

Calculation of Wave-functions with Frozen Orbitals in Mixed Quantum Mechanics/Molecular Mechanics methods. II. Application of the Local Basis Equation.

György G. Ferenczy
MTA-SE Molecular Biophysics Research Group
Semmelweis University Budapest
1094 Budapest, Tűzoltó u. 37-47, Hungary

November 8, 2012

Abstract

The application of the local basis equation¹ in mixed quantum mechanics/molecular mechanics (QM/MM) and quantum mechanics/quantum mechanics (QM/QM) methods is investigated. This equation is suitable to derive local basis nonorthogonal orbitals that minimize the energy of the system and it exhibits good convergence properties in a self-consistent field solution. These features make the equation appropriate to be used in mixed QM/MM and QM/QM methods to optimize orbitals in the field of frozen localized orbitals connecting the subsystems. Calculations performed for several properties in diverse systems show that the method is robust with various choices of the frozen orbitals and frontier atom properties. With appropriate basis set assignment it gives results equivalent with those of a related approach² using the Huzinaga equation. Thus the local basis equation can be used in mixed QM/MM methods with small size quantum subsystems to calculate properties in good agreement with reference Hartree-Fock-Roothaan results. It is shown that bond charges are not necessary when the local basis equation is applied, although they are required for the self-consistent field solution of the Huzinaga equation based method. On the other hand, the deformation of the wave-function near to the boundary is observed without bond charges and this has a significant effect on deprotonation energies but a less pronounced effect when the total charge of the system is conserved. The local basis equation can also be used to define a two layer quantum system with nonorthogonal localized orbitals surrounding the central delocalized quantum subsystem.

Keywords: mixed QM/MM, QM/QM, frozen localized orbitals, local basis equation, nonorthogonal orbitals ■

INTRODUCTION

Mixed QM/QM and QM/MM methods allow a potentially accurate treatment of extended systems with an affordable computational work. The basic idea is to perform higher quality calculations for a central part of the system where a chemical or physical event takes place, while a more approximate method is applied to the environment whose effect onto the central part is taken into account. Methods to separate the system into subsystems are discussed in ref.². Here we note that the separation requires special considerations when the two subsystems are covalently bound. Covalently bound subsystems are typical in calculations for systems built from chemically bound monomers like proteins, DNA, zeolites,... A possible way of separation is to assign localized or strictly localized orbitals to bonds at the boundary of the subsystems³⁻¹⁰. Their advantage is that they allow to keep apart the electrons of the two subsystems thus making possible to treat them at different levels of approximations. Atoms participating in a (strictly) localized orbital of a subsystem may have other bonds belonging to the other subsystem. Owing to the artificial environment of the localized orbitals at the boundary, typically they are not optimized, rather their coefficients are taken from model systems and are kept frozen when other orbitals are optimized.

The usual approach to the optimization of orbitals in the field of frozen orbitals includes the explicit orthogonalization of the basis functions to the frozen orbitals. This is necessary in computational schemes that, on one hand, assume orthogonality among the orbitals, but, on the other hand, do not guaranty orthogonality to frozen orbitals not included in the optimization.

In a previous paper² it was shown that the Huzinaga equation¹¹, with an appropriate basis set assignment, makes it possible to optimize orbitals of a central subsystem in the field of frozen localized orbitals at the boundary of the subsystems. The application of the Huzinaga equation results in orbitals orthogonal to the frozen orbitals without basis set orthogonalization. In the present contribution an alternative to the Huzinaga equation, the local basis

equation¹ and its application to optimize the orbitals is investigated. While the Huzinaga equation keeps the frozen and optimized orbitals orthogonal and this has to be respected in the basis set assignment, the local basis equation allows nonorthogonality and thus its application allows more flexibility in the basis set assignment. Moreover, while the self-consistent solution of the Huzinaga equation is possible only with an appropriately chosen system Hamiltonian and frozen orbitals, the application of the local basis equation is more flexible in this respect, as well (see later). The local basis equation is not only suitable to optimize orbitals in the field of frozen orbitals at the boundary, but it also allows to define a layer of *a priori* localized orbitals with local basis sets in mixed QM/QM or QM/QM/MM schemes.

The paper is organized as follows. First, the main features of the local basis equation are recapitulated and are compared to related schemes including the Huzinaga equation. Then calculations are presented in which the local basis equation is used to optimize orbitals in the field of frozen orbitals and of surrounding atoms represented by point charges. Other calculations apply two QM regions, one with full QM treatment, and another with *a priori* localized orbitals. Calculated properties as obtained with various approximate schemes are compared to reference results.

THE LOCAL BASIS EQUATION

The starting point for the derivation of the local basis equation is the separation of the electrons into groups and to optimize the orbitals of a selected group in the field of other groups that are kept frozen. A principal feature of the equation is that optimized orbitals are not required to be orthogonal to the frozen orbitals. This makes possible to use group specific (local) basis sets. The local basis equation was derived in ref.¹ and is written in a slightly modified notation as

$$(\mathbf{I}_a - \mathbf{S}_a \mathbf{R}^{\bar{a}}) \mathbf{F} (\mathbf{I}_a - \mathbf{R}^{\bar{a}} \mathbf{S}_a) \mathbf{C}_{aA} = (\mathbf{I}_a - \mathbf{S}_a \mathbf{R}^{\bar{a}}) \mathbf{S}_a \mathbf{C}_{aA} \mathbf{E}_{AA}. \quad (1)$$

where \mathbf{I} is a unit matrix, \mathbf{S} is the basis overlap matrix, $\mathbf{R}^{\bar{a}}$ projects to all orbitals not in group a , \mathbf{C} is the coefficient matrix and \mathbf{E} is the diagonal matrix of eigenvalues. Lower indices refer to the dimensions of the matrices. The first index specifies the number of rows and the second index specifies the number of columns. 'a' and 'A' as indices refer to dimensions equal to the number of basis functions and to the number of occupied molecular orbitals in group a , respectively. A dot or a missing lower index indicate full dimension (e.g. all basis functions) of a matrix. Owing to the possible nonorthogonality of the orbitals $\mathbf{R}^{\bar{a}} = \mathbf{C}^{\bar{a}}(\sigma^{\bar{a}})^{-1}(\mathbf{C}^{\bar{a}})^{\dagger}$ with $\sigma^{\bar{a}}$ being the overlap matrix of molecular orbitals. Both $\sigma^{\bar{a}}$ and $\mathbf{C}^{\bar{a}}$ refers to all orbitals not in group a .

