

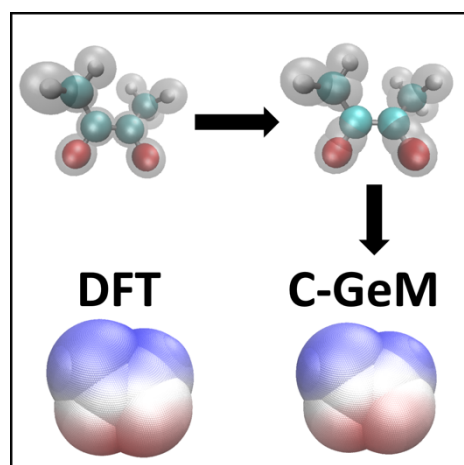
C-GeM: Coarse-Grained Electron Model for Predicting the Electrostatic Potential in Molecules

Itai Leven^{1,2,5} and Teresa Head-Gordon^{1,2,3,4,5}

¹*Kenneth S. Pitzer Center for Theoretical Chemistry*, ²*Department of Chemistry*,
³*Department of Bioengineering*, ⁴*Department of Chemical and Biomolecular Engineering*
University of California Berkeley
⁵*Chemical Sciences Division, Lawrence Berkeley National Laboratory*
Berkeley, California 94720, USA

We have developed a new coarse-grained electron model, C-GeM, in which atoms are represented by a positive core and an electron shell described by Gaussian charge distributions, with the interaction energy between the core and shell reflecting the electronegativity of a given atomic element. By minimizing the electronic shell positions in the field of atomic core positions, the model can provide accurate electrostatic properties of molecules and their interactions. We **have tested** the performance of the C-GeM model for a set of molecules containing H, C, O and Cl atoms to show that it can predict the electrostatic potential with high accuracy, and correctly describe the dissociation of HCl into ionic fragments in solution and to neutral atoms in the gas phase. The resulting C-GeM approach offers many advantages over expensive *ab initio* methods and reactive force field charge equilibration methodologies: it can rapidly predict the electrostatic potential surfaces of molecules, molecules dissociate into integer charge fragments so that redox reactions are easily described, there is no unphysical long-range charge transfer, it can account for out-of-plane polarization, and charges are not required to be centered on atoms, thereby accounting for electrostatic features such as sigma holes.

†Corresponding author: thg@berkeley.edu



Prediction of electrostatic properties for molecules is of vital importance in numerous research disciplines. In biochemistry, the electrostatic potential is a dominant factor determining the preference for functional states in biomolecules such as ligand-binding or protein-protein interactions.¹⁻⁴ In material science, the function of nanoporous crystals such as zeolites and metal-organic frameworks for gas storage and separation applications rely on their ability to absorb polar molecules.⁵⁻⁷ In electrochemistry, the function of electrochemical cells relies on the diffusion of ions and the double-layer formation at the electrode surface.⁸⁻¹¹ Computational modeling of these systems thereby requires an accurate description of the electrostatic interaction between the different components of these complex system. While first principles methodologies are desirable due to their high accuracy, they are limited to small systems and require a great amount of computational resources, making them unsuitable for many applications. Classical models which rely on electrostatic potential fitted charges (EPFC)¹²⁻¹⁴ or empirically derived charges (EDC)¹⁵⁻¹⁸ are therefore utilized.

For EPFC models, the partial charges on atoms are derived by either direct fitting the charges to minimize the error with respect to an *ab initio* electrostatic potential or using partial charges derived from a partitioning scheme of the electronic wave functions such as Mulliken,¹⁹ Hirshfeld-I^{20,21}, DDEC/c3²² or natural population analysis²³. Such generated partial charges are the primary electrostatic description used in all the major fixed charge force fields such as AMBER²⁴ and CHARMM²⁵, or those utilized in Poisson Boltzmann equation solvers to solve for the electrostatic potentials of large molecules.²⁶⁻²⁸ To include many-body effects that is crucial for heterogeneous environments, advanced force fields require parameterization of polarization and/or charge transfer along with the permanent multipolar electrostatics.²⁹ In all cases these approaches require an expensive first principles calculation of the electrostatic potential surface and then parameter tuning for each new molecular system, making them difficult or unsuitable for high-throughput screening applications. Furthermore, the EPFC models are largely restricted to applications for which the electron density changes are relatively small, and thus are precluded from describing any reactive chemistry.

