Calculation of Wave-functions with Frozen Orbitals in Mixed Quantum Mechanics/Molecular Mechanics methods. I. Application of the Huzinaga Equation.

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

November 8, 2012

Abstract

Mixed quantum mechanics/quantum mechanics (QM/QM) and quantum mechanics/molecular mechanics (QM/MM) methods make computations feasible for extended chemical systems by separating them into subsystems that are treated at different level of sophistication. In many applications the subsystems are covalently bound and the use of frozen localized orbitals at the boundary is a possible way to separate the subsystems and to ensure a sensible description of the electronic structure near to the boundary. A complication in these methods is that orthogonality between optimized and frozen orbitals has to be warranted and this is usually achieved by an explicit orthogonalization of the basis set to the frozen orbitals. An alternative to this approach is proposed by calculating the wave-function from the Huzinaga-equation that guaranties orthogonality to the frozen orbitals without basis set orthogonalization. The theoretical background and the practical aspects of the application of the Huzinaga equation in mixed methods is discussed. Forces have been derived to perform geometry optimization with wave-functions from the Huzinaga-equation. Various properties have been calculated applying the Huzinaga-equation for the central QM subsystem, representing the environment by point charges and using frozen strictly localized orbitals to connect the subsystems. It is shown that a 2-3 bond separation of the chemical or physical event from the frozen bonds allows a very good reproduction (typically around 1 kcal/mol) of standard Hartree-Fock-Roothaan results. The proposed scheme provides an appropriate framework for mixed QM/QM and QM/MM methods.

Keywords: mixed QM/MM, QM/QM, frozen localized orbital, Huzinaga equation,

■

INTRODUCTION

Extended electronic systems are advantageously treated with mixed quantum mechanics/quantum mechanics (QM/QM) or quantum mechanic/molecular mechanics (QM/MM) methods. These schemes make it possible to perform an accurate QM calculation for a part of the system while other parts are treated at a more approximate level. Mixed methods are suitable to investigate various phenomena including chemical and biochemical reactions[1–5]. The separation of the system into subsystems is performed so that a part relevant to the phenomena forms the central subsystem treated with an appropriate QM method while the rest of the system is considered as the environment and a lower level method is applied. The separation is a critical issue motivated by different factors. A small central QM subsystem encompassing the relevant structural motif is beneficial for the required computational work but the proximity of the subsystem boundary may cause artifacts. Performing comparative studies, e.g. on reaction mechanisms in enzyme and in water[6,7], is assumed to alleviate these artifacts. However, numerous applications of mixed methods[1,4] use no reference system and the proper treatment of the boundary significantly affects the quality of the results.

Certain QM/QM schemes, like the fragment molecular orbital method[8,9], the divide and conquer method[10,11], its molecular orbital variant[12,13] and also the frozen DFT method[14,15] inherently define the way of separating the subsystems. Other methods require special considerations when the boundary crosses chemical bonds that is often inevitable for example in studying macromolecules like biopolymers. The separation of covalently bound subsystems was developed basically along too lines. The link atom method cuts the bond and introduces extra capping atoms to fill the dangling bonds. These extra link atoms are most often hydrogen-atoms, but can be another type of atoms or chemical groups. This scheme is conceptually simple and is easy to implement in computational codes. Its disadvantage is that the addition of link atoms creates artifacts that have to

be corrected for. The electron density of the central QM system is deformed owing to the introduction of the link atoms. In addition, these link atoms are close to some atoms in the other subsystem. Nevertheless, the link atom approach is widely used and it can provide useful results when appropriate care is taken in its application[16].

An alternative way of separation along a chemical bond does not introduce extra atoms, but applies a special treatment for the electronic properties near to the QM/MM boundary. Several variants of the boundary atom approach use appropriately parametrized potentials to describe the interaction of the boundary atom with the quantum system while the atom exhibits an MM-like character towards the MM system (see early proposed methods in refs.[17,18] and a review in ref.[4]). By contrast, the frozen orbital approach uses predefined frozen localized orbitals to connect the subsystems. The present contribution proposes a novel way to calculate the QM wave-function in the frozen orbital approach, the latter is presented in some details below.

The basic idea is to connect covalently bound atoms belonging to different subsystems with localized orbitals, while the other bonds of these atoms are treated in a way consistent with the rest of the subsystem they belong to. The original proposal comes from Warshel and Levitt[19] within a QM/MM scheme using a semiempirical QM method with hybrid orbitals. Hybrid orbital basis and strictly localized orbitals were also used in mixed QM/QM methods at the CNDO/2[20] and NDDO[21] levels. These methods naturally yield strictly localized orbitals as connections between the subsystems. These QM/QM methods were transformed into a QM/MM scheme in which the strictly localized orbitals connecting the QM and MM parts are taken from a model system and are frozen in the course of the calculation of the QM wave-function[22]. This Local Self-Consistent Field (LSCF) method was later extended to ab initio level[23,24].

The idea of connecting two sybsytems with frozen localized orbitals is illustrated in Figure 1. Atom A (the frontier atom) has a strictly localized bond orbital that connects it to atom B in the central QM subsystem. Other bonds

of atom A are directed towards the environment that will be referred to as MM subsystem, although it can be also treated with a QM method typically at a more approximate level than the central subsystem. Atom B is bound to other atoms in the central QM subsystem and all electrons of atom B except those participating in the bond to atom A are handled at a QM level that is the single determinant approximation in the present approach. Thus the separation of the two subsystems is possible at an atom (A) that binds to the central QM subsystem with a bond that can be reasonably described by a strictly localized molecular orbital. It is worth noting that several frozen orbitals may appear in the system even when only two subsystems are present if the subsystems are connected by several bonds.

Frozen orbitals are also used in other related methods. Friesner and co-workers performed intensive parametrization of the interaction terms of atoms near to the subsystem boundary and achieved accurate description of geometries and energies[25,26]. The method of Generalized Hybrid Orbitals (GHO) places hybrids on atoms at the boundary, and those pointing towards the QM system are included in the self-consistent solution while those pointing towards MM atoms are frozen[27,28]. While the original GHO method assigns the same occupation number to all frozen hybrids of an atom, a more flexible variant allows the variation of the occupation numbers[29]. These approaches were also extended to post-Hartree-Fock methods[26,30,31].

A common feature of the frozen orbital approaches is that the calculation for the central quantum mechanical subsystem takes into account the interaction between the central system and the frozen MOs. Typically, the equations used for determining the central wave-function assume interactions among orthonormal orbitals in line with the ability of these equations to provide us with orthonormal optimized orbitals. However, they do not guarantee that the optimized orbitals are orthogonal to the frozen orbitals. While orthogonality is typically guaranteed at a semiempirical level owing to the neglect of certain orbital overlaps, orthogonality is not automatically achieved at the ab initio level.

