

Manual for Hybrid Quantum Mechanical/Molecular Mechanics calculations within the Local Self-Consistent Field method and the Link-Atom scheme

Xavier Assfeld,^{1,*} Nicolas Ferré,^{2,†} Vincent Rooy, Yohann Moreau, and Pierre-François Loos

¹*Laboratoire de Chimie et Biochimie Théoriques,
UMR 7565 CNRS-UHP, Institut Jean Barriol (FR CNRS 2843),
Faculté des Sciences et Techniques, Nancy-Université, B.P. 239,
54506 Vandoeuvre-lès-Nancy Cedex, France.*

²*Universités Aix-Marseille I, II et III - CNRS
UMR 6517 - LCTMM, Site de Saint Jérôme - Case 521
Av. Escadrille Normandie-Niemen, F-13397 Marseille Cedex 20, France.*

(Dated: October 7, 2008)

The purpose of this manual is to provide an introduction to hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) calculations using the Local Self-Consistent Field (LSCF) method and the Link-Atom (LA) scheme. The implementation is based on a coupling between the Gaussian 03 software (Quantum mechanics calculations) and the Tinker 4.2 program (Molecular Mechanics calculations). The theory of these two QM/MM methodologies, as well as the implementation considerations, are briefly detailed, and an exhaustive list of examples are given at the end of the manuscript.

Keywords: Quantum Mechanics/Molecular Mechanics, Gaussian 03, Tinker 4.2, Local Self-Consistent Field, Link-Atom, Electrostatic Embedding, Microiterations, Self-Consistent Reaction Field, Single-point calculations, Geometry optimization, Molecular Dynamic

Contents

I. Introduction	2
II. Notations et Definitions	3
A. Notations	3
B. Summation range	3
III. Theory	3
A. QM/MM Hamiltonian	3
B. QM/MM energy	4
C. The Local Self-Consistent Field method	5
1. The LSCF equations	5
D. Orthogonalization of the FOs	6
1. Orthogonalization of the AOs	6
2. Modification of the Roothaan equations	7
3. Algorithm	8
4. Energy gradients	8
5. SLBOs derivatives	10
E. The Link-Atom scheme	10
F. The localization criteria	10
1. External criteria	10
2. The internal criteria	11
IV. Implementation and Computations	11
A. Molecular Mechanics: Tinker	11
1. A brief introduction to the Tinker 4.2 software	11
2. Installation	12

*Corresponding author; Electronic address: Xavier.Assfeld@cbt.uhp-nancy.fr

†Electronic address: Nicolas.Ferre@univ-provence.fr

3. Tinker files	12
4. Building macromolecular structure with Tinker	13
5. MM calculations with Tinker	16
6. QM/MM calculations with Tinker	16
B. Quantum Mechanics: Gaussian 03	18
1. Implementation considerations	18
2. Installation	19
3. The Gaussian input file	20
4. How can I prepare my QM input file from my Tinker xyz file ?	24
5. Where I can find information about my QM/MM job in the Gaussian output file ?	25
C. LSCF	28
1. Localization procedures	29
2. FO	30
3. SLBO	31
4. Orthogonalization schemes	33
5. SCSLBO	34
D. Molecular dynamic	35
1. QM:MM MD job	35
2. LSCF/MM MD job	36
3. HLA/MM MD job	37
V. Test Calculations	38
A. MM	38
B. QM	39
1. Localization	39
2. LSCF	39
C. QM/MM	40
1. QM:MM	40
2. LSCF/MM	41
3. HLA/MM	42
Acknowledgments	43
References	43

I. INTRODUCTION

During the past decade, the interest of theoreticians for macromolecular systems, or more generally large systems, has grown enormously. This is a direct consequence of the tremendous improvement of computers, both from the storage (memory and/or hard drive) and the CPU points of view. However, one has to bear in mind that studying chemical reactions in such systems (i.e. locating transition states) is still out of reach for the casual methods of quantum chemistry, in spite of the computer enhancement. To treat such large molecular systems, the community of quantum chemists has developed new methods that combine different levels of theory. The total system is divided in several parts, each one described at a given level of theory. The levels of theory are chosen such that, for each part, the most important physical phenomenon is correctly handled. Various schemes (combinations) are possible. For example, the part of the macromolecule undergoing the chemical reaction (generally few atoms) needs to be treated with high level (for example, correlated *ab initio*) techniques. This is called Quantum Mechanics (QM) level. The neighboring region, larger than the QM one, where consecutive reactions can take places for example proton transfer rearrangement like in the bacteriorhodopsin molecules can be described with a less time demanding method, typically a semiempirical one. Let us call this level QM' to avoid confusion with the former QM level. The remaining atoms of the macromolecule, which plays a nontrivial and nonisotropic role by means of electrostatic and mechanic interactions, are often handled with Molecular Mechanics (MM) force fields. Finally the surrounding solvent can be included explicitly via a classical force field (MM') or implicitly using a polarizable continuum (Self-Consistent Reaction Field, SCRF). The combination of these four levels of theory gives rise to a QM/QM'/MM:MM' or QM/QM'/MM:SCRF method, if one uses the "/" symbol to represent the separation between two parts connected by a chemical bond and the ":" character when no chemical bond connects the two parts. One could of course think of more complicated combinations, but the most widely known and used schemes of hybrid method are certainly the QM/MM and the

QM:MM ones [1–4]. Let us focus our attention on the methods that need the “/”, i.e. when covalent chemical bonds are formally cut to divide the system. Many solutions have been proposed and applied to circumvent the so-called dangling bond problem. If used carefully each scheme will correctly handle the connection between the two parts. We refer the interested reader to recent reviews for further details [5–7].

II. NOTATIONS ET DEFINITIONS

In this section, we define the overall notations and definitions that we will use in the QM/MM manual.

A. Notations

	Coefficients	Indices
AOs		$\mu, \nu, \lambda, \sigma$
x		Derivation wrt a nuclear displacement x
(x)		Derivation wrt a nuclear displacement x of the AOs integrals only
α		α MOs or SLBOs, α MOs energies, ...
β		β MOs or SLBOs, β MOs energies, ...
Occupied SLBOs	$(\mathbf{L})_{\mu I} = l_{\mu I}$	I, J, K, L
Occupied MOs	$(\mathbf{C})_{\mu i} = c_{\mu i}$	i, j, k, l
Virtual SLBOs	$(\mathbf{L})_{\mu A} = l_{\mu A}$	A, B, C, D
Virtual MOs	$(\mathbf{C})_{\mu a} = c_{\mu a}$	a, b, c, d
Occ. or virt. SLBOs	$(\mathbf{L})_{\mu P} = l_{\mu P}$	P, Q, R, S
Occ. or virt. MOs	$(\mathbf{C})_{\mu p} = c_{\mu p}$	p, q, r, s
MOs or SLBOs	$(\mathbf{D})_{\mu t} = d_{\mu t}$	t, u, v, w

B. Summation range

- M : Number of atoms
- n : Number of electrons
- N : Number of the AOs basis functions
- L : Number of FOs
- occ : Summation over all the occupied MOs or occupied SLBOs
- virt : Summation over all the virtual MOs or virtual SLBOs
- all : Summation over all the MOs or SLBOs (occupied and virtual)

III. THEORY

A. QM/MM Hamiltonian

In the case of the QM/MM methodology, the total hamiltonian of the system $\hat{\mathcal{H}}$ is written in the conventional form:

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}^{\text{QM}} + \hat{\mathcal{H}}^{\text{QM/MM}} + \hat{\mathcal{H}}^{\text{MM}} \quad (1)$$

where $\hat{\mathcal{H}}^{\text{QM}}$ and $\hat{\mathcal{H}}^{\text{MM}}$ are the QM and MM hamiltonian, respectively. $\hat{\mathcal{H}}^{\text{QM/MM}}$ correspond to the QM/MM interaction between the QM and the MM part. We can separate the $\hat{\mathcal{H}}^{\text{QM}}$ hamiltonian in two parts which correspond to the electronic and the nuclear part ($\hat{\mathcal{H}}_{\text{elec}}^{\text{QM}}$ and $\hat{\mathcal{H}}_{\text{nuc}}^{\text{QM}}$):

$$\hat{\mathcal{H}}^{\text{QM}} = \hat{\mathcal{H}}_{\text{elec}}^{\text{QM}} + \hat{\mathcal{H}}_{\text{nuc}}^{\text{QM}} \quad (2)$$

where

$$\hat{\mathcal{H}}_{\text{elec}}^{\text{QM}} = -\frac{1}{2} \sum_i^n \Delta_i - \sum_i^n \sum_A^M \frac{Z_A}{|\mathbf{R}_A - \mathbf{r}_i|} + \sum_i^n \sum_{j>i}^n \frac{1}{|\mathbf{r}_j - \mathbf{r}_i|} \quad (3)$$

$$\hat{\mathcal{H}}_{\text{nuc}}^{\text{QM}} = \sum_A^M \sum_{B>A}^M \frac{Z_A Z_B}{|\mathbf{R}_B - \mathbf{R}_A|} \quad (4)$$

$\hat{\mathcal{H}}^{\text{QM/MM}}$ is then written:

$$\hat{\mathcal{H}}^{\text{QM/MM}} = \hat{V}_{\text{elec}}^{\text{QM/MM}} + V_{\text{nuc}}^{\text{QM/MM}} + V_{\text{vdw}}^{\text{QM/MM}} \quad (5)$$

where $\hat{V}_{\text{elec}}^{\text{QM/MM}}$, $V_{\text{nuc}}^{\text{QM/MM}}$, $V_{\text{vdw}}^{\text{QM/MM}}$ are the electron-charge, nuclei-charge et van der Waals (vdw) interaction operator and potentials, respectively. $V_{\text{vdw}}^{\text{QM/MM}}$ is the vdw interaction between QM and MM atoms, given by the relation (6):

$$V_{\text{vdw}} = \sum_{A>B} 4\varepsilon_{AB} \left[\left(\frac{\sigma_{AB}}{R_{AB}} \right)^{12} - \left(\frac{\sigma_{AB}}{R_{AB}} \right)^6 \right] \quad (6)$$

It requires the knowledge of the $\{\varepsilon_A, \sigma_A\}_{A \in \text{QM}}$ parameters for the QM atoms. The most common choice is to use the parameters of the corresponding force field for the QM atoms. Another strategies could be used [8]. $V_{\text{nuc}}^{\text{QM/MM}}$ is given by the relation:

$$V_{\text{nuc}}^{\text{QM/MM}} = \sum_A^M \sum_{B \in \text{MM}} \frac{Z_A q_B}{|\mathbf{R}_B - \mathbf{R}_A|} \quad (7)$$

For the electrostatic interaction between the QM and the MM atoms, it exists several levels of approximation. In your implementation, we use the *Electrostatic Embedding* (EE) scheme, which explicitly polarized the QM wavefunction.

Within the EE scheme, the electrostatic interaction is calculated at the high level of theory:

$$\hat{V}_{\text{elec}}^{\text{QM/MM}} = - \sum_i^n \sum_{B \in \text{MM}} \frac{q_B}{|\mathbf{R}_B - \mathbf{r}_i|} \quad (8)$$

This interaction is added to the core hamiltonian of the QM part and allows the polarization of the wavefunction by the classical point charges of the MM force field [7].

B. QM/MM energy

According to the previous hamiltonian decomposition, we define the total energy E of a QM/MM system by the sum of the QM energy (E^{QM}), the MM energy (E^{MM}) and the interaction energy between the two parts:

$$\begin{aligned} E &= E^{\text{QM}} + E^{\text{MM}} + E^{\text{QM/MM}} \\ &= E_{\text{elec}}^{\text{QM}} + E_{\text{nuc}}^{\text{QM}} + E^{\text{MM}} + E_{\text{elec}}^{\text{QM/MM}} + E_{\text{nuc}}^{\text{QM/MM}} + E_{\text{vdw}}^{\text{QM/MM}} \end{aligned} \quad (9)$$

$E^{\text{QM/MM}}$ is defined by:

$$E^{\text{QM/MM}} = \langle \Psi | - \sum_i^n \sum_{B \in \text{MM}} \frac{q_B}{|\mathbf{R}_B - \mathbf{r}_i|} | \Psi \rangle + V_{\text{nuc}}^{\text{QM/MM}} + V_{\text{vdw}}^{\text{QM/MM}} \quad (10)$$

where the first term consist to calculate the electrostatic interaction within the EE scheme. The electronic energy of the QM part $E_{\text{elec}}^{\text{QM}}$ is given by the relation (11):

$$E_{\text{elec}}^{\text{QM}} = \sum_{\mu\nu}^N P_{\mu\nu}^T H_{\mu\nu} + \frac{1}{2} \sum_{\mu\nu\lambda\sigma}^N P_{\mu\nu}^T P_{\lambda\sigma}^T G_{\mu\nu\lambda\sigma} \quad (11)$$

Its expression differs depending on the method which is used to treat the QM/MM frontier.

C. The Local Self-Consistent Field method

The particularity of the Local Self-Consistent Field (LSCF) method is that it optimizes a wave function in the presence of frozen molecular orbitals [9–14]. The QM subsystem can be treated at both the spin-restricted or unrestricted Hartree-Fock (HF) and density-functional theory (DFT) level for single point calculations and geometry optimizations. Post-HF LSCF single-point calculations are also possible and extension to geometry optimizations is underway. In QM/MM calculations [15–21], the connecting bond is described by a Strictly Localized Bond Orbital (SLBO). This SLBO is determined on a model compound that possesses the chemical bond of interest and applied to the system under investigation according to the transferability principle. The two atoms defining the connection between the QM and the MM subsystems are called X and Y for the atom on the QM side and on the MM side respectively. The Y atom can be called a quantum-classical atom (QCA) since, while on one side it possesses basis functions and a nuclear charge, on the other side it has a classical atomic point charge, van der Waals parameters, and force field parameters corresponding to bonding interactions with the MM part. The classical parameters are those of the force field used to describe the MM part without any fudging adjustment for QM/MM calculations. The QM/MM interactions are summarized in Table I. One shall not forget that, in addition to these non-electronic QM/MM interactions, the MM point charges interact with the electrons of the QM fragment giving rise to the polarization of the wave-function. This polarization is the most important component of QM/MM methods.

For the derivation of the Local Self-Consistent Field (LSCF) equations and methodologies, we report the reader to the Ref. [9, 10, 14, 19, 22, 23]. The Ref. [9] concerns the definition of the LSCF energy, and the modification of the SCF process to perform LSCF calculations. Ref. [10] gathers the energy gradients for geometry optimization, and the definition of the Frontier Bond Potential (FBP). Ref. [19, 22, 23] contain details about several methodologies, used to ensure an accurate description of the QM/MM frontier for biological systems. Ref. [14] derives a new development of the LSCF method, called Optimized LSCF (OLSCF). Some applications of the LSCF method can be found in Ref. [11, 13, 15–18, 20, 21].

TABLE I: Details of the QM/MM interactions within the LSCF method. The atom of the quantum part are labelled Q and the ones of the MM part C. Z_A and q_A represent, respectively, the nuclear charge and the classical atomic point charge of atom A. $P_{\mu\nu}$ is the $\mu\nu^{th}$ element of the density matrix.

		Bonded terms
Bonds		Y-C ₁
Angles		X-Y-C ₁ , Y-C ₁ -C ₂
Torsions		Q ₁ -X-Y-C ₁ , X-Y-C ₁ -C ₂ , Y-C ₁ -C ₂ -C ₃
		Non-bonded terms
vdw ^a		$Q_i-Y_k, Q_i-C_B, Y_k-C_B, Y_k-Y_l^b$
Electrostatics	Z-Z	$Z_{Q_i}-Z_{Y_k}, Z_{Y_k}-Z_{Y_l}^b$
	Z-q	$Z_{Q_i}-q_{Y_k}, Z_{Q_i}-q_{C_B}, Z_{Y_k}-q_{C_B}, Z_{Y_k}-q_{Y_l}^b$
	q-q ^a	$q_{Y_k}-q_{C_B}, q_{Y_k}-q_{Y_l}^b$
	e-Z	$\sum_k \sum_{\mu\nu} P_{\mu\nu}^T \left\langle \mu \left \frac{Z_{Y_k}}{r_k - \mathbf{r}} \right \nu \right\rangle$
	e-q	$\sum_k \sum_{\mu\nu} P_{\mu\nu}^T \left\langle \mu \left \frac{q_{Y_k}}{r_k - \mathbf{r}} \right \nu \right\rangle, \sum_{A \in \text{MM}} \sum_{\mu\nu} P_{\mu\nu}^T \left\langle \mu \left \frac{q_A}{\mathbf{R}_A - \mathbf{r}} \right \nu \right\rangle$

^aStandard 1-4 condition of the force fields.

^b $k \neq l$

The LSCF equations have been implemented in the Gaussian 03 package [24]. Single-point calculation and geometry optimization have been allowed within HF and DFT methods. Only single-point calculations have been implemented for post-HF methods.

1. The LSCF equations

The LSCF method is devoted to solve the general problem of optimizing a wave function subject to the constraint that some predefined spin-orbitals remain unchanged, i.e., frozen. In the next paragraph, we present how the usual iterative procedure, the Self-Consistent Field (SCF) method, must be modified to account for this constraint. Within the LCAO approximation, the only data needed to solve the Hartree-Fock equations – or equivalently, the Kohn-Sham equations in Density Functional Theory (DFT) framework – are the expansion coefficients of the frozen spin-orbitals on the given basis set functions and their occupation numbers. In fact, no restriction is applied on these spin-orbitals,

which can be of any shape and occupied or not. This could be useful for the excited or ionized states of molecules. Because we predefine spin-orbitals, we can perform calculations in the restricted or the unrestricted formalism – or equivalently, the spin polarized method in DFT.

