

Communication: Quantum polarized fluctuating charge model: A practical method to include ligand polarizability in biomolecular simulations

S. Roy Kimura, Ramkumar Rajamani, and David R. Langley

Citation: *J. Chem. Phys.* **135**, 231101 (2011); doi: 10.1063/1.3671638

View online: <http://dx.doi.org/10.1063/1.3671638>

View Table of Contents: <http://jcp.aip.org/resource/1/JCPSA6/v135/i23>

Published by the [American Institute of Physics](#).

Additional information on *J. Chem. Phys.*

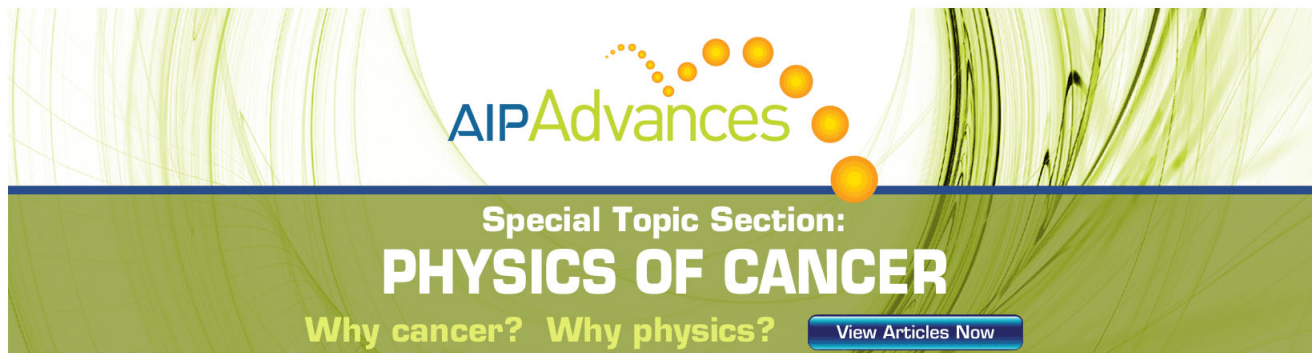
Journal Homepage: <http://jcp.aip.org/>

Journal Information: http://jcp.aip.org/about/about_the_journal

Top downloads: http://jcp.aip.org/features/most_downloaded

Information for Authors: <http://jcp.aip.org/authors>

ADVERTISEMENT



AIP Advances

Special Topic Section:
PHYSICS OF CANCER

Why cancer? Why physics? [View Articles Now](#)

Communication: Quantum polarized fluctuating charge model: A practical method to include ligand polarizability in biomolecular simulations

S. Roy Kimura, Ramkumar Rajamani, and David R. Langley^{a)}

Department of Computer-Assisted Drug Design, Bristol-Myers Squibb R & D, 5 Research Parkway, Wallingford, Connecticut 06492, USA

(Received 20 October 2011; accepted 2 December 2011; published online 21 December 2011)

We present a simple and practical method to include ligand electronic polarization in molecular dynamics (MD) simulation of biomolecular systems. The method involves periodically spawning quantum mechanical (QM) electrostatic potential (ESP) calculations on an extra set of computer processors using molecular coordinate snapshots from a running parallel MD simulation. The QM ESPs are evaluated for the small-molecule ligand in the presence of the electric field induced by the protein, solvent, and ion charges within the MD snapshot. Partial charges on ligand atom centers are fit through the multi-conformer restrained electrostatic potential (RESP) fit method on several successive ESPs. The RESP method was selected since it produces charges consistent with the AMBER/GAFF force-field used in the simulations. The updated charges are introduced back into the running simulation when the next snapshot is saved. The result is a simulation whose ligand partial charges continuously respond in real-time to the short-term mean electrostatic field of the evolving environment without incurring additional wall-clock time. We show that (1) by incorporating the cost of polarization back into the potential energy of the MD simulation, the algorithm conserves energy when run in the microcanonical ensemble and (2) the mean solvation free energies for 15 neutral amino acid side chains calculated with the quantum polarized fluctuating charge method and thermodynamic integration agree better with experiment relative to the Amber fixed charge force-field. © 2011 American Institute of Physics. [doi:10.1063/1.3671638]

