J Mol Model
DOI 10.1007/s00894-008-0306-z

ORIGINAL PAPER

Optimization of cutting schemes for the evaluation of molecular electrostatic potentials in proteins via Moving-Domain QM/MM

Lochana C. Menikarachchi · José A. Gascón

Received: 13 February 2008 / Accepted: 19 March 2008

Abstract This work presents new developments of the moving-domain QM/MM (MoD-QM/MM) method for modeling protein electrostatic potentials. The underlying goal of the method is to map the electronic density of a specific protein configuration into a point-charge distribution. Important modifications of the general strategy of the MoD-QM/MM method involve new partitioning and fitting schemes and the incorporation of dynamic effects via a single-step free energy perturbation approach (FEP). Selection of moderately sized QM domains partitioned between C_α and C (from C=O), with incorporation of delocalization of electrons over neighboring domains, results in a marked improvement of the calculated molecular electrostatic potential (MEP). More importantly, we show that the evaluation of the electrostatic potential can be carried out on a dynamic framework by evaluating the free energy difference between a non-polarized MEP and a polarized MEP. A simplified form of the potassium ion channel protein *Gramicidin-A* from *Bacillus brevis* is used as the model system for the calculation of MEP.

Keywords Molecular electrostatic potential · Moving-Domain QM/MM · Polarization

Introduction

Electrostatic interactions play a critical role in many biologically important processes involving proteins [1–3]. Examples of events that are controlled or modulated by electrostatic effects are enzymatic reactions [4–6], protein-ligand binding [7–9], and ion-selectivity in transmembrane channels [10–12]. The use of first principles methods to describe such events is clearly prohibited. This problem motivated the development of classical molecular mechanics (MM) force fields over the past 20 years. Although MM force fields are today an invaluable tool in modeling biologically relevant events [13–18], current implementations are based on empirical force fields that only effectively describe the quantum mechanical energy surface. One of the most critical approximations concerns the treatment of electrostatic effects, which are simply defined by Coulombic interactions between static charges. The static nature of this approximation does not allow electronic response (fluctuations of multipoles) due to changes in the chemical environment. In a number of cases, such shortcomings preclude a quantitative comparison and interpretation of experimental observables [19, 20]. In recent years, development of a new generation of polarizable force fields, able to treat electronic response, has been the subject of intense exploration [19–27]. A common philosophy of these methods is that any required parametrization (e.g., atom polarizabilities, electronegativities, etc.) is done *a priori* for a target set, with the assumption that polarization phenomena will emerge for a general biomolecular system.

A different approach lies within the field of linear-scaling semiempirical methods, which by construction include electronic response to the changes in the chemical environment (including bond breaking) [28–31] and are able to treat thousand of atoms. A liability of semiempirical methods is that their implementation on a dynamical framework for large biomolecular systems can lead to unrealistic conformations (e.g., distorted protein secondary structure, untwist of protein α helices) [32, 33]. Neverthe-

L. C. Menikarachchi · J. A. Gascón (✉)
Department of Chemistry, University of Connecticut,
55 North Eagleville Rd, Unit 3060,
Storrs, CT 06269, USA
e-mail: jose.gascon@uconn.edu

less, further development of semiempirical methods applied to protein conformation and MEP derived properties is a subject of continued interest [34, 35].

With the goal of obtaining accurate electrostatic potentials that are specific to a particular protein and are derived from first principles methods (density based or wave-function based), we have [36] recently proposed a self-consistent protocol that exploits the intrinsic architecture of QM/MM calculations, in particular that of the ONIOM method [37–39]. The method is analogous to that proposed by Gao for the simulation of liquids, in which QM/MM is used to determine the wave function of individual molecules [40–42]. The QM/MM protocol proposed by Gascón et al., termed moving-domain QM/MM (MoD-QM/MM), partitions a large molecular system into a number of molecular domains (typically one or two amino acid residues per partition) and iteratively performs a set of QM/MM calculations redefining the QM region in each iteration. Since each QM/MM calculation is performed within the so-called electronic embedding (EE) approach [43], the QM regions are *polarized* by the MM region. Therefore, a mapping of the electronic density of each domain into a distribution of monopoles (or multipoles in general) provides electronic response to the specific chemical environment of each domain. As expected, MoD-QM/MM inherits the critical problem associated with boundary effects typical of QM/MM methods. In addition, the MEP calculated from MoD-QM/MM is non-transferable to other protein conformations beyond those near the reference conformation used in the self-consistent procedure, since implementation of the method at each step of a molecular dynamics (MD) simulation would be impractical. To address these issues, this work presents new strategies to optimize boundary schemes in QM/MM calculations and explores optimal conditions to incorporate dynamical effects via a free energy perturbation approach (FEP). The system chosen to examine these new strategies is a reduced model of the gramicidin A (gA) channel, whose structure has been solved at high resolution [44, 45]. The interest in exploring polarization effects in biological ion channels is based on the need of obtaining quantitative comparison with conductivity measurements, which are currently reproduced only semi-quantitatively by standard force fields [11, 46].

Section QM/MM method of this paper describes the moving-domain QM/MM method and outlines different cutting schemes for the QM-MM boundary. Section Protein channel model describes the ion channel model used for the calculation of MEP, and Sect. MD simulations describes the MD simulations carried out with the MM reference potential. Section Result explores the various cutting schemes as judged by comparing the MEP to full quantum mechanical calculations and Sect. Towards the incorporation of finite temperature effects outlines the strategies to perform a single step free energy perturbation evaluation to change from the MM surface to the MoD-QM/MM surface. This section also discusses convergence issues associated with the free energy perturbation approach. Section Conclusion summarizes the results and concludes.

