A Mixed Quantum Mechanics/Molecular Mechanics (QM/MM) Method for Large-Scale Modeling of Chemistry in Protein Environments

R. B. MURPHY,[1] D. M. PHILIPP,[2] R. A. FRIESNER[3]

[1] *Schrodinger Inc., 1 Exchange Place, Jersey City, New Jersey 07302*
[2] *Beckman Research Institute, Caltech, Pasadena, California*
[3] *Columbia University, Department of Chemistry, New York City, New York*

Received 30 April 2000; accepted 13 July 2000

ABSTRACT: A QM–MM method, using our previously developed frozen orbital QM–MM interface methodolgy, is presented as a general, accurate, and computationally efficient model for studying chemical problems in a protein environment. The method, its parametrization, and a preliminary application to modeling cytochrome P-450 chemistry are presented. © 2000 John Wiley & Sons, Inc. J Comput Chem 21: 1442–1457, 2000

Keywords: QM/MM; density functional theory (DFT); enzyme chemistry; P450

Introduction

The development of accurate and reliable methods for mixing quantum mechanical and molecular mechanics (QM/MM) representations of a chemical system has been a major goal of computational chemistry over the past decade.[1–10] Such a representation allows one to treat a small portion of a large systems accurately (e.g., permitting modeling of reactive chemistry within that region) while incorporating the remainder of the system at a reasonable level of precision without the rapid growth in computational cost associated with a fully quantum mechanical treatment.

The key problem in defining a suitable QM/MM representation is the construction of an accurate interface between the QM and MM regions of the system. A number of different approaches have been pursued by various research groups; while significant progress has been made, the problem is a very difficult one, and development of an automated methodology that reproduces fully quantum mechanical calculations with negligible errors has been difficult to accomplish. Consequently,

Correspondence to: R. A. Friesner; e-mail: rich@chem.columbia.edu

Contract/grant sponsor: NIH
Contract/grant sponsor: Institute of General Medical Sciences; contract/grant number: GM-40526
Contract/grant sponsor: Division of Research Resources via the Columbia Center for Biomolecular Simulation; contract/grant number: P41-RR06892

QM/MM methodology remains an important active area of computational chemistry research.

In previous articles,[10, 11] we have presented a QM/MM approach for *ab initio* quantum chemical calculations based on the use of localized, frozen molecular orbitals as the interface between the QM and MM regions. Our approach differs from others in the literature in a number of important methodological details, but more generally in that extensive parametrization of the interface for each chemical group to be modeled, based on fitting to quantum chemical data for a model small molecule, is mandated. This parametrization requires more initial effort but, in our view, yields results that are more stable and accurate. Our QM/MM methodology has been implemented by building a tight coupling between the Jaguar suite of *ab initio* electronic structure programs[12] and the IMPACT code of Levy and coworkers; the OPLS-AA protein force field of Jorgensen and coworkers[13] is used for the molecular mechanics part of the calculations. Results were presented for both Hartree–Fock (HP) and density functional (DFT) quantum chemical models. Small peptide test cases demonstrated that the methodology provided good results for both deprotonation energies in the QM region and relative conformational energetics; the high polarity of the amide group makes QM/MM modeling of peptides a challenging problem (e.g., compared to a system primarily composed of hydrocarbons).

As mentioned above, the approach we have taken involves considerable initial effort due to the parametrization required for each functional group to be studied. Such a methodology will be useful to others only when it has been automated to model an important class of chemical systems, via a library of frozen orbitals and interface parameters that can be easily deployed to study the particular system of interest. We have chosen to focus here on the modeling of protein active sites; the study of reactive chemistry in the active site region of enzyme is of great importance in biochemistry. Our strategy has been to develop QM/MM parameters specific to the 20 amino acids, under the assumption that cofactors, metal ions, ligands, and other species present in the active site will be treated at the QM level. Indeed, a major motivation of the approach we have taken is to avoid construction of a molecular mechanics force field for these systems, which is often a highly labor intensive undertaking (and, in the case of metals or of reactive species in general, of questionable accuracy, at least with currently available force field technology).

We present in this article a complete QM/MM parameter set for the 20 normal amino acids in which the QM/MM interface can be defined either in the backbone (between the alpha-carbon and the amide group) or between the backbone and the side chain (i.e., between the alpha-carbon and the beta-carbon). Both DFT and Hartree–Fock parameter sets have been developed and are discussed below. A local MP2 version of the theory has been implemented but not parametrized. The interface parameters are developed by fitting fully QM data for blocked dipeptides. Beginning with molecular mechanics approximations to dipeptide rotamer states, we have developed a high-quality QM database for fitting containing several thousand data points in all, thus permitting an accurate rendering of each important torsional surface. Furthermore, we have constructed a second database consisting of hydrogen bonded pairs of small molecule side-chain analogues. A total of ~200 such pairs is used to determine van der Waals radii for atoms in the QM region. These data sets are considerably more extensive, and of higher quality, than any that have been used previously in developing or testing QM/MM models of peptides and proteins.

Parameter development, however, is only one aspect of building a QM/MM approach powerful enough to treat enzyme active site chemistry realistically in reasonable amounts of CPU time. We have invested an equal amount of effort in optimizing algorithms for the QM/MM calculations themselves, focusing in the present work on geometry optimization as the initial applications of our methodology will be to compute relative energetics of proposed reactive intermediates in active site catalytic processes. A critical technique is to adiabatically minimize the MM region after each QM geometry step; without this, the number of QM steps would become prohibitively large and the calculations out of range of all but the most powerful supercomputers. We have parallelized our code so that reasonable throughput can be achieved for the relatively large (100–200 atom) QM regions that in our view are necessary to reliably model active site reactive chemistry. Finally, we have incorporated a continuum dielectric treatment of aqueous solvation, one that is capable of handling the QM and MM regions simultaneously, and contains an analytical gradient, so that geometry optimizations in solution can be performed. Inclusion of solvation in QM/MM modeling efforts will be critical in some (but not all) applications, an obvious case being calculation of pKas of ionizable protein side-chain and ligand groups.

Given an accurate parametrization and efficient computational algorithms, the third requirement for the methodology to be useful in a practical context is ease of use. Setting up a QM/MM calculation is, in principle, a formidable undertaking, with the requirement for specifying the interface and associated parameters in a complex three dimensional topology. We have developed automated software to perform all of these functions. The user simply specifies the QM regions of the protein in terms of residues and whether the interface is to be placed in the backbone or side chain, and the correct interface parameters are implemented with no human intervention. Because our parametrization is basis set specific, the specified basis set (6-31G* in the present article) must be used in the interface region; however, we have also developed a flexible interface allowing the use of mixed basis sets. For example, one can use a large, high-quality basis set in a reactive region to determine accurate single point energies, while retaining a smaller basis set at the QM/MM interface which, in an optimal implementation, will be located around the periphery of the reactive chemistry. We have used this strategy very successfully in our QM based modeling[14] of the protein methane monooxygenase (MMO), and expect it to work for a wide range of active site modeling efforts. These features are in the process of being implemented in a graphical user interface, so that QM/MM calculations can be prepared via point-and-click methods using a visual representation of the active site.

This article is organized into the following sections. In the next section, we briefly review our QM/MM formalism, which has been described in detail in other publications. In the subsequent section, we discuss our parametrization strategy and present results for the 20 amino acids, including extensive testing of the accuracy of the parameters. The Validation and Testing section describes the algorithmic and software implementation of the methodology, including various ease of use features that have been automated. In the Implementation section, we present an actual application to a metalloprotein active site, that of cytochrome P450cam. Cytochrome P450s play a critical role in the metabolism of all organisms,[15] and is important in pharmaceutical applications due to its impact on toxicity; when accurate structural models of the human enzyme become available (as is anticipated in several years), modeling of these systems will become highly relevant to drug discovery efforts as well as of interest in a basic research context. The results we report here are of a preliminary nature (a more extensive investigation of the catalytic cycle in this system will be presented elsewhere), but are sufficient to demonstrate the power of our new QM/MM approach. Finally, in the conclusion, we discuss future directions.