Molecular mechanics (MM) potentials are widely used in drug discovery research to model binding of small molecule ligands to proteins or DNA.^{1,2} A recent example is the ground-breaking long timescale simulations published by Dror *et al.*³ The potentials used in these studies partly rely on fixed partial charges assigned to atom centers to describe non-bonded electrostatic interactions through Coulomb's law. Partial charges are usually fit to reproduce the electrostatic potential field around the molecule of interest determined via quantum mechanical (QM) calculations in vacuum or constant dielectrics. Because molecules are polarized by their conformation and external electric fields, macromolecular force-fields^{1,4-6} are fit to the mean field expected around multiple low energy conformers of each amino acid or nucleic acid residue building block. In a simple protein or DNA calculation, such a description provides reasonable mean accuracies as the residues adopt different conformations and face heterogeneous environments within a single system. In contrast, higher accuracy is often required to make ligand-receptor binding calculations useful in drug research, which focuses on specific atomic interactions in and around the binding pocket. Indeed many studies^{3,7} report mean errors in calculated binding energies on the order of 1 to 2 kcal/mol using state-of-the-art fixed charge force-fields. Although much of this error likely stems from inadequate configurational sampling, the neglect of polar-

ization is also potentially a significant source of systematic error.⁸

Various strategies to dynamically incorporate electronic polarizability into MM calculations have been previously developed⁹ including fluctuating charges,^{10,11} inducible dipoles,^{10,12,13} and classical Drude oscillators.¹⁴ While parameter fitting for biomolecular systems, with well defined residues, is a tractable problem, the exercise can become extremely labor intensive if not prohibitive for adequate coverage of small molecule chemical space. Another approach is the hybrid QM/MM method,^{15,16} where the ligand and optionally nearby protein atoms are described with QM, while the rest of the system is treated with MM. In principle, this approach is accurate if the QM calculations are performed at an adequate level of theory and is consistent with the MM force-field used (e.g., HF/6-31G*). In practice, simulation timescales needed for protein-ligand modeling (>10 ns) can only be achieved today using semi-empirical methods or possibly using HF/3-21G* or density functional theory with a small basis set. A third approach is to explicitly model the expected polarization that exists in a particular structure by pre-fitting partial charges through a self-consistent iteration of QM calculations on each residue and ligand.^{17,18} Such an approach would yield a specialized charge set useful for in-depth studies of particular systems. Finally, implementations of linear scaling methods for semi-empirical QM treatments of macromolecules are beginning to enable longer timescale fully polarized molecular dynamics (MD) simulations. They currently represent the closest approximation to an

^{a)}Electronic mail: david.langley@bms.com.

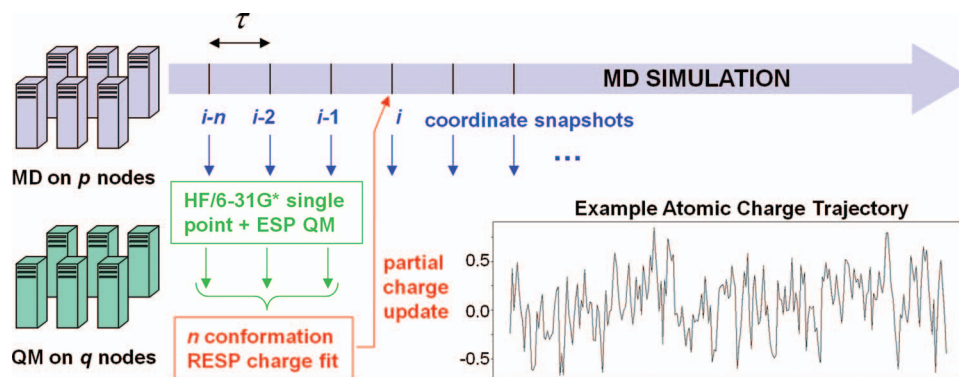


FIG. 1. Summary of the quantum polarized fluctuating charge (QPFC) methodology. In the current implementation, the QM charge fitting is done on the ligand at each interval, τ .

