

JCTC

Journal of Chemical Theory and Computation

A Combined QM/MM Poisson–Boltzmann Approach

Seth A. Hayik, Ning Liao, and Kenneth M. Merz, Jr.*

*Department of Chemistry, Quantum Theory Project, University of Florida,
P.O. Box 118435, Gainesville, Florida 32611-8435*

Received September 24, 2007

Abstract: A method of solving the mixed quantum mechanical/molecular mechanical (QM/MM) Hamiltonian in solution, using the Poisson–Boltzmann (PB) equation to calculate partial charges and solvation free energies, is presented. This method combines a linear scaling divide and conquer semiempirical algorithm with the PB equation in a QM/MM framework, allowing only a specified region's charges to be polarized by the solvent while using fixed charges from a MM force field for the remaining system. This can save time over a full QM implementation, only requiring self-consistency to be achieved in a small QM region, while giving comparable results. The solvation free energy of pentapeptides capped with an acetyl group (ACE) at the N-terminus and an N-methylamine group at the C-terminus (NME) was used to study the accuracy of this method as well as several small protein systems. The solvation free energies for the QM/MM implementation compare well with a full QM treatment of the same system, giving reasonable representations for the solvation free energy of the entire system independent of the QM region's size. In the case of the pentapeptides, the average error was only 4.9 kcal/mol with the smallest QM region. This mixed method will allow an accurate description of solvation effects in an area of interest, such as an active site, using mixed QM/MM Hamiltonians. Possible applications for this method include protein–ligand binding and reaction mechanism studies.

Introduction

The effect of solvent on biological molecules is known to be very significant,^{1–5} and in order to model biomacromolecules accurately one has to include this effect to properly account for electrostatic interactions. These solvation effects are often necessary in molecular dynamics simulations and protein ligand binding studies to get accurate free energies of macromolecular systems where a difference of 1 or 2 kcal/mol is substantial.^{6–8} Computationally, the effect of solvent can be included explicitly, by including water molecules surrounding the solute,⁹ or implicitly by adding terms to the Hamiltonian using a continuum method.¹⁰ Explicitly modeling solvent is expensive because it adds more atoms to the system, whose numbers increase greatly with a protein's overall size and typically requires an extended equilibration of the system. To counter this potential increase in cost and sampling, the generalized Born¹¹ method (GB) is sometimes used in QM/MM and pure MM studies to provide an estimate

of solvent effects on the system^{12–15} with a lower cost than some other more accurate methods, such as the Poisson–Boltzmann approach. Inclusion of solvation effects allows a more accurate representation of the system, since most biological systems do not exist in a vacuum. Both explicit and implicit methods have been proven to be useful, and each has their own strengths and weaknesses and are essential for accurately modeling biomolecules.^{1–5}

Continuum solvation methods were designed to cheaply and accurately approximate the statistical sampling of solvent configurations at the interface between a solute and solvent.^{10,16–18} There are many continuum solvent models available, but this paper will focus on the Poisson–Boltzmann (PB) finite difference method as originally implemented in Delphi by Nicholls and Honig,¹⁹ extended to QM methods by Friesner, Goddard, Honig, and co-workers,²⁰ adapted for the linear scaling semiempirical program DivCon by Gogonea and Merz,² and further elaborated on by Liao and Merz.²¹ This method was chosen because it is suitable for use on biological macromolecules in a quantum mechanical frame-

* Corresponding author e-mail: merz@qtp.ufl.edu.

work both in terms of cost and accuracy.² These methods have been effectively used to account for solvation in previous studies of binding free energy of small molecules and provide an accurate estimate for solvation free energies.^{6,22,23} The PB method is an efficient solvation model, making it useful in calculations involving large macromolecules, and has been shown to give good solvation energies when used with a quantum mechanical method.^{20,24,25} Combined with DivCon's linear scaling QM capabilities the PB method presents an efficient and accurate way to calculate the solvation free energy of an entire macromolecular system with a reasonable amount of time invested in the calculation.

The mixed quantum mechanical(QM)/molecular mechanical(MM) method was originally formulated in 1976 by Warshel and Levitt who were exploring the catalytic mechanism of lysozyme.²⁶ Interest in the QM/MM approach was revitalized in 1990 via application and validation work reported by Field, Bash, and Karplus on the basic QM/MM strategy and its application to chemical problems.²⁷ Mixed QM/MM methods are widely used now in a number of programs such as AMBER²⁸ and CHARMM²⁹ and can help reduce the cost of large calculations while allowing important areas of the protein to be treated with a QM Hamiltonian.^{26,27,30} This has proven extremely useful in the study of reaction mechanisms when incorporated with molecular dynamics^{31–33} and also for protein ligand studies in some cases.^{22,34} Another advantage of these methods is that they can be used to save parametrization time in the case of nonstandard residues such as metal ions or drug molecules. In these instances the nonstandard residue can be included in the QM region so that the step of generating accurate force field parameters for the residue(s) in questions does not have to be undertaken. A third advantage of a mixed QM/MM method is that the QM region can properly represent the polarization effects of the surrounding MM and solvent regions through the QM-MM interactions. This polarization changes the QM region's electronic properties and can have a large impact on a reaction pathway or ligand binding.³⁴