We start with a system containing an even number of electrons ($2n$) within the restricted formulation of Hartree-Fock equation. The variational MOs ($\{|p\rangle\}_{1 \leq p \leq (N-L)}$) of the system are developed on a gaussian basis set containing N AOs $\{|\mu\rangle\}_{1 \leq \mu \leq N}$:

$$|p\rangle = \sum_{\mu}^N c_{\mu p} |\mu\rangle \quad (12)$$

In the same manner, we consider L frozen orbitals (FOs) $\{|l_P\rangle\}_{1 \leq P \leq L}$ developed on the same AOs basis:

$$|l_P\rangle = \sum_{\mu}^N l_{\mu P} |\mu\rangle \quad (13)$$

D. Orthogonalization of the FOs

To avoid the factorial number of integrals to compute with nonorthogonal molecular orbitals, we require that the variationally optimized molecular orbitals be, of course, mutually orthogonal and, orthogonal to the frozen one. A simple recipe to fulfill this requirement is to expand the molecular orbitals on a set of functions already orthogonal to the frozen orbitals (FOs). We define a three-steps procedure to transform the original nonorthogonal basis set, composed of N functions (atomic orbitals), into a set of functions, mutually orthogonal, and orthogonal to the frozen orbitals. First, the set of L frozen orbitals $\{|l_P\rangle\}_{1 \leq P \leq L}$ is orthonormalized. It exists several methods which allow to orthogonalize these functions with the help of a orthogonalization matrix \mathbf{X} :

- The most common method is the Löwdin orthogonalization [25]. It consists to use as orthogonalization matrix \mathbf{X} the inverse square root of the FOs overlap matrix: $D_{PQ} = \langle l_P | l_Q \rangle$. The Löwdin matrix is then equal to $\mathbf{X} = \mathbf{D}^{-1/2}$ [26] where \mathbf{D} is the overlap matrix between FOs. This procedure is employed in the conventional SCF process in order to work in a AOs basis mutually orthogonal.
- The Gram-Schmidt (GS) orthogonalization procedure consists to take each vector and to subtract successively the projections of the previous vectors to form a vector orthogonal to the previous ones. This procedure is very useful because it allows the keep the first vector unchanged.

Because there is no restriction on the occupation numbers of the FOs, the Löwdin method lead to rotations between FOs. In the case where we define both occupied and virtual FOs, it results to rotations between the occupied and the virtual spaces. It has the effect to partially fill the virtual FOs space, leading to a wrong description of the system. In this case, one must use a different orthogonalization approach, as the GS method. This procedure is used for the study of the core excited and core ionized states of macromolecular systems [13].

1. Orthogonalization of the AOs

In the following, we assume that the FOs ($\{|l_{\mu P}\rangle\}_{1 \leq P \leq L}$) have been orthogonalized as previously described. The next step consists to build a new basis set which is mutually orthogonalized and orthogonal to the L FOs. As pioneered by Huzinaga [27], we project out of the subspace defined by the orthogonalized FOs each original basis function, and normalize it. We get an intermediate set of N functions orthogonal to the FOs ($|\tilde{\mu}\rangle$):

$$|\tilde{\mu}\rangle = \left(1 - \sum_P^L S_{\mu P}^2\right)^{-1/2} \left(|\mu\rangle - \sum_P^L |l_P\rangle \langle l_P | \mu\rangle\right) \quad (14)$$

where $S_{\mu P}$ is the overlap integral $\langle \mu | l_P \rangle$. It allows to define the matrix \mathbf{M} which transform the starting AOs $\{|\mu\rangle\}_{1 \leq \mu \leq N}$ in a basis set which is orthogonal to the FOs:

$$M_{\mu\nu} = \left(1 - \sum_P^L S_{\mu P}^2\right)^{-1/2} \left(\delta_{\mu\nu} - \sum_P^L \sum_{\lambda}^N l_{\nu P} l_{\lambda P} S_{\lambda\mu}\right) \quad (15)$$

Because it exists L linear dependencies introduced by the FOs and because the functions $\{|\tilde{\mu}\rangle\}_{1 \leq \mu \leq N}$ do not form a mutually orthogonalized set of basis functions, we proceed to a canonical orthogonalization [28] to obtain a set of $(N - L)$ functions $\{|\mu'\rangle\}_{1 \leq \mu' \leq (N-L)}$ on which the MOs can be expanded:

$$\mathbf{X} = \mathbf{U} \cdot \tilde{\mathbf{S}}^{-1/2} \quad (16)$$

where \mathbf{U} is the matrix containing the eigenvectors of the $\tilde{\mathbf{S}}$ matrix (overlap matrix between the $\{|\tilde{\mu}\rangle\}_{1 \leq \mu \leq N}$) and $\tilde{\mathbf{S}}$ is the same matrix in the eigenvectors basis set (*i.e.* the diagonal matrix). The combination of these two transformations defines the \mathbf{B} matrix of dimension $(N \times (N - L))$:

$$\mathbf{B} = \mathbf{M} \cdot \mathbf{X} \quad (17)$$

The \mathbf{B} matrix plays the same role than the $\mathbf{S}^{-1/2}$ matrix in the conventional SCF procedure. The Fock matrix \mathbf{F} is defined following the conventional relation:

$$F_{\mu\nu} = H_{\mu\nu}^c + \sum_{\lambda\sigma} P_{\lambda\sigma}^T \left[(\mu\nu|\lambda\sigma) - \frac{1}{2}(\mu\sigma|\lambda\nu) \right] \quad (18)$$

where we have use the chemist notation for the two-electron integrals $(\mu\nu|\lambda\sigma)$. \mathbf{P}^T is the density matrix of the system which is constituted by two contributions: a contribution due to the FOs (\mathbf{P}^F) and one due to the variational MOs (\mathbf{P}^Q).

$$P_{\mu\nu}^T = P_{\mu\nu}^Q + P_{\mu\nu}^F = 2 \sum_i^{\text{occ}} c_{\mu i} c_{\nu i} + 2 \sum_J^{\text{occ}} l_{\mu J} l_{\nu J} \quad (19)$$

The Fock matrix \mathbf{F} of dimension $(N \times N)$ is then transformed in the basis $\{|\mu'\rangle\}_{1 \leq \mu' \leq (N-L)}$ with the help of the \mathbf{B} matrix to give the \mathbf{F}' matrix of dimension $((N - L) \times (N - L))$ ($\mathbf{F}' = \mathbf{B}^\dagger \cdot \mathbf{F} \cdot \mathbf{B}$). Then, we diagonalize the \mathbf{F}' matrix to obtain the matrix of the $(N - L)$ eigenvectors \mathbf{C}' and the $(N - L)$ eigenvalues (ϵ) , and we back-transform to the initial basis set ($\mathbf{C} = \mathbf{B} \cdot \mathbf{C}'$) to calculate the new density matrix corresponding to the MOs part (\mathbf{P}^Q). By adding the contribution to the FOs (\mathbf{P}^G), we obtain the total density matrix of the system.

2. Modification of the Roothaan equations

The stationnarity condition of the energy with respect to the variational MOs coefficients $\{c_{\mu p}\}_{\substack{1 \leq \mu \leq N \\ 1 \leq p \leq N-L}}$ is,

$$\frac{\delta E}{\delta c_{\mu p}} = 0 \quad (20)$$

and the orthogonality constraint between the whole orbitals of the system (MOs and FOs),

$$\sum_{\mu\nu} c_{\mu p} S_{\mu\nu} c_{\nu q} = \delta_{pq} \quad (21)$$

$$\sum_{\mu\nu} l_{\mu P} S_{\mu\nu} l_{\nu Q} = \delta_{PQ} \quad (22)$$

$$\sum_{\mu\nu} c_{\mu p} S_{\mu\nu} l_{\nu Q} = 0 \quad (23)$$

lead to the modification of the Roothaan equations [10, 29]:

$$\mathbf{F} \cdot \mathbf{C} = \mathbf{S} \cdot \mathbf{C} \cdot \mathbf{E} + \mathbf{S} \cdot \mathbf{L} \cdot \mathbf{\Lambda} \quad (24)$$

where \mathbf{L} is the matrix of the FOs and $\{\Lambda_{pQ}\}_{\substack{1 \leq p \leq (N-L) \\ 1 \leq Q \leq L}}$ are the Lagrange multipliers of the MOs-FOs block:

$$\Lambda_{pQ} \equiv \varepsilon_{pQ} = \sum_{\mu\nu} c_{\mu p} F_{\mu\nu} l_{\nu Q} \quad (25)$$

Using the relations $\mathbf{C} = \mathbf{B} \cdot \mathbf{C}'$ and $\mathbf{X}^\dagger \cdot \mathbf{S} \cdot \mathbf{X} = \mathbb{I}$, and multiplying at left by \mathbf{B}^\dagger , we obtain:

$$\mathbf{F}' \cdot \mathbf{C}' = \mathbf{C}' \cdot \mathbf{E} \quad (26)$$

with $\mathbf{F}' = \mathbf{B}^\dagger \cdot \mathbf{F} \cdot \mathbf{B}$ and $\mathbf{C}'^\dagger \cdot \mathbf{S} \cdot \mathbf{L} = \mathbb{O}$ (orthogonality between MOs and FOs). It leads to a hermitian problem which can be solved by diagonalizing the \mathbf{F}' matrix. We refer the readers to the Ref. [29] and [10] for more information about the modifications of the Roothaan equations and the introduction of the Lagrange multipliers.

3. Algorithm

We summarize here the LSCF algorithm implemented the Gaussian 03 package:

1. Orthogonalize the FOs.
2. Build the \mathbf{M} matrix which orthogonalize the AOs basis to the FOs.
3. Build the \mathbf{X} matrix. Transform the functions obtained in 2. in a orthogonalized basis set, linearly independent.
4. Compute $\mathbf{B} = \mathbf{M} \cdot \mathbf{X}$.
5. Define the Fock matrix \mathbf{F} in the initial basis set.
6. Compute $\mathbf{F}' = \mathbf{B}^\dagger \cdot \mathbf{F} \cdot \mathbf{B}$.
7. Diagonalize \mathbf{F}' : $\epsilon = \mathbf{C}'^\dagger \cdot \mathbf{F}' \cdot \mathbf{C}'$ where ϵ is the eigenvalue matrix.
8. Transform the eigenvectors in the initial basis set: $\mathbf{C} = \mathbf{B} \cdot \mathbf{C}'$.
9. Compute \mathbf{P}^Q .
10. Exit test. If not satisfied, go back to 5.

4. Energy gradients

In order to determine the forces which act on the atoms and to perform geometry optimizations, the expression of the energy first derivatives with respect to the nuclear coordinates x is necessary [28, 30–32]. In the following, we use the notations introduced in the book of Yamaguchi *et al.* [33]. The x exponent corresponds to the derivation of a quantity following the nuclear displacement x , while the $^{(x)}$ exponent correspond to the derivation of the AOs integrals only (not the coefficients). For example:

$$\begin{aligned} F_{\mu\nu}^x &= H_{\mu\nu}^{(x)} + \sum_{\lambda\sigma} P_{\lambda\sigma}^T G_{\mu\nu\lambda\sigma}^{(x)} + \sum_{\lambda\sigma} (P_{\lambda\sigma}^T)^x G_{\mu\nu\lambda\sigma} \\ &= F_{\mu\nu}^{(x)} + \sum_{\lambda\sigma} (P_{\lambda\sigma}^T)^x G_{\mu\nu\lambda\sigma} \\ H_{\mu\nu}^x &\equiv H_{\mu\nu}^{(x)} \end{aligned} \quad (27)$$

where $\{d_{\mu t}\}_{1 \leq \mu \leq N, 1 \leq t \leq N}$ are the coefficients of MOs or FOs. The determination of the first derivatives of the LSCF energy imposes several modifications compared to the conventional approach [34] due to the non-variationality of the FOs. Indeed, the MOs coefficients $\{c_{\mu p}\}_{1 \leq \mu \leq N, 1 \leq p \leq (N-L)}$ are obtained with respect to the minimization of the energy, while the FOs coefficients $\{l_{\mu P}\}_{1 \leq \mu \leq N, 1 \leq P \leq L}$ are kept unchanged. Following the chain rule derivative, the expression of the energy derivative with respect to x (E^x) is then:

$$\begin{aligned} E^x &= E^{(x)} + \sum_{\mu} \sum_{t \text{ all}} \frac{\partial E}{\partial d_{\mu t}} d_{\mu t}^x \\ &= E^{(x)} + \sum_{\mu} \sum_{p \text{ all}} \frac{\partial E}{\partial c_{\mu p}} c_{\mu p}^x + \sum_{\mu} \sum_{P \text{ all}} \frac{\partial E}{\partial l_{\mu P}} l_{\mu P}^x \end{aligned} \quad (28)$$

However, as $\partial E/\partial c_{\mu P} = 0$ and $\partial E/\partial l_{\mu P} \neq 0$, it follows:

$$E^x = E^{(x)} + \sum_{\mu} \sum_P^{\text{all}} \frac{\partial E}{\partial l_{\mu P}} l_{\mu P}^x \quad (29)$$

The expression of E^x then requires the FOs derivatives with respect to x . The derivation of the relation (11) with respect to x gives [28]:

$$E^x = \sum_{\mu\nu} P_{\mu\nu}^T H_{\mu\nu}^{(x)} + \frac{1}{2} \sum_{\mu\nu\lambda\sigma} P_{\mu\nu}^T P_{\lambda\sigma}^T G_{\mu\nu\lambda\sigma}^{(x)} + \sum_{\mu\nu} (P_{\mu\nu}^T)^x F_{\mu\nu} \quad (30)$$

which can be written using the relation (24):

$$\begin{aligned} E^x = & \sum_{\mu\nu} P_{\mu\nu}^T H_{\mu\nu}^{(x)} + \frac{1}{2} \sum_{\mu\nu\lambda\sigma} P_{\mu\nu}^T P_{\lambda\sigma}^T G_{\mu\nu\lambda\sigma}^{(x)} - \sum_{\mu\nu} W_{\mu\nu} [I] S_{\mu\nu}^{(x)} \\ & + \sum_{\mu\nu} \sum_I^{\text{occ}} l_{\mu I}^x F_{\mu\nu} l_{\nu I} + \sum_{\mu\nu} \sum_I^{\text{occ}} l_{\mu I} F_{\mu\nu} l_{\nu I}^x \\ & - \sum_{\mu\nu} \sum_P^{\text{all}} \left[l_{\mu P}^x (\mathbf{S} \cdot \mathbf{P}^Q \cdot \mathbf{F})_{\mu\nu} l_{\nu P} + l_{\mu P} (\mathbf{F} \cdot \mathbf{P}^Q \cdot \mathbf{S})_{\mu\nu} l_{\nu P}^x \right] \end{aligned} \quad (31)$$

where

$$W_{\mu\nu} [I] = \sum_i^{\text{occ}} \varepsilon_i c_{\mu i} c_{\nu i} + 2 \sum_i^{\text{occ}} \sum_P^{\text{all}} \varepsilon_i P c_{\mu i} l_{\nu P} \quad (32)$$

is the energy-weighted density matrix corresponding to the MOs-MOs and MOs-FOs blocks. It gives:

$$\begin{aligned} E^x = & \sum_{\mu\nu} P_{\mu\nu}^T H_{\mu\nu}^{(x)} + \frac{1}{2} \sum_{\mu\nu\lambda\sigma} P_{\mu\nu}^T P_{\lambda\sigma}^T G_{\mu\nu\lambda\sigma}^{(x)} - \sum_{\mu\nu} W_{\mu\nu} [I] S_{\mu\nu}^{(x)} \\ & + \sum_{\mu\nu} (P_{\mu\nu}^G)^x F_{\mu\nu} - \sum_{\mu\nu} \sum_i^{\text{occ}} \sum_P^{\text{all}} [\varepsilon_i P l_{\mu P}^x S_{\mu\nu} c_{\nu i} + \varepsilon_i P c_{\mu i} S_{\mu\nu} l_{\nu P}^x] \end{aligned} \quad (33)$$

The expression of E^x can be expand as:

$$\begin{aligned} E^x = & \sum_{\mu\nu} P_{\mu\nu}^T H_{\mu\nu}^{(x)} + \frac{1}{2} \sum_{\mu\nu\lambda\sigma} P_{\mu\nu}^T P_{\lambda\sigma}^T G_{\mu\nu\lambda\sigma}^{(x)} - \sum_{\mu\nu} W_{\mu\nu} [I] S_{\mu\nu}^{(x)} \\ & + \sum_{\mu\nu} (P_{\mu\nu}^G)^x F_{\mu\nu} - \sum_{\mu\nu} (W_{\mu\nu} [II])^x S_{\mu\nu} \end{aligned} \quad (34)$$

with

$$(P_{\mu\nu}^G)^x = \sum_I^{\text{occ}} l_{\mu I}^x l_{\nu I} + \sum_I^{\text{occ}} l_{\mu I} l_{\nu I}^x \quad (35)$$

$$(W_{\mu\nu} [II])^x = \sum_i^{\text{occ}} \sum_P^{\text{all}} [\varepsilon_i P l_{\mu P}^x c_{\nu i} + \varepsilon_i P c_{\mu i} l_{\nu P}^x] \quad (36)$$

where $(W_{\mu\nu} [II])^x$ corresponds to a part of the derivative of the energy-weighted density matrix associated with the MOs-FOs block. We note that the expression (34) require the knowledge of the FOs derivatives with respect to the nuclear coordinates.