A possible remedy to this problem is to explicitly orthogonalize the basis set of the active orbitals to the frozen MOs[23,28]. The neglect of nonorthogonality[28] was also proposed. Philipp and Friesner[25] derived a self-consistent field equation with a non-hermitian, modified Fock matrix to calculate orthogonal orbitals. In the present contribution another approach is proposed. The central wave-function is calculated from the Huzinaga equation[32] that is able to give orbitals orthogonal to the frozen orbitals without explicit orthogonalization. The equation and its application in the framework of mixed methods with frozen orbitals is presented below.

THE HUZINAGA EQUATION IN THE FROZEN LOCALIZED ORBITAL APPROACH

The Huzinaga equation was derived to determine a set of active orbitals interacting with a set of frozen orbitals within the Hartee-Fock approximation[32]

$$(\hat{F} - \hat{\rho}^f \hat{F} - \hat{F}\hat{\rho}^f)\phi_i^a = \epsilon_i^a \phi_i^a. \tag{1}$$

where $\hat{F}$ is the Fock operator, $\hat{\rho}^f$ projects to the space of frozen orbitals, ϕ_i^a is the ith active orbital and ϵ_i^a is the associated eigenvalue.

Former applications of the Huzinaga equation include the calculation of valence orbitals in the field of fixed core orbitals and it also served as starting point for additional approximations like replacing the core electrons with model potentials[33,34]. An analysis of the Huzinaga equation has been presented in ref.[35] and its conclusions relevant to our present topic is recapitulated below.

Eq. 1 has been derived with the assumption that the active orbitals are orthogonal to the frozen orbitals[32]. This is automatically satisfied when the same basis set is used for both the active and frozen sets of orbitals. When group specific basis sets are used then care has to be taken to select basis sets that allow orthogonality among groups. As frozen orbitals in mixed QM/MM methods are

localized by having basis functions centered on the atoms at the boundary the basis set of the active orbitals has to include these orbitals, as well.

Frozen orbitals do not have to be eigenfunctions of the Fock operator. The idempotency of $\hat{\rho}^f$, built from the frozen orbitals assures that it commutes with the Huzinaga operator, $\hat{F} - \hat{\rho}^f \hat{F} - \hat{F}\hat{\rho}^f$, on the left of Eq. (1). These two operators can be diagonalized on a common set of functions. Thus appropriate linear combinations of the frozen orbitals are eigenfunctions of the Huzinaga operator. Moreover, when the nonzero eigenvalues of $\hat{\rho}^f \hat{F} \hat{\rho}^f$ are negative, then the frozen space eigenvalues of the Huzinaga equation are positive while the eigenvalues of the occupied active orbitals are negative. Then the frozen and active orbitals are guaranteed to be orthogonal owing to the hermicity of the Huzinaga operator. The nonzero eigenvalues of $\hat{\rho}^f \hat{F} \hat{\rho}^f$ are negative when the frozen orbitals are appropriately chosen exact eigenfunctions of $\hat{F}$ (i.e. occupied orbitals) and they are expected to be negative when the frozen orbitals are reasonable approximations to the exact eigenfunctions of $\hat{F}$. On the other hand, when an orbital in the frozen space is associated with a positive eigenvalue of $\hat{\rho}^f \hat{F} \hat{\rho}^f$ then this orbital appears as a negative eigenvalue solution of Eq. (1) and this prevents the usual self-consistent field solution for orbitals orthogonal to the frozen space.

At this point the introduction of basis functions is appropriate. The Huzinaga equation in terms of basis functions reads as

$$\left[\mathbf{F} - \mathbf{S}\mathbf{R}^f\mathbf{F} - \mathbf{F}\mathbf{R}^f\mathbf{S}\right]\mathbf{C}^{\mathrm{a}} = \mathbf{S}\mathbf{C}^{\mathrm{a}}\mathbf{E}^{\mathrm{a}}, \tag{2}$$

where $\mathbf{F}$ is the Fock matrix, $\mathbf{S}$ is the basis overlap matrix, $\mathbf{C}^a$ includes the coefficients of the active orbitals, $\mathbf{E}^a$ is the diagonal matrix of the corresponding eigenvalues and $\mathbf{R}^f$ projects to the frozen orbitals. Assuming orthonormal frozen orbitals $\mathbf{R}^f = \mathbf{C}^f(\mathbf{C}^f)^\dagger$. In this equation two groups appear. In fact, several frozen and active groups can be defined and the Huzinaga equation can be used to optimize orbitals for several active groups. Then an equation for each active group has to be solved.

$$\left[\mathbf{F} - \mathbf{S}\mathbf{R}^{\overline{a}}\mathbf{F} - \mathbf{F}\mathbf{R}^{\overline{a}}\mathbf{S}\right]\mathbf{C}^{\mathrm{a}} = \mathbf{S}\mathbf{C}^{\mathrm{a}}\mathbf{E}^{\mathrm{a}} \qquad a = 1, 2, ..., N_a \tag{3}$$

Here a refers to an active group, N_a is the number of active groups and $\mathbf{R}^{\overline{a}}$ projects to all orbitals not in group a. As the equations are coupled via the the projectors they have to be solved repeatedly until self-consistency. In the present contribution the application of the Huzinaga equation is restricted to the use of a single active group. As active groups have to be orthogonal to all other (active and frozen) groups in order to be consistent with the derivation of the Huzinaga equation a common basis set is practical for all active groups and then they are naturally treated as a single group. The use of several active groups is justified in cases where group specific basis sets are used (e.g. orthogonal basis sets obtained by explicit orthogonalization or by the neglect of orbital overlaps).

Frozen groups can use local, group specific basis sets and no orthogonality among the frozen groups is required. Indeed, the frozen orbital approach in a mixed method may include several covalent bonds that cross the subsystem boundary as in applications presented later. Their nonorthogonality does not appear explicitly in Eq. (3) but it affects the calculation of the $\mathbf{R}^{\overline{a}}$ matrix that takes the form of $\mathbf{R}^{\overline{\mathbf{a}}} = \mathbf{C}^{\overline{\mathbf{a}}}(\sigma^{\overline{\mathbf{a}}})^{-1}(\mathbf{C}^{\overline{\mathbf{a}}})^{\dagger}$, where $\sigma^{\overline{\mathbf{a}}}$ is the overlap matrix of molecular orbitals not in group a.

Geometry optimization of a system with molecular orbitals satisfying the Huzinaga equation can be efficiently performed by calculating the forces acting on atoms. Formulas for the forces are derived in the Appendix. Several terms of the forces do agree with those of a system with canonical orbitals and few terms are different. Existing codes can be easily adapted to calculate forces for systems with Huzinaga orbitals.