In our QM/MM implementation the linear scaling semiempirical QM program DivCon^{35,36} has been integrated into AMBER.²⁸ This allows the solvation energy for an entire protein to be calculated utilizing atomic charges from semiempirical calculations in the QM region and fixed MM point charges in the surrounding protein. These can be Mulliken, CM1,³⁷ or CM2³⁸ charges, but due to accuracy CM1 or CM2 are typically used in DivCon calculations. This results in a method that can predict solvation free energies at a reduced cost compared to full QM calculations by centering specifically on a region of interest, an active site for example, for the expensive calculations and using the less expensive MM force calculations and charges on the rest of the protein. Our approach also takes advantage of DivCon's linear scaling, which allows a large QM region to be chosen in a QM/MM calculation without becoming exceedingly computationally expensive.

Methods

QM/MM Implementation. In the QM/MM method the entire system is first divided into two subsystems. One region

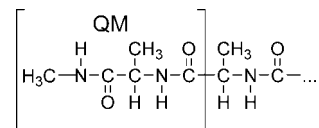


Figure 1. Representative view of the QM/MM region and its cut off points. This also shows an example of the “1 peptide” designation used in Table 1 (ACE cap plus one Ala residue).

(QM) contains a limited number of atoms to be treated with the chosen QM potential, while the second region (MM) is treated with an appropriate force field. In a protein, covalent bonds are cut at the interface between these two regions, and these bonds are treated by the addition of a hydrogen link atom to the QM region in order to ensure a closed shell calculation. This is a widely used method and is reliable given proper placement of the link atom away from polarized bonds.³⁰ In the AMBER implementation the interactions with atoms in the MM region are included, except those bonded to a QM atom (called the MM link pair), so that artificially high forces and polarization effects are not introduced into the calculation. For MM atoms directly bonded to QM atoms, electrostatic interactions are replaced by those of the link atom and the VDW interactions remain unchanged. A representative view of the QM region cut is shown in Figure 1. In our applications the peptide bonds are included in the QM region to avoid cutting these polarizable bonds.

The Hamiltonian of the system is divided into three terms describing the different interactions and is written as

$$H = H_{QM} + H_{MM} + H_{QM/MM} \quad (1)$$

where H_{QM} is the QM Hamiltonian, H_{MM} is the MM Hamiltonian, and $H_{QM/MM}$ is the Hamiltonian describing the interactions between the two subsystems. This Hamiltonian allows the QM region to interact with its surroundings, which allows the MM region to affect the electronic structure of the QM region, thereby, affecting reaction barrier heights, binding free energies, and many other properties.

The Divide and Conquer (D&C) Algorithm. In our implementation the linear-scaling D&C algorithm can be used to solve a semiempirical Hamiltonian for the large molecules of interest.^{35,36} In this method the QM region is divided into subsystems, a Fock matrix for the entire system is built, and then each subsystem's Fock matrix is individually diagonalized giving the equation

$$F^\alpha C^\alpha = C^\alpha E^\alpha, \alpha = 1, \dots, n_{sub} \quad (2)$$

This leads to a global density matrix, $P_{\mu\nu}^{total}$, that is a sum over the density matrix of the subsystems, $P_{\mu\nu}^\alpha$

$$P_{\mu\nu}^{total} = \sum_{\alpha} P_{\mu\nu}^{\alpha} \quad (3)$$

where

$$P_{\mu\nu}^{\alpha} = D_{\mu\nu}^{\alpha} \sum_{i=1}^{M^{\alpha}} n_i^{\alpha} (c_{\mu i}^{\alpha})^* c_{\nu i}^{\alpha} \quad (4)$$

In eq 4 $D_{\mu\nu}^{\alpha}$ is equal to 0 or $1/n_{\mu\nu}$, where n is the number of times a basis function appears in a subsystem. With the

Table 1. Solvation Free Energy of Pentapeptides with Varying QM Region Size^a

residue	DivCon	5 peptides	4 peptides	3 peptides	2 peptides	1 peptide
ALA	-30.5	-30.4	-29.0	-28.8	-28.7	-28.6
ARG	-572.6	-572.6	-572.9	-572.0	-570.2	-560.1
ASN	-55.9	-55.9	-55.7	-57.7	-58.9	-61.0
ASP	-698.2	-698.4	-694.2	-695.0	-698.0	-701.6
CYS	-29.1	-29.1	-29.0	-30.5	-31.1	-32.5
GLU	-653.9	-653.9	-652.2	-652.9	-653.3	-653.5
GLN	-69.3	-69.4	-71.5	-72.3	-70.5	-71.9
GLY	-39.1	-39.1	-37.7	-37.1	-36.5	-35.5
HIS	-81.0	-80.9	-74.5	-70.8	-67.9	-64.3
ILE	-24.9	-24.9	-24.9	-25.0	-24.1	-24.3
LEU	-24.2	-24.2	-24.1	-24.7	-24.4	-24.8
LYS	-624.1	-624.0	-619.6	-616.8	-607.4	-602.9
MET	-34.7	-34.6	-35.0	-33.9	-34.3	-34.3
PHE	-32.6	-32.6	-32.6	-32.7	-33.0	-32.0
PRO	-39.5	-39.4	-36.8	-34.1	-30.0	-27.4
SER	-41.1	-41.2	-39.6	-41.5	-43.3	-45.2
THR	-52.6	-52.6	-53.5	-51.5	-52.0	-52.2
TRP	-53.4	-53.4	-50.8	-52.3	-49.8	-48.3
TYR	-46.4	-46.5	-46.9	-47.6	-47.4	-48.4
VAL	-22.1	-22.1	-22.8	-21.9	-22.9	-23.7