Equations to calculate nonorthogonal orbitals with local basis sets have been derived previously. Stoll *et al.*¹² presented an equation for the orbitals that make the energy stationary without requiring the orthogonality of the orbitals. They also derived an eigenvalue equation to determine these orbitals. Their eigenvalue equation in the present notation reads as

$$(\mathbf{I}_a - \mathbf{S}_a \mathbf{R} + \mathbf{S}_{aa} \tilde{\mathbf{R}}_a^{a\dagger}) \mathbf{F} (\mathbf{I}_a - \mathbf{R} \mathbf{S}_a + \tilde{\mathbf{R}}_a^a \mathbf{S}_{aa}) \mathbf{C}_{aA} = \mathbf{S}_{aa} \mathbf{C}_{aA} \mathbf{E}_{AA}. \quad (2)$$

where

$$\tilde{\mathbf{R}}_a^a = \mathbf{C}_{\sigma^{\bar{a}}}^{-1} \mathbf{C}_{Aa}^{\dagger} \quad (3)$$

and thus it is different from the local basis equation (1). Note that \mathbf{C} and \mathbf{E} in Eq. (2) are in general different from those in Eq. (1). Nevertheless, the solutions of both equations span the same space and satisfy the equation of stationary energy^{1,12}.

It was reported that the iterative solution of Eq. (2) is impractical owing to its extremely slow convergence¹². A proposed reason for the slow convergence is that the matrix to be diagonalized according to Eq. (2) contains the orbital coefficients up to six order. The local basis equation, Eq. (1), on the other hand, includes the orbital coefficients to be optimized only in the Fock matrix and thus up to second order similarly to the standard Hartree-Fock-Roothaan equation. Indeed, we have not observed convergence difficulties in the solution of the local

basis equation.

A set of eigenvalue equations equivalent to Eq. (1) was derived¹³ for subsystems with mutually exclusive basis sets to exclude BSSE from molecular interaction calculations. Later the equations were modified to allow group of electrons with shared basis functions^{14,15}. These latter equations are similar to Eq. (2) in that they contain the coefficients to be optimized up to six order, and convergence difficulties were reported. The method was applied^{15,16} to derive strictly localized orbitals to use as frozen orbitals in the LSCF method^{4,5,17}. It was also used¹⁵ in a QM/MM type calculation, similar to those presented in the current contribution, but instead of a self-consistent field iteration, a computationally less efficient gradient optimization was applied to calculate the orbitals. The use of first, and occasionally second, derivatives with respect to the coefficients to calculate nonorthogonal orbitals was also proposed in refs.^{12,18,19}. Such schemes aim at overcoming self-consistent field convergence difficulties. Owing to the advantageous convergence properties of Eq. (1) the self-consistent field solution was exclusively applied in the present contribution.

Equations for nonorthogonal group functions that do not minimize the energy of the system have also been proposed²⁰⁻²². These equations offer computationally economical approximations to Eq. (1).

The local basis equation results in nonorthogonal orbitals in general. However, the total wave-function may be equivalent with that of the Huzinaga equation, the latter can be written as

$$[\mathbf{F} - \mathbf{SR}\bar{\mathbf{a}}\mathbf{F} - \mathbf{FR}\bar{\mathbf{a}}\mathbf{S}]\mathbf{C}^a = \mathbf{SC}^a\mathbf{E}^a \quad (4)$$

Note that the dimensions of the matrices are typically equal to the total number of basis functions or molecular orbitals as it is discussed below. This is the reason why matrix dimensions are not indicated in Eq. (4) and the group index appears as a superscript. As it is discussed in refs.^{1,2} orthogonality is implicit in the derivation of the Huzinaga equation and this has to be respected in the assignment of basis sets to groups. The inclusion of all basis functions of the

frozen groups into the basis set of the group(s) to be optimized is a valid choice. One can easily see that when basis set assignment allows orthogonality then the solutions of Eq. (4) are also solutions of Eq. (1). However, the molecular orbitals obtained by Eq. (1) are not necessarily orthogonal, even when basis sets make orthogonality possible.

It is discussed in ref.² that if orbitals in the frozen space are not good approximations to occupied orbitals then negative eigenvalue solutions of Eq. (4) appear in the frozen space and the self-consistent solution that includes the selection of lowest eigenvalue orbitals fails. Indeed, it was observed that certain combinations of core charges and frozen orbitals at the QM/MM boundary prevent the self-consistent solution of Eq. (4). By contrast, Eq. (1) is less sensitive in this respect as orbitals in the frozen space always appear as 0 eigenvalue solutions. Examples are presented in the subsequent sections where Eq. (1) is used to calculate the wave-function of a QM subsystem, and the solution of Eq. (4) fails.

SAMPLE CALCULATIONS

QM/MM type calculations were performed with a model described in ref.². The principal features of the model include the division of the system into a central QM part and an environment, the latter is represented by multipole derived point charges^{23,24}. The two subsystems are connected by one or several strictly localized bonds (SLMO). These SLMOs connect two atoms: *A* is the frontier atom, whose other bonds are directed towards the MM region, while bonds of *B* belong to the QM region. Bond charges are placed at the midpoints of the bonds of atom *A* in the MM region (Figure 1).

The objective of the calculations presented below is to explore what advantages are offered by the increased flexibility first, in the selection of core charges and frozen localized orbitals and second, in basis set assignments.

Calculations to solve Eq. (1) were performed with a locally modified version

of GAMESS-US²⁵. Multipole derived charges were obtained from a distributed multipole analysis with the program GDMA²⁶ followed by a charge fitting with the program Mulfit²⁷. Frozen orbitals were obtained with model molecules containing the relevant chemical motifs. The procedures agree with those applied in ref.².

Efficient geometry optimization with wave-functions obtained from the local basis equation requires the calculation of forces. Formulas for forces are presented in the Appendix. Most terms agree with those commonly used for canonical orbitals and existing codes can easily be adapted for geometry optimizations with local basis equation orbitals.

All calculations applied the 6-31G* basis set²⁸. Standard Hartree-Fock-Roothaan (HFR) calculations were used as reference.

As it was discussed above, the local basis equation gives a wave-function equivalent to that obtained with the Huzinaga equation, Eq. (4), when a basis set assignment compatible with the latter is used. Therefore, results presented in ref.² can also be obtained with the local basis equation, Eq. (1). The calculations were indeed repeated with the local basis equation in order to collect more experience concerning its convergence properties. Calculations were performed for the deprotonation energy of C₅H₁₁COOH, for the conformational energy of the same molecule as a function of the rotation of the -COOH group, for the conformational energy of the Gly-His-Gly peptide as a function of the imidazole rotation and for the proton transfer energy curve between sidechains of Asp and His residues. The systems are described in more detail in ref.² and also in the subsequent sections where calculations with varying parameters are presented. Here we note that with identical system setup (charges and basis set assignments) the results obtained in ref.² were perfectly reproduced and no convergence difficulties were observed. Accordingly, a QM/MM boundary separated by 2-3 bonds from a protonation site allows a good reproduction of the deprotonation energy. Geometrical parameters obtained from gradient optimizations are in excellent

agreement with reference results. It was also found that conformational energy curves are well reproduced with a 2-3 bond separation of the boundary from the rotating bond. A similarly small quantum system is able to well reproduce the proton transfer energy curve between aspartate and histidine residues.