Review of QM/MM Formalism

REVIEW OF HARTREE–FOCK FORMALISM

Because a detailed presentation of the mathematical formalism for our Hartree–Fock implementation of QM/MM has been given in ref. 10, we shall give only a brief overview of that formalism here. Application of the method begins by specifying QM and MM regions of the molecule; the interface between these regions is mediated by frozen orbitals that we restrict to being defined at single bonds. The frozen orbitals are obtained from small model molecule calculations as described below. The QM and MM regions interact via electrostatics (MM point charges interacting with the QM wave function) and van der Waals terms. The quantum chemical wave function at a specific geometry is determined by solving the Hartree–Fock equations for the QM region using the frozen orbitals in essence as pseudopotentials and in the field of the MM point charges. The energy in the MM region is given by the usual MM energy expression. The total energy of the system is thus formally decribed by the following components:

$$E_{\mathrm{QM}} = \sum_{\mu\nu} P_{mn} H_{\mu\nu}^{\mathrm{core}*} + \frac{1}{2}\sum_{\mu\nu} P_{\mu\nu}[2J_{\mu\nu} - K_{\mu\nu}] + \sum_{AC} \frac{Z_a Z_c}{R_{ac}} + \sum_{AM} \frac{Z_A q_M}{R_{AM}} \quad (1)$$

$$E_{\mathrm{MM}} = \sum_{i=\mathrm{stretches}} k_i (r_i - r_0)^2 + \sum_{j=\mathrm{bends}} k_j (\theta_j - \theta_0)^2 + \sum_{k=\mathrm{torsions}} \frac{V_{k,1}}{2}(1+\cos\phi_k) + \frac{V_{k,2}}{2}(1-\cos 2\phi_k) + \frac{V_{k,3}}{2}(1+\cos 3\phi_k) + \sum_{MN} \frac{q_M q_N}{R_{MN}} + \sum_{MN} 4\varepsilon_{mn}\left[\left(\frac{\sigma_{MN}}{R_{MN}}\right)^{12} - \left(\frac{\sigma_{MN}}{R_{MN}}\right)^{6}\right] \quad (2)$$

Indices M, N refer to MM atoms, Q to quantum atoms. The first E_{QM} is the quantum energy (for a closed shell case) defined by the electronic Hamiltonian including the point charges of the MM atoms whose potential is absorbed into the one electron

Hamiltonian $H^{\text{core}*}$. The quantum region is, therefore, polarized self-consistently by the MM charges. We have implemented the open-shell version of the theory as well. The second term, E_{MM}, is the OPLS-AA MM energy among the MM atoms.

The frozen orbital is obtained straightforwardly from calculations on the model molecule. The molecule is optimized into its lowest energy conformation, and Boys localization is used to generate localized orbitals. The orbital coefficients are truncated so that basis functions are restricted to atoms defined as part of the chemical functionality being parametrized (this is the principal restriction on transferability), and the orbital is then renormalized to have unit electron density. The orbital coefficients and the orientation of the orbital with respect to the QM and MM fragments is stored; when used to build an interface, the orbital is appropriately translated and rotated to be in the correct position.

Double occupation of the frozen orbital places a charge density of two electrons between the QM and MM atom. Half of this density is assignable to the QM atom and completes the approximately spherical charge shell around that atom. However, the description of the charge density around the MM atom terminating the frozen bond is asymmetric in that the frozen orbital explicitly represents the charge density in one specific direction, whereas in the remaining directions the charge distribution is approximated as a (spherical) point charge. This breaking of symmetry leads to qualitatively incorrect results unless corrected for. Our approach is to place a "bond charge" at the midpoint of the frozen bond, which compensates for the charge density of the frozen orbital assignable to the MM side, thus approximately restoring spherical symmetry around the MM atom. The magnitude of this point charge is optimized by fitting to the energies of point charge interactions with the molecule, and then tested by calculating deprotonation energies for various hydrogen locations. As part of this process, charges are also placed on the QM and MM atoms themselves; the sum of these charges is determined by electrical neutrality but the partitioning is determined as part of the optimization process described above.

As will be shown below, the results indicate that the bond charge does very well at correcting the asymmetry introduced by the frozen orbital as long as one is at a sufficient distance (typically ~3 Å or more) from the bond charge. This distance criterion simply indicates that at short distances the electrostatic field from the corrected model is not identical to a fully quantum chemical charge distribution. Improvements are possible via a more sophisticated model for the correction, and we are presently investigating approaches involving a more delocalized distribution of the bond charge, which does appear to reduce the errors for these cases.

Once the bond charge and charges on the QM and MM atoms at the interface are determined, MM like stretch, bend, and torsion parameters in the expression $E_{\text{MM-corr}}$ in eqs. (3)–(4) below have to be optimized to ensure that the conformational energetics around the frozen bond behave correctly. These include torsional, stretching, and bending parameters associated with coordinates within a specified proximity of the frozen bond. Accurate quantum chemical values are obtained for the model molecule by displacements of the relevant coordinates (followed by restrained geometry optimization for the torsional variations), and the parameters are then fit to reproduce the fully QM results using standard least squares methodology. It is also useful in achieving good accuracy for conformational energetics to scale the close lying electrostatic interactions around the frozen bond as is typically done for one to four interactions in standard molecular mechanics implementations; the scaling factors that we use, and the rules for applying them, are presented in ref. 10. The final energy expression for the correction terms is given by:

$$\begin{aligned} E_{\text{elec-corr}} = &-\sum_{MQ}\sigma_{MQ}\frac{q_M q_Q^*}{R_{MQ}} - 2\sum_{M,\,a=frz}\sigma_{Ma}\left\langle a\left|\frac{q_M}{R_{Ma}}\right|a\right\rangle \\ &-\sum_{M,B=b.c.}\sigma_{MB}\frac{q_M q_B}{R_{MB}} - \sum_{Q,B=b.c.}\sigma_{QB}\frac{q_Q^* q_B}{R_{QB}} \\ &-2\sum_{B=b.c.,\,a=frz}\sigma_{Ba}\left\langle a\left|\frac{q_B}{R_{Ba}}\right|a\right\rangle \end{aligned} \tag{3}$$

$$\begin{aligned} E_{\text{MM-corr}} = &\sum_{i=\text{stretches}} k_i(r_i - r_0)^2 + \sum_{j=\text{bends}} k_j(\theta_i - \theta_0)^2 \\ &+\sum_{k=\text{torsions}}\frac{V_{k,1}}{2}(1+\cos\phi_k) \\ &+\frac{V_{k,2}}{2}(1-\cos 2\phi_k) + \frac{V_{k,3}}{2}(1+\cos 3\phi_k) \end{aligned} \tag{4}$$

In these equations q_B denotes the bond charge, and a the frozen orbital. The first term $E_{\text{elec-corr}}$ contains ectrostatic corrections between QM and MM atoms near the frozen orbitals. The function of this correction is to reduce or eliminate short-range electrostatic interactions. This is analogous to the MM practice of neglecting one to two and one to three electrostatic interactions and scaling one to four in-

teractions by 0.5. The σ parameters in this third term are nonzero only within next-neighbor atoms of the frozen orbital. Screening of the short-range portion of the frozen orbital potential is accomplished as well. The second term, $E_{\text{MM-corr}}$, are MM like stretch, bend, and torsion correction terms between MM and QM atoms. The stretch terms are between a pair of atoms where at least one of the two atoms are one of the bond adjacent atoms (atoms adjacent to the frozen orbital atoms), while the bend terms are for all bond angles for which the two bond adjacent atoms are two of the three atoms involved in the bond angle. The torsion terms include all torsion angles where the bond adjacent atoms are two of the four atoms involved. The purpose of these MM-like correction terms is to be able to adjust the potential near the QM–MM boundary to reproduce quantum potential surfaces.