^a All energies in kcal/mol. Number of peptides refers to number of peptides included in the QM region. For example with glycine, 5 peptides refers to 5 glycine and the two capping residues (ACE and NME) in the QM region. Four peptides refers to 4 peptides and the ACE capping residue. "DivCon" refers to an AM1 calculation on the entire pentapeptide done in DivCon.

density matrix in hand we can directly calculate the energy using standard expressions. Our implementation of the divide and conquer algorithm is explained in detail in the extant literature,^{35,36,39} hence, we will focus on the specifics of the QM/MM Poisson–Boltzmann implementation.

Poisson–Boltzmann Combined with DivCon. Implicit solvent models, like those based on the Poisson–Boltzmann approach, involve a trade off between accuracy and computational efficiency. In the PB model, the system is divided into regions with different dielectric constants, separated by the solvent accessible surface of the solute molecule. The solute is then described as a collection of discrete point charges, which are calculated by the QM/MM method. The potential of the electrostatic field (called the reaction field) generated by the polarized solvent due to the introduction of the solute is obtained by solving the Poisson–Boltzmann equation

$$\nabla[\epsilon(r)\phi \nabla(r)] - \kappa(r)2\sinh(\phi(r)) = -4\rho\pi(r) \quad (5)$$

where $\rho(r)$ is the charge distribution of the solute, calculated either via semiempirical QM methods using Mulliken or CM1³⁷/CM2³⁸ charge models for the QM region or carried over from the AMBER force field used for the MM region. $\phi(r)$ is the electrostatic potential that needs to be determined, and $\epsilon(r)$ is the dielectric constant distribution in space, which is normally defined as ϵ_{out} for the solvent and ϵ_{in} for the solute. $\kappa(r)$ is a modified Debye–Huckel parameter that reflects the salt concentration and temperature.

The PB solver in our semiempirical QM program, DivCon, uses the Finite Difference Method (FDM)^{40–42} to solve the PB equation. After solving the PB equation, the electrostatic potential is added into the QM Hamiltonian used in the treatment of the solute. In this way, the polarization of the solute by the reaction field is incorporated into the QM calculation. The polarized solute now yields a new set of atomic point charges, which will "repolarize" the surrounding solvent generating a new electrostatic potential upon solution

of the PB equation again. This iterative Self Consistent Reaction Field (SCRf) calculation is continued until convergence is reached. Upon convergence the electrostatic potential is used to calculate the electrostatic or polar part of the solvation free energy, while the nonpolar part is calculated from the solvent accessible surface area of the solute. A more detailed discussion of this method and its use within DivCon will be reported on in a future publication.²¹

Solvation Free Energy in a QM/MM Calculation. A QM/MM implementation of the PB model within DivCon allows specific regions of the protein to be polarized by solvent dynamically, while ignoring the solvent's influence on regions of lesser importance. The implementation described is briefly mentioned by Murphy et al.⁴³ but is further elaborated upon and applied to our own program in the following. A similar method using DFT methods is discussed by Li et al.⁴⁴ where the protein and solvent are continua. They explore the redox potential of two-sulfur, two-iron clusters and find that including the protein field with the solvent makes a large contribution toward the accurate description of the potential in the QM region. The QM/MM PB implementation described herein will, in general, reduce the overall computational expense versus a full QM treatment, because the Hartree–Fock equations are solved for a much smaller region. The current implementation also presents the opportunity to use the Divide and Conquer linear scaling semiempirical method for the calculation of both the protein and solvent while including the effects of all of the MM atoms on the QM region.

After the normal QM/MM setup of the protein system, the point charges of the protein are built up from the QM and MM regions. The MM charges are treated as fixed atomic point charges and do not fluctuate during the SCRf process. These are obtained from the AMBER atom types and passed into DivCon through the QM/MM interface. The QM charges are calculated from the density matrix of the

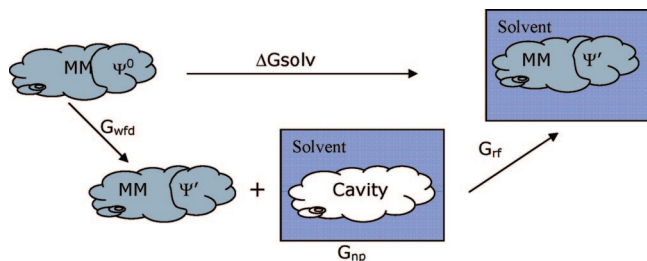


Figure 2. The terms involved in obtaining the free energy of solvation.