By contrast, the EDC models offer the major advantage of being general, in the sense that they only use atomic parameters of the electronegativity and atomic hardness, to predict the partial charges on atoms in any molecular environment based on the electronegativity equalization family of methods (EEM).³⁰ Therefore, EDC models and EEM methods are often the preferred choice for

fast electrostatic screening applications of large molecular databases^{17,31,32}, and as electrostatic and charge transfer models for reactive force fields.^{33–35} This however comes at the expense of accuracy, since EDC models are known for their deficiency in obtaining non-integer charge transfer at long molecular separation, unphysical metallic polarization^{36,37}, and not being able to describe out of plane polarization relevant to aromatic systems. There has been recent progress to address these deficiencies, such as the split charge equilibration^{38,39} (SQE) and the atom condensed Kohn-Sham DFT approximated to second order (ACKS2)¹⁶ models that impede unphysical long range charge transfer; however, these methods still require assigning a reference charge state to the atom that prohibits their use in redox reactions and electrochemistry where a dynamic reference state is required. A more recent version of the ReaxFF model, eReaxFF, includes explicit electrons in addition to replacing the EEM model with the more recent ACKS2 model⁴⁰ so that it can now describe integer charge transfer between molecular fragments. Another method developed to account for explicit charge transfer within the framework of the split charge model is redoxSQE, such that the oxidation state of the atom can change according to an electron transfer probability.⁴¹ Other force fields that incorporate explicit electrons such as LEWIS⁴² and eFF⁴³ have been developed to represent the full atomic interaction potential surface.

In this paper we present a novel approach for predicting the electrostatic properties of molecules through a coarse-grained electron model (C-GeM) that has the same generality as EDC models solved with EEM methods. In C-GeM the atoms are divided into a positive core and a negative shell, where the interaction energy between the core and shell recapitulates the electronegativity of a given atomic element. The basic principle in C-GeM is to minimize the positions of the electronic shells while keeping the positions of the nuclei fixed, in much the same way as the electronic wave function is minimized in *ab initio* methods to find the ground state electronic configuration under the Born Oppenheimer approximation. We provide a proof of principle of the C-GeM approach on a set of molecules containing H, C, O and Cl atoms. We first train our chemical atoms by fitting to the electrostatic potential surfaces (EPS) obtained from density functional theory (DFT) for various molecules containing these atomic elements, and then validate the predictive capabilities of C-GeM on a completely different set of molecules with the same chemical atoms but different bonding and hybridization features compared to the training set.

We show that the EPS obtained with our model is in very good agreement with the *ab initio* EPS and is found to systematically outperform the EPS results of EEM using a reactive force field

developed for aqueous chloride reactions⁴⁴. Furthermore we show the applicability of our model for redox reactions by correctly predicting the different dissociation outcomes of hydrogen chloride in water and gas phase environments. C-GeM is simple, fast, intuitive and can be utilized in a broad range of applications from fast virtual electrostatic screening of drug molecules to alternate formulations of polarization and charge transfer in force fields used in molecular dynamics.

THEORY

In C-GeM the atom is divided into a positively charged core and a negatively charged shell, which are represented as smeared Gaussian charges as used in various schemes such as PQeq⁴⁵ or in advanced force field models^{46–51}. We note that our model treats atoms as isolated chemical species which differs from force field atom typing that depends on hybridization or bonding. The advantage of the Gaussian functional form is well-appreciated since the long-range interaction between the two charge distributions behaves like $1/r_{ij}$, while at short-range the interaction converges to a finite value eliminating the singularities associated with point charges^{45,47,50,51}. More specifically, the Gaussian charge distribution of the core of atom i , ρ_c^i , and a generic electron shell ρ_s is given by:

$$\rho_c^i(\vec{r}) = q_c \cdot \left(\frac{\alpha_c^i}{\pi}\right)^{\frac{3}{2}} e^{-\alpha_c^i(|\vec{r}-\vec{r}_c|^2)} \quad (1a)$$

$$\rho_s(\vec{r}) = q_s \cdot \left(\frac{\alpha_s}{\pi}\right)^{\frac{3}{2}} e^{-\alpha_s(|\vec{r}-\vec{r}_s|^2)} \quad (1b)$$

where $\vec{r}$ is a position vector for the central position of the core ($\vec{r}_c$) and shell ($\vec{r}_s$), respectively, and in this work we set the core and shell charges, q_c and q_s , to 1 and -1 respectively. We note in principle that we can introduce multiple shells with additional charges. The spread of the Gaussian distributions is controlled by α_c^i for core atom type i and α_s is the width of the gaussian distribution of the electronic shell