Given the energy expressions[1–4] and parameters for the correction terms, one can derive expressions for analytical gradients. This is quite complicated, due to constraints introduced by the frozen orbital; however, we have carried out the derivation in detail for every term in ref. 10. Here, we simply note that in terms of computational effort, the additional terms in the gradient are negligible, and that we have extensively debugged and tested our code to the point where it performs geometry optimizations reliably and in a number of steps comparable to that required for fully QM optimization. With a gradient methodology at hand, it is possible to optimize structures and thus apply the approach to a wide variety of chemical problems.

DFT FORMALISM

The extension of the Hartree–Fock equations to DFT is simply a matter of using the standard DFT quantum energy expression in place of the Hartree–Fock. In other words, E_{QM} of eq. (1) becomes,

$$E_{\text{QM-DFT}} = \sum_{\mu\nu} P_{mn} H_{\mu\nu}^{\text{core}*} + \frac{1}{2}\sum_{\mu\nu} P_{\mu\nu}\left[2J_{\mu\nu} - K^{xc}\right] + \sum_{AC}\frac{Z_a Z_c}{R_{ac}} + \sum_{AM}\frac{Z_A q_M}{R_{AM}} \quad (5)$$

with K^{xc} the DFT exchange correlation matrix:[16]

$$K_{\mu\nu}^{xc} = \int dr \left(\frac{\partial f_{xc}[\rho]}{\partial \rho}\right)\chi_\mu(r)\chi_\nu(r) \quad (6)$$

It is a simple matter to shown that the QM–MM gradient within DFT is derived with an analogous extension of the Hartree–Fock method. For DFT functionals of the B3-LYP form,[17] K^{xc} will include a HF exchange matrix. We have implemented an open-shell QM–MM DFT method as well. Our pseudospectral implementation of DFT is approximately as fast as the HF code, and even faster for functionals without HF exchange. Thus, there is no significant additional cost in CPU time in using QM–MM DFT compared to HF.

LOCAL MP2 FORMALISM

We have implemented the QM–MM method described above to work with our pseudospectral local MP2 (LMP2) method,[18] but as of this writing have not completed interface parameters for this method. The calculation of the LMP2 correlation energy within this QM–MM scheme is identical to a normal LMP2 calculation with the caveat that the frozen occupied orbital space cannot be excited from. This restriction is simple to implement. We presently have only implemented single-point LMP2 energies, although a gradient implementation is feasible, given that we have LMP2 gradients.

Development of QM/MM Parameters for Protein Active-Site Modeling

OVERVIEW

In this section we present our protocol for developing QM/MM interface parameters. The protocol is identical for both DFT and Hartree–Fock Hamiltonians, and is applied to both quantum chemical methods in an equivalent fashion. In both cases we have developed parameters specifically for the 6-31G* basis set, a reasonable compromise between accuracy and efficiency for geometry optimization. This basis set is not, of course, adequate for computing accurate energy differences for reactive chemistry using DFT methods; large basis sets of at least triple zeta quality are required. Our approach is to employ a mixed basis set for more accurate calculations, 6-31G* at the QM/MM interface, and a large basis set in the reactive region of the molecule. We and others have used such a mixed basis set for large scale DFT calculations with great success, for example[14] in studying the catalytic intermediates of the protein methane monooxygenase (MMO).

A final subtlety involves the question of what level of quantum chemical theory to use in the fully QM calculations to which the QM/MM parameters are fit. At first glance, it would seem natural to fit to the level of QM theory being used in the calculation, for example, HF if the QM region is to be described at the HF level. However, the

assignment of QM van der Waals parameters introduces an additional term into the Hamiltonian that could arguably be said to contain MP2-like information. This applies to the DFT model as well, to which one can also view the QM/MM van der Waals interaction as a supplement. More generally, there is no reason not to use the torsional terms to improve upon an HF or DFT description of conformational energetics, given that one is not in any case dealing with a pure Hamiltonian of any sort. We, therefore, fit both HF and DFT QM/MM parameters using a fully QM data set calculated at the local MP2 (LMPZ) level[18] using the Dunning correlation consistent triple zeta basis[19] without f functions (cc-pVTZ(-f) basis). We have previously shown[18] that LMP2(cc-pVTZ(-f)) provides excellent agreement with experiment for conformational energetics (average deviation ~0.25 kcal/mol for a database of 36 small molecules, and a considerable improvement over either HF or currently available DFT functionals). At the current level of agreement of the QM/MM and QM results, this decision is probably not crucial; one is principally trying to achieve good structures and reasonable conformational energetics in the interface region, as there will be some reliance on cancellation of error. However, as the method is improved and agreement with the fitting data becomes tighter, our belief is that this approach provides the correct path on which to go forward. Note also that in our on-going MM force-field development work we are fitting the MM parameters to the same data set, which will facilitate increased consistency across the interface.

AVAILABLE LOCATIONS OF THE QM/MM INTERFACE

To carry out protein active site modeling, we require a minimum parametrization for the 20 common amino acids (note that unusual amino acids critical to active site functioning can be treated at the QM level if necessary). The first step in formulating a systematic approach to this parameter development effort is to specify the chemical locations at which a QM/MM frozen orbital interface can be placed. Below, we describe parametrization at two different types of locations. First, the interface can be placed in two locations in the peptide backbone; between the alpha-carbon and the carbonyl carbon in the amide group, and between the alpha-carbon and the amide nitrogen (these locations are shown in Figure 1a–e). These parameters sets were developed using the blocked alanine dipeptide as a template molecule, but are transferable to an arbitrary amino acid pair. Second, side-chain specific parameters have been developed, allowing the interface to be constructed between the alpha-carbon and the beta-carbon (Fig. 1e). We have developed parameters of this type only for a subset of amino acids where the side chain is of sufficient length; for cases with very small side chains, an interface of this types suffers from both diminished accuracy (due to interference between the QM side chain and MM backbone) and minimal gains of efficiency compared to the backbone cuts. We again use a blocked dipeptide containing the relevant side chain as the template molecule for this interface. Residues listed in Table I can be treated with side-chain cuts. All residues not listed in Table I (Gly, Ala, Ser, Thr, Pro) can be treated by making two cuts in the backbone between which is the QM side chain, as depicted in Figure 1f.

PARAMETRIZATION OF THE QM VAN DER WAALS RADII

Prior to developing the stretching, bending, torsional, and bond charge parameters for the interface, it is necessary to specify the QM van der Waals parameters, as these have a significant influence on internal hydrogen bonds formed in the template blocked dipeptides. To do this, we have calculated optimized structures and binding energies for ~200 pairs of small molecule ammo acid side-chain and backbone analogues, using a reasonably accurate quantum chemical methodology (HF/6-31G* geometry optimization followed by a LMP2/cc-pVTZ single-point energy calculation. There is a question as to what type of results the QM/MM binding energies should be fit to: in parametrizing MM force fields to be used in the condensed base, for example, gas phase pair binding energies are typically adjusted to be ~15–20% higher than accurate QM values so as to implicitly incorporate the effects of medium polarization in an average fashion. In a hydrogen bonding structure in which one partner is QM and the other MM, a reasonable hypothesis is that one should aim for a binding energy that is in between the QM and MM results, as the QM side contributes explicit polarization, whereas the MM side does not. We have adopted this strategy, using the OPLS-AA results as the MM values (note that the above relation is not always systematically obeyed for OPLS-AA, indicating that irregularities in parametrization protocol are more the rule than the exception).