Frozen orbitals and bond charges

It was reported in ref.² that the application of Eq. (4) requires the introduction of bond charges in order to achieve proper convergence in the self-consistent field solution. These bond charges are placed at the midpoints of those bonds of the frontier atom that are directed towards the MM subsystem (Figure 1). This allows the use of an increased core charge for the frontier atom that ensures the proper behavior of Eq. (4) in the self-consistent field solution, and the good reproduction of reference results.

The local basis equation exhibit good convergence even without bond charges and with a small core charge of the frontier atom. Thus the deprotonation energy calculation for molecule $C_5H_{11}COOH$ was repeated without bond charges. The C_{17} -atom of the terminal CH_3 group was selected as the frontier atom and the C_{14} - C_{17} bond as the frozen bond (Figure 2). (This separation corresponds to 'cut1' in ref.².) Thus the 3 terminal H-atoms were not part of the quantum system. A core charge of +2.861 was assigned to the frontier atom (C_{17}). This value was obtained as the difference between +3 (number of explicit electrons) and -0.139, the multipole derived charge for the atom. Multipole derived charges were also assigned to the MM atoms. Charges were calculated for the neutral molecule and were also applied for the charged one. The calculated deprotonation energy differs from the reference value by 6.1 kcal/mol. (We recall that the error is 0.2 kcal/mol with appropriately chosen bond charges.) It was also observed that the Mulliken atomic charges²⁹ increased on atom C_{14} (-0.075) and decreased on atoms connecting to C_{14} (C_{11} -0.326, H-atoms connecting to C_{14} 0.074) relative to the HFR reference (C_{14} -0.304, C_{11} -0.307, H-atom connecting to C_{14} 0.158).

By contrast, Mulliken charges are close to the HFR Mulliken charges when bond charges are present. This suggests that the frozen orbital itself is unable to fully compensate for the unphysical environment near to the subsystem boundary and this causes a significant error in calculated deprotonation energies. On the other hand, the frozen orbital together with appropriately chosen frontier atom core charge and bond charges are able to restore the correct polarity of the wave-function near to the boundary.

The combined effect of the frozen orbital and bond charges was further analyzed in calculations with varying basis functions for the active orbitals. (The term active orbital refers to those orbitals whose coefficients are optimized in contrast to the frozen orbitals whose coefficients are kept fixed.) While the application of the Huzinaga equation requires the inclusion of the frozen orbital basis into the active basis set this is not necessary when the local basis equation is used. The frozen SLMO uses the basis orbitals of the atoms connected, namely the frontier atom whose other bonds are directed towards the MM region and the connected atom in the QM region. The fixed core orbital on the frontier atom also uses the basis functions of this atom. On the other hand, the active orbitals do not necessarily use the basis functions on the frontier atom. The computational saving one may realize with the exclusion of frontier atom functions from the active basis set is minor, but it affects the wave-function and potentially the deprotonation energy. However, no significant variation of the deprotonation energy is observed using the reduced basis set either with (0.2 kcal/mol error) or without (6 kcal/mol error) bond charges.

These calculations were repeated with a more extended frozen orbital at the boundary (Figure 3). The frozen localized orbital includes the basis functions centered on the terminal C₁₁-C₁₄-C₁₇ moiety including the two H-atoms connecting to C₁₄. By contrast, the active basis set does not include functions on the terminal C₁₄-C₁₇ moiety and its H-atoms. In this way, the perturbation caused by the QM/MM boundary is expected to be better modulated by the frozen orbital as the active orbitals do not extend to the terminal -C₁₄-C₁₇ group. We

found, however, that the deprotonation energy without bond charges is still in an error of about 5 kcal/mol. An analysis of the Mulliken charges reveals that the C₁₄ methylene group (C₁₄ -0.389, connecting H-atoms 0.163) is more similar to the reference (C₁₄ -0.304, connecting H-atoms 0.158) than before (C₁₄ -0.075, connecting H-atoms 0.074) as a consequence of keeping the orbitals using the C₁₄ basis functions frozen. On the other hand, the connecting C₁₁-atom (-0.078), closer to the carboxyl end, exhibits larger deviation from the reference (-0.307) than before (-0.326).

The importance of the deformation of the wave-function near to the boundary without bond charges is expected to depend on the property studied. While deprotonation energies are significantly influenced, a less pronounced effect is expected when the charge of the systems to be compared is unaltered. In order to test this hypothesis calculations for conformational energies and for a proton transfer energy curve previously performed with bond charges² were repeated without bond charges. In these calculations the core charge of the frontier atom was set to the sum of +3 (the number of explicit electrons) and the multipole derived charge for this atom.

The energy as a function of the rotation of the carboxyl group in the C₅H₁₁COOH molecule around the C5-C3 bond was calculated with two subsystem separations. One (cut1) that corresponds to that in Figure 2 and another (cut2) with a smaller QM subsystem, where the frozen SLMO bounds atoms C₁₁ and C₁₄ were selected. These separations were previously found to yield a good reproduction of the energy curve when bond charges were used (Figure 2 in ref.²). Figure 4 shows that the energy curve obtained without bond charges well follows the reference curve with both system separations. The energy difference with respect to the reference is larger without bond charges than with bond charges but does not exceed 0.5 kcal/mol.

The energy of the rotation of the imidazol group in the Gly-His-Gly tripeptide as obtained with and without bond charges together with the reference results are

shown in Figure 5. The shape of the curve is again well described by the QM/MM calculations. Results without bond charges show again a larger deviation from the reference values. The largest difference with bond charges is within 1 kcal/mol and it amounts to nearly 2 kcal/mol without bond charges.

The energy variation in the proton transfer between aspartate and histidine was calculated as a function of the position of the proton. Results obtained with and without bond charges together with the reference is shown in Figure 6. The shape of the curves are very similar for all three calculations. The deviation from the reference is smaller with bond charges than without bond charges. The curves are superimposed at their minimum energy value that corresponds to the protonated aspartate. The other minimum corresponds to the protonated histidine. The difference between the two minima is reproduced with an error of 1.3 kcal/mol with bond charges and with an error of 3.5 kcal/mol without bond charges.