QM region using AM1⁴⁵ and are updated every SCRF cycle. DivCon allows the use of Mulliken, CM1,³⁷ or CM2³⁸ charges.

The entire collection of atomic point charges, from both the QM and MM regions, are then passed to the PB solver in DivCon, together with the coordinates of the protein. The dielectric constants are assigned based on the grid point location either in the solvent or protein as determined by the solvent accessible surface of the protein. The solver then solves the PB equation numerically based on the finite difference method and obtains the electrostatic potential. The resulting electrostatic potential is used to generate a set of virtual surface charges located on the solvent accessible surface of the protein. This set of virtual surface charges represents the electrostatic field (reaction field) produced by the polarized solvent. The Hamiltonian of the QM region in vacuum (H^0) is then perturbed by the interaction of the virtual surface charges with the solute electrons and nuclei and integrated over the solvent accessible surface of the protein, giving eq 6, the new potential energy operator. In this equation, $\sigma(r')$ is the virtual surface charge located at r' , and a' is an area of the solute surface (S):

$$H = H^0 + \int_S da' \frac{\sigma(r')}{|r - r'|} \quad (6)$$

The perturbed Hamiltonian results in a new set of atomic point charges in the QM region, which reflects the polarization of the QM solute. This updated set of QM charges, together with the fixed MM charges, is then fed into the PB solver for another cycle of the SCRF calculation and this continues until the polarization between solute and solvent achieves convergence. The converged reaction field (represented by the virtual surface charges) and solute charge distribution (wave functions in the QM representation) are used to calculate the polar part of the solvation free energy (G_{rf}):²

$$G_{rf} = \frac{1}{2} \int_V d\rho(r) \int_S dr' \frac{\sigma(r')}{|r - r'|} \quad (7)$$

A schematic view of a solvation free energy calculation, along with the terms represented in the above equations, is illustrated in Figure 2. During the dissolution of the protein into solvent, the solvent is polarized and generates a reaction field. The solute is also polarized by this reaction field, hence, the wave function of the QM portion (MM portion is fixed) is modified. We describe the energy difference between the two states (the vacuum and dissolved states) of the solute via the wave function distortion (wfd) term given by G_{wfd} .

The electrostatic interaction between the reaction field and the dissolved state of the solute is called the reaction field (rf) energy (G_{rf}). The G_{wfd} and G_{rf} constitute the polar part of the solvation free energy. There is also an entropy penalty associated with making a cavity in the solvent, which is usually combined together with the nonpolar (np) interaction between the solute and solvent. These two terms are both proportional to the solvent accessible surface area and are lumped together in the G_{np} term. Further technical details of how to calculate these terms are given in ref 2.

In summary, the QM/MM PB method makes it possible to study the protein polarized by solvent only in the area of interest without the expense of having to perform the SCRF calculations on the entire protein while still yielding a reliable solvation free energy for the entire macromolecule.

Results and Discussion

We evaluated the QM/MM PB method by comparing QM/MM calculations of various sizes of the QM region with full QM SCRF calculations. The Poisson–Boltzmann solver within our semiempirical QM program, DivCon, has been described and validated,^{2,21} so we will compare our QM/MM results with full QM results from this approach. Results comparing 65 small molecule's experimental solvation free energy^{46,47} to calculated solvation free energy in DivCon can be found in the Supporting Information. DivCon allows the internal and external dielectric values to be set to user-defined values. For these calculations we will use the values of 1.0 for the internal dielectric (the QM and MM regions) and 80.0 for external dielectric value, the surrounding "water" environment. A grid spacing of 2.5 gridpts/Å, and a probe radius of 1.40 Å is used throughout. These parameters were found to give the best cost to performance benefit in these calculations. We use 1.0 for the internal dielectric to avoid double counting the polarization in the QM region. Since the QM atoms are already polarized by the environment through the PB approach, a higher dielectric to approximate the polarization effect, as is utilized in fixed point charge PB approaches, is unnecessary. Pentapeptides, capped with acetyl (ACE) and N-methylamine (NME) on their respective termini, were calculated, and their solvation free energy was compared to a full QM calculation using the AM1⁴⁵ Hamiltonian on the same conformation. The QM region cut was made between the carbonyl carbon and the α carbon of the next alanine or between the nitrogen and the α carbon of the previous alanine so that the peptide bond was not cut. An example of the QM/MM cut and a demonstration of the "1 peptide" designation found below can be seen in Figure 1.

These structures were minimized with 10 steps of steepest decent followed by a maximum of 4990 steps of conjugate gradient using the AMBER minimizer and the ff99 force field. All calculations were done as single point calculations at the minimized geometry, and the AM1 Hamiltonian is used throughout. This level of minimization was found to be sufficient because, on average, the rmsd between the AM1 and the molecular mechanical force field was found to be only 0.3 Å for these small systems. Moreover, since the absolute solvation free energy was of less interest than the

relative solvation free energies of the different methods, this level of minimization was deemed appropriate. The ending point of the minimization is unimportant for the purpose of this study; all we required was a low energy structure that we could make comparisons from. We are not comparing directly to the experimental conformation for this peptide, so sampling and conformational issues were not a concern. The minimized coordinates for all of the pentapeptide systems are available in the Supporting Information.