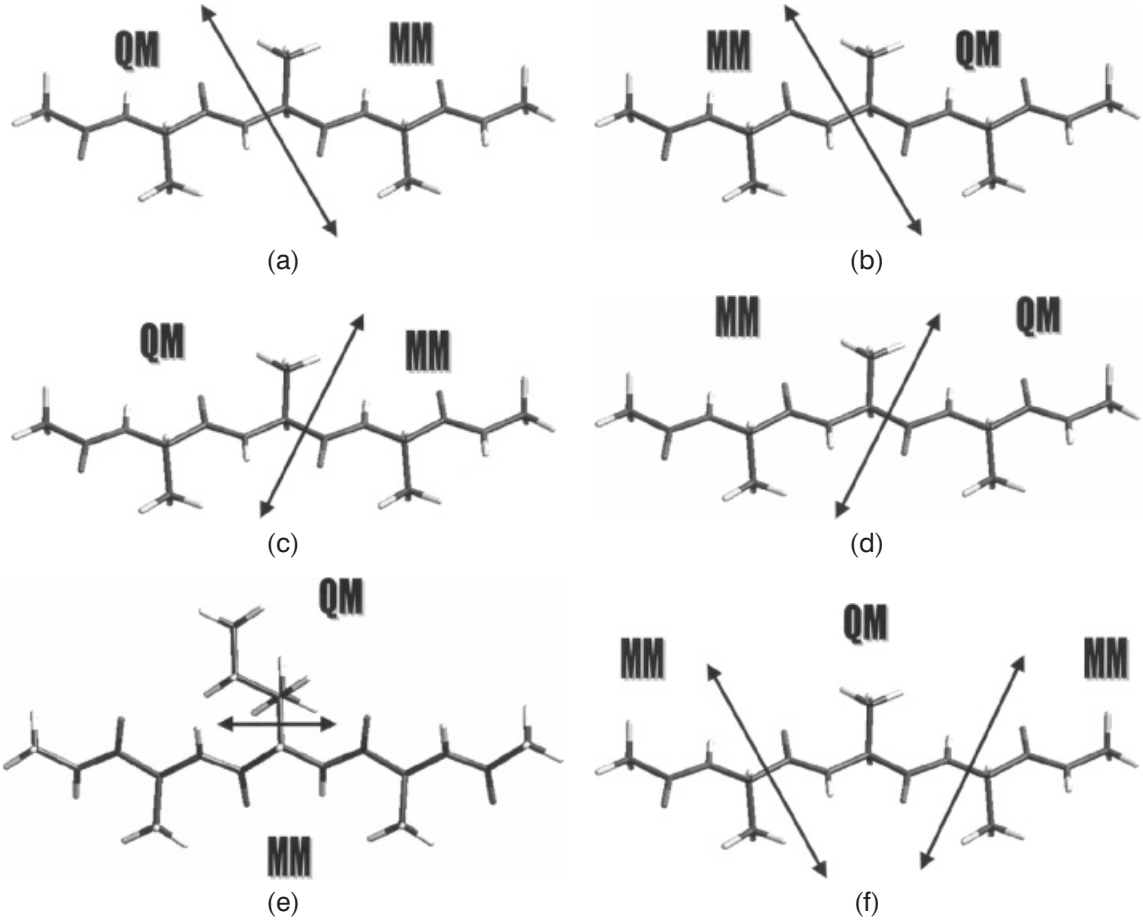

FIGURE 1. (a–f) Possible locations for QM–MM interfaces. In addition, pure QM ligands and ions can be specified.

TABLE I.
RMS Deviations (kcal/mol) of Rotamer Side-Chain Conformational Energies for HF/6-31G* QM–MM, B3-LYP/6-31G* QM–MM, and OPLS-AA Relative to cc-pVTZ(-f) LMP2.

Residue	QM–MM QM = HF	QM–MM QM = DFT	OPLS-AA
Phe	0.18	0.14	0.18
His	1.14	1.10	1.05
Asn	0.38	1.36	2.29
Val	0.21	0.41	0.62
Trp	0.71	0.25	0.75
Tyr	0.37	0.29	0.40
Leu	0.77	0.90	0.40
Met	1.10	1.21	1.82
Gln	1.48	1.03	2.70
Glu	2.61	2.69	3.34
Cys	0.61	0.62	3.52
Lys	1.44	1.29	4.22
Ile	1.03	0.56	1.19
Asp	1.11	1.91	2.51
Arg	1.67	3.50	2.90

Table II presents a subset of the molecular pairs that we have investigated, along with QM, MM, and QM/MM binding energetics (in all cases shown, good agreement was achieved between the QM and QM/MM structures for the dimer). Table III gives representative optimized Lennard–Jones parameters that we have determined for OPLS-AA atom types. It was only necessary to vary the van der Walls radii and not the well depths. In general, there is a 5 to 10% increase in the QM–MM radii relative to OPLS-AA. This increase of the radii is largely necessary to keep the QM region from being overly attracted by the MM charges, because the MM parameters present insufficient Pauli repulsion to the QM region.

Although our optimization protocol is far from perfect, the binding energies are typically within 0.6 kcal/mol of the desired target—a much better result than that obtained without the extensive parametrization effort deployed here (e.g., by simply using the OPLS-AA van der Waals parameters for the QM region, which leads to many grossly incorrect binding energies). As the active site modeling

TABLE II.
QM–MM Dimer Binding Energies (kcal/mol) Relative to OPLS-AA, Hartree–Fock (HF/6-31G), and Local MP2 (LMP2-cc-pVTZ(-f)).**

Dimer	LMP2	HF	OPLS-AA	QM/MM QM = HF	MM/QM QM = HF	QM/MM QM = DFT	MM/QM QM = DFT
cf3cooh-ch3coh	8.77	9.31	7.97	8.16	9.48	7.83	8.19
ch3coh-nh2coch3	5.40	5.68	8.47	7.43	7.08	7.15	6.10
ch3cooh-ch3sh	5.27	4.36	5.28	6.05	6.51	5.80	6.34
ch3oh-ch3oh	5.72	5.58	5.39	4.38	5.60	5.00	5.53
ch3cooh-ch3oh	5.74	5.80	4.81	5.42	5.36	4.83	5.22
ch4-nh2coch3	0.97	0.80	1.14	0.66	0.88	0.71	0.89
ch3coh-ch3oh	5.72	5.58	5.39	4.38	5.60	5.00	5.53
c6h6-ch4	0.84	0.23	1.19	1.07	1.22	1.02	1.21
ch4-ch3coh	0.64	0.42	0.42	0.51	0.71	0.49	0.70
ch3cooh-nh2ch3	11.31	10.61	10.50	10.76	11.59	10.40	11.22
nh2ch3-nh2ch3	3.48	2.86	4.61	4.72	3.91	4.59	3.86
nh2ch3-nh2coch3	4.95	4.34	4.63	4.26	3.74	4.20	3.16
nh2ch3-ch3cooh	4.96	4.57	4.09	4.31	3.67	3.92	3.62
nh2coch3-ch3sh	3.42	2.06	4.33	4.13	3.90	4.03	3.76
ch3cooh-h2o	3.90	3.35	3.13	4.61	3.13	4.32	3.02
nh3ch3-histidine	8.44	6.99	7.53	8.85	8.26	8.61	8.10
ch3oh-histidine	7.44	6.80	5.76	6.54	6.47	6.40	5.86

Both choices for the QM monomer are displayed (QM/MM, MM/QM).

plan specifies placing the QM/MM interface some distance from the active-site chemistry, one can hope that extremely high accuracy is not required in these values, because there will be a significant cancellation of error at the interface when comparing several different structures. On the other hand, qualitatively incorrect interactions would quite possibly lead to highly distorted geometries and induce strain energy that could perturb the reactive region; we believe that our parametrization results are of sufficient quality to avoid or at least mitigate this sort of problem.

