partial \mathbf{q}\}'$; and
- (6) construct the next point of the path by following the gradient of the free energy along a small distance ds ,

$$\xi^{k+1} = \xi^k + ds \frac{\mathbf{grad} A}{|\mathbf{grad} A|}.$$

Such a procedure has been used to study the addition of CCl_2 to ethylene and is explained in Sec. IV.

B. Special types of constraints

The preceding equations [(23), (24), (42), and (55)] have been derived without considering the actual form of the constraints. Therefore, they are general but they look quite difficult to compute efficiently. Despite the fact that a general procedure has already been given by Darve and Pohorille,¹⁴ we would like to show here that many common constraints lead to considerable simplifications of the previous expressions. Four cases are described here in which Z_ξ is a constant. In those cases, the previous equations become

$$\frac{\partial A}{\partial \xi} = \langle \lambda_\xi \rangle_{\xi^*}, \quad (56)$$

$$\frac{\partial A}{\partial q_n} = \frac{\langle \delta_{q_n} (-Z_{q_n} \ddot{q}_n) \rangle_{\xi^*}}{\langle \delta_{q_n} \rangle_{\xi^*}}. \quad (57)$$

1. Bond distance

If the reaction coordinate is chosen to be the bond distance between atoms i and j , we have¹⁹

$$\xi = d_{ij} = \sqrt{(x_i - x_j)^2 + (x_{i+1} - x_{j+1})^2 + (x_{i+2} - x_{j+2})^2},$$

$$Z_\xi = \left(\frac{1}{m_i} + \frac{1}{m_j} \right). \quad (58)$$

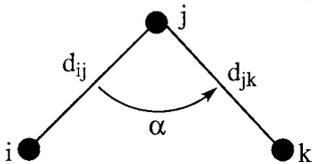
Care must be taken when constraining many bond distances. In this case, Z_ξ might no longer be a constant, and the above simplifications should not be applied blindly. Let us consider a simulation with two constrained bond distances. Two cases can be met depending on whether these two bonds share a common atom or not.

We consider first the case where the two bonds do *not* share a common atom, for example, bonds between atoms i and j and between k and l . The matrix Z_ξ is

$$\mathbf{Z}_\xi = \begin{pmatrix} Z_{d_{ij}} & 0 \\ 0 & Z_{d_{kl}} \end{pmatrix}, \quad (59)$$

with $Z_{d_{ij}} = 1/m_i + 1/m_j$ and $Z_{d_{kl}} = 1/m_k + 1/m_l$. As a consequence, $|\mathbf{Z}_\xi|$ is equal to $|Z_{d_{ij}}||Z_{d_{kl}}|$ and is a constant; one can use simplified Eqs. (56) and (57).

Let us consider now the case where the constrained bonds share a common atom, for example, bonds between atoms i and j and between atoms j and k . We also denote by α as the angle between the three atoms (see Scheme 1).



The matrix Z_ξ reads

$$\mathbf{Z}_\xi = \begin{pmatrix} Z_{d_{ij}} & \cos \alpha/m_j \\ \cos \alpha/m_j & Z_{d_{jk}} \end{pmatrix}. \quad (60)$$

As a consequence, $|\mathbf{Z}_\xi|$ is no longer a constant and Eqs. (41) and (55) must be used.

Another commonly used reaction coordinate is the difference of two distances. Once again, one has to consider the case where the two distances share a common atom or not. When the two bonds are not sharing any atom, i.e., when they are involving atoms i , j , k , and l , we have

$$\xi = d_{ij} - d_{kl}, \quad (61)$$

$$Z_\xi = \frac{1}{m_i} + \frac{1}{m_j} + \frac{1}{m_k} + \frac{1}{m_l}.$$

On the other hand, when the two distances share a common atom j , the previous equations become

$$\xi = d_{ij} - d_{jk}, \quad (62)$$

$$Z_\xi = \frac{1}{m_i} + \frac{1}{m_k} + \frac{2}{m_j}(1 - \cos \alpha).$$

So that Z_ξ is not a constant and the correction terms should be evaluated explicitly.

2. Generalized distance

Another quite common reaction coordinate is the mass-weighted distance between one reference geometry and the current geometry of the system.^{19,22} We note $\{y_i; i = 1, \dots, 3N\}$ as the Cartesian coordinates of the reference geometry. We have

$$\xi = \sqrt{\sum_i m_i (x_i - y_i)^2}, \quad (63)$$

$$Z_\xi = 1.$$

3. Bond angle

We now consider constraining the angle α between the atoms i , j , and k , that is the angle between the bonds ij and jk . We denote by d_{ij} , d_{jk} , and d_{ik} as the distances between these atoms. With these notations, Z_ξ reads²³

$$Z_\xi = Z_\alpha = \frac{1}{m_i d_{ij}^2} + \frac{1}{m_k d_{jk}^2} + \frac{d_{ik}^2}{m_j d_{ij}^2 d_{jk}^2}. \quad (64)$$

Therefore, Z_ξ is not a constant and one must use the complete formulas. However, if the bond distances d_{ij} and d_{jk} are also constrained, the formulas can be simplified. When considering a simulation, with α , d_{ij} , and d_{jk} all constrained, Z_ξ reads²³

$$\mathbf{Z}_\xi = \begin{pmatrix} Z_\alpha & -(\sin \alpha/m_j d_{jk}) & -\sin \alpha/m_j d_{ij} \\ -(\sin \alpha/m_j d_{jk}) & Z_{d_{ij}} & \cos \alpha/m_j \\ -(\sin \alpha/m_j d_{ij}) & \cos \alpha/m_j & Z_{d_{jk}} \end{pmatrix}. \quad (65)$$

Developing $|\mathbf{Z}_\xi|$, one finds

$$|\mathbf{Z}_\xi| = Z_\alpha \left(\frac{1}{m_i m_k} + \frac{1}{m_i m_j} + \frac{1}{m_j m_k} \right). \quad (66)$$

Thus, the derivatives $(\partial/\partial x_i)|\mathbf{Z}_\xi|$ of $|\mathbf{Z}_\xi|$ are *not* zero but $|\mathbf{Z}_\xi|$ is constant during a simulation. Using this fact, Eq. (41) reads

$$\frac{\partial A}{\partial \xi_k} = \frac{1}{Q_\xi^c} \left(\langle \lambda_{\xi_k} \rangle + \frac{k\langle T \rangle}{2Z_\alpha} B_k \right), \quad (67)$$

with

$$B_k = \sum_{j=1}^{j=r} (\mathbf{Z}_\xi^{-1})_{kj} \left\{ \sum_{i=1}^{i=3N} \frac{\partial \xi_k}{\partial x_i} \frac{\partial Z_\alpha}{\partial x_i} \right\}. \quad (68)$$

As α , d_{ij} , and d_{jk} are all constrained, Z_α and thus B_k are easily computed during (or after) the simulation. We conclude the discussion of this case by noting that despite the fact that one should take the corrective terms $(k\langle T \rangle/2Z_\alpha)B_k$ into account, they are of the order of magnitude of some tenth of kT and thus one might wonder if they are negligible or not. This point will be discussed in greater details in Sec. IV.

4. Linear constraint

Linear constraint can be written as $\xi = \sum_i a_i x_i$ and thus, $Z_\xi = \sum_i m_i^{-1} a_i^2$ is a constant.

This type of constraint has already been used in order to calculate the free-energy change along the intrinsic reaction path constructed on the PES.²⁴

These constraints are of considerable practical importance. First, in a simulation, it is quite common to constrain also the global rotation and the global translation of the molecule. Equations similar to the previous equations [(23) and (42)] can be derived by replacing $|Z_\xi|$ by $|Z_{\xi, T, R}|$. By construction, the global rotation and translation are orthogonal to all internal coordinates. Therefore, we have $=|Z_\xi| |Z_{T, R}|$.

We show in Appendix C that constraining the overall rotation and translation corresponds to applying six additional *linear constraints* on the system.

As a consequence, the term $|Z_{T, R}|$ is a constant that can be ignored when calculating the derivatives of the free energy, leading to the following formulas:

$$\left(\frac{\partial A}{\partial \xi} \right)_{\xi^*} = \frac{\langle Z_\xi^{-1/2} [(\partial V / \partial \xi) - kT(\partial \ln |J| / \partial \xi)] \rangle_{\xi^*, T, R}}{\langle Z_\xi^{-1/2} \rangle_{\xi^*, T, R}}, \quad (69)$$

$$\frac{\partial A}{\partial \xi} = \frac{1}{\langle Z_\xi^{-1/2} \rangle_{\xi^*}} \left\langle Z_\xi^{-1/2} \left(\lambda_\xi + \frac{1}{2\beta Z_\xi^2} \times \left[\sum_{i=1}^{i=3N} \frac{\partial \xi}{\partial x'_i} \frac{\partial Z_\xi}{\partial x'_i} \right] \right) \right\rangle_{\xi^*, T, R}. \quad (70)$$

Second, let us consider a simulation with many active coordinates, *all* described by linear constraints (which may include the translation and rotation constraints). Let us denote by $\{\xi_i; i=1, \dots, r\}$ the r linear constraints applied during the simulation. We have

$$\xi_i = \sum_{j=1}^{j=3N} c_{ij} x_j \quad \text{for } i \text{ in } 1, \dots, r. \quad (71)$$

The Lagrangian associated with this simulation is

$$\mathcal{L}^* = \frac{1}{2} \dot{\mathbf{x}}^t \mathbf{M} \dot{\mathbf{x}} - V(\mathbf{x}) + \sum_{i=1}^r \lambda_i (\xi_i - \xi_i^*). \quad (72)$$

The equations of motion are then

$$m_j \ddot{x}_j = - \frac{\partial V}{\partial x_j} + \sum_{i=1}^{i=r} \lambda_i \frac{\partial \xi_i}{\partial x_j} \quad \text{for } j \text{ in } 1, \dots, 3N. \quad (73)$$

Demanding that $\ddot{\xi}_i = 0$ for all i in $1, \dots, r$ leads to a set of coupled linear equations,

$$\sum_{j=1}^{3N} - \frac{1}{m_j} \frac{\partial V}{\partial x_j} c_{ij} + \sum_{k=1}^{k=r} \lambda_k \sum_{j=1}^{j=3N} \frac{c_{kj} c_{ij}}{m_j} = 0 \quad \text{for } i \text{ in } 1, \dots, r. \quad (74)$$

To find the forces acting on the constraints, i.e., to find the values of all λ_k , we have to solve this linear system. We define

$$D_i = \sum_j \frac{1}{m_j} \frac{\partial V}{\partial x_j} c_{ij} \quad \text{for } i \text{ in } 1, \dots, r. \quad (75)$$

Using the definition of Z_ξ , it is easily seen that the term $\sum_{j=1}^{j=3N} (c_{kj} c_{ij} / m_j)$ corresponds to the element $[Z_\xi]_{ik}$. In matrix notation, Eq. (74) reads

$$\mathbf{Z}_\xi \boldsymbol{\lambda} = \mathbf{D}, \quad (76)$$

which is easily solved by inverting the matrix \mathbf{Z}_ξ which depends only on the definition of the linear constraints ξ_i . This last expression can be further simplified when the constraints are expressed in the mass-weighted Cartesian coordinates frame,

$$\xi_i = \sum_{j=1}^{j=3N} \frac{c_{ij}}{\sqrt{m_j}} x'_j \quad \text{for } i \text{ in } 1, \dots, r. \quad (77)$$

Requesting that the linear constraints form an orthonormal set in the mass-weighted basis leads to

$$\langle \xi_i | \xi_k \rangle = \sum_{j=1}^{j=3N} \frac{c_{ij} c_{kj}}{m_j} = \delta_{ik}. \quad (78)$$

Therefore, in the case of orthogonal constraints, the inverse of the mass matrix Z_ξ reduces to the identity matrix and the previous equations [(75) and (76)] become

$$\lambda_i = \sum_{j=1}^{j=3N} \frac{1}{m_j} \frac{\partial V}{\partial x_j} c_{ij}. \quad (79)$$

Moreover, Eq. (41), that should be used to calculate the free-energy derivatives in the case of the multiple active coordinates, reduces to its simplest form, similar to (56):

$$\frac{\partial A}{\partial \xi_k} = \langle \lambda_{\xi_k} \rangle_{\xi^*} \quad \text{for } k \text{ in } 1, \dots, r. \quad (80)$$

This illustrates one convenient property of orthogonal linear constraints; they are all decoupled which leads to considerable simplification for the calculation of the free-energy derivatives.

However, the most interesting aspect of these constraints appears when *all* generalized coordinates are linear constraints,

$$\xi_i = \sum_{j=1}^{j=3N} \frac{c_{ij}}{\sqrt{m_j}} x'_j \quad \text{for } i \text{ in } 1, \dots, r, \quad (81)$$

$$q_k = \sum_{j=1}^{j=3N} \frac{c_{kj}}{\sqrt{m_j}} x'_j \quad \text{for } k \text{ in } r+1, \dots, 3N. \quad (82)$$

Using the previous equations, one finds that the force acting on an inactive coordinate q_k during the MD simulation is

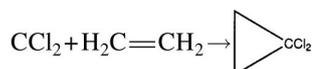
$$F_{q_k} = \sum_{j=1}^{j=3N} \frac{1}{m_j} \frac{\partial V}{\partial x_j} c_{kj} \quad \text{for } k \text{ in } r+1, \dots, 3N. \quad (83)$$

Comparing Eqs. (79) and (83) shows that the effect of the Lagrange multiplier λ_i is to exactly compensate the force acting on the active coordinate ξ_i during the MD simulation, thus ensuring that it remains constant.

These last equations show the advantage of linear constraints over the other constraints: one can estimate the gradient of the free energy for the full set of the generalized coordinates by analyzing the data of *one* MD simulation. The only condition is to have sufficient sampling along the inactive coordinates. Therefore, one can launch a simulation with a small active set, ideally comprising only the reaction coordinate, without losing any information.

IV. APPLICATION

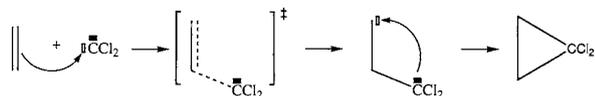
It is worth noticing that once we have an expression from which to calculate the free-energy derivatives, we can apply the same algorithms as those used in quantum chemistry in connection with potential-energy gradients. For example, it is possible to optimize a structure directly on the free-energy surface in the subset of the active coordinates. It is further possible to find a transition state along a path, to calculate the Hessian by finite difference, and thus to characterize the structures anywhere on the path. This will be applied in this section to the addition of CCl_2 to ethylene:



We will first optimize the structure of the transition state and the product at 300 K and compare the geometrical parameters to those of the 0-K geometries. We will then focus on constructing the minimum-free-energy path at 0 and 300 K.

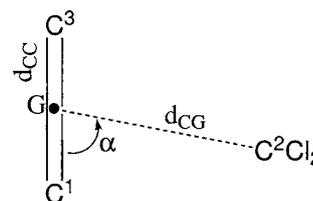
This reaction has already theoretically been studied in our group,²⁵ as well as in other groups.^{26,27} In particular, possible deficiencies of the density-functional theory (DFT) methods to describe the long-range interactions have already been stressed. However, our goal here is to construct the reaction path on the free-energy surface for this reaction and to compare it with the path obtained with a predefined reaction coordinate. In such a comparison, the accuracy of the DFT is not the main issue.

Previous studies have shown that the reaction proceeds in two steps: the first phase corresponds to the electrophilic addition of the carbene to *one* of the carbons making up the double bond. This phase proceeds through a transition state that has been optimized. The final phase is a nucleophilic attack on the second carbon of the double bond to close the cycle. This phase proceeds without any barrier, directly after the first one (see Scheme 2).



As already noticed,²⁵ the distance between the center of the double bond and the carbon of the carbene is a good reaction coordinate for the first phase, but is not sufficient to accurately describe the second phase. Therefore, we have decided to include three coordinates into the active set: d_{CC} as the ethylene CC bond distance, d_{CG} as the distance between the carbon atom of the carbene and the center of the

double bond denoted by G , and α as the angle between the double bond and this last distance. This is depicted on Scheme 3.



For further reference, the previous study using only the d_{CG} distance as a reaction coordinate will be referred to as the 1D study by opposition to the present study that uses a three-coordinate active set and will thus be referred to as the three-dimensional (3D) scheme.

The \mathbf{Z}_ξ matrix corresponding to the constraints d_{CG} , d_{CC} , and α reads

$$\mathbf{Z}_\xi = \begin{pmatrix} Z_{d_{\text{CG}}} & 0 & 0 \\ 0 & Z_{d_{\text{CC}}} & 0 \\ 0 & 0 & Z_\alpha \end{pmatrix}. \quad (84)$$

This form of the \mathbf{Z}_ξ matrix is quite different from that obtained in Eq. (65) for a system of two bonds sharing a common atom. This comes from the fact that the elements of this matrix are defined with respect to the coordinates of the atoms C^1 , C^2 , and C^3 whereas α and d_{CG} are defined with respect to C^1 , C^3 , and G . As G is the center of mass of C^1C^3 , it plays a symmetric role in the expressions of the elements of the \mathbf{Z}_ξ matrix, leading to compensating terms which sum up to zero. As an example, let us consider the off-diagonal term $Z_{d_{\text{CC}}d_{\text{CG}}}$ between the double bond C^1C^3 and the distance C^2G . Simple algebra gives

$$\frac{\partial d_{\text{CC}}}{\partial \mathbf{x}_1} = -\frac{\partial d_{\text{CC}}}{\partial \mathbf{x}_3}, \quad (85)$$

$$\frac{\partial d_{\text{CC}}}{\partial \mathbf{x}_2} = 0, \quad (86)$$

$$\frac{1}{m_1} \frac{\partial d_{\text{CG}}}{\partial \mathbf{x}_1} = \frac{1}{m_3} \frac{\partial d_{\text{CG}}}{\partial \mathbf{x}_3}, \quad (87)$$

which leads to

$$\begin{aligned} Z_{d_{\text{CC}}d_{\text{CG}}} &= \sum_{i=1}^{i=3} \frac{1}{m_i} \frac{\partial d_{\text{CC}}}{\partial \mathbf{x}_i} \frac{\partial d_{\text{CG}}}{\partial \mathbf{x}_1} \\ &= \frac{\partial d_{\text{CC}}}{\partial \mathbf{x}_1} \left(\frac{1}{m_1} \frac{\partial d_{\text{CG}}}{\partial \mathbf{x}_1} - \frac{1}{m_3} \frac{\partial d_{\text{CG}}}{\partial \mathbf{x}_3} \right) = 0. \end{aligned} \quad (88)$$

Similar cancellations appear for the other nondiagonal terms.

We will now give the expression of $Z_{d_{\text{CG}}}$, $Z_{d_{\text{CC}}}$, and Z_α . Even though in our case all three atoms have the same mass, we will write the following formula for the general case where all three atoms have different masses and G is the center of mass of C^1C^3 . Tedious but straightforward algebra leads to

$$\mathbf{x}_G = \frac{m_1 \mathbf{x}_1 + m_3 \mathbf{x}_3}{m_1 + m_3}, \quad (89)$$

$$Z_{d_{CG}} = \frac{1}{m_2} + \frac{1}{m_1 + m_3}, \quad (90)$$

$$Z_{d_{CC}} = \frac{1}{m_1} + \frac{1}{m_3}, \quad (91)$$

$$Z_\alpha = \frac{Z_{d_{CG}}}{d_{CG}^2} + \frac{Z_{d_{CC}}}{d_{CC}^2}, \quad (92)$$

$$|\mathbf{Z}_\xi| = Z_\alpha Z_{d_{CG}} Z_{d_{CC}}. \quad (93)$$

Finally, the derivatives of the free energy read

$$\frac{\partial A}{\partial d_{CG}} = \frac{1}{Q} \left(\langle \lambda_{d_{CG}} \rangle - \frac{k\langle T \rangle Z_{d_{CG}}}{Z_\alpha d_{CG}^3} \right), \quad (94)$$

$$\frac{\partial A}{\partial d_{CC}} = \frac{1}{Q} \left(\langle \lambda_{d_{CC}} \rangle - \frac{k\langle T \rangle Z_{d_{CC}}}{Z_\alpha d_{CC}^3} \right), \quad (95)$$

$$\frac{\partial A}{\partial \alpha} = \frac{1}{Q} \langle \lambda_{d_{CG}} \rangle. \quad (96)$$

A. Stationary points

1. Transition state

Optimization of the transition state structure was carried out by employing the quasi-Newton scheme²⁸ using the formula proposed by Boffill to update the Hessian.²⁹ This procedure has been described in detail elsewhere.³⁰ This quasi-Newton scheme is an iterative procedure that requires initial values of the three parameters d_{CG} , d_{CC} , and α . These values were taken from the previous 1D study,²⁵ in which the transition state (TS) was located at $d_{CG}=4.5a_0$; this was taken as the initial value of d_{CG} for our 3D optimization. The thermal average of the ethylene bond length and of the angle during the previous 1D simulation with $d_{CG}=4.5a_0$ were taken as initial values for d_{CC} and for the angle α . The Hessian at this initial geometry was calculated employing finite differences of the gradients. Diagonalization lead to only one negative eigenvalue proving the transition state nature of the starting point. The Hessian matrix was also calculated and diagonalized for the final geometry: we found only one negative eigenvalue, corresponding to an eigenvector directed mainly along the d_{CG} variable.

The resulting geometry is reported in Fig. 1, together with the geometry obtained at 300 K when only d_{CG} is constrained, as well as the geometry obtained on the PES at 0 K. The main geometrical parameters are given Table I.

All three structures are nonsymmetric; the α angle is approximately equal to 115° instead of 90° . This is in agreement with the Woodward–Hoffman rules and with previous calculations^{27,31–34} and experimental results.³⁴ Moreover, all geometries correspond to an early transition state in which the ethylene molecule is only slightly distorted. The double

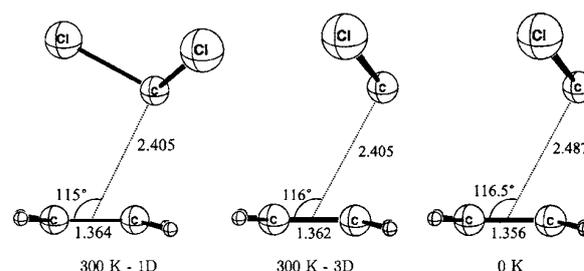


FIG. 1. Transition states geometries for the addition of dichlorocarbene CCl_2 to ethylene, optimized at 0 K, at 300 K with only the d_{CG} distance constrained (1D), and at 300 K with d_{CG} , d_{CC} , and α all constrained (3D). G is the midpoint of the double bond. Distances in Å and angles in deg.

bond length equals to approximately 1.36 \AA , close to the equilibrium distance in the free molecule: $d_{CC}=1.33 \text{ \AA}$.

The two structures found at 300 K are almost identical. This is not surprising; as already stated the distance d_{CG} between the middle of the double bond and the carbon atom of the carbene is a good reaction coordinate up to this point.

The transition state geometry at 0 K is in fairly good agreement with previous *ab initio* calculations^{27,31–33} except for the d_{CG} distance which is slightly overestimated here: $d_{CG}=2.487 \text{ \AA}$ compared to $d_{CG}=2.37 \text{ \AA}$ at the B3LYP/6-31G* level,³³ or $d_{CG}=2.38 \text{ \AA}$ at the MP2/6-31G* level.^{27,32,33} This is due to the fact that this region of the potential-energy surface is shallow so that the location of the transition state depends strongly on the functional and on the basis set used. Similarly, the α angle is also a bit overestimated; it equals 116.5° in our study, whereas it equals, respectively, 111.7° and 112.2° at the MP2/6-31G* and B3LYP/6-31G* levels of calculation.^{27,32,33}

Going from the transition state structure found at 0 K to that obtained at 300 K leads to an increase of the double bond length due to thermal vibrations. On the other hand, the d_{CG} distance is smaller at high temperature. This comes from the fact that the barrier originates mainly from rotational entropy lost when the transition state is formed. As the average rotational momentum increases with the temperature, we expect this barrier to move to smaller distance as the temperature rises. This is in agreement with previous studies.^{25,30}

2. 1,1'-dichlorocyclopropane

The main geometrical parameters for the 1,1'-dichlorocyclopropane are reported in Table II and the

TABLE I. Main geometrical parameters for the transition state geometries for the addition of the dichlorocarbene CCl_2 to ethylene, optimized at 0 K, at 300 K with one constraint (1D), and at 300 K with three constraints (3D). The uncertainty for structures on the potential-energy surface is much smaller and thus not quoted.

	d_{CG} (Å)	d_{CC} (Å)	α (deg)
1D	2.405(2) ^a	1.364(2) ^b	115(1) ^b
3D	2.405(2)	1.362(2)	116(1)
0 K	2.487	1.356	116.5

^aEstimate of the uncertainty.

^bAverage values.

TABLE II. Main geometrical parameters for the 1,1'-dichlorocyclopropane, optimized at 0 K, at 300 K with one constraint (1D), and at 300 K with three constraints (3D).

	d_{CG} (Å)	d_{CC} (Å)	α (deg)
1D	1.289(2) ^a	1.535(2) ^b	90(1) ^b
3D	1.289(2)	1.532(2)	90(1)
0 K	1.288	1.527	90.1

^aEstimate of the uncertainty.

^bAverage values.

corresponding structures are given in Fig. 2. All three structures are very symmetric: $\alpha=90^\circ$ which means that the carbon atoms form an isocel triangle, its base being the former ethylenic bond. All three carbon-carbon bond lengths are now close to that of a single bond: $d_{C^1C^3}=1.53$ Å and $d_{C^1C^2}=d_{C^3C^2}=1.50$ Å.

Similar to what was observed for the transition state structures, there are very little difference between the room-temperature 1D and 3D structures. This comes from the fact that the gradient is zero for a minimum, so that the definition of the reaction coordinate does not matter anymore. To further assess this point, we have launched a simulation with no constraints. The average values of d_{CG} , d_{CC} , and α are in very good agreement with the constrained simulations: $d_{CG}=1.291\pm 0.002$ Å, $d_{CC}=1.533\pm 0.002$ Å, and $\alpha=90.0\pm 1^\circ$.

The 0-K structure is in good agreement with previous calculations; the difference in distances and angles is less than 0.01 Å and 1° , respectively.³¹ When going from 0 to 300 K, the C-C bonds elongate, while d_{CG} and α remain constant. This comes from the fact that this molecule has a C_{2v} symmetry which imposes that the average angle should be close to its value on the potential-energy surface.

B. Reaction path

We focus now on the core of this application; the construction of the reaction path connecting the reactants to the product, directly on the free-energy surface. The aim of this part is dual: first, we show how to use the previous formula on a real example. Second, we compare the 3D path constructed here to the path obtained with only one “chemically intuitive” reaction coordinate. Starting from the transition state, we have constructed the forward reaction path leading to the product and the backward path leading to the reactants.

In this work, we have moved along the gradient employing a small step size; at each point \mathbf{x}^k of the path, a simulation is launched while constraining all three active coordinates d_{CG} , d_{CC} , and α . The previous formulas [(94)–(96)] are used to compute the free-energy gradient: $\mathbf{g}=(\partial A/\partial d_{CG}, \partial A/\partial d_{CC}, \partial A/\partial \alpha)^t$. We then convert this gradient into a normalized mass-weighted gradient: $\mathbf{g}_{MW}=\mathcal{N}\mathbf{Z}\mathbf{g}$, with $\mathcal{N}=(\mathbf{g}^t\mathbf{Z}\mathbf{g})^{-1/2}$. The next point \mathbf{x}^{k+1} is calculated by following the gradient on a small distance ds ,

$$\mathbf{x}^{k+1}=\mathbf{x}^k-ds\times\mathbf{g}_{MW}. \quad (97)$$

An alternative way is to employ the scheme derived by Gonzalez and Schlegel³⁵ which allows for the use of a much bigger stepsize. The forward path corresponds to the closure of the cycle, that is to say to the formation of the second

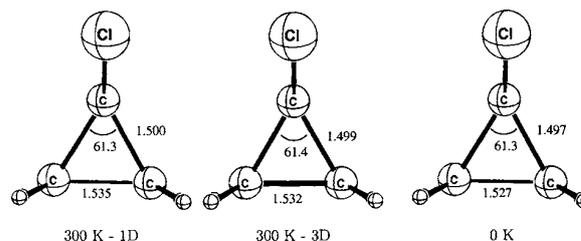


FIG. 2. Geometries of the 1,1'-dichloro cyclopropane optimized at 0 K, at 300 K with only the d_{CG} distance constrained (1D), and at 300 K with d_{CG} , d_{CC} , and α all three constrained (3D). G is the midpoint of the double bond. Distances in Å and angles in deg.

carbon-carbon bond. The gradient along this path is large, and we used a step size of 0.5 a.u.³⁶ The backward path corresponds to the departure of the carbene, which is the reverse of the electrophilic addition. The change in free energy is small in that direction, and we had to employ a smaller step size of 0.2 a.u. to minimize the statistical noise.

1. Free-energy profile

The resulting free-energy profile (FEP) is reported in Fig. 3 together with the profile generated with one constraint.²⁵ In order to compare them, both paths have been plotted using the d_{CG} distance as an approximate reaction coordinate. The FEP obtained in the present work is given as an inset in Fig. 3. The two profiles are very similar: first, the free energy increases smoothly from the isolated reactants to the transition state. Then, it decreases abruptly when the cyclopropane is formed.