5. SLBOs derivatives

As shown in the section III D 4, the energy gradients require the knowledge of the FOs derivatives with respect to a nuclear displacement x , $\{l_{\mu P}^x\}_{1 \leq P \leq L}$. When using SLBOs, this quantity can be determined by numerical differentiation [35, 36]:

$$l_{\mu P}^x(x) = \frac{l_{\mu P}(x + \delta x) - l_{\mu P}(x - \delta x)}{2\delta x} \quad (37)$$

where $l_{\mu P}(x + \delta x)$ is the coefficient of the AOs μ in the SLBO $|l_P\rangle$ in regard to a displacement x . Indeed, the coefficients of the SLBOs $|l_P\rangle$ depends only of the coordinates of the two atoms X_P and Y_P . Moreover, as their number are relatively restraint, the determination of their numerical derivatives is not time-consuming.

E. The Link-Atom scheme

The most simple and widely used scheme, called the Link-Atom (LA), use a monovalent atom to cap the unsaturated QM atoms. The LA is usually an hydrogen atom [3, 4, 37–44]. Details about the LA scheme can be found on Ref. [37, 38, 40, 42, 43, 45–52]. For some examples of the LA applications, we refer the readers to the Ref. above and Ref. [8, 53–62]. On December 2007, N. Ferré has included Morokuma’s scheme for link atom positions and gradients [40, 45]. The $R(Q_1 - \text{HLA})$ distance is then modified following the equation:

$$R(Q_1 - \text{HLA}) = c_{\text{HLA}} \cdot R(Q_1 - C_1) \quad (38)$$

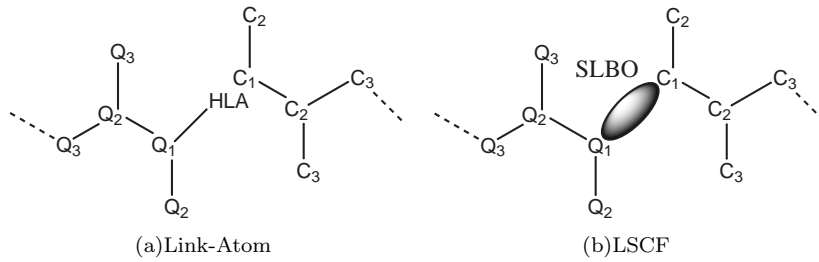
with

$$c_{\text{HLA}} = \frac{R_o(Q_1 - \text{HLA})}{R_o(Q_1 - C_1)} \quad (39)$$

Note that the scaling factor is currently hard-coded, using standard atomic radii. In the LA scheme, special treatments of the MM boundary charges are used because of the over-polarization of the wave function due to the nearby point charges on the QM/MM frontier [3, 6, 44, 63–67]. These MM point charges might be either zeroed, scaled, redistributed [67] or partially delocalized [44] to avoid the unphysical polarization.

The LA scheme has been implemented in the Gaussian 03 [24] and the Molcas 7.0 [68] package. Single-point and geometry optimization calculations are allowed within HF, DFT and post-HF methods, yet. In order to perform QM/MM calculations using Molcas 7.0 (for CASSCF and CASPT2 calculations mainly), we refer the reader to the website of N. Ferré for more details (http://sites.univ-provence.fr/lcp-ct/ferre/nf_tinker_qmmm.html)

FIG. 1: QM/MM methodologies: Link-Atom and LSCF approach.



F. The localization criteria

1. External criteria

The two first criteria that we will be briefly detailed correspond to external criteria, i.e. the user must specified the A and B atoms where he wants the MOs to be localized.

a. Weinstein-Pauncz The Weinstein-Pauncz criterion [69, 70] (WP) consists to maximize the quantity $\langle \Omega \rangle_{\text{WP}}$ which represents the Mulliken overlap population [71] between the two atoms A et B :

$$\langle \Omega \rangle_{\text{WP}} = \sum_{\mu \in A} \sum_{\nu \in B} c_{\mu i} S_{\mu\nu} c_{\nu i} \quad (40)$$

b. Magnasco-Perico For the Magnasco-Perico criterion [72, 73] (MP), it maximizes the quantity $\langle \Omega \rangle_{\text{MP}}$ given by the relation (41).

$$\langle \Omega \rangle_{\text{MP}} = \sum_{\mu \in \{A, B\}} \sum_{\nu \in \{A, B\}} c_{\mu i} S_{\mu\nu} c_{\nu i} \quad (41)$$

It corresponds to maximizing the overlap population between the atoms A and B in the same manner that the WP criterion, but takes into account the electronic population of the two atoms.

2. The internal criteria

a. Boys-Foster The Boys-Foster criterion [74–76] (BF) minimizes the squared distance between two electrons:

$$\langle \Omega \rangle_{\text{BF}} = \sum_i^n \left\langle \phi'_i \phi'_i \left| (\mathbf{r}_2 - \mathbf{r}_1)^2 \right| \phi'_i \phi'_i \right\rangle \quad (42)$$

Physically, it corresponds to determine a set of orbitals that are the less spatially spread. In other words, the BF method consist in maximizing the distance between the orbital centroids. Unlike the Pipek-Mezey method, the BF method does not conserve the separation between the $|\sigma\rangle$ and $|\pi\rangle$ orbitals. Indeed, the method produces "banana" bonds" (τ et τ') which are symmetric and anti-symmetric combination of the σ and π bonds:

$$|\tau\rangle = \frac{1}{\sqrt{2}} (|\sigma\rangle + |\pi\rangle) \quad (43)$$

$$|\tau'\rangle = \frac{1}{\sqrt{2}} (|\sigma\rangle - |\pi\rangle) \quad (44)$$

b. Pipek-Mezey The Pipek-Mezey (PM) localization criterion [77, 78] maximizes the sum of the Mulliken atomic charges [71]. The atomic charge of the i^{th} MO centered on the atom A is given by:

$$q_i(A) = \sum_{\mu \in A} \sum_{\nu} c_{\mu i} c_{\nu i} S_{\mu\nu} \quad (45)$$

The maximized function is then:

$$\langle \Omega \rangle_{\text{PM}} = \sum_A^M [q_i(A)]^2 \quad (46)$$

IV. IMPLEMENTATION AND COMPUTATIONS

In this section, we will refer to the terminal prompt by `user@localhost:`, following by the absolute path where you are currently working. The absolute path where the Tinker and Gaussian 03 programs are located will be referred by `$TK4QMMM` and `$LSCFMM` variables, respectively.

A. Molecular Mechanics: Tinker

1. A brief introduction to the Tinker 4.2 software

The Tinker molecular modeling software is a complete and general package for molecular mechanics and dynamics, with some special features for biopolymers. Tinker has the ability to use any of several common parameter sets, such as Amber [79–85] (ff94, ff96, ff98 and ff99), CHARMM [86–91] (CHARMM19 and CHARMM27), Allinger MM [92–98, 98–105] (MM2-1991 and MM3-2000) and OPLS [106–112] (OPLS-UA, OPLS-AA and OPLS-AA/L).

2. Installation

The Tinker 4.2 software [113] is used to perform the MM part of the QM/MM computation. We encourage the reader to first read the Tinker user's guide that can be found in <ftp://dasher.wustl.edu/pub/tinker/doc/guide.pdf>. The Tinker package can be freely download at <http://dasher.wustl.edu/tinker/>. Then, you must install it and apply the patch that include all the modifications of the Tinker subroutine, following the details given at http://sites.univ-provence.fr/lcp-ct/ferre/nf_tinker_qmmm.html. Actually, this part of the installation is the easiest to do.

Do not forget to add the path of the Tinker binary file to your PATH variable, as mentioned in the Tinker manual.

First, you must move in the Tinker source code directory,

```
user@localhost:~ cd $TK4QMMM/source
```

To build the Tinker binary files, do not forget to compile the additional source files, as listed above in the case of the xlf IBM compiler.

```
xlf -c -O elecpot.f
xlf -c -O county.f
xlf -c -O qmmmsetup.f
xlf -c -O tkr2qm.f
xlf -c -O runqm.f
xlf -c -O hess4qm.f
xlf -c -O dynaqmmm.f
```

These lines may be added in the `compile.make` file which is located in the `source` directory of the Tinker program. Moreover, the lines listed below may be added to the `library.make` file,

```
elecpot.o \
county.o \
qmmmsetup.o \
hess4qm.o \
tkr2qm.o \
runqm.o
```

as well as these two lines in the `link.make` file in order to build the two binary files which are necessary to run QM/MM jobs with Tinker.

```
xlf -o tkr2qm.x tkr2qm.o -L. -ltinker
xlf -o dynaqmmm.x dynaqmmm.o -L. -ltinker
```

Do not forget also to add these two lines in the `rename.make` file in order to move the binary files in the `bin` directory.

```
mv tkr2qm.x ../bin/tkr2qm
mv dynaqmmm.x ../bin/dynaqqmm
```

At this point the Tinker binary files could be built with the following command lines:

```
user@localhost:~ cd $TK4QMMM/source
user@localhost:$TK4QMMM/source ./compile.make
user@localhost:$TK4QMMM/source ./library.make
user@localhost:$TK4QMMM/source ./link.make
user@localhost:$TK4QMMM/source ./rename.make
```

We refer the reader to the `qmmm.i` file in the `source` directory of the Tinker program for information concerning the QM/MM variables. Commands for other OS and/or compiler can be found in the main Tinker directory. For example, makefiles and compiler command for Linux OS are located in the `linux` directory.

3. Tinker files

Several files are of first importance to run a MM and/or a QM/MM calculation using Tinker:

- The **xyz** file is the basic Tinker Cartesian coordinates file type. It contains a title line followed by one line for each atom in the structure. Each line contains: the sequential number within the structure, an atomic symbol or name, X-, Y-, and Z-coordinates, the force field atom type number of the atom, and a list of the atoms connected to the current atom.
- The keyword parameter file always has the extension **key**. It contains values for any of a wide variety of switches and parameters that are used to change the course of the computation from the default.
- The potential energy parameter files **prm**. Each of these files contains a definition of the potential energy functional forms for that force field as well as values for individual energy parameters.
- The **dyn** file contains values needed to restart a molecular or stochastic dynamics computation. It stores the current position, current velocity and current and previous accelerations for each atom, as well as the size and shape of any periodic box or crystal unit cell. This information can be used to start a new dynamics run from the final state of a previous run.

For example, we present the **xyz** file of the ethane molecule using the OPLS-AA force field (**oplsaa.prm**) on FIG. 2:

FIG. 2: Tinker **xyz** file of the ethane molecule using the OPLS-AA force field (**oplsaa.prm**).

```
8 ethane
  1 C    0.000000    0.000000    0.000000    1    2    3    4    5
  2 H    0.000000    0.000000    1.089000    6    1
  3 H    1.026719    0.000000   -0.363000    6    1
  4 H   -0.513360   -0.889165   -0.363000    6    1
  5 C   -0.683537    1.183920   -0.483333    1    1    6    7    8
  6 H   -0.683537    1.183920   -1.572333    6    5
  7 H   -1.710256    1.183920   -0.120333    6    5
  8 H   -0.170177    2.073085   -0.120333    6    5
```

and its corresponding **key** file (FIG. 3).

FIG. 3: Tinker **key** file of the ethane molecule using the OPLS-AA force field (**oplsaa.prm**).

```
PARAMETERS oplsaa.prm
DIGITS 8
VERBOSE
ARCHIVE
```

We recommend to use the **DIGITS 8** keyword in order to increase the digits of precision during a MM calculation. In certain cases, the **DEBUG**, **VERBOSE** and **ARCHIVE** keywords can also be used (see the Tinker manual for more details about these keywords). The **CHARGE** keyword is widely used to modify the point charge value of the MM atoms [118]. For an example of **prm** file, we refer the reader to the **params** directory of the Tinker software. Additional examples for Tinker job can be also found in the **example** directory of the Tinker software.

4. Building macromolecular structure with Tinker

We recommend to use the **protein** and **nucleic** program that you can find in Tinker (see the manual for more details). Here is two examples of how to build a polypeptide (FIG. 4 and 5) and a double-stranded DNA fragment (FIG. 6 and 7).

For nucleic acid, be very careful of the notation corresponding to the nucleoside (desoxyribose ou ribose). For example, the deoxycytidine is noted (DCY), while the cytidine is noted (CYT).

FIG. 4: Creation of a Ace-Lys-NMe tripeptide with the **protein** program of the Tinker software using the CHARMM27 force field (charmm27.prm).

```
#####
#####
##
##          TINKER  ---  Software Tools for Molecular Design      ##
##
##                      Version 4.2   June 2004                    ##
##
##          Copyright (c)  Jay William Ponder  1990-2004          ##
##                      All Rights Reserved                        ##
##
#####
#####

Enter Name to be used for Output Files : Lys

Enter Title :  Ace-Lys-NMe

Enter Potential Parameter File Name :  charmm27.prm

Enter One Residue per Line, Free Format:  1 or 3 Letter Code, then
Phi/Psi/Omega (3F), Chi Angles (4F), then Disulfide Partner if a
CYS (I) and D/L Chirality as desired (A1), <CR> ends Residue entry

The allowed N-Cap Residues are Acetyl=ACE or Formyl=FOR, and the
possible C-Cap Residues are N-MethylAmide=NME or Amide=NH2

Enter Residue   1 :  ACE

Enter Residue   2 :  LYS

Enter Residue   3 :  NME

Enter Residue   4 :

Cyclize the Polypeptide Chain [N] :  N
```

FIG. 5: Ace-Lys-NMe tripeptide built with the **protein** program.

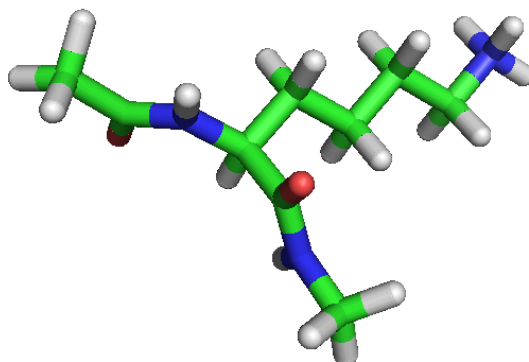


FIG. 6: Creation of a double-stranded DNA fragment of sequence 5'-CG-CG-CG-3' with the `nucleic` program of the Tinker software using the AMBER *ff99* force field (`amber99.prm`). The CHARMM parameters for nucleic acids are not yet implemented in the Tinker software.

```
#####
#####
##
##          TINKER  ---  Software Tools for Molecular Design      ##
##
##                      Version 4.2  June 2004                    ##
##
##          Copyright (c)  Jay William Ponder  1990-2004          ##
##                      All Rights Reserved                        ##
##
#####
#####
```

Enter Name to be used for Output Files : CCC

Enter Title : CG-CG-CG

Enter Potential Parameter File Name : amber99.prm

Enter A-, B- or Z-Form Helix for the Structure [B] : B

Enter One Nucleotide per Line, 5' to 3': Give 3 Letter Code,
followed by Backbone Torsions (6F) and Glycosidic Torsion (1F)

Use Residue=MOL to Begin a New Strand, Residue=<CR> to End Entry

Enter Residue 1 : DCY

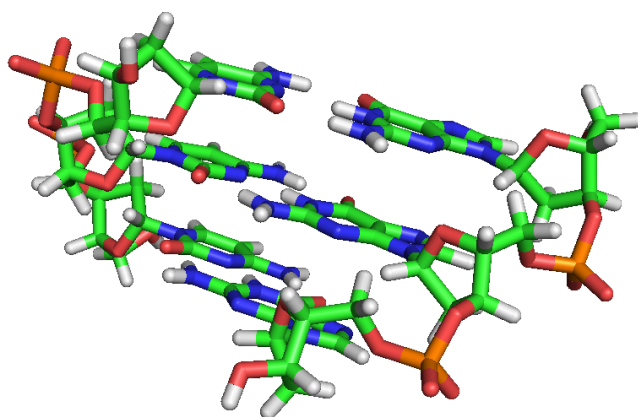
Enter Residue 2 : DCY

Enter Residue 3 : DCY

Enter Residue 4 :

Build a Double Helix using Complimentary Bases [N] : Y

FIG. 7: Double-stranded DNA of sequence 5'-CG-CG-CG-3' building with the **nucleic** program.



If your starting geometry comes from a **pdb** file from the Protein Data Bank, you can use the Tinker program **pdxyz**, that prepares a **xyz** file from the **pdb** one. We would like to point out that the Molden software [114] is very useful to handle some modifications on the **xyz** and **pdb** files, and also to prepare files in order to run QM/MM calculations [119].

5. MM calculations with Tinker

Several Tinker potential energy programs are of prime importance: **analyze** (in order to check the validity of your **xyz** file), **minimize** (see also **optimize** and **newton**) to perform geometry optimization, **dynamic** (MD or stochastic dynamics), **pss** and **scan** (search algorithm for global optimization of molecular conformation). The most widely used programs for QM/MM calculations are **analyze** (check the **xyz** file), **minimize** (to obtain a good starting geometry for the QM/MM calculations from the MM results, if possible) and **xyzedit** (prepare your Gaussian input file).