TABLE III.
Representative QM–MM van der Waals Parameters.

Atom Type	σ–QM–MM	σ–OPLS-AA
C Trp-Gamma	1.90	1.775
CA Aromatic carbon	1.90	1.775
CR Histidine C-epsilon	1.90	1.775
CT Tetrahedral carbon	1.90	1.75
N Amide	1.788	1.625
NB Unprotonated aromatic	1.90	1.625
NT Amine	1.788	1.65
O C=O In peptide	1.628	1.48
OC2 Aldehyde/ketone	1.776	1.48
OC3 Carboxylic acid	1.776	1.48
S Sulfides	1.90	1.80
SH Thiols	1.90	1.80

In general there is a 5–10% increase in the QM–MM radii relative to OPLS-AA.

HYDROGEN BOND CORRECTIONS

As alluded to above, a proper parametrization of the QM–MM potential requires a somewhat stronger Pauli repulsion term between QM–MM regions than encountered for pure MM models. The physical reason for this is clearly that the QM wave function sees no Pauli repulsion from MM charges. Part of the enhanced QM–MM Pauli repulsion appears in the enlarged (relative to OPLS-AA) van der Waals radii presented in the previous section. We have found that strong hydrogen bonds require an additional Pauli repulsion, presumably because of the large polarization characteristic of hydrogen bonds. In particular, the hydrogen bonds that appear to need a stronger Pauli correction are the N—H...(O,S) hydrogen bonds prevalent in proteins. The following exponential functional performs very well at providing the necessary QM–MM Pauli repulsion,

$$V_{\text{H-bond}} = \frac{\varepsilon e^{-(\sigma r^2)}}{r^2} \tag{7}$$

with r the H-acceptor atom distance and σ, ε parameters determined to minimize errors in peptide protein conformations with strong H bonds. The σ, ε parameters are 0.15 $Å^{-2}$ and 20 kcal/mol, respectively, for all QM–MM N—H. . .(O,S) hydrogen bond interactions. These could be further refined to particular pairs of atom types. This term is similar to a repulsive core pseudopotential ecountered in quantum chemistry.

DEVELOPMENT OF FROZEN BOND INTERFACE PARAMETERS

With the QM van der Waals parameters specified, we next proceed to the development of the QM/MM frozen interface parameters. For this development, we employ blocked dipeptides as model molecules. These systems are small enough to allow an extensive high-quality quantum chemical potential energy surface to be produced, yet large enough to properly capture the chemical environment of each frozen bond. The alanine dipeptide is used to develop the backbone frozen orbital parameters; the appropriate dipeptide is employed for cuts in specific side chains.

We have described[10] the methodology for fitting the stretching, bending, and bond charge parameters in previous articles, and will only briefly review this here. The bond charge is parametrized to reproduce quantum mechanical deprotonation energies. Small dipeptides are used and the deprotonation site occurs from the side-chain atom ~4 bonds or more away from the cut location. Bond stretch and angle bend parameters are fit by independent displacements of the stretch/bend internal coordinates and fitting the energy differences to quantum values.

The final step in constructing the potential function is to fit torsional parameters to reproduce quantum chemical conformational energies of the model molecules. Our approach here was discussed in ref. 10; points are as follows:

1. We begin with a set of ~200 side-chain rotamer states generated via molecular conformational search by Jorgensen and coworkers. The rotamer states were generated by an extensive conformational search of the side-chain degrees of freedom. Each of these conformations is then optimized at the HF/6-31G** level and followed by a single-point LMP2/cc-pVTZ(-f) calculation. Reproduction of the relative energetics of these minima provide a rigorous test of the validity of our side chain-based frozen orbital parametrization. For the backbone, we fit to the alanine dipeptide surface and then test the results by calculating relative energies of 10 tetrapeptides conformations we have studied previously at the same level of quantum chemical methodology.
2. We generate potential energy surfaces suitable for fitting torsional parameters by taking one-dimensional cuts along the relevant torsional degrees of freedom through each minimum. This allows reliable treatment of coupled torsions, yet is much less time consuming than generating full two or even three or four-dimensional surfaces. Overall, on the order of 4000 data points are included in the data set—by far the largest set of high-quality peptide data that has been used in QM/MM or indeed any kind of force-field development.
3. A specially developed weighting scheme, involving the gradient of the potential, has been developed, and is very helpful in reliably fitting the entire data set. This methodology will be discussed in detail in a subsequent publication.

Validation and Testing of the Methodology

DEPROTONATION TESTS

One important measure of how well a QM/MM model reproduces accurate quantum mechanics is chemical reaction energetics. In this article, we examine a simple reaction, removal of a proton from the QM region. Deprotonation energies are calculated for a number of polyalanine structures, and the QM and QM/MM results are compared. Tables IV and V present the test cases that we have examined. Tables IV and V compare the QM and QM/MM results, and also display the distance between the proton to be removed and the frozen interface orbital. It can be seen that when this distance is greater than ~5 Å—a distance that one would want to maintain between a reactive chemical event and the QM/MM interface in any event—the error is on the order of 0.1–0.2 kcal/mol, a negligible amount when compared to the hundreds of kcal/mol total reaction energy, and small even by the standards of the intrinsic error in HF or DFT calculations compared to the experiment (the latter with a large basis set does quite well, but still makes errors on the order of 1–2 kcal/mol).

TABLE IV.
B3-LYP 6-31G* QM–MM Absolute Deprotonation Energy Differences Relative to Full 6-31G* B3-LYP Values.

Peptide MM region	QM region	Frozen orb.–H distance (Å)	Error (kcal/mol)
ace-ala-LEU[a]-nma	H1	4.1	0.70
ace-ala-LEU[a]-nma	H2	4.3	0.52
ace-ala-LEU[a]-nma	H3	4.3	0.53
ace-ala-ALA-leu[a]-nma	H1	5.2	0.40
ace-ala-ALA-leu[a]-nma	H2	5.5	0.34
ace-ala-ALA-leu[a]-nma	H3	6.3	0.29
ace-ala-ALA-ala-leu[a]-nma	H1	8.7	0.20
ace-ala-ALA-ala-leu[a]-nma	H2	9.7	0.15
ace-ala-ALA-ala-leu[a]-nma	H3	8.2	0.23
ace-ala-ala-ala-ala-ala-LEU[a]-ala-nma	H1	4.1	1.21
ace-ala-ala-ala-ala-ala-LEU[a]-ala-nma	H1	4.3	0.93
ace-ala-ala-ala-ala-ala-LEU[a]-ala-nma	H1	4.3	0.82

[a] Denotes the deprotonated QM leucine residue (H of $C\gamma H_3$). The capitalized residue denotes the QM/MM boundary with the capitalized residue in the QM region.

When the reactive event is very close to the frozen bond, the errors are somewhat larger but still not disastrous. In fact, errors can be reduced substantially even for these cases with a more sophisticated charge distribution for the bond charge; we will discuss this further in another publication. The simple single-bond charge formalism is adequate for our first generation model (other errors are larger with any sensible definition of the QM/MM interface location) and has the advantage of being easy to parametrize.

CONFORMATIONAL ENERGETICS

Backbone Interface

Tables VI and VII presents relative energetics of 10 conformations of the alanine tetrapeptide, which we have extensively studied in previous

TABLE V.
HF 6-31G* QM–MM Absolute Deprotonation Energy Differences Relative to Full 6-31G* HF Values.