ab initio QM treatment of the entire simulated system (macromolecule + ligand + water) but are still limited to the 100 ps timescale.^{19,20}

Although higher levels of theory such as resolution-of-identity second-order Møller-Plesset perturbation theory (RI-MP2) with Dunning's aug-cc-pvtz basis set²¹ have seen widespread adoption for *ab initio* QM calculations, most force-fields today rely on semi-empirical or HF/6-31G* (Refs. 22 and 23) basis sets for partial charge calculations. It is also well known that vacuum HF/6-31G* calculations result in over-polarization of the ligand which fortuitously makes the fitted partial charges more consistent with use in the aqueous environment.⁴ Fitting charges in the presence of protein and solvent external fields may result in additional over-polarization since the water and the protein are not allowed to instantaneously polarize themselves, although allowing the atomic nuclei to move during MD should account for part of this effect.

Notwithstanding these issues, we have implemented a simple hybrid scheme that enables dynamic treatment of ligand electronic polarizability (Figure 1). A traditional MD simulation is run on p processors, while coordinate snapshots are periodically saved at a simulation time interval, τ . As each snapshot is saved, a QM HF/6-31G* (or other appropriate basis set) single-point electrostatic potential (ESP) grid calculation is submitted to an extra set of q processors. To prevent processor idling, τ , p , and q are tuned so that the QM calculation on the ligand is completed within the wall-clock time necessary for the MD to proceed by τ (typically 10–20 ps of simulation time depending on the ligand size). The polarizing effects of the protein and water are included in the QM calculations as fixed external charges from the MD snapshot. When n ESP grids are available, a multi-conformer restrained electrostatic potential (RESP) fit^{22,23} calculation is launched and ligand partial charges are refit. The charges are fed back into the simulation when the next snapshot is saved and the cycle is repeated. Although there is a small time delay in the polarization response equal to τ , this error is not expected to be significant as our goal is to reproduce the mean polarization of the ligand in the ~ 100 ps timescale.

Validation and preliminary calibration of the methodology was performed through calculation of solvation free energies of 15 neutral amino acid side chains (Table I, Figure 2).

For each side chain, a quantum polarized fluctuating charge (QPFC) simulation was first performed in an approximately $22 \times 22 \times 22$ Å TIP3P water box (~ 350 water molecules). The mean charges determined from the equilibrated 500 ps portions of the QPFC simulations were used for subsequent fixed charge thermodynamic integration (TI) calculations. TI was performed by increasing a charge scaling factor, λ , from 0 to 1, in five equally spaced increments and using a simple rectangular integration scheme. Details of the simulations are summarized in Table II. Non-polar contributions were taken from Shirts *et al.*²⁴ For comparison, calculations were repeated using generalized amber force-field (GAFF) AM1-

TABLE I. Amino acid solvation-free energy calculations versus experiment.