These results show that bond charges applied together with an increased core charge contribute to a better reproduction of reference results. It has to be noted that the magnitude of charges were not carefully optimized, rather they were selected by a trial and error procedure to well describe the deprotonation energy of $C_5H_{11}COOH$ at a particular QM-MM subsystem separation², and were used in all calculations performed for various systems and subsystem separations. Thus the application of bond charges with increased core charge is beneficial for accuracy. Moreover, our results suggests that the core charge is fairly transferable as the same core charge together with bond charges adjusted to the total multipole derived charge of the atom could be effectively used in different chemical environments. On the other hand, calculations without bond charges have the advantage of having less parameters. The price one has to pay for it is the lower accuracy. It is likely that accuracy can be improved by increasing the size of the quantum subsystem, thus increasing the separation between the chemical event and the perturbed electron density near to the subsystem boundary.

Nonorthogonal localized orbitals in multilayer QM systems

The local basis equation was shown to be able to calculate *a priori* localized orbitals¹. These orbitals can use local basis sets and they are nonorthogonal. Their advantage is that they require the solution of reduced dimension equations than the standard, full basis Hartee-Fock-Roothaan equations. This can be exploited in the framework of mixed methods by defining a QM/QM system built from a central subsystem with the usual delocalized orbitals and a localized subsystem that includes several localized molecular orbitals. The orbitals of both subsystems of this QM/QM system are optimized. Eq. 1 is solved for each group, namely for the central part, and for each group of localized orbitals, except for the frozen one. These equations are coupled through $\mathbf{R}^{\bar{a}}$ that is built from the coefficients of other groups, therefore the equations are solved in an iterative way. A more detailed description of the solution for orbitals in several groups is given in ref.¹.

An alternative to the above approach is to use fixed localized orbitals in the outer QM subsystem. This option was not tested but we note that models using a QM environment with constrained or fixed electron density interacting with the central QM subsystem were proposed within the framework of the density functional theory. A variant of the divide-and conquer method³⁰ uses a preoptimized density that is not changed when the geometry of the central QM subsystem is optimized. The constrained³¹ and frozen DFT³² approaches calculate the interactions between the subsystems quantum mechanically and those within the outer subsystem classically. These methods, within the framework of DFT, provide an alternative way of solving the nonorthogonality problem between the central subsystem and its environment.

A QM/QM system built from delocalized and localized orbitals can be complemented with a MM subsystem. Such a multilayer system as defined for the $\text{C}_5\text{H}_{11}\text{COOH}$ molecule is shown in Figure 7. The deprotonation energy calculated

for this QM/QM/MM system reproduces the standard Hartree-Fock-Roothaan result within 1 kcal/mol.

The energy curve of the rotation of the carboxyl group in the $C_5H_{11}COOH$ was also investigated with a 3 layer model shown in Figure 8. The localized QM system is smaller and the MM system is larger than in the previous deprotonation study. This separation was selected, since a 2 layer system with the same MM subsystem was shown to describe the energy curve very well². This latter curve was calculated with the Huzinaga equation but it can be obtained also from the local basis equation (with a single QM region and with basis functions of all atoms including the frozen orbitals). Figure 9 shows the standard QM energy curve together with the 2 layer and 3 layer QM/MM results. The energy curve of the 3 layer model is practically indistinguishable from the 2 layer curve showing that the use of localized orbitals introduces no error into the model. This is remarkable taking into account the small extension of the localized orbitals and also their vicinity to the rotating bond.

CONCLUSION

The local basis equation, Eq. (1)¹, is proposed for use in mixed QM/MM and QM/QM calculations as it is appropriate to calculate the QM wave-function interacting with frozen orbitals^{5-10,17}. The equation results in orbitals that make the energy stationary and it proved to be well suited for an iterative self-consistent field solution. The resulting wave-function is equivalent with that obtained with the Huzinaga equation, Eq. (4)^{2,11}, when a basis set consistent with the latter is used. It well reproduced deprotonation energy, conformational energy curves and a proton transfer energy curve when the QM/MM boundary separated by 2-3 bonds from a protonation site or from the rotating bond. Under these conditions, geometrical parameters obtained from gradient optimizations are in excellent agreement with reference results.

The local basis equation is more flexible than the Huzinaga equation for use

in mixed methods for two reasons. First, it is compatible with the assignment of local basis sets owing to its ability to calculate nonorthogonal orbitals. Second, orbitals in the frozen space that are not good approximations to the occupied orbitals prevent the self-consistent solution for active orbitals with the Huzinaga equation but they are compatible with the local basis equation.

Calculations were performed to study the possibility to omit bond charges and an increased core charge at the QM/MM boundary found to be necessary in calculations applying the Huzinaga equation. It was shown that the artificial environment near to the boundary is only partially compensated by the frozen bond and an increased core charge with bond charges is necessary for a more complete compensation. The use of bond charges is required when energies of systems with different charges are to be compared like in deprotonation energy calculations. The beneficial effect of bond charges was found to be less pronounced but still significant in conformational energy calculations and in the description of a proton transfer energy curve. Thus, although the omission of bond charges allows the use of a decreased number of system specific parameters, it results in lower accuracy that possibly can be compensated by a larger size QM system.

The local basis equation can be used to derive *a priori* localized nonorthogonal orbitals at a reduced computational effort. This feature combined with its ability to optimize orbitals interacting with frozen orbitals can be used to define a two layer QM system with a delocalized inner and a localized outer layer and they can be complemented with an MM layer resulting in a QM/QM/MM system. Such systems were realized in studies of deprotonation and conformational energies. Results were found to agree very well with that of the two layer QM/MM model.

The molecular orbitals obtained with the local basis equation are in general nonorthogonal even if a basis set compatible with orthogonal orbitals is used and this is in contrast to the solutions of the Huzinaga equation. The nonorthogonal orbitals can be orthogonalized *a posteriori* that may be advantageous e.g. when they are used in post-Hartree-Fock methods.

ACKNOWLEDGMENTS

The support by the project TÁMOP-4.2.2.C-11/1/KONV-2012-0010 "Supercomputer - the national virtual laboratory" which is supported by the European Union and co-financed by the European Regional Fund is gratefully acknowledged.

APPENDIX

Derivatives of the One-determinant Wave-function of Nonorthogonal Orbitals

The energy of a closed shell system is written as

$$E = \frac{1}{2} \sum_{\alpha\beta} P_{\alpha\beta} (h_{\beta\alpha} + F_{\beta\alpha}) \quad (5)$$

where $P_{\alpha\beta}$, $h_{\beta\alpha}$ and $F_{\beta\alpha}$ are elements of the density matrix, the core Hamiltonian matrix and the Fock matrix, respectively, and the summations extend over the basis set.