Peptide MM region	QM region	Frozen orb.–H distance (Å)	Error (kcal/mol)
ace-ala-ALA-nma[a]	H1	4.5	0.2
ace-ala-ALA-nma[a]	H2	4.2	0.0
ace-ala-ALA-nma[a]	H3	4.2	0.1
ace-ala-LYS[a]-nma	H1	6.4	0.3
ace-ala-LYS[a]-nma	H2	6.4	0.3
ace-ala-LYS[a]-nma	H3	7.0	0.0
ace-ala-ALA-ala-leu[a]-ala-nma	H1	8.7	0.2
ace-ala-ALA-ala-leu[a]-ala-nma	H2	9.7	0.0
ace-ala-ALA-ala-leu[a]-ala-nma	H3	8.2	0.2
ace-ala-ALA-leu[a]-ala-nma	H1	5.2	0.6
ace-ala-ALA-leu[a]-ala-nma	H2	5.7	0.7
ace-ala-ALA-leu[a]-ala-nma	H3	6.3	0.5
ace-ala-ala-ala-ala-ALA-leu[a]-ala-nma	H1	5.2	1.1
ace-ala-ala-ala-ala-ALA-leu[a]-ala-nma	H2	5.7	1.1

[a] Denotes the deprotonated residue with the H atom taken from the end of the side chain. The capitalized residue denotes the QM/MM boundary with the capitalized residue in the QM region.

TABLE VI.
Conformational Energies (kcal/mol) of Alanine Tetrapeptide Calculated with B3-LYP/6-31G* QM–MM for the Four QM–MM Boundaries Depicted in Figure 1a–d.

Conformation	Cut 1	Cut 2	Cut 3	Cut 4	LMP2	OPLS-AA
1	2.24	2.80	2.61	2.56	2.71	3.05
2	2.36	2.35	1.63	2.31	2.84	2.69
3	−0.60	−0.20	−0.80	−0.90	0.00	−1.14
4	3.85	4.22	2.55	4.17	4.13	3.83
5	3.73	3.26	4.84	2.97	3.88	4.80
6	1.74	−0.80	0.86	2.09	2.20	−0.69
7	4.53	6.64	7.41	6.30	5.77	4.36
8	6.95	5.72	3.76	5.37	4.16	6.88
9	6.85	8.26	7.67	6.59	6.92	6.30
10	7.50	7.08	9.18	8.33	6.99	9.52
RMS	1.03	1.21	1.11	0.75	0.00	1.64

The zero point is adjusted in all calculations to minimize the RMS with respect to LMP2 [cc-pVTZ(-f)].

articles,[26] comparing with quantum chemistry at the LMP2/cc-pVTZ(-f) level. Figure 1a–e indicates where the QM/MM interface is placed for each of the results reported in Table VI. The QM/MM results are comparable in quality to the MM results.

Side-Chain Interfaces

Table I summarizes the RMS deviations between the QM/MM, MM, and QM results for the relative energetics of the side-chain rotamer dataset described above. Again, the QM/MM results are comparable in quality to those from the MM force field. The robustness of performance across a wide range of interfaces demonstrates that the methodology we have developed is reliable and relatively straightforward to parametrize. The QM/MM rotamer structures also display good fidelity to the QM structures. As in the case of the MM results themselves, there are occasional errors that are larger than one would like, in both energy and structure; however, our judgment is that improvement of these cases will have to await improvement of the MM force field. The present results represent a reasonable first-generation model that should be adequate to study a wide variety of protein active site modeling problems.

TABLE VII.
Conformational Energies (kcal/mol) of Alanine Tetrapeptide Calculated with HF/6-31G* QM–MM for the Four QM–MM Boundaries Depicted in Figure 1a–d.

Conformation	Cut 1	Cut 2	Cut 3	Cut 4	LMP2	OPLS-AA
1	3.19	1.96	2.17	2.47	2.71	3.05
2	3.66	2.00	1.49	2.43	2.84	2.69
3	−0.10	−0.80	−0.40	−0.01	0.00	−1.14
4	2.90	3.03	2.17	3.26	4.13	3.83
5	3.73	4.13	4.81	3.25	3.88	4.80
6	3.30	0.96	−0.20	0.81	2.20	−0.69
7	2.32	5.84	6.14	6.49	5.77	4.36
8	3.46	8.30	6.23	4.56	4.16	6.88
9	8.33	8.45	8.57	7.87	6.92	6.30
10	8.31	5.30	9.08	8.81	6.99	9.52
RMS	1.13	1.64	1.56	0.90	0.00	1.64

The zero point is adjusted in all calculations to minimize the RMS with respect to LMP2 [cc-pVTZ(-f)].

Implementation of the QM/MM Model for Protein Simulations

OVERVIEW

The data presented above demonstrates that the frozen orbital-based QM/MM model that we have developed provides an accurate representation of conformational, elecstrostatic, and nonbonded interactions of peptide backbone and side-chain groups. However, to realistically model proteins, which are several orders of magnitude larger than the test systems we have discussed above, a substantial effort in computational implementation was required. We have carried out this implementation via a tight coupling of the Jaguar suite of *ab initio* programs[12] with the IMPACT molecular modeling code, originally developed by Levy and coworkers. The major features of the implementation are as follows:

1. Ease of use is critical for QM/MM applications, where specification of the interface can be extremely complex and time consuming. We have developed an automated approach in which the user simply specifies which protein residues are to be treated quantum mechanically, and whether bonds are to be cut in the side chain or backbone, and the program builds the interface model and drives all of the calculations. In addition, ligands and ions can be specified as quantum mechanical. This feature will allow nonexpert users to employ the program without an unacceptable learning curve.
2. Computational efficiency and time to solution have been extensively addressed, for exampe, by developing a parallel version of the QM/MM code and by constructing an efficient minimization algorithm in which the MM part of the calculation is adiabatically minimized after each QM step. We also allow freezing of a significant part of the protein, a valuable technique when dealing with very large systems.
3. We have incorporated a continuum solvation model, including an analytical gradient, into the QM/MM Hamiltonian. The model allows treatment of the entire protein by the solvation model, which is solved self-consistently (via a self-consistent reaction field approach) for each QM step. Inclusion of solvation is critical in some (although not all) applications.

In addition to these three major areas, many other infrastructure developments were needed to allow the QM and MM codes to interoperate smoothly. Below, we outline the principal developments in the three major areas discussed above.

EASE OF USE/PROGRAM FLOW

The QM/MM program is driven by a Jaguar driver code that calls IMPACT and Jaguar executables as needed. The IMPACT and Jaguar programs share necessary data via a file on disk. IMPACT does the necessary setup of the protein/ligand structure, and the user specifies within IMPACT the QM region simply using residue numbers, QM/MM cut locations, and nondefault basis sets if desired. IMPACT will subsequently write a Jaguar input file from which Jaguar begins the quantum calculations. We are presently writing a graphical interface from which the user can pick the QM regions at the screen and launch the job.

COMPUTATIONAL EFFICIENCY

Pseudospectral QM/Parallelization

Because the QM portion of a QM/MM calculation dominates the CPU usage and relatively large ligand/active centers (~1000 basis functions) are commonly encountered, it is essential that the QM methods perform at peak performance. The DFT, Hartree–Fock, and LMP2 methods in Jaguar have been documented[20] to achieve superior performance on large systems through the use of novel pseudospectral integration techniques.[21] In addition, we have implemented parallel versions of these algorithms that scale up well with the number of processors.[22] The combination of the pseudospectral and parallel capabilities makes QM optimizations of systems with ~120–150 atoms or 1000–1500 basis functions feasible in a matter of a week over several processors. Detailed timings of the QM methods can be found in ref. 20.