The CM1 charge model³⁷ was used for the partial charges in the PB calculations. In these calculations a subset of the pentapeptides was tested with the CM2 charge model,³⁸ and a similar trend between CM1 and CM2 solvation free energies was observed. Hence, we will only show the CM1 results. The solvation free energy of a subset of pentapeptides was also examined using the PM3 Hamiltonian, and as with the CM1/CM2 models there was no appreciable difference in performance between AM1 and PM3. Hence, the QM level we use throughout will be AM1.

Increasing QM Size. Results from increasing QM region size on the pentapeptides used above can be found in Table 1. For this, the number of atoms included in the QM region was gradually decreased so the accuracy of the QM/MM PB method could be determined. As the QM region gets smaller it is expected that the accuracy, compared to a full AM1 treatment of the entire pentapeptide, should decrease. Since only the QM atomic charges are allowed to fluctuate, the number of atoms that are allowed to polarize in the solvent field decreases, and the number of fixed charges provided by the force field increases potentially leading to less accurate solvation free energies. This means that the amount of quantum information in the SCRF calculations shrinks, and a gradual shift out of agreement with full AM1 calculations might be expected. For these systems the solvation energy remains accurate with the 1 peptide systems, with the smallest QM region, having an unsigned average difference of only 4.9 kcal/mol. Not unexpectedly, the size of the QM region has a greater impact on charged residues than aliphatic ones. Charged or highly polar residues, such as proline and histidine, are more polarizable and as these residues/atoms are assigned fixed MM charges the polarization experienced by the SCRF procedure is diminished, hence, the solvation free energy is expected to change the most as the number of QM residues is decreased. This is certainly true for the positively charged ARG and LYS residues, while the negatively charged GLU and ASP residues are affected to a lesser extent.

Table 2 shows the unsigned differences between the QM/MM calculation and the full QM DivCon solvation free energy. As expected, the accuracy of the QM/MM PB results decreases as the QM region decreases. This demonstrates the importance of the polarizable charges in the QM region against the fixed MM charges. As the MM region grows the solvation energy becomes less accurate relative to the full QM calculations but can still get pertinent information regarding the solvation free energy. Applying this to a protein could give valuable insights into active site polarization while still producing reliable solvation free energies.

Table 2. Unsigned Differences between Full QM and QM/MM Solvation Free Energy Calculations^a

	5 peptide difference	4 peptide difference	3 peptide difference	2 peptide difference	1 peptide difference
ALA	0.0	1.5	1.6	1.7	1.9
ARG	0.0	0.2	0.6	2.4	12.6
ASN	0.0	0.3	1.7	3.0	5.1
ASP	0.1	4.0	3.2	0.2	3.4
CYS	0.0	0.1	1.4	2.0	3.4
GLU	0.0	1.8	1.1	0.7	0.4
GLN	0.0	2.2	2.9	1.1	2.5
GLY	0.0	1.4	2.0	2.6	3.6
HIS	0.1	6.6	10.2	13.1	16.7
ILE	0.0	0.0	0.1	0.8	0.6
LEU	0.1	0.1	0.4	0.1	0.5
LYS	0.0	4.4	7.3	16.6	21.2
MET	0.1	0.3	0.8	0.4	0.4
PHE	0.0	0.1	0.1	0.4	0.6
PRO	0.1	2.7	5.4	9.5	12.1
SER	0.1	1.5	0.4	2.2	4.1
THR	0.0	0.9	1.1	0.5	0.3
TRP	0.0	2.6	1.1	3.7	5.1
TYR	0.1	0.6	1.2	1.0	2.0
VAL	0.0	0.7	0.2	0.8	1.6
average	0.0	1.6	2.1	3.2	4.9

^a Unsigned differences in solvation energy between QM/MM PB calculations and full QM PB calculations, in kcal/mol. All energies are compared to an entire pentapeptide in DivCon solvation free energy using the AM1 Hamiltonian.