On an energetic ground, in agreement with the fact that the free energy is a state function, both paths lead to a similar change in free energy equal to $\Delta A_{\text{corr}}=-47.6$ kcal mol⁻¹ in this work and $\Delta A_{\text{corr}}=-46.5$ kcal mol⁻¹ in our previous work, when employing 40 000 steps. Moreover, as d_{CG} is a good reaction coordinate for the first phase, the barrier is also similar in both cases: $\Delta A_{\text{corr}}^{\text{TS}}\approx 11.5$ kcal mol⁻¹ in this work, compared to $\Delta A_{\text{corr}}^{\text{TS}}\approx 11.7$ kcal mol⁻¹ previously. These values are in good agreement with the previous studies.^{25,32}

2. Geometrical parameters

Even though the two free-energy profiles are similar, the two paths are actually quite different in terms of geometrical parameters. We discuss here the variations of the structural parameters d_{CC} and α for both paths. The evolution of the double bond distance d_{CC} and of the α angle obtained in this study are reported on Fig. 4 together with the average values $\langle d_{CC} \rangle$ and $\langle \alpha \rangle$ obtained in the previous study using only d_{CG} as a reaction coordinate.

The first feature worth noting is that the forward path, corresponding to the the formation of the second C-C bond, is not parallel to the d_{CG} axis but acquires contributions along both d_{CC} and α . This confirms that an accurate description of this step requires a more complex reaction coordinate than just the d_{CG} distance. As a consequence, despite the fact that d_{CG} varies only from 2.40 to 1.29 Å for the second phase, the actual path length is of the same order of magnitude as for the first phase: $s\approx 10$ a.u.

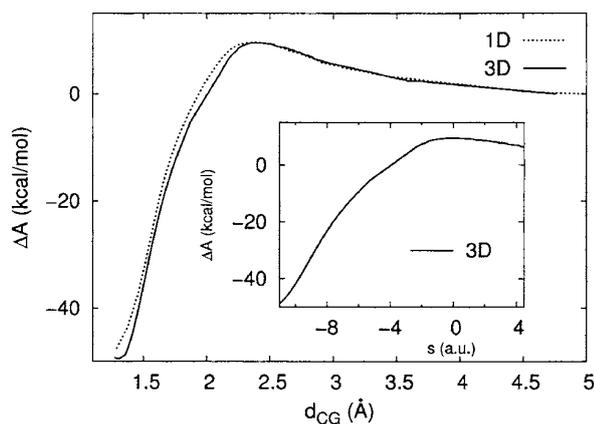


FIG. 3. Free-energy profile (FEP) for the addition of dichlorocarbene CCl_2 to ethylene. The 1D profile is calculated using d_{CG} as the predefined reaction coordinate, whereas the 3D profile uses an active set of three coordinates (d_{CG} , d_{CC} , and α). The inset shows the FEP plotted against the curvilinear distance from the transition state along the 3D path. G is the midpoint of the double bond.

The evolution of the two structural parameters d_{CC} and α can be divided into three zones. The first one corresponds to $d_{CG} \gtrsim 2.6$ Å. In this zone, there is little interaction between the two reactants: d_{CC} is constant and equals to 1.33 Å, which is the standard double bond length, and α tends to 90° .³⁷ This comes from the fact that, for the unconstrained system, at large separation there is free rotation of the carbene around the ethylene molecule. As a consequence, α should vary freely between 0° and 180° , with an average value of 90° .

The second zone corresponds to the electrophilic addition, that is the formation of the first bond between the carbene and the ethylene molecule. In this zone, d_{CG} evolves from ca. 1.8 to ca. 2.6 Å. As the two molecules start to interact the ethylenic bond elongates from 1.33 to ca. 1.48 Å. As expected, this last value is intermediate between a single and a double carbon-carbon bond length. Simultaneously, α increases to 128° . This is a consequence of the fact that the symmetric approach for which $\alpha=90^\circ$ is forbidden and is thus associated with a very high barrier.³⁴ During this phase, the C^1C^2 bond is formed: this distance decreases from 2.4 to 1.56 Å, which is close to a single bond.

The last zone corresponds to $d_{CG} \leq 1.8$ Å and represents the closure of the cyclopropane ring. The C^1C^3 bond (formerly the double bond) length increases to its final value of 1.538 Å, while α drops to 90° . Comparing the path constructed using a three-coordinate active space to the path previously obtained shows that α is a much better reaction coordinate for this phase than d_{CG} . Indeed, during this phase the C^1C^2 distance is almost constant and close to 1.54 Å, while the C^3C^2 distance decreases strongly from 2.3 to 1.54 Å. This illustrates that the second phase of the reaction is actually the bending of the C^1C^2 bond towards the C^3 carbon atom. For this type of movement α is a good reaction coordinate whereas d_{CG} is a poor one. As a consequence, when only d_{CG} is used as a reaction coordinate, α drops abruptly from 115° to 90° around $d_{CG}=1.36$ Å. On the other hand, the variations of α are much smoother with our active set.

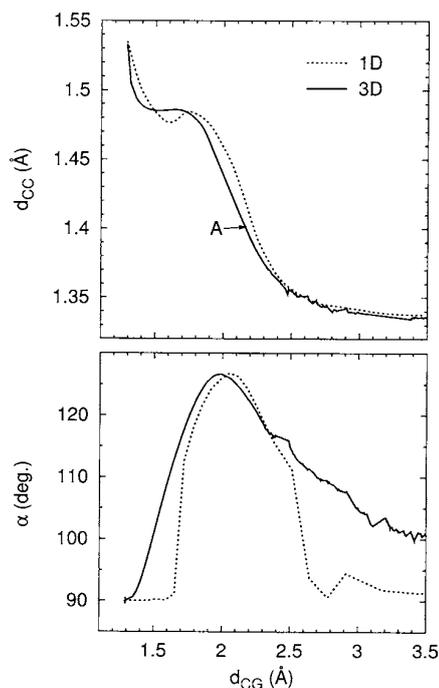


FIG. 4. Reaction path for the addition of dichlorocarbene to ethylene. Two projections are shown in the d_{CC} - d_{CG} and α - d_{CG} subspaces, G being the midpoint of the double bond. The 1D path is obtained using d_{CG} as the reaction coordinate, whereas the 3D path uses an active set of three coordinates (d_{CG} , d_{CC} , and α).

To conclude this section, we would like to point out that, as expected, both paths are qualitatively similar, with the same overall evolutions of the geometrical parameters. However, they differ significantly on a quantitative basis, especially in the third zone. This is due to the fact that d_{CG} is not a good approximation to the reaction coordinate in this zone. To illustrate how this explains why the two paths are different, let us consider a force acting on the d_{CC} variable at the point A (Fig. 4) located in this zone. We denote the local coordinate set by (ξ, u, v) , with ξ being the reaction coordinate. From the definition of a reaction coordinate, we have

$$\frac{\partial A}{\partial \xi} = g \neq 0, \quad (98)$$

$$\frac{\partial A}{\partial u} = 0, \quad (99)$$

$$\frac{\partial A}{\partial v} = 0. \quad (100)$$

As d_{CG} is not a good approximation to the reaction coordinate, ξ depends not only on d_{CG} but also on d_{CC} and α . As a result, the derivative of the free energy along d_{CC} is related to the gradient of the free energy,

$$\frac{\partial A}{\partial d_{CC}} = \frac{\partial A}{\partial \xi} \frac{\partial \xi}{\partial d_{CC}} + \frac{\partial A}{\partial u} \frac{\partial u}{\partial d_{CC}} + \frac{\partial A}{\partial v} \frac{\partial v}{\partial d_{CC}} = \frac{\partial A}{\partial \xi} \frac{\partial \xi}{\partial d_{CC}} \neq 0. \quad (101)$$

Thus, if d_{CC} is not constrained, the system will evolve in order to minimize the free energy in that direction, and the two paths will not coincide.

C. Importance of the correction to evaluate the gradients

It is worth analyzing the importance of the B terms in Eqs. (94)–(96). They are reported on Fig. 5 together with the average value of the Lagrange multipliers. It appears that those terms are usually smaller than the Lagrange multiplier by at least an order of magnitude. However, as the reaction path is following the free-energy gradient, these small differences are accumulated along the path, leading to a non-negligible correction. For example, for the studied reaction, the free-energy correction along the path can reach 1 kcal mol⁻¹. The larger correction for the geometrical parameters is observed for the evolution of the double bond length: $\Delta(d_{CC}) \approx 0.01$ Å. However, the structural differences for the product and the transition state are much smaller: $\Delta(d) \leq 0.005$ Å and $\Delta(\alpha) \leq 0.5^\circ$. The correction for the total free-energy difference is only $\Delta(\Delta A_{\text{corr}}) = 0.1$ kcal mol⁻¹.

We would like to draw attention to the fact that, even though those terms are generally not negligible, one should not forget that we are using classical mechanics, and that all quantum effects are missing for the description of the nuclei motion. In particular, zero-point energy (ZPE) and tunneling are completely neglected.

D. Effect of the temperature

Following the procedure of Gonzalez and Schlegel,³⁵ we have constructed the reaction path at 0 K. As the differences between the potential-energy profile and the free-energy profile have already been discussed,²⁵ we shall focus here on the differences between the path at 0 and 300 K. The two paths are reported on Fig. 6.

At large separation, that is for the electrophilic addition, d_{CG} is a good reaction coordinate at both temperatures. Moreover, as long as the two reactants interact only weakly, the thermal motions are mainly vibrations of the two molecules. As the potential-energy surface is almost harmonic for such vibrations, the evolutions of the double bond distance and of the angle α are similar at 300 and at 0 K. On the potential-energy surface, as the distance between the two fragments increases, the interaction depends less and less on the angle, leading to some spurious evolutions. On the free-energy surface, this independence of the interaction on the angle leads to an average value of 90° instead.

When the interaction starts to be stronger, these oscillations of the angle disappear and both paths are qualitatively similar. However, due to the vibrational thermal energy, at 300 K the parameter changes are larger.

E. Computational details

The Car–Parrinello projector augmented-wave^{38,39} (CP-PAW) program by Blöchl was used for all *ab initio* molecular-dynamics calculations. In the CP-PAW calculations, periodic boundary conditions were used in all examples with an orthorhombic unit cell described by the lat-

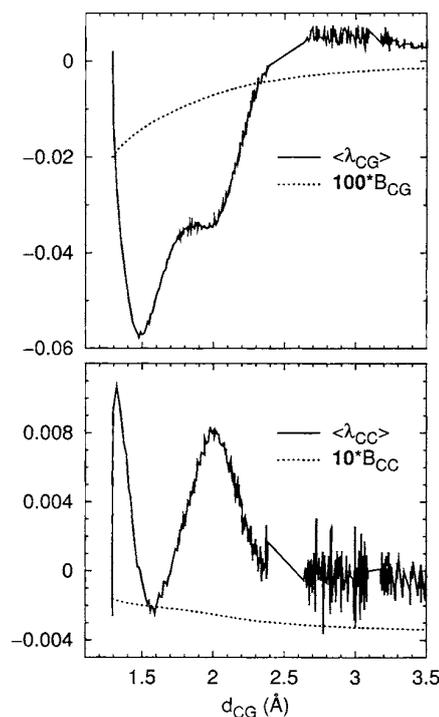


FIG. 5. Importance of the correction terms in the evaluation of the free-energy derivative. $\langle \lambda_x \rangle$ is the average of the Lagrange multiplier. $B_{CG} = -kT(Z_{d_{CG}}/Z_{\alpha d_{CG}^2})$ and $B_{CC} = -kT(Z_{d_{CC}}/Z_{\alpha d_{CC}^2})$ are the correction to the derivative of the free energy along d_{CG} and d_{CC} , respectively. G is the midpoint of the double bond.

tice vectors ($[0, 14.74, 14.74]$, $[14.74, 0, 14.74]$, $[14.74, 14.74, 0]$) (bohr, 7.8 Å). The energy cutoff used to define the basis set was 30 Ry (15 a.u.) in all cases. Because the systems of interest are all isolated molecules, only the Γ point in k space was included and the interaction between images was removed by the method proposed by Blöchl.³⁹ The approximate DFT used here consisted of the combination of the Perdew–Wang parametrization of the electron gas⁴⁰ in combination with the exchange gradient correction presented by Becke⁴¹ and the correlation correction of Perdew.⁴² The SHAKE algorithm²⁰ was employed in order to impose the constraints. The mass of the hydrogen atoms was taken to be that of deuterium, and normal masses were taken for all other elements.

Room-temperature CP-PAW calculations were performed at a target temperature of 300 K. The Andersen thermostat⁴³ was applied to the nuclear motion by reassigning the velocity of N randomly chosen nuclei every n steps where N and n are chosen to maintain the desired temperature. In our case this amounted to one velocity reassignment every 20 steps. Thermostat settings were monitored and adjusted if necessary during the equilibration stage, with the main criteria for adequate thermostating being the mean temperature lying within a range of 300 ± 10 K and a temperature drift lower than 1 K/ps. In combination with the Andersen thermostat, a constant friction was applied to the wave function with a value of 0.001. Following the conclusions of the previous study, for each simulation, we performed between 35 000 and 50 000 steps in order to ensure that the

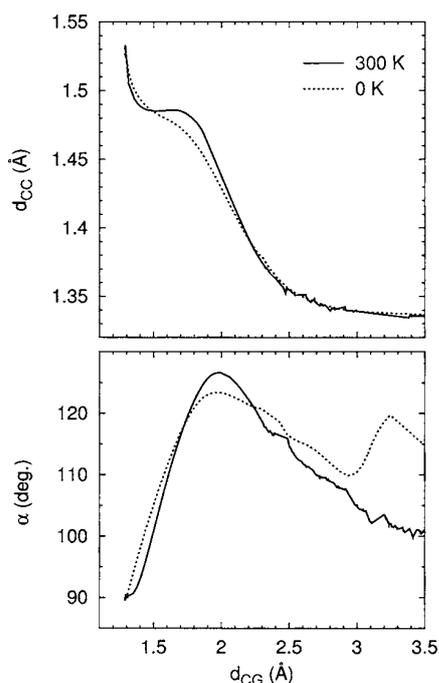


FIG. 6. Reaction path for the addition of the dichlorocarbene to ethylene at 0 and 300 K. Two projections are shown in the d_{CC} - d_{CG} and α - d_{CG} subspaces, G being the midpoint of the double bond.

system was fully equilibrated and that the temperature and the free-energy gradient were fully converged.

The free-energy profiles were obtained by numerical integration of the gradient along the path, using a procedure similar to the pointwise thermodynamic integration (PTI) method.⁹ As the overall rotation and translation of the molecule are frozen during a simulation, one has to correct the free energy obtained from a simulation. We have used the procedure of Kelly *et al.*²⁵ To summarize, the overall correction for the entropy is the sum of the translational and rotational entropies,

$$\Delta S_{\text{corr}}^{AB}(s) = S_R^{AB}(s) + S_T^{AB}(s) - S_T^A(\infty) - S_T^B(\infty), \quad (102)$$

where $S_R^{AB}(s)$ is the rotational entropy at $RC=s$ which is geometry dependent, and $S_T^{AB}(s)$ is the translational entropy at $RC=s$. The last two terms represent the translational entropy of the isolated species A and B . These terms are calculated using standard formula for the partition functions. Finally, the total free-energy change $\Delta A_{\text{corr}}^{AB}(s)$ is obtained from a CP-PAW simulation with the constraints described above as

$$\Delta A_{\text{CM}}^{AB}(s) = \Delta A_{\text{PAW}}^{AB}(s) - T\Delta S_{\text{corr}}^{AB}(s),$$

where $\Delta A_{\text{PAW}}^{AB}(s)$ is the change in free energy obtained directly from the simulation, and CM (classical mechanics) refers to the fact that the motion of the nuclei is described using classical mechanics. It should be mentioned that the ZPE correction is not included in our simulations. This should not seriously hamper our objective which is to analyze the qualitative differences between the paths obtained with one and three constraints.

V. CONCLUSION

In this work, we have proposed a new look at the standard formulas for evaluating the derivatives of the free energy along a reaction coordinate.

First, we recollected the different formulas available in a uniform approach. These formulas allow one to compute accurately and efficiently the gradient of the free energy for an *unconstrained* system using a *constrained* molecular-dynamics simulation.

The main finding of this investigation is a set of equations that makes it possible to construct a minimum-free-energy reaction path instead of calculating the free-energy changes along a predefined path. Indeed, we believe that one can use the free-energy gradients in the same way potential-energy gradients have been used in the past 20 years in quantum chemistry calculations.

The addition of the dichlorocarbene to ethylene was studied as a numerical example. It was shown that a simulation using only one constraint is not sufficient to describe the whole path. Using the free-energy gradient in a subset of three active coordinates leads to a smoother path, refining the understanding of this process.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Michael Seth and Dr. Shengyong Yang for helpful and fruitful discussions. This work was supported by the National Sciences and Engineering Research Council of Canada (NSERC). Calculations were performed in part on the Westgrid cluster and the MACI Alpha cluster located at the University of Calgary. One of us (T.Z.) thanks the Canadian government for a Canada Research Chair.

APPENDIX A: PROPERTIES OF THE MASS MATRIX

Using $\mathbf{A}_{q\xi}^{-1}\mathbf{A}_{q\xi}=\mathbf{I}$, one readily finds

$$\mathbf{X}_q\mathbf{A}_q + \mathbf{Y}_\xi\mathbf{B}_\xi^t = \mathbf{I}_{3N-1}, \quad (A1)$$

$$\mathbf{X}_q\mathbf{B}_\xi + \mathbf{Y}_\xi\mathbf{C}_\xi = \mathbf{0}, \quad (A2)$$

$$\mathbf{Y}_\xi^t\mathbf{A}_q + \mathbf{Z}_\xi\mathbf{B}_\xi^t = \mathbf{0}, \quad (A3)$$

$$\mathbf{Y}_\xi^t\mathbf{B}_\xi + \mathbf{Z}_\xi\mathbf{C}_\xi = 1. \quad (A4)$$

Plugging Eq. (A3) into Eq. (A1) leads to

$$\mathbf{A}_q^{-1} = \mathbf{X}_q - \mathbf{Y}_\xi\mathbf{Z}_\xi^{-1}\mathbf{Y}_\xi^t. \quad (A5)$$

The last relation we need derives from

$$\begin{pmatrix} \mathbf{A}_q & \mathbf{0} \\ \mathbf{B}_\xi^t & 1 \end{pmatrix} = \mathbf{A}_{q\xi}\mathbf{A}_{q\xi}^{-1}\begin{pmatrix} \mathbf{A}_q & \mathbf{0} \\ \mathbf{B}_\xi^t & 1 \end{pmatrix} = \mathbf{A}_{q\xi}\begin{pmatrix} \mathbf{I}_{3N-1} & \mathbf{Y}_\xi \\ \mathbf{0} & \mathbf{Z}_\xi \end{pmatrix}. \quad (A6)$$

Equating the determinant of the first and last terms, we get

$$|\mathbf{A}_{q\xi}| = |\mathbf{A}_q||\mathbf{Z}_\xi|^{-1} = |\mathbf{J}^t\mathbf{M}\mathbf{J}| = |\mathbf{J}|^2|\mathbf{M}|. \quad (A7)$$

APPENDIX B: PROPERTIES OF GAUSSIAN INTEGRALS

We recall here the main properties of Gaussian integrals,

$$\int \mathbf{d}\mathbf{u} e^{(-\mathbf{u}^t \mathbf{A} \mathbf{u})} \propto |\mathbf{A}|^{-1/2}, \quad (\text{B1})$$

$$\int \mathbf{d}\mathbf{u} e^{(-\mathbf{u}^t \mathbf{A} \mathbf{u})} \mathbf{u}^t \mathbf{B} \mathbf{u} \propto \frac{1}{2} \text{Tr}(\mathbf{A}^{-1} \mathbf{B}) \int \mathbf{d}\mathbf{u} e^{(-\mathbf{u}^t \mathbf{A} \mathbf{u})}. \quad (\text{B2})$$

APPENDIX C: CONSTRAINING THE OVERALL ROTATION AND TRANSLATION

In this section, we derive the expressions used to constrain the overall translation and rotation of the system. For this, we start from a reference geometry \mathbf{x}^0 and we seek the conditions that should be satisfied by the new geometry \mathbf{x} . The center of mass G is defined by

$$\mathbf{x}_G = \frac{\sum_{i=1}^N m_i \mathbf{x}_i}{\sum_{i=1}^N m_i},$$

where m_i is the mass of the atom i with coordinate \mathbf{x}_i . In the following, the notation $\tilde{\mathbf{x}}_i = \mathbf{x}_i - \mathbf{x}_G$ will be used. Let us denote by \hat{x} , \hat{y} , and \hat{z} the unit vectors of our laboratory coordinate system. For the sake of clarity, the component of x_i along the \hat{x} , \hat{y} , and \hat{z} axes will be denoted by $x_{i,1}$, $x_{i,2}$, and $x_{i,3}$ respectively.

Constraining the translation is equivalent to freeze the movement of the center of mass, i.e., to apply the following linear constraints:

$$\sigma_1 = \sum_{i=1}^N \frac{m_i}{M} (x_{i,1} - x_{i,1}^0) = 0, \quad (\text{C1a})$$

$$\sigma_2 = \sum_{i=1}^N \frac{m_i}{M} (x_{i,2} - x_{i,2}^0) = 0, \quad (\text{C1b})$$

$$\sigma_3 = \sum_{i=1}^N \frac{m_i}{M} (x_{i,3} - x_{i,3}^0) = 0. \quad (\text{C1c})$$

Constraining the rotation can be done by using the second Eckart conditions.⁴⁴ These conditions minimize the angular momentum due to small displacements: they provide an approximate way to constrain the global rotation during a molecular-dynamics simulation,

$$\sigma_4 = \hat{x} \cdot \left[\sum_{i=1}^N m_i (\tilde{\mathbf{x}}_i^0 \times \tilde{\mathbf{x}}_i) \right] = 0, \quad (\text{C2a})$$

$$\sigma_5 = \hat{y} \cdot \left[\sum_{i=1}^N m_i (\tilde{\mathbf{x}}_i^0 \times \tilde{\mathbf{x}}_i) \right] = 0, \quad (\text{C2b})$$

$$\sigma_6 = \hat{z} \cdot \left[\sum_{i=1}^N m_i (\tilde{\mathbf{x}}_i^0 \times \tilde{\mathbf{x}}_i) \right] = 0, \quad (\text{C2c})$$

that is

$$\sigma_4 = \sum_{i=1}^N m_i (\tilde{x}_{i,2}^0 \tilde{x}_{i,3} - \tilde{x}_{i,2} \tilde{x}_{i,3}^0) = 0, \quad (\text{C3a})$$

$$\sigma_5 = \sum_{i=1}^N m_i (\tilde{x}_{i,3}^0 \tilde{x}_{i,1} - \tilde{x}_{i,3} \tilde{x}_{i,1}^0) = 0, \quad (\text{C3b})$$

$$\sigma_6 = \sum_{i=1}^N m_i (\tilde{x}_{i,1}^0 \tilde{x}_{i,2} - \tilde{x}_{i,1} \tilde{x}_{i,2}^0) = 0. \quad (\text{C3c})$$

These last equations can be written as linear constraints,

$$\sigma_4 = \sum_{i=1}^N m_i \left\{ x_{i,3} \left(\tilde{x}_{i,2}^0 - \sum_{j=1}^N \frac{m_j}{M} \tilde{x}_{j,2}^0 \right) - x_{i,2} \left(\tilde{x}_{i,3}^0 - \sum_{j=1}^N \frac{m_j}{M} \tilde{x}_{j,3}^0 \right) \right\} = 0, \quad (\text{C4a})$$

$$\sigma_5 = \sum_{i=1}^N m_i \left\{ x_{i,1} \left(\tilde{x}_{i,3}^0 - \sum_{j=1}^N \frac{m_j}{M} \tilde{x}_{j,3}^0 \right) - x_{i,3} \left(\tilde{x}_{i,1}^0 - \sum_{j=1}^N \frac{m_j}{M} \tilde{x}_{j,1}^0 \right) \right\} = 0, \quad (\text{C4b})$$

$$\sigma_6 = \sum_{i=1}^N m_i \left\{ x_{i,2} \left(\tilde{x}_{i,1}^0 - \sum_{j=1}^N \frac{m_j}{M} \tilde{x}_{j,1}^0 \right) - x_{i,1} \left(\tilde{x}_{i,2}^0 - \sum_{j=1}^N \frac{m_j}{M} \tilde{x}_{j,2}^0 \right) \right\} = 0. \quad (\text{C4c})$$

APPENDIX D: ACTUALLY CALCULATING EQ. (55)

In this appendix, we propose one way to calculate the derivatives of the free energy A along q_n , using a simulation in which only ξ is constrained. As an example, we consider that the coordinates q_i to q_n were inactive at step $k-1$ and become active at step k .

The first step is to use the expressions for the generalized coordinates to obtain the values for \mathbf{Z}_ξ , Z_{q_n} , $\partial Z_{q_n} / \partial x'$, and $\partial q_n / \partial x'$.

The main difficulty in using Eq. (55) is that one must ensure that the sampling of q_n around q_n^k is sufficient. In practice, this imposes to have long MD simulations, with approximately 100 000 steps for each considered inactive coordinate. One way to circumvent this problem is to use a Taylor expansion of the derivative of the potential around q_i^k, \dots, q_n^k ,

$$\begin{aligned} \frac{\partial V}{\partial q_n} = & \left(\frac{\partial V}{\partial q_n^k} \right)_{q_1^k, \dots, q_n^k, \xi^*} + \sum_{j=i}^{j=n} \left(\frac{\partial^2 V}{\partial q_n \partial q_j} \right)_{q_1^k, \dots, q_n^k, \xi^*} (q_j - q_j^k) \\ & + \frac{1}{2} \sum_{j=i}^{j=n} \sum_{l=i}^{l=n} \left(\frac{\partial^3 V}{\partial q_n \partial q_j \partial q_l} \right)_{q_1^k, \dots, q_n^k, \xi^*} (q_j - q_j^k)(q_l - q_l^k). \end{aligned} \quad (D1)$$

The simulation data is then used to fit the coefficients of this expression. The resulting equation is then plugged into Eq. (55).

- ¹M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon, Oxford, England, 1987).
²C. Chipot and D. A. Pearlman, *Mol. Simul.* **28**, 1 (2002), and references therein.
³S. Park and K. Schulten, *J. Chem. Phys.* **120**, 5946 (2004).
⁴F. M. Ytreberg and D. M. Zuckerman, *J. Chem. Phys.* **120**, 10876 (2004).
⁵C. Jarzynski, *Phys. Rev. Lett.* **78**, 2690 (1997).
⁶C. Jarzynski, *Phys. Rev. E* **56**, 5018 (1997).
⁷R. W. Zwanzig, *J. Chem. Phys.* **22**, 1420 (1954).
⁸G. M. Torrie and J. P. Valleau, *J. Comput. Phys.* **23**, 187 (1977).
⁹T. P. Straatsma and J. A. McCammon, *J. Chem. Phys.* **95**, 1175 (1991).
¹⁰W. F. van Gunsteren, in *Computer Simulations of Biomolecular Systems: Theoretical and Experimental Applications*, edited by W. F. van Gunsteren and P. K. Weiner (ESCOM, Leiden, The Netherlands, 1989), Vol. 1, p. 27.
¹¹E. A. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral, *Chem. Phys. Lett.* **156**, 472 (1989).
¹²M. Sprik and G. Ciccotti, *J. Chem. Phys.* **109**, 7737 (1998).
¹³W. K. den Otter and W. J. Briels, *J. Chem. Phys.* **109**, 4139 (1998).
¹⁴E. Darve and A. Pohorille, *J. Chem. Phys.* **115**, 9169 (2001).
¹⁵K. Fukui, *J. Phys. Chem.* **74**, 4161 (1970).
¹⁶K. Fukui, *Acc. Chem. Res.* **14**, 363 (1981).
¹⁷W. K. den Otter and W. J. Briels, *Mol. Phys.* **98**, 773 (2000).
¹⁸E. Darve, M. A. Wilson, and A. Pohorille, *Mol. Simul.* **28**, 113 (2002).
¹⁹J. Schlitter and M. Klähn, *Mol. Phys.* **101**, 3439 (2003).

- ²⁰J.-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, *J. Comput. Phys.* **23**, 327 (1977).
²¹G. Vilasi, *Hamiltonian Dynamics* (World Scientific, London, England, 2001), Chap. 2, p. 44.
²²J. Schlitter, W. Swegat, and T. Mülders, *J. Mol. Model.* **7**, 171 (2001).
²³E. B. Wilson, Jr., J. C. Decius, and P. C. Cross, *Molecular Vibrations* (McGraw-Hill, New York, 1955).
²⁴A. Michalak and T. Ziegler, *J. Phys. Chem. A* **105**, 4333 (2001).
²⁵E. Kelly, M. Seth, and T. Ziegler, *J. Phys. Chem. A* **108**, 2167 (2004).
²⁶N. G. Rondam, K. N. Houk, and R. A. Moss, *J. Am. Chem. Soc.* **102**, 1770 (1980).
²⁷K. Krogh-Jespersen, S. Yan, and R. A. Moss, *J. Am. Chem. Soc.* **121**, 6269 (1999).
²⁸H. B. Schlegel, in *Encyclopedia of Computational Chemistry*, edited by P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III, and P. R. Schreiner (Wiley, Chichester, 1998), p. 1136, and references therein.
²⁹J. M. Bofill, *J. Comput. Chem.* **15**, 1 (1994).
³⁰S. Yang, I. Hristov, P. Fleurat-Lessard, and T. Ziegler, *J. Phys. Chem. A* **109**, 197 (2005).
³¹K. N. Houk, N. G. Rondam, and J. Mareda, *J. Am. Chem. Soc.* **106**, 4291 (1984).
³²J. F. Blake, S. G. Wiershke, and W. L. Jorgensen, *J. Am. Chem. Soc.* **111**, 1919 (1989).
³³J. J. Blavins, D. L. Cooper, and P. B. Karadakov, *Int. J. Quantum Chem.* **98**, 465 (2004).
³⁴A. E. Keating, S. R. Merrigan, D. A. Singleton, and K. N. Houk, *J. Am. Chem. Soc.* **121**, 3933 (1999).
³⁵C. Gonzalez and H. B. Schlegel, *J. Chem. Phys.* **90**, 2154 (1989).
³⁶Even though this value might seem quite large, because of the two heavy chlorine atoms, a step of 0.5 a.u. corresponds to a decrease of the d_{CG} distance by approximately 0.005 Å only.
³⁷The convergence is slow and is not shown in the figure.
³⁸P. E. Blöchl, *Phys. Rev. B* **50**, 17953 (1994).
³⁹P. E. Blöchl, *J. Phys. Chem.* **99**, 7422 (1995).
⁴⁰J. P. Perdew and Y. Wang, *Phys. Rev. B* **45**, 13244 (1992).
⁴¹A. Becke, *Phys. Rev. A* **38**, 3098 (1988).
⁴²J. P. Perdew, *Phys. Rev. B* **33**, 8822 (1986).
⁴³H. C. Andersen, *J. Chem. Phys.* **72**, 2384 (1980).
⁴⁴C. Eckart, *Phys. Rev.* **47**, 552 (1935).