QM/MM method

QM/MM hybrid methods partition the system into QM and MM layers. Within the so called electronic embedding approach (EE) [43], the electrostatic influence of the surrounding molecular environment is explicitly considered. The particular methodology used in our study is the ONIOM-EE approach [37–39]. In this approach, a small molecular domain, called region *X*, is treated according to a rigorous *ab initio* quantum chemistry method, while the rest of the system, region *Y*, is treated according to a MM force field. For systems where both regions are covalently bonded, the covalency of the frontier QM atoms is completed according to the link-hydrogen atom scheme [37]. The QM/MM energy is obtained via an extrapolation approach according to three independent calculations:

$$E(QM/MM) = E(QM)_X + E(MM)_{X+Y} - E(MM)_X, \qquad (1)$$

where $E(QM)_X$ is the energy of region X at the QM level, $E(MM)_{X+Y}$ is the energy of the entire system at the MM level, and $E(MM)_X$ is the energy of X at the MM level. In the EE approach electrostatic interactions between X and Y are considered in each of the terms of the right hand side of Eq. (1), so that electrostatic interactions are cancelled out at the MM level, but remain at the QM level. On the other hand, Van der Waals interactions between X and Y are only considered at the MM level (in the term $E(MM)_{X+Y}$). Since the QM region interacts with the MM environment, calculation of atomic charges, derived from the QM electronic density, are *polarized* by the MM region.

MoD-QM/MM

MoD-QM/MM involves a decomposition of the system into molecular domains. These domains are sequentially treated as quantum mechanical layers embedded in a molecular mechanics environment that polarizes the QM region. The main goal is to obtain a distribution of atomic charges that describes the proper polarization of the individual molecular fragments according to the total electric field of the specific protein of interest. For systems for which the monopole approximation is not appropriate, the method can be easily extended to obtain higher multipoles. It is, in fact, well documented that an atom can present well marked positive and negative regions [47]. In any case, the underlying philosophy is to compute the *most reliable*

electrostatic interactions in the *specific* biomolecule of interest, as opposed to obtaining average atomic charges that only *approximately* describe the electrostatic interactions in a *large* number of proteins. Therefore, while accurate, the computed electrostatic potential is most useful for describing electrostatic interactions associated with the system near the configuration chosen in the implementation of the protocol. The MoD-QM/MM protocol begins with the partitioning of the system into *n* molecular domains of suitable sizes for QM/MM calculations. To describe the different cutting schemes, it is assumed that each molecular domain consists of a molecular fragment of an average sized amino acid. However, there is no restriction on the size and chemical nature of the domains. Starting at the residue R_1 as the first QM domain, while the rest is MM, the distribution of electrostatic-potential (ESP) atomic charges in R_1 is obtained by performing a single point QM/MM calculation using electronic embedding. The QM layer is then *moved* from R_1 to another residue R_2 and the polarized distribution of charges in R_2 is obtained analogously. The calculation on R_2 considers the updated charge distribution previously computed for R_1. The QM/MM polarization is implemented for all residues in the protein, R_1, …R_n and the cycle is repeated until the values of atomic charges converge [36]. Figure 1 illustrates the MoD-QM/MM protocol as applied to a simple molecular structure with *n* residues. The surfaces represent the QM regions in QM/MM calculations. Regions with updated charges are represented with ball-and-sticks diagramas.

Preliminary results obtained by applying the method to various protein systems show that the distribution of charges converges in a small number of cycles, independent of the size of the protein, and independent of the order in which individual domains are *polarized* [36]. Thus, provided that a finite cutoff for non-bonded interactions is implemented, the resulting method scales linearly with the size of the protein. The current application of MoD-QM/MM exploits the architecture of the ONIOM method [37–39], as implemented in *Gaussian 03* [48]. The program that carries out the self-consistent polarization, MODQ3M, was written in our laboratory.

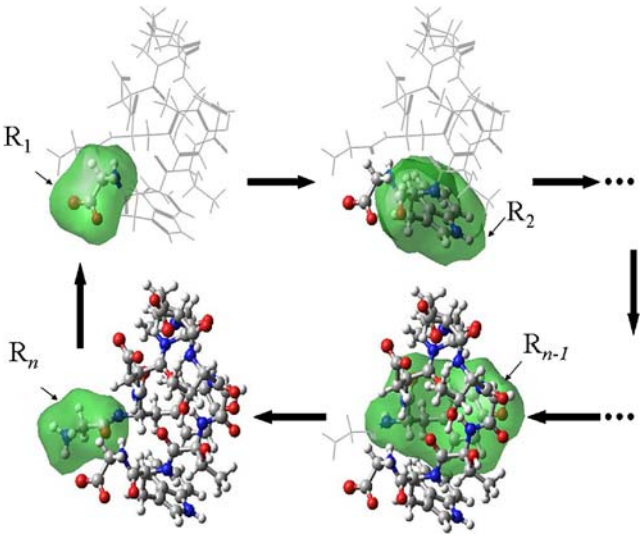

Fig. 1 Schematic representation of the MoD-QM/MM method. Green surfaces represent the QM region in QM/MM calculation. Colored balls and sticks represent regions with updated charges

Cutting schemes in QM/MM

In the ONIOM-EE approach used in this study, the QM domain is treated at the HF/6–31g* level while the rest of the system is treated with molecular mechanics, starting from the AMBER-99 assignment of charges for the amino acid residues. The final converged charge distribution is actually independent of the initial distribution of charges. On the other hand, as discussed later, one can also start with a particular *unpolarized* distribution of charges and restrain the charges derived by MoD-QM/MM to this initial distribution. Partitioning a system into individual molecular domains requires a proper treatment of the boundary between the layers. In MoD-QM/MM, the covalency of frontier atoms is completed according to the standard link-hydrogen atom scheme.