FROZEN ORBITALS AND INTERACTIONS WITH THE ENVIRONMENT

Frozen orbitals are taken from calculations for model molecules, for example for a carbon-carbon bond in apolar environment it can be obtained from a calculation

performed for ethane; A standard Hartree-Fock-Roothaan calculation with the relevant basis is followed by orbital localization[36] and the omission of the tails, i.e. keeping only the renormalized coefficients on the two carbon atoms. All other electrons of atom B in Figure 1 are parts of the central QM system. The situation is different for atom A. Core electrons can be optimized or they can be kept fixed at values taken from model calculations. A further possibility is that they do not appear explicitly but they decrease the positive charge of the nucleus of A. All these possibilities are consistent with a QM/MM system. In the calculation presented core electrons were treated explicitly. The use of both fixed and optimized core orbitals were tested similarly to the treatments reported in refs.[37] and[38], respectively. As no benefit from optimization was observed core orbitals were fixed in all calculations reported. Further electrons on atom A can be assigned to the bonds directing towards atoms in the environment. Their treatment is described later.

The present contribution concentrates on the proper description of the central QM subsystem in the frozen orbital approach. The treatment of the environment can be either at a QM or MM level and the choice affects the parametrization of the interaction energy terms at the boundary. In the calculations presented below the environment is represented by point charges that allows the inclusion of electrostatic and polarization interactions. Further terms are to be included for a more complete description of the system, but the applied point charge model can be considered as a part of any more elaborate description. We also note that the use of charges reproducing the electrostatic potential of the wave-function can model a QM/QM treatment and is also compatible with a QM/MM treatment using the AMBER force field[39]. The polarization of the MM region can also be accounted for in this scheme, e.g. with the model proposed in ref.[40], although this option was not investigated in the current work.

Point charges are placed at the atomic positions. Multipole derived atomic charges[41,42] were used. These charges reproduce the electrostatic potential of

distributed multipoles[43] obtained from the wave-function and in this respect they are similar to potential derived charges. The calculation of multipole derived charges is fast and advantageous as it results in well defined charges even in the interior of the molecules and charges reflect the symmetry of the molecule.

In all calculations presented below the frontier atom A was chosen to be an sp^3 carbon with a single localizable bond directed towards the QM system and three bonds towards the MM system. Thus explicit electrons on atom A are the core electron pair and the valence electron contributing to the frozen bond. Accordingly, as a first attempt, the core charge of atom A was set to be equal to the multipole derived charge obtained for A increased by +3. It was observed, however, that the iterative solution of Eq. (3) led to molecular orbitals with negative eigenvalues in the frozen space. As it was discussed above this happens when the frozen orbital space is not a good approximation to a subspace of the occupied orbitals of the Fock operator. This situation arises from the conflict between the near to +3 core charge of atom A and the frozen orbital taken from a bond of sp^3 carbon-atoms. Indeed, it was found that the expected behavior of Eq. (3) can be restored by an increased core charge that is compensated by placing negative charges in the middle of the bonds of atom A directed towards atoms in the MM region (Figure 1). These latter charges will be called bond charges. The details of the charge determination is presented later.

It is worth noting that the compensation for the unphysical environment at the boundary by the combination of frozen orbitals and point charges was applied in related methods. Philipp and Friesner[25] placed a point charge at the midpoint of the frozen bond and compensating charges on the nearby MM atoms to reproduce deprotonation energies. Ferré *et al.*[24] increased the charge of the frontier C-atom and decreased the charge of the connecting H-atoms in order to improve optimized geometrical parameters and the energy profile of chemical reactions calculated with the LSCF method. The redistributed charge scheme of Lin and Truhlar[44] has also to be mentioned. This link atom scheme places charges on

bonds connecting the boundary atom (the one replaced by a link atom in the QM calculation) to atoms in the MM region.

The use of an increased core charge with bond charges will be further investigated using a related approach that allows the self-consistent determination of the orbitals even when the solution of Eq. (3) fails[45].

SAMPLE CALCULATIONS

In the calculations presented below various properties as obtained by standard Hartree-Fock-Roothaan (HFR) calculations and by methods mimicking QM/MM schemes are compared. The latter calculations include the same system as the reference HFR calculations but a part of the system described quantum chemically and another part by point charges. The QM and point charge systems are connected by strictly localized molecular orbitals. The QM wave-function is calculated by Eq. (3). The aim of the calculations is to investigate how various properties are reproduced by the QM/MM scheme and how the extension of the QM system affects the results.

Calculations to solve Eq. (3) were performed with a locally modified version of GAMESS-US[46]. Multipole derived charges were obtained first by performing a distributed multipole analysis with the program GDMA[43] and then by fitting charges with the program Mulfit[47]. (Frontier atom and MM charges are given as Supplementary material.) The 6-31G* basis set[48] was used in all calculations.

Examples presented first are calculations for the $C_5H_{11}COOH$ molecule. Events at the carboxyl end are investigated and the boundary of the QM/MM system is moved away systematically from the carboxyl group. This procedure allows us an estimation of the effect of QM system size on the quality of calculated properties.

Deprotonation energy

The deprotonation energy of $C_5H_{11}COOH$ was calculated with the standard HFR method as reference and with the QM/MM scheme. The extension of the QM system was systematically varied in the latter calculations. The various separations of the system are shown in Figure 2. As it was discussed above, although the number of explicit electrons of atom A is 3 an increased core charge and compensating negative bond charges were introduced to set up a system compatible with Eq. (3). The core charge was determined with the largest QM subsystem ('cut1' in Figure 2) in a trial and error procedure so that a good reproduction of the deprotonation energy could be achieved. The bond charges were adjusted so that the total charge be equal to the multipole derived charge of atom A. The total charge includes the three equal bond charges, the core charge and a charge of -3 the latter accounts for the two core electrons and the single valence electron by which A contributes to the SLMO. It was found that a core charge of +5.6 results in a reproduction of the deprotonation energy with an error of 0.2 kcal/mol. Note that multipole derived charges were calculated for the neutral molecule and were applied also for the deprotonated one, as well. The core charge of +5.6 and bond charges as required by the multipole derived charge of atom A in the actual molecule were then applied in all subsequent calculations.

Deprotonation energies are presented in Table 1. The calculations include geometry optimizations for both the protonated and deprotonated molecules. In the QM/MM geometry optimizations all QM atoms including atom B were optimized and all MM atoms and atom A at the frontier (see Figure 1) were kept fixed at positions obtained in the reference full QM optimization. The RMS difference in geometrical parameters with respect to the reference optimized HFR structures are also presented in Table 1.

The error in deprotonation energies is very small for the largest system (cut1) and its absolute value increases with decreasing QM system size. The error is between 1 and 2 kcal/mol for systems where the protonation site is separated

by 4-6 Å from the midpoint of the frozen bond. The geometrical parameters are very well reproduced showing that even in this sensitive deprotonation test optimized molecular geometries are hardly affected by the approximations.

It is worth emphasizing that the reproduction of the deprotonation energy is a highly sensitive test as it includes species with different charges and thus it is highly dependent on long range electrostatic interactions. Indeed, deprotonation energies were found to depend significantly on the core charge of the frontier atom. On the other hand, in most applications the total charge of the systems compared do not change and the choice of the core charge is expected to have a modest effect on the calculated properties. Calculations of other properties in the subsequent sections were performed without attempting to adjust the core charge of the frontier atom and thus a charge of +5.6 was used throughout.