Side chain	Nonpolar ^a	AMBER ^b	GAFF ^c	QPFC full ^d	QPFC 0.9 ^e	Expt. ^f
Ala	2.57	2.57	2.57	2.55	2.56	1.94
Val	2.71	2.69	2.70	2.70	2.71	1.99
Leu	2.78	2.73	2.78	2.73	2.74	2.28
Ile	2.85	2.85	2.85	2.84	2.84	2.15
Ser	1.76	-4.37	-3.30	-7.83	-4.89	-5.06
Thr	1.94	-3.83	-3.21	-7.38	-4.53	-4.88
Phe	2.45	0.10	-0.26	-1.22	-0.24	-0.76
Tyr	1.96	-4.23	-4.74	-7.81	-4.96	-6.11
Trp	1.52	-4.87	-5.57	-7.37	-5.01	-5.88
Cys	2.15	0.11	-0.08	-1.35	-0.24	-1.24
Met	2.52	0.91	0.58	-0.95	0.10	-1.48
Asn	1.47	-7.80	-8.15	-18.29	-12.28	-9.68
Gln	1.66	-7.69	-8.41	-17.31	-11.64	-9.38
His	1.08	-8.43	-5.57	-16.48	-11.24	-10.27 ^g
Hie	1.09	-8.98	-6.88	-14.41	-9.63	-10.27 ^g
RMS Error ^h		1.35	1.63	3.78	1.18	

^aNonpolar solvation free energies calculated via thermodynamic integration taken from Shirts *et al.* (Ref. 24).

^bTotal solvation free energies calculated using AMBER charges; values taken from Shirts *et al.* (Ref. 24).

^cTotal solvation free energies calculated using GAFF charges, this study, plus nonpolar energies from Shirts *et al.* (Ref. 24).

^dTotal solvation free energies calculated from QPFC simulations with full polarization, this study.

^eTotal solvation free energies calculated from QPFC simulation with polarization reduced by factor of 0.9, this study.

^fExperimental solvation free energies taken from Wolfenden *et al.* (Ref. 27).

^gExperimental values for imidazole shown in both.

^hRoot mean square error between calculated and experimental values.

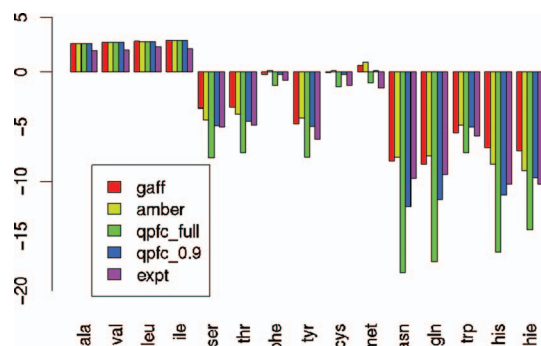


FIG. 2. Calculated and experimental solvation free energies (kcal/mol). qpfc_full refers to the results when the un-scaled full RESP charges were used in the FEP calculations. qpfc_0.9 refers to the case when those charges were scaled by a factor of 0.9 during the QPFC simulations, and the resulting mean charges were used in the FEP calculations. Experimental values taken from Ref. 27.

BCC charges. The free energy calculations using assisted model building with energy refinement (AMBER) force-field charges from Ref. 24 are also reproduced in the table. All bonded and Lennard-Jones parameters were taken from the AMBER force-field. The NAMD (Ref. 25) and Q-Chem (Ref. 26) packages were used for the MD and QM calculations, respectively.

We initially found that full QPFC polarization overestimated the solvation free energies (rms error 3.8 kcal/mol). However, simply scaling the RESP charges by a factor of 0.9 during the QPFC simulations reduced the over-polarization and resulted in better reproduction of experiment (rms error 1.2 kcal/mol). The inclusion of polarization may account for the slight improvement in accuracy compared to fixed charge force-fields (AMBER, GAFF). However, even with charge-scaling, the amide groups in Asn and Gln side chains are still over-polarized. Closer inspection revealed increased polarization in the nitrogen and the carbonyl carbon atoms of the amide groups, in agreement with the explicit polarization (X-Pol) QM calculations performed in Ref. 20. The lack of polarization of Cys and Met side chains are likely due to the inherent problem of representing sulfur polarizability using a single point charge at the center of the atom.