Adiabatic QM/MM Minimization

Because molecular mechanics minimizations of proteins typically take thousands of steps, it is essential to reformulate the minimization procedure for QM–MM to avoid having to evaluate many QM energy/gradient cycles. To this end we have implemented an adiabatic minimization procedure for QM–MM minimizations. In the adiabatic method, the MM space is first minimized with a frozen QM

region. Following this MM minimization, the forces on the QM atoms are calculated, including the force from the MM atoms, and a single geometry step is taken in the QM region with the MM region held frozen. The cycle then returns to the MM minimization with a frozen QM region. The optimization of the QM steps takes advantage of the algorithms used in Jaguar to optimize small molecules. This avoids wasting expensive gradient calculations on the cruder optimizers of the MM codes, which take many iterations relative to quantum codes using Hessian updating.

During the MM minimization the MM atoms experience a QM force from the MM–QM coupling terms. To avoid the explicit calculation of this QM–MM coupling force on each MM step we have defined a good approximation to this coupling force to account for changes in this force when the MM atoms change positions during MM optimization. The approximate QM–MM force used during the MM optimization consists of the exact QM–MM force F_{ex} evaluated at the initial geometry of the MM minimization plus an approximate correction force ΔF, which accounts for changes in the exact force due to changes in the positions of the MM atoms. The approximation made in calculating ΔF involves approximating the QM region by point charges derived from a fit to the QM electrostatic potential at the end of each QM step. The change in force ΔF is simply the difference in the force (relative to the initial reference geometry) from the MM point charge–QM ESP point charge interaction. That is ΔF is the change in QM–MM point charge model force from its value at the initial reference geometry of the MM minimization relative to the value of this force at a given geometry of the MM minimization. This approximation works very well in practice, and of course, as the minimization converges ΔF approaches zero and the approximation becomes exact.

We have found that the adiabatic minimization procedure can minimize QM–MM systems in of order 50–200 cycles for large (>20 atoms) QM regions and proteins with of order 7000 atoms. One hundred steps for an optimization of the corresponding pure QM region is not uncommon. This indicates that the adiabatic procedure is within a factor of 2 as efficient as a pure QM optimization. Furthermore, the MM minimizations get progressively less expensive at later cycles, typically only taking ~50 MM steps in a marginal amount of time. Timings for a minimization of cytochrome P-450 are discussed below.

SOLVATION MODEL

QM–MM solvation single-point and optimization calculations can be performed within the context of the PBF continuum solvent method. In this method,[23, 24] the QM region is self-consistently polarized by the surrounding dielectric. The Poisson equation is solved numerically on a grid to provide a self-consistent solution for the potential of the solute + solvent with the QM-solute approximated by ESP fit charges. The surface charges at the solute/solvent interface obtained by this solution are subsequently used as a polarizing potential in an optimization of the QM wave function. As shown in ref. 24 the PBF solvation method gives accurate solvation energies (to within ~0.3 kcal/mol) for small molecules. The QM–MM solvation procedure is quite similar, the main difference being that the MM charges are not self consistently polarized. A solvated QM/MM optimization follows the same adiabatic method described above with the addition of the solution for the reaction field. In the quantum steps, the full reaction field from the protein + QM region is calculated and the QM wave function is self-consistently optimized with the reaction field as it is for small molecule calculations.[23] In the MM steps, the QM region is replaced by the QM ESP fit charges.

Preliminary Calculations on Cytochrome P450-cam

Cytochrome P-450 is an enzyme whose variants are ubiquitously distributed across a wide variety of organisms.[15] Human isoforms of the enzyme in the liver are of great importance for pharmaceutical development because they are involved in a significant fraction of toxicity and drug metabolic pathways. Although a high-resolution structure of a human enzyme does not yet exist, such structures are likely to be produced in the next 5 years (either experimentally or via homology modeling), which will open the possibility of computer modeling of these critical processes. However, it is very difficult to study the chemistry or ligand binding of cytochrome P450 with conventional molecular modeling techniques. The existence of a reactive metal center, and the centrality of the reactive chemistry in the interaction of the enzyme with various drug candidates, mandate a quantum chemical treatment of the active site; on the other hand, the protein structure is clearly important in selectivity, binding affinity, and reaction kinetics, and cannot be incorporated to any

great degree with conventional quantum chemical techniques. A QM/MM methodology of the type we have developed is well suited to addressing this problem with relatively modest computational resources.

STRUCTURE AND QM REGION

The particular systems studied are available as files "1PHF" and "1AKD" from the public PDB archives. The former has a coordinating phenyl imidazole ligand, and the latter is the so-called substrate free state with a camphor ligand near the heme.

P-450 and Phenyl-Imidazole Ligand

The 1PHF structure consists of cytochrome P-450 containing a 4-phenyl imidazole complexed with Fe. On the opposite side of the heme Fe is coordinated to a cysteinate sulfur (R—S^-). Thus, the Fe is in a sixfold coordination site and assumed to be in a formal charged state of Fe^{3+} given the S^- formal charge and the two negative charges in the heme ring system. The net spin state is a either a doublet with the Fe—S moiety low spin coupled as in the coordinated dioxygen state of P-450 or a high-spin quartet. There are a total of 7075 atoms in the system.

The QM region depicted in Figure 2a was chosen to be the full heme ring, the Fe, the full coordinating cysteine residue including residues on either side of this cysteine, and the 4-phenyl imidazole. The two QM–MM cuts in the protein occur in the residues adjacent to the coordinating cysteine. The net charge of the QM region is −2 from the two carboxylate groups on the heme. There are 125 total quantum atoms and 1138 6-31G* basis functions. The QM method used was B3-LYP DFT, which has been shown to give reasonable results for transition metal containing systems and in particular for related enzymes such as MMO.[14] The outer shell of the protein was frozen during whole optimization, leaving 3960 of the 7075 atoms free to optimize. This procedure is commonly used to avoid irrelevant energy differences caused by rearrangements of the outer parts of the protein.

P-450 and Camphor

The specification of this system depicted in Figure 2b is identical to that above, with a quantum camphor molecule replacing the phenyl-imidazole ligand. The camphor is not directly bound to the Fe, and has forced water out of the heme binding region. The main purpose of running this system was to find the lowest energy spin state of the substrate free P-450.

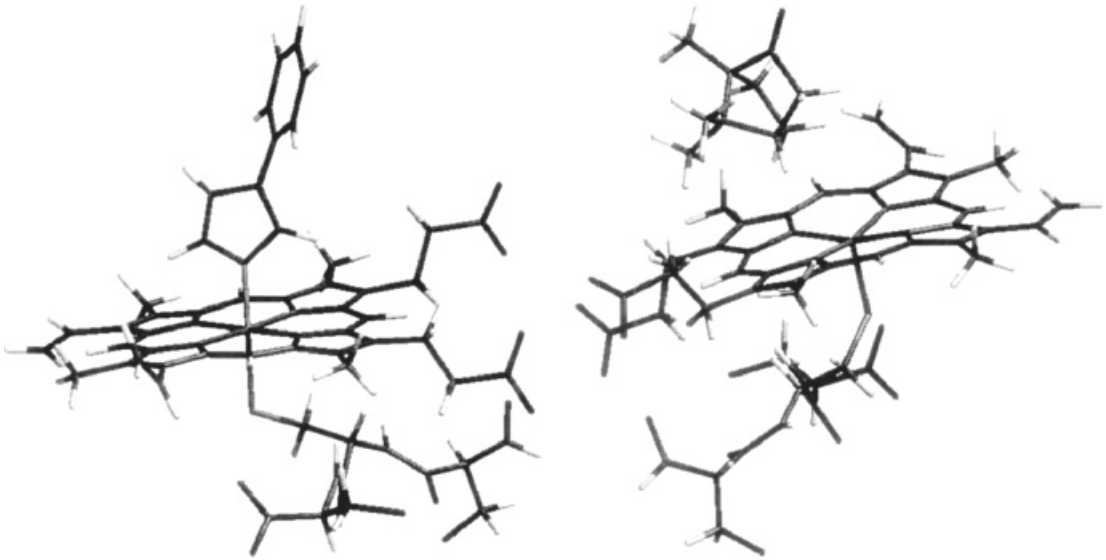

FIGURE 2. (a) QM region of P-450 QM–MM calculation consisting of P-450 heme region and a coordinating phenyl-imidazole ligand. (b) QM region of P-450 calculation consisting of P-450 heme region and bound camphor.