Conformational energetics

The energy of the $C_5H_{11}COOH$ molecule was calculated as the function of the rotation of the carboxyl group as defined by the dihedral angle O_2-C_3-C_5-C_8 (Figure 2). The rest of the molecule is in the fully extended conformation. The separations of the system agree with those in the deprotonation test except that 'cut4' was excluded as it would require the inclusion of MM torsional potential. The energy curves obtained in the standard reference calculation and in those performed for the three different system separations are presented in Figure 3. Perfect reproduction of the reference with 'cut1' and 'cut2' obtained; the difference is within 0.1 kcal/mol throughout the whole dihedral angle range. The overall shape is reproduced with 'cut3', as well, with a maximal difference slightly over 0.5 kcal/mol. We note, that no MM van der Waals potential is included in these calculations, although it may have a small effect for 'cut3' where the separation between the rotating O-atoms and the closest MM atom is below 5 Å.

These results are highly encouraging as they show that the selection of a small central subsystem is appropriate to well describe the conformational en-

ergy change. A qualitatively correct description is obtained in all cases and the reproduction of the energy curve is perfect when at least 2 bonds separate the rotating bond and the SLMO at the QM/MM boundary.

The second example of conformational energetics is the rotation of the imidazol group in a Gly-His-Gly tripeptide. The QM system includes the central histidine extended with the amide bonds (Figure 4). Note that excluding the bond adjacent to the rotating bond, the chosen boundaries are the nearest ones having sp^3 carbon atoms as frontier atoms in the SLMO (*cf.* Figure 1). With this choice it is reasonable to use the core charge for the frontier atom set in the deprotonation study of the $C_5H_{11}COOH$ molecule.

The energy of the reference full molecule QM calculation and that of the QM/MM system is shown in Figure 4. The two curves are superimposed at their lowest energy values. The QM/MM curve well follows the reference and reproduces the three minima of the latter. The two higher energy minima are shifted by about 1 kcal/mol lower with respect to the absolute minimum in the QM/MM curve. On the other hand, their relative heights on the QM/MM curve is practically equal to that in the reference.

The rotating bond and the SLMO are separated by 3 bonds and this corresponds to 'cut1' of the $C_5H_{11}COOH$ system. This latter gave a perfect description of the energy curve. By contrast, the error of 1 kcal/mol in a region of the energy curve in the Gly-His-Gly system is likely to be related to the spatial proximity ($\sim$5 Å) of the imidazole group to MM atoms (*cf.* Figure 4). Further refinement of the description is expected by the inclusion of van der Waals terms between the QM and MM atoms.

Proton transfer energy curve

The transfer of the side chain COOH proton from Asp to His is investigated. The arrangement of the two molecules and the QM and MM systems are shown in

Figure 5. The boundaries were chosen so that the frozen SLMOs bind the C_α and C_β atoms. This is an advantageous choice since these are the furthest boundaries that allow the use of a single SLMO in each molecule. Moreover, these bonds are expected to be similar to the C-C SLMO in the $C_5H_{11}COOH$ deprotonation example for which the core charge of frontier atom A has been set.

The proton transfer energy curves are shown in Figure 5. Both the reference QM and the QM/MM curves have their minimum at the 1 Å proton - Asp oxygen separation and these points were used to superimpose the two curves. The curves raise to a maximum near 1.5 Å. They are practically indistinguishable in the sharply raising region and a small difference of less than 1 kcal/mol appears at the maximum. The position of the second minimum - corresponding to the protonated His - is perfectly reproduced by the QM/MM calculation while a slightly increased difference (1 kcal/mol) is observed at this minimum.

The description of the proton transfer curve by the QM/MM calculation is highly satisfactory. The minimum energy positions of the proton (Asp-H...His *vs.* Asp...H-His) are perfectly reproduced and the corresponding 27.2 kcal/mol energy difference is reproduced with an error of 1.3 kcal/mol. The transition between these two positions is also well described. It is worth noting that the SLMOs connecting the QM and MM systems are separated by only 2 bonds from the atoms connecting to the proton. The use of such a small size QM system is computationally beneficial. Moreover, the separation using C_α as a frontier atom is expected to be generally advantageous in calculations for proteins, a major field of application of QM/MM methods.

CONCLUSION

The application of the Huzinaga equation[32] in mixed methods with frozen orbitals is proposed. This equation guaranties orthogonality of the optimized molecular orbitals to the frozen orbitals. The basis set of the optimizable orbitals has to be chosen so that orthogonality to the frozen orbitals could be achieved[35].

This can be realized by a basis set that includes the basis of the frozen orbitals. Thus the explicit orthogonalization of the basis set to the frozen orbitals is not required. This approach is also valid when several nonorthogonal frozen orbitals are present.

A QM/MM type scheme with molecular orbitals from the Huzinaga equation, strictly localized frozen orbitals at the boundary and point charges in the MM system was implemented. This scheme requires system specific parameters that include SLMO coefficients taken from calculations for model systems and multipole derived charges[41,42]. They both can be obtained with a reasonable effort. It was found that an additional parameter, the frontier atom core charge together with compensating bond charges, is required to achieve a self-consistent solution and a good reproduction of reference properties. It is important to emphasize that in all calculations the same orbital coefficients and core charges were used, although the boundary carbon-carbon bonds have fairly different environments in the systems investigated. This suggests a good transferability of these parameters and their wide range applicability e.g. in biochemical systems. The implementation focuses on the central QM system and is intended to serve as a framework for mixed QM/MM and QM/QM methods. Further parametrization is expected to be necessary (e.g. for force field parameters including interactions with the frontier atoms) if changes in the environment is also to be taken into account.

The accuracy of calculated properties as a function of the central QM systems was studied. The deprotonation energy in the $C_5H_{11}COOH$ molecule was reproduced with an error of 1-2 kcal/mol as far as the protonated O-atom is separated from the frozen bond by at least 3 C-C bonds corresponding to a distance over 4 Å. Optimized geometries do not exhibit significant deviations from those obtained in reference calculations and this appears to be independent from the QM system size. The energy curve obtained by rotating the carboxyl group in the same molecule was also calculated with various QM system sizes. Perfect reproduction of the reference energy curve was found when the rotating and frozen

bonds are separated by at least 2 C-C bonds. The rotation of the imidazole group in the Gly-His-Gly was also investigated. The QM system with a rotating bond separated by 3 bonds from the two frozen bonds resulted in a highly satisfactory description of the energy curve. The positions of the minima are in perfect agreement with the reference results. The maximal error in the energy difference between two minima amounts to 1 kcal/mol that is attributed to the spatial proximity of the imidazole group and one of the frozen bonds. The transfer of the side chain proton from an aspartic acid to a histidine was also studied. A small QM system was used in which the atoms directly participating in the proton transfer were separated from the frozen bonds by 2 bonds. The shape of the proton transfer curve was perfectly reproduced and the energy difference between the two minima is obtained with an error of 1 kcal/mol.