6. QM/MM calculations with Tinker

Additional Tinker QM/MM keywords: the following keywords must be placed in the **key** file, in the same way than the other Tinker keywords (see above):

- **QMMM N M**: This keyword means that **N** atoms of the whole molecular system are also defined in the Gaussian input. **M** is only required for LSCF calculations to specify the number of LSCF frontier atoms.
- **[QM,MM,LA,YA] ATOM_I**: **ATOM_I** is defined as a quantum mechanical (QM), molecular mechanical (MM), link atom (LA) or LSCF frontier atom (YA).
- **QMMM-ELECTROSTATICS [DIRECT,ESPF]**: This keyword selects the method for computing the QM/MM electrostatic interactions. The **DIRECT** method computes the electrostatic integrals over the MM point charges, as it is common in the majority of the QM/MM methodologies. The **ESPF** method replaces this scheme by an approximated one, more adapted to a sophisticated MM force field which involves for instance multipoles, periodic boundary conditions, ...
- **QMMM-MICROITERATION [ON,OFF]**: This keyword turns **ON** or **OFF** the MM geometry minimization for each Gaussian geometry optimization step (it is the so-called microiteration scheme). The QM-derived charges are possibly Mulliken charges (default), derived from the electrostatic potential (**MK**, **CHELP** or **CHELPG**) or NPA charges. This specification is set by the **POPULATION** keyword in the Gaussian input file (http://gaussian.com/g_ur/k_population.htm).

- **QMMM-HESSIAN [ON,OFF]**: This keyword turns ON or OFF the computation of the full Hessian of the system, used to extract the part involving only the atoms defined into Gaussian. This can be used by Gaussian to get a better model hessian. This feature is very memory consuming and should be considered only for "small" molecular systems.
- **QMMM-EXTERNAL [GAUSSIAN,MOLCAS] SHELL_COMMAND**: This keyword indicates that now Tinker is the driver of the QM/MM calculation. This allows the use of some Tinker programs like `minimize`. It allows also the use of `dynaqmmm`, a QM/MM-oriented molecular dynamics program. First, select the QM program (currently Gaussian 03 or Molcas) that has to be run. Second, write the `SHELL_COMMAND` able to run the QM program.

On FIG. 8, we present an example of the Tinker `key` file for a LSCF/MM calculation on the ethane molecule. The first methyl group is described at the QM level, while the second one is modeled using the OPLS-AA force field. The QM/MM frontier is located on the central C-C bond. The `xyz` file on FIG. 2 can be used for such calculations.

FIG. 8: Tinker `key` file for a LSCF/MM calculation on the ethane molecule using the OPLS-AA force field (`oplsaa.prm`).

```
PARAMETERS oplsaa.prm
DIGITS 8
VERBOSE
QMMM 8 1
QM 1
QM 2
QM 3
QM 4
YA 5
MM 6
MM 7
MM 8
QMMM-ELECTROSTATICS DIRECT
QMMM-MICROITERATION OFF
QMMM-HESSIAN ON
```

The FIG. 9 gathers the Tinker `key` file in the case of a QM/MM calculation using the LA scheme:

FIG. 9: Tinker `key` file for a QM/MM calculation on the ethane molecule using the LA scheme and the OPLS-AA force field (`oplsaa.prm`). The MM charges have been set to zero with the help of the `CHARGETERM NONE` keyword. The `ATOM` keyword is used to define the HLA.

```
PARAMETERS oplsaa.prm
DIGITS 8
VERBOSE
CHARGETERM NONE
ATOM 2999 1 HA "HLA" 1 1.008 0
QMMM 9
QM 1
QM 2
QM 3
QM 4
MM 5
MM 6
MM 7
MM 8
LA 9
QMMM-ELECTROSTATICS DIRECT
QMMM-MICROITERATION OFF
QMMM-HESSIAN ON
```

For QM/MM calculation using the LA scheme, the `xyz` file must be modified to take into account the HLA, as shown on FIG. 10:

FIG. 10: Tinker xyz file of the ethane molecule in the case of a QM/MM computation using the LA scheme. The OPLS-AA force field (`oplsaa.prm`) has been used to model the MM level. HLA is the last atom of the file.

```

9  ethane
1  CT      .00000000      .00000000      .00000000      1      2      3      4      5
2  HC      .00000000      .00000000      1.08900000      6      1
3  HC      1.02672000      .00000000      -.36300000      6      1
4  HC      -.51336000      -.88916000      -.36300000      6      1
5  CT      -.68354000      1.18392000      -.48333000      1      1      6      7      8
6  HC      -.68354000      1.18392000      -1.57233000      6      5
7  HC      -1.71026000      1.18392000      -.12033000      6      5
8  HC      -.17018000      2.07308000      -.12033000      6      5
9  HA      -.50159000      .86878000      -.35468000      2999      5      1

```

Before to move on the QM calculations, we would like to point out a very important thing about the electroneutrality of the MM part within QM/MM calculations. Indeed, the users must be check that the overall charge of the MM surrounding sum as an integer. A fractional or non-zero MM charge may induce artifical polarization of the QM wavefunction [7]. In this case, the MM charges may be modified with the help of the RESP (*Restreint Electrostatic Potential*) procedure [115, 116] for example.

B. Quantum Mechanics: Gaussian 03

We encourage the reader to first read the Gaussian 03 user manual that can be found in http://www.gaussian.com/g_ur/g03mantop.htm.

1. Implementation considerations

To run usual jobs (single-point calculations, geometry optimizations, ...), the Gaussian 03 software is considered as the leader program (Gaussian 03 call Tinker 4.2 for informations about the MM part of the QM/MM job). In the following, we describe the content of the modified Gaussian links which are called during a QM/MM calculation:

- 11: Additional QM/MM keywords (LSCF and LA), see above.
- 1101: Modification of the Gaussian input file for QM/MM calculations (LSCF and LA).
- 1202: Define the coordinates of the HLA (LA). Read the new coordinates of the QM atoms during a molecular dynamic simulation (LSCF and LA).
- 1301: Recover the MM energy, MM energy gradient and MM hessian matrix from the Tinker calculations (LSCF and LA). Definition of the frontier atom (LSCF and LA). Set the nuclear charge of the Y atom and the corresponding atomic basis set (LSCF). Frontier bond potential (LSCF). Compute the QM/MM electrostatic interaction (LSCF and LA).
- 1397: ESPF calculation for QM/MM electrostatic interaction [117].
- 1401: Small modification for LSCF calculations: set to zero the number of frozen electrons in the Y frontier atoms within the frozen core approach.
- 1502: Modification of the SCF process to perform LSCF calculations within HF formalism and DFT (LSCF).
- 1601: Localization of MOs in order to obtain LOs (LSCF).
- 1701: Modification of the HF and DFT energy gradients (LSCF). Computation of the QM/MM electrostatic interaction derivatives (LSCF and LA).
- 1716: Modification of the energy gradient for LA scheme (constraint position of the HLA).
- 1801: Post-SCF calculations within the LSCF formalism (MP2, CC, CIS, CISD and TD-DFT single-point calculations). Exclude FOs from the post-HF calculations in a way similar to the Frozen Core approach.

- 19999: Export the QM energy gradients and QM charges during a MD job (in this case, Tinker is the leader program).

2. Installation

The first step of the installation is to create your own directory to compile the modified Gaussian links. In the following, we refer to this directory by the environment variable `LSCFMM`. The second step is to copy the native Gaussian links that you want to modify (listed above) in this directory, as well as the file `ertgen.inc`.

```
user@localhost:~ mkdir $LSCFMM
user@localhost:~ cd $LSCFMM
user@localhost:$LSCFMM for i in l1.F l101.F l202.F l301.F l397.F l401.F l502.F l601.F l701.F
l716.F l801.F 19999.F ertgen.inc ; do cp $g03root/g03/$i $LSCFMM/ ; done
```

Then, one must create a directory for each link. For example, create a directory 1502 which will be associated with the link 1502.F.

```
user@localhost:$LSCFMM mkdir $LSCFMM/1502
```

Copy this file in the directory 1502 and use the `fsplit` command to split the file 1502.F:

```
user@localhost:$LSCFMM cp $LSCFMM/1502.F $LSCFMM/1502/
user@localhost:$LSCFMM cd $LSCFMM/1502
user@localhost:$LSCFMM/1502 fsplit $LSCFMM/1502.F
```

It has the effect to create several fortran files associated with each subroutine contained in the 1502.F file. Then, you can remove the file 1502.F to the 1502 directory.

```
user@localhost:$LSCFMM/1502 rm $LSCFMM/1502/1502.F
```

In addition, several Gaussian files are necessary. Create a `bsd` directory in the `$LSCFMM` directory and copy the four Gaussian files that can be found in the `$g03root/g03/bsd` directory:

```
user@localhost:$LSCFMM cd $LSCFMM
user@localhost:$LSCFMM mkdir $LSCFMM/bsd
user@localhost:$LSCFMM cd $LSCFMM/bsd
user@localhost:$LSCFMM/bsd for i in g03.make main.F testrt.F updatelink1 ;
do cp $g03root/g03/bsd/$i $LSCFMM/bsd ; done
```

At the stage, the QM/MM patch (`g03_for_qmmm.diff`) could be used to modify the source code. The command is:

```
user@localhost:$LSCFMM/bsd cd $LSCFMM
user@localhost:$LSCFMM patch<g03_for_qmmm.diff
```

The modified links listed above (except `l1.exe`) are compiled with the `makeexe` script which is depicted on Figure 11:

FIG. 11: `makeexe` script used to compile the modified Gaussian links.

```
#!/bin/bash
. $g03root/g03/bsd/g03.profile
if [ $# == 0 ]
then echo " Usage: makeexe [-e x|a|ax|xa] N"; exit;
elif [ $# == 1 ]
then
    typeset -r link=$1;
    echo "Compiling l${link}.F";
    rm -f l${link}.a l${link}.exe l${link}.res ;
    cat l${link}/*.f > l${link}.F;
elif [ $1 == "-e" ]
then
    typeset -r link=$3;
    if [ $2 = "x" ]
    then rm -f l${link}.exe;
    elif [ $2 = "ax" ] || [ $2 = "xa" ] || [ $2 = "a" ]
    then rm -f l${link}.a l${link}.exe;
    else echo "Arguments for -e unknown" $2; exit 3;
    fi
    rm l${link}.res
    echo "Compiling l${link}.F";
else
    echo "Option unknown" $1; exit 2;
fi
make -f ./bsd/g03.make l${link}.exe 1>>l${link}.res 2>&1
grep -i "Error" l${link}.res
echo "Finished ! Check l${link}.res"
chmod o-rwx l${link}.*
```

The command to compile these links is `makeexe "link number"`. For example, to compile the link 1502, one must type the line:

```
user@localhost:$LSCFMM ./makeexe 502
```

In addition to these links, `l1.exe`, `testrt` and `g03` binaries are built with the help of the `makeg03` script depicted on Figure 12. To do so, type the command `makeg03`:

```
user@localhost:$LSCFMM ./makeg03
```

FIG. 12: `makeg03` script used to build the `l1.exe`, `testrt` and `g03` binaries.

```
#!/bin/bash
# To compile the QM/MM version of g03
. $g03root/g03/bsd/g03.profile
cat l1/*.f > l1.F
rm -f l1.a g03 testrt.o
make -f ./bsd/g03.make l1.exe 1>l1.res 2>&1
make -f ./bsd/g03.make g03 1>>l1.res 2>&1
make -f ./bsd/g03.make testrt 1>>l1.res 2>&1
chmod o-rwx l1.* g03 testrt
grep -i "Error" l1.res
```

You are now ready to use the QM/MM code.

3. The Gaussian input file

The basic structure of a Gaussian input file includes several different sections:

- Link 0 Commands: Locate and name scratch files (not blank line terminated).
- Route section (# lines): Specify desired calculation type, model chemistry and other options (blank line terminated).
- Title section: Brief description of the calculation (blank line terminated).
- Molecule specification: Specify molecular system to be studied (blank line terminated).
- Optional additional sections: Additional input needed for specific job types (usually blank line terminated).

In order to use the modified links, you must add the following line in the % section of the Gaussian input file:

```
%SUBST 1101 $LSCFMM
%SUBST 1202 $LSCFMM
%SUBST 1301 $LSCFMM
%SUBST 1397 $LSCFMM
%SUBST 1401 $LSCFMM
%SUBST 1502 $LSCFMM
%SUBST 1601 $LSCFMM
%SUBST 1701 $LSCFMM
%SUBST 1716 $LSCFMM
%SUBST 1801 $LSCFMM
%SUBST 19999 $LSCFMM
```

The absolute path where the gaussian executable can be found is placed after the %SUBST keyword and the name of the corresponding link. Moreover, we provide information about the additional Gaussian keywords introduced in order to perform QM/MM calculations using the LSCF method and the LA scheme. The use of these keywords are strictly equivalent to the usual Gaussian keywords:

- QMMM:
 - SLBO=Read: perform QM/MM calculation by means of the LSCF methodology. The SLBOs are read at the end of the Gaussian input file.
 - LinkAtom1: perform QM/MM calculation by means of the LA scheme.
 - LinkAtom2: LA scheme with $\forall A \in \text{MM}, \left\langle \mu \left| \frac{q_A}{|\mathbf{R}_A - \mathbf{r}|} \right| \nu \right\rangle = 0$ if $\mu \cup \nu \in \text{HLA}$
 - LinkAtom3: LA scheme with $\forall A \in \text{MM}, \left\langle \mu \left| \frac{q_A}{|\mathbf{R}_A - \mathbf{r}|} \right| \nu \right\rangle = 0$ if $\mu \cap \nu \in \text{HLA}$
 - Tinker1: compute QM/MM electrostatic interaction with the DIRECT method.
 - Tinker2: obsolete (same as Tinker1).
 - Tinker3: compute QM/MM electrostatic interaction by means of the ESPF formalism [117].
- LSCF:
 - FO=MM: perform calculation with MM FOs, which are read at the end of the Gaussian input file.
 - SLBO=NN: perform calculation with NN SLBOs, which are read at the end of the Gaussian input file.
- NoSym: This is *not* a QM/MM keyword but it is required during a QM/MM calculation in order to keep the input orientation for both LSCF and LA computations.

LinkAtom2 and LinkAtom3 were coded by N. Ferré for technical considerations [53]. In the case QMMM=SLBO=Read, you must include the LSCF=(FO=MM,SLBO=NN) keyword where MM is the number of FOs and NN the number of SLBOs.

After the keyword line and the title line (separate by a blank line), you should add the charge and multiplicity. These two quantities stand for those of the QM part only (the MM charges and multiplicity are not taken into account). Next, the geometry specifications are gathered following a cartesian or Z-matrix format. We recommend to use cartesian coordinates for large systems. To specify the type of the atom, the ONIOM notation is used (see the Gaussian 03 manual for more details at the address http://www.gaussian.com/g_ur/k_oniom.htm). The syntax of a line is: "Atomic number or Atomic symbol" X Y Z "QM/MM type" where QM/MM type" is given by the letter H for a QM atom, L for a MM atom and M for an HLA or LSCF frontier atom. Between the atomic number and the cartesian coordinates can be placed the integer 0 or -1 to specify that the cartesian coordinates of the corresponding

atom is allowed to change or is kept fixed, respectively. Its option is particularly suitable for partial optimization of large systems (protein backbone frozen for example).

After a blank line, one must specify the name of the **xyz** Tinker file (less or equal to 8 characters), followed by the number of atoms in this file.

The remaining of the Gaussian input file contains informations for LSCF calculations (see below). Here is an example of Gaussian input file to perform a QM/MM calculation on the ethane molecule using the LA scheme (FIG. 13) and the LSCF method (FIG. 14):

FIG. 13: Gaussian input file of a QM/MM calculation using the LA scheme. The molecule considered here is the ethane molecule.

```
%SUBST L101 $LSCFMM
%SUBST L202 $LSCFMM
%SUBST L301 $LSCFMM
%SUBST L701 $LSCFMM
%SUBST L716 $LSCFMM
%CHK=ethane
#P RHF ST0-3G Opt SCF=Tight Pop=Full GFInput
  NoSym QMMM=(LinkAtom1,Tinker1)

TITLE HERE

0 1
C 0 .00000 .00000 .00000 H
H 0 .00000 .00000 1.08900 H
H 0 1.02672 .00000 -.36300 H
H 0 -.51336 -.88916 -.36300 H
C 0 -.68354 1.18392 -.48333 L
H 0 -.68354 1.18392 -1.57233 L
H 0 -1.71026 1.18392 -.12033 L
H 0 -.17018 2.07308 -.12033 L
H 0 -.50159 .86878 -.35468 M

ethane 9
```

FIG. 14: Gaussian input file of a QM/MM calculation using the LSCF scheme. The molecule considered here is the ethane molecule.

```
%SUBST L101 $LSCFMM
%SUBST L301 $LSCFMM
%SUBST L502 $LSCFMM
%SUBST L701 $LSCFMM
%CHK=ethane
#P RHF STO-3G Opt SCF=Tight Pop=Full GFInput
  NoSym QMMM=(SLBO=Read,Tinker1) LSCF=SLBO=1

TITLE HERE

O 1
C 0 .00000 .00000 .00000 H
H 0 .00000 .00000 1.08900 H
H 0 1.02672 .00000 -.36300 H
H 0 -.51336 -.88916 -.36300 H
C 0 -.68354 1.18392 -.48333 M
H 0 -.68354 1.18392 -1.57233 L
H 0 -1.71026 1.18392 -.12033 L
H 0 -.17018 2.07308 -.12033 L

ethane      8

XY_POT
1 5 -26.3145 5.2319 -1.0393 -0.3589 20.1330
****

SLBO
11 1 5 2 ROTATE
  0.020662 -0.035787 0.014610
-0.704198 1.219707 -0.497943
  0.032606 -0.056475 1.100392
  -0.0100366 0.3562863-0.1995480 0.3456273-0.1411018
  -0.0100366 0.3562863 0.1995480-0.3456273 0.1411018
****
```

For LSCF/MM calculations, additional information are required (XY_POT keyword). After the atom number of the X and Y atoms (**strictly in this order**), the user must specify the frontier bond parameters in order to ensure an accurate description of the LSCF frontier. The frontier bond potential has the form:

$$E_{X-Y} = (A + Br + Cr^2) e^{Dr} + \frac{E}{r} \quad (47)$$

where r is the distance between the X and Y atoms. The five parameters $\{A, B, C, D, E\}$ have been determined on the basis of pure QM calculations. These parameters depend of the atom type (mainly C, N or O) of the two atoms defining the QM/MM boundary, as well as their hybridation (sp^2 or sp^3). However, their are mainly basis set independent. All the parameters can be found in Ref. [10] and in the Table II:

TABLE II: Parameters of the frontier bond potentials (in a.u.) for various types of frontier bonds (X-Y).