4) The one with the Adaptive Biasing Density of States

During my project with T. Ziegler, we showed that one can, in principle, optimize a transition state directly on the FES. Then, starting from the TS, one can follow the Intrinsic Reaction Path. This requires to find the TS and then, at each step to compute the free energy gradients. However, for complex reactions, finding the TS is the limiting step ! More, in usual formula, such as the one given in the previous chapter, one needs to compute both first and second derivatives of the reaction coordinates which is technically demanding for complex coordinates.

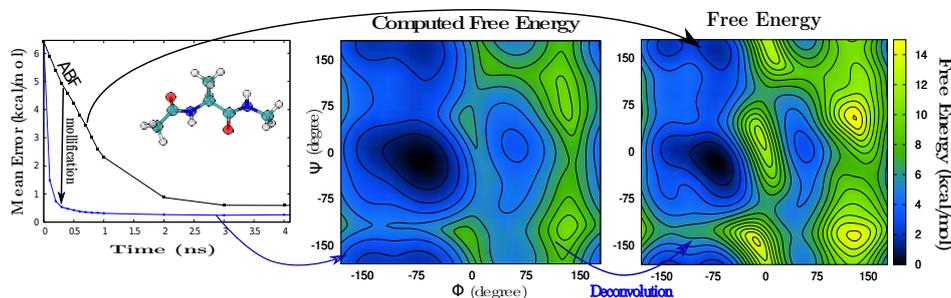
With this in mind, in collaboration with the applied mathematicians of the École Nationale des Ponts et Chaussées, we developed a new free energy sampling approach based on biased simulations. Even though it was not designed in this way, one can see it as an extension of the Wang-Landau¹ method coupled to metadynamics.² Indeed, the bias is constructed by adding Gaussians, as in metadynamics, to the Density of States (DOS), as in Wang-Landau. In practice, this is done by replacing the delta function of the Free Energy definition by a finite Gaussian:

$$e^{-\beta A(\xi^*)} \propto \int_{\mathcal{X}} \exp\left(-\frac{|\xi|^2}{\alpha^2}\right) e^{-\beta V(x)} dx,$$

This gives our method two clear advantages: (i) on the practical side, it is very easy to implement (see the appendix of the following article), (ii) on the fundamental side, the approximate DOS is a convolution of the exact DOS and the finite Gaussian:

$$e^{-\beta A_\alpha(\xi^*)} \propto \int_{\Omega} \delta_\alpha(\bar{\xi} - \xi^*) e^{-\beta A(\bar{\xi})} d\bar{\xi}$$

Therefore, the exact free energy can be recovered at the end of the simulation, at no cost, by a simple deconvolution.



¹a) Wang, F.; Landau, D. P. *Phys. Rev. Lett.* **2001**, *86*, 2050–2053. b) Wang, F.; Landau, D. P. *Phys. Rev. E* **2001**, *64*, 056101

²a) Laio, A.; Parrinello, M. *Proc. Natl. Acca. Sci.* **2002**, *99*, 12562–12566. b) Barducci, A.; Bussi, G.; Parrinello, M. *Phys. Rev. Lett.* **2008**, *100*, 020603.

Free Energy Calculations: An Efficient Adaptive Biasing Potential Method

Bradley M. Dickson,^{*,†,‡} Frédéric Legoll,[§] Tony Lelièvre,[†] Gabriel Stoltz,[†] and Paul Fleurat-Lessard^{*,‡}

Université Paris Est, CERMICS, Project MICMAC Ecole des Ponts ParisTech - INRIA, 6 & 8 Avenue Blaise Pascal, 77455 Marne-la-Vallée Cedex 2, France, Université de Lyon, Laboratoire de Chimie UMR 5182, École Normale Supérieure de Lyon, 46 allée d'Italie, 69364 Lyon Cedex 7, France, and Université Paris Est, Institut Navier, LAMI, Project MICMAC Ecole des Ponts ParisTech - INRIA, 6 & 8 Avenue Blaise Pascal, 77455 Marne-la-Vallée Cedex 2, France

Received: January 31, 2010; Revised Manuscript Received: March 13, 2010

We develop an efficient sampling and free energy calculation technique within the adaptive biasing potential (ABP) framework. By mollifying the density of states we obtain an approximate free energy and an adaptive bias potential that is computed directly from the population along the coordinates of the free energy. Because of the mollifier, the bias potential is “nonlocal”, and its gradient admits a simple analytic expression. A single observation of the reaction coordinate can thus be used to update the approximate free energy at every point within a neighborhood of the observation. This greatly reduces the equilibration time of the adaptive bias potential. This approximation introduces two parameters: strength of mollification and the zero of energy of the bias potential. While we observe that the approximate free energy is a very good estimate of the actual free energy for a large range of mollification strength, we demonstrate that the errors associated with the mollification may be removed via deconvolution. The zero of energy of the bias potential, which is easy to choose, influences the speed of convergence but not the limiting accuracy. This method is simple to apply to free energy or mean force computation in multiple dimensions and does not involve second derivatives of the reaction coordinates, matrix manipulations nor on-the-fly adaptation of parameters. For the alanine dipeptide test case, the new method is found to gain as much as a factor of 10 in efficiency as compared to two basic implementations of the adaptive biasing force methods, and it is shown to be as efficient as well-tempered metadynamics with the postprocess deconvolution giving a clear advantage to the mollified density of states method.

1. Introduction

Many interesting physical systems can be categorized as rare-event systems. The unifying feature of these systems is that the dynamics involved require a time resolution much smaller than the time scale on which interesting events take place. Of central importance to the evolution of such systems is the free energy. Low lying regions of the free energy and the barriers separating these regions dictate the thermodynamics and, to some extent, the kinetics¹ of the system. Because free energy barriers are only rarely crossed, efficient exploration of the free energy landscape is practically impossible with straightforward integration of the equations of motion.

Recently, a number of approaches have used generalized ensemble simulations to accelerate exploration of the free energy landscape.^{2–14} In these methods, information about the free energy is estimated during simulation, and that information is fed back to the dynamics as a statistical bias. While there are many variations on this idea, the common aim is to minimize the time spent sampling regions of the free energy that have been sampled in the past. These schemes may be classified into two categories: adaptive biasing force (ABF) methods,^{7,15} which use an approximation of the mean force to bias the dynam-

ics, and adaptive biasing potential (ABP) methods,^{2,6,8,9,13,14} which use an approximation of the free energy as a bias potential.

The underlying idea for all adaptive methods is that it is computationally more efficient to sample the distribution associated with a flattened free energy than it is to sample the density associated with the actual, very rough free energy. In the spirit of the work of Berg et al.,² Wang and Landau,^{5,6} Eisenmenger et al.,^{4,16} and Procacci et al.,⁹ we propose an ABP method that builds an approximate density of states (DOS) and uses that approximation to define a bias potential. However, in contrast to these approaches, we use a mollification of the underlying DOS to produce the desired approximation. This leads to a smooth, adaptive bias potential whose gradient admits a simple analytic expression and that can be computed without knowledge of the actual DOS. Because the actual and approximate free energies are related by a convolution, it is easy to recover the former from the latter via deconvolution. Our framework is not restricted to one-dimensional or orthogonal reaction coordinates. Moreover, it avoids second derivatives of the reaction coordinate.

This paper is organized as follows. We describe our ABP method in section 2 (see, in particular, eqs 9, 10, and 11), and comment on its convergence. We contrast this method to existing ones in section 3, and present some numerical validation on a benchmark system in section 4. Our conclusions are summarized in section 5.

* To whom correspondence should be addressed. E-mail: adynata@gmail.com, Paul.Fleurat-Lessard@ens-lyon.fr.

† Université Paris Est, CERMICS.

‡ Université de Lyon.

§ Université Paris Est, Institut Navier.

2. Description of the Method

2.1. Free Energy and Its Mollified Version. Consider a system whose configuration is described by a variable $x \in \mathcal{X}$, where \mathcal{X} is the configuration space. We denote by V the potential energy function. Assume that we are given an N -dimensional reaction coordinate $\xi(x)$ with values in Ω , which characterizes some physical event. The DOS at a value ξ^* of the reaction coordinate is defined as

$$e^{-\beta A(\xi^*)} = Z^{-1} \int_{\mathcal{X}} \delta(\xi(x) - \xi^*) e^{-\beta V(x)} dx \quad (1)$$

where $\beta = 1/k_B T$, and Z is a normalization constant, chosen such that

$$\int_{\Omega} e^{-\beta A(\xi)} d\xi = 1$$

Equation 1 defines the free energy $A(\xi^*)$. Recall that, in practice, the free energy needs only be known up to an additive constant since the important quantities to describe the relative likelihoods of physical states are free-energy differences.

In general, the free energy is unknown and has to be approximated. The method we propose in this work is based on the following limit:

$$e^{-\beta A(\xi^*)} = \lim_{\alpha \rightarrow 0} e^{-\beta A_\alpha(\xi^*)}$$

where

$$e^{-\beta A_\alpha(\xi^*)} = Z^{-1} \int_{\mathcal{X}} \delta_\alpha(\xi(x) - \xi^*) e^{-\beta V(x)} dx \quad (2)$$

with, for instance, a Gaussian approximation of the Dirac delta function:

$$\delta_\alpha(\xi) = \left(\frac{1}{\alpha\sqrt{\pi}} \right)^N \exp\left(-\frac{|\xi|^2}{\alpha^2} \right)$$

Equation 2 defines an approximate free energy A_α , obtained by sampling the DOS at a finite α , i.e., by sampling a mollified DOS. Notice that the approximation resulting from finite α can in fact be rewritten as a convolution of the actual DOS $e^{-\beta A}$ with δ_α . Indeed,

$$\begin{aligned} e^{-\beta A_\alpha(\xi^*)} &= Z^{-1} \int_{\mathcal{X}} \delta_\alpha(\xi(x) - \xi^*) e^{-\beta V(x)} dx \\ &= Z^{-1} \int_{\Omega} \int_{\mathcal{X}} \delta_\alpha(\xi - \xi^*) \delta(\xi(x) - \xi) e^{-\beta V(x)} dx d\xi \\ &= \int_{\Omega} \delta_\alpha(\xi - \xi^*) e^{-\beta A(\xi)} d\xi \end{aligned} \quad (3)$$

This remark is the basis for an extraction of the actual free energy A from A_α through a deconvolution procedure (see section 4.1). While we make this presentation with a scalar α , this could easily be generalized to the case where α takes different values in different dimensions of the reaction coordinate.

Equation 3 is also helpful in assessing the errors introduced in simulations employing harmonic constraint potentials to compute the free energy. The corresponding error is analogous

to the convolution errors discussed in this paper. Note that the parameter α can be converted to a force constant for a harmonic potential via $k = 2k_B T/\alpha^2$, where k is the force constant. Errors resulting from finite k can be identified as resulting from a convolution between the true DOS and a known Gaussian function. Typically, the harmonic constraints are tight enough for A_α to be a good approximation of A , but any persisting bias can, at least in principle, be removed by deconvolution as shown below.

2.2. Interest of the Mollified Free Energy. In this work, we use A_α to define an adaptive bias. The first interest of this approach is that the gradient of A_α is much easier to compute than the gradient of A . Indeed, the latter reads (see refs 17–19)

$$F_j(\xi^*) = \left\langle \sum_{i=1}^N \nabla V \cdot G_{ji}^{-1} \nabla \xi_i - \beta^{-1} \nabla \cdot (G_{ji}^{-1} \nabla \xi_i) \right\rangle_{\xi^*} \quad (4)$$

where $\langle \cdot \rangle_{\xi^*}$ denotes a canonical average for a fixed value of the reaction coordinate, and G is the Gram matrix. The latter matrix is defined as $G = JJ'$ with $J_{ij} = \partial \xi_j / \partial x_i$ (x_i are the Cartesian coordinates on which the reaction coordinates are defined). The computation of the free energy gradient therefore requires the computation of second derivatives of the reaction coordinate, which is cumbersome in many cases. The gradient of the mollified free energy has a much simpler expression:

$$\frac{\partial A_\alpha(\xi^*)}{\partial \xi_j^*} = -k_B T \frac{\int_{\mathcal{X}} \partial_{\xi_j} \delta_\alpha(\xi(x) - \xi^*) e^{-\beta V(x)} dx}{\int_{\mathcal{X}} \delta_\alpha(\xi(x) - \xi^*) e^{-\beta V(x)} dx} \quad (5)$$

where j is a reaction coordinate index and

$$\partial_{\xi_j} \delta_\alpha(\xi(x) - \xi^*) = \frac{2}{\alpha^2} (\xi_j(x) - \xi_j^*) \delta_\alpha(\xi(x) - \xi^*)$$

In particular, no derivative of the reaction coordinates are required.

Another interest of the mollified free energy lies in the nonlocality of δ_α , which allows a single observation of ξ to contribute to A_α for a range of values ξ^* , leading to a faster convergence. The question is then whether there is a range of α for which (i) α is sufficiently large so that A_α could be estimated with fewer samples than what would be required to compute A and (ii) α is sufficiently small, so that A_α is close enough to A to efficiently bias the dynamics. We show in section 4.3 that a large range of α satisfies these two conditions on a paradigmatic test case.

2.3. A New ABP Method. 2.3.1. Construction of the Method. To compute approximations of eqs 3 and 5 as time averages along a trajectory x_t driven by the potential function $V(x)$, we first assume that x_t is ergodic with respect to the canonical ensemble. We may take x_t as a solution to the Langevin equation driven by the potential V , for example. Using trajectory averages, eq 3 can be approximated by the following longtime limit:

$$e^{-\beta A_\alpha(\xi^*, t)} = Z_t^{-1} \left(1 + \int_0^t \delta_\alpha(\xi(x_s) - \xi^*) ds \right) \quad (6)$$

where the normalization constant Z_t is

$$Z_t = \int_{\Omega} (1 + \int_0^t \delta_{\alpha}(\xi(x_s) - \xi^*) ds) d\xi^* = |\Omega| + t$$

The normalization constant ensures that

$$\int_{\Omega} e^{-\beta A_{\alpha}(\xi, t)} d\xi = 1 \quad (7)$$

at all times $t \geq 0$. Notice that we implicitly assumed that the reaction coordinate has values in a finite space Ω . This is indeed the case when angles are considered. For unbounded reaction coordinates, it is always possible to restrict the sampling to important values of $\xi(x)$. In practice, the range of the reaction coordinate needs to be truncated anyway.

From eq 6, we obtain

$$\frac{\partial A_{\alpha}(\xi^*, t)}{\partial \xi_j^*} = -k_B T \frac{\int_0^t \partial_{\xi_j} \delta_{\alpha}(\xi(x_s) - \xi^*) ds}{1 + \int_0^t \delta_{\alpha}(\xi(x_s) - \xi^*) ds} \quad (8)$$

which, in the longtime limit converges to eq 5.

Now, a simple ergodic average such as eq 6 or eq 8 can of course not be used in practice since the dynamics at hand are usually metastable for complex systems, and the convergence of the time averages eq 6 and eq 8 is very slow. We therefore need to bias the dynamics in order to remove the metastability.

In what follows, we will consider a trajectory x_t obtained from the equations of motion with the biased potential $V + V_b$. The idea behind adaptive method is to use the opposite of some current approximation of the free energy as a biasing potential, and to update the estimate as time goes on, in a way such that the bias eventually converges to the correct free energy. Here, we consider an ABP method, defined through the following update of the biasing potential V_b :

$$e^{\beta V_b(\xi, t)} = e^{-\beta \Delta A_{\alpha}(\xi, t)} e^{\beta c} \quad (9)$$

where the renormalized current approximation of the mollified free energy $e^{-\beta \Delta A_{\alpha}(\xi, t)}$ is

$$e^{-\beta \Delta A_{\alpha}(\xi, t)} = \frac{e^{-\beta A_{\alpha}(\xi, t)}}{\max_{\xi^*} [e^{-\beta A_{\alpha}(\xi^*, t)}]}$$

The parameter c in eq 9 is an important quantity in our method, which allows one to tune the convergence rate of the method. We discuss its choice in section 2.3.3. With these definitions, $V_b = -A_{\alpha}$ up to an additive constant that is chosen such that $\max[V_b] = c$. Similarly, $\Delta A_{\alpha} = A_{\alpha}$, again, up to an additive constant that is such that $\min[\Delta A_{\alpha}] = 0$.

Departing from standard ABP/ABF frameworks, we use ideas from importance sampling to write eqs 6 and 8 as time averages over biased trajectories

$$e^{-\beta A_{\alpha}(\xi^*, t)} = Z_t^{-1} \left(1 + \int_0^t \delta_{\alpha}(\xi(x_s) - \xi^*) e^{\beta V_b(\xi(x_s), s)} ds \right) \quad (10)$$

where Z_t is still a normalization constant ensuring eq 7, and

$$\frac{\partial A_{\alpha}(\xi^*, t)}{\partial \xi_j^*} = -k_B T \frac{\int_0^t \partial_{\xi_j} \delta_{\alpha}(\xi(x_s) - \xi^*) e^{\beta V_b(\xi(x_s), s)} ds}{1 + \int_0^t \delta_{\alpha}(\xi(x_s) - \xi^*) e^{\beta V_b(\xi(x_s), s)} ds} \quad (11)$$

The ABP method we discuss here is based on the biasing potential eq 9, updated with the current estimate of the free energy eq 10. New configurations are obtained by integrating in time the biased equations of motion using the simple estimate eq 11 for the biasing force. The convergence of this method is discussed in section 2.3.3.

In fact, eq 10 is a way to evaluate the convolution in eq 3 at each point ξ^* using a biased trajectory. This gives us a precise understanding of how using finite α introduces error in the estimate A_{α} and how to remove that error. This is a strength of our method. If we try to draw analogy with metadynamics, the framework of eq 10 would imply the continuous deposition of the Gaussians δ_{α} at each point $\xi(x_t)$ along the trajectory. Notice that, in this analogy, the Gaussians would be added to the DOS rather than to the bias potential, precluding us from going any further with the analogy.

2.3.2. Time Discretization. Let us briefly discuss the time discretization of the method based on eqs 9, 10, and 11. Assume that we have a suitable discretization where time is broken into parts of duration Δt so that $t = n\Delta t$ and $x_{i\Delta t}$ is written x_i . The biasing potential is now updated as

$$e^{\beta V_b(\xi, n)} = e^{-\beta \Delta A_{\alpha}(\xi, n)} e^{\beta c} \quad (12)$$

where $e^{-\beta \Delta A_{\alpha}(\xi, n)} = e^{-\beta A_{\alpha}(\xi, n) / \max_{\xi^*} [e^{-\beta A_{\alpha}(\xi^*, n)}]}$, and eqs 10 and 11 are respectively replaced by

$$e^{-\beta A_{\alpha}(\xi^*, n+1)} = Z_n^{-1} \left(1 + \sum_{i=0}^n \delta_{\alpha}(\xi(x_i) - \xi^*) e^{\beta V_b(\xi(x_i), i)} \right) \quad (13)$$

and

$$\frac{\partial A_{\alpha}(\xi^*, n+1)}{\partial \xi_j^*} = k_B T \frac{\sum_{i=0}^n \partial_{\xi_j} \delta_{\alpha}(\xi(x_i) - \xi^*) e^{\beta V_b(\xi(x_i), i)}}{1 + \sum_{i=0}^n \delta_{\alpha}(\xi(x_i) - \xi^*) e^{\beta V_b(\xi(x_i), i)}} \quad (14)$$

At $t = 0$ we have $\exp[-\beta A_{\alpha}(\xi, 0)] = 1/Z_0$. Let us emphasize again that the trajectory x_t is generated from a biased equation of motion associated with the biased potential $V + V_b$.

The implementation only requires storing the current value of the numerator and denominator of eq 14 at the points ξ^* . In particular, Z_n is never needed in practice (see Appendix A). The biasing force $-\nabla V_b$, needed for instance to integrate the Langevin dynamics, is obtained through eq 14.

2.3.3. Convergence and Consistency. It can be checked that, if the biasing potential V_b converges in the long-time limit, then it converges to $-A_{\alpha}$ up to an additive constant. Indeed, denoting by $\bar{A}_{\alpha}(\xi) = \lim_{t \rightarrow \infty} A_{\alpha}(\xi, t)$, the trajectory x_t is sampled according to the limiting canonical measure associated with the potential $V - \bar{A}_{\alpha} + C$ (where C is an unimportant constant), so that eq 10 leads to

$$\begin{aligned}
e^{-\beta \bar{A}_\alpha(\xi^*)} &= \lim_{t \rightarrow +\infty} \frac{1 + \int_0^t \delta_\alpha(\xi(x_s) - \xi^*) e^{\beta V_b(\xi(x_s), s)} ds}{\int_\Omega (1 + \int_0^t \delta_\alpha(\xi(x_s) - \xi') e^{\beta V_b(\xi(x_s), s)} ds) d\xi'} \\
&= \frac{\int_{\mathcal{Q}} \delta_\alpha(\xi(x) - \xi^*) e^{\beta(V(x)+C)} dx}{\int_\Omega \int_{\mathcal{Q}} \delta_\alpha(\xi(x) - \xi') e^{\beta(V(x)+C)} dx d\xi'} \\
&= e^{-\beta A_\alpha(\xi^*)}
\end{aligned}$$

The fact that, if a limit exists, then it is the correct one, is an important consistency check of the method. However, we were not able to prove that the biasing potential indeed converges (this seems difficult for ABP methods, while such an analysis can rigorously be done for some ABF methods²⁰).

Let us now look more carefully at the first iterations of the algorithm, in order to understand the role of the constant c in eq 9 or eq 12. We base our considerations on the numerical discretization eq 13 to simplify the argument. First, recall that the constant c does not change the longtime limit of the algorithm. However, it helps accelerate the convergence during the initial transient regime. The first iteration of eq 13 indeed shows that

$$e^{\beta V_b(\xi^*, 1)} = e^{\beta c} \frac{1 + \delta_\alpha(\xi(x_0) - \xi^*) e^{\beta c}}{1 + \delta_\alpha(0) e^{\beta c}}$$

When c is such that $e^{\beta c}$ is small, $V_b(\xi^*, 1)$ is raised by a small amount, and the gradient of V_b encourages trajectories to move away from ξ^* to some small extent. By increasing the value of c , we obtain a bias potential that pushes trajectories away from ξ^* more strongly, hence increasing the efficiency of the bias potential, in particular at the early stages of the process. We therefore conclude that the value of c should be as large as possible while maintaining numerical stability. Not all ABP methods update their biases according to this rule (see the comparison between our approach and the standard self-healing umbrella sampling (SHUS) algorithm in section 3.1).

To conclude this section, let us note that, once the bias is converged, it can be fixed and used as a bias to compute thermodynamic properties as in any importance sampling approach.

3. Comparison with Other Methods

3.1. Self-Healing Umbrella Sampling. SHUS⁹ can be seen as a special case of the method presented here. SHUS can be written in terms of eq 12 using the following time-dependent constant:

$$e^{\beta c(n)} = \max_{\xi^*} [e^{-\beta A_\alpha(\xi^*, n)}]$$

With this choice we have $V_b = -A_\alpha$ and $\int_\Omega e^{\beta V_b(\xi, t)} d\xi = 1$. This choice for $c(n)$ appears to be suboptimal which may be related to the analysis of section 2.3.3 that shows that the value of c should be as large as possible. Notice also that, when the reaction coordinate space is discretized into a finite number of bins, the normalization condition eq 7 should be restated as a sum over bin indexes and the maximal value of $\exp[-\beta A_\alpha(\xi, n)]$ is therefore less than 1. This corresponds to a negative value of $c(n)$. We checked for the test case considered in section 4 that our method outperforms SHUS for precisely this reason.

3.2. Adaptive Biasing Force. We compare numerically our approach to two ABF formulations in section 4.1. ABF is a good reference for comparison because there are no model parameters to choose. Errors arise only through time and reaction coordinate discretization. Two exact formulations of the free energy gradient are eq 4 above, and

$$F(\xi^*) = - \left\langle \frac{d}{dt} \left(M_\xi \frac{d\xi}{dt} \right) \right\rangle_{\xi^*} \quad (15)$$

where $M_\xi^{-1} = JMJ'$ with M being the mass matrix and J being defined in eq 4 (see ref 15 for this second expression). We point out that, in practice, $F(\xi)$ is approximated by a trajectory average $F(\xi, t)$, which is then used to bias the dynamics. For further details on the expressions eq 4 and eq 15 or their numerical implementation, we refer the reader to the cited works.

With ABF, one must address constructing the free energy from an estimation of its gradient, the calculated field F . While there are specific solutions to this problem^{15,21,22} we employ a standard variational formulation. We recast this question as an optimization problem where the objective function

$$I(u) = \int_\Omega \|F(\xi) - \nabla_\xi u\|^2 d\xi \quad (16)$$

is to be minimized. The corresponding Euler–Lagrange equation is

$$\Delta_\xi u(\xi) = \nabla_\xi \cdot F \quad (17)$$

which is just Poisson’s equation, to be supplemented with appropriate boundary conditions (depending on the problem at hand). The solution $u(\xi)$ is the best representation of the free energy given the vector field $F(\xi)$. This is solved via finite difference in the present work, but finite elements (or any Galerkin method) could be used as well.

3.3. Metadynamics. Because we have developed a method within the adaptive bias potential paradigm, we also make a comparison to well-tempered metadynamics.²³ In this formulation of metadynamics, the bias potential in one dimension is given by

$$V_b^{\text{meta}}(\xi, \tau) = \sum_{t' \leq \tau} G(\xi - \xi_{t'}, h(\xi, t'), w) \quad (18)$$

where the functions $G(X, H, W)$ are Gaussians of width W and height H , centered on X . We write V_b^{meta} to indicate that this is the bias potential generated by metadynamics. The Gaussian height in well-tempered metadynamics is both dependent on time and position along the reaction coordinate $h(\xi, t) = \omega \exp[-V_b^{\text{meta}}(\xi, t)/k_B \Delta T] \tau_G$. For details of this version of metadynamics, we refer the reader to ref 23. We compare to this particular formulation because it requires less interaction with the user and a choice of parameter values is given in the cited reference.

4. Numerical Examples

4.1. Simulation Details and Results. Alanine dipeptide is a familiar system for benchmarking sampling methods.^{15,21,23–26} Here, we employ AMBER with a half femtosecond time step, no constraints, solvent effects are modeled with the generalized

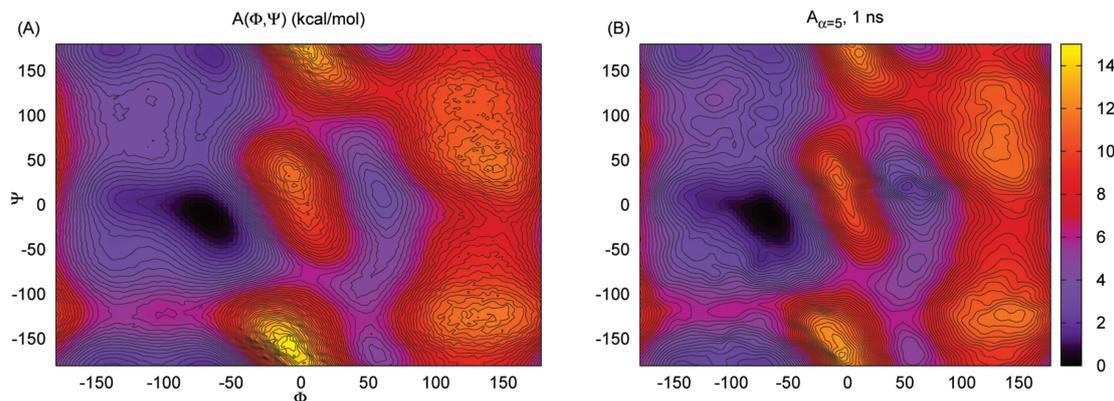


Figure 1. Contours are placed every $1/2k_B T$ (kcal/mol). (A) An estimate of the exact free energy (see text) which compares well to ref 15. (B) The free energy estimate after 1 ns of biased dynamics ($\alpha = 5$).

Born model, and we use the ff94 parametrization. The temperature was maintained at $T = 300$ K with Langevin dynamics where the collision frequency is 1 ps^{-1} . We select the common backbone dihedral angles $(\xi_1, \xi_2) = (\Phi, \Psi)$ as reaction coordinates.

When discretizing the reaction coordinate, it is common to use a small bin size to be sure that the free energy is correctly captured. Here, we use 300 bins of width 1.2° . We will also consider a bin width of 3.6° for eq 15 to examine the influence of bin size on ABF. In practice, for eqs 13 and 14, the current configuration along a trajectory may contribute only to an $m \times m$ grid centered around $(\xi_1(x_t), \xi_2(x_t))$, which amounts to truncating the range of the Gaussian function δ_α . The number of bins m were chosen so that $\delta'_\alpha(\xi - \xi^*)$ is negligible for ξ^* outside this box. For example, when $\alpha = 5^\circ$ we use $m = 20$. In practice we neglect the normalization Z_n as well as the normalization of δ_α . We give a schematic algorithm in Appendix A.