All these results suggest that the wave-functions as obtained with the Huzinaga equation in the field of frozen orbitals and multipole derived charges give a highly satisfactory description of a large system at a reduced computational cost. The presented implementation focuses on the description of the central system in the field of a constant environment. It forms an appropriate basis to implement QM/QM and QM/MM schemes with frozen orbitals[23–29].

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.

Additional Supporting Information may be found in the online version of this article.

APPENDIX

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

The energy of a closed shell system is written as

$$E = \frac{1}{2}\sum_{\alpha\beta} P_{\alpha\beta}(h_{\beta\alpha} + F_{\beta\alpha}) \tag{4}$$

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.

The molecular orbitals are divided into groups, each group having its own basis set. The orbital coefficients of frozen groups are kept fixed at certain predefined values, while the coefficients of orbitals in active groups are calculated from Eq. (3). Thus orbitals of an active group form an orthonormal set. Moreover, they are orthogonal to all orbitals of all other groups in line with the derivation of the Huzinaga equation[32,35]. In typical applications this is achieved by defining a single active group with a basis that includes all frozen basis orbitals. By contrast, several frozen groups can be present and they are not required to be orthogonal to each other. This is the case in the applications presented above, where the frozen core(s) and SLMO(s) are overlapping. Nevertheless, in the following derivation all frozen orbitals are considered as a single group of nonorthogonal orbitals. The advantage of such a treatment is that in this way there are two orthogonal groups present and the nonorthogonality of the frozen orbitals can be taken into account by using the appropriate form for its density ($\mathbf{P^f}$) and for the matrix ($\mathbf{R^f}$) projecting to the frozen group: $\mathbf{P^f} = 2\mathbf{R^f} = 2\mathbf{C^f}(\sigma^\mathbf{f})^{-1}(\mathbf{C^f})^\dagger$. $\mathbf{C^f}$ is the coefficient matrix and $\sigma^\mathbf{f}$ is the overlap matrix of the frozen molecular orbitals.

We write the derivative with respect to a nuclear coordinate q_i as

$$\frac{\mathrm{d}E}{\mathrm{d}q_i} = A + B \tag{5}$$

with

$$A = \sum_{\alpha\beta} P_{\alpha\beta}\frac{\partial h_{\alpha\beta}}{\partial q_i} + \frac{1}{2}\sum_{\alpha\beta\gamma\delta} P_{\alpha\beta}P_{\gamma\delta}\left(\frac{\partial\langle\alpha\beta|\gamma\delta\rangle}{\partial q_i} - \frac{1}{2}\frac{\partial\langle\alpha\beta|\delta\gamma\rangle}{\partial q_i}\right) \tag{6}$$

and

$$B = \sum_{\alpha\beta} \frac{\mathrm{d}P_{\alpha\beta}}{\mathrm{d}q_i} F_{\alpha\beta} \tag{7}$$

A is well known in the formula for the derivative of the canonical MOs. Owing to the group orthogonality B can be written as a sum of group contributions

$$B = 2tr \left\{ \frac{\mathrm{d}R^a}{\mathrm{d}q_i} F + \frac{\mathrm{d}R^f}{\mathrm{d}q_i} F \right\} \tag{8}$$

where a refers to the active group, and tr stands for the trace of the matrix in curly braces. The separate treatment of active and frozen groups is justified by the different form of their projectors.

Let's first consider the first term of Eq. (8) that contains the contributions from the active group.

$$4tr \left\{ \frac{\mathrm{d}(C^a)^\dagger}{\mathrm{d}q_i} F C^a \right\} = 4tr \left\{ \frac{\mathrm{d}(C^a)^\dagger}{\mathrm{d}q_i} \left(S C^a E^a + S R^f F C^a \right) \right\} \tag{9}$$

where the right hand side is obtained by invoking Eq. (3).

The two terms in Eq. (9) can be handled separately. Making use of the orthonormality relation for group a one can write

$$\frac{\mathrm{d}(C^a)^\dagger}{\mathrm{d}q_i} S C^a + (C^a)^\dagger \frac{\mathrm{d}S}{\mathrm{d}q_i} C^a + (C^a)^\dagger S \frac{\mathrm{d}C^a}{\mathrm{d}q_i} = 0 \tag{10}$$

Then the first sum in the right hand side of Eq. (9) can be shown to take the form

$$-2tr \left\{ \frac{\mathrm{d}S}{\mathrm{d}q_i} C^a E^a (C^a)^\dagger \right\} \tag{11}$$

which includes the usual energy weighted density matrix, now for the active orbitals only and with the eigenvalues of the Huzinaga matrix.

The second term in Eq. (9) can be manipulated using the derivative of the orthogonality relation between groups a and f to yield

$$-2tr \left\{ \frac{\mathrm{d}S}{\mathrm{d}q_i} \left(R^f F R^a + R^a F R^f \right) \right\} \tag{12}$$

The only term remains to be treated is the second one of Eq. (8) that includes the projector of the nonorthogonal frozen orbitals. Since

$$\frac{\partial \sigma^{-1}}{\partial q_i} = -\sigma^{-1} C^+ \frac{\partial S}{\partial q_i} C \sigma^{-1} \tag{13}$$

$2tr\left\{\frac{\mathrm{d}R^f}{\mathrm{d}q_i}F\right\}$ can be written as

$$-2tr\left\{\frac{\mathrm{d}S}{\mathrm{d}q_i}R^f F R^f\right\} \tag{14}$$

Making use of Eqs. (11,12,14), B of Eq. (5) takes the form

$$B = -2tr\left\{\frac{\mathrm{d}S}{\mathrm{d}q_i}\left(C^a E^a (C^a)^\dagger + R^f F R^f + R^f F R^a + R^a F R^f\right)\right\} \tag{15}$$

This makes the expression for the energy derivative complete. Note, that the above derivation assumes that frozen orbital coefficients are independent from the molecular geometry. See ref.[24] for handling the variation of the frozen orbital coefficients with changes in atomic positions.

References

1. Schmidt, T. C.; Paasche, A.; Grebner, C.; Ansorg, K.; Becker, J.; Lee, W.; Engels, B. *Top. Curr. Chem.* **2012,** DOI: 10.1007/128_2011_309.

2. Acevedo, O.; Jorgensen, W. L. *Acc. Chem. Res.* **2010,** *43,* 142-151.

3. Mata, R. A. *Phys. Chem. Chem. Phys.* **2010,** *12,* 5041-5052.

4. Senn, H. M.; Thiel, W. *Angew. Chem. Int. Ed.* **2009,** *48,* 1198-1229.

5. Kamerlin, S. C. L.; Haranczyk, M.; Warshel, A. *J. Phys. Chem. B* **2009,** *113,* 1253-1272.

6. Hu, H.; Yang, W. *Annu. Rev. Phys. Chem.* **2008,** *59,* 573-601.

7. Kamerlin, S. C. L.; Vicatos, S.; Dryga, A.; Warshel, A. *Annu. Rev. Phys. Chem.* **2011,** *62,* 41-64.

8. Kitaura, K.; Ikeo, E.; Asada, T.; nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999,** *301,* 701-706.

9. Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. A* **2007,** *111,* 6904-6914.