Solvation Free Energies of Small Proteins. This QM/MM method is intended to be used on biomolecular systems, more specifically proteins. While the previous test shows the method's ability to perform on small systems, tests on protein systems need to be performed to gauge its efficacy on the systems of interest. In order to examine this, a series of solvation free energy calculations was done on small proteins Hepatocyte nuclear factor 1- α (DCoH),⁴⁸ SARS-coronavirus accessory protein (Orf7a),⁴⁹ bovine pancreatic trypsin inhibitor (BPTI),⁵⁰ and T4 lysozyme⁵¹ containing 1284, 1034, 892, and 2603 total atoms, respectively (PDB ids: 1usm, 1xak, 4pti, 182l). First, the system was minimized within DivCon using the AM1 Hamiltonian and the LBFGS⁵² method. After this the solvation free energy of the entire minimized system was calculated with the AM1 Hamiltonian. Following these calculations, a small QM region was picked and expanded in 2–3 Å increments giving QM sizes as small as 33 atoms and as large as 1734 atoms. The results of these calculations are given in Table 3 with differences between full QM solvation energy and QM/MM solvation energies shown in Table 4. The proteins give unsigned average differences from the full AM1 QM calculations of 19.13 kcal/mol, 19.67 kcal/mol, 11.02 kcal/mol, and 9.08 kcal/mol for DCoH, Orf7a, BPTI, and lysozyme, respectively. These are in good agreement with the expected solvation free energies given from full QM calculations. There is only a weak to nonexistent trend between including more atoms in the QM region and increased accuracy in the solvation free energy as might be expected for these systems. The solvation free energy differences are generally acceptable in most cases and show that even reasonably sized QM regions (less than 100 atoms) in a QM/MM calculation should be expected to give solvation free energies in good accord with full QM calculations. Also, supplied in the Supporting Information,

Table 3. Solvation Free Energy of Increasing QM Size for Four Different Protein Systems^a

Å (no. of QM atoms)	solvation free energy			
	1usm	1xak	4pti	182l
3(33,40,62,40)	−452.62	−400.19	−780.44	−1411.92
5(59,98,107,54)	−467.23	−399.38	−779.70	−1412.87
8(187,161,176,182)	−462.01	−399.72	−805.40	−1391.85
10(236,243,308,262)	−459.19	−402.66	−802.21	−1403.46
12(421,281,373,342)	−464.64	−402.19	−777.16	−1399.31
15(696,405,490,630)	−458.71	−416.09	−818.12	−1387.74
18(959,525,621,925)	−495.69	−416.11	−804.70	−1385.38
21(1169,685,676,1290)	−456.99	−396.71	−801.81	−1401.54
24(1260,859,789,1734)	−442.26	−386.32	−798.18	−1417.58
27(913)	−	−392.52	−	−
AMBER Full QM	−443.24	−381.52	−803.50	−1401.69

^a Solvation free energies for the different proteins examined in this study. The left column is the distance expanded from a specific atom. The values in parentheses are the number of QM atoms, not counting link atoms, in the QM/MM calculations for 1usm, 1xak, 4pti, and 182l, respectively. All energies are in kcal/mol.

Table 4. Solvation Free Energy Differences for Four Proteins^a

Å (no. of QM atoms)	1usm	1xak	4pti	182l
3(33,40,62,40)	−9.38	−18.67	23.06	−10.23
5(59,98,107,54)	−23.99	−17.86	23.80	−11.18
8(187,161,176,182)	−18.77	−18.20	−1.90	9.84
10(236,243,308,262)	−15.95	−21.14	1.29	−1.77
12(421,281,373,342)	−21.40	−20.67	26.34	2.38
15(696,405,490,630)	−15.47	−34.57	−14.62	13.95
18(959,525,621,925)	−52.45	−34.59	−1.20	16.31
21(1169,685,676,1290)	−13.75	−15.19	1.69	0.15
24(1260,859,789,1734)	0.98	−4.80	5.32	−15.89
27(913)	−	−11.00	−	−

^a Differences in solvation free energies for the proteins examined. The left column is the distance expanded from a specific atom in Angstroms. The values in parentheses are the number of QM atoms, not counting link atoms, in the QM/MM calculations for 1usm, 1xak, 4pti, and 182l, respectively. All energies are in kcal/mol.

timings for 1usm for both standard and divide and conquer calculations can be found.

Conclusion

We have presented a QM/MM implementation of the SCRF Poisson–Boltzmann solver available in DivCon to calculate solvation free energies of large molecules within a semiempirical framework. A full QM version of this Poisson–Boltzmann solver has been validated by Liao and Merz,²¹ and we find that it is also applicable using a QM/MM approach as implemented in the Sander program of the AMBER suite of programs. This method allows the solvation free energy of an entire protein to be calculated, including only part of the system in the quantum region, thus allowing only these charges to be polarized by the solvent while keeping the remaining system at fixed charges given by the force field. Using a divide and conquer technique, a large quantum region may be selected due to DivCon's linear scaling nature allowing the link atoms to be farther from the active site than other methods may allow due to computational expense. This also allows the solvation free energy to be calculated using Mulliken, CM1, or CM2 charges for the QM region. This method might be applied, in particular, to drug design due to its ability to allow polarization and charge transfer effects in a region of interest while including solvation terms and finding the solvation free energy of a solute. It could also be used to determine pK_a and thermodynamic quantities

and generate improved force fields for nonstandard ligands/residues and many other applications relevant to biological systems.

Acknowledgment. We thank the NIH (GM066859) for financial support of this research.