In simulations with eqs 13, 4, and 15, we use a “ramp function” $R(N_{l,k}) = \min[1, N_{l,k}/N_0]$ to scale the biasing force (see, for instance, ref 15), where $N_{l,k}$ is the population in the bin (l, k) and the parameter $N_0 = 10$ was optimized for eqs 4 and 15. The ramp function scales the biasing force so that the initially noisy observations of the force do not induce nonequilibrium effects. The biasing force for the method presented here is given by eq 14. The biasing force for the ABF methods are given in eqs 4 and 15. The biasing forces (and biasing potential) are updated at each time step.

To study sampling efficiency we use the average difference

$$d(t) = \frac{1}{n^2} \sum_{k=1}^n \sum_{l=1}^n |A_{\text{ref}}(k, l) - \hat{A}(k, l, t)| \quad (19)$$

between the estimated free energy and a reference to be defined below. n is the number of bins in each coordinate, k and l are bin indices. For eqs 4 and 15 \hat{A} is the solution of eq 17. In our method, \hat{A} is either the left-hand side of eq 13, A_α , or its deconvoluted version A_α^{deconv} . Finally, for eq 18, $\hat{A} = -(T + \Delta T)V_b^{\text{meta}}/\Delta T$. The reported results for $d(t)$ are found by using only a single trajectory with each method. We do not report the results obtained with SHUS since the convergence was found to occur much slower than for cases where $c > 0$.

We use the Richardson–Lucy algorithm^{27,28} to deconvolute A_α because of its simplicity, but another method of decon-

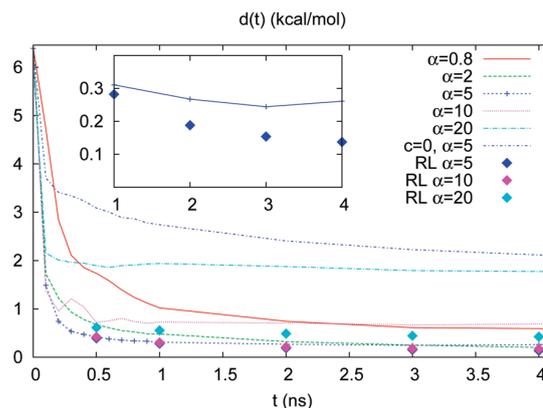


Figure 2. Error (eq 19) as a function of time for the method presented in this paper, for various α with and without deconvolution. Unless otherwise stated, $c = 15k_B T$. The $\alpha = 5^\circ$, $c = 0$ simulation demonstrates slow convergence due to a suboptimal choice of c , as described in the text. In the inset we show the last 4 ns of the $\alpha = 5^\circ$ results.

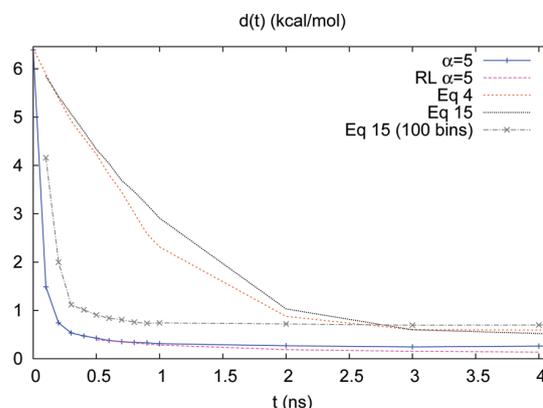


Figure 3. Error (eq 19) as a function of time for $\alpha = 5^\circ$ and $\alpha = 10^\circ$ with the method proposed in this paper, and comparison with ABF results obtained from eqs 4 and 15.

volution could be used, particularly if δ_α is not defined as a Gaussian. This algorithm is denoted by “RL” throughout. The RL algorithm uses the following iterative procedure:

$$f_{i+1}(\xi) = f_i(\xi) \int_{\Omega} \frac{e^{-\beta A_\alpha(\xi^*)}}{\int_{\Omega} \delta_\alpha(\hat{\xi} - \xi^*) f_i(\hat{\xi}) d\hat{\xi}} \delta_\alpha(\xi^* - \xi) d\xi^* \quad (20)$$

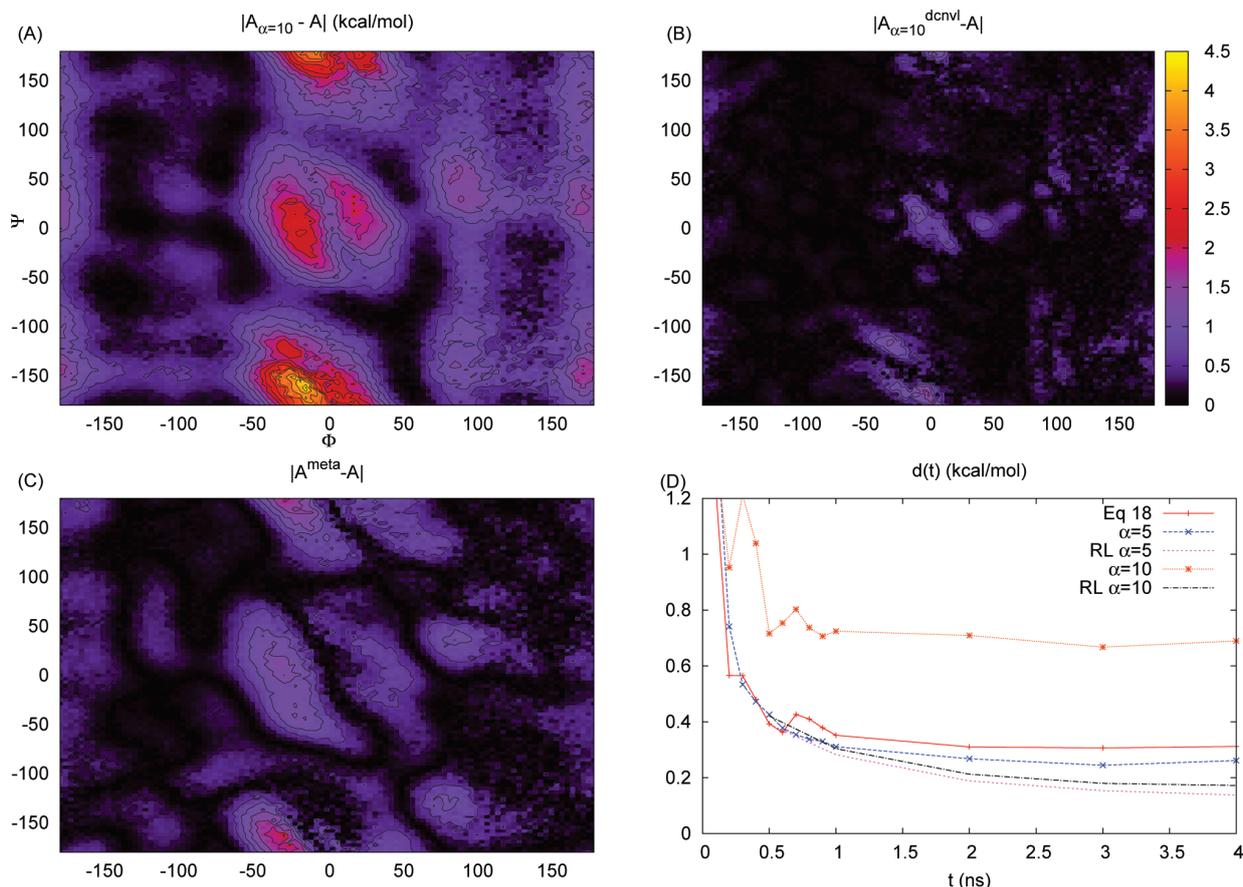


Figure 4. In panel A we show the absolute difference between the computed $A_{\alpha=10}$ and the reference A . Most of the error is, as expected, due to the regions of large curvature. In panel B the absolute difference between the deconvoluted free energy $A_{\alpha=10}^{\text{dcnvl}}$ and the reference A is shown. In panel C we show the absolute difference between $A^{\text{meta}} = -(T + \Delta T)V_b^{\text{meta}}/\Delta T$ and the reference A . The free energy estimates in panels A–C were taken at the end of a 4 ns trajectory. In panel D eq 19 is shown for the well-tempered metadynamics method (eq 18) for comparison with results from the mollified DOS method.

where $f_0(\xi) = \exp[-\beta A_\alpha(\xi)]$, which is given by eq 13. To begin the algorithm, δ_α and f_i must be normalized. The fixed-point iteration eq 20 suggests that $f_n(\xi) \rightarrow \exp[-\beta A(\xi)]$ as $n \rightarrow +\infty$. We use 10 iterations in the reported results.

The reference free energy was computed by reweighting a long biased trajectory (120 ns) as

$$A_{\text{ref}}(k, l) = -k_B T \ln \left(\sum_i \delta[\xi_k - \xi_1(x_i)] \times \delta[\xi_l - \xi_2(x_i)] \exp[\beta V_b(\xi(x_i), \tau)] \right)$$

where $V_b(\xi, \tau)$ was constructed from 4 ns of simulation with the mollified DOS method. The bias was not updated during the construction of the reference free energy. This produces a result free from errors associated with the choice of α . The reference profile A_{ref} is shown in Figure 1A, and in Figure 1B we show $A_\alpha(\xi, t)$ at 1 ns of sampling with $\alpha = 5^\circ$. The average difference $d(t)$ is shown in 1 for A_α with different values of α . To show how the zero of energy of the bias potential controls the speed of convergence, in Figure 2 we plot eq 13 with $c = 0$ in eq 12 and we set $c = 15k_B T$ for the remaining simulations. In Figure 3 we show $d(t)$ for eqs 4 and 15 (ABF methods). Results for eq 18 (well-tempered metadynamics) are shown in Figure 4D.

4.2. Efficiency of the Results As a Function of α . For small α , the nonlocality of the formulation disappears, and in Figure 2 we see slow convergence for $\alpha = 0.8^\circ$. For intermediate values of α , nonlocality allows the bias potential to equilibrate much faster. For $\alpha = 2^\circ$ and $\alpha = 5^\circ$, A_α is a good approximation of A , $d(t)$ falls well under 1 kcal/mol, and we observe high efficiency. With the value $\alpha = 10^\circ$, $d(t)$ plateaus at roughly 1 kcal/mol; α is now too large for A_α to be a good approximation of A . After applying the RL deconvolution to $A_{\alpha=10^\circ}$, $d(t)$ drops to match the accuracy obtained with $\alpha = 2^\circ$ or $\alpha = 5^\circ$. The correspondence between A_α and A has deteriorated but not enough to decelerate the sampling: $A_{\alpha=10^\circ}$ is still a good biasing potential and A can be recovered with deconvolution even at very short times.

For large α , eq 14 approaches zero, leaving only a small biasing force to accelerate the dynamics. To assess whether $\alpha = 20^\circ$ is so large as to slow down the sampling, we apply the RL deconvolution. The results in Figure 2 demonstrate that A can be recovered to high accuracy for $\alpha = 20^\circ$ at long times but that sampling efficiency is affected.

In Figure 3 we show $d(t)$ for eq 4 and eq 15 with a bin size of 1.2° and also for eq 15 with a bin width of 3.6° . If we compare the time to reach $d(t) = 1$ kcal/mol, simulation with eq 13 is roughly 3–10 times faster than that with eqs 4 and 15 for $2 \leq \alpha \leq 20$ at the bin size of 1.2° . For the larger bin size 3.6° , ABF sampling speed becomes competitive with the

mollified DOS approach, but it is impossible to remove the error. The 3.6° bin width coincides with the Gaussian half-width of δ_α when $\alpha = 2^\circ$. A larger bin size can enhance sampling speed for ABF but at a cost in accuracy. Note that $\alpha = 20^\circ$ corresponds to a δ_α with a half width that spans 33.3° in one dimension. This is a very large effective bin width for the accuracy of the results; a similar bin size with eq 4 or eq 15 would produce large, irreparable errors.

In Figure 4 we show results for the metadynamics simulations. We use the values $\Delta T = 1800$ K, $\omega = 0.24$ cal mol $^{-1}$ fs $^{-1}$, and $\tau_G = 120$ fs, as suggested in ref 23. We could not improve the results by choosing different parameters. In panels A and B of Figure 4, we show the absolute difference between the computed $A_{\alpha=10}$ and the reference A with and without deconvolution, respectively. Clearly, the bulk of error is due to the misrepresentation of the very negatively curved regions of the free energy, and the ability to deconvolute A_α drastically reduces this error. In panel C we show the absolute difference between the free energy computed via eq 18 and the reference. We see again that the error is concentrated in the regions of large negative curvature but there is no simple and obvious way to reduce these errors with some postprocess. Panel D confirms that the metadynamics promotes extremely rapid sampling but that the long time accuracy, especially in strongly curved regions, is limited.

The results summarized in Figures 2, 3, and 4 imply that a wide range of values $2 \leq \alpha \leq 20$ lead to good efficiency. The ability to use the simple deconvolution algorithm is a clear advantage of the method.

4.3. Choosing α A Priori. We now discuss how to *a priori* choose α based on some rough error estimates. Taking ξ as a scalar, we may expand $e^{-\beta A(\xi^*)}$ as a Taylor series. Equation 3 yields

$$e^{-\beta(A_\alpha(\xi^*) - A(\xi^*))} \approx 1 + \frac{\alpha^2}{4} \left[\left(\frac{A'(\xi^*)}{k_B T} \right)^2 - \frac{A''(\xi^*)}{k_B T} \right] \quad (21)$$

where we keep terms up to the second moment of δ_α . Assuming that $A(\xi)$ is harmonic near the minimum $\xi = q$, the curvature can be estimated as $A''(q) = k_B T / \sigma^2$, where σ^2 is the variance of the reaction coordinate at temperature T . From eq 21,

$$\exp[-\beta(A_\alpha(q) - A(q))] \approx 1 - \frac{\alpha^2}{4\sigma^2}$$

While the higher order terms and the regions where $A' \neq 0$ are certainly important to the total error, this motivates defining α as a function of σ if little is known about the free energy; we can always calculate σ in the initial state.

We calculate the variance of the reaction coordinates to be about $\sigma^2 = 340$ deg 2 for both Φ and Ψ . In terms of the values of α discussed above, this implies that $\sigma/9 \leq \alpha \leq \sigma/2$ is a good range for fixing α from calculation of σ . Of course, different α 's may also be used for different coordinates.

5. Conclusion

In conclusion, we have developed and tested an efficient ABP scheme. The nonlocality of δ_α leads to a bias potential and a bias force that equilibrate rapidly. Shifting the zero of energy on the bias potential was shown to result in efficient importance sampling. The parameter c has influence on only the efficiency

of the importance sampling but not on the limiting error of A_α . Because the bias potential is related to a convoluted free energy, deconvolution can be applied at the end of a simulation to remove all of the errors associated with the choice of the model parameter α - a unique feature and strength of this approach. This is limited only by the extent of sampling and the spacial discretization. This scheme easily accommodates the computation of the free energy surface and free energy gradient in several dimensions. We also suggest a simple means of *a priori* specifying α and c that should be quite general in applicability.

Acknowledgment. This work is supported by the Agence Nationale de la Recherche, under grants ANR-09-BLAN-0216-01 (MEGAS) and ANR-06-CIS-014 (SIRE).

A Schematic Algorithm

To help illustrate the simplicity of implementing eq 14 from the text

$$\frac{\partial A_\alpha(\xi^*, n+1)}{\partial \xi_j^*} = k_B T \frac{\sum_{i=0}^n \partial_{\xi_j^*} \delta_\alpha(\xi(x_i) - \xi^*) e^{\beta V_b(\xi(x_i), i)}}{1 + \sum_{i=0}^n \delta_\alpha(\xi(x_i) - \xi^*) e^{\beta V_b(\xi(x_i), i)}} \quad (22)$$

for a two-dimensional computation, we give a schematic algorithm here. We first define some array names. Let the array named "pop(k, l)" store the population at the (k, l) grid point (this is just the denominator of eq 22 above), where k corresponds to the bin index of $\xi_1(x_i)$ and l corresponds to the bin index of $\xi_2(x_i)$. Let the array named "dpop(j, k, l)" hold the derivative along the $\xi_{j=1,2}$ direction at the point (k, l). The array "dpop" is simply the numerator of eq 22 above. We use "dA(k)" to store the gradient of the free energy at the present point (k, l). We assume that α has been calculated and c has been specified. We let $\delta_\alpha(\xi) = e^{-|\xi|^2/\alpha^2}$, which amounts to ignoring the normalization of the Gaussian functions. Lastly, we denote the trajectory in phase space as x_i , $F(n')$ is the force along the n' th degree of freedom, d/dn' is the derivative with respect to the n' th degree of freedom, and we use $V(x)$ for the potential energy.

First we initialize the arrays.

$$\begin{aligned} t = 0, \text{ pop}(k, l) &= 1 \quad \forall k, l \\ \text{and dpop}(j, k, l) &= 0 \quad \forall k, l, j \\ \text{and } M &= 1 \end{aligned}$$

where $M = \max_{k,l} [\text{pop}(k, l)]$. Each time the molecular dynamics forces are computed, we must also compute the current biasing information. Notice that we define everything in terms of the "pop" and "dpop" arrays so that no array is needed for the bias potential and that $M = \max_{k,l} [\text{pop}(k, l)]$ can be updated without looping over the full reaction coordinate domain.

! evaluate free energy gradient at (k, l) for $j = 1, 2$

dA(j) = dpop(j, k, l)/pop(k, l)

! add bias forces to the existing forces and use a

! "Ramp function" R as described in the text

$R = \min(1, \text{pop}(k, l)/10)$

$F(n') = F(n') + R \sum_{j=1}^2 \text{dA}(j) d\xi_j/dn'$

! evaluate the weighting factor W for updating "pop" and "dpop"

```

W = exp[β Vb] = exp[β c]pop(k,l)/M
! update "pop" and "dpop" on an m by m grid
loop k' = k - m/2, k + m/2
  loop l' = l - m/2, l + m/2
    pop(k',l') = pop(k,l) + δα(ξ1(xi) - ξ1,k'*)δα(ξ2(xi) - ξ2,l'*)W
    if pop(k',l') > M then M = pop(k',l')
    loop j = 1,2
      dpop(j,k',l') = dpop(j,k,l) + δξj*[δα(ξ1(xi) - ξ1,k'*)δα(ξ2(xi) - ξ2,l'*)]W

```

We have defined k' and l' so that "pop" and "dpop" are updated on an $m \times m$ grid as discussed in the text. The treatment of (k', l') should reflect whether the domain is assumed to be periodic or not. The approximate free energy A_α is recovered (up to an additive constant) with $A_\alpha = k_B T \ln[\text{pop}(k, l)/M]$.

The dynamics will now evolve in the presence of the biasing force $dA(j)$, while the arrays "pop" and "dpop" hold unbiased estimates of the population and the derivatives of the population. Notice that the free energy gradient is reduced to a simple ratio, and the only difficulty lies in the careful treatment of the loops over the grid points k' and l' . The often mathematically complex computation of the free energy and free energy gradient is reduced to simple bookkeeping.

References and Notes

- Hänggi, P.; Talkner, P.; Borkovec, M. *Rev. Mod. Phys.* **1990**, *62*, 251–341.
- Berg, B. A.; Neuhaus, T. *Phys. Lett. B* **1991**, *267*, 249–253.
- Berg, B. A.; Neuhaus, T. *Phys. Rev. Lett.* **1992**, *68*, 9–12.
- Hansmann, U.; Okamoto, Y.; Eisenmenger, F. *Chem. Phys. Lett.* **1996**, *259*, 321–330.
- Wang, F.; Landau, D. P. *Phys. Rev. E* **2001**, *64*, 056101:1–16.
- Wang, F.; Landau, D. P. *Phys. Rev. Lett.* **2001**, *86*, 2050–2053.
- Darve, E.; Pohorille, A. *J. Chem. Phys.* **2001**, *115*, 9169–9183.
- Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.
- Marsili, S.; Barducci, A.; Chelli, R.; Procacci, P.; Schettino, V. *J. Phys. Chem. B* **2006**, *110*, 14011–14013.
- Maragliano, L.; Vanden-Eijnden, E. *Chem. Phys. Lett.* **2006**, *426*, 168–175.
- Lelièvre, T.; Rousset, M.; Stoltz, G. *J. Chem. Phys.* **2007**, *126*, 134111:1–8.
- Zheng, L.; Yang, W. *J. Chem. Phys.* **2008**, *129*, 014105.
- Zheng, L.; Chen, M.; Yang, W. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 20227–20232.
- Zheng, L.; Chen, M.; Yang, W. *J. Chem. Phys.* **2009**, *130*, 234105.
- Darve, E.; Rodriguez-Gomez, D.; Pohorille, A. *J. Chem. Phys.* **2008**, *128*, 144120:1–13.
- Hansmann, U.; Eisenmenger, F.; Okamoto, Y. *Chem. Phys. Lett.* **1998**, *297*, 374–382.
- den Otter, W.; Briels, W. *J. Chem. Phys.* **1998**, *109*, 4139–4148.
- Sprik, M.; Ciccotti, G. *J. Chem. Phys.* **1998**, *109*, 7737–7744.
- Ciccotti, G.; Lelièvre, T.; Vanden-Eijnden, E. *Commun. Pure Appl. Math.* **2008**, *61*, 371–408.
- Lelièvre, T.; Rousset, M.; Stoltz, G. *Nonlinearity* **2008**, *21*, 1155–1181.
- Maragliano, L.; Vanden-Eijnden, E. *J. Chem. Phys.* **2008**, *128*, 184110:1–10.
- Kästner, J. *J. Chem. Phys.* **2009**, *131*, 034109:1–8.
- Barducci, A.; Bussi, G.; Parrinello, M. *Phys. Rev. Lett.* **2008**, *100*, 020603:1–4.
- Bartels, C.; Karplus, M. *J. Comput. Chem.* **1997**, *18*, 1450–1462.
- Strodel, B.; Wales, D. *Chem. Phys. Lett.* **2008**, *466*, 105–115.
- Li, X.; Latour, R. A.; Stuart, S. J. *J. Chem. Phys.* **2009**, *130*, 174106:1–9.
- Richardson, W. *J. Opt. Soc. Am.* **1972**, *62*, 55–59.
- Lucy, L. *Astron. J.* **1974**, *79*, 745–754.

JP100926H

Part III

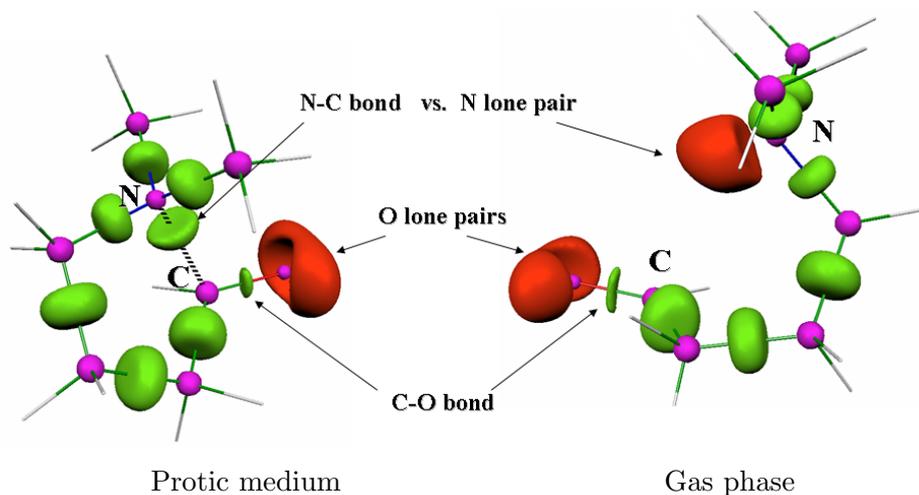
The one with some applications

5) The one with the unusual bond

The human immunodeficiency virus aspartic protease (HIV-1 PR) is one of the major targets for the design of anti-AIDS medicines.¹ Indeed, this enzyme was early shown to be essential for proper virion assembly and maturation, and its inhibition leads to the production of non-infectious particles. Since 1995, several HIV PR inhibitors were approved by the FDA and enabled to prolong significantly the life expectancy of HIV infected patients. Nevertheless, the rapid emergence of drug resistance via multiple HIV PR mutations leads to limited potency. One way to overcome this serious hurdle could be to provide new drugs that exhibit unusual binding modes.

Hasserodt *et al* proposed a new concept of aspartic proteases inhibitor based on a non covalent N \cdots CO bond,² that should be able to interact strongly with a localized region of the active site. They assumed that this unusual interaction would provide a better transition state analog than the usual hydroxyethylene moiety, that is involved in most of HIV PR inhibitors. Unfortunately, up to now, the inhibiting power of these compounds is quite low.

We thus studied the electronic structure of the N \cdots CO weak bond by means of Electron Localization Function and ab initio calculations. We have shown that the bond formation is driven by the enhancement of the ionic contribution C⁺-O⁻ induced by the strong polarization effect of the near N lone pair. This ionic configuration is stabilized by a polar solvent, and even more by a protic solvent via small but cumulative charge transfer from the negatively charge oxygen to the surrounding water molecules.



¹a) Wlodawer, A.; Vondrasek, J. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 249-284. b) Fitzgerald, P. M. D.; Springer, J. P. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 299-320.

²a) Gautier, A.; Pitrat D.; Hasserodt, J. *Bioorg. Med. Chem.* **2006**, *14*, 3835-3847. b) Waibel, M.; Hasserodt, J. *J. Org. Chem.* **2008**, *73*, 6119-6126. c) Waibel, M.; Pitrat, D.; Hasserodt, J. *Bioorg. Med. Chem.* **2009**, *17*, 3671-3679

DOI: 10.1002/chem.200601250

Unusual Bond Formation in Aspartic Protease Inhibitors: A Theoretical Study

 Julien Pilmé, Hélène Berthoumieux, Vincent Robert, and Paul Fleurat-Lessard*^[a]

Abstract: The origin of the formation of the weak bond $N|\cdots C=O$ involved in an original class of aspartic protease inhibitors was investigated by means of the electron localization function (ELF) and explicitly correlated wavefunction (MRCI) analysis. The distance between the electrophilic C and the nucleophilic N centers appears to be controlled directly by the polarity and proticity of the medium. In light of these investigations, an unusual dative N–C

bonding picture was characterized. Formation of this bond is driven by the enhancement of the ionic contribution $C^+ - O^-$ induced mainly by the polarization effect of the near N lone pair, and

to a lesser extent by a weak charge delocalization $N \rightarrow CO$. Although the main role of the solvating environment is to stabilize the ionic configuration, the protic solvent can enhance the $C^+ - O^-$ configuration through a slight but cumulative charge transfer towards water molecules in the short N–C distance regime. Our revisited bond scheme suggests the possible tuning of the N–CO interaction in the design of specific inhibitors.

Keywords: ab initio calculations · chemical bonding · density functional calculations · ELF (electron localization function) · protease inhibitors

Introduction

Understanding the nature of chemical bonding has always fascinated chemists.^[1] Even though strong bonds are now well understood, the debate around the origin of weak bonds is always considerable, a prototype of such being the hydrogen bond.^[2] In this context, the interaction of a tertiary amino-group nitrogen with a carbonyl double bond, that is, $N|\cdots C=O$, has attracted much attention over several decades. This particular bond was first postulated by Kermack and Robinson to explain the low activity of the carbonyl

group in the cryptopine and protopine alkaloids.^[3] Later, it was seen in other alkaloids as a flexible ring clasp. Since then, this intriguing interaction has appeared in many areas of chemistry and biochemistry. In particular, it was used by Bürgi et al. to map the reaction coordinate of the nucleophilic addition to a carbonyl group.^[4] It was also postulated in the challenging issue of the enantioselective heterogeneous hydrogenation of aldehydes, in which it links the CO moiety to the chiral modifier of the catalyst.^[5] Because peptide bonds are the key to protein structure and activity, biological analogues based on this interaction can be anticipated.^[6] Recently, the $N|\cdots C=O$ interaction was used as the central part of new candidates for inhibitors of aspartic proteases.^[7] These enzymes are known to have key functions in the replication of the HIV virus and the development of Alzheimer's disease. The $N|\cdots C=O$ interaction is interesting as it should lead to inhibitors with not only steric but also electronic properties that mimic those of the transition-state structure. Indeed, the standard description of this bond involves a $N^{\delta+} - C - O^{\delta-}$ charge separation that closely mimics that arising during an amide-bond hydrolysis (Figure 1). Because this first candidate revealed a low inhibiting power ($IC_{50} > 1 \mu\text{mol}$), one might wonder what the source (electronic or steric) of this deficiency might be. Therefore, a detailed understanding of this particular bond would be highly desirable in the design of this new class of inhibitors.

[a] Dr. J. Pilmé, H. Berthoumieux, Dr. V. Robert, Dr. P. Fleurat-Lessard
 Laboratoire de Chimie, UMR CNRS 5182
 Ecole Normale Supérieure de Lyon
 46 Allée d'Italie, 69364 Lyon Cedex 07 (France)
 Fax: (+33)472-728-860
 E-mail: Paul.Fleurat-Lessard@ens-lyon.fr

Supporting information for this article is available on the WWW under <http://www.chemeurj.org/> or from the author: Table S1 lists the main structural parameters for the $\text{Me}_2\text{N}-(\text{CH}_2)_3-\text{CH}=\text{O}$ model system optimized at the B3LYP/6-31+G(d) level solvated by up to four water molecules with and without PCM; Table S2 gathers the distances and binding energies for the bimolecular system at different levels of calculation (see text); listing of cartesian coordinates for the main optimized structures (gas phase, PCM) obtained at the B3LYP/6-31+G(d) level; listing of cartesian coordinates at the 6-31+G* and aug-cc-pVTZ for the bimolecular test system $\text{NMe}_3|\cdots\text{H}_2\text{CO}$.