Bond type	A	B	C	D	E
C(sp ³)-C(sp ³)	-26.3145	5.2319	-1.0393	-0.3589	20.1330
C(sp ³)-O	-25.0856	5.0907	-1.2836	-0.3848	18.8380
O-C(sp ³)	-26.8871	4.8309	-0.9887	-0.3693	20.5643
C(sp ³)-N	-28.3881	6.0076	-1.2983	-0.3736	20.9211
N-C(sp ³)	-27.1145	5.0382	-1.0480	-0.3705	20.7263
C(sp ²)-N	-37.0482	8.9376	-2.0228	-0.4089	24.3843
N-C(sp ²)	-30.9690	6.2066	-1.3907	-0.3892	22.0465
C(sp ²)-C(sp ³)	-28.9594	6.1121	-1.2288	-0.3687	21.2616
C(sp ³)-C(sp ²)	-26.7094	5.3146	-1.0772	-0.3634	20.2739

This section, which starts with the XY_POT keyword, must be ended by ****. The last part of the input file contain the SLBO notifications: we refer the reader to Section IV C of additional information.

4. How can I prepare my QM input file from my Tinker xyz file ?

With the help of the xyzedit program followed by the name of your xyz file, the QM input file can be generated from the Tinker xyz file.

```
#####
#####
##                                     ##
##           TINKER  ---  Software Tools for Molecular Design           ##
##                                     ##
##           Version 4.2  June 2004                                     ##
##                                     ##
##           Copyright (c)  Jay William Ponder  1990-2004             ##
##           All Rights Reserved                                         ##
##                                     ##
#####
#####
```

This is a QM/MM calculation

5 is defined as a LSCF/MM frontier Y atom.

The TINKER XYZ Edit Facility can Provide :

- (1) Offset the Numbers of the Current Atoms
- (2) Deletion of Individual Specified Atoms
- (3) Deletion of Specified Types of Atoms
- (4) Deletion of Atoms outside Cutoff Range
- (5) Insertion of Individual Specified Atoms
- (6) Replace Old Atom Type with a New Type
- (7) Assign Connectivities based on Distance
- (8) Convert Units from Bohrs to Angstroms
- (9) Invert thru Origin to give Mirror Image
- (10) Translate All Atoms by an X,Y,Z-Vector
- (11) Translate Center of Mass to the Origin
- (12) Translate a Specified Atom to the Origin
- (13) Translate and Rotate to Inertial Frame
- (14) Move to Specified Rigid Body Coordinates

- (15) Move Stray Molecules into Periodic Box
- (16) Create and Fill a Periodic Boundary Box
- (17) Soak Current Molecule in Box of Solvent
- (18) Append another XYZ file to Current One
- (19) Prepare a QM/MM input (G03/Tinker)
- (20) Prepare a QM/MM input (Molcas/Tinker)

Enter the Number of the Desired Choice :

At the stage, we can choose to generate a Gaussian (19) or a Molcas (20) input file. Be careful, your input files are not complete, and you must add or modify some information in order to run a QM/MM calculation (the QM and QM/MM keywords for instance, as well as the atom layers). Here is an generated Gaussian input from the `xyzedit` utils for the ethane molecule:

```
%SUBST L101
%SUBST L301
%SUBST L302
%SUBST L397
%SUBST L701
%CHK=ethane
%MEM=
#P KEYWORDS HERE
NoSym QMMM=(LinkAtom1,Tinker1)

TITLE HERE

CHARGE & MULTIPLICITY HERE
C 0 .00000 .00000 .00000 H
H 0 .00000 .00000 1.08900 H
H 0 1.02672 .00000 -.36300 H
H 0 -.51336 -.88916 -.36300 H
C 0 -.68354 1.18392 -.48333 M
H 0 -.68354 1.18392 -1.57233 L
H 0 -1.71026 1.18392 -.12033 L
H 0 -.17018 2.07308 -.12033 L

ethane      8
```

5. Where I can find information about my QM/MM job in the Gaussian output file ?

To answer this question, we will consider the `test2000` calculation. Details about the calculation can be found in the Section V C 2 and the Gaussian input file is depicted on Figure 15

FIG. 15: Gaussian input file of the `test2000` calculation.

```

%SUBST L101 $LSCFMM
%SUBST L301 $LSCFMM
%SUBST L502 $LSCFMM
%SUBST L701 $LSCFMM
%CHK=test2000
%MEM=100MB
#P RHF/6-31G* Opt=(GDIIS,Big)
  NoSym QMMM=(SLBO=Read,Tinker1) LSCF=SLBO=1

TITLE HERE

1 1
C 0  -.10655  -.60010  .27316  L
C 0  -.26389  -.65943  1.74293  L
O 0   .40212  -1.42386  2.42801  L
H 0   .68757  -1.31314  -.04031  L
H 0   .18837   .42210  -.04428  L
H 0  -1.05594  -.88403  -.22829  L
N 0  -1.17779   .15423  2.28790  L
C 0  -1.42259   .20181  3.71902  M
C 0  -2.38643  -.89850  4.16367  L
O 0  -3.53817  -.64884  4.51712  L
H 0  -1.72616   .74317  1.70641  L
H 0   -.48606   .03166  4.23493  L
C 0  -1.94803  1.60075  4.13836  H
C 0  -1.96839  1.83376  5.66098  H
C 0  -2.67400  3.14464  6.04443  H
C 0  -2.81404  3.31375  7.56043  H
N 0  -3.58564  4.53809  7.87188  H
H 0  -2.97177  1.73756  3.72352  H
H 0  -1.29332  2.37117  3.67305  H
H 0   -.92383  1.82482  6.04591  H
H 0  -2.51151   .98852  6.14476  H
H 0  -3.68408  3.12559  5.57209  H
H 0  -2.10361  3.99383  5.60509  H
H 0  -1.81307  3.39483  8.03782  H
H 0  -3.35336  2.44304  7.99539  H
H 0  -3.68089  4.64653  8.90216  H
H 0  -4.53231  4.46575  7.44593  H
H 0  -3.09739  5.36878  7.48067  H
N 0  -1.91276  -2.15757  4.15620  L
C 0  -2.71120  -3.28999  4.56180  L
H 0   -.98926  -2.32468  3.80369  L
H 0  -2.12901  -4.23286  4.46066  L
H 0  -3.62385  -3.37472  3.92877  L
H 0  -3.03138  -3.18804  5.62328  L

test2000      34

XY_POT
13 8 -26.3145 5.2319 -1.0393 -0.3589 20.1330
****

SLBO
11 13 8 18 ROTATE
  0.018235  -0.031584   0.012894
 -0.701772   1.215504  -0.496228
  0.034689  -0.060084   1.097989
-0.0100239 0.1990300-0.1448146 0.2508263-0.1023994 0.1764988-0.0676370 0.1171506
-0.0478266-0.0002957 0.0133157-0.0036985-0.0136114 0.0055568-0.0096247
-0.0100239 0.1990300 0.1448146-0.2508263 0.1023994 0.1764988 0.0676370-0.1171506
  0.0478266-0.0002957 0.0133157-0.0036985-0.0136114 0.0055568-0.0096247
****

```

The first information is provided in the l101.

(Enter \$LSCFMM/l101.exe)

```
*****
* This is a QM/MM calculation *
*   with the LSCF/MM method   *
*****
```

The message may change depending on the QM/MM that you are using. The l301 provides information on the MM part of the QM/MM calculation:

(Enter \$LSCFMM/l301.exe)

```
#####
#####
##                                     ##
##          TINKER  ---  Software Tools for Molecular Design          ##
##                                     ##
##                               Version 4.2  June 2004                ##
##                                     ##
##          Copyright (c)  Jay William Ponder  1990-2004             ##
##                               All Rights Reserved                   ##
##                                     ##
#####
#####
```

This is a QM/MM calculation

8 is defined as a LSCF/MM frontier Y atom.

MM energy components passed:

```
eb  =      .00111096 ua =      .69713823 kcal/mol
ea  =      .00240304 ua =      1.50793322 kcal/mol
eub =      .00013881 ua =      .08710404 kcal/mol
eid =      .00004759 ua =      .02986379 kcal/mol
et  =      .00668359 ua =      4.19401440 kcal/mol
ev  =      .00015655 ua =      .09823969 kcal/mol
ec  =     -0.04328081 ua =     -27.15911751 kcal/mol
```

End of this Tinker run

It gives the MM energy components in Hartree and in kcal/mol. It stands for the bond stretching, angle bending, Urey-Bradley, improper dihedral, torsional, van der Waals and Coulomb parameters (CHARMM27 force field). See the Tinker manual for more details about the different MM components. Following the Tinker section, the MM part of the QM/MM energy is given in Hartree (-0.032740 in this case). This contribution is added to the nuclear repulsion energy. The two following sentences indicates that frontier bond potential parameters have been found between frontier atoms 13 and 8: the corresponding energy (EXY) has been added to the previous calculated MM energy (EMM).

```
MM part of the QM/MM energy =      -.032740 ... to be added to the nuclear repulsion energy
Input QM/MM frontier bond potential
Frontier bond potential for atoms 13 and 8
EMM + EXY =      -.117202
```

The main LSCF part of the QM/MM calculation has been done in the l502. At the beginning of this link, there are some information about the calculations and, especially the number of FOs or SLBOs (see next Section).

(Enter /w30/QMMM/LSCFMM/l502.exe)

```
This is the QM part of a LSCF/MM calculation
This is a LSCF calculation: 0 frozen orbitals and 1 SLBOs
Warning: only s, p, d, f functions can be rotated
```

Within LSCF calculations, only s, p, d, f AOs functions can be rotated.

The rest of the information provided by the 1502 corresponds to details about occupation, rotation, normation and orthogonalization of the FOs and/or SLBOs. In particular, the output provides the AOs coefficients of the SLBOs after rotation and normalization.

```

This SLBO 1 may be rotated ...
... Read the old (X,Y,R) coordinates
... Rotation OK, check the norm
1 SLBOs read succesfully
There are 1 alpha frozen orbitals in LSCF
Occupation numbers for alpha FOs are:
2
For ALPHA FOs:
FO 1: norm = .9913144063
Maybe you should check this FO
For this geometry, FOs are :
11 13 8 18
-.0100677 .1999000 .1044529 -.2780972 -.0833613 .1772703 .0487857 -.1298878
-.0389346 -.0036071 .0178564 -.0048871 -.0108378 -.0032487 .0086494
-.0100677 .1999000 -.1044529 .2780972 .0833613 .1772703 -.0487857 .1298878
.0389346 -.0036071 .0178564 -.0048871 -.0108378 -.0032487 .0086494
NOLA = 1
NOLB = 0
Lowdin for Occupied and Virtual SLBO

```

The energy given at the end of the 1502 corresponds to the total QM/MM energy (Eq. (9)):

```

SCF Done: E(RHF) = -212.297266134      A.U. after 13 cycles
          Conv = .5773D-08             -V/T = 2.0013
          S**2 = .0000

```

C. LSCF

First, we will consider calculations using Frozen Orbitals (FOs) and/or Strictly Localized Bond Orbitals (SLBOs). The LSCF calculation is done by 1502 for the energy calculation and 1701 for energy gradients computations.

Do not forget to substitute this link to perform geometry optimization or forces computations within the LSCF framework.

At this stage, we would like to point out that geometry optimization is not enabled for FOs calculations due to their delocalized character.

First, one must specify the number of SLBOs and/or FOs. It is specified by I0p(5/199=MMNN) or by the keyword LSCF=(FO=NN,SLBO=MM) (see Section IV B 3 for more details) where NN and MM are the numbers of FOs and SLBOs, respectively.

```

C I0p(199) ... LSCF computation
C 0 ... Default (No)
C NN ... NN SLBO (only the X-Y coefficients are given in input)
C MM00 ... MM FO (all coefficients are given in input)

```

The LSCF computations using FOs and/or SLBOs also require the occupation and the spin state of these orbitals. 11 stands for an α occupied orbital and 10 for an empty one. β -spin orbitals are indicated using the minus symbol: -11 and -10 for occupied and empty ones, respectively. In the case of restricted calculations, only α -spin orbitals are required: each one are then doubly occupied or empty. For unrestricted calculations, the user must specify the spatial part of the FOs and/or the SLBOs for each spin state. We would like to point out that calculations with both FOs and SLBOs can also be performed. First, we detail how to obtain such FOs and SLBOs by means of localization criteria:

1. Localization procedures

It exists four localization criteria in our modified version of the Gaussian 03, implemented in 1601 and specified with `IOP(198)`: Weinstein-Pauncz (WP) [69, 70], Magnasco-Perico (MP) [72, 73], Boys-Foster (BF) [74–76] and Pipek-Mezey [78] (PM). The WP and MP criteria allow to determine SLBOs, while the BF and PM are used for both FOs and SLBOs. In order to obtain SLBOs in the case of the BF and PM criteria, one must remove the orthogonalization-delocalization tails.

`C2IOP(198)`

```
C   IOP(198) ... Localization criterion.
C       0 ... Default (No).
C       1 ... Weinstein-Pauncz (WP)
C       2 ... Weinstein-Pauncz + Magnasco-Perico (WP + MP)
C       3 ... Boys-Foster (BF)
C       4 ... Pipek-Mezey (PM)
C       5 ... Boys-Foster + Pipek-Mezey (BF + PM)
```

In addition, you can choose the MOs window that you want to localize: it is controlled by `IOP(196)` and `IOP(197)`:

`C2IOP(196)`

```
C   IOP(196) ... MOs Window for localization for BF and/or PM
C       0 ... Occupied MOs (default)
C       1 ... Occupied and Virtual MOs (separately)
```

`C2IOP(197)`

```
C   IOP(197) ... Splitting option in localization criteria for BF and/or PM
C       0 ... No splitting in Occupied MOs (default)
C       1 ... Split Core/Valence in Occupied MO during localization
```

The two options are commonly used for OLSCF calculations [14] to finely tune the determination of the SLBO and SLABO in two well-defined subspaces (see Section IV C 5 for more details). We present an example of Gaussian input file for the localization (using WP criteria) of the two σ O-H bonds of the water molecule (FIG. 16). The structure has been optimized at the RHF/STO-3G level of theory.

FIG. 16: Gaussian input file for the localization of the two σ O-H bonds of the water molecule using the WP localization criterion. The molecular geometry has been obtained at the RHF/STO-3G level of theory.

```
%Subst 1601 $LSCFMM
#p RHF/STO-3G SP SCF=Tight
NoSym IOp(6/198=1)

TITLE HERE

O 1
O          -.047439      .000000      -.034838
H          .033512      .000000      .951323
H          .909596      .000000      -.286152

1 2
1 3
```

Compared to a standard Gaussian input file, you should add the path of the modified 1601 and choose the localization criterion IOp(6/198). When one wants to use the WP or the MP criteria, the list of the atoms between which you want to localize the orbital must be specified at the end of the file. This last requirement is not necessary for the BF and PM criteria since they are global. Below, we present two examples of output files (parts) obtained after WP localization procedure:

```
Start of MO localization:
...
Input format for SLBOs:
SLBO 1 between atoms 1 and 2:
  -.0432304 .4053887 .0051270 .0000000 .5005524
  .4820168
SLBO 2 between atoms 1 and 3:
  .0000008 -.0000044 -.4120131 .0000000 .5819783
  -.5228053
End of MO localization
```

The WP and MP criteria provide MOs coefficients in the AOs basis of the two atoms defined by the input. For the BF criterion (and PM criterion), the output is of the form:

```
BOYS-FOSTER L0s
Boys Occupied L0s
LocMO: Using Boys method
Initial Trace= .21450016D+00 Initial TraceA= .69982772D-01
RMSG= .49003652D-10
```

	1	2	3	4	5
1	1.01299	-.06667	.06667	.09721	.09721
2	-.06497	.25023	-.25023	-.66105	-.66105
3	.02668	.06783	-.56851	.30614	.30614
4	.00000	.00000	.00000	-.70711	.70711
5	.01889	.57885	.12835	.21674	.21674
6	-.00254	.53060	.10471	.10160	.10160
7	-.00254	-.10471	-.53060	.10160	.10160

Next, we will consider calculations using FOs and SLBOs.

2. FO

The FOs are specified by the keyword FO at the end of the input file. The end of the section is indicated by a line containing ****. The FOs are defined in columns (five per row max.). The first row indicates the spin and the occupation number. FIG. 17 and 18 depict common input files for restricted and unrestricted LSCF calculations using FOs.