Supporting Information Available: Average time per SCF cycle in vacuum and solution for the 1usm protein along with the number of SCF cycles required for convergence in vacuum and solution, experimental versus calculated solvation free energies from DivCon for 39 neutral, 19 anions, and 17 cations along with the coordinates of the 20 pentapeptides used, and coordinates for all the pentapeptides. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- (1) Friesner, R. A. Modeling Polarization in Proteins and Protein-ligand Complexes: Methods and Preliminary Results. *Adv. Protein Chem.* **2005**, 72, 79–104.
- (2) Gogonea, V.; Merz, K. M. Fully quantum mechanical description of proteins in solution. Combining linear scaling quantum mechanical methodologies with the Poisson–Boltzmann equation. *J. Phys. Chem. A* **1999**, 103 (26), 5171–5188.
- (3) Honig, B.; Nicholls, A. Classical Electrostatics in Biology and Chemistry. *Science* **1995**, 268 (5214), 1144–1149.
- (4) Sharp, K. A.; Honig, B. Electrostatic interactions in macromolecules: theory and applications. *Annu. Rev. Biophys. Chem.* **1990**, 19, 301–32.

- (5) Schaefer, P.; Ricciardi, D.; Cui, Q. Reliable treatment of electrostatics in combined QM/MM simulation of macromolecules. *J. Chem. Phys.* **2005**, *123* (1), 014905.
- (6) Simonson, T.; Archontis, G.; Karplus, M. Free energy simulations come of age: protein-ligand recognition. *Acc. Chem. Res.* **2002**, *35* (6), 430–47.
- (7) Simonson, T. Macromolecular electrostatics: continuum models and their growing pains. *Curr. Opin. Struct. Biol.* **2001**, *11* (2), 243–52.
- (8) Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L. 3rd, Assessing scoring functions for protein-ligand interactions. *J. Med. Chem.* **2004**, *47* (12), 3032–47.
- (9) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935.
- (10) Roux, B.; Simonson, T. Implicit solvent models. *Biophys. Chem.* **1999**, *78* (1–2), 1–20.
- (11) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
- (12) Bashford, D.; Case, D. A. Generalized born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129–52.
- (13) Tsui, V.; Case, D. A. Molecular dynamics simulations of nucleic acids with a generalized born solvation model. *J. Am. Chem. Soc.* **2000**, *122* (11), 2489–2498.
- (14) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Struct., Funct., Bioinf.* **2004**, *55* (2), 383–394.
- (15) Edinger, S. R.; Cortis, C.; Shenkin, P. S.; Friesner, R. A. Solvation free energies of peptides: Comparison of approximate continuum solvation models with accurate solution of the Poisson-Boltzmann equation. *J. Phys. Chem. B* **1997**, *101* (7), 1190–1197.
- (16) Cramer, C. J.; Truhlar, D. G. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.* **1999**, *99* (8), 2161–2200.
- (17) Tomasi, J.; Persico, M. Molecular Interactions in Solution: An Overview of Methods Based on Continuous Distributions of the Solvent. *Chem. Rev.* **1994**, *94* (7), 2027–2094.
- (18) Fogolari, F.; Brigo, A.; Molinari, H. The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *J. Mol. Recognit.* **2002**, *15* (6), 377–92.
- (19) Nicholls, A.; Honig, B. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J. Comput. Chem.* **1991**, *12* (4), 435–445.
- (20) Tannor, D. J.; Marten, B.; Murphy, R.; Friesner, R. A.; Sitkoff, D.; Nicholls, A.; Ringnalda, M.; Williams, A.; Goddard, I.; Honig, B. *J. Am. Chem. Soc.* **1994**, *116*, 11875–82.
- (21) Liao, N.; Merz, K. M. Manuscript in preparation.
- (22) Grater, F.; Schwarzl, S. M.; Dejaegere, A.; Fischer, S.; Smith, J. C. Protein/ligand binding free energies calculated with quantum mechanics/molecular mechanics. *J. Phys. Chem. B* **2005**, *109* (20), 10474–83.
- (23) Archontis, G.; Simonson, T.; Karplus, M. Binding free energies and free energy components from molecular dynamics and Poisson-Boltzmann calculations. Application to amino acid recognition by aspartyl-tRNA synthetase. *J. Mol. Biol.* **2001**, *306* (2), 307–27.
- (24) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98* (18), 10037–41.
- (25) Smith, P. E.; Pettitt, B. M. Modeling Solvent in Biomolecular Systems. *J. Phys. Chem.* **1994**, *98* (39), 9700–9711.
- (26) Warshel, A.; Levitt, M. Theoretical studies of enzymatic reactions: Dielectric, electrostatic and steric stabilization of a carbonium ion in the reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.
- (27) Field, M. J.; Bash, P. A.; Karplus, M. A Combined Quantum-Mechanical and Molecular Mechanical Potential for Molecular-Dynamics Simulations. *J. Comput. Chem.* **1990**, *11* (6), 700–733.
- (28) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Matthews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. *AMBER 9*, University of California: San Francisco, 2006.
- (29) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4* (2), 187–217.
- (30) Senn, H. M.; Thiel, W. QM/MM methods for biological systems. In *Atomistic Approaches in Modern Biology: From Quantum Chemistry to Molecular Simulations*; Reiher, M., Ed.; Springer: Berlin, Heidelberg, 2007; Vol 268, pp 173–290.
- (31) Li, G. H.; Cui, Q. Mechanochemical coupling in myosin: A theoretical analysis with molecular dynamics and combined QM/MM reaction path calculations. *J. Phys. Chem. B* **2004**, *108* (10), 3342–3357.
- (32) Noodleman, L.; Lovell, T.; Han, W. G.; Li, J.; Himo, F. Quantum chemical studies of intermediates and reaction pathways in selected enzymes and catalytic synthetic systems. *Chem. Rev.* **2004**, *104* (2), 459–508.
- (33) Ranaghan, K. E.; Ridder, L.; Szeftczyk, B.; Sokalski, W. A.; Hermann, J. C.; Mulholland, A. J. Transition state stabilization and substrate strain in enzyme catalysis: ab initio QM/MM modelling of the chorismate mutase reaction. *Org. Biomol. Chem.* **2004**, *2* (7), 968–980.
- (34) Hensen, C.; Hermann, J. C.; Nam, K.; Ma, S.; Gao, J.; Holtje, H. D. A combined QM/MM approach to protein-ligand interactions: polarization effects of the HIV-1 protease on selected high affinity inhibitors. *J. Med. Chem.* **2004**, *47* (27), 6673–80.
- (35) Dixon, S. L.; Merz, K. M. Semiempirical molecular orbital calculations with linear system size scaling. *J. Chem. Phys.* **1996**, *104* (17), 6643–6649.
- (36) Dixon, S. L.; Merz, K. M. Faster, accurate semiempirical molecular orbital calculations for macromolecules. *J. Chem. Phys.* **1997**, *107* (3), 879–893.
- (37) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. Class-Iv Charge Models - a New Semiempirical Approach in Quantum-Chemistry. *J. Comput.-Aided Mol. Des.* **1995**, *9* (1), 87–110.