10. Yang, W. *Phys. Rev. Lett.* **1991,** *66,* 1438-1441.

11. Yang, W.; Lee, T.-S. *J. Chem. Phys.* **1995,** *103,* 56745678.

12. Dixon, S. L.; Merz, K. M. *J. Chem. Phys.* **1996,** *104,* 66436649.

13. He, X.; Merz, K. M. *J. Chem. Theory Comput.* **2010,** *6,* 405411.

14. Wesołowski, T. A.; Warshel, A. *J. Phys. Chem.* **1993,** *97,* 80508053.

15. Wesołowski, T. A. *Phys. Rev. A* **2008,** *77,* 012504.

16. Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000,** *104,* 1720-1735.

17. Antes, I.; Thiel, W. *J. Phys. Chem. A* **1999,** *103,* 92909295.

18. Zhang, Y.; Lee, T.-S.; Yang, W. *J. Chem. Phys.* **1999,** *110,* 4654.

19. Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976,** *103,* 227-249.

20. Náray-Szabó, G.; Surján, P. *Chem. Phys. Lett.* **1983,** *96,* 499-501.

21. Ferenczy, G.G.; Rivail, J. L.; Surján, P. R.; Náray-Szabó, G. *J. Comput. Chem.* **1992,** *13,* 830-837.

22. Théry, V.; Rinaldi, D,; Rivail, J. L.; Maigret, B.; Ferenczy, G.G. *J. Comput. Chem.* **1994,** *14,* 269-282.

23. Assfeld, X.; Rivail, J. L. *Chem. Phys. Lett.* **1996,** *263,* 100-106.

24. Ferré, N.; Assfeld, X.; Rivail, J.-L. *J. Comput. Chem.* **2002,** *23,* 610-624.

25. Philipp, D. M.; Friesner R. A. *J. Comput. Chem.* **1999,** *20,* 1468-1494.

26. Murphy, R. B.; Philipp, D. M.; Friesner R. A. *J. Comput. Chem.* **2000,** *21,* 14421457.

27. Gao, J.; Amara, P.; Alhambra C.; Field, M. J. *J. Phys. Chem.* **1998,** *102,* 4714-4721.

28. Pu, J.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. A* **2004,** *108,* 632650.

29. Jung, J.; Choi C. H.; Sugita, Y.; Ten-no, S. *J. Chem. Phys.* **2007,** *127,* 204102.

30. Jung, J.; Sugita, Y.; Ten-no, S. *J. Chem. Phys.* **2010,** *132,* 084106.

31. Kitagawa, Y.; Akinaga, Y.; Kawashime, Y.; Jung, J.; Ten-no, S. *J. Chem. Phys.* **2012,** *401,* 95-102.

32. Huzinaga, A.; Cantu, A. A. *J. Chem. Phys.* **1971,** *55,* 5543-5549.

33. Sakai, Y.; Miyoshi, E.; Klobukowski, M.; Huzinaga, S. **1987,** *J. Comput. Chem. 8,* 226-255.

34. Sakai, Y.; Miyoshi, E.; Klobukowski, M.; Huzinaga, S. **1987,** *J. Comput. Chem. 8,* 256-264.

35. Ferenczy, G. G.; Adams, W. H. *J. Chem. Phys.* **2009,** *130,* 134108-134119.

36. Pipek, J.; Mezey, P. G. *J. Chem. Phys.* **1989,** *90,* 4916-4926.

37. Fornili, A.; Loos, P.-F.; Sironi, M.; Assfeld X. *Chem. Phys. Lett.* **2006,** *427,* 236-240.

38. Loos, P.-F.; Fornili, A.; Sironi, M.; Assfeld X. *Comput. Lett.* **2007,** *3,* 473-486.

39. Ponder, J. W.; Case, D. A. *Adv. Prot. Chem.* **2003,** *66,* 27-85.

40. Illingworth,C. J. R.; Parkes, K. E. B.; Snell, C. R.; Ferenczy, G. G.; Reynolds, C. A. *J. Phys. Chem. A* **2008,** *112,* 1215112156.

41. Ferenczy, G. G. *J. Comput. Chem.* **1991,** *12,* 913-917.

42. Winn, P. J.; Ferenczy, G. G.; Reynolds, C. A. *J. Phys. Chem. A* **1997,** *101,* 5437-5445.

43. Stone, A. J. *J. Chem Theory Comp.* **2005,** *1,* 1128-1132.

44. Lin, H.; Truhlar, D. G. *J. Phys. Chem. A* **2005,** *109,* 3991-4004.

45. Ferenczy G. G. *submitted for publication.*

46. Schmidt, M.W.; Baldridge, 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.

47. 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

48. Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972,** *56,* 2257-2261.

Table 1: QM/MM errors in deprotonation energies and in geometrical parameters for the $C_5H_{11}COOH$ molecule with various system separations (see Figure 2). Reference is the Hartree-Fock-Roothaan results for the full molecule.

System	O-SLMO distance[a] (Å)	No of separating bonds[b]	ΔE error (kcal/mol)	RMS error in bond length (Å)	RMS error in bond angle (°)	RMS error in torsion angle (°)
cut1	6.8	5	-0.17	0.000	0.5	0.1
cut2	5.6	4	-1.43	0.001	0.4	0.3
cut3	4.3	3	-1.58	0.003	0.4	0.7
cut4	3.0	2	-3.78	0.003	0.6	0.2

[a]Distance between the O-atom holding the proton and the midpoint of the bond with SLMO

[b]Number of bonds between the O-atom holding the proton and the bond with SLMO

FIGURE CAPTIONS

Figure 1: Separation of a system into subsystems. A strictly localized molecular orbital connects atoms A (frontier atom) and B. Other bonds of atom B are directed towards atoms in the central (QM) system and other bonds of atoms A are directed towards atoms in the environment (MM system). Bond charges (q2) are negative charges mimicking electrons and are placed at the midpoint of the bonds connecting frontier atom A and atoms in the MM system. The positive charge on the core of atom A is q1 and was set to 5.6 in all calculations presented. The net charge of 3×q2+q1-3 (-3 accounts for the 2 core electrons and the valence electron in the SLMO) is equal to the multipole derived charge for atom A.

Figure 2: Various separations of the $C_5H_{11}COOH$ molecule into subsystems. Dashed lines cross the SLMOs at the QM/MM boundary. They are closer to atom A than to atom B.

Figure 3: Energy of the $C_5H_{11}COOH$ molecule as a function of the rotation of the -COOH group. Dihedral angle of rotation is indicated by lines parallel to bonds. System separations are shown by dashed lines.

Figure 4: Energy of the Gly-His-Gly molecule as a function of the rotation of the histidine group. System separation is also indicated.