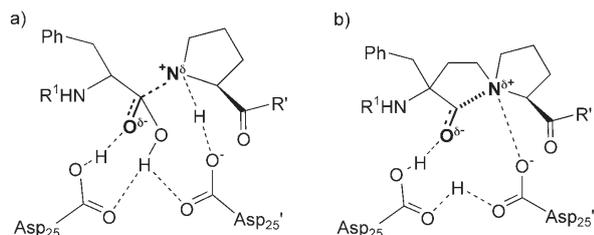


Figure 1. a) Transition state of the enzymatic reaction. b) Inhibitor candidate employing the N \cdots CO bond.

Starting with the milestone studies of Leonard,^[8] many experimental groups have investigated this interaction by using various techniques.^[9–12] Surprisingly, the N–C distance is very sensitive to the environment (e.g., solvation effects),^[12c] with variations between 1.6 and 2.9 Å being reported.^[4,13] Experimental values are split into two groups, corresponding to a short-distance regime ($d_{\text{CN}} \leq 2.0$ Å) and a long-distance regime ($d_{\text{CN}} \geq 2.5$ Å). From the theoretical side, all parameterization using standard force fields have been unsuccessful so far.^[6a,9d] Ad hoc empirical corrections were necessary to reproduce approximately molecular structures of cryptopine and clivorine, excluding any transferable parameters.^[6a] This, added to the extreme variability of the bond length, may suggest a more subtle bonding picture than the traditional $n_{\text{N}}(\sigma\text{-donor}) \rightarrow \pi^*_{\text{CO}}(\text{acceptor})$ scheme.^[8,9]

Considering the tremendous importance and unresolved character of this particular interaction, the purpose of the current work is to provide a further understanding of the intrinsic nature of this bond and to extract information on the relationship between the solvation effects and the observed distances regimes. Thus, a model system $\text{Me}_2\text{N}-(\text{CH}_2)_3-\text{CH}=\text{O}$ was chosen as a realistic candidate of amino aldehyde molecules^[12c] to scrutinize the nature of the $\text{N}|\cdots\text{C}=\text{O}$ bond (Figure 2).

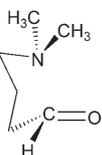


Figure 2. Model system of the tertiary amine-carbonyl bond.

In this context, one might wish to call for a detailed electronic description based on state-of-the-art calculations. The topological analysis of the electron localization function (ELF)^[14] based on density functional theory (DFT) calculations appears as a natural and powerful tool to describe the chemical bond. In addition to the ELF analysis, explicitly correlated wave-function analysis obtained from ab initio multi reference configuration interaction (MRCI) calculations were also considered.

Computational Details

DFT calculations were performed by using the Gaussian03 program.^[15] The hybrid density functional B3LYP was used^[16] with the standard basis set 6-31+G(d)^[17] for all atoms. The diffuse functions are compulsory to

correctly describe the ionic form of the carbonyl moiety. Comparative studies have been also carried out with the aug-cc-pVTZ basis sets.^[18] Both basis sets display very similar results in the geometry optimizations. To validate the use of this particular functional, tests have been performed with BPW91^[19] and MPWB1K^[20] functionals. Furthermore, MP2 and CCSD(T) calculations were carried out to validate our DFT approach. The polarizable continuum model (PCM) of Tomasi et al.,^[21] using the integral equation formalism, was used for calculations of the solvated model.

The binding energy (BE) values were corrected of the basis set superposition error (BSSE). As BE and BSSE are not easily accessed for an intramolecular bond, they were estimated by using the bimolecular test system $\text{Me}_3\text{N}|\cdots\text{H}_2\text{CO}$. By using the standard counterpoise correction,^[22] BSSE was estimated to be 1.2 kcal mol⁻¹ at the B3LYP/6-31+G(d) level.

By using the DFT optimized structures, we then performed MRCI wave-function analysis to investigate the contributions of polarization and charge-transfer effects. Complete active space self-consistent (CASSCF) simulations were performed by using the MOLCAS package, optimizing simultaneously the orbitals and the wave-function coefficients for a four-electron–three-molecular-orbital (MO) active space.^[27] Dynamical correlation effects were then estimated by including single and double excitations in a variational procedure on top of the CASSCF wave function, as available in the CASDI code.^[28] The Cowan–Griffin ab initio pseudopotential and basis sets of double zeta+diffuse+polarization quality were used on N (5s5p1d)→[2s3p1d], C (5s5p1d)→[2s3p1d]→[2s4p1d] atoms,^[29] whilst H atoms were depicted with minimal basis set (STO-3G) [1s].^[30] The CAS(4,3) validity was checked by including more active orbitals and electrons up to a CAS(8,7). The added orbitals remained either doubly occupied or vacant

Bonding Analysis

Two distinct approaches were considered for the bonding analysis. First, in addition to the DFT wave function, the ELF approach has been used extensively for the analysis of chemical bonding or chemical reactivity.^[23] The relationship of the ELF function to pair functions has been demonstrated,^[24] however, ELF values can be calculated more easily than the characteristics of pair functions. The ELF function can be interpreted as a signature of the electronic-pair distribution, split into an intuitive chemical picture: core- (labeled C(A)), bonding- (labeled V(A, B)), and nonbonding- (labeled V(A)) region pairs (the so-called basins). This is to be contrasted with the “atoms in molecules theory”,^[25] in which basins are localized on the atoms only. These ELF regions match closely the domains of the VSEPR model.^[26] Therefore, the ELF analysis should enable us to follow the change in interaction nature. The basin (atomic or ELF) populations are calculated by integrating the charge density over the basin volume.

Second, the correlated wave function extracted from our MRCI calculations was analyzed in terms of the ionic, covalent, and charge-transfer components.

All the topological analyses were carried out from a Kohn–Sham wave function with the TopMod^[14c] package. The ELF and MO isosurfaces were visualized by using the Molekel^[31] software.

Results and Discussion

B3LYP calculations of our model system lead us to the qualitative and nonambiguous conclusion that the N–C distance appears clearly controlled by environmental effects. Indeed, in the gas phase, the DFT optimization of our model system exhibits rather long N–C (2.86 Å) and short C–O (1.22 Å) distances (see Table S1 in the Supporting Information). Mimicking the physiological medium by using a polarizable continuum model (PCM with $\epsilon_r=78.4$) shortens the N–C distance to 2.54 Å. Nevertheless, such a value is much larger than the typical N–C experimental distances observed in amino ketone derivatives.^[12c] This result strongly suggests that one has to take into account explicitly the protic character of the medium. Indeed, calculations using an embedding continuum ($\epsilon_r=78.4$) together with a single water molecule to account explicitly for the protic character of the solvent exhibit a drastic shortening of the N–C distance (1.81 Å). This single result supports the stabilizing effect of the protic environment favoring the short-distance regime. Simultaneously, unusually long C–O distances (see Table S1 in the Supporting Information) and significant pyramidalization of the carbon atom (>30% sp^3 character) are observed. This is to be contrasted with the Bürgi–Dunitz NCO angle^[4] that remains close to 110° for all N–C distances. These structural effects are related to an increase in the binding energy from 2 kcal mol⁻¹ (gas phase) to 9 kcal mol⁻¹ (PCM and one water molecule), that is, twice the average hydrogen-bond value.

To confirm this trend, several additional calculations were performed by using other methods (BPW91, MPWB1K, and MP2) with 6-31+G(d) and aug-cc-pVTZ basis sets for the bimolecular model system NMe₃⋯H₂CO. The main results (distances and binding energies) are presented in Table S2 in the Supporting Information and reveal qualitatively similar results: upon addition of one water molecule and a continuum, the system switches from the long-distance regime (>2.5 Å) to the short-distance regime (<2.0 Å), together with a significant increase in the binding energy. These conclusions were confirmed by a reference CCSD(T) optimization of the bimolecular system with one water molecule embedded in a continuum: $d_{CN}=1.653$ Å, BE=11 kcal mol⁻¹ (BSSE corrected), compared to B3LYP values $d_{CN}=1.715$ Å and BE=9 kcal mol⁻¹ (BSSE corrected). All of these calculated parameters are in excellent accord with experimental studies.^[4,13]

To understand the role of the environment, we then conducted some calculations by using different models. Previous theoretical studies have shown that the first solvation shell of the carbonyl-O lone pairs consists of two to three water molecules. This is to be contrasted with the environment of the hydroxide ion HO⁻, which involves up to four water molecules.^[32] Thus, we carried out calculations of our system in the presence of three or four water molecules. In each case, we also tested the effect of adding an embedding PCM, leading to four calculations. All optimized structures display short N–C distances (from 1.87 to 1.63 Å) and quite

large C–O distances (from 1.25 to 1.34 Å), whatever the environment (see Table S1 in the Supporting Information). Furthermore, convergence of the critical N–C distance is achieved (variation $\approx 2\%$) for an environment consisting of four water molecules with PCM. However, as the number of water molecules is increased, the effect of the PCM embedding is reduced. With four water molecules, the difference between the N–C distances of the two models (gas phase and PCM) is smaller (0.06 Å) than the dispersion observed experimentally^[4,12,13] in the short regime (0.20 Å). Thus, to analyze the bonding properties, we felt that the inclusion of only four explicit water molecules would be a reasonable model to describe both the polar and protic effects of the solvating environment (see Figure 3).

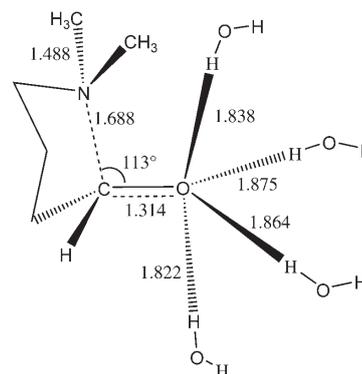


Figure 3. Model system in the presence of four water molecules. The distances [Å] and the Bürgi–Dunitz NCO angle [°] are given for the optimized B3LYP/6-31+G(d) calculation without PCM.

At this stage, the fundamental question of the intrinsic electronic nature of the intramolecular NCO bond in the two regimes deserves special attention. As our aim was to unravel the nature of the NCO interaction, we selected five N–C distances representative of the long-distance regime ($d_{CN}=3.0$ Å) and of the short-distance regime ($d_{CN}=1.90, 1.81, 1.70,$ and 1.63 Å, Table 1). This set of distances allows us to cover the experimental range of N–C distances observed for different compounds,^[4,12] as well as optimized values obtained with various computational approaches. For each selected distance, partial geometry optimizations were carried out both in the gas phase for the model system (see Figure 2) alone and in the presence of four water molecules (see Figure 3), the C–N distance remaining constant.

We then performed ELF analysis of the DFT wave function for both model systems. Explicitly correlated MRCI wave-function analysis (CASSCF+singles and doubles CI) were conducted to validate the conclusions of the ELF analysis. Such a strategy allows one to concentrate on the intimate nature of this interaction as a function of structural parameters controlled by environmental effects. Figure 4 displays the ELF localization domains of the model system for two typical N–C distances (1.81 and 3.0 Å).

Table 1. C–O bond length, ELF population, and correlated wave-function analysis for the optimized gas-phase model system for five selected N–C distances. The values given in parenthesis were calculated in the presence of four water molecules solvating the carbonyl group.

$d_{\text{CN}} [\text{\AA}]^{\text{[a]}}$	3.00	1.90	1.81	1.70	1.63
	CO bond-length evolution				
$d_{\text{CO}} [\text{\AA}]$	1.21 (1.23)	1.25 (1.28)	1.26 (1.30)	1.27 (1.31)	1.28 (1.32)
	ELF population analysis				
$\bar{N}V(\text{O})^{\text{[b]}}$	5.21 (5.26)	5.64 (5.70)	5.73 (5.85)	5.87 (5.94)	5.95 (6.03)
$\bar{N}V(\text{C},\text{O})^{\text{[c]}}$	2.38 (2.30)	2.05 (1.89)	1.94 (1.73)	1.82 (1.63)	1.71 (1.55)
$\delta q_{\text{N}\rightarrow\text{CO}}^{\text{[d]}}$	0(0)	0.1(0.1)	0.12(0.13)	0.14(0.15)	0.18(0.19)
	Correlated wave-function analysis				
% ionic	34	49	53	58	67
C–O/N–C ^{covalent}	>60	4.30	2.92	2.06	1.95

[a] Frozen during the optimization. [b] Oxygen lone-pair population [c] C–O bonding population. [d] Relative net charge transfer N→CO calculated with the atomic populations.^[16] [e] Covalent ratio (see text and ref. [35]).

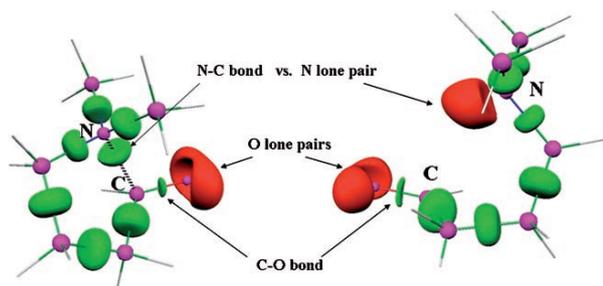


Figure 4. ELF localization domains (protonated bonds have been omitted for clarity) of the tertiary amine–carbonyl model system for two typical distances: left) $d_{\text{CN}}=1.81 \text{ \AA}$, right) $d_{\text{CN}}=3.0 \text{ \AA}$. Color code: magenta: core; red: nonbonding; green: bonding.

As illustrated in Figure 4, the pure nitrogen lone pair is observed at large N–C distances. There is no covalent interaction with the carbonyl moiety, suggesting a weak Van der Waals bond. Conversely, as the N–C distance shortens, the lone pair becomes progressively involved in a partially dative N–C bond, as one may expect.^[19] However, the calculated bonding basin is still centered close to the nitrogen atom and, interestingly, only a small contribution of the carbon atomic domain to the bonding population is observed (16% at $d_{\text{CN}}=1.63 \text{ \AA}$). This is much smaller than in a typical covalent peptide bond, such as $\text{Me}_2\text{N}-\text{CH}=\text{O}$, in which this contribution reaches up to 40%. Thus, our ELF analysis of the model system (Figure 2) supports a weak covalent character for the N–C interaction as opposed to the well-known scenario $n_{\text{N}} \rightarrow \pi_{\text{CO}}^*$. Let us stress that the global charge transfer from nitrogen towards the carbonyl fragment is positive, but rather small ($\leq 0.19 \text{ e}$). This is to be compared with the protonated system $\text{Me}_2\text{N}(\text{CH}_2)_3-\text{C}^+\text{H}-\text{OH}$ in which the equilibrium C–N distance equals 1.58 \AA , and the charge transfer is 50% larger (0.27 e). Meanwhile, the carbonyl moiety undergoes deep charge reorganization under the polarizing effect of the N lone pair. As seen in

Table 1, a simultaneous increase in the oxygen lone pair ($5.21 \text{ e} \rightarrow 5.95 \text{ e}$) and reduction in the C–O bond populations ($2.38 \text{ e} \rightarrow 1.71 \text{ e}$) are to be noticed along formation of the N–C bond. Therefore, the ELF population analysis of the model system sheds new light on the nature of the $\text{N}|\cdots\text{C}=\text{O}$ interaction, which, in our view, should be described as an enhancement of the ionic character of the C–O bond as the N–C distance decreases, ruling out the traditional $\text{N}^+-\text{C}-\text{O}^-$ picture resulting from the density transfer $n_{\text{N}} \rightarrow \pi_{\text{CO}}^*$.^[8,9]

At this stage, one can easily grasp why solvation plays such a role in determining the distance regime: a polar medium is needed to stabilize the ionic form. However, to fully clarify the influence of the water molecules, we performed an ELF population analysis of our system surrounded by four water molecules (Table 1). The populations of the O lone pair and N–C bond are affected mainly by the vicinity of the N atom, regardless of the presence of water molecules. Indeed, valence population changes induced by the addition of the water molecules are five times smaller than those due to shortening of the N–C bond (Table 1). This shows that the first role of the solvent is to stabilize the charge separation. As a consequence, the electronic distribution on the $\text{Me}_2\text{N}-(\text{CH}_2)_3-\text{CH}=\text{O}$ moiety does not depend upon the presence of the water molecules. This in turn fully supports our previous conclusion that the solvated NCO interaction is best described as $[\text{N}|\text{C}=\text{O} \leftrightarrow \text{N}|\cdots\text{C}^+-\text{O}^-]$. On top of this main effect, a very small net charge transfer (0.04 e) is observed from the CO bond to each water molecule at $d_{\text{CN}}=1.63 \text{ \AA}$. By cumulative effect, the C–O bond is depopulated ($1.71 \text{ e} \rightarrow 1.55 \text{ e}$), which induces a destabilization of the C–O covalent bond and consequently, an increase in the C–O distance. Finally, this analysis shows that the protic environment is able to enhance slightly the C^+-O^- configuration in addition to its main electrostatic stabilizing role.

Our previous ELF analysis confirmed that the main role of the medium is essentially electrostatic, the electronic distribution of the model system being very similar, whatever the protic environment. Consequently, MRCI calculations, using the DFT geometries, were conducted only on the model system of Figure 2. The particular splitting into ionic and covalent contributions is accessible directly from explicitly correlated methods that are very insightful into the analysis of bond-formation phenomena. Large configuration interaction (CI) calculations have been used with great success in the study of electronic properties of molecular and solid-state materials.^[33] It has been reported that DFT tends to systematically exaggerate delocalization effects,^[34] whereas CI approaches account accurately for subtle charge-reorganizations processes. Thus, with the goal of microscopic interpretation in mind, we felt that state-of-the-art ab initio calculations would be complementary to our combined DFT/ELF analysis. A complete active-space (four electrons/three orbitals) self-consistent field method was used to generate MOs for our subsequent correlation calculations, including single and double excitations. The speculated bond-formation mechanism $n_{\text{N}} \rightarrow \pi_{\text{CO}}^*$ suggests the active participation

of the N lone pair and the frontier orbitals of the carbonyl fragment, which are very similar to the π_{CO} and π^*_{CO} MOs (Figure 5).

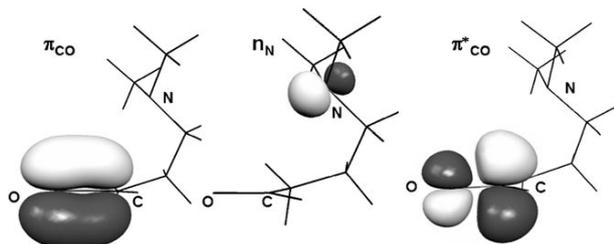


Figure 5. Frontier MOs of the tertiary amine-carbonyl model system.

These frontier MOs were reallocated on the C and O centers to grasp the charge reorganization in a valence-bond-type analysis. In agreement with the small charge transfer found in our ELF analysis ($\delta q_{\text{N} \rightarrow \text{CO}} \leq 0.18$), the n_{N} orbital remains almost doubly occupied (at least 1.997 e), excluding any significant $n_{\text{N}} \rightarrow \pi^*_{\text{CO}}$ mechanism. The correlated wave function (see Table 1) is dominated by the ionic contributions (i.e., C^+O^- and C^-O^+) at the short-distance regimes ($d_{\text{CN}} \leq 1.90 \text{ \AA}$). As for the covalent part, the participation of the carbon atom was analyzed by using the ratio of the covalent contributions^[35] of the C–O and N–C bonds. Even though bond formation is strongly suggestive of a four-electron/three-center scheme, the remarkable feature is a significant reduction (60 \rightarrow 1.95, Table 1) of this covalent-contribution ratio: as the C–N distance decreases, the carbon covalent participation is partly redirected from the CO bond to the forming CN bond. This in turn seems to be a determining process. In addition, this analysis of the correlated wave function confirms the nondispersive character of the N–C–O bond. Let us stress that the validity of the DFT approach is thus demonstrated a posteriori.

Conclusion

The nature of the $\text{N}|\cdots\text{CO}$ interaction was reconsidered extensively on the basis of combined ELF analysis and ab initio calculations. Our ELF topological analysis differentiates clearly the long and short N–C distance regimes. The four-electron/three-center picture that has been conveyed so far in the literature suffers from the absence of any strong $n_{\text{N}} \rightarrow \pi^*_{\text{CO}}$ charge transfer. From our point of view, the bond formation is driven clearly by the enhancement of the ionic contribution C^+O^- induced by the strong polarization effect of the near N lone pair. Conversely, the covalent character of the CN bond is still much weaker ($\approx 1/3$) than that of the CO bond. Our results obtained by including water molecules validate the $\text{N}|\cdots\text{C}^+\text{O}^-$ bonding scheme and clarify the fundamental role of the solvating environment, which stabilizes mainly the ionic C^+O^- configuration, but also enhances this configuration at the very short-distance

regimes. A suitable parameterization for such nondispersive interaction necessarily relies on a high-level description of the intimate intramolecular-charge reorganization. Our work supports strongly the role of the medium, whose effects can be theoretically tuned to account for the crucial water-excluding folding processes of biological systems. Finally, our results provide a revisited scheme for the $\text{N}|\cdots\text{CO}$ interaction. Thus, we hope that this bonding scheme will assist in the synthesis of more-efficient inhibitors.

- [1] G. N. Lewis, *J. Am. Chem. Soc.* **1916**, 38, 762.
- [2] M. E. Tuckerman, D. Marx, M. L. Klein, M. Parrinello, *Science* **1997**, 275, 817; S. B. Suh, J. C. Kim, Y. C. Choi, S. Yun, K. S. Kim, *J. Am. Chem. Soc.* **2004**, 126, 2186.
- [3] W. O. Kermack, R. J. Robinson, *J. Chem. Soc. Trans.* **1922**, 121, 427.
- [4] H. B. Bürgi, J. D. Dunitz, E. Scheffter, *J. Am. Chem. Soc.* **1973**, 95, 5065; H. B. Bürgi, J. D. Dunitz, E. Scheffter, *Nature New Biol.* **1973**, 244, 186; H. B. Bürgi, J. D. Dunitz, *Acc. Chem. Res.* **1983**, 16, 153.
- [5] J. W. de M. Carneiro, C. da S. B. de Oliveira, F. B. Passos, D. A. G. Aranda, P. Rogério N. de Souza, O. A. C. Antunes, *J. Mol. Catal. A* **2001**, 170, 235; M. Studer, H.-U. Blaser, C. Exner, *Adv. Synth. Catal.* **2003**, 345, 45.
- [6] a) R. Griffith, J. B. Bremner, S. J. Titmuss, *J. Comput. Chem.* **1997**, 18, 1211; b) R. Griffith, B. F. Yates, J. B. Bremner, S. J. Titmuss, *J. Mol. Graph. Model.* **1997**, 15, 91; c) R. Griffith, J. B. Bremner, *J. Comput.-Aided Mol. Des.* **1999**, 13, 69.
- [7] A. Gautier, D. Pitrat, J. Hasserodt, *Bioorg. Med. Chem.* **2006**, 14, 3835.
- [8] a) N. J. Leonard, *Rec. Chem. Prog.* **1956**, 17, 243, and references therein; N. J. Leonard, *Acc. Chem. Res.* **1979**, 12, 423, and references therein.
- [9] a) G. Spanka, P. Rademacher, *J. Org. Chem.* **1986**, 51, 592; b) G. Spanka, P. Rademacher, H. Duddeck, *J. Chem. Soc. Perkin Trans. 2* **1988**, 2119; c) P. Rademacher, *Chem. Soc. Rev.* **1995**, 143; d) G. Spanka, R. Boese, P. Rademacher, *J. Org. Chem.* **1987**, 52, 3362.
- [10] M. R. Bell, S. Archer, *J. Am. Chem. Soc.* **1960**, 82, 151; M. R. Bell, S. Archer, *J. Am. Chem. Soc.* **1960**, 82, 4638.
- [11] R. McCrindle, A. J. McAlees, *J. Chem. Soc. Chem. Commun.* **1983**, 61.
- [12] a) A. J. Kirby, I. V. Komarov, N. Feeder, *J. Chem. Soc. Perkin Trans. 2* **2001**, 522; b) A. J. Kirby, I. V. Komarov, V. A. Bilenko, J. E. Davies, J. M. Rawson, *Chem. Commun.* **2002**, 2106; c) J. E. Davies, A. J. Kirby, I. V. Komarov, *Helv. Chim. Acta* **2003**, 86, 1222.
- [13] a) S. R. Hall, F. R. Ahmed, *Acta Cryst. B* **1968**, 24, 337; S. R. Hall, F. R. Ahmed, *Acta Cryst. B* **1968**, 24, 346; b) J. A. Wunderlich, *Acta Cryst.* **1967**, 23, 846; c) G. I. Birnbaum, *J. Am. Chem. Soc.* **1974**, 96, 6165.
- [14] a) A. D. Becke, K. E. Edgecombe, *J. Chem. Phys.* **1990**, 92, 5397; b) B. Silvi, A. Savin, *Nature* **1994**, 371, 683; c) S. Noury, X. Krokidis, F. Fuster, B. Silvi, *Comput. Chem.* **1999**, 23, 597.
- [15] Gaussian 03, Revision D.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, H. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W.

- Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Wallingford CT, **2004**.
- [16] a) C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785; b) A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648.
- [17] P. C. Hariharan, J. A. Pople, *Theor. Chim. Acta* **1973**, *28*, 213; T. Clark, J. Chandrasekhar, G. W. Spitznagel, P. von R. Schleyer, *J. Comput. Chem.* **1983**, *4*, 294.
- [18] a) T. H. Dunning, Jr., *J. Chem. Phys.* **1989**, *90*, 1007; b) R. A. Kendall, T. H. Dunning, Jr., R. J. Harrison, *J. Chem. Phys.* **1992**, *96*, 6796.
- [19] A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098; J. P. Perdew, Y. Wang, *Phys. Rev. B* **1992**, *45*, 13244.
- [20] Y. Zhao, D. G. Truhlar, *J. Phys. Chem. A* **2004**, *108*, 6908.
- [21] M. T. Cancès, B. Mennucci, J. Tomasi, *J. Chem. Phys.* **1997**, *107*, 3032; M. Cossi, G. Scalmani, N. Rega, V. Barone, *J. Chem. Phys.* **2002**, *117*, 43.
- [22] S. F. Boys, F. Bernardi, *Mol. Phys.* **1970**, *19*, 553.
- [23] a) X. Krokidis, S. Noury, B. Silvi, *J. Phys. Chem. A* **1997**, *101*, 7277; b) J. Pilme, B. Silvi, E. A. Alikhani, *J. Phys. Chem. A* **2005**, *109*, 10028; c) I. Fourré, B. Silvi, P. Chaquin, A. Sevin, *J. Comput. Chem.* **1999**, *20*, 897.
- [24] B. Silvi, *J. Phys. Chem. A* **2003**, *107*, 3081.
- [25] R. F. W. Bader, *Atoms in Molecules: A Quantum Theory*, Oxford University Press, Oxford, **1990**.
- [26] R. J. Gillespie, R. S. Nyholm, *Q. Rev. Chem. Soc.* **1957**, *11*, 339.
- [27] K. Anderson, M. P. Fluscher, G. Karlström, R. Lindh, P. A. Malqvist, J. Olsen, B. Roos, A. J. Sadlej, M. R. A. Blomberg, P. E. M. Siegbahn, V. Kello, J. Noga, M. Urban, P. O. Widmark, *MOLCAS, Version 5.4, University of Lund, Lund* (Sweden).
- [28] N. Ben Amor, D. Maynau, *Chem. Phys. Lett.* **1998**, *286*, 211.
- [29] Z. Barandiaran, L. Seijo, *Can. J. Chem.* **1992**, *70*, 409.
- [30] W. J. Hehre, R. F. Stewart, J. A. Pople, *J. Chem. Phys.* **1969**, *51*, 2657.
- [31] P. Flukiner, H. P. Luthi, S. Portman, J. Weber, *Molekel*, Swiss Center for Scientific Computing, Switzerland, **2000–2002**.
- [32] a) S. Chalmet, M. F. Ruiz-Lopez, *J. Chem. Phys.* **1999**, *111*, 1117; b) T. Ishida, P. J. Rossky, E. W. Castner, Jr., *J. Phys. Chem. B* **2004**, *108*, 17585; c) A. Botti, F. Bruni, S. Imberti, M. A. Ricci, A. K. Soper, *J. Mol. Liquids* **2005**, *117*, 81; d) J. Pilmé, P. Fleurat-Lessard, unpublished results.
- [33] C. J. Calzado, J. Cabrero, J.-P. Malrieu, R. Caballol, *J. Chem. Phys.* **2002**, *116*, 2728; C. J. Calzado, J. Cabrero, J.-P. Malrieu, R. Caballol, *J. Chem. Phys.* **2002**, *116*, 3985.
- [34] J. Cabrero, C. J. Calzado, D. Maynau, R. Caballol, J.-P. Malrieu, *J. Phys. Chem. A* **2002**, *106*, 8146.
- [35] By using the correlated wave function on the basis of localized orbitals, one can split the wave function into covalent and ionic (i.e., charge transfers $N^{\cdot-}C^+O^{\cdot-}, N^+C^{\cdot-}O$) contributions.