Figure 2 presents the different types of partitioning schemes explored in this study. In Scheme 1 the boundary between the QM and MM region is placed on the amide bond (Fig. 2, top left). In Scheme 2 the boundary is placed on the bond between the C_α and carbonyl carbon (Fig. 2, top right). In both Scheme 1 and 2 the fitting procedure is carried out by restraining the total charge of region 1 to be equal to the total charge of the QM domain (Q). Therefore, charges on region 2 must add up to zero. These conditions are expressed in the following equations:

$$\sum_{i=1}^{M} q_i = Q \tag{2}$$

$$\sum_{i=M+1}^{N} q_i = 0 \tag{3}$$

where N is the total number of atoms in the QM domain (amino acid plus link atoms), and M is the number of atoms in region 1. An alternative fitting method is one in which the calculation of ESP charges is carried out with the single restrain $\sum_{i=1}^{N} q_i = Q$. However, in order to conserve charge after subsequent QM/MM calculations on neighboring domains, the total charge of region 1 is forced, *a posteriori*, to add up to Q. This can be simply accomplished by shifting the charge of the link atoms into the atoms to which they are attached. This shifting scheme turns out to be

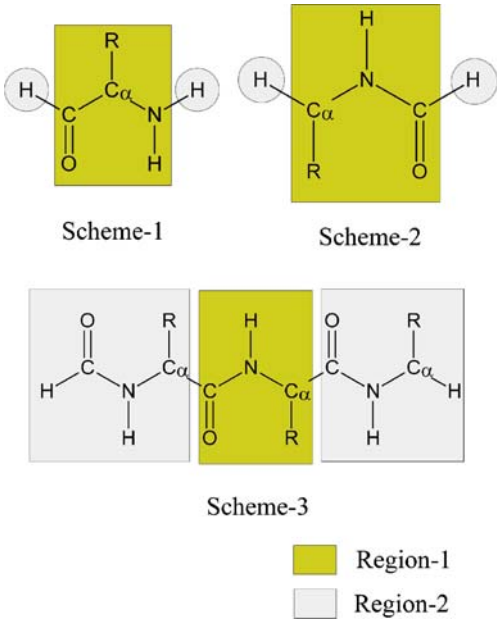

Fig. 2 Representations of the regions used in the fitting procedure for each QM/MM calculation. Although domains are shown here containing one amino acid, region 1 may actually contain more than one amino acid, or any prosthetic residue

substantially better than the scheme represented by Eqs. 2 & 3, as far as reproducing the MEP of a polypeptide compared to a full QM benchmark. This is the fitting procedure used in the present work for Schemes 1 and 2. Scheme 3 utilizes a cutting scheme similar to Scheme 2 with an additional feature that enables, in an implicit way, delocalization of electrons over neighboring domains. In Scheme 3 partitions are redefined by grouping three consecutive domains (or two if the original domains include a terminal residue). The fitting of ESP charges is done by restraining the total charge of the central domain to the charge of that amino acid. More importantly, updating of charges is done only on that central domain. Although Scheme 3 is computationally more expensive, it avoids boundary effects by implicitly incorporating delocalization of electrons over neighboring domains. Delocalization is only implicit because the total charge on the central domain is still restrained to be an integer. Similarly to Eqs. 2 & 3 the charge constraints for Scheme 3 are:

$$\sum_{i=1}^{M} q_i = Q \tag{4}$$

$$\sum_{i=M+1}^{N} q_i = Q_2 \tag{5}$$

where Q_1 is the charge of the central domain (region 1), Q_2 is the total charge of neighboring domains (region 2), and $Q = Q_1 + Q_2$.

Protein channel model

The gramicidin model is based on the X-ray structure with access code 1MAG. To compare with full QM calculations, only one monomer was considered, while Trp, Leu, and Val amino acids were mutated to Ala (modification of the side chains, only affects the outside of the channel). The resulting model contains 149 atoms suitable for a full QM benchmark calculation (see Fig. 3). The sequence of amino acids in the model is A-GA-AA-AA-AA-AA-AA-AA. As indicated by "-", the channel is partitioned into eight domains. The first domain is A, the second is GA, and all remaining six domains consist of two Ala residues. Despite that all side chains are non-polar, the backbone is still susceptible to polarization by the ion. It is important to remark that the purpose of this study is to fine tune the moving-domain QM/MM computational protocol to reproduce ab-initio quality MEPs, and not to model specific features of gramicidin channels. All calculations were performed along the internal axis of the channel with potassium ion located at the center of mass of the channel. The geometry chosen for obtaining charges and calculating the MEP corresponds to snapshots of the MD simulations described below. The channel's principle axis of inertia was placed along the Z axis. The top of the channel was placed at z = 0. Thus, the MEP along the Z axis is defined as:

$$u(z) = \sum_{i=1}^{N} \frac{q_i}{|\mathbf{R}_i - \mathbf{r}_z|}, \tag{6}$$

where R_i are the positions of the channel atoms, r_z are position vectors along the Z axis, and the sum goes over the atoms on the peptide chain (N = 149). The benchmark quantum chemistry calculation is an ONIOM-EE calculation with the entire polypeptide in the QM layer while the potassium atom was defined as the only MM atom. Since the MEP is calculated from the charges that belong to the polypeptide, allowing charge transfer to and from the ion would prevent a fair comparison with a force field represented by a fixed set of charges. Although any level of theory can be applied for the QM part, for a fair comparison with the Amber force field, we use here HF/6–31g*, which is the method and basis set used in the force field parametrization [16].