Within the LSCF calculation, the number of α electrons must be always superior to the number of β electrons.

FIG. 17: Gaussian input file for restricted calculations using FOs at the RHF/STO-3G level of theory. The FOs are the two σ O-H bonds of the water molecule localized by the BF criterion. The first σ O-H bond is doubly occupied, while the second one is empty. The molecular geometry has been obtained at the RHF/STO-3G level of theory.

```
%Subst 1502 $LSCFMM
#p RHF/STO-3G SP SCF=Tight
NoSym LSCF=FO=2

TITLE HERE

0 1
      8      0      -.047439      .000000      -.034838
      1      0      .033512      .000000      .951323
      1      0      .909596      .000000      -.286152

FO
11 10
-.06667      .06667
 .25023     -.25023
 .06783     -.56851
 .00000      .00000
 .57885      .12835
 .53060      .10471
-.10471     -.53060
****
```

FIG. 18: Gaussian input file for unrestricted calculations using FOs at the UHF/STO-3G level of theory. The FOs are the two σ O-H bonds of the water molecule localized by the BF criterion. The first σ O-H bond is doubly occupied (one α and one β electron), while the second one is singly occupied with one α electron. The molecular geometry has been obtained at the RHF/STO-3G level of theory.

```
%Subst 1502 $LSCFMM
#p UHF/STO-3G SP SCF=Tight
NoSym LSCF=FO=4

TITLE HERE

1 2
      8      0      -.047439      .000000      -.034838
      1      0      .033512      .000000      .951323
      1      0      .909596      .000000      -.286152

FO
11 11 -11 -10
-.06667      .06667  -.06667      .06667
 .25023     -.25023  .25023     -.25023
 .06783     -.56851  .06783     -.56851
 .00000      .00000  .00000      .00000
 .57885      .12835  .57885      .12835
 .53060      .10471  .53060      .10471
-.10471     -.53060  -.10471     -.53060
****
```

3. SLBO

The syntax of the SLBO section is :

```
"spin and occupation" "X atom number" "Y atom number" "R atom number" ROTATE
```

Coordinates of the X atom in the model system (X, Y, Z)
 Coordinates of the Y atom in the model system (X, Y, Z)
 Coordinates of the R atom in the model system (X, Y, Z)
 list of the AOs coefficients of the X atom
 list of the AOs coefficients of the Y atom

The first row contains four integers and one character string. The keyword **ROTATE** indicates the SLBOs has to be rotated from the model system to the current molecular geometry [10]. The three lines gather the cartesian coordinates of the X, Y and R atoms defining the reference frame for the geometry of the model system. The two following lists which contain the X and Y coefficients are real number (eight per line max.).

FIG. 19: Gaussian input file of the methanediol molecule using restricted calculations. The SLBOs are determined at the RHF/STO-3G level of theory on the water molecule using the WP criterion and describe the two σ O-H bonds. The first σ O-H bond is doubly occupied, while the second one is empty. The molecular geometry has been obtained at the RHF/STO-3G level of theory.

```
%Subst 1502 $LSCFMM
#p RHF/STO-3G SP SCF=Tight
NoSym LSCF=SLBO=2

TITLE HERE

0 1
6      0      -.014554      .000000      -.028447
1      0      .000878      .000000      1.069835
1      0      1.030075      .000000      -.390225
8      0      -.692438      -1.193908      -.426892
1      0      -.622334      -1.203652      -1.416024
8      0      -.692438      1.193908      -.426892
1      0      -.622334      1.203652      -1.416024

SLBO
11 4 5 1 ROTATE
-.047439      .000000      -.034838
.033512      .000000      .951323
.909596      .000000      -.286152
-.0432304 .4053887 .0051270 .0000000 .5005524
.4820168
10 6 7 1 ROTATE
-.047439      .000000      -.034838
.909596      .000000      -.286152
.033512      .000000      .951323
.0000008 -.0000044 -.4120131 .0000000 .5819783
-.5228053
****
```


FIG. 20: Gaussian input file of the methanediol molecule using restricted calculations. The SLBOs are determined at the RHF/STO-3G level of theory on the water molecule using the WP criterion and describe the two σ O-H bonds. The first σ O-H bond is doubly occupied (one α and one β electron), while the second one is singly occupied with one α electron. The molecular geometry has been obtained at the RHF/STO-3G level of theory.

```
%Subst 1502 $LSCFMM
#p UHF/STO-3G SP SCF=Tight
NoSym LSCF=SLBO=4

TITLE HERE

2 1
6      0      -.014554      .000000      -.028447
1      0      .000878      .000000      1.069835
1      0      1.030075      .000000      -.390225
8      0      -.692438      -1.193908      -.426892
1      0      -.622334      -1.203652      -1.416024
8      0      -.692438      1.193908      -.426892
1      0      -.622334      1.203652      -1.416024

SLBO
11 4 5 1 ROTATE
-.047439      .000000      -.034838
.033512      .000000      .951323
.909596      .000000      -.286152
-.0432304 .4053887 .0051270 .0000000 .5005524
.4820168
11 6 7 1 ROTATE
-.047439      .000000      -.034838
.033512      .000000      .951323
.909596      .000000      -.286152
-.0432304 .4053887 .0051270 .0000000 .5005524
.4820168
-11 4 5 1 ROTATE
-.047439      .000000      -.034838
.909596      .000000      -.286152
.033512      .000000      .951323
.0000008 -.0000044 -.4120131 .0000000 .5819783
-.5228053
-10 6 7 1 ROTATE
-.047439      .000000      -.034838
.909596      .000000      -.286152
.033512      .000000      .951323
.0000008 -.0000044 -.4120131 .0000000 .5819783
-.5228053
****
```

4. Orthogonalization schemes

When using both occupied and virtual FOs and/or SLBOs, you must be very careful about the orthogonalization scheme that you used to avoid rotation and subsequent mixing between occupied and virtual orbitals. For more details about this point, see Ref. [13]. The orthogonalization scheme is chosen with the IOp(198), as listed below:

```
CLSCF-Begin
C      IOp(198) ... Orthogonalization scheme for SLBOs and/or FOs
C      0 ... Lowdin scheme for Occupied and Virtual simultaneously (Default)
C      1 ... Gram-Schmidt for Occupied and Virtual simultaneously
C      2 ... Lowdin for Virtual followed by Gram-Schmidt for Occupied
C      3 ... Gram-Schmidt for Occupied followed by Lowdin for Virtual
```

5. SCSLBO

In this part, we present a Gaussian input file to perform Optimized LSCF calculation [14] (OLSCF), and to obtain Self-Consistent SLBO (SCSLBO). We consider the alanine molecule: its geometry has been optimized at the B3LYP/6-311++G** using pure-QM calculation. Then, we perform OLSCF calculations with one SLBO and one Strictly Localized Anti-Bonding Orbital (SLABO) localized on the central $C_\alpha-C_\beta$ bond of the molecule with the help of the PM criterion. Since this part is still under development, the OLSCF code is located in an other version of link 1502. OLSCF calculations required the total number of SLBO and SLABO (LL) defined in the OSLBO section. This information is defined with the IOP(5/199=LL0000). The dimension of the reduced OLSCF Fockien which is diagonalized during the calculation is also required, and defined by the IOp(5/197=K): K stand for the number of mixing orbitals during OLSCF calculations.

FIG. 21: Gaussian input file of an OLSCF calculation on the alanine molecule. The SLBO and the SLABO are located on the $C_\alpha-C_\beta$ bond and defined in the OSLBO section in the same way than the SLBOs.

```
%Subst 1502 /w30/QMMM/OLSCF
#p B3LYP/6-311++G** SP SCF=Tight
NoSym IOp(5/197=2,5/199=20000)

t

0 1
7      0      0.023637 -0.070923 -0.028478
6      0     -0.008514  0.030283  1.478353
6      0      1.461312 -0.021528  1.985213
8      0      1.643763  0.383096  3.152917
8      0      2.304561 -0.493058  1.180146
1      0      0.763824 -0.744363 -0.288095
6      0     -0.843285 -1.109326  2.056206
1      0     -0.450719  1.000355  1.720432
1      0     -1.877221 -1.062592  1.702778
1      0     -0.850645 -1.026413  3.143554
1      0     -0.422254 -2.082894  1.788232
1      0     -0.876050 -0.358807 -0.439359
1      0      0.285091  0.830078 -0.454870

OSLBO
11 2 7 8 ROTATE
 0.018789 -0.032561  0.013304
-0.702326  1.216481 -0.496637
 0.037048 -0.064137  1.106463
-0.03288 -0.06694 -0.07984 0.13828 -0.05645 0.23001 -0.11955 0.20706
-0.08454 0.13023 -0.07293 0.12632 -0.05157 0.01458 -0.00221 0.00383
-0.00156 -0.00700 0.00571 -0.00990 -0.00808 -0.01400
-0.03288 -0.06694 0.07984 -0.13828 0.05645 0.23001 0.11955 -0.20706
0.08454 0.13023 0.07293 -0.12632 0.05157 0.01458 0.00221 -0.00383
0.00156 -0.00700 0.00571 -0.00990 -0.00808 -0.01400
10 2 7 8 ROTATE
 0.018789 -0.032561  0.013304
-0.702326  1.216481 -0.496637
 0.037048 -0.064137  1.106463
 0.04788 0.15799 0.09548 -0.16523 0.06742 -0.49676 0.44498 -0.77105
0.31477 -1.66362 0.52452 -0.90844 0.37078 0.45808 -0.14150 0.24508
-0.10003 0.15847 -0.12962 0.22351 0.18291 0.31732
-0.04788 -0.15799 0.09541 -0.16523 0.06755 0.49676 0.44505 -0.77106
0.31467 1.66362 0.52440 -0.90842 0.37101 -0.45808 -0.14147 0.24507
-0.10008 -0.15852 0.12978 -0.22461 -0.18277 -0.31654
****
```

D. Molecular dynamic

When running QM/MM molecular dynamic jobs and/or geometry optimizations using periodic boundary conditions (PBC), Tinker is the leader program. MD trajectories are then performed with the `dynaqmmm` program and geometry optimization with the `minimize` program. `DIRECT` method to compute the electrostatic interaction is not allowed for MD of periodic or non-periodic systems: only `ESPF` method is working.

1. QM:MM MD job

QM/MM MD trajectory is introduced by the `QMMM-EXTERNAL` keyword in the Tinker `key` file (FIG. 22), as well as the input and output files (`g03input.com` and `g03output.log`, respectively). One must also give the number and the list of the QM atoms after the `QMMM` keyword.

FIG. 22: Tinker `key` file of a MD trajectory of the water dimer using the Velocity Verlet algorithm.

```
PARAMETERS amber99.prm
DIGITS 8
ARCHIVE
VERBOSE
RANDOMSEED 123456789
INTEGRATE VERLET
QMMM 3
QM 583
QM 584
QM 585
QMMM-EXTERNAL GAUSSIAN $LSCFMM/g03 < g03input.com > g03output.log
```

The Gaussian input file (`g03input.com`) for these kind of calculations is much more complicated than the one for static calculations, as gathered on Figure 23. It contains only the QM atoms.

FIG. 23: Gaussian input file of a MD trajectory of the water dimer. The electrostatic interaction is computed with the `ESPF` method.

```
%Subst 1101 $LSCFMM
%Subst 1202 $LSCFMM
%Subst 1301 $LSCFMM
%Subst 1397 $LSCFMM
%Subst 1701 $LSCFMM
%Subst 19999 $LSCFMM
#t RHF/ST0-3G SCF=Tight Force ExtraLink=397
NoSym IOp(2/198=1,3/198=-2,99/198=1)

QM/MM

O 1
O -.07504437 .00000000 .01334682
H .31769794 .00000000 .92114815
H .75638863 .00000000 -.52969298
```

The purpose of the QM calculation is to give the QM energy (`SCF=Tight`) and gradients (`Force`) to the Tinker program, which then performs the integration of the equations of motion. The QM/MM electrostatic interaction is computed with the `ESPF` method (`ExtraLink=397` and `IOp(3/198=-2)`). The options `IOp(2/198=1)` and `IOp(99/198=1)` allow to import the new QM coordinates after a MD step and export the QM energy and gradients to the Tinker software, respectively [120]. Then, the MD job is sent with the `dynaqmmm` program. The user must specify the desired number of dynamics steps, a time interval for the dynamics steps, and an interval between coordinate saving. At the

same time that a point along the trajectory is saved, the complete information needed to restart the trajectory from that point is updated and stored in the **dyn** file.

2. LSCF/MM MD job

Within the LSCF/MM method, the Y frontier atom must be added in the **key** file. On Figure 24, we present the Tinker **key** file for the ethane molecule.

FIG. 24: Tinker **key** file of a MD trajectory within the LSCF/MM method. The molecule considered here is the ethane molecule.

```
PARAMETERS oplsaa.prm
DIGITS 8
ARCHIVE
VERBOSE
RANDOMSEED 123456789
INTEGRATE VERLET
QMMM 5 1
QM 1
QM 2
QM 3
QM 4
YA 5
QMMM-EXTERNAL GAUSSIAN $LSCFMM/g03 < g03input.com > g03output.log
```

In the case of a LSCF/MM calculation, the Gaussian input file contains several additional information relative to the LSCF computation which should be placed in the **XY_POT** and the **SLBO** sections. Moreover, the atomic basis set must be defined in a **GEN** section to specify the atomic basis set of the Y atom (FIG. 25).

FIG. 25: Gaussian input file of a MD trajectory of the ethane molecule within the LSCF/MM method. The electrostatic interaction is computed with the ESPF method.

```
%SUBST L101 $LSCFMM
%SUBST L202 $LSCFMM
%SUBST L301 $LSCFMM
%SUBST L397 $LSCFMM
%SUBST L502 $LSCFMM
%SUBST L701 $LSCFMM
%SUBST L9999 $LSCFMM
%CHK=ethane
#t RHF GEN SCF=Tight Force ExtraLink=397
NoSym IOp(2/198=1,3/198=-2,3/199=10,99/198=1) LSCF=SLB0=1

TITLE HERE

O 1
C 0 .00000 .00000 .00000
H 0 .00000 .00000 1.08900
H 0 1.02672 .00000 -.36300
H 0 -.51336 -.88916 -.36300
H 0 -.68354 1.18392 -.48333

XY_POT
1 5 -26.3145 5.2319 -1.0393 -0.3589 20.1330
****
1 2 3 4 0
STO-3G
****
5 0
C 0
S 3 1.00
71.6168370 0.15432897
13.0450960 0.53532814
3.5305122 0.44463454
SP 3 1.00
2.9412494 -0.09996723 0.15591627
0.6834831 0.39951283 0.60768372
0.2222899 0.70011547 0.39195739
****

SLB0
11 1 5 2 ROTATE
0.020662 -0.035787 0.014610
-0.704198 1.219707 -0.497943
0.032606 -0.056475 1.100392
-0.0100366 0.3562863 -0.1995480 0.3456273 -0.1411018
-0.0100366 0.3562863 0.1995480 -0.3456273 0.1411018
****
```

3. HLA/MM MD job

As a LSCF/MM computation, the HLA must be specified in the Tinker **key** file (FIG. 26).

FIG. 26: Tinker key file of a MD trajectory within the LA scheme. The molecule considered is the ethane molecule.

```
PARAMETERS ./oplsaa.prm
DIGITS 8
VERBOSE
RANDOMSEED 123456789
INTEGRATE VERLET
ARCHIVE
ATOM 2999 1 HA "HLA" 1 1.008 0
CHARGE TERM NONE
QMMM 5
  QM 1
  QM 2
  QM 3
  QM 4
  LA 5
QMMM-EXTERNAL GAUSSIAN $LSCFMM/g03 < g03input.com > g03output.log
```

In the case of the LA scheme, the Gaussian input file contains the coordinates of the HLA.

FIG. 27: Gaussian input file of a MD trajectory of the ethane molecule within the LA scheme. The electrostatic interaction is computed with the ESPF method.

```
%SUBST L101 $LSCFMM
%SUBST L202 $LSCFMM
%SUBST L301 $LSCFMM
%SUBST L397 $LSCFMM
%SUBST L701 $LSCFMM
%SUBST L716 $LSCFMM
%SUBST L9999 $LSCFMM
%CHK=ethane
#t RHF/ST0-3G SCF=Tight Force ExtraLink=397
NoSym IOp(2/198=1,3/198=-2,99/198=1)

TITLE HERE

O 1
C 0 .00000 .00000 .00000
H 0 .00000 .00000 1.08900
H 0 1.02672 .00000 -.36300
H 0 -.51336 -.88916 -.36300
H 0 -.68354 1.18392 -.48333
```

The scaling factor c_{HLA} (EQ. (39)) is specified when running the `dynaqmmm` program, after the conventional required information. In the next section, we present an exhaustive list of examples.

V. TEST CALCULATIONS

All calculations have been tested on IBM RS6000 AIX 5.3:

A. MM

- `test0001` Geometry optimization of the ethane molecule using the OPLS-AA force field.
- `test0002` Geometry optimization of the Ace-Lys-NMe tripeptide using the CHARMM27 force field.