Figure 5: Energy of the Asp - His system as a function of the separation of the proton from the oxygen of Asp. System separation is also indicated.

FIGURES

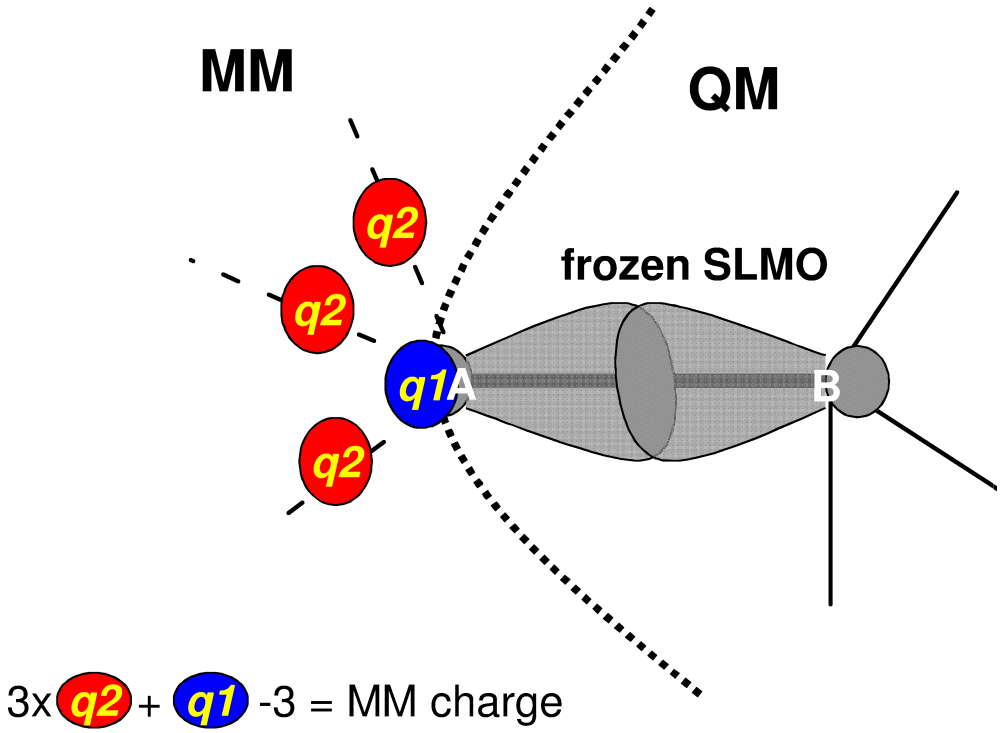

Figure 1
G. G. Ferenczy
J. Comput. Chem.

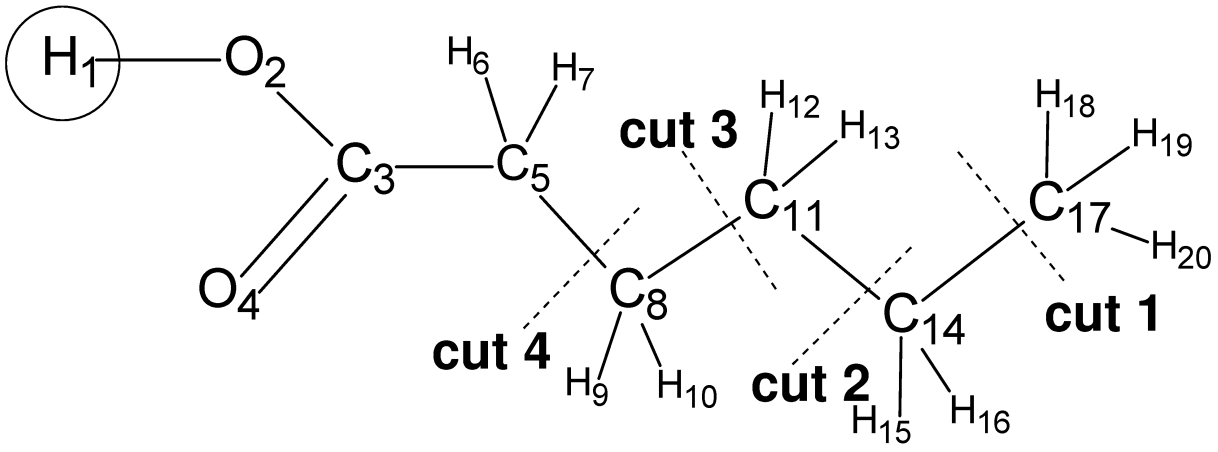

Figure 2
G. G. Ferenczy
J. Comput. Chem.