MD simulations

Molecular dynamics simulations were carried out at 300 K using the *NVT* ensemble for 300 picoseconds. After the first 100 ps, snapshots were saved at intervals of 0.1 ps. To maintain the structural integrity of the channel, atoms were restrained near their original X-ray positions. Atoms within a 1 Å radius from their original structure are free to move.

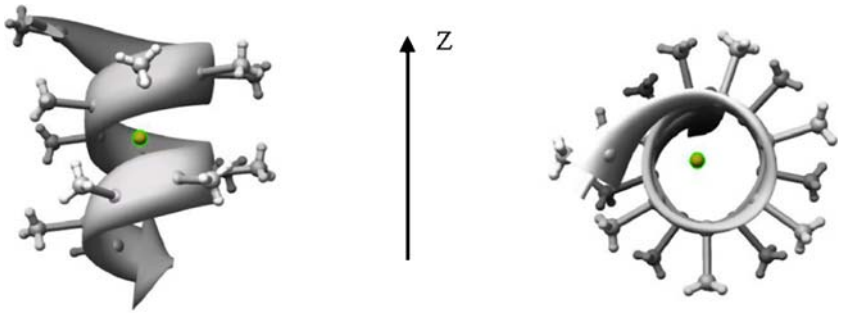

Fig. 3 Side and top views of the ion channel model used in the present study. Although the Z-axis is shown here outside the channel, it passes through the center of mass of the channel, located in its interior

A potassium atom was placed on the z-axis, roughly at the center of mass of the channel. When atoms leave that sphere they are restrained with a spring centered at their original position with a force constant of 100 kcal/(mol Å^2). These 2000 saved configurations are used later for the evaluation of the free energy difference of changing from the reference MM potential given by the *unpolarized* set of charges to the MM potential given by the MoD-QM/MM charges. For evaluation of the MEP in the comparison of different cutting schemes a snapshot of this MD simulation is used. For the evaluation of the free energy profile, a potassium ion was constrained on N_P points along the Z-axis at intervals of 0.2 Å. The free energy along the Z-axis is defined as

$$F(z) - F_0 = \int_0^z \left\langle \frac{\partial E}{\partial z'} \right\rangle_{z'} dz' \approx \sum_{i=1}^{N_P} \left\langle \frac{\partial E}{\partial z'} \right\rangle_{z_i'} \Delta z_i', \qquad (7)$$

where F_0 is defined as $F(z = 0)$ and $f_i \equiv \left\langle \frac{\partial E}{\partial z'} \right\rangle_{z_i'}$ is the average of the z component of the force, at position z_i', along the MD simulation. After 40 picoseconds of equilibration, averages were calculated over 80 picoseconds for each point on the Z-axis. The error in the free energy at position z_i is determined as $\sigma(z_i) = \sqrt{\sum_{j=1}^{i} \Delta z_j^2 \sigma_{fj}^2}$, where σ_{fi}^2 is the error of the average force at position z_i. This error was estimated as the standard deviation of the mean value calculated over 10 different trajectories. Typical values of $\sigma(z_i)$ are between 0.1 to 0.5 kcal mol^{-1}.

Results

In general, the MEPs generated by the various MoD-QM/MM schemes show a substantial improvement with respect to the electrostatic potentials given by the force field Amber, as compared to the benchmark quantum calculation. As shown in Fig. 4, the MEP obtained via MoD-QM/MM Scheme 1 closely reproduces the shape of the MEP curve obtained from the benchmark calculation. Although the overall root mean square deviation (RMSD) with respect to the full QM calculation is just 1.2 kcal/mol/C smaller than for the standard force field, the slope of the curve is better reproduced by MoD-QM/MM. The curve generated via the fixed-charge force field Amber deviates significantly in the middle region of the channel where the positively charged potassium ion is located. This clearly demonstrates the importance of polarization effects in determining accurate molecular electrostatic potentials. It is evident from the figure that, at the start of the channel, the MEP is poorly reproduced by MoD-QM/MM (differences can be as large as 10 kcal mol^{-1}). This region coincides with the location of the smallest of the domains (a single Ala residue). This is expected since boundary effects, inherent to the QM/MM ONIOM-EE method, have a larger influence the smaller the QM domain is. These boundary effects are mainly due to two different problems. The first is the overpolarization of the bond containing the link atom [49], which originates in the treatment of the charge near the boundaries.