We permit the molecular orbitals to overlap, i.e., $\sigma_{ij} = \langle \phi_i | \phi_j \rangle$ is not assumed to equal zero when $i \neq j$. This leads to two notable differences with respect to the derivatives of canonical orbitals. First, the density matrix of a closed shell system of nonorthogonal orbitals takes the form of $\mathbf{P} = 2\mathbf{R}$ where $\mathbf{R} = \mathbf{C}\sigma^{-1}\mathbf{C}^\dagger$. Second, in contrast to the canonical case where the orthonormality of the molecular orbitals has to be taken into account no such constraint appears in the nonorthogonal case.

The energy derivative with respect to a nuclear coordinate (q_i) can be written as³³

$$\frac{dE(\mathbf{C}(\mathbf{q}), \mathbf{q})}{dq_i} = \frac{\partial E(\mathbf{C}(\mathbf{q}), \mathbf{q})}{\partial q_i} \quad (6)$$

i.e. the dependence on the coefficients does not enter. This holds for our case, since coefficients are either optimized or are kept fixed. (See ref.¹⁷ for taking into account the variation of the frozen orbital coefficients with changes in atomic positions.)

Then

$$\frac{dE}{dq_i} = A + B \quad (7)$$

with

$$A = \sum_{\alpha\beta} P_{\alpha\beta} \frac{\partial h_{\alpha\beta}}{\partial q_i} + \frac{1}{2} \sum_{\alpha\beta\gamma\delta} P_{\alpha\beta} P_{\gamma\delta} \left(\frac{\partial \langle \alpha\beta | \gamma\delta \rangle}{\partial q_i} - \frac{1}{2} \frac{\partial \langle \alpha\beta | \delta\gamma \rangle}{\partial q_i} \right) \quad (8)$$

and

$$B = \sum_{\alpha\beta} \frac{\partial P_{\alpha\beta}}{\partial q_i} F_{\alpha\beta} \quad (9)$$

A appears also in the formula for canonical orbitals. On the other hand, B includes the derivatives of the density matrix and thus it is affected by the nonorthogonality. σ^{-1} in $\mathbf{P} = 2\mathbf{C}\sigma^{-1}\mathbf{C}^\dagger$ does depend on the nuclear coordinates. Since

$$\frac{\partial \sigma^{-1}}{\partial q_i} = -\sigma^{-1} C^\dagger \frac{\partial S}{\partial q_i} C \sigma^{-1} \quad (10)$$

B can be written as

$$B = -2 \sum_{\alpha\beta\gamma\delta} \frac{\partial S_{\gamma\delta}}{\partial q_i} R_{\delta\beta} F_{\alpha\beta} R_{\alpha\gamma} \quad (11)$$

It is worth noting that if R is built from the canonical orbitals then the right hand side of Eq. (11) is equal to $-\sum_{\alpha\beta} \frac{\partial S_{\alpha\beta}}{\partial q_i} W_{\alpha\beta}$, where $W_{\alpha\beta} = 2 \sum_i C_{\alpha i} E_i C_{\beta i}$ is an element of the energy weighted density matrix. Then Eqs. 7-11 reduce to the derivatives of the canonical orbitals.

Computer programs available for the calculation of the derivatives with canonical orbitals can be adapted easily for nonorthogonal orbitals. The energy weighted density matrix has to be replaced by the $2\mathbf{RFR}$ matrix and the projector \mathbf{R} has to be evaluated with the nonorthogonal formula.

References

1. Ferenczy G. G.; Adams, W. H. *J. Chem. Phys.* **2009**, *130*, 134108.
2. Ferenczy G. G. *submitted for publication*.
3. Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227-249.
4. Théry, V.; Rinaldi, D.; Rivail, J. L.; Maignet, B.; Ferenczy, G.G. *J. Comput. Chem.* **1994**, *14*, 269-282.
5. Assfeld, X.; Rivail, J. L. *Chem. Phys. Lett.* **1996**, *263*, 100-106.
6. Philipp, D. M.; Friesner R. A. *J. Comput. Chem.* **1999**, *20*, 1468-1494.
7. Murphy, R. B.; Philipp, D. M.; Friesner R. A. *J. Comput. Chem.* **2000**, *21*, 14421457.
8. Gao, J.; Amara, P.; Alhambra C.; Field, M. J. *J. Phys. Chem.* **1998**, *102*, 4714-4721.
9. Pu, J.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 632650.
10. Jung, J.; Choi C. H.; Sugita, Y.; Ten-no, S. *J. Chem. Phys.* **2007**, *127*, 204102.
11. Huzinaga, A.; Cantu, A. A. *J. Chem. Phys.* **1971**, *55*, 5543-5549.
12. Stoll, H.; Wagenblast, G.; Preuß, H. *Theoret. Chim Act.* **1980**, *57*, 169-178.
13. Gianinetti, E.; Raimondi, M.; Tornaghi, E. *Int. J. Quant. Chem.* **1996**, *60*, 157-166.
14. Sironi, M.; Famulari, A. *Theor. Chem. Acc.* **2000**, *103*, 417-422.
15. Fornili, A.; Sironi, M.; Raimondi, M. *J. Mol. Struct. (Theochem)* **2003**, *632*, 157-172.

16. Fornili, A.; Moreau, Y.; Sironi, M.; Assfeld, X. *J. Comput. Chem.* **2006**, *27*, 515-523.
17. Ferré, N.; Assfeld, X.; Rivail, J.-L. *J. Comput. Chem.* **2002**, *23*, 610-624.
18. Smits, G. F.; Altona, C. *Theor. Chem. Acc.* **1985**, *67*, 461-475.
19. Couty, M.; Bayse, C. A.; Hall, M. B. *Theor. Chem. Acc.* **1997**, *97*, 96-109.
20. Matsuoka, O. *J. Chem. Phys.* **1977**, *66*, 1245-1254.
21. Fülischer, M. P.; Mehler, E. L. *J. Comput. Chem.* **1991**, *12*, 811-828.
22. Szekeres, Z.; Surján, P. R. *Chem. Phys. Lett.* **2003**, *369*, 125-130.
23. Ferenczy, G. G. *J. Comput. Chem.* **1991**, *12*, 913-917.
24. Winn, P. J.; Ferenczy, G. G.; Reynolds, C. A. *J. Phys. Chem. A* **1997**, *101*, 5437-5445.
25. Schmidt, M.W.; Baldrige, K.K.; Boatz, J.A.; Elbert, S.T.; Gordon, M.S.; Jensen, J.J.; Koseki, S.; Matsunaga, N.; Nguyen, K.A.; Su, S.; Windus, T.L.; Dupuis, M.; Montgomery, J.A. *J. Comput. Chem.* **1993**, *14*, 1347-1363.
26. Stone, A. J. *J. Chem Theory Comp.* **2005**, *1*, 1128-1132.
27. Ferenczy, G. G.; Reynolds, C. A.; Winn, P. J.; Stone, A. J. Mulfit program included in the GDMA distribution <http://www-stone.ch.cam.ac.uk/programs/gdma.html>
28. Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257-2261.
29. Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 2343-2346.
30. Lee, T-S.; Yang, W. *Int. J. Quant. Chem.* **1998**, *69*, 397-404.
31. Hong, G.; Strajbl, M.; Wesolowski, T. A.; and Warshel, A. *J. Comput. Chem.* **2000**, *21*, 1554-1561.