$$\alpha_c^i = \frac{\lambda}{2 * R_i^2} \quad (2a)$$

$$\alpha_s = \frac{\lambda}{2 * R_s^2} \quad (2b)$$

where R_i is the atomic covalent radius of atom type i , R_s is the effective radius of the shell, and λ is a global fitting parameter (Figure 1a). Integration of the Coulombic interactions of the Gaussian

densities yields the analytical form for the electrostatic energy between arbitrary core-core, core-shell, and shell-shell interactions:

$$E_{ij}^{elec}(r_{ij}) = \frac{q_i \cdot q_j}{r_{ij}} \operatorname{erf} \left(\sqrt{\frac{\alpha_i \cdot \alpha_j}{\alpha_i + \alpha_j}} r_{ij} \right) \quad (3a)$$

where r_{ij} is the distance such that in the limit $r_{ij} \rightarrow 0$:

$$\lim_{r_{ij} \rightarrow 0} E_{ij}^{elec}(r_{ij}) = \frac{2q_i \cdot q_j}{\pi} \left(\sqrt{\frac{\alpha_i \cdot \alpha_j}{\alpha_i + \alpha_j}} \right) \quad (3b)$$

In addition to the electrostatic interaction between two Gaussian densities we add an additional Gaussian term between all possible i and j core and shell interactions given by:

$$E_{ij}^{Gauss}(r_{ij}) = A_i e^{-\gamma_i r_{ij}^2} \quad (4)$$

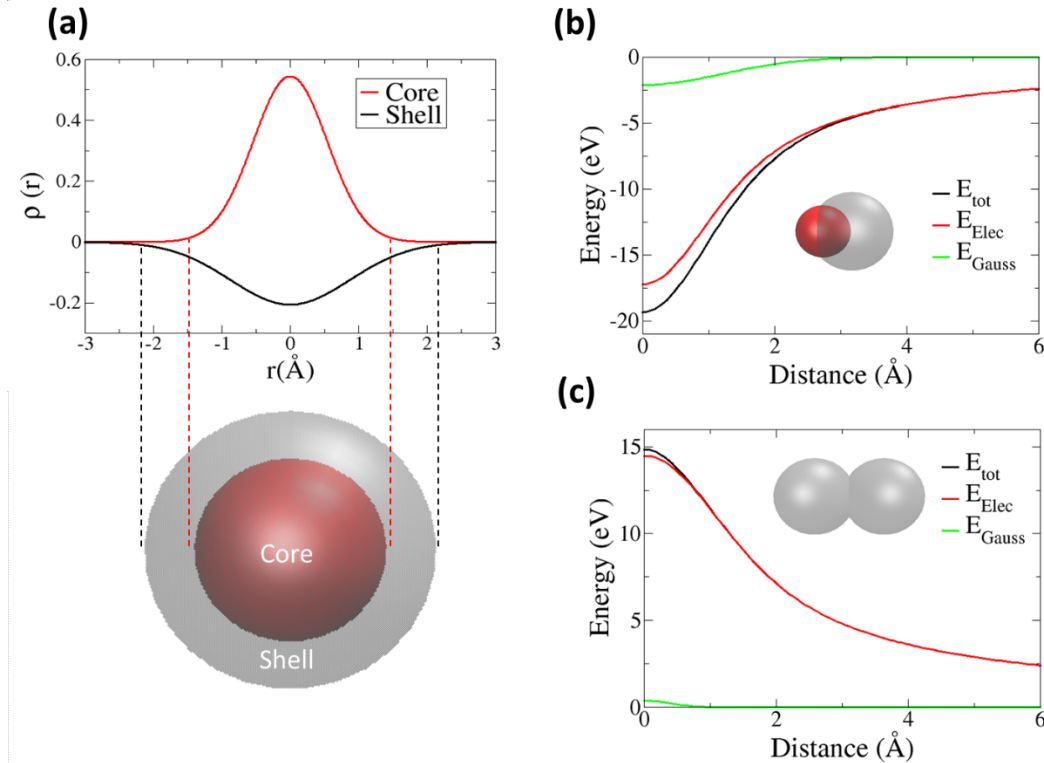


Figure 1. Core-shell representation and their interaction energies in C-GeM. (a) Gaussian charge distribution of a positive core oxygen (red) and a negative shell (black). (b) The distance dependence of core and shell electrostatic (red), gaussian (green) and total (black) interactions energies. (c) Interaction energy between two shells for electrostatic (red), gaussian (green) and total (black).