Results

The P-450 optimization was run on 6 SGI-R10k nodes in ~1 week wallclock time. The calculation took ~30 QM–MM cycles. Approximately 10 h was spent on the initial MM minimization, in which the QM region is frozen. The final geometry has an rms deviation of 0.6 Å with respect to the nonhydrogen atoms of the crystal structure, a reasonable level of accuracy. The rms deviation in the quantum subspace is 0.5 Å. In a general study of a system like this, the initial cost of this minimization would be amortized over similar runs in which the QM region would be perturbed (by changing the ligand, for example), because these subsequent runs would have a good initial geometry. As an initial calibration of the energetics we optimized the doublet and quartet spin states of the system with the imidazole ligand. The doublet was found to be ~14 kcal/mol lower in energy. This is in accord with EPR data[25] that indicates that Fe is in a low spin state when in a sixfold Fe^{3+} coordination site involving dative bonding to the ligand. The substrate-free system with the camphor above the heme ring was found after full QM–MM optimization to have a high-spin quartet ground state with the doublet ~15 kcal/mol higher in energy. This ordering of spin states is also in agreement with the experiment.[15]

Conclusion

We have developed a novel implementation of mixed QM/MM methods, which is specifically targeted at modeling protein active sites. The results shown above for intermolecular interactions, conformational energetics, and deprotonation energies provide in-depth validation of the accuracy of the model compared to high-level fully QM calculations on the same systems. Furthermore, the cytochrome P450cam computations demonstrate that the method is readily applicable to large, complex protein active sites, where the computational requirements for such calculations are relatively modest considering the difficulty of the endeavor, and that the results obtained compare favorably with the experiment in terms of both structural and energetic predictions. Further tests of the method are still required to demonstrate robustness over a wide range of problems, and to characterize the accuracy that is obtainable for a diverse sample of active-site chemical reactions. Finally, we have not, in this article presented results for solvated calculations, other than to describe the methodology that has been implemented; computations employing the solvation model discussed above will be given elsewhere.

Assuming that the methodology is, in fact, robust, and provides accuracy at the level expected from current DFT functionals (~2–5 kcal/mol for reactive energy differences), the approach described in this article opens the possibility of studying a large number of transition metal containing enzymes (there are now several hundred high-resolution crystal structures of such enzymes in the PDB) to elucidate in atomic detail the active-site catalytic pathways. While one can often formulate a reasonable hypotheses for catalytic intermediates and transition states by examining the crystal structure, there are typically a nontrivial number of alternatives, and the selected correct structures and specifying energies from among the manifold of possibilities is very difficult in the absence of a model capable of carrying out geometry optimization and supplying reliable energy differences. What QM/MM provides, as contrasted with QM-only calculations, is the ability to carry out this program at a reasonable computational cost without having to often make unacceptable compromises in the size of the model being studied. The QM/MM interface we have constructed surely has quantitative errors, but represents a very substantial improvement compared to radical truncation of the size of the model that will often affect structures in an uncontrolled fashion. The ease-of-use features we have implemented are crucial in this regard as well, as the need for carrying out a large number of computational experiments on different types of models, which is very demanding in terms of the human effort required, is significantly reduced.

Extension of the methods presented here to chemical groups other than amino acids does not present any problem in principle. All that is necessary is to specify suitable model molecules and to accumulate a library of parameters. Extensive tests do need to be carried out to determine how much chemical similarity is require to yield reliable results in transfering a set of interface parameters from a given template into a larger system. These tests are straightforward to perform (via comparison with accurate QM calculations), but surveying a wide range of chemical functionalities, as is required in the present approach, is time consuming. Nevertheless, our expectation is that in several years a general library will be available and the method applicable to other problems such as catalysis in polymeric systems by organometallic catalysts.

Acknowledgments

We thank the NPACI program, supported by the NSF, for computation time at the NCSA in Urbana–Champaign.

References

1. Warshel, A.; Lewitt, M. J Mol Biol 1976, 103, 227.
2. Singh, U. C.; Kollman, P. A. J Comput Chem 1986, 7, 718.
3. Bash, P. A.; Field, M. J.; Karplus, M. JACS 1987, 109, 8092.
4. Gao, J.; Xia, X. Science 1992, 258, 631.
5. Luzhkov, V.; Warshel, A. J Comput Chem 1992, 13, 199.
6. Stanton, R. V.; Hartsough, D. S.; Merz, K. M. J Phys Chem 1993, 97, 11868.
7. Tunon, I.; Martins–Costa, M. T. C.; Millot, C.; Ruiz–Lopez, M. F.; Rivail, J.-L. J Comput Chem 1996, 17, 19.
8. Bakowiec, D.; Thiel, W. J Phys Chem 1996, 100, 10580.
9. Maseras, F.; Morokuma, K. J Comput Chem 1995, 16, 1170.
10. Philipp, D. M.; Friesner, R. A. J Comput Chem 1999, 20, 1468.
11. Murphy, R. B.; Philipp, D. M.; Friesner, R. A. Chem Phys Lett, to appear.
12. Jaguar v4.0, Schrodinger Inc.
13. Jorgensen, W. A.; Maxwell, D. S.; Tirado–Rives, J. J Am Chem Soc 1996, 118, 11225.
14. Dunietz, B. D.; Beachy, M. D.; Cao, Y.; Whittington, D. A.; Lippard, S. J.; Friesner, R. A. J Am Chem Soc, to appear.
15. Porter, T. D.; Coon, M. J. J Biol Chem 1991, 266, 13469.
16. Johnson, B. G.; Gill, P. M. W.; Pople, J. A. J Chem Phys 1993, 98, 5612.
17. Becke, A. D. J Chem Phys 1993, 98, 1372.
18. Murphy, R. B.; Beachy, M. D.; Friesner, R. A.; Ringnalda, M. N. J Chem Phys 1995, 103, 1481.
19. Dunning, T. H. J Chem Phys 1989, 90, 1007.
20. Friesner, R. A.; Murphy, R. B.; Beachy, M. D.; Cao, Y.; Pollard, W. T.; Ringnalda, M. N. J Phys Chem A 1999, 103, 1913.
21. Friesner, R. A. J Phys Chem 1988, 92, 3091.
22. Beachy, M. D.; Chasman, D.; Friesner, R. A.; Murphy, R. B. J Comput Chem 1998, 19, 1030.
23. Tannor, D. J.; Marten, B.; Murphy, R. B.; Friesner, R. A.; Sitkoff, D.; Nicholls, A.; Ringnalda, M.; Goddard, W. A., III; Honig, B. J Am Chem Soc 1994, 116, 11875.
24. Marten, B.; Kim, K.; Cortis, C.; Friesner, R.; Murphy, R.; Ringnalda, M.; Sitkoff, D.; Honig, B. J Phys Chem 1996, 100, 11775.
25. Lippard, S. J.; Berg, J. M. In Principles of Bionorganic Chemistry; University Science Books: Mill Valley, CA, 1994.
26. Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. J Am Chem Soc 1997, 119, 5908.