32. Wośowski, T. A.; Warshel, A. *J. Phys. Chem.* **1993**, *97*, 80508053.

33. Pulay, P. *Mol. Phys.* **1968**, *17*, 197-204.

FIGURE CAPTIONS

Figure 1: Schematic representation of the separation of the system into QM and MM subsystems. Atom A is the frontier atom. Bond charges are represented by circles on the bonds of A within the MM region.

Figure 2: Separation of the $C_5H_{11}COOH$ molecule into subsystems. Subsystems are connected with an SLMO. See text for details.

Figure 3: Separation of the $C_5H_{11}COOH$ molecule into subsystems. Subsystems are connected with an LMO. See text for details.

Figure 4: Energy of the $C_5H_{11}COOH$ molecule as a function of the rotation of the $-COOH$ group calculated with 2 different system separations with and without bond charges. See text for the system separations designated by 'cut1' and 'cut2'.

Figure 5: Energy of the Gly-His-Gly molecule as a function of the rotation of the histidine group calculated with and without bond charges.

Figure 6: Energy of the Asp - His system as a function of the separation of the proton from the oxygen of Asp calculated with and without bond charges.

Figure 7: Separation of the $C_5H_{11}COOH$ molecule into subsystems. The central delocalized QM subsystem includes the C-COOH moiety with 12 doubly occupied MOs. The localized QM subsystem includes 17 localized MOs out of which 4 are core orbitals. The frozen MOs are a C-C bond and a core orbital at the QM/MM boundary. The H-atoms of the terminal CH_3 group are represented by point charges.

Figure 8: Separation of the $C_5H_{11}COOH$ molecule into subsystems. The central delocalized QM subsystem includes the C-COOH moiety with 12 doubly occupied MOs. The localized QM subsystem includes 13 localized MOs out of which 3 are core orbitals. The frozen MOs are a C-C bond and a core orbital at the QM/MM boundary. The terminal CH_3 group and the H-atoms of the connecting CH_2 group are represented by point charges.

Figure 9: Energy of the $C_5H_{11}COOH$ molecule as a function of the rotation of the -COOH group calculated with standard QM and with a 2 layer and a 3 layer model the latter is shown in Figure 8. See text for details.

FIGURES

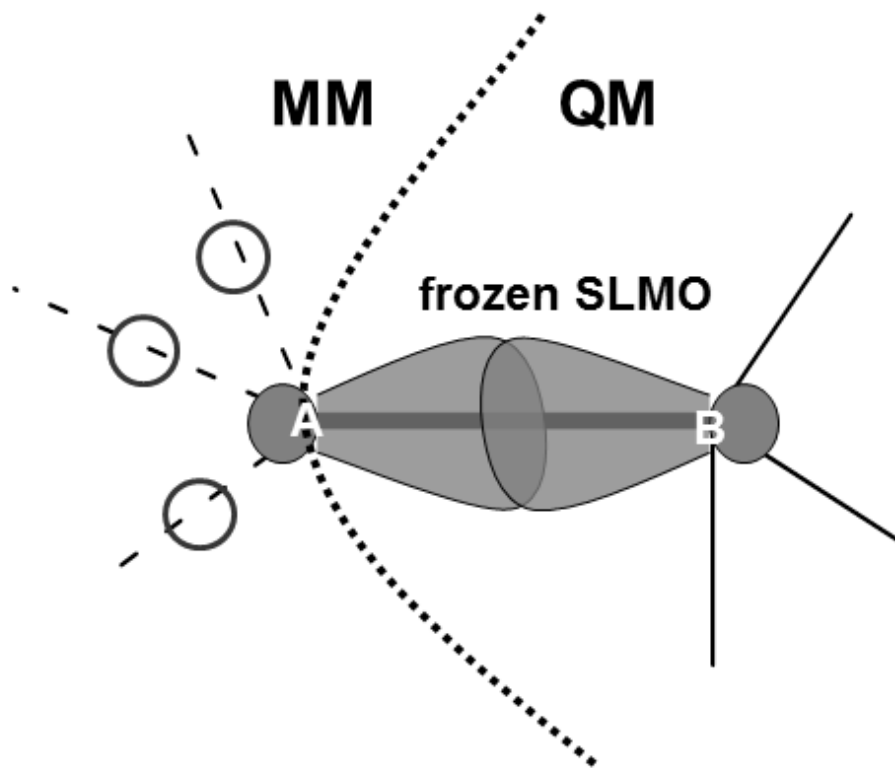


Figure 1
G. Ferenczy
J. Comput. Chem.

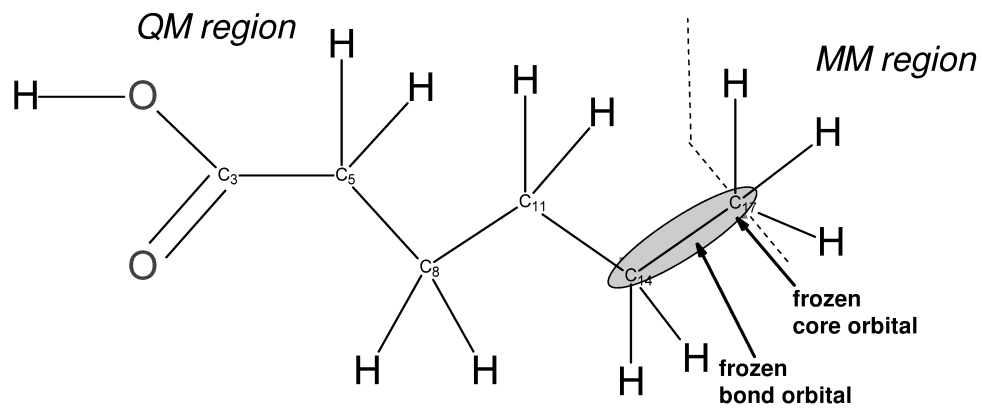


Figure 2
G. Ferenczy
J. Comput. Chem.