Various approaches aimed to reduce this overpolarization effect have been proposed [50–55]. In particular, here we use the default scheme implemented in *Gaussian 03* in which charges on atoms within one, two, and three bonds from the QM boundary are set equal to zero. It is also expected that this overpolarization effect be more critical if the QM/MM cut is placed on a polar bond, such as the

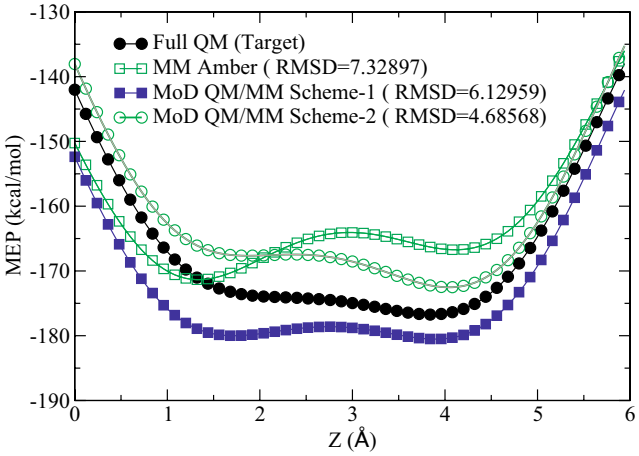

Fig. 4 Molecular electrostatic potential (MEP) generated by MoD-QM/MM using cutting Scheme 1 (cut between amide bond) and Scheme 2 (cut between C_α and C)

amide bond used in Scheme 1. It has, in fact, been pointed out that cutting amide bonds can cause severe overpolarization [56]. The second problem is not allowing for charge transfer onto neighboring domains. These two problems are, in some extent, addressed in the design of Schemes 2 and 3. MoD-QM/MM Scheme 2 partitions the amino acids by breaking C_α-carbonyl carbon bonds. The MEP evaluated using Scheme 2 shows a considerable improvement with respect to Scheme 1, both in terms of RMSD and relative fluctuations in the MEP (see Fig. 4). MEP calculated from MoD-QM/MM Scheme 3 (see Fig. 5), which incorporates charge transfer implicitly, shows a substantial improvement over the other MoD-QM/MM schemes. The computational cost of using Scheme 3 is about two to three times the cost of using Schemes 1 or 2 (still within linear scaling). To make a final comparison with a non-polarizable force field, Fig. 6 shows the molecular electrostatic potential obtained via Amber, MoD-QM/MM using Scheme 3, and the benchmark calculation. Clearly, there is a major improvement as a result of including polarization and delocalization effects via MoD-QM/MM.

Toward the incorporation of finite temperature effects

The MoD-QM/MM method is in principle applicable for a well-defined configuration of the protein or for configurations near the reference conformation used in the implementation of the method. Thus, due to the geometry dependency of the method, the question is: can it be implemented on a dynamic framework? Polarization of a large size biomolecule (> 1000 atoms) at each step of a MD simulation would be computationally very expensive, only attainable with massive parallelization. On the other hand, one can resort to free energy approaches that combine fast sampling using approximate MM force fields with a less

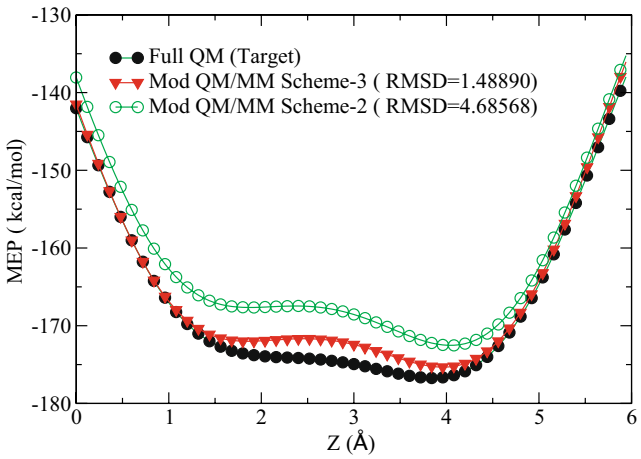

Fig. 5 Molecular electrostatic potential (MEP) generated by MoD-QM/MM using cutting Scheme 2 (cut between C_α and C) and Scheme 3 (same cut as 2 plus implicit delocalization)

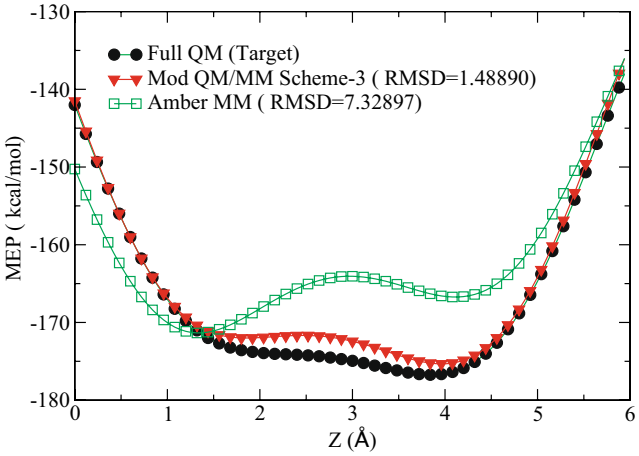

Fig. 6 Molecular electrostatic potential (MEP) generated by MoD-QM/MM according to cutting Scheme 3 (including implicit delocalization) and the fixed charged force field Amber

frequent but more accurate evaluation using higher theory levels such as QM or QM/MM [57–62]. The relevant quantity to evaluate in these approaches is the average

$$F^* - F^0 = -kT \ln \left\langle e^{-\beta\left(V^* - V^0\right)} \right\rangle_0 \tag{8}$$