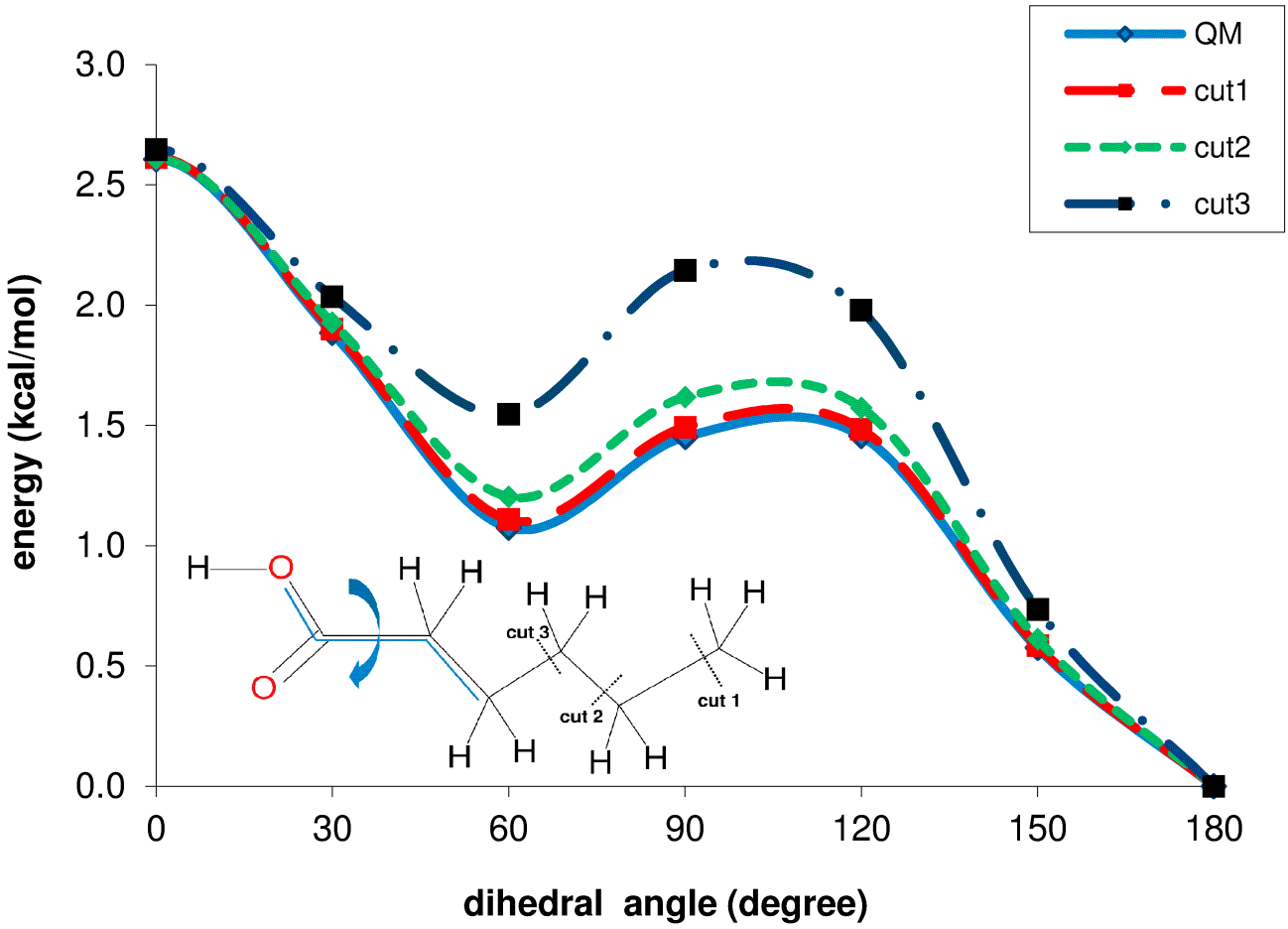

Figure 3
G. G. Ferenczy
J. Comput. Chem.

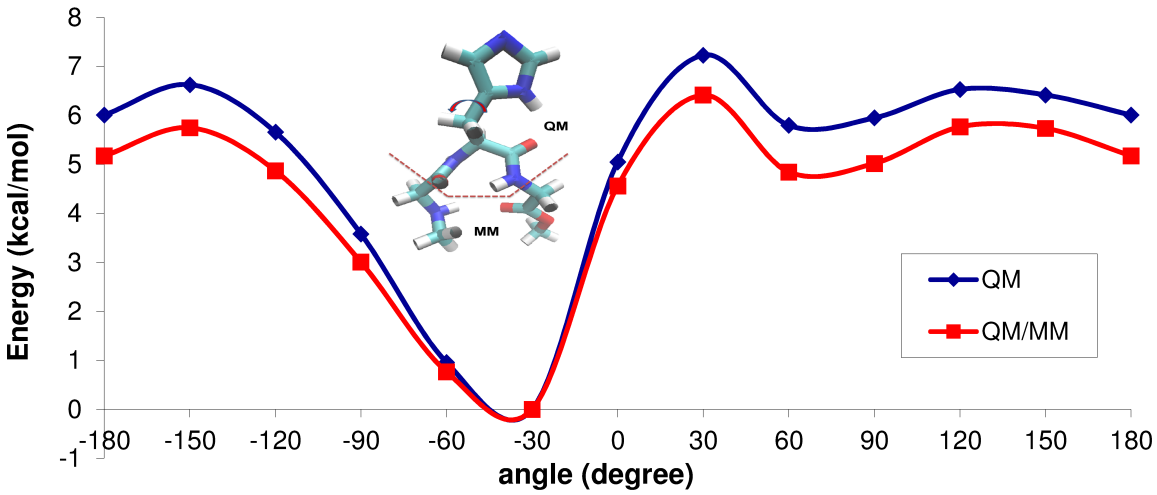

Figure 4
G. G. Ferenczy
J. Comput. Chem.

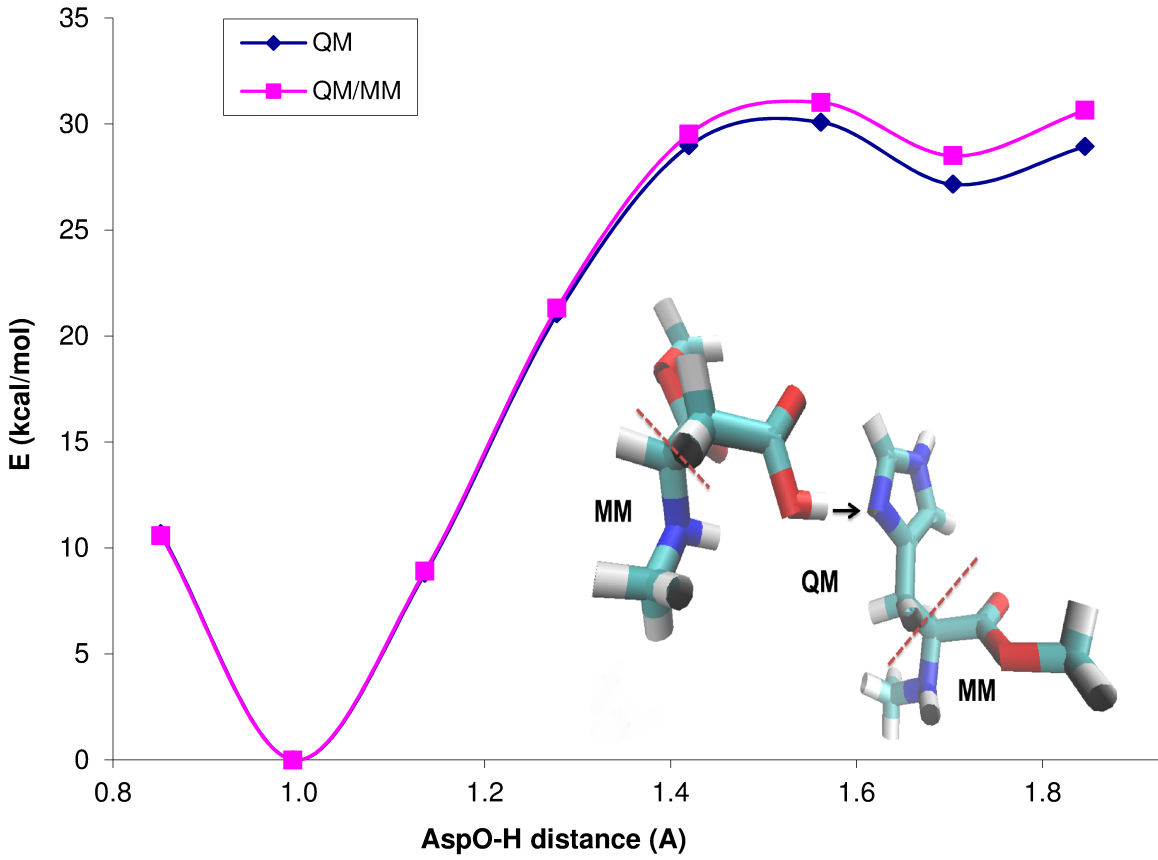

Figure 5
G. G. Ferenczy
J. Comput. Chem.