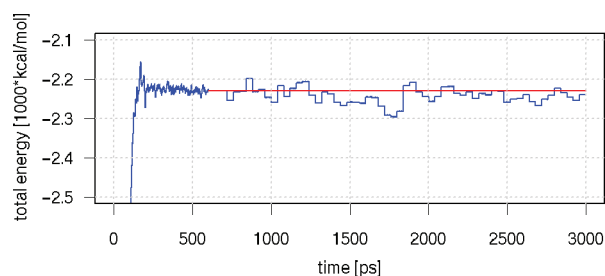


FIG. 3. Total energy trajectory showing heating and equilibration phases in NVT ensembles, and QPFC simulation in NVE ensemble starting from 600 ps into the simulation (blue line). The discontinuities in the energy (blue line) during the QPFC NVE phase correspond to the replacement of ligand charges with newly calculated ones. The red line shows the result of adding back the change in electrostatic energy ($\Delta E_{\text{Coulombic}}$) to the total QPFC NVE energy each time the charges are updated. Note that the total energy is now conserved (red line).

Updating the charges on the ligand introduces a new potential energy change exactly equal to the change in the Coulombic potential resulting from changes in the ligand-self and ligand-system interaction energies. We verified that this change in potential energy is exactly balanced by changes in the kinetic energy and the QPFC method conserves energy when run in the constant-NVE ensemble (Figure 3).

Although the benefits of the QPFC method can be immediately realized by applying it to study any protein-ligand system of interest, a few enhancements may further improve the applicability and accuracy of the method. First, a systematic optimization of the charge re-scaling factor (0.9) may improve the accuracy of the polarization contribution. Second, QPFC could be combined with polarizable force-fields to model polarization of the full system. This will take advantage of the parameter optimizations that have been performed so far in the field, while simultaneously and accurately accounting for small-molecule polarization effects. Third, the QPFC region could be expanded to also include the biomolecular residues neighboring the ligand. This will better describe the polarization between the ligand and binding site residues. Finally, it may be possible to extend QPFC to enable dynamic polarization of all the residues in the system by applying a self-consistent iterative scheme similar to the one devised by Zhang and Ji¹⁷ for static systems.

TABLE II. Simulation details.

Number of water molecules	~300–400 depending on amino acid side chain
Force-field	Amber99
Initial box dimensions	~22 × 22 × 22 Å
Integration scheme	Velocity verlet (no multi-timescale updates)
Simulation ensemble	NPT, 300K, 1 atm
Start-up protocol	Minimization: 250 steps conjugate gradient; Heating: NVT ensemble, 10 K increments every 100 steps up to 300K; Equilibration: 50 ps, NPT ensemble
Electrostatics treatment	10–8 Å taper w/ PME
Boundary conditions	Periodic cubic box
QM charging frequency	Every 10 ps
QM averaging window	Three snapshots (30 ps)
Thermodynamic integration scheme	Five steps, 500 ps each stage
Average wall-clock performance	50 min. / ns on 16 (MD) + 4 (QM) CPUs

Overall, we find that the QPFC provides a practical method that complements the strengths and weaknesses of QM/MM and polarizable force-fields for MD simulation of ligand-bound biomolecular systems. We show that (1) by incorporating the cost of polarization back into the potential energy of the MD simulation, the algorithm conserves energy when run in the microcanonical ensemble and (2) the mean solvation free energies for 15 neutral amino acid side chains calculated with the QPFC method and thermodynamic integration agree better with experiment relative to the Amber fixed charge force-field.