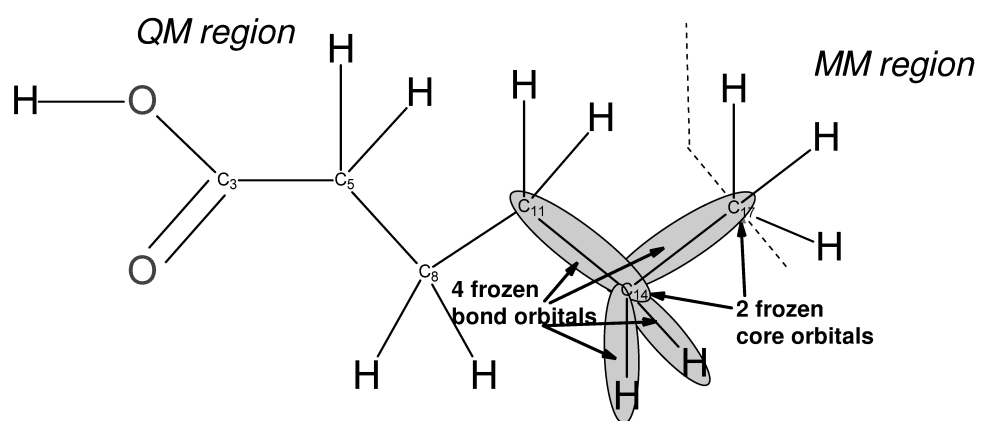


Figure 3
G. Ferenczy
J. Comput. Chem.

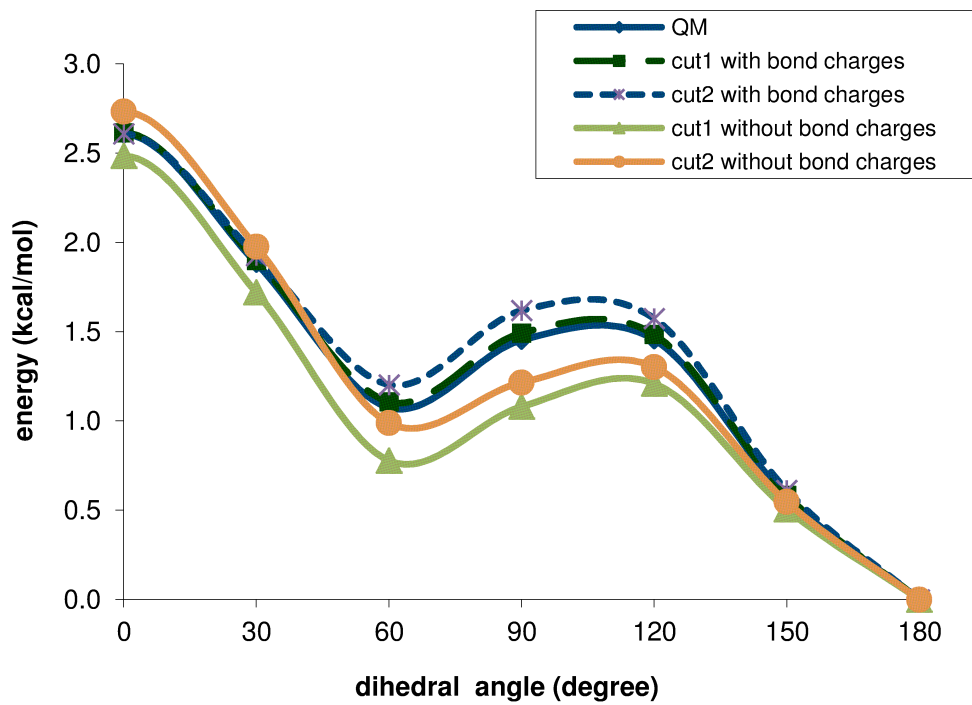


Figure 4
G. Ferenczy
J. Comput. Chem.

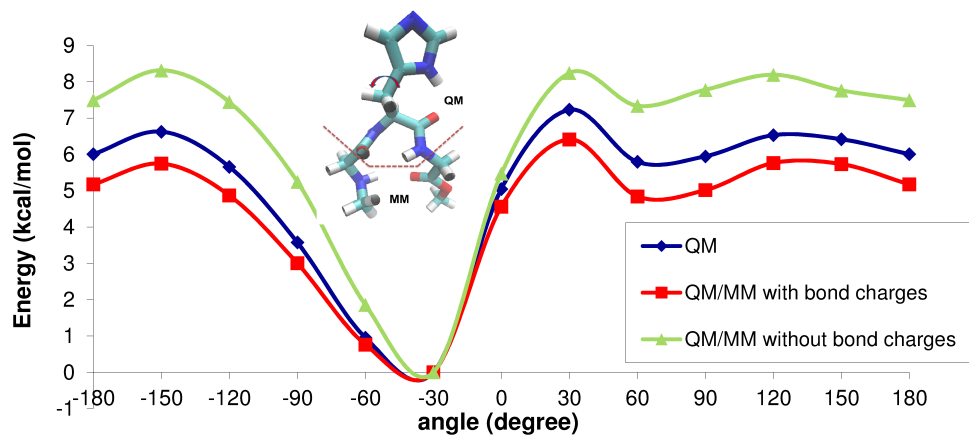


Figure 5
G. Ferenczy
J. Comput. Chem.

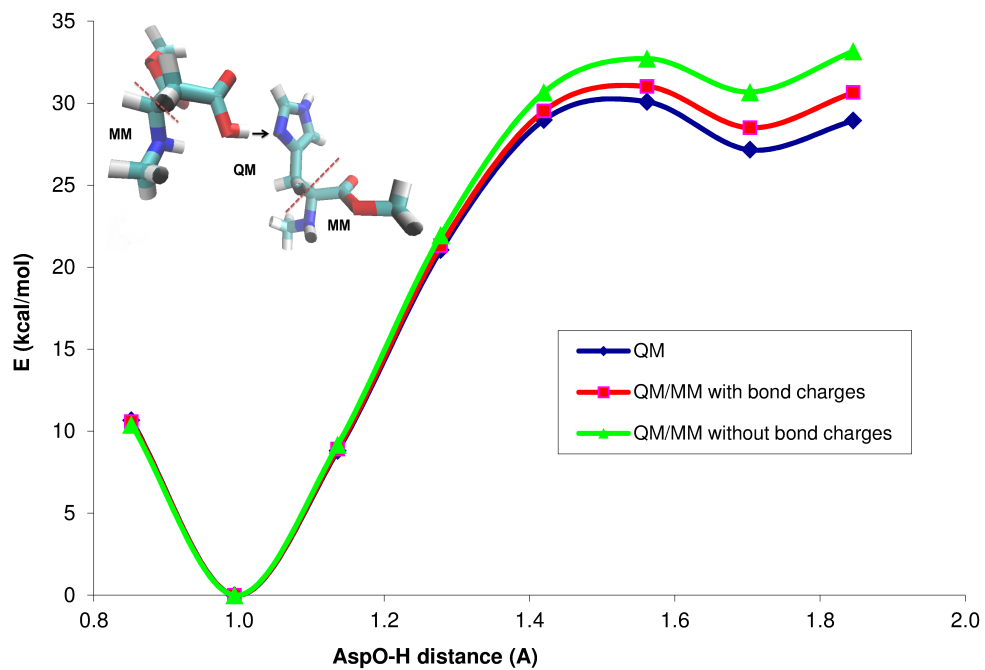


Figure 6
G. Ferenczy
J. Comput. Chem.