- **test0003** Geometry optimization of the 5'-CG-CG-CG-3' double-stranded DNA fragment using the AMBER99 force field. An continuum solvation model is used to represent the solvent.
- **test0004** MD trajectory (10000 steps of 0.1 fs at 298 K) of the water dimer using the TIP3P model (`amber99.prm`). Starting geometry: minimization at the MM level (threshold 0.01).
- **test0005** The system consists of 216 rigid TIP3P water molecules in a 18.643 Å periodic box, 9.0 Å shifted energy switch cutoffs for nonbonded interactions (geometry optimization using `minimize`)
- **test0006** The system consists of 216 rigid TIP3P water molecules in a 18.643 Å periodic box, 9.0 Å shifted energy switch cutoffs for nonbonded interactions (10000 dynamics steps of 1.0 fs each using the Velocity Verlet integrator). Starting geometry: minimization at the MM level (threshold 0.01).

B. QM

1. Localization

- **test0010** Geometry optimization of the water molecule at the HF/STO-3G level of theory and localization of the two σ O-H bonds with the WP and WP+MP criteria, and the whole occupied MOs with the help of the BF, MP and BF+PM criteria.
- **test0011** Geometry optimization of the water molecule at the HF/STO-3G level of theory and localization of the MOs using the BF criterion We have considered the localization (i) of the occupied MOs window, (ii) of the occupied and virtual MOs windows (separately), and (iii) of the occupied MOs windows (core and valence orbitals separately) and the virtual MOs windows.
- **test0012** Geometry optimization of the ethane molecule at the HF/6-31G*, HF/6-31+G* and HF/6-311G* levels of theory and localization of the σ C-C bond using the WP criterion at the same levels of theory.

2. LSCF

a. FO

- **test0100** Single-point on the optimized HF/6-31G* structure of the water molecule at the HF/6-31G* level of theory. The FOs are determined with the BF criterion and represent the two σ O-H bonds.
- **test0101** Unrestricted single-point on the optimized HF/6-31G* structure of the water molecule at the HF/6-31G* level of theory. FOs are determined with the BF criterion and represent the two σ O-H bonds. Cationic state (Charge +1, Multiplicity 2) of the water molecule. The first σ O-H bond is doubly occupied and the second one is singly occupied with one α -electron.
- **test0102** Core ionisation examples without core orbital (CO) relaxation Unrestricted single-point on the optimized B3LYP/6-311++G** structure of the propyne molecule at the PBE0/6-311++G** level of theory. FOs are determined with the BF criterion. Cationic state (Charge +1, Multiplicity 2) of the propyne molecule. The calculation contains two COs in the input file: one singly occupied with an α electron and the second one is empty (β -MOs space). The spatial part is the same for the two FOs. It corresponds to the core hole of the central carbon of the propyne molecule.
- **test0103** Core ionisation examples with core orbital (CO) relaxation Unrestricted single-point on the optimized B3LYP/6-311++G** structure of the propyne molecule at the PBE0/6-311++G** level of theory. FOs is determined with the BF criterion. Cationic state (Charge +1, Multiplicity 2) of the propyne molecule. The calculation contains one CO which is empty (β -MOs space). It corresponds to the core hole of the central carbon of the propyne molecule.

b. SLBO

- **test0110** Geometry optimization of the water molecule at the HF/6-31G* level of theory. The SLBO are determined according to the WP criterion (σ O-H bonds).
- **test0111** Geometry optimization of the water molecule at the HF/6-31G* level of theory. Anionic state of the water molecule (Charge -1, Multiplicity 2). The SLBO are determined according to the WP criterion (σ O-H bonds) and are doubly occupied (one α - and one β -electron).

c. *SCSLBO*

- **test0130** Single-point on the optimized B3LYP/6-311++G** structure of the alanine (zwitterionic form) at the B3LYP/6-311++G** level of the theory. The SLBO and the SLABO are determined at the same level of theory on the ethane molecule using the PM criterion. See Ref. [14] for more details.
- **test0131** Same as **test0130** with SCRF model (IEFPCM with UA0 radii, Water).
- **test0132** Unrestricted single-point on the optimized B3LYP/6-311++G** structure of the alanine (zwitterionic form) at the B3LYP/6-311++G** level of the theory. The SLBO and the SLABO are determined at the B3LYP/6-311++G** level of theory on the ethane molecule, using the PM criterion. Anionic state of the alanine molecule (Charge -1, Multiplicity 2).

C. QM/MM

1. QM:MM

- **test1000** Geometry optimization at the HF/6-31G*:TIP3P level of theory. Water dimer (1 QM + 1 MM). Microiterations and Hessian calculation disabled. Starting geometry: minimization at the MM level (threshold 0.001). Electrostatic interaction has been computed with the **DIRECT** method.
- **test1001** Geometry optimization at the HF/6-31G*:TIP3P level of theory. Water dimer (1 QM + 1 MM). Microiterations and Hessian calculation disabled. Starting geometry: minimization at the MM level (threshold 0.001). Electrostatic interaction has been computed with the **ESPF** method.
- **test1002** Geometry optimization at the HF/6-31G*:TIP3P level of theory. Water dimer (1 QM + 1 MM). Microiterations enabled with MK charges for the QM atoms (Hessian calculation enabled). Starting geometry: minimization at the MM level (threshold 0.001). Electrostatic interaction has been computed with the **DIRECT** method.
- **test1003** Geometry optimization at the HF/6-31G*:TIP3P level of theory. Water dimer (1 QM + 1 MM). Microiterations enabled with MK charges for the QM atoms (Hessian calculation enabled). Starting geometry: minimization at the MM level (threshold 0.001). Electrostatic interaction has been computed with the **ESPF** method.
- **test1004** Geometry optimization at the HF/6-31G*:TIP3P level of theory. Water cluster containing 27 molecules (1 QM + 26 MM). Microiterations enabled with CHELPG charges for the QM atoms (Hessian calculation disabled). Starting geometry: minimization at the MM level (threshold 0.001). Electrostatic interaction has been computed with the **DIRECT** method.
- **test1005** Geometry optimization at the HF/6-31G*:TIP3P level of theory. Water cluster containing 27 molecules (1 QM + 26 MM). Microiterations enabled with CHELPG charges for the QM atoms (Hessian calculation disabled). Starting geometry: minimization at the MM level (threshold 0.001). Electrostatic interaction has been computed with the **ESPF** method.
- **test1006** Geometry optimization at the HF/6-31G*:TIP3P level of theory. Water cluster containing 27 molecules (1 QM + 26 MM). Microiterations enabled with CHELPG charges for the QM atoms (Hessian calculation enabled). Starting geometry: minimization at the MM level (threshold 0.001). Electrostatic interaction has been computed with the **DIRECT** method.
- **test1007** Geometry optimization at the HF/6-31G*:TIP3P level of theory. Water cluster containing 27 molecules (1 QM + 26 MM). Microiterations enabled with CHELPG charges for the QM atoms (Hessian calculation enabled). Starting geometry: minimization at the MM level (threshold 0.001). Electrostatic interaction has been computed with the **ESPF** method.
- **test1008** Geometry optimization at the HF/6-31G*:TIP3P level of theory. Water cluster containing 27 molecules (1 QM + 26 MM). Microiterations and Hessian calculation disabled. Starting geometry: minimization at the MM level (threshold 0.001). The MM atoms are constraint to their initial positions by the Gaussian optimizer. The options **Opt=(Big,GDIIS)** have been add in the Gaussian input file to speed up the calculation.

- **test1009** Geometry optimization at the HF/6-31G*:TIP3P level of theory. Water cluster containing 27 molecules (1 QM + 26 MM). Microiterations disabled and Hessian calculation. Starting geometry: minimization at the MM level (threshold 0.001). The MM atoms are constraint to their initial positions. Their are not been added to the input Gaussian file.
- **test1010** Geometry optimization at the HF/6-31G*:TIP3P level of theory. Water cluster containing 27 molecules (1 QM + 26 MM). Microiterations and Hessian calculation disabled. Starting geometry: minimization at the MM level (threshold 0.001). The first five water molecules (1 QM + 4 MM) have been considered in the gaussian input file. The remaining part is frozen during the calculations. Concerning these five molecules, two are considered frozen by the gaussian optimizer, and the three others are fully optimized. The options `Opt=(Big,GDIIS)` have been add in the Gaussian input file to speed up the calculation.
- **test1011** Geometry optimization at the HF/6-31G*:TIP3P level of theory. Water cluster containing 27 molecules (1 QM + 26 MM). Microiterations and Hessian calculation enabled. Starting geometry: minimization at the MM level (threshold 0.001). The first five water molecules (1 QM + 4 MM) have been considered in the gaussian input file and fully optimized by the gaussian optimizer. The remaining part is micro-optimized by the Tinker minimizer (CHELP charges). The options `Opt=(Big,GDIIS)` have been add in the Gaussian input file to speed up the calculation.
- **test1012** Molecular dynamic trajectory (100 steps of 0.1 fs at 298 K) at the HF/STO-3G:TIP3P level of theory. Water dimer (1 QM + 1 MM). Starting geometry: optimized geometry at the HF/STO-3G:TIP3P level of theory.
- **test1013** Minimization of a cubic periodic box containing 216 water molecules (18.643 Å) at the HF/STO-3G:TIP3P level of theory (1 QM + 215 MM). Shifted energy switch cutoffs (9.0 Å) for nonbonded interactions.
- **test1014** Molecular Dynamic trajectory (NVT) of a cubic periodic box containing 216 water molecules (18.643 Å) at the HF/STO-3G:TIP3P level of theory (1 QM + 215 MM). (100 dynamics steps of 1.0 fs using the Velocity Verlet integrator) Shifted energy switch cutoffs (9.0 Å) for nonbonded interactions. Starting geometry: minimization at the same level of theory (threshold 0.01).

2. LSCF/MM

- **test2000** Geometry optimization of the Ace-Lys-NMe tripeptide at the HF/6-31G* level of theory using the CHARMM27 force field. The boundary is located between the C_α and C_β atoms of the lysine residue. The methodology described in Ref. [10] is used. The SLBO is determined on the ethane molecule with the WP criterion at the same level of theory. Values of Ref. [10] have been used for the FBP. Starting geometry: minimization at the MM level (threshold 0.001).
- **test2001** Geometry optimization of the Ace-Lys-NMe tripeptide at the HF/STO-3G level of theory using the CHARMM27 force field. The boundary is located between the C_α and C_β atoms of the lysine residue. The methodology described in Ref. [22] is used. The SLBO is determined on the ethane molecule with the WP criterion at the same level of theory. Starting geometry: minimization at the MM level (threshold 0.001).
- **test2002** Geometry optimization of the Ace-Lys-NMe tripeptide at the HF/6-311G** level of theory using the CHARMM27 force field. The boundary is located on the two peptidic bonds. The methodology described in Ref. [23] is used. The SLBO is determined on the methylacetamide molecule with the WP criterion at the same level of theory. Starting geometry: minimization at the MM level (threshold 0.01).
- **test2003** Geometry optimization of the Ace-Trp-NMe tripeptide at the B3LYP/6-31G* level of theory using the CHARMM27 force field. The boundary is located the C_α and C_β atoms of the tryptophan residue. The methodology described in Ref. [10] is used. The SLBO is determined on the ethane molecule with the WP criterion at the HF/6-31G* level of theory. Starting geometry: minimization at the MM level (threshold 0.01).
- **test2004** TD-PBE0/6-31+G* calculation on the Ace-Trp-NMe tripeptide (B3LYP/6-31G* geometry) using the CHARMM27 force field. The boundary is located the C_α and C_β atoms of the tryptophan residue. The methodology described in Ref. [10] is used. The SLBO is determined on the ethane molecule with the WP criterion at the HF/6-31+G* level of theory.

- **test2005** Geometry optimization of the Ace-Lys-NMe tripeptide at the HF/6-31G* level of theory using the CHARMM27 force field. The boundary is located between the C_α and C_β atoms of the lysine residue. The methodology described in Ref. [10] is used. The SLBO is determined on the ethane molecule with the WP criterion with the same level of theory. Starting geometry: minimization at the MM level (threshold 0.01). A SCRF method is used to model the aqueous solution (IEFPCM with UA0 radii).
- **test2006** Single-point calculation on the Ace-Trp-NMe tripeptide (B3LYP/6-31G* geometry) at the MP2/6-311G* level of theory using the CHARMM27 force field. The boundary is located the C_α and C_β atoms of the tryptophan residue. The methodology described in Ref. [10] is used. The SLBO is determined on the ethane molecule with the WP criterion at the HF/6-311G* level of theory.
- **test2007** Geometry optimization of the Ace-Lys-NMe tripeptide at the HF/6-31G* level of theory using the CHARMM27 force field. The boundary is located between the C_α and C_β atoms of the lysine residue. The methodology described in Ref. [10] is used. The SLBO is determined on the ethane molecule with the WP criterion at the same level of theory. Values of Ref. [10] have been used for the FBP. Starting geometry: minimization at the MM level (threshold 0.001). Electrostatic interaction has been computed with the ESPF method [117].
- **test2008** Geometry optimization of the Thioredoxin protein (1TOF.pdb) at the HF level of theory using a mixed basis set (S=6-311+G** and C,H=6-311G*) and the CHARMM27 force field to represent the MM part. The methodology described in Ref. [62] is used. The SCCO scheme has been used for the QM/MM frontier [22]. The SLBO is determined on the ethane molecule with the WP criterion at the HF/6-311G* level of theory. Electrostatic interaction has been computed with the ESPF method. Starting geometry: first structure of the 1TOF.pdb file and minimization at the MM level (threshold 0.01).
- **test2009** Molecular dynamic (100 steps of 0.5 fs at 298 K) of the Thioredoxin protein (anionic state) at the HF level of theory using a mixed basis set (S=6-311+G** and C,H=6-311G*) and the CHARMM27 force field to represent the MM part. The methodology described in Ref. [62] is used. The SCCO scheme has been used for the QM/MM frontier [22]. The SLBO is determined on the ethane molecule with the WP criterion at the HF/6-311G* level of theory. Electrostatic interaction has been computed with the ESPF method Starting geometry: optimized structure of the neutral form at the same level of theory.
- **test2010** Geometry optimization of the Thioredoxin protein (1TOF.pdb) at the HF level of theory using a mixed basis set (S=6-311+G** and C,H=6-311G*) and the CHARMM27 force field to represent the MM part. The methodology described in Ref. [62] is used. The SCCO scheme has been used for the QM/MM frontier [22]. The SLBO is determined on the ethane molecule with the WP criterion at the HF/6-311G* level of theory. Anionic state of the Thioredoxin protein (Charge -1, Multiplicity 2). Electrostatic interaction has been computed with the ESPF method. Starting geometry: first structure of the 1TOF.pdb file and minimization at the MM level (threshold 0.01).

3. HLA/MM

- **test3000** Geometry optimization of the ethane molecule at the HF/STO-3G level of theory using the LA scheme. The MM surrounding has been modelled with the OPLS-AA force field. The MM charges have been set to zero to avoid overpolarization of the QM wavefunction. Electrostatic interaction has been computed with the DIRECT method.
- **test3001** Geometry optimization of the ethane molecule at the HF/STO-3G level of theory using the LA scheme. The MM surrounding has been modelled with the OPLS-AA force field. The MM charges have been set to zero to avoid overpolarization of the QM wavefunction. Electrostatic interaction has been computed with the ESPF method.
- **test3002** Geometry optimization of the Thioredoxin protein (1TOF.pdb) at the MP2 level of theory using a mixed basis set (S=6-31+G** and C,H=6-31G*) and the CHARMM27 force field to represent the MM part. The methodology described in Ref. [62] is used. Electrostatic interaction has been computed with the DIRECT method and the MM charges of the C_1 atoms have been redistributed on the C_2 atoms. See Ref. [62] for more details. Starting geometry: first structure of the 1TOF.pdb file and minimization at the MM level (threshold 0.01).

- **test3003** Geometry optimization of the Thioredoxin protein (1TOF.pdb) at the MP2 level of theory using a mixed basis set (S=6-31+G** and C,H=6-31G*) and the CHARMM27 force field to represent the MM part. The methodology described in Ref. [62] is used. Electrostatic interaction has been computed with the ESPF method and the MM charges of the C₁ atoms have been redistributed on the C₂ atoms. See Ref. [62] for more details. Starting geometry: first structure of the 1TOF.pdb file and minimization at the MM level (threshold 0.01).
- **test3004** Geometry optimization of the Thioredoxin protein (1TOF.pdb) at the MP2 level of theory using a mixed basis set (S=6-31+G** and C,H=6-31G*) and the CHARMM27 force field to represent the MM part. The methodology described in Ref. [62] is used. Electrostatic interaction has been computed with the DIRECT method and the MM charges of the C₁ atoms have been redistributed on the C₂ atoms. See Ref. [62] for more details. Microiterations enabled with MK charges for the QM atoms. Starting geometry: first structure of the 1TOF.pdb file and minimization at the MM level (threshold 0.01).
- **test3005** Geometry optimization of the Thioredoxin protein (1TOF.pdb) at the MP2 level of theory using a mixed basis set (S=6-31+G** and C,H=6-31G*) and the CHARMM27 force field to represent the MM part. The methodology described in Ref. [62] is used. Electrostatic interaction has been computed with the ESPF method and the MM charges of the C₁ atoms have been redistributed on the C₂ atoms. See Ref. [62] for more details. Microiterations enabled with MK charges for the QM atoms. Starting geometry: first structure of the 1TOF.pdb file and minimization at the MM level (threshold 0.01).
- **test3006** Geometry optimization of the Thioredoxin protein (1TOF.pdb) at the HF level of theory using a mixed basis set (S=6-31+G** and C,H=6-31G*) and the CHARMM27 force field to represent the MM part. The methodology described in Ref. [62] is used. Electrostatic interaction has been computed with the ESPF method and the MM charges of the C₁ atoms have been redistributed on the C₂ atoms. See Ref. [62] for more details. Starting geometry: first structure of the 1TOF.pdb file and minimization at the MM level (threshold 0.01).
- **test3007** Molecular dynamic (100 steps of 1.0 fs at 298 K using the Velocity Verlet integrator) of the Thioredoxin protein (anionic state) at the HF level of theory using a mixed basis set (S=6-31+G** and C,H=6-31G*) and the CHARMM27 force field to represent the MM part. The methodology described in Ref. [62] is used. Electrostatic interaction has been computed with the ESPF method and the MM charges of the C₁ atoms have been redistributed on the C₂ atoms. See Ref. [62] for more details. Starting geometry: optimized structure of the neutral form at the same level of theory.