where A_i is a parameter reflecting the magnitude of the interaction energy, and γ_i that controls the radial range of the interaction given by

$$\gamma_i = \frac{\omega_i}{2 * R_i} \quad (5)$$

for which ω_i is a global parameter corresponding to atomic cores or shells. For a core(i)-shell(j) interaction we define A_i to meet the criteria:

$$\chi_i = E_{ij}^{Gauss}(r_{ij} = 0) + E_{ij}^{elec}(r_{ij} = 0) \quad (6)$$

where χ_i is the electronegativity of the core atom type i , and which leads to the following equation:

$$A_i = \frac{\chi_i - E_{ij}^{elec}(r_{ij} = 0)}{e^{-\gamma_i r_{ij}^2}} = \frac{\chi_i - \frac{2q_i \cdot q_j}{\pi} \left(\sqrt{\frac{\alpha_i \cdot \alpha_j}{\alpha_i + \alpha_j}} \right)}{e^{-\gamma_i r_{ij}^2}} \quad (7)$$

Designed such that the total interaction energy between a core and a shell mimics the electronegativity of a given core atom. In the case of shell-shell interaction χ_i is replaced by an effective shell-shell interaction energy χ_{shell} , requiring a fitting parameter obtained by reducing the error with respect to the *ab initio* EPS over a finite training set of molecules. For core-core interactions we only consider electrostatic interaction and set $A_i = 0$. Hence the total interaction energy of a system of n particles (cores +shells) is then given by:

$$E_{CGeM} = \sum_i^n \sum_{j<i}^n \left(E_{ij}^{elec}(r_{ij}) + E_{ij}^{Gauss}(r_{ij}) \right) \quad (8)$$

and shown in Figure 1b for core-shell and Figure 1c for shell-shell interactions. The basic principle of C-GeM is to minimize the C-GEM energy (Eq. 8) with respect to the position of the shells while keeping the core positions fixed. The corresponding optimized shell configuration is used to generate an electrostatic potential (ESP) utilizing Eq. (3a), and schematically shown in Figure 2.

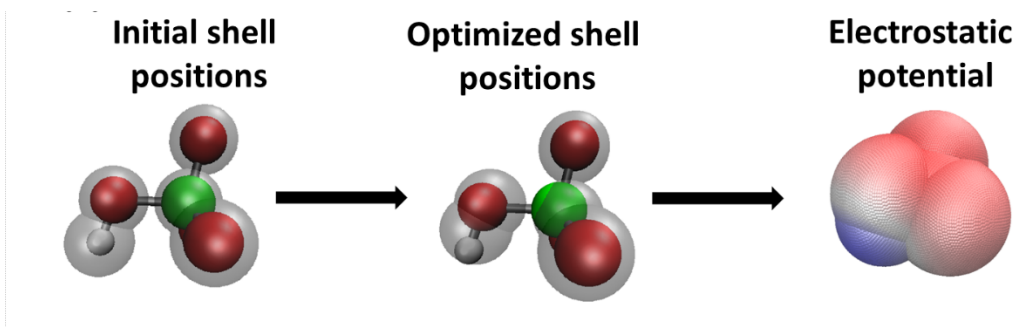


Figure 2. Schematic representation of how C-GeM is used to generate the electrostatic potential. The positions of the shells are geometrically optimized with respect to fixed atomic cores and which are then used to generate the ESP of the corresponding molecule.

C-GeM is designed to be integrated with MD force fields in such a way that in every MD time step a minimization of shell positions is performed. Since the shells are allowed to dissociate from the atomic core, explicit charge transfer between different molecular segments is easily accounted for in the final optimized state.

METHODS

To find the optimal shell configuration of a system, we perform a minimization of the position of the Gaussian shells while keeping the position of the Gaussian cores fixed using the FIRE⁵² algorithm. The parameters of the C-GeM model are the global parameters, λ , ω_{core} , ω_{shell} , χ_{shell} , R_{shell} and atomic element specific parameters χ_i and R_i are trained and validated with an *ab initio* EPS determined on a grid at a $2.0 \cdot vdW_{radius}$ distance using Density Function Theory (DFT). Using the Q-chem⁵³ simulation package, we generated the EPS using the $\omega b97bX-V/def2-qzvpp$ ⁵