Received: August 31, 2006

Revised: December 20, 2006

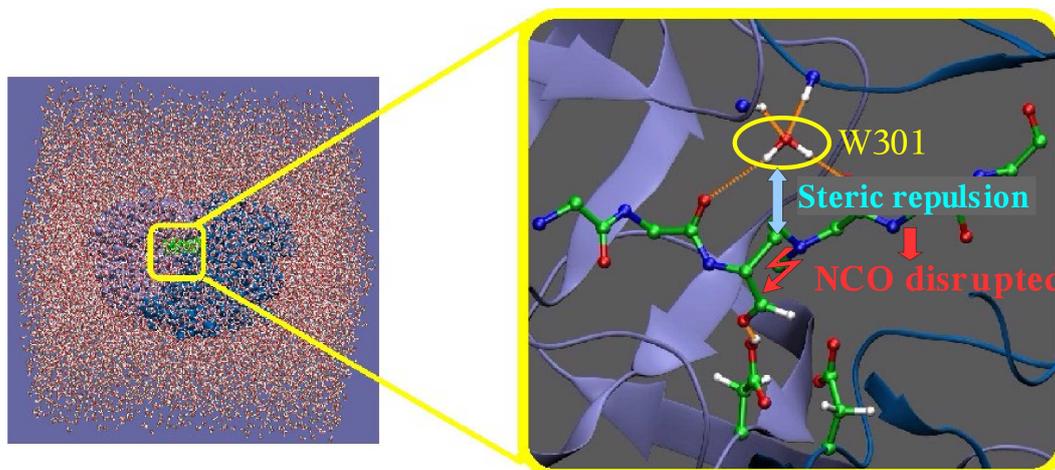
Published online: March 27, 2007

6) The one with biochemistry

The second step of this biochemical study aims at elucidating the origin of the low inhibiting power of the first molecules synthesized by J. Hasserodt et al.¹ For this, we use molecular dynamics simulations of the solvated inhibitor-enzyme complex. Due to the complexity of the N \cdots CO bond, an accurate quantum level of theory is necessary for this part of the system. Thus we use an hybrid Car-Parrinello/molecular mechanics (QM/MM) scheme:² the active site of the enzyme (formed by the side chains of the Aspartic acids 25 and 25') and the N \cdots CO moiety is described using the BLYP functional, while the rest of the system (ca. 38000 atoms) is described with the AMBER ff03 force field.

These calculations show that the N \cdots CO interaction is not stable in the active site of the protein. In particular, it turns out that there is a competition between the formation of this interaction and the conservation of an hydrogen bonds network that is known to be crucial for the binding of both substrates and inhibitors. During our simulations, these H-bonds are first partly disrupted and the characteristic C-N distance of the N \cdots CO moiety exhibits a drastic lengthening at the same time. Then, the hydrogen bonds network is restored and the inhibitors is finally stabilized in a conformation in which the N \cdots CO bond is disrupted.

However, our simulations indicate that the interfering part of the inhibitor is not the N \cdots CO core but the aliphatic bridge linking the amino part to the carbonyl group. Therefore, we suggest that alternative candidates that displace the W301 water molecule should be more efficient.



Solvated HIV-1 PR/Inhibitor complex

Zoom on the steric interaction

¹Gautier, A.; Pitrat D.; Hasserodt, J. *Bioorg. Med. Chem.* **2006**, *14*, 3835–3847.

²Laio, A.; VandeVondele, J.; Rothlisberger, U. *J. Chem. Phys.* **2002**, *116*, 6941-6947

Low Inhibiting Power of N···CO Based Peptidomimetic Compounds against HIV-1 Protease: Insights from a QM/MM Study

Julian Garrec,^{†,||} Michele Cascella,[‡] Ursula Rothlisberger,[§] and Paul Fleurat-Lessard^{*,†}

Université de Lyon, École Normale Supérieure de Lyon, Laboratoire de Chimie – UMR 5182, 46 allée d'Italie, 69364 Lyon Cedex 07, France, Department of Chemistry and Biochemistry, University of Bern, Freiestrasse 3, CH-3012 Bern, Switzerland, and Laboratory of Computational Chemistry and Biochemistry, Institute of Chemical Sciences and Engineering, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

Received September 8, 2009

Abstract: Recently, Hasserodt et al. proposed new HIV-1 drug candidates based on a weak N···CO interaction, designed to be a close transition state analog (Gautier et al. *Bioorg. Med. Chem.* **2006**, *14*, 3835–3847; Waibel et al. *J. Bioorg. Med. Chem.* **2009**, *17*, 3671–3679). They suggested that further improvement of these compounds could take advantage of computational approaches. In the present work, we propose an atomistic model based on a QM/MM description of the N···CO core embedded in an amino-aldehyde peptidic inhibitor. We focus on the existence of the N···CO interaction in the aqueous and enzymatic media. We show that the N···CO bond holds in water, while in the protein, there is a competition between the formation of the weak N···CO bond and the conservation of the hydrogen bond network around the structural water molecule W301 that is known to be crucial for the binding of both substrates and inhibitors. This competition hampers the inhibitor to provide strong stabilizing interactions with all the key parts of the protein at the same time. Our calculations indicate that this competition we observed in peptidic compounds might be avoided by the proper design of nonpeptidic ones, following a similar strategy to that for cyclic urea derivatives and the FDA approved drug Tipranavir. Hence, our results encourage further development of the nonpeptidic hydrazino-urea derivatives suggested recently by Hasserodt et al.

1. Introduction

The human immunodeficiency virus type 1 aspartic protease (HIV-1 PR, Figure 1) is one of the major targets for the

design of anti-AIDS drugs.^{1,2} This enzyme catalyzes the hydrolysis at specific sites of the polyprotein encoded by the virus genome yielding separate functional proteins.^{3–5} This function was shown early to be crucial to virion assembly and maturation, and its disruption by either active-site mutation or inhibition leads to the production of viral particles that lack infectious ability.^{6–8}

Several HIV-1 PR inhibitors have been approved by the FDA and significantly prolong the life expectancy of HIV infected patients.^{1,9–11} Nevertheless, the rapid emergence of resistance caused by multiple HIV-1 PR mutations decreases the effectiveness of these drugs.^{9,12,13} Almost all FDA-approved drugs are peptidomimetic active-site inhibitors that contain a hydroxyl group designed to interact with the central

* To whom correspondence should be addressed. Phone: +33 4 7272 8154, Fax: +33 4 7272 8860. E-mail: Paul.Fleurat-Lessard@ens-lyon.fr.

[†] École Normale Supérieure de Lyon.

[‡] University of Bern.

[§] École Polytechnique Fédérale de Lausanne.

^{||} Present address: Laboratory of Computational Chemistry and Biochemistry, Institute of Chemical Sciences and Engineering, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland.

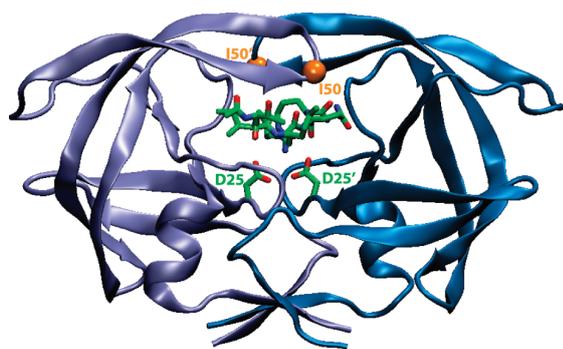


Figure 1. Crystallographic structure of HIV-1 PR in complex with the peptidomimetic inhibitor MVT-101^{36,34} (4HVP entry in the PDB data bank⁴⁴). The *aspartyl dyad* (residues 25 and 25') is located in the lower part of the active site. Ile 50 and 50' are located at the flap tips and are indicated with orange spheres. Drawings were made with the VMD program.⁸³

aspartyl dyad of HIV-1 PR (see the Abbreviations section for a list of abbreviations used in this work).^{1,2}

Freire et al. suggested that the competitive advantage of HIV-1 PR inhibitors over substrate binding is probably due to their higher rigidity, providing a more favorable entropic change upon binding. However, because of their rigidity, these inhibitors are less amenable to adapt to shape modifications of the enzymatic binding site induced by mutations.¹⁴ Hence, the design of new potent inhibitors that exhibit both a better binding free energy and an increased flexibility remains a challenging task.^{13,15–17}

In that context, finding functional groups that yield stronger interactions with the aspartyl dyad than the usual hydroxyl group would be of great interest.¹⁸ Toward this aim, Hasserodt et al.^{19–21} proposed a new concept of aspartic protease inhibitors based on a noncovalent interaction of a tertiary amine nitrogen with a carbonyl group, the so-called $N\cdots CO$ bond, that they believed to be a better transition state analog than the commonly used hydroxyl moiety (Figure 2). As a first attempt, they synthesized a series of dipeptide mimics containing the $N\cdots CO$ core,¹⁹ the amine and aldehyde fragments being bridged by an ethylene moiety (referred to as the *aliphatic bridge* hereafter, Figure 2a).

The best candidate of these amino-aldehyde peptides (AAP) exhibited an inhibition constant of 97 μM , far from the typical picomolar value that is desirable for a potent inhibitor.^{15,17,22–25}

In order to design more efficient inhibitors, they synthesized new compounds based on hydrazyno-urea heterocycles containing both the $N\cdots CO$ core and a carbonyl group aimed at forming hydrogen bonds with two NH groups of the upper part of the HIV-1 PR active site,^{20,21} the so-called *flaps*. This strategy is similar to that involved in the design of cyclic urea derivatives²⁶ and the FDA-approved drug Tipranavir.²⁷ The hydrazyno-urea derivatives indeed proved to interact slightly stronger with HIV-1 PR ($K_i \approx 29 \mu\text{M}$ for the best candidate). Hasserodt et al. concluded that a hit to lead optimization could now be initiated with the help of computational methods.^{20,21} Such an approach, however, needs the definition of a proper model describing the $N\cdots CO$ core in a biological environment. The purpose of the present

work is to design this model and to apply it to AAPs in order to help elucidate the origin of their low inhibition power.

Using a mixed quantum mechanics/molecular mechanics (QM/MM)²⁸ approach, we have designed a model that explicitly includes the solvated enzyme complexed with an AAP. An accurate quantum level of theory is crucial to the correct description of the $N\cdots CO$ core interacting with the aspartyl dyad, due to the intrinsic complexity of this system. No accurate transferable force field parameters exist for the $N\cdots CO$ bond,^{29,30} in particular because the stability of the $N\cdots CO$ is highly sensitive to the nature of the surrounding medium.^{19,31–33} For instance, the $N\cdots CO$ bond is unstable in apolar-aprotic media, leaving the tertiary amine and the aldehyde groups essentially independent. The structure of a complex between HIV-1 PR and an AAP has not been reported yet. Our study aims at exploring the feasibility of $N\cdots CO$ bond formation in this enzyme. In order to compare the behavior of the AAP, in particular the $N\cdots CO$ bond, in the protein and in aqueous media, we have also designed a similar QM/MM model of the AAP in water.

This article is organized as follows: in the first part, we detail our structural model and the computational procedure used to model the AAP in the enzymatic and aqueous media. Results are reported and discussed in the second part, while the last part summarizes our main findings.

2. Materials and Methods

2.1. Initial Structure. No experimental structure of an AAP complexed with HIV-1 PR ($E\cdot AAP$) is available. Nonetheless, on the basis of known structural features of HIV-1 PR–substrate complexes and previous kinetics/inhibition experiments of Hasserodt et al., it is possible to construct a starting structure for molecular dynamics simulations. Indeed, inhibition profiles show a competitive mechanism strongly suggesting that the AAP truly binds to the active site of the enzyme.¹⁹ In addition, HIV-1 PR is known to bind a variety of peptide substrates in the same extended conformation,^{34,35} the substrate backbone exhibiting many hydrogen bonds with the enzyme and the side chains being accommodated in a series of binding site subpockets. Due to the very high peptidic character of AAPs, their backbone and side chains should bind the active site in a way very similar to that of the corresponding peptides. Since AAPs are designed to be transition state analogs,¹⁹ the reaction intermediate ($E\cdot INT$) that connects the two TSs along the reaction pathway of HIV-1 PR appears as a natural starting structure for $E\cdot AAP$ modeling.

From both ¹⁸O isotope exchange experiments³⁶ and X-ray structures that captured key intermediate stages of the catalytic reaction,^{37,38} there is evidence that the substrate peptide bond cleavage involves the nucleophilic attack of a water molecule onto the scissile peptide bond, leading to a tetrahedral intermediate (Figure 2c). Note that, despite this commonly accepted picture, the protonation state of the intermediate is still a matter of controversy. Indeed, while *ab initio* calculations suggested that the intermediate is a neutral gem-diol,^{39,40} an empirical valence bond (EVB) model—calibrated against DFT calculations involving model

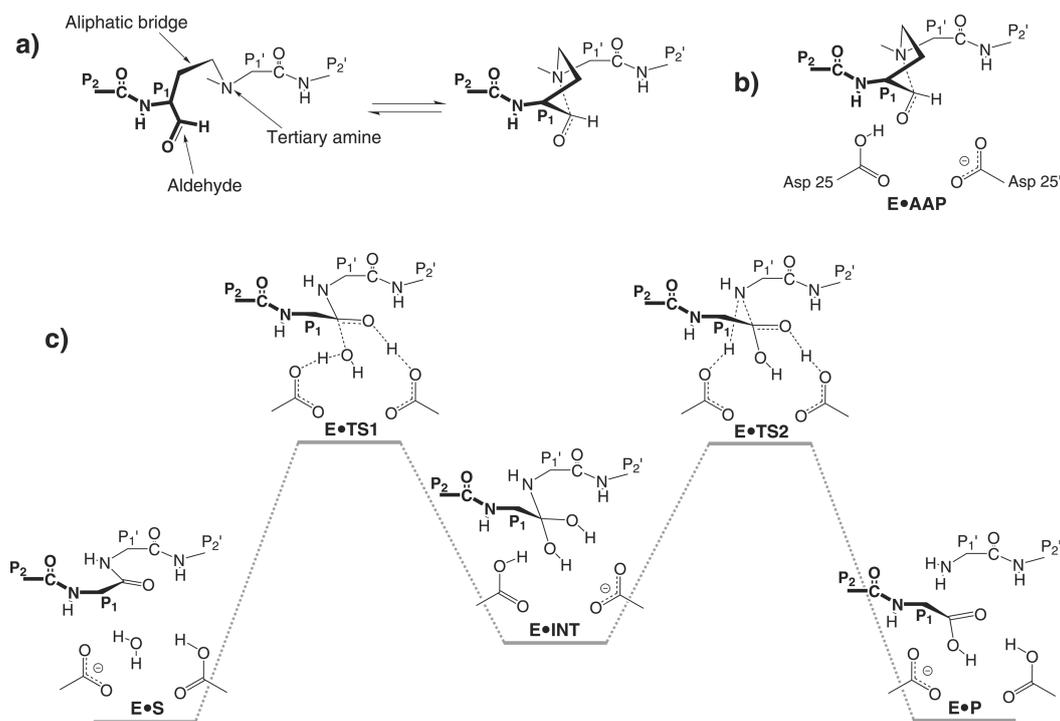


Figure 2. N \cdots CO bond as a transition state mimic. (a) N \cdots CO bond formation within an amino-aldehyde peptide (AAP). Side chains of the AAP are not represented for the sake of simplicity. Instead, their positions are indicated using the notation of Schechter and Berger (P₂, P₁, P₁', P₂').⁸⁴ The N \cdots CO core is introduced at the P₁–P₁' junction, the proximity of the amine and aldehyde fragments being ensured by an *aliphatic bridge*. (b) HIV-1 PR aspartyl dyad complexed with an AAP. (c) Catalytic mechanism of HIV-1 PR involving a tetrahedral intermediate (E•INT), which is represented here as a gem-diol. Note that some authors have reported that this intermediate could be an oxyanion (see text).^{41,42}

compounds in the gas phase—provided evidence for a charged oxyanion.⁴¹ Other authors calibrated their EVB Hamiltonian without including the gem-diol in their resonance structure set.⁴² Since we have chosen a computational setup involving an *ab initio* approach similar to that of ref 39, we have considered a neutral gem-diol in the present study. However, we stress that this choice should not affect our results dramatically, since the tetrahedral intermediate is just chosen as a starting structure to perform a first equilibration of the system. More specifically, we chose the complex between HIV-1 PR (E) and the Thr–Ile–Met–Met–Gln–Arg peptide substrate⁴³ in its hydrated form (INT; Figure 2c). E•INT was constructed from the X-ray structure of HIV-1 PR complexed with the highly peptidic MVT-101 inhibitor^{34,36} (4HVP entry in the PDB data bank⁴⁴), a compound that exhibits the sequence N-acetyl-Thr–Ile–Nle-Ψ[CH₂NH]–Nle–Gln–Arg–amide (Nle = norleucine). The Asp dyad was assumed to be monoprotonated according to the commonly accepted mechanism of HIV-1 PR.^{36,39,41,42,45} The protein was immersed in a 90 × 71 × 74 Å³ water box, and the entire system was neutralized by adding six chloride counterions. The whole system was composed of about 38 000 atoms.

2.2. Equilibration of E•INT. **2.2.1. Classical MD Simulations.** The complex was first equilibrated at the classical level using the AMBER9 suite of programs.⁴⁶ The parameters for the solute, apart from the amide hydrate (–C(OH)₂–NH–) moiety of the gem-diol intermediate, were taken from the AMBER 03 force field,⁴⁷ and the TIP3P

model was used to describe water molecules.⁴⁸ Bonded and van der Waals parameters of the amide hydrate were extracted from the generalized AMBER force field (GAFF).⁴⁹ The charges of the amide hydrate were obtained using a standard RESP procedure.⁵⁰ The electrostatic potential was computed at the HF/6-31G(d) level with the Gaussian 03 package,⁵¹ from a model compound including the hydrated Met–Met sequence capped with acetyl and N-methyl groups, i.e. Ace–Met–[C(OH)₂–NH]–Met–Nme. Note that these additional parameters are used for a small part only of the entire system, which is then described within the QM part during further equilibration at the QM/MM level (see next section).

Long-range electrostatic interactions were computed using the Ewald particle mesh method.^{52,53} A cutoff of 10 Å was applied for the van der Waals interactions and the real part of the electrostatic interactions. A time step of 1.5 fs was used, and all bonds containing hydrogen were constrained using the SHAKE algorithm. Constant temperature was achieved using Langevin dynamics⁵⁴ with a collision frequency of 5 ps^{–1}, while the pressure was maintained using a Berendsen's barostat⁵⁵ with a relaxation time of 1.0 ps. The system was first heated to 150 K over 15 ps and then to 300 K over a further 15 ps. Then, an equilibration of 1 ns at 1 atm and 300 K was carried out.

2.2.2. QM/MM MD Simulations. Starting from the equilibrated E•INT structure obtained at the classical level, we switched to a hybrid quantum mechanics/molecular mechanics (QM/MM)²⁸ description for the system. We used the

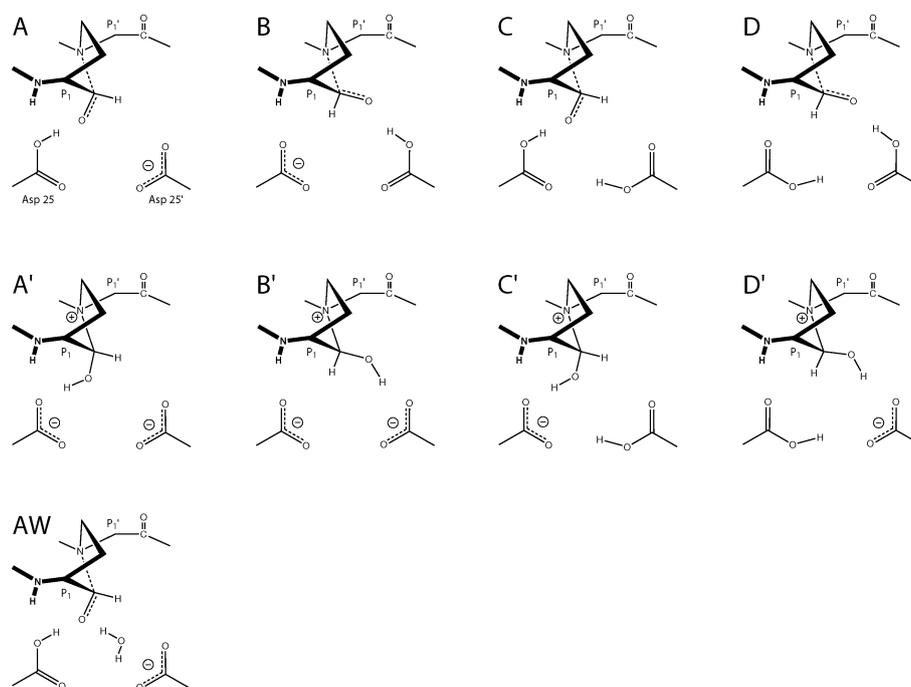


Figure 3. E•AAP isomers and protonation states considered in this study.

approach developed by Rothlisberger and co-workers.^{56–58} The QM region encompassed the Asp25(25') side chains and the amide hydrate moiety and was described at the DFT/BLYP^{59,60} level of theory. Dangling bonds were saturated with hydrogen atoms. The Kohn–Sham orbitals were expanded in plane waves with a cutoff of 70 Ry and a quantum cell of $17.2 \times 14.8 \times 14.8 \text{ \AA}^3$. A fictitious electron mass of 600 au and a time step of 5 au ($\approx 0.12 \text{ fs}$) were used.

The E•INT system was first minimized using a simulated annealing-like procedure: starting from a temperature of 50 K, atomic velocities were rescaled at each time step by a factor of 0.99. Then, the system was heated up to 300 K over 15 ps. Subsequently, 12 ps of NVT simulation were performed, using a Nosé–Hoover chain thermostat^{61–63} of 900 cm^{-1} frequency.

2.3. Transformation of E•INT into E•AAP. A structure obtained after 10.3 ps of simulation was chosen to perform the E•INT \rightarrow E•AAP transformation. The hydrated substrate was modified at the Met–Met junction to incorporate the N•••CO core and the aliphatic bridge (Figure 2b). The starting value of the characteristic C–N distance of the N•••CO moiety was chosen to be 1.7 \AA , which is a typical value for a N•••CO bond in a polar-protic medium.³³ At this stage, we should stress that the only groups that differ from the initial crystallographic structure are the N•••CO core and the aliphatic bridge. Once the peptide backbone of the AAP is accommodated in the active site, there are few possibilities left for the positioning of the cycle made of the aliphatic bridge and the N•••CO group. The main uncertainty lies in the configuration of the carbon atom of the N•••CO moiety. Thus we decided to study both possible configurations (Figure 3A,B).

Once the starting position of heavy atoms was chosen, we had to address the question of the active site protonation

state within the E•AAP complex. Indeed, HIV-1 PR exhibits a wide range of protonation patterns according to the presence and the nature of the ligand.^{39,45,64–71} The prediction of such a pattern can be done, for instance, by fitting kinetic data recorded at different pH to rate equations,^{45,64} by computational pK_a estimation from an experimental structure,^{65,66} by NMR titration,^{67,68} by constructing computational models that aim at reproducing experimental data such as crystallographic structures^{70,71} or NMR chemical shifts,⁶⁹ or by combining X-ray and neutron crystallography.⁷²

The *a priori* determination of the E•AAP active-site protonation pattern is challenging. We decided to apply a systematic strategy, in which a series of protonation states for each configuration of the N•••CO carbon atom was considered. When bound to neutral ligands, the aspartyl dyad of HIV-1 PR is usually monoprotonated^{39,45,71,73} or diprotonated,⁶⁷ while positively charged ligands may yield an unprotonated dyad.^{45,65}

Following this, we generated a series of nine isomers in different protonation states that are depicted in Figure 3.

First, we considered complexes A,B (monoprotonated dyad with neutral AAP) and C,D (diprotonated dyad with neutral AAP). In addition, the N•••CO interaction can also be described by the limiting $N^+ - C - O^-$ form. The negatively charged oxygen atom can be seen as an alcoholate and can thus be very basic. We have thus considered the possible proton transfer from the neighboring carboxyl group to the N•••CO core leading to a positively charged AAP containing a $N^+ - CO - H$ moiety. A',B' (unprotonated dyad with positively charged AAP) and C',D' (monoprotonated dyad with positively charged AAP) were obtained from A,B and C,D, respectively, by shifting the closest proton of the Asp dyad to the N•••CO oxygen atom.

To our knowledge, the catalytic water molecule that is tightly H-bonded to the aspartyl dyad in the HIV-1 PR-substrate complex (Figure 2c) is systematically displaced by any active site inhibitor bound to the enzyme. Indeed, it has only been observed in crystallographic structures of the free enzyme,^{74,75} but never in any X-ray or NMR structure of the bound enzyme. However, we found it interesting to check if an AAP can displace this water molecule. Thus, we considered a last complex, denoted by AW, in which a water molecule is added close to the aspartyl dyad.

2.4. QM/MM Modeling of E•AAP Isomers in Different Protonation States. Once the E•INT \rightarrow E•AAP transformation was performed, each E•AAP structure underwent a mild annealing-like protocol allowing a slow relaxation and minimization of the newly introduced aliphatic bridge–N \cdots CO moiety and the surrounding protein medium. This enabled us to scrutinize (i) the stability of the N \cdots CO bond within the protein and (ii) the interactions between the AAP and the protein.

We used the same level of calculation as the one we used to equilibrate E•INT. The QM region included the Asp25(25') side chains, the N \cdots CO moiety, and the aliphatic bridge. We stress that the ability of density functional theory to describe the N \cdots CO interaction was demonstrated previously.³³ The quantum cell size was adapted to that of the QM region, leading to a $16.9 \times 14.3 \times 16.9 \text{ \AA}^3$ box.

Each E•AAP complex underwent a first minimization, using the same simulated annealing-like procedure that we used previously to optimize the geometry of the E•INT complex, i.e., using a starting temperature of 50 K and a scaling factor of 0.99. These calculations were stopped as soon as the temperature reached a value below 3 K. Then, the system was let free to relax during 8 ps of NVE QM/MM molecular dynamics. During this run, each complex heated up to about 40 K due to remaining bad contacts. Finally, the minimization of the system was achieved using a second annealing with a scaling factor of 0.999.

To check the validity of this relaxation/optimization protocol, we performed further calculations using constraints. These calculations are described in the Supporting Information.

2.5. QM/MM Modeling of AAP in Water. The simulations in water were started from the same starting structure of the AAP as the one used in complex A. The inhibitor was immersed in a $49 \times 49 \times 49 \text{ \AA}^3$ water box equilibrated at room temperature at the classical level. All of the atoms of the AAP were fixed during the first stages of the simulation. For the QM/MM runs, we used the same description as the one for the E•AAPs. The quantum cell size was adapted to that of the QM region, leading to a $11.0 \times 11.3 \times 10.4 \text{ \AA}^3$ box.

The water box was further equilibrated using the following protocol: A first annealing was performed using a starting temperature of 50 K and a velocity scaling factor of 0.99. Then, the solvent underwent 4.7 ps of MD simulation at 300 K using the Berendsen weak coupling algorithm.⁵⁵ Finally, the target temperature was decreased linearly from 300 to 1 K in 3 ps.

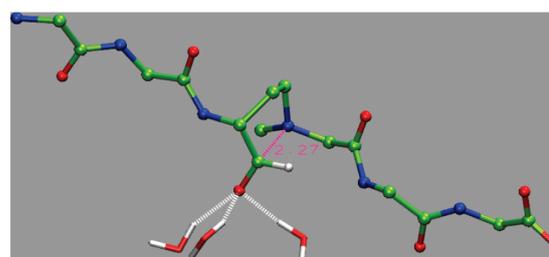


Figure 4. Optimized geometries of the AAP in water at the QM/MM level of theory. Only the inhibitor backbone and water molecules hydrogen bonded to the oxygen atom of the N \cdots CO core are represented. The N \cdots CO interaction is depicted in magenta.

At this stage, the constraints on the inhibitor were removed, and we minimized the QM and the MM part of the system consecutively. Then, the system underwent 4 ps of NVT simulation using a Nosé–Hoover chain thermostat.^{61–63} Finally, a minimized geometry was obtained by performing an annealing with a velocity scaling factor of 0.999.

3. Results and Discussion

3.1. AAP in Water. We start the analysis of our results by describing the behavior of the N \cdots CO bond embedded in an AAP in an aqueous medium. During our 4 ps of NVT simulation at 300 K, a weak N \cdots CO interaction was maintained, with an average value of the C–N distance of 2.41 Å and a standard deviation of 0.18 Å. After minimization using our annealing-like protocol, this distance decreased to 2.27 Å. Figure 4 shows the corresponding optimized geometry. The nitrogen lone pair of the N \cdots CO core is directed toward the aldehyde, which has lost its coplanarity. Three water molecules are hydrogen bonded to the aldehyde, which stabilizes the N \cdots CO bond.

The fact that the N \cdots CO bond is maintained over the course of the simulation shows that our computational approach is able to reproduce the closed configuration reported by Hasserodt et al.¹⁹ On the basis of NMR measurements in methanol, they estimated that 70% of the AAPs exhibit a N \cdots CO bond. We expect that much longer simulations would provide opening and closing events.

Nevertheless, the average C–N distance we observe is longer than the typical value of 1.8 Å of a N \cdots CO bond in a polar-protic medium.³³ This is in agreement with the computational study of Pilmé et al., who have shown that water molecules H-bonded to the N \cdots CO core stabilize the N⁺–C–O[–] form by accepting part of its electronic density.³³ Even though the charge transfer is small (ca. 0.06 e/molecule), it was shown to be sufficient to stabilize short CN distances. In our simulations, the interaction between the N \cdots CO core and the surrounding water molecules is described through the QM/MM interface, which does not account for charge transfer effects, hence leading to a longer C–N distance. Note, however, that this problem does not hold for our simulations in HIV-1 PR (next sections), because the polar protein group close to the N \cdots CO core was included in the QM part.