where F^* is the free energy at the higher theory level (QM or QM/MM), F^0 is the free energy at the MM level, V^*-V^0 is the difference in energy between the two levels at a given conformation, k is the Boltzmann constant, T is the temperature and β=kT. The symbol $\langle\rangle_0$ indicates that the average is evaluated using the partition function of the reference potential V^0. It is important to remark that the formulation expressed by Eq. (8) is exact. Practical evaluation of Eq. (8) is, however, extremely difficult due to convergency problems [63]. Using a less accurate reference potential comes to the expense of having a poor correlation with the high theory energy surface, which leads to extremely large standard deviations in the evaluation of the statistical average of Eq. (8). Only when the energy surfaces are highly correlated (V^* is a perturbation with respect to V^0) then the average converges. In such cases, Eq. (8) is usually used as synonymous of free energy perturbation theory. The problem is that such situations do not occur often in the transition from MM to QM or QM/MM unless certain approximations are made [60–62]. We propose here that, since MoD-QM/MM reevaluates the electrostatic interactions, leaving all bonded interactions and Van der Waals interactions intact, a free energy perturbation approach would apply. Thus, the hypothesis to explore is whether the average in Eq. (8) converges as a result of using the charges obtained via MoD-QM/MM. As a first attempt, we only evaluate the difference V^*-V^0 taking into account the intermolecular coulombic interaction

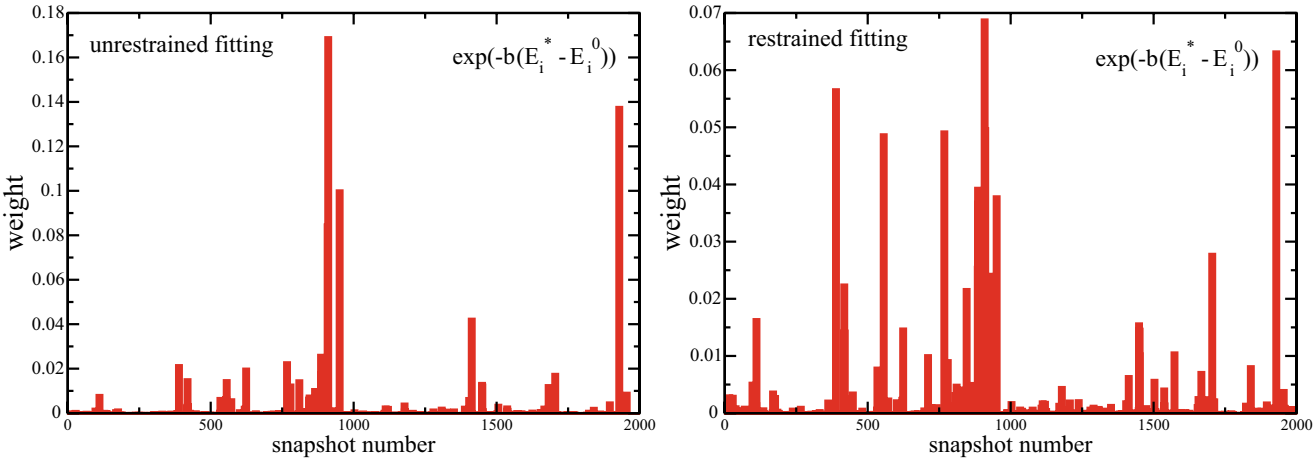

Fig. 7 Weight of each term in the average of Eq. (6) for each of the collected snapshots. *Left*, fitting is carried out using Scheme 1 as described in Sect. MoD-QM/MM. *Right*, fitting is carried out as in Scheme 1, constraining the charges to be close to those in the Amber reference potential

between the channel and the ion, which are the terms more largely affected by polarization effects. Since all other terms in the MM force field are left unchanged, the difference V^*-V^0 is equal to E^*-E^0, where E^* (E^0) is the intermolecular electrostatic energy (polypeptide-ion electrostatic interactions). Figure 7 (*left*) shows the contribution of each of the snapshots collected along the MD simulation (using Amber) to the average of Eq. (8). In the calculation of MoD-QM/MM charges for all these snapshots, fitting Scheme 1 was used. The distribution of weights shows a marginal convergence state, with about 8 configurations contributing to 60% of the average $\left\langle e^{-\beta\left(E^*-E^0\right)}\right\rangle_0$.

In order to improve convergence of this average we added an additional constrain to the fitting Scheme 1. This additional constrain is represented in the following equation:

$$\chi^2 = \sum_{j=1}^{N_g}\left(u_j - U_j\right)^2 + \sum_{i=1}^{M} w_i\left(q_i - q_i^0\right)^2, \tag{9}$$

where U_j is the QM electrostatic potential at a grid point j, N_g is the number of grid points around the QM domain used to fit the MEP, u_j is the MEP generated by the point charges q_i (to be determined), q_i^0 are the initial Amber charges, and w_i is the weight for each of these additional constraints. The fitting procedure is based on the analytical minimization of χ^2:

$$\frac{\partial\chi^2}{\partial q_k} = \sum_{j=1}^{N_g} 2\left(u_j - U_j\right)\frac{\partial u_j}{\partial q_k} + 2w_k\left(q_k - q_k^0\right). \tag{10}$$

Solution of Eq. (10) for all q_k, subject to the constraint $\sum_{i=1}^{N} q_i = Q$, is referred as *restrained* fitting.