- ¹D. R. Langley, *J. Biomol. Struct. Dyn.* **16**(3), 487 (1998).
- ²S. R. Kimura, A. J. Tebben, and D. R. Langley, *Proteins: Struct., Funct. Genet.* **71**(4), 1919 (2008).
- ³R. O. Dror, A. C. Pan, D. H. Arlow, D. W. Borhani, P. Maragakis, Y. Shan, H. Xu, and D. E. Shaw, *Proc. Natl. Acad. Sci. U.S.A.* **108**(32), 13118 (2011).
- ⁴W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, *J. Am. Chem. Soc.* **117**(19), 5179 (1995).
- ⁵A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, *J. Phys. Chem. B* **102**(18), 3586 (1998).
- ⁶J. Tirado-Rives and W. L. Jorgensen, *J. Am. Chem. Soc.* **112**(7), 2773 (1990).
- ⁷C. R. W. Guimarães and M. Cardozo, *J. Chem. Inf. Model.* **48**(5), 958 (2008).
- ⁸Y. Tong, Y. Mei, Y. L. Li, C. G. Ji, and J. Z. Zhang, *J. Am. Chem. Soc.* **132**(14), 5137 (2010).
- ⁹P. E. M. Lopes, B. Roux, and A. D. MacKerell, Jr., *Theor. Chem. Acc.* **124**(1–2), 11 (2009).
- ¹⁰S. Patel and C. L. Brooks III, *J. Comput. Chem.* **25**(1), 1 (2004).
- ¹¹S. W. Rick, S. J. Stuart, and B. J. Berne, *J. Chem. Phys.* **101**(7), 6141 (1994).
- ¹²Y. P. Liu, K. Kim, B. J. Berne, R. A. Friesner, and S. W. Rick, *J. Chem. Phys.* **108**(12), 4739 (1998).
- ¹³B. Ma, J. H. Lii, and N. L. Allinger, *J. Comput. Chem.* **21**(10), 813 (2000).
- ¹⁴G. Lamoureux and B. Roux, *J. Chem. Phys.* **119**(6), 3025 (2003).
- ¹⁵J. Gao and X. Xia, *Science* **258**(5082), 631 (1992).
- ¹⁶A. Warshel and M. Levitt, *J. Mol. Biol.* **103**(2), 227 (1976).
- ¹⁷C. G. Ji and J. Z. Zhang, *J. Phys. Chem. B* **113**(49), 16059 (2009).
- ¹⁸Y. Xiang, L. Duan, and J. Z. Zhang, *J. Chem. Phys.* **134**(20), 205101 (2011).
- ¹⁹A. van der Vaart, V. Gogonea, S. L. Dixon, and K. M. Merz, Jr., *J. Comput. Chem.* **21**(16), 1494 (2000).
- ²⁰W. Xie, M. Orozco, D. G. Truhlar, and J. Gao, *J. Chem. Theor. Comput.* **5**(3), 459 (2009).
- ²¹T. H. Dunning, Jr. and P. J. Hay, in *Modern Theoretical Chemistry*, edited by H. F. Schaefer III (Plenum, New York, 1976), pp. 1–28.
- ²²C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman, *J. Phys. Chem.* **97**(40), 10269 (1993).
- ²³W. D. Cornell, P. Cieplak, C. I. Bayly, and P. A. Kollman, *J. Am. Chem. Soc.* **115**(21), 9620 (1993).
- ²⁴M. R. Shirts, J. W. Pitner, W. C. Swope, and V. S. Pande, *J. Chem. Phys.* **119**(11), 5740 (2003).
- ²⁵M. T. Nelson, W. Humphrey, A. Gursoy, A. Dalke, L. V. Kale, R. D. Skeel, and K. Schulten, *Int. J. High Perform. Comput. Appl.* **10**(4), 251 (1996).
- ²⁶Y. Shao, L. F. Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, R. A. DiStasio, Jr., R. C. Lochan, T. Wang, G. J. O. Beran, N. A. Besley, J. M. Herbert, C. Yeh Lin, T. Van Voorhis, S. Hung Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C. P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. Min Rhee, J. Ritchie, E. Rosta, C. David Sherrill, A. C. Simmonett, J. E. Subotnik, H. Lee Woodcock III, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer III, J. Kong, A. I. Krylov, P. M. W. Gill, and M. Head-Gordon, *Phys. Chem. Chem. Phys.* **8**(27), 3172 (2006).
- ²⁷R. Wolfenden, L. Andersson, P. M. Cullis, and C. C. B. Southgate, *Biochemistry* **20**(4), 849 (1981).