- (38) Li, J. B.; Zhu, T. H.; Cramer, C. J.; Truhlar, D. G. New class IV charge model for extracting accurate partial charges from wave functions. *J. Phys. Chem. A* **1998**, *102* (10), 1820–1831.
- (39) Yang, W. T.; Lee, T. S. A Density-Matrix Divide-and-Conquer Approach for Electronic-Structure Calculations of Large Molecules. *J. Chem. Phys.* **1995**, *103* (13), 5674–5678.
- (40) Davis, M. E.; Mccammon, J. A. Solving the Finite-Difference Linearized Poisson-Boltzmann Equation - a Comparison of Relaxation and Conjugate-Gradient Methods. *J. Comput. Chem.* **1989**, *10* (3), 386–391.
- (41) Holst, M.; Kozack, R. E.; Saied, F.; Subramaniam, S. Treatment of electrostatic effects in proteins: multigrid-based Newton iterative method for solution of the full nonlinear Poisson-Boltzmann equation. *Proteins* **1994**, *18* (3), 231–45.
- (42) Warwicker, J.; Watson, H. C. Calculation of the electric potential in the active site cleft due to α helix dipoles. *J. Mol. Biol.* **1982**, *157* (4), 671–9.
- (43) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments. *J. Comput. Chem.* **2000**, *21* (16), 1442–1457.
- (44) Li, J.; Nelson, M. R.; Peng, C. Y.; Bashford, D.; Noodleman, L. Incorporating protein environments in density functional theory: A self-consistent reaction field calculation of redox potentials of [2Fe2S] clusters in ferredoxin and phthalate dioxygenase reductase. *J. Phys. Chem. A* **1998**, *102* (31), 6311–6324.
- (45) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. P. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
- (46) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. J. Group Contributions to the Thermodynamic Properties of Non-ionic Organic Solutes in Dilute Aqueous Solution. *J. Solution Chem.* **1981**, *10* (8), 563–595.
- (47) Wolfenden, R.; Anderson, L.; Cullis, P. M.; Southgate, C. C. Affinities of Amino Acid Side Chains for Solvent Water. *Biochemistry* **1981**, *20*, 849–855.
- (48) Tahirov, T. H.; Inagaki, E. Unpublished PDB entry 1usm. <http://www.rcsb.org>(accessed Oct 3, 2007).
- (49) Nelson, C. A.; Lee, C. A.; Fremont, D. H.; Joachimiak, A. Structure and intracellular targeting of the SARS-coronavirus Orf7a accessory protein. *Structure* **2005**, *13*, 75–85.
- (50) Huber, R.; Kukla, D.; Ruehlmann, A.; Epp, O.; Formanek, H.; Deisenhofer, J.; Steigemann, W. The Geometry of the Reactive Site and of the Peptide Groups in Trypsin, Trypsinogen and its Complexes and Inhibitors. *Acta Crystallogr.* **1983**, *39*, 480.
- (51) Morton, A.; Matthews, B. W. Specificity of ligand binding in a buried nonpolar cavity of T4 lysozyme: linkage of dynamics and structural plasticity. *Biochemistry* **1995**, *34* (27), 8576–88.
- (52) Liu, D. C.; Nocedal, J. On the limited memory BFGS method for large scale optimization. *Math. Prog.* **1989**, *45*, 503–528.

CT700245A