Fitting without the additional restrain to the Amber charges is referred as *unrestrained* fitting. Figure 7 (*right*) shows the statistical contribution to the average using the *restrained* fitting. Here, a unique value for the weights has been used $(w_i = w = 0.15)$.. Although adding more restrains in the fitting of the electrostatic potential goes in detriment of reproducing the target MEP, the value of the weight used here still guarantees that the MEP of a full QM

Table 1 Statistical properties of the average $\left\langle e^{-\beta\left(E^*-E^0\right)}\right\rangle_0$ at $z = 6$ Å. $\sigma_{\Delta F}$ and $\sigma_{\Delta E}$ are the standard deviations of ΔF and ΔE, respectively

	$\Delta F = F^* - F^0$	$\sigma_{\Delta F}$	$\Delta E = E^* - E^0$	$\sigma_{\Delta E}$
Unrestrained fitting	-33.2	7.1	-29.0	3.2
Restrained fitting	-29.0	4.6	-19.8	2.1

All values are in kcal mol^{-1}

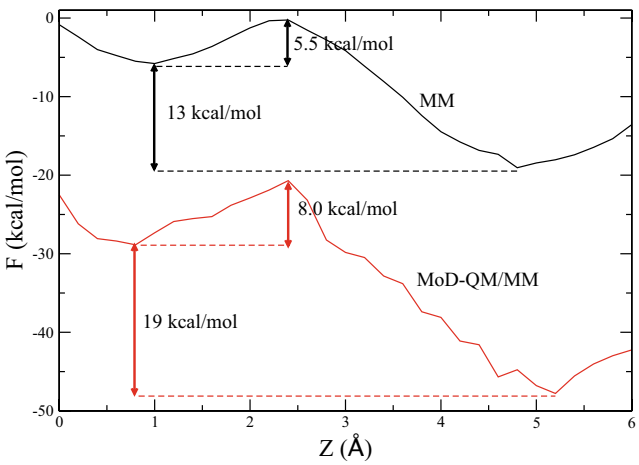

Fig. 8 Free energy profile for a potassium atom along the z-axis of the gramicidin model channel according to the Amber force field (top curve). The lower curve corresponds to the same force field in which electrostatic interactions are corrected via MoD-QM/MM

calculation is closely reproduced. As seen in Fig. 7 (*right*), there is a substantial improvement in the quality of the statistics, with about 16 configurations contributing to 60% of the average. It is important to emphasize that using poorly correlated potentials would result in a extremely poor statistic, having in general only one configuration contributing to the total average [60]. The statistical properties of the average in Eq. (8) for both fitting schemes are summarized in Table 1. As judged by the somewhat large standard deviations, it is clear that there is still room for improvement toward better statistics. A simple goal would be to reduce the standard deviation in the free energy difference (ΔF) to be at least as large as the standard deviation for the difference in energy (ΔE). Nevertheless, these results suggest that the calculation of a free energy profile with the improved potential given by MoD-QM/MM can be, in practice, carried out using a MM force field within a free energy perturbation (FEP) approach. Figure 8 shows the free energy profile of a potassium ion along the z-axis computed using the force field Amber via thermodynamic integration and the corrected free energy using Eq. (8). Although the free energy profile at the higher level of theory is more rugged than it is at the MM level, there is a clear distinction in the values of energy barriers and energy difference between bound states. Moreover, since all bonded and Van der Waals interactions are the same for both potentials, these differences originated exclusively from the treatment of electrostatic interactions.

Conclusions

We have discussed a number of improvements made to the linear scaling MoD-QM/MM computational protocol that can be used to calculate accurate molecular electrostatic potentials of biomacromolecules. Selection of moderately sized QM domains partitioned between C_α and C (Scheme 2), with incorporation of delocalization of electrons over neighboring domains (Scheme 3), results in a marked improvement of the calculated molecular electrostatic potential (MEP). Although MoD-QM/MM is more suitable for applications that do not require a dynamic representation of the protein, we have discussed a practical implementation to incorporate finite temperature effects. We showed that a standard force field can be used as a reference potential to perform dynamics, followed by evaluation of the free energy difference between the standard MM surface and the MoD-QM/MM surface. Such an approach could be of major relevance to the calculation of conductivity properties of ion channels, which currently do not provide a quantitative comparison with experimental data.

Acknowledgements J.A.G acknowledges financial support from the Camille and Henry Dreyfus New Faculty Award, start-up package funds from University of Connecticut, and supercomputer time from the National Energy Research Scientific Computing (NERSC) Center.