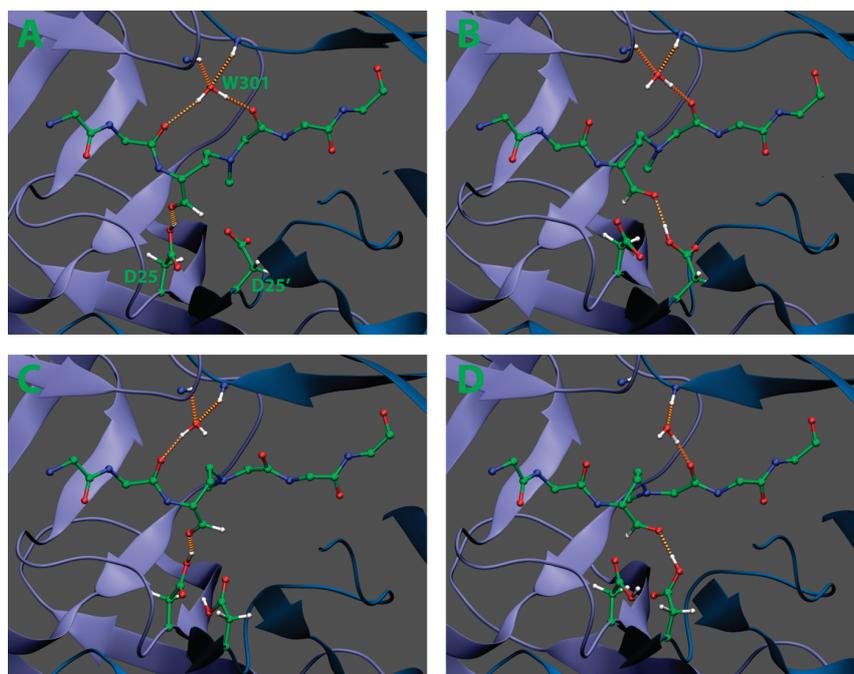


Figure 5. Optimized geometries of complexes A–D at the QM/MM level of theory. Only polar hydrogens belonging to the H-bond networks around W301, the N \cdots CO core, and the aspartyl dyad are represented.

3.2. Protonation State of the N \cdots CO Oxygen Atom in the Protein. The first question we addressed concerning the behavior of AAPs within HIV-1 PR was whether the protein is able to provide a stabilizing medium for a protonated, positively charged AAP. Indeed, due to the enhanced ionic character of the C–O bond within the N \cdots CO core, the basicity of the oxygen is increased, and thus, a sufficiently acidic medium could stabilize a covalent N $^+$ –CO–H moiety.

In the early stages of the minimization of isomers A'–D', we observed that the proton linked to the N \cdots CO oxygen was systematically shifted to the closest aspartate, leading to isomers A–D, respectively. Thus, despite the difficulty of providing an a priori accurate estimation of the protonation pattern in the active site of our system, our calculations consistently converge to an AAP N \cdots CO core that is unprotonated within HIV-1 PR. Therefore, in the following, only isomers A–D and AW will be considered.

3.3. N \cdots CO Bond Stability in the Active Site and Interaction with the Asp Dyad. Optimized geometries of complexes A–D are depicted in Figure 5. Selected structural parameters are reported in Table 1. For each complex, the characteristic C–N distance of the N \cdots CO moiety exhibits a drastic lengthening from the starting value of 1.7 Å. Note that, since this opening event was already observed during geometry optimization, we did not attempt to run molecular dynamics simulations at room temperature, as we did for the modeling of AAPs in water. Instead, we applied a mild relaxation/minimization protocol in order to assess the effect of the enzymatic media on the geometry of AAPs.

Monoprotonated complexes A and B are stabilized in a fully opened conformation, with $d_{\text{CN}} = 3.75$ and 3.43 Å in A and B, respectively. This distance is larger than the

Table 1. Main Geometrical Parameters Resulting from the Minimization of Complexes A–D^a

	A	B	C	D
d_{CN}^b	3.75	3.43	3.01	2.51
NCO \cdots H $\delta_{\text{Asp25(25)'}}$ ^c	1.66	1.76	1.60	1.61
NH _{Ile50} \cdots O _{W301} ^d	2.20	2.39	2.34	3.42
NH _{Ile50} \cdots O _{W301} ^d	2.26	2.01	2.08	1.87
CO _{P2} \cdots H _{W301} ^d	2.30	3.67	1.77	4.28
CO _{P1} \cdots H _{W301} ^d	1.75	1.67	3.83	1.64
Ω^e	94.25	69.44	37.77	-43.12
Φ^f	-125.77	-84.81	-74.34	-51.06

^a Distances are given in Å and angles in degrees. ^b Distance between the tertiary amine nitrogen and the aldehyde. ^c Hydrogen bond between the N \cdots CO oxygen and the H δ of the aspartyl dyad. Residue 25 or 25' is considered, depending on the protonation pattern of the aspartyl dyad. ^d Hydrogen bond network around the structural water molecule W301. ^e Dihedral angle between each Asp oxygen of the aspartyl dyad. ^f Dihedral angle of the aliphatic bridge.

characteristic value of even a weak N \cdots CO interaction in which $d_{\text{CN}} \approx 2.8$ Å.³³ In both structures, the N \cdots CO nitrogen lone pair is no longer directed toward the planar aldehyde group, indicating that the N \cdots CO interaction is completely disrupted. Furthermore, the aspartyl dyad coplanarity is lost according to the values of the dihedral angle between each Asp oxygen, i.e., $\Omega = 94.2$ and 69.4° for A and B, respectively. This structural feature plays a crucial role in the binding of both substrate and inhibitors.^{4,76,77} The highly distorted conformation of the aspartyl dyad is a clear indicator of an unfavorable interaction between the N \cdots CO core, Asp25 and Asp25'. Despite the hydrogen bond between the proton of Asp25(25') and the carbonyl oxygen of the AAP in isomer A(B) (see Table 1), the AAP does not provide sufficient shielding to stabilize the strong electrostatic Asp–Asp repulsion.⁷⁰

The AW complex (results not shown) follows a similar route, leading to a drastic lengthening of the C–N distance and the loss of the aspartyl dyad coplanarity. Moreover, the opening of the N \cdots CO core is accompanied by the departure of the catalytic water molecule out of the aspartyl dyad, yielding a structure close to complex A. In fact, we observe that this water molecule is stabilized by a hydrogen bond with the Gly27 carbonyl group. Despite the short time scale that is accessible at the QM/MM level, we expect that this is a transient state, prior to a move of the catalytic water molecule to the bulk. We have already observed this behavior in a classical simulation of a HIV-1 PR-substrate complex (not presented here), which is consistent with observations reported by others.⁴⁰

Diprotonated complex D exhibits the smallest C–N distance ($d_{\text{CN}} = 2.51 \text{ \AA}$), which corresponds to a weak N \cdots CO interaction.³³ The nitrogen lone pair is still directed toward the carbonyl group which is less planar and more tightly H-bound to Asp25' than in the monoprotonated complexes, according to the $\text{NCO}\cdots\text{H}\delta_{\text{Asp25}'}$ distance. The Asp dyad remains almost coplanar, suggesting that the complex is much more stable than the monoprotonated ones. This is in agreement with the fact that the shielding introduced by the additional proton in diprotonated states makes the aspartyl dyad more amenable to accepting the accumulated negative charge of the N \cdots CO oxygen.

Complex C represents an intermediate situation between complexes D and A,B. As in D, it is characterized by a coplanar aspartyl dyad and a tight hydrogen bond between the oxygen of the N \cdots CO core and Asp25. However, similarly to A and B, the C–N distance equals 3.01 \AA , which corresponds at best to a shallow N \cdots CO bond.

Our systematic approach reflects the geometric behavior of the N \cdots CO core within the enzyme, over a set of different conditions related to the protonation state and the starting configuration of the N \cdots CO carbon atom. As we will show in the next section, our model enables one to establish a correlation between the disruption of the N \cdots CO bond and other interactions that play a key role in the affinity between the AAP and HIV-1 PR.

3.4. Origin of the N \cdots CO Opening. A detailed analysis of the hydrogen bond network inside the HIV-1 PR active site sheds some light on the origin of the instability of the N \cdots CO bond within the enzyme. In HIV-1 PR-substrate and in most HIV-1 PR-inhibitor complexes,⁷⁸ a tetrahedrally coordinated structural water molecule, commonly labeled W301,^{1,79,80} bridges Ile50(50') NH groups belonging to the upper part of the active site cleft (the so-called *flaps*, Figure 1) and P₂ and P₁' CO groups of the substrate/inhibitor. Hence, this hydrogen bond network plays a crucial role in the correct positioning of both substrate and peptidomimetic inhibitors in the active site. Inhibitors that do not exhibit these interactions are those that have been designed to displace W301, such as cyclic urea derivatives²⁶ or the FDA-approved drug Tipranavir.²⁷

Figure 6a depicts the starting geometry (just after the E•INT→E•AAP transformation) of W301 and its surrounding atoms within complex A. Note that the position of heavy atoms is the same as that of complex C and AW and is very

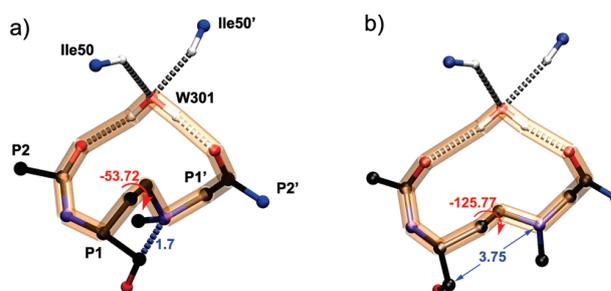


Figure 6. Competition between the N \cdots CO bond formation and the conservation of the H-bond network around the structural water molecule W301. Starting (a) and final (b) geometries of W301 and its surrounding atoms within complex A. Side chains and hydrogen atoms not belonging to the H-bond network around W301 are not represented for the sake of simplicity. The “macrocycle” composed of W301, P₂, and P₁' CO groups and the aliphatic bridge is represented with orange transparent tubes. The H bonds around W301 are depicted with gray-dashed tubes, while the N \cdots CO interaction is represented with a blue-dashed tube in the starting structure (a).

similar to that of complexes B and D, the only difference being the configuration of the N \cdots CO carbon atom. The tetrahedral H-bond network around W301 is present, prior to any step of our minimization protocol. In the starting configuration of our complexes, the “macrocycle” composed of W301, P₂, and P₁' CO groups and the aliphatic bridge (represented with orange transparent tubes in Figure 6) is in a quite compact conformation, while the C–N distance of the N \cdots CO moiety is 1.7 \AA .

During the NVE molecular dynamics run of complex A, the H bonds around W301 are first partly disrupted, and as the C–N distance lengthens, the hydrogen bond network is progressively restored. Figure 6b represents the final (optimized) structure of complex A. The “macrocycle” is stabilized in an extended conformation, in which the aliphatic bridge has been pushed away from W301. This structural reorganization occurs together with a drastic increase of both the characteristic dihedral angle Φ of the aliphatic bridge (from -51 to -126°) and the C–N distance (from 1.7 to 3.75 \AA). Hence, the disruption mechanism of the N \cdots CO bond observed in complex A may be formulated as follows: The hydrogen bond network that W301 tends to form with two backbone carbonyl groups of the AAP tightens the “macrocycle”, which in turn reduces the steric hindrance by extending the aliphatic bridge, thus lengthening the C–N distance of the N \cdots CO core.

Complex B exhibits a behavior similar to that of complex A, sharing the same location of W301 and the evolution of the structure toward an extended “macrocycle”. The major difference lies in the value of the dihedral angle of the aliphatic bridge (Table 1), which is lower in B, i.e., $\Phi = -84.8^\circ$. Thus, the aliphatic bridge remains closer to the center of the “macrocycle”, and W301 cannot establish an optimal hydrogen bond with the P₂ carbonyl.

In order to analyze the link between the hydrogen bond network around W301 and the CN distance, we have

conducted a constrained optimization. Starting from the optimized geometry of complex B, we changed the hydrogen bond network around the W301 molecule from that in complex B to that in complex A. The CN bond was not frozen, and we found that it increased to reach a final length of 3.70 Å, close to the value observed in complex A.

The situation is rather different for diprotonated complexes. In the optimized complexes C and D, W301 is no longer tetrahedrally coordinated, as depicted in Figure 5C,D. The highest disruption occurs for complex D, which conserves only two hydrogen bonds. On the other hand, the aliphatic bridge remains in a conformation closer to that of the starting structure, as indicated by the Φ values: -74.3 and -51.1° for C and D, respectively. This is consistent with the C–N distance values discussed in the previous section. As previously noted, a diprotonated dyad is a better hydrogen bond donor than a monoprotonated one and is thus more favorable to the N \cdots CO interaction, which in turn, is more competitive against the conservation of the H-bond network around W301.

The behavior observed for complexes A–D can be summarized stating that a N \cdots CO bond and a proper H-bond network around W301 cannot be realized both at the same time. Clearly, the introduction of the N \cdots CO core at the scissile peptide bond location induces a systematic competition between the N \cdots CO bond formation and the interaction network involving the structural water molecule W301. Barillari et al.⁸¹ estimated the binding free energy of W301 to a series of complexes between HIV-1 PR and peptidomimetic inhibitors using the double-decoupling free energy simulation method. They found that the binding free energy ranged from -7 to -10 kcal mol $^{-1}$. If one removes the entropy contribution, in which the upper bound has been estimated to be about 2 kcal mol $^{-1}$,⁸² one obtains a binding energy of $-(9-12)$ kcal mol $^{-1}$. This is comparable to the energy of the N \cdots CO bond, which has been estimated to be -11 kcal mol $^{-1}$ at the CCSD(T) level.³³ These energetic considerations support the competition that we observed between the conservation of the hydrogen bond network around W301 and N \cdots CO bond formation.

Since the N \cdots CO moiety has been designed to interact strongly with the aspartyl dyad,¹⁹ our simulations show that an AAP cannot provide stabilizing interactions with all the key parts of the HIV-1 PR binding site at the same time. The low inhibition power of AAPs might originate, at least partly, from this competition. Interestingly, our results suggest that further development of N \cdots CO containing inhibitors should focus on nonpeptidic compounds that could displace W301. This in line with the hydrazino-urea compounds proposed recently by Hasserodt et al.^{20,21} These derivatives contain both a N \cdots CO bond and a hydrazino-urea group designed to interact directly with the flaps of HIV-1 PR, similar to cyclic urea derivatives²⁶ and the FDA-approved drug Tipranavir.²⁷

Note that hydrazino-urea compounds synthesized by Hasserodt et al. bind only slightly stronger to HIV-1 PR than AAPs ($K_i \approx 29$ μ M and 97 μ M, respectively). However, the former contain only three groups aimed at filling the subpockets of the enzyme binding cleft, while cyclic urea

derivatives usually have four.²⁶ Hence, we encourage the development of optimized N \cdots CO-containing hydrazino-urea inhibitors that would contain proper groups aimed at interacting with the same binding subpockets as cyclic urea derivatives. Such a design could be assisted by a computational study based on the same approach we developed in the present work.

4. Conclusions

We have proposed here a computational procedure to tackle the difficult theoretical description of HIV-1 PR inhibitors based on the unusual N \cdots CO bond. This procedure consists of an explicitly solvated model of the ligand-enzyme complex (E•AAP) described at the atomistic level with a QM/MM approach. The N \cdots CO core and part of the enzyme active site are described using an accurate QM level, while the rest of the protein and the solvent are described using the classical AMBER force field.

In this work, we have applied this model to investigate the origin of the low inhibiting power of the recently proposed amino-aldehyde peptide (AAP) compounds against HIV-1 PR. Our calculations provide detailed information on the feasibility of the N \cdots CO bond formation within the enzymatic environment along with crucial interactions that govern the stability of the complex.

Considering all the possible protonation patterns of the active site aspartyl dyad, we have shown that N \cdots CO bond formation/dissociation takes place in a competitive mechanism, in which the structural water molecule W301 tends to establish a hydrogen bond network that indirectly penalizes the shortening of the distance between the nitrogen atom and the CO group of the N \cdots CO core. We conclude that the reported poor inhibition power of AAPs¹⁹ originates, at least partly, from this competition.

Despite this, a N \cdots CO \cdots H $\delta_{\text{Asp25(25)}}$ hydrogen bond was observed for each protonation state. In the case of complex D, this H bond is tighter, and a weak N \cdots CO interaction is formed at the same time. This means that, under appropriate conditions, a N \cdots CO core could interact strongly with the aspartyl dyad of HIV-1 PR. Hence, the design of N \cdots CO-containing candidates that could displace the water molecule W301 would be an interesting alternative to AAPs. This supports the idea that a hydrazino-urea core^{20,21} is an interesting template for the design of potent anti-AIDS drugs. In particular, it would be interesting to extend the recent work of Hasserodt et al. to hydrazino-urea derivatives containing peripheral groups aimed at filling the HIV-1 PR subpockets P₂, P₁, P₁', P₂' properly, similar to cyclic urea derivatives. Such a design could be done *in silico*, using the approach we have developed in the present study.

Acknowledgment. We thank the “Centre Informatique National de l’Enseignement Supérieur” and the “Institut du développement et des ressources en informatique Scientifique” for computational support and resources. We also thank Dr. M. Waibel and Pr. J. Hasserodt for useful

discussions. The financial support from the “Cluster de recherche Chimie de la Région Rhone-Alpes” is duly acknowledged.

Abbreviations

HIV-1, Human Immunodeficiency Virus Type 1; AIDS, Acquired ImmunoDeficiency Syndrome; PR, Protease; E, catalytically active dimeric form of HIV-1 protease; INT, Thr-Ile-Met-Met-Gln-Arg peptide substrate in its hydrated form; QM/MM, Quantum Mechanics/Molecular Mechanics; FDA, Food and Drug Administration; AAP, Amino-Aldehyde Peptide; CPMD, Car-Parrinello Molecular Dynamics; DFT, Density Functional Theory.

Supporting Information Available: The results of the additional simulations with various constrained distances (d_{CN} , $\text{NH}_{\text{Ile50}}\cdots\text{O}_{\text{W301}}$, $\text{NH}_{\text{Ile50}}\cdots\text{O}_{\text{W301}}$, $\text{CO}_{\text{P2}}\cdots\text{H}_{\text{W301}}$, $\text{CO}_{\text{P1}}\cdots\text{H}_{\text{W301}}$) are described. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- Wlodawer, A.; Vondrasek, J. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 249–284.
- Wlodawer, A.; Erickson, J. W. *Annu. Rev. Biochem.* **1993**, *62*, 543–585.
- Debouck, C.; Gorniak, J. G.; Strickler, J. E.; Meek, T. D.; Metcalf, B. W.; Rosenberg, M. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 8903–8906.
- Fitzgerald, P. M. D.; Springer, J. P. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 299–320.
- Poorman, R.; Tomasselli, A.; Heinrikson, R.; Kezdy, F. *J. Biol. Chem.* **1991**, *266*, 14554–14561.
- Kohl, N. E.; Emini, E. A.; Schleif, W. A.; Davis, L. J.; Heimbach, J. C.; Dixon, R. A. F.; Scolnick, E. M.; Sigal, I. S. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 4686–4690.
- Seelmeier, S.; Schmidt, H.; Turk, V.; von der Helm, K. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 6612–6616.
- McQuade, T. J.; Tomasselli, A. G.; Liu, L.; Karacostas, V.; Moss, B.; Sawyer, T. K.; Heinrikson, R. L.; Tarpley, W. G. *Science* **1990**, *247*, 454–456.
- Richman, D. D. *Nature* **2001**, *410*, 995–1001.
- Hammer, S. M.; Squires, K. E.; Hughes, M. D.; Grimes, J. M.; Demeter, L. M.; Currier, J. S.; Eron, J. J.; Feinberg, J. E.; Balfour, H. H.; Deyton, L. R.; Chodakewitz, J. A.; Fischl, M. A.; Phair, J. P.; Pedneault, L.; Nguyen, B.-Y.; Cook, J. C. *N. Engl. J. Med.* **1997**, *337*, 725–733.
- Gulick, R. M.; Mellors, J. W.; Havlir, D.; Eron, J. J.; Gonzalez, C.; McMahon, D.; Richman, D. D.; Valentine, F. T.; Jonas, L.; Meibohm, A.; Emini, E. A.; Chodakewitz, J. A.; Deutsch, P.; Holder, D.; Schleif, W. A.; Condra, J. H. *N. Engl. J. Med.* **1997**, *337*, 734–739.
- Condra, J. H.; Schleif, W. A.; Blahy, O. M.; Gabryelski, L. J.; Graham, D. J.; Quintero, J.; Rhodes, A.; Robbins, H. L.; Roth, E.; Shivaprakash, M.; Titus, D.; Yang, T.; Teplert, H.; Squires, K. E.; Deutsch, P. J.; Emini, E. A. *Nature* **1995**, *374*, 569–571.
- Martinez-Cajas, J. L.; Wainberg, M. A. *Antiviral Res.* **2007**, *76*, 203–221.
- Luque, I.; Todd, M. J.; Gómez, J.; Semo, N.; Freire, E. *Biochemistry* **1998**, *37*, 5791–5797.
- Velázquez-Campoy, A.; Kiso, Y.; Freire, E. *Arch. Biochem. Biophys.* **2001**, *390*, 169–175.
- Velázquez-Campoy, A.; Luque, I.; Freire, E. *Thermochim. Acta* **2001**, *380*, 217–227.
- Vega, S.; Kang, L.-W.; Velázquez-Campoy, A.; Kiso, Y.; Amzel, M.; Freire, E. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 594–602.
- Leung, D.; Abbenante, G.; Fairlie, D. P. *J. Med. Chem.* **2000**, *43*, 305–341.
- Gautier, A.; Pitrat, D.; Hasserodt, J. *Bioorg. Med. Chem.* **2006**, *14*, 3835–3847. Kinetics and inhibition measurements are given in their Supplementary Data.
- Waibel, M.; Hasserodt, J. *J. Org. Chem.* **2008**, *73*, 6119–6126.
- Waibel, M.; Pitrat, D.; Hasserodt, J. *Bioorg. Med. Chem.* **2009**, *17*, 3671–3679.
- Shuman, C. F.; Hämäläinen, M. D.; Danielson, U. H. *J. Mol. Recognit.* **2004**, *17*, 106–119.
- Velázquez-Campoy, H. O. A.; Xie, D.; Freire, E. *Protein Sci.* **2002**, *11*, 1908–1916.
- Markgren, P.-O.; Schaal, W.; Hämäläinen, M.; Karlén, A.; Hallberg, A.; Samuelsson, B.; Danielson, U. H. *J. Med. Chem.* **2002**, *45*, 5430–5439.
- Velázquez-Campoy, A.; Freire, E. *J. Cell. Biochem.* **2001**, *37*, 82–88.
- Lam, P. Y. S.; Ru, Y.; Jadhav, P. K.; Aldrich, P. E.; DeLucca, G. V.; Eyermann, C. J.; Chang, C.-H.; Emmett, G.; Holler, E. R.; Daneker, W. F.; Li, L.; Confalone, P. N.; McHugh, R. J.; Han, Q.; Li, R.; Markwalder, J. A.; Seitz, S. P.; Sharpe, T. R.; Bachelier, L. T.; Rayner, M. M.; Klabe, R. M.; Shum, L.; Winslow, D. L.; Kornhauser, D. M.; Jackson, D. A.; Erickson-Viitanen, S.; Hodge, C. N. *J. Med. Chem.* **1996**, *39*, 3514–3525.
- Doyon, L.; Tremblay, S.; Bourgon, L.; Wardrop, E.; Cordingley, M. G. *Antiviral Res.* **2005**, *68*, 27–35.
- Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227–249.
- Spanka, G.; Boese, R.; Rademacher, P. *J. Org. Chem.* **1987**, *52*, 3362–3367.
- Griffith, R.; Bremner, J. B.; Titmuss, S. J. *J. Comput. Chem.* **1997**, *18*, 1211–1221.
- Leonard, N. J.; Oki, M. *J. Am. Chem. Soc.* **1954**, *76*, 3463–3465.
- Kirby, A. J.; Komarov, I. V.; Bilenko, V. A.; Davies, J. E.; Rawson, J. M. *Chem. Commun.* **2002**, 2106.
- Pilmé, J.; Berthoumieux, H.; Robert, V.; Fleurat-Lessard, P. *Chem.—Eur. J.* **2007**, *13*, 5388–5393.
- Miller, M.; Schneiderand, J.; Sathyanarayana, B.; Tothand, M.; Marshalland, G.; Clawsonand, L.; Selkand, L.; Kent, S.; Wlodawer, A. *Science* **1989**, *246*, 1149–1152.
- Prabu-Jeyabalan, M.; Nalivaika, E.; Schiffer, C. A. *Structure* **2002**, *10*, 369–381.
- Hyland, L. J.; Tomaszek, T. A., Jr.; Roberts, G. D.; Carr, S. A.; Magaard, V. W.; Bryan, H. L.; Michael, S. A. F.; Moore, L.; Minnich, M. D.; Culp, J. S.; DesJarlais, R. L.; Meek, T. D. *Biochemistry* **1991**, *30*, 8441–8453.

- (37) Kovalevsky, A. Y.; Chumanovich, A. A.; Liu, F.; Louis, J. M.; Weber, I. T. *Biochemistry* **2007**, *46*, 14854–14864.
- (38) Kumar, M.; Prashar, V.; Mahale, S.; Hosur, M. V. *Biochem. J.* **2005**, *389*, 365–371.
- (39) Piana, S.; Bucher, D.; Carloni, P.; Rothlisberger, U. *J. Phys. Chem. B* **2004**, *108*, 11139–11149.
- (40) Carnevale, V.; Raugeri, S.; Piana, S.; Carloni, P. *Comput. Phys. Commun.* **2008**, *179*, 120–123.
- (41) Trylska, J.; Grochowski, P.; McCammon, J. A. *Protein Sci.* **2004**, *13*, 513–528.
- (42) Bjelic, S.; Aqvist, J. *Biochemistry* **2006**, *45*, 7709–7723.
- (43) Ratner, L.; Haseltine, W.; Patarca, R.; Livak, K. J.; Starcich, B.; Josephs, S. F.; Doran, E. R.; Rafalski, J. A.; Whitehorn, E. A.; Baumeister, K.; Ivanoff, L.; Petteway, S. R., Jr.; Pearson, M. L.; Lautenberger, J. A.; Papas, T. S.; Ghraryeb-parallel, J.; Changparallel, N. T.; Gallo, R. C.; Wong-Staal, F. *Nature* **1985**, *313*, 277–284.
- (44) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (45) Hyland, L. J.; Tomaszek, T. A., Jr.; Meek, T. D. *Biochemistry* **1991**, *30*, 8454–8463.
- (46) Case, D.; Darden, T.; Cheatham, T., III; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Merz, K.; Pearlman, D.; Crowley, M.; Walker, R.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D.; Schafmeister, C.; Ross, W.; Kollman, P. *AMBER 9*; University of California: San Francisco, 2006.
- (47) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
- (48) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (49) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (50) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269–10280.
- (51) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, Revision 02; Gaussian, Inc.: Wallingford, CT, 2004.
- (52) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (53) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (54) Izaguirre, J. A.; Catarella, D. P.; Wozniak, J. M.; Skeel, R. D. *J. Chem. Phys.* **2001**, *114*, 2090–2098.
- (55) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (56) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471–2474.
- (57) Laio, A.; VandeVondele, J.; Rothlisberger, U. *J. Chem. Phys.* **2002**, *116*, 6941–6947.
- (58) Laio, A.; Gervasio, F. L.; VandeVondele, J.; Sulpizi, M.; Rothlisberger, U. *J. Phys. Chem. B* **2004**, *108*, 7963–7968.
- (59) Becke, A. D. *Phys. Rev. A* **1998**, *38*, 3098–3100.
- (60) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
- (61) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511–519.
- (62) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.
- (63) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635–2643.
- (64) Ido, E.; ping Han, H.; Kezdy, F. J.; Tang, J. *J. Biol. Chem.* **1991**, *266*, 24359–24366.
- (65) Czodrowski, P.; Sotriffer, C. A.; Klebe, G. *J. Chem. Inf. Model.* **2007**, *47*, 1590–1598.
- (66) Trylska, J.; Antosiewicz, J.; Geller, M.; Hodge, C.; Klabe, R.; Head, M.; Gilson, M. *Protein Sci.* **1999**, *8*, 180–195.
- (67) Yamazaki, T.; Nicholson, L. K.; Wingfield, D. A. T. P.; Stahl, S. J.; Kaufman, J. D.; Eyermann, C. J.; Hodge, C. N.; Lam, P. Y. S.; Ru, Y.; Jadhav, P. K.; Chang, C.-H.; Webers, P. C. *J. Am. Chem. Soc.* **1994**, *116*, 10791–10792.
- (68) Wang, Y.-X.; Freedberg, D. I.; Yamazaki, T.; Wingfield, P. T.; Stahl, S. J.; Kaufman, J. D.; Kiso, Y.; Torchia, D. A. *Biochemistry* **1996**, *35*, 9945–9950.
- (69) Piana, S.; Sebastiani, D.; Carloni, P.; Parrinello, M. *J. Am. Chem. Soc.* **2001**, *123*, 8730–8737.
- (70) Piana, S.; Carloni, P. *Proteins* **2000**, *39*, 26–36.
- (71) Harte, W. E., Jr.; Beveridge, D. L. *J. Am. Chem. Soc.* **1993**, *115*, 3883–3886.
- (72) Adachi, M.; Ohhara, T.; Kurihara, K.; Tamada, T.; Honjo, E.; Okazaki, N.; Arai, S.; Shoyama, Y.; Kimura, K.; Matsumura, H.; Sugiyama, S.; Adachi, H.; Takano, K.; Mori, Y.; Hidaka, K.; Kimura, T.; Hayashi, Y.; Kiso, Y.; Kuroki, R. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 4641–4646.
- (73) Tawa, G. J.; Topol, I. A.; Burt, S. K.; Erickson, J. W. *J. Am. Chem. Soc.* **1998**, *120*, 8856–8863.
- (74) Pillai, B.; Kannan, K.; Hosur, M. V. *Proteins* **2001**, *43*, 57–64.
- (75) Kumar, M.; Hosur, M. V. *Eur. J. Biochem.* **2003**, *270*, 1231–1239.
- (76) Davies, D. R. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 189–215.
- (77) Todd, M. J.; Semo, N.; Freire, E. *J. Mol. Biol.* **1998**, *283*, 475–488.
- (78) Vondrasek, J.; Wlodawer, A. *Proteins* **2002**, *49*, 429–431.
- (79) Swain, A. L.; Miller, M. M.; Green, J.; Rich, D. H.; Schneider, J.; Kent, S. B. H.; Wlodawer, A. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8805–8809.