Acknowledgments

The authors are grateful to Professor Jean-Louis Rivail for the seminal idea of the LSCF method leading to these developments.

-
- [1] A. Warshell and M. Levitt, *J. Mol. Biol.* **103**, 227 (1976).
 - [2] G. F. Smits and C. Altona, *Theor. Chem. Acc.* **67**, 461 (1985).
 - [3] U. C. Singh and P. A. Kollman, *J. Comput. Chem.* **7**, 718 (1986).
 - [4] M. J. Field, P. A. Bash, and M. Karplus, *J. Comput. Chem.* **11**, 700 (1990).
 - [5] J. Gao, *Methods and Applications of Combined Quantum Mechanical and Molecular Mechanical Potentials* (1967), vol. 7 of *Reviews in Computational Chemistry*, p. 119.
 - [6] N. Reuter, A. Dejaegere, B. Maigret, and M. Karplus, *J. Phys. Chem. A* **104**, 1720 (2000).
 - [7] H. Lin and D. Truhlar, *Theor. Chem. Acc.* **117**, 185 (2007).
 - [8] N. Ferré, A. Cembran, M. Garavelli, and M. Olivucci, *Theor. Chem. Acc.* **112**, 335 (2004).
 - [9] X. Assfeld and J.-L. Rivail, *Chem. Phys. Lett.* **263**, 100 (1996).
 - [10] N. Ferré, X. Assfeld, and J.-L. Rivail, *J. Comput. Chem.* **23**, 610 (2002).
 - [11] N. Ferré and X. Assfeld, *J. Chem. Phys.* **117**, 4119 (2002).
 - [12] A. Fornili, Y. Moreau, M. Sironi, and X. Assfeld, *J. Comput. Chem.* **27**, 515 (2006).
 - [13] P.-F. Loos and X. Assfeld, *Int. J. Quantum Chem.* **107**, 2243 (2007).
 - [14] P.-F. Loos and X. Assfeld, *J. Chem. Theor. Comput.* **3**, 1047 (2007).
 - [15] N. Ferré and X. Assfeld, *J. Mol. Struct. (THEOCHEM)* **632**, 83 (2003).
 - [16] X. Assfeld, N. Ferré, and J.-L. Rivail, *Theor. Chem. Acc.* **111**, 328 (2004).

- [17] Y. Moreau, P.-F. Loos, and X. Assfeld, *Theor. Chem. Acc.* **112**, 228 (2004).
- [18] Y. Moreau and X. Assfeld, *Lecture Series on Computer and Computational Sciences* **3**, 1 (2005).
- [19] A. Fornili, P.-F. Loos, M. Sironi, and X. Assfeld, *Chem. Phys. Lett.* **427**, 236 (2006).
- [20] P.-F. Loos, X. Assfeld, and J.-L. Rivail, *Theor. Chem. Acc.* **118**, 165 (2007).
- [21] P.-F. Loos, J. Preat, A. D. Laurent, C. Michaux, D. Jacquemin, E. A. Perpète, and X. Assfeld, *J. Chem. Theor. Comput.* **4**, 637 (2008).
- [22] P.-F. Loos, A. Fornili, M. Sironi, and X. Assfeld, *Comp. Lett.* **4**, 473 (2007).
- [23] P.-F. Loos and X. Assfeld, *AIP Conf. Proc.* **963**, 308 (2007).
- [24] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, T. V. Jr., K. N. Kudin, J. C. Burant, et al., *Gaussian 03, Revision B.05*, Gaussian, Inc., Wallingford, CT, 2004.
- [25] P. O. Lwdin, *Phys. Rev.* **97**, 1474 (1955).
- [26] J.-L. Rivail, *Eléments de Chimie Quantique l'Usage des Chimistes* (EDP Sciences/CNRS Editions, 1999).
- [27] S. Huzinaga, *J. Chem. Phys.* **55**, 5543 (1971).
- [28] A. Szabo and N. Ostlund, *Modern Quantum Chemistry : Introduction to Advanced Structure Theory* (Dover publications Inc., Mineola, New-York, 1989).
- [29] N. Ferré, Ph.D. thesis, Laboratoire de Chimie Thorique, UMR CNRS-UHP 7565, Universit Henri Poincar, Nancy I (2001).
- [30] P. Pulay, *Mol. Phys.* **17**, 197 (1969).
- [31] P. Pulay, *Analytical Derivative Methods in Quantum Chemistry* (Plenum, New-York, 1977), p. 153, Applications of Electronic Structure Theory.
- [32] P. Pulay, *Analytical Derivative Methods in Quantum Chemistry* (John Wiley & Sons, 1987), vol. 69 of *Advances in Chemical Physics: ab initio Methods in Quantum Chemistry (Part II)*, p. 241.
- [33] Y. Yamaguchi, O. Osamura, J. D. Goddard, and H. F. S. III, *A new Dimension to Quantum Mechanics. Analytical Derivative Methods in ab initio Molecular Electronic Structure Theory* (Oxford University Press, Oxford, 1994).
- [34] P. Pulay, *J. Chem. Phys.* **78**, 5043 (1983).
- [35] D. M. Philipp and R. A. Friesner, *J. Comput. Chem.* **20**, 1468 (1999).
- [36] R. B. Murphy, D. M. Philipp, and R. A. Friesner, *J. Comput. Chem.* **21**, 1442 (2000).
- [37] F. Maseras and K. Morokuma, *J. Comput. Chem.* **16**, 1170 (1995).
- [38] M. Svensson, S. Humbel, D. J. Froese, T. Matsubara, S. Sieber, and K. Morokuma, *J. Phys. Chem.* **100**, 19357 (1996).
- [39] S. Ranganathan and J. Gready, *J. Phys. Chem. B* **101**, 5614 (1997).
- [40] S. Dapprich, I. Komárino, K. S. Byun, K. Morokuma, and M. J. Frisch, *J. Mol. Struct. (THEOCHEM)* **461**, 1 (1999).
- [41] P. Cummins and J. Gready, *J. Phys. Chem. B* **104**, 4503 (2000).
- [42] T. Vreven, B. Mennucci, C. O. da Silva, K. Morokuma, and J. Tomasi, *J. Chem. Phys.* **115**, 62 (2001).
- [43] T. Kerdcharoen and K. Morokuma, *Chem. Phys. Lett.* **355**, 257 (2002).
- [44] D. Das, K. Eurenus, E. Billings, P. Sherwood, D. Chatfield, M. Hodoseek, and B. Brooks, *J. Chem. Phys.* **117**, 10534 (2002).
- [45] S. Humbel, S. Sieber, and K. Morokuma, *J. Chem. Phys.* **105**, 1959 (1996).
- [46] M. Svensson, S. Humbel, and K. Morokuma, *J. Chem. Phys.* **105**, 3664 (1996).
- [47] T. Matsubara, F. Maseras, N. Koga, and K. Morokuma, *J. Phys. Chem.* **100**, 2573 (1996).
- [48] R. D. Froese, D. G. Musaev, and K. Morokuma, *J. Am. Chem. Soc.* **120**, 1581 (1998).
- [49] T. Vreven and K. Morokuma, *J. Comput. Chem.* **21**, 1419 (2000).
- [50] T. Vreven and K. Morokuma, *Theor. Chem. Acc.* **109**, 125 (2003).
- [51] T. V. abnd K. Morokuma, O. Farkas, H. B. Schlegel, and M. J. Frisch, *J. Comput. Chem.* **24**, 760 (2003).
- [52] T. Vreven, K. S. Byun, I. Komáromi, S. Dapprich, J. A. M. Jr., K. Morokuma, and M. J. Frisch, *J. Chem. Theor. Comput.* **117**, 815 (2006).
- [53] N. Ferré and M. Olivucci, *J. Mol. Struct. (THEOCHEM)* **632**, 71 (2003).
- [54] N. Ferré and M. Olivucci, *J. Am. Chem. Soc.* **125**, 6868 (2003).
- [55] T. Andruniow, N. Ferré, and M. Olivucci, *Proc. Nati. Acad. Sci. USA* **101**, 17908 (2004).
- [56] A. Cembran, F. Bernardi, M. Olivucci, and M. Garavelli, *Proc. Nati. Acad. Sci. USA* **102**, 6255 (2005).
- [57] P. B. Coto, A. Sinicropi, N. Ferré, and M. Olivucci, *Proc. Nati. Acad. Sci. USA* **103**, 17154 (2006).
- [58] L. M. Frutos, T. Andruniow, F. Santoro, N. Ferré, and M. Olivucci, *Proc. Nati. Acad. Sci. USA* **104**, 7764 (2007).
- [59] A. Strambi, P. B. Coto, N. Ferré, and M. Olivucci, *Theor. Chem. Acc.* **118**, 185 (2007).
- [60] A. Sinicropi, T. Andruniow, N. Ferré, R. Basoci, and M. Olivucci, *J. Am. Chem. Soc.* **127**, 11534 (2005).
- [61] A. Strambi, P. B. Coto, L. M. Frutos, N. Ferré, and M. Olivucci, *Theor. Chem. Acc.* **130**, 3382 (2008).
- [62] E. E. Dumont, P.-F. Loos, A. D. Laurent, and X. Assfeld, *J. Chem. Theor. Comput.* **in press** (2008).
- [63] I. Antes and W. Thiel, *J. Phys. Chem. A* **103**, 9290 (1999).
- [64] P. Amara and M. J. Field, *Theor. Chem. Acc.* **109**, 43 (2003).
- [65] D. Bakowies and W. Thiel, *J. Phys. Chem.* **100**, 10580 (1996).
- [66] A. de Vries, P. Sherwood, S. Collins, A. Rigby, M. Rigutto, and G. Kramer, *J. Phys. Chem. B* **103**, 6133 (1999).
- [67] H. Lin and D. Truhlar, *J. Phys. Chem. A* **109**, 3991 (2005).
- [68] G. Karlström, R. Lindh, P.-A. Malmqvist, B. O. Roos, U. Ryde, V. Veryazov, P.-O. Widmark, M. Cossi, B. Schimelpennig, P. Neogady, et al., *Computational Material Science* **28**, 222 (2003).
- [69] H. Weinstein and R. Pauncz, *Symp. Faraday Soc.* **2**, 23 (1968).
- [70] H. Weinstein, R. Pauncz, and M. Cohen, *Adv. Atomic. Molec. Phys.* **7**, 97 (1971).
- [71] R. S. Mulliken, *J. Chem. Phys.* **23**, 1833 (1955).

- [72] V. Magnasco and A. Perico, *J. Chem. Phys.* **47**, 971 (1967).
- [73] V. Magnasco and A. Perico, *J. Chem. Phys.* **48**, 800 (1967).
- [74] S. F. Boys, *Rev. Mod. Phys.* **32**, 296 (1960).
- [75] J. M. Foster and S. F. Boys, *Rev. Mod. Phys.* **32**, 300 (1960).
- [76] S. F. Boys (Academic, New-York, 1966), p. 253, *Quantum Theory of Atoms, Molecules and the Solid State*.
- [77] J. Pipek, *Int. J. Quantum Chem.* **36**, 487 (1989).
- [78] J. Pipek and P. Mezey, *J. Chem. Phys.* **90**, 4916 (1989).
- [79] P. Kollman, R. Dixon, W. Cornell, T. Fox, C. Chipot, and A. Pohorille, *The Development/Application of a 'Minimalist' Organic/Biochemical Molecular Mechanic Force Field using a Combination of ab Initio Calculations and Experimental Data* (1997), vol. 3 of *Computer Simulation of Biomolecular Systems*, p. 83.
- [80] J. Aqvist, *J. Phys. Chem.* **94**, 8021 (1990).
- [81] W. S. Ross and C. C. Hardin, *J. Am. Chem. Soc.* **116**, 6070 (1994).
- [82] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. M. Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, *J. Am. Chem. Soc.* **117**, 5179 (1995).
- [83] T. E. C. III, P. Cieplak, and P. A. Kollman, *J. Biomol. Struct. Dyn.* **16**, 845 (1999).
- [84] G. Moyna, H. J. Williams, R. J. Nachman, and A. I. Scott, *Biopolymers* **49**, 403 (1999).
- [85] J. Wang, P. Cieplak, and P. A. Kollman, *J. Comput. Chem.* **21**, 1049 (2000).
- [86] A. D. M. Jr., C. L. B. III, L. Nilsson, B. Roux, Y. Won, and M. Karplus, *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program* (John Wiley & Sons: Chichester, 1998), vol. 1 of *The Encyclopedia of Computational Chemistry*, p. 271.
- [87] N. Foloppe and A. D. M. Jr., *J. Comput. Chem.* **21**, 86 (2000).
- [88] E. Neria, S. Fischer, and M. Karplus, *J. Chem. Phys.* **105**, 1902 (1996).
- [89] L. Nilsson and M. Karplus, *J. Comput. Chem.* **7**, 591 (1986).
- [90] B. R. Brooks, R. E. Bruccoleri, D. J. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.* **4**, 187 (1983).
- [91] A. D. M. Jr., D. Bashford, M. Bellott, R. L. D. Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, et al., *J. Phys. Chem. B* **102**, 3586 (1998).
- [92] M. T. T. N. L. Allinger, M. A. Miller, and D. H. Wertz, *J. Am. Chem. Soc.* **93**, 1637 (1971).
- [93] N. L. Allinger, *J. Am. Chem. Soc.* **99**, 3279 (1977).
- [94] N. L. Allinger, Y. H. Yuh, and J.-H. Lii, *J. Am. Chem. Soc.* **111**, 8551 (1989).
- [95] N. L. Allinger, H. J. Geise, W. Pyckhout, L. A. Paquette, and J. C. Gallucci, *J. Am. Chem. Soc.* **111**, 1106 (1989).
- [96] J.-H. Lii and N. L. Allinger, *J. Am. Chem. Soc.* **111**, 8567 (1989).
- [97] J.-H. Lii and N. L. Allinger, *J. Am. Chem. Soc.* **111**, 8576 (1989).
- [98] J.-H. Lii and N. L. Allinger, *J. Phys. Org. Chem.* **7**, 591 (1994).
- [99] N. L. Allinger, F. Li, and L. Yan, *J. Comput. Chem.* **11**, 848 (1990).
- [100] N. L. Allinger, F. Li, L. Yan, and J. C. Tai, *J. Comput. Chem.* **11**, 868 (1990).
- [101] J.-H. Lii and N. L. Allinger, *J. Comput. Chem.* **19**, 1001 (1998).
- [102] N. L. Allinger, K. Chen, and J.-H. Lii, *J. Comput. Chem.* **14**, 642 (1996).
- [103] N. Nevins, K. Chen, and N. L. Allinger, *J. Comput. Chem.* **14**, 669 (1996).
- [104] N. Nevins, K. Chen, and N. L. Allinger, *J. Comput. Chem.* **14**, 695 (1996).
- [105] N. L. Allinger, K. Chen, J. A. Katzeellenbogen, S. R. Wilson, and G. M. Anstead, *J. Comput. Chem.* **14**, 747 (1996).
- [106] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, *J. Am. Chem. Soc.* **117**, 11225 (1996).
- [107] D. S. Maxwell, J. Tirado-Rives, and W. L. Jorgensen, *J. Comput. Chem.* **16**, 984 (1995).
- [108] W. L. Jorgensen and N. A. McDonald, *J. Mol. Struct. (THEOCHEM)* **424**, 145 (1998).
- [109] N. A. McDonald and W. L. Jorgensen, *J. Phys. Chem. B* **102**, 8049 (1998).
- [110] R. C. Rizzo and W. L. Jorgensen, *J. Am. Chem. Soc.* **121**, 4827 (1999).
- [111] M. L. P. Price, D. Ostrovsky, and W. L. Jorgensen, *J. Comput. Chem.* **22**, 1340 (2001).
- [112] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, *J. Phys. Chem. B* **105**, 6474 (2001).
- [113] J. W. Ponder, *Tinker, version 4.2*, washington University: St. Louis, MO, 2004.
- [114] G. Schaftenaar and J. H. Noordik, *J. Comput.-Aided Mol. Design* **123**, 14 (2000).
- [115] C. I. Bayly, W. D. Cornell, and P. A. Kollman, *J. Phys. Chem.* **97**, 10269 (1993).
- [116] W. D. Cornell, P. Cieplak, C. I. Bayly, and P. A. Kollman, *J. Am. Chem. Soc.* **115**, 9620 (1993).
- [117] N. Ferré and J. Angyán, *Chem. Phys. Lett.* **415**, 100 (2002).
- [118] CHARGETERM NONE is also widely used to set to zero the overall MM point charges.
- [119] To avoid the renumbering of the atoms, the -A option from Molden is required.
- [120] The Guess=Read keyword could be used to speed up the calculation.