References

1. Warshel A, Sharma PK, Kato M, Parson WW (2006) Biochim Biophys Acta 1764:1647–1676
2. Warshel A (1981) Acc Chem Res 14:284–290
3. Sharp KA, Honig B (1990) Annu Rev Biophys Biophys Chem 19:301–332
4. Warshel A, Sharma PK, Kato M, Xiang Y, Liu HB, Olsson MHM (2006) Chem Rev 106:3210–3235
5. Warshel A (1998) J Biol Chem 273:27035–27038
6. Roca M, Marti S, Andres J, Moliner V, Tunon M, Bertran J, Williams AH (2003) J Am Chem Soc 125:7726–7737
7. Sheinerman FB, Norel R, Honig B (2000) Curr Opin Struct Biol 10:153–159
8. Norel R, Sheinerman F, Petrey D, Honig B (2001) Protein Sci 10:2147–2161
9. Lee LP, Tidor B (2001) Nat Struct Biol 8:73–76
10. Roux B, MacKinnon R (1999) Science 285:100–102
11. Allen TW, Andersen OS, Roux B (2004) Proc Natl Acad Sci USA 101:117–122
12. Bastug T, Kuyucak S (2007) J Chem Phys 126:105103–105104
13. Jorgensen WL, Maxwell DS, TiradoRives J (1996) J Am Chem Soc 118:11225–11236
14. Kaminski G, Duffy EM, Matsui T, Jorgensen WL (1994) J Phys Chem 98:13077–13082
15. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) J Phys Chem B 102:3586–3616
16. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) J Am Chem Soc 117:5179–5197
17. Halgren TA (1999) J Comput Chem 20:730–748
18. Scott WRP, Hunenberger PH, Tironi IG, Mark AE, Billeter SR, Fennen J, Torda AE, Huber T, Kruger P, van Gunsteren WF (1999) J Phys Chem A 103:3596–3607
19. Halgren TA, Damm W (2001) Curr Opin Struct Biol 11:236–242
20. Patel S, Brooks CL (2006) Mol Simul 32:231–249
21. Kaminski GA, Stern HA, Berne BJ, Friesner RA, Cao YX, Murphy RB, Zhou R, Halgren TA (2002) J Comput Chem 23:1515–1531
22. Maple JR, Cao Y, Damm W, Halgren TA, Kaminski GA, Zhang LY, Friesner RA (2005) J Chem Theor Comput 1:694–715
23. Piquemal JP, Williams-Hubbard B, Fey N, Deeth RJ, Gresh N, Giessner-Prettre C (2003) J Comput Chem 24:1963–1970
24. Stern HA, Kaminski GA, Banks JL, Zhou R, Berne BJ, Friesner RA (1999) J Phys Chem B 103:4730–4737
25. Yu H, Hansson T, Gunsteren WFv (2003) J Chem Phys 118:221–234
26. Ren PY, Ponder JW (2002) J Comput Chem 23:1497–1506
27. Ren P, Ponder JW (2003) J Phys Chem B 107:5933–5947
28. Stewart JJP (1996) Int J Quantum Chem 58:133–146
29. York DM, Lee TS, Yang WT (2005) J Am Chem Soc 127:7215–7226
30. Gogonea V, Suarez D, van der Vaart A, Merz KW (2001) Curr Opin Struct Biol 11:217–223
31. Van der Vaart A, Gogonea V, Dixon SL, Merz KM (2000) J Comput Chem 21:1494–1504

32. Elstner M, Frauenheim T, Suhai S (2003) J Mol Struct THEOCHEM 632:29–41
33. Mohle K, Hofmann HJ, Thiel W (2001) J Comput Chem 22:509–520
34. Titmuss SJ, Cummins PL, Bliznyuk AA, Rendell AP, Gready JE (2000) Chem Phys Lett 320:169–176
35. Zuegg J, Bliznyuk AA, Gready JE (2003) Mol Phys 101:2437–2450
36. Gascon JA, Leung SSF, Batista ER, Batista VS (2006) J Chem Theor Comput 2:175–186
37. Maseras M, Morokuma K (1995) J Comput Chem 16:1170–1179
38. Svensson M, Humbel S, Froese RDJ, Matsubara T, Sieber S, Morokuma K (1996) J Phys Chem 100:19357–19363
39. Vreven T, Morokuma K (2000) J Comput Chem 16:1419–1432
40. Gao JL (1997) J Phys Chem B 101:657–663
41. Gao JL (1998) J Chem Phys 109:2346–2354
42. Wierzchowski SJ, Kofke DA, Gao JL (2003) J Chem Phys 119:7365–7371
43. Bakowies D, Thiel W (1996) J Phys Chem 100:10580–10594
44. Ketchem RR, Roux B, Cross TA (1997) Structure 5:1655–1669
45. Townsley LE, Tucker WA, Sham S, Hinton JF (2001) Biochemistry 40:11676–11686
46. Bastug T, Kuyucak S (2006) Chem Phys Lett 424:82–85
47. Politzer P, Murray JS, Lane P (2007) Int J Quantum Chem 107:3046–3052
48. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgonery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2003) Gaussian 03. Gaussian Inc., Wallingford CT
49. Lin H, Truhlar D (2007) Theor Chem Acc 117:185–199
50. Singh UC, Kollmann PA (1986) J Comput Chem 7:718–729
51. Antes I, Thiel W (1998) On the treatment of link atoms in hybrid methods. In: Gao J, Thompson MA (Eds.) Hybrid Quantum Mechanical and Molecular Mechanical Methods Proc. ACS Symp Ser, Washington DC, 1998, Am Chem Soc 712:50–65
52. Sinclair PE, de Vries A, Sherwood P, Catlow CRA, van Santen RA (1998) J Chem Soc Faraday Trans 94:3401–3408
53. Das D, Eurenius KP, Billings EM, Sherwood P, Chatfield DC, Hodoscek M, Brooks BR (2002) J Chem Phys 117:10534–10547
54. Amara P, Field MJ (2003) Theor Chem Acc 109:43–52
55. Lin H, Truhlar DG (2005) J Phys Chem A 109:3991–4004
56. Ferre N, Olivucci M (2003) J Mol Struct Theochem 632:71–82
57. Jorgensen WL (1989) Acc Chem Res 22:184–189
58. Bentzien J, Muller RP, Florian J, Warshel A (1998) J Phys Chem B 102:2293–2301
59. Zhang YK, Liu HY, Yang WT (2000) J Chem Phys 112:3483–3492
60. Rod TH, Ryde U (2005) J Chem Theor Comput 1:1240–1251
61. Rod TH, Ryde U (2005) Phys Rev Lett 94:138302–138305
62. Rosta E, Klahn M, Warshel A (2006) J Phys Chem B 110:2934–2941
63. Strajbl M, Florian J, Warshel A (2000) J Am Chem Soc 122:5354–5366