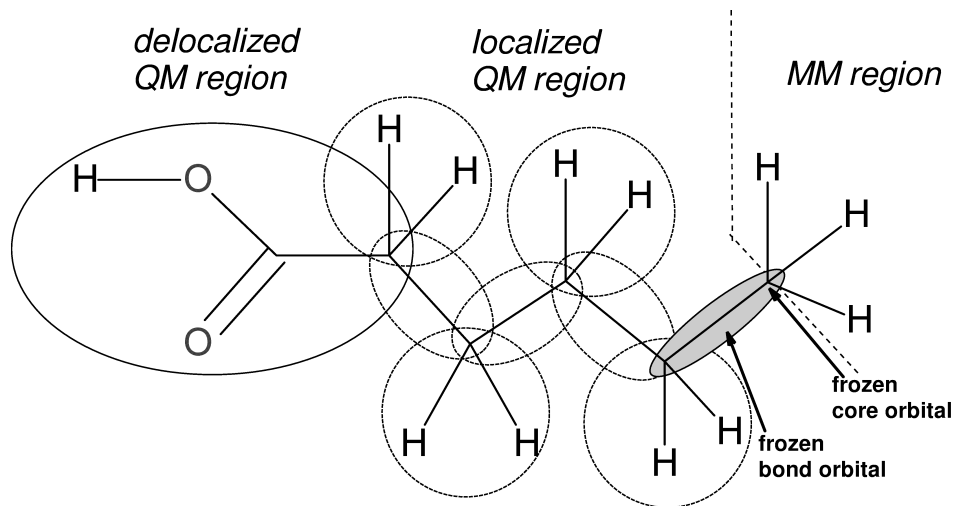


Figure 7
G. Ferenczy
J. Comput. Chem.

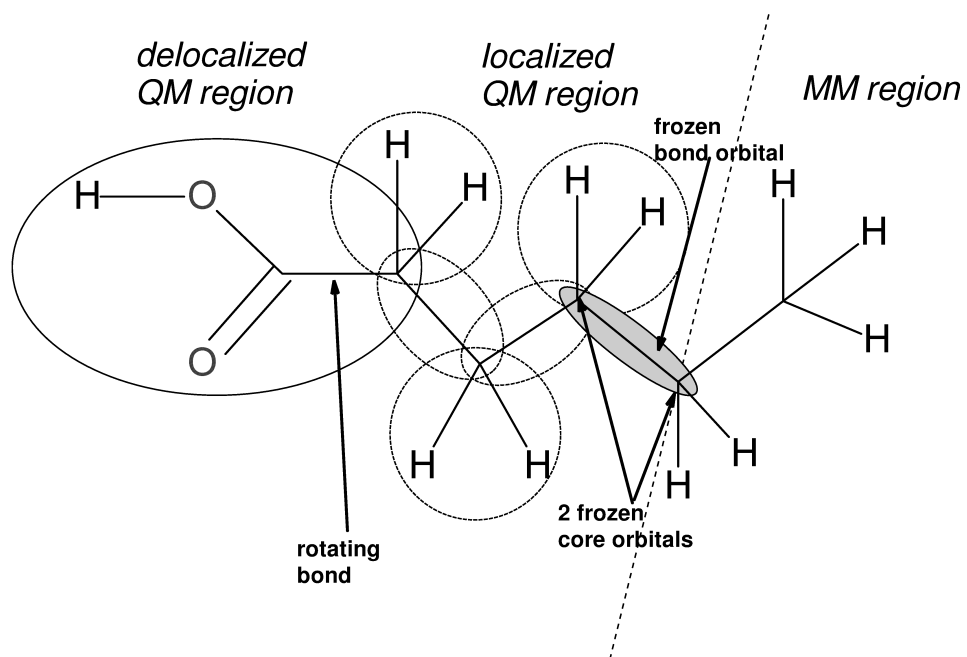


Figure 8
 G. Ferenczy
 J. Comput. Chem.

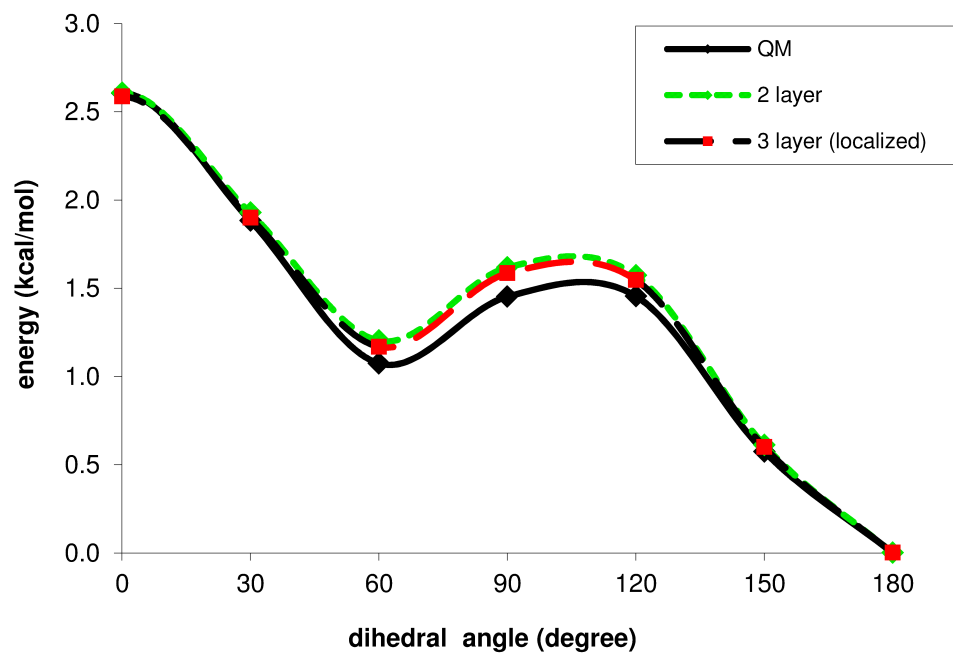


Figure 9
G. Ferenczy
J. Comput. Chem.