N•••CO Based Peptidomimetic Compounds

J. Chem. Theory Comput., Vol. 6, No. 4, 2010 **1379**

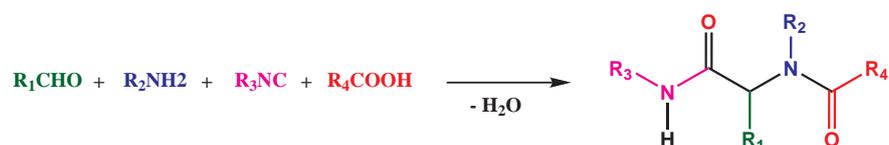
- (80) Jaskolski, M.; Tomasselli, A. G.; Sawyer, T. K.; Staples, D. G.; Heinrikson, R. L.; Schneider, J.; Kent, S. B. H.; Wlodawer, A. *Biochemistry* **1991**, *30*, 1600–1609.
- (81) Barillari, C.; Taylor, J.; Viner, R.; Essex, J. W. *J. Am. Chem. Soc.* **2007**, *129*, 2577–2587.
- (82) Dunitz, J. D. *Science* **1994**, *264*, 670–671.
- (83) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (84) Schechter, I.; Berger, A. *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157–162.

CT9004728

7) The one with organic reactions

In drug design, reactions that lead to variety of products with increased molecular diversity and complexity are the target of many studies. Amongst those, multicomponent reactions (MCRs)[1] constitute a very promising path as they represent efficient and general ways of producing families of products. During a multicomponent reaction, three reactants or more are put together simultaneously and lead quantitatively to one product.

The most interesting MCR are probably those including isocyanides,[2, 3, 4] embodied by the four component Ugi reaction[4, 5, 6, 7]. Discovered in 1958, it allows the formation of peptides derivatives by the coupling of isocyanides with carbonyl compounds, primary amines and carboxylic acids:



These reactions are quantitative due to the final irreversible acyl transfer coined as Mumm rearrangement.[8] Rossen *et al.* have recently proposed to use the Ugi reaction as an alternative path for an intermediate along the synthesis of the Merck HIV protease inhibitor Crixivan[®] (Indinavir[™]).[9]

Changing one partner in a known multicomponent coupling is an interesting strategy for the discovery of new reactions and reactivity with increasing molecular scaffolds. When considering carboxylic acids, their replacement represents a challenging problem because of the central role they play in the many steps of the reaction mechanism (Figure 7.1).

Indeed, their Bronst ed properties allow faster imine formation and addition of the moderately nucleophilic isocyanide to the activated iminium. The carboxylate is then involved in the trapping of the nitrilium species, and finally the structure of the acid allows the Mumm rearrangement to settle and displace all the equilibria (Figure 7.1). Therefore efficient replacement of the carboxylic acids might imply new reaction mechanism and rearrangement to occur.

El Ka im and Grimaud have recently reported the first use of Smiles rearrangement in Ugi type reactions.[10, 11] In these new multicomponent couplings, electron-deficient phenols, such as *ortho*- and *para*-nitro phenol, are used as acidic partners leading to N-aryl and O-aryl amides (Figure 7.2).

They have further demonstrated that these couplings could be extended to various heterocyclic derivatives.[12] This extension is very promising in terms of drug design as these heterocycles are key compounds in many biological systems with active compounds such as thymine or cytosine. Such compounds have been reported to be potent therapeutic agents for the treatment of inflammatory diseases (asthma, rheumatoid arthritis. . .),[13] HBV infection,[14] Creutzfeldt-Jacob disease,[15] epilepsy, cancer.[16, 17] Consequently such heterocyclic compounds remain major targets for the pharmaceutical industry.

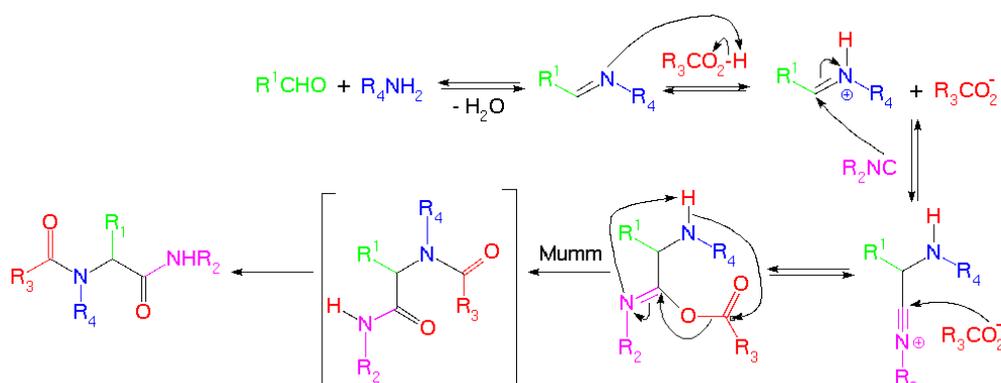


Figure 7.1: Proposed mechanism of the Ugi reaction.

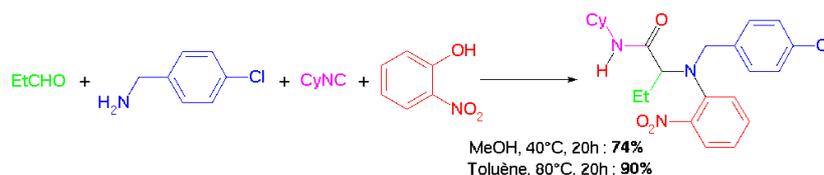


Figure 7.2: Ugi-Smiles coupling.

Project Motivation

One of the main drawback of the Ugi and Ugi-Smiles reactions is that none of them is enantioselective. Therefore, their application to drug design is always coupled to an additive asymmetric step. For example, Rossen *et al.* needed to use a chiral hydrogenation after the Ugi step.[9] Finding a chiral inductor or catalyst for these reactions would represent a major achievement not only for organic chemistry synthesis but also for the pharmaceutical industry. Until now, only Passerini-kind reactions (three components reaction) were made enantioselective. However, there are only few published possibilities (Denmark [18, 19], Schreiber [20], Wang [21]) and the proposed conditions are difficult to use. As stated by Ugi, "until today the problem of control of the newly formed stereocenter has not been solved conclusively".[4]

Thus, the long term goal of this project is to better understand the mechanism of the Ugi-Smiles coupling in order to find of a chiral variant of this reaction.

Many experimental results were gathered by El Kaïm and Grimaud during their ongoing studies on the chemistry of Ugi-like reactions. Some results still lack full rationalization:

- Role of the solvent: the usually proposed mechanism for the Ugi reaction involves ionic species (Figure 7.1). As the use of a nitrophenol appears as a simple variation of a reactant, one could expect that the mechanism of the Ugi-Smiles coupling should also involves ionic species. However, El Kaïm *et al.* found that the yield of the Ugi-Smiles reaction in toluene (an apolar medium) is greater than in methanol.
- Variations of the nitrophenol reactants: the proposed mechanism involves the protonation of an imine and then a phenolate attack on a nitrilium: the phenol's pKa and the nucleophilicity of the associated base should then be two parameters of great importance. However, the use of species more acid and with associated bases more nucleophilic (thiophenols against phenols) lead to very low yields.

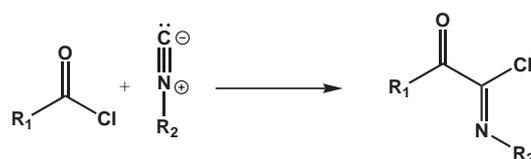
Our approach

In order to unravel this complex mechanism, we combine theory and experiments in a complementary way. The theoretical study use the large database of experimental results obtained by the ENSTA group to propose some new mechanisms for the Ugi-Smiles reaction.

Let us denote by **A** the Amino group, **B** the aldehyde, **C** the isoCyanide and **D** the phenol Derivative. Considering the Ugi-Smiles reaction, one can imagine six paths leading from **A+B+C+D** to the adduct **A-B-C-D**, formed before the Smile rearrangement:

- 1) $A+B+C+D \rightarrow A-B+C+D \rightarrow A-B-C+D \rightarrow A-B-C-D$
- 2) $A+B+C+D \rightarrow A-B+C+D \rightarrow A-B+C-D \rightarrow A-B-C-D$
- 3) $A+B+C+D \rightarrow A+B-C+D \rightarrow A-B-C+D \rightarrow A-B-C-D$
- 4) $A+B+C+D \rightarrow A+B-C+D \rightarrow A+B-C-D \rightarrow A-B-C-D$
- 5) $A+B+C+D \rightarrow A+B+C-D \rightarrow A-B+C-D \rightarrow A-B-C-D$
- 6) $A+B+C+D \rightarrow A+B+C-D \rightarrow A+B-C-D \rightarrow A-B-C-D$

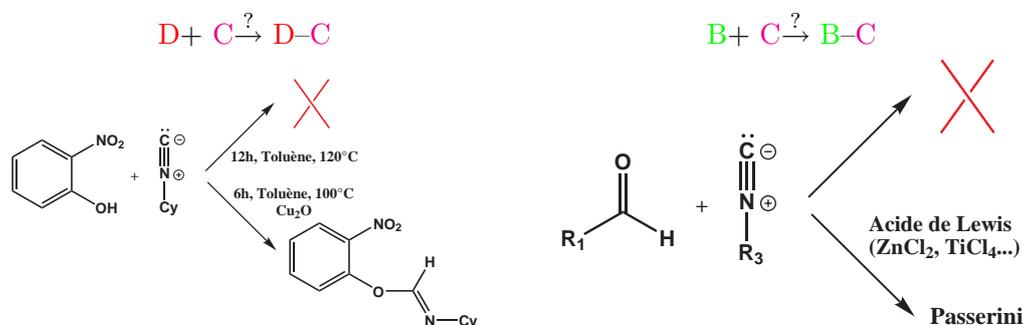
The first path corresponds to the mechanism proposed by Ugi.[5, 7] On top of these paths, one can also consider the insertion of a compound into a bond previously formed, as in the Nef insertion reaction:[22]



This leads to six additional paths:

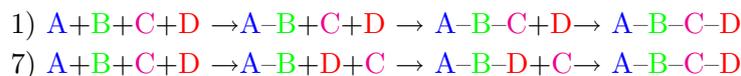
- 7) $A+B+C+D \rightarrow A-B+D+C \rightarrow A-B-D+C \rightarrow A-B-C-D$
- 8) $A+B+C+D \rightarrow A+B-D+C \rightarrow A-B-D+C \rightarrow A-B-C-D$
- 9) $A+B+C+D \rightarrow A+C-D+B \rightarrow A-C-D+B \rightarrow A-B-C-D$
- 10) $A+B+C+D \rightarrow A-C+D+B \rightarrow A-C-D+B \rightarrow A-B-C-D$
- 11) $A+B+C+D \rightarrow A-D+B+C \rightarrow A-B-D+C \rightarrow A-B-C-D$
- 12) $A+B+C+D \rightarrow A-D+B+C \rightarrow A-C-D+B \rightarrow A-B-C-D$

In order to check which paths are possible, we conducted some experiments corresponding to different steps along the retro-synthetic path. We first tried to add the isocyanide group onto the phenol or the aldehyde group:



As none of these steps work under the usual experimental conditions of the Ugi-Smiles reaction, we can eliminate paths 2, 3, 4, 5, 6 and 9 from our scheme. It is known that the aldehyde and the phenol derivative cannot react together to form the **B-D** compound: this

rules out path 8. Last, up to date, there is no example of the insertion of an aldehyde group into a chemical bond. Therefore, out of the 12 initial paths, only two remain, corresponding to the mechanism proposed by Ugi[5, 7] and to the insertion of isocyanide into the CO bond of the aminol $A-B-D$:



This is summarized in Figure 7.3.

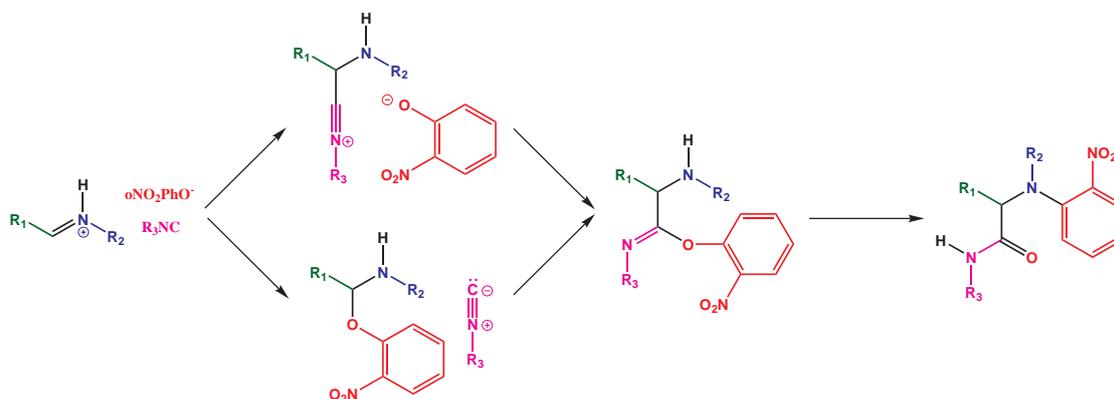
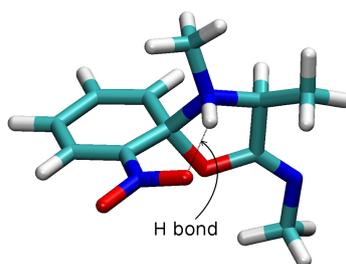


Figure 7.3: Two possible mechanisms for the Ugi-Smiles reaction

Our theoretical study thus focuses on these two paths, and more specifically on two steps: the Smiles rearrangement and the Nef insertion. Indeed, when considering the mechanism proposed by Ugi, all the reactions are equilibrated under usual conditions except the final rearrangement. Consequently, we have first focused our efforts on the Smiles rearrangement. Even though we have not yet unraveled the influence of all substituents, we have explained why heteroatoms are so important in the ortho position: they form a hydrogen bond with the hydrogen atom of the approaching amino group. This is depicted on Scheme 1. This H-bond stabilizes the spiro compound and thus facilitates the rearrangement.



Scheme 1

References

- [1] Zhu, J.; Bienaymé, H. *Multicomponent Reactions*; Wiley-VCH: .
- [2] Armstrong, R. W.; Combs, A. P.; Tempest, P. A.; Brown, S. D.; Keating, T. A. *Acc. Chem. Res.* **1996**, *29*, 123-131.

- [3] Dömling, A. *Chem. Rev.* **2006**, *106*, 17-89.
- [4] Dömling, A.; Ugi, I. *Angew. Chem. Int. Ed.* **2000**, *39*, 3168-3210.
- [5] Ugi, I.; Meyr, R. *Angew. Chem.* **1958**, *70*, 702.
- [6] Ugi, I.; Meyr, R. *Chem. Ber.* **1960**, *93*, 239.
- [7] Ugi, I.; Steinbrückner, C. *Angew. Chem.* **1960**, *72*, 267-268.
- [8] Mumm, O. *Ber. Dtsch. Chem. Ges.* **1910**, *43*, 886-893.
- [9] Rossen, K.; Pye, P.; DiMichelc, L.; Volante, R.; Reider, P. *Tetrahedron Lett.* **1998**, *39*, 6823-6826.
- [10] El Kaïm, L.; Grimaud, L.; Oble, J. *Angew. Chem. Int. Ed.* **2005**, *44*, 7961-7964.
- [11] El Kaïm, L.; Gizolme, M.; Grimaud, L. *Org. Lett.* **2006**, *8*, 5021.
- [12] El Kaïm, L.; Gizolme, M.; Grimaud, L.; Oble, J. *Org. Lett.* **2006**, *8*, 4019.
- [13] Doherty, G. A.; Kamenecka, T.; McCauley, E.; Van Riper, G.; Mumford, R. A.; Tong, S.; Hagmann, W. K. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 729-731.
- [14] Chen, H.; Zhang, W.; Tam, R.; Raney, A. K. *PCT Int Appl.* **2005**, *WO 2005058315 A1 20050630*.
- [15] Perrier, V.; Wallace, A. C.; Kanedo, K.; Safar, J.; Prusiner, S. B.; Cohen, F. E. *Proc. Natl. Acad. Sci.* **2000**, *97*, 6073-6078.
- [16] Fredholm, B. B.; Ijzerman, A. P.; Jacobson, K. A.; Klotz, K.-N.; Linden, J. *Pharmacol. Rev.* **2001**, *53*, 527-552.
- [17] Meijer, L.; Raymond, E. *Acc. Chem. Res.* **2003**, *36*, 417-425.
- [18] Denmark, S.; Fan, Y. *J. Am. Chem. Soc.* **2003**, *125*, 7825.
- [19] Denmark, S. E.; Fan, Y. *J. Org. Chem.* **2005**, *70*, 9667.
- [20] Andreana, P.; Liu, C.; Schreiber, S. *Org. Lett.* **2004**, *6*, 4231.
- [21] Wang, S.-X.; Wang, M.-X.; Wang, D.-X.; Zhu, J. *Org. Lett.* **2007**, *9*, 3615.
- [22] Nef, J. U. *Ueber das zweiwerthige Kohlenstoffatom* *Justus Liebig's Annalen der Chemie* **1892**, *270*, 267-335.

Part IV

The one with the future plans

In the coming years, I plan to keep my dual approach: combining development of new tools and their applications to study chemical reactivity.

1 Tools development

1.1 Describing the solvent

On the methodological side, on top of continuing my current projects, I also want to start working on the description of the solvent as a reactant. Indeed, the solvent is playing an important role in the applications that I have approached up to now. Describing all the solvent molecules at the QM level is neither possible nor interesting in molecular dynamics simulations of solvated systems. However, when using hybrid methods, one is confronted to the subset definition: where to put the solvent molecules? If one considers the solvation of a $N \cdots CO$ core in water, on the average, there are three water molecules around the carbonyl group. However, these molecules are moving during the simulation time and are replaced by some others preventing standard approach to describe them in QM or in MM. One way to deal with such a system is to have a dynamic definition of the QM and MM subsets. One usually defines a QM core that is then used to define the full QM subset: all molecules closer than a given distance from this core are described using the QM method. The first proposal goes back to the *Hot Spot* method of Rode et al. designed to study metallic ions solvation in 1996.[1] Many methods have been proposed since then, among which are the ONIOM-XS,[2] Time-Dependent Groups,[3] the Adaptive Partitioning method and its Sorted version[4], and the difference-based adaptive solvation approach.[5] By construction, the number of QM molecules is fluctuating along the course of the simulation and that one has to take care of the conservation of momentum and total energy. Out of all previous works, only the Adaptive Partitioning (AP) method of Heyden et al. is doing so. Let us note however, that standard QM/MM CPMD dynamics do not conserve energy exactly as one usually uses approximations to compute the electrostatic interactions between the QM and the MM part.[6] The Sorted Adaptive Partitioning (SAP) method and the difference-based adaptive solvation (DAS) of Bulo et al. ensure only continuity of the forces and are thus not adequate for microcanonical simulations. However, ensuring the forces continuity has been proved sufficient to sample the canonical ensemble.[4, 5, 7] Most of these methods are based on the introduction of a buffer region between the QM core and the MM environment. Then, all possible configurations are generated by considering each solvent molecule in the buffer region either QM or MM. The adaptive QM/MM energy is defined as a combination of the energies of these configurations. However, if there are N solvent molecules in the buffer region, this lead to 2^N calculations (as in the AP method) and this would not be usable for large system. One has thus to find some criterion to eliminate most of the configurations. The published methods usually differ in the way they eliminate configurations as well as in the function relating the adaptive energy to the configurations energy. At the end, they all lead to a computational cost that scales approximately linearly with N . In the SAP and in DAS, only configurations in which there is no MM solvent molecule closer to the QM core than a QM molecule are kept, leading to a computation cost scaling linearly. This idea is shown in Figure 1.

Sorted AP and DAS differ in the way they weight the remaining configurations. In collaboration with R. Bulo, we will first couple the DAS method with Gaussian to study the Ugi-Smile reaction. Then we will implement it in CPMD for our QM/MM biochemical studies.

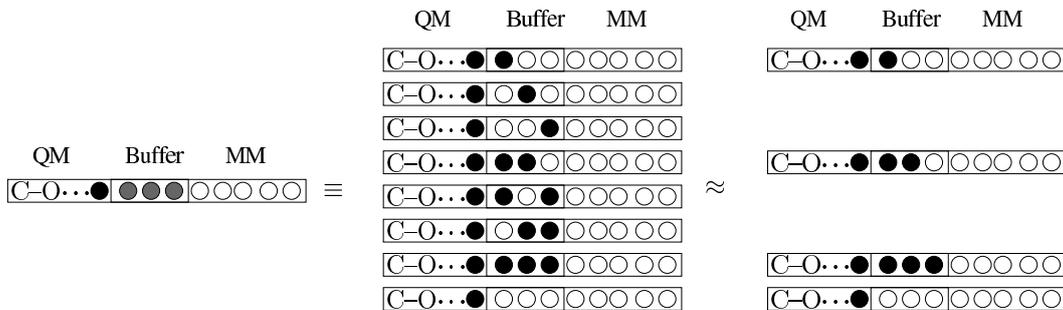


Figure 1: Schematic representation of the configurations selection for SAP and DAS. ● represent a solvent molecule that could be described either at the QM or at the MM level, ● represent QM solvent molecules, while ○ represent MM solvent molecules.

2 Applications

2.1 Biochemistry

The aim of our biochemical simulations is to help designing new *lead* for the HIV-1 PR inhibition. For this, an ideal tool would be *ab initio docking*: a fast way to predict accurately the binding affinity of HIV-1 PR with an inhibitor candidate. During the Ph. D of J. Garrec, we have used CPMD QM/MM simulations to study such interactions. However, these simulations remain much too time consuming to be used as a docking tool. Another approach is to use more approximate methods to estimate the binding free energy. One such method is the extension of the MM/PBSA method to QM/MM approach. The QM/MM PBSA method has been developed by Fisher et al.[8] as an extension of the MM/PBSA method of Kollman et al.[9] In this method, the binding free energy ΔG_{bind} between a protein P and a ligand L is defined as:

$$\Delta G_{bind}(P/L) = G_{aq}(P - L) - G_{aq}(P) - G_{aq}(L),$$

where $G_{aq}(X)$ is the free energy of the solvated X system.* In practice, $\Delta G_{aq}(X)$ is computed as:

$$G_{aq}(X) = \langle E_{gas}(X) \rangle - TS_{gas}(X) + \Delta G_{solvation}(X)$$

$E_{gas}(X)$ is the potential energy of the X system, computed using a QM/MM approach. The average is computed over a set of conformations generated using QM/MM Molecular Dynamics simulations. $S_{gas}(X)$ is the entropy of X and can be computed using normal mode analysis for example. Last, $\Delta G_{solvation}(X)$ is computed using the PBSA method: in these methods, the solvation free energy is decomposed into its polar and non polar contributions : $\Delta G_{solvation}(X) = \Delta G_p(X) + \Delta G_{np}(X)$. The polar term is computed using a numerical solution of the Poisson-Boltzmann equation, while the non polar term estimated using the solvent-accessible surface area (SASA) approximation: $\Delta G_{np}(X) = \gamma \times SASA + \beta$, where γ and β are parameters.[10]

Even though the approximations might seem gross, they usually lead to the correct trends. For example, H. Jamet et al. have recently used QM/MM PBSA to compute the inhibition constants of a family of CK2 inhibitors.[11] They could reproduce the experimental results and even propose new candidates.

*We assume here that we can define a unique free energy reference for all species, so that ΔG_{bind} has a physical meaning.

At the end of his Ph.D., J. Garrec has finished equilibrated the DMP-323 (Figure 2a) complexed by the HIV-1 PR. Our goal is now to modify this complex in order to incorporate the $N \cdots CO$ core in it to obtain a compound close to those synthesized by J. Hasserodt.[12, 13] We will then try different molecules by varying the substituents of positions 1 to 5 (Figure 2b).

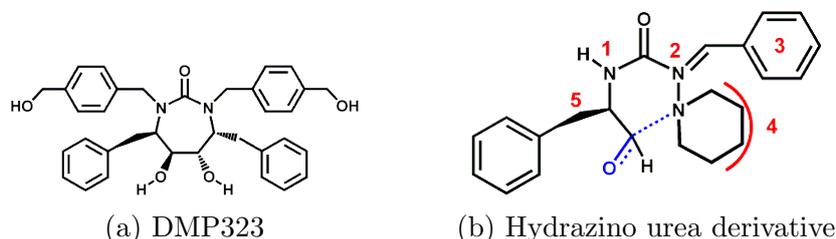
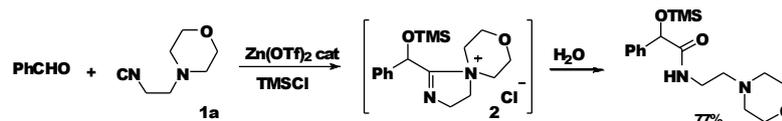


Figure 2: a) Cyclic urea (DMP323) b) Cyclic urea derivative based on the $N \cdots CO$ core.

Once the first hydrazino urea derivative would have been equilibrated using standard QM/MM simulations, we will use QM/MM PBSA methods to estimate the binding affinities of different derivatives. This will be done in collaboration with H. Jamet.

2.2 Organic reactions with isocyanides

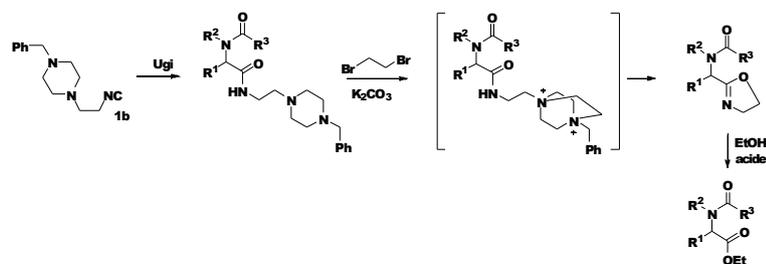
The Ugi reactions (and their derivatives) are the main use of isocyanides in organic synthesis. One of the key step in these reactions is the nucleophilic addition of the isocyanide to an activated imine group. If one wants to use the isocyanides in different conditions, it is important to better characterize their nucleophilicity with respect to their substituents. Some experimental studies have been conducted,[14] and we would like to focus on the donor substituent in β or γ position. Indeed Ganem et al. have shown that the coupling yield greatly depends on the isocyanide used: they are good with isocyanomorpholin (**1a**) but quite small with the standard Ugi reactant cyclohexylisocyanide (Scheme 2).[15] Ganem proposed that this peculiar behavior is due to the formation a spiro intermediate **2**.



Scheme 2

In this project, that will be conducted in collaboration with L. Grimaud and L. El Kaïm (Laboratoire de Chimie et Procédés, ENSTA), we will analyze the peculiar reactivity of these isocyanides substituted with donor heteroatoms. The first step will be to use topological analysis of the ELF function, as well as Fukui descriptor to obtain some quantitative information about the nucleophilicity of these new isocyanides. We will then extend this work to the reactivity of isocyanide **1b** (Scheme 3) and couple the Ugi reaction with 1,2-dibromoethane. This might open a new route to oxazole derivatives. These species are important synthesis intermediates as they can easily be converted into unsaturated esters.

This amide to ester conversion directly links to an important topic in isocyanide based multicomponent reactions. Indeed, many groups have already tried to find “universal” isocyanides that could easily be modified after the Ugi reaction.



Scheme 3

References

- [1] Kerdcharoen, T.; Liedl, K. R.; Rode, B. M. *Chemical Physics* **1996**, *211*, 313 - 323.
- [2] Kerdcharoen, T.; Morokuma, K. *Chemical Physics Letters* **2002**, *355*, 257 - 262.
- [3] Salazar, M. R. *The Journal of Physical Chemistry A* **2005**, *109*, 11515–11520.
- [4] Heyden, A.; Lin, H.; Truhlar, D. G. *The Journal of Physical Chemistry B* **2007**, *111*, 2231–2241.
- [5] Bulo, R. E.; Ensing, B.; Sikkema, J.; Visscher, L. *Journal of Chemical Theory and Computation* **2009**, *5*, 2212–2221.
- [6] Laio, A.; VandeVondele, J.; Rothlisberger, U. *J. Chem. Phys.* **2002**, *116*, 6941–6947.
- [7] Guthrie, M. G.; Daigle, A. D.; Salazar, M. R. *Journal of Chemical Theory and Computation* **2010**, *6*, 18–25.
- [8] Gräter, F.; Schwarzl, S. M.; Dejaegere, A.; Fischer, S.; Smith, J. C. *The Journal of Physical Chemistry B* **2005**, *109*, 10474–10483.
- [9] Massova, I.; Kollman, P. A. *Journal of the American Chemical Society* **1999**, *121*, 8133–8143.
- [10] Sitkoff, D.; Sharp, K. A.; Honig, B. *The Journal of Physical Chemistry* **1994**, *98*, 1978–1988.
- [11] Retegan, M.; Milet, A.; Jamet, H. *Journal of Chemical Information and Modeling* **2009**, *49*, 963–971.
- [12] Waibel, M.; Hasserodt, J. *J. Org. Chem.* **2008**, *73*, 6119–6126.
- [13] Waibel, M.; Pitrat, D.; Hasserodt, J. *Bioorg. Med. Chem.* **2009**, *17*, 3671–3679.
- [14] Tumanov, V. V.; Tishkov, A. A.; Mayr, H. *Angew. Chem. Int. Ed.* **2007**, *46*, 3563–3566.
- [15] Ganem, B.; Xia, Q. *Org. Lett.* **2002**, *4*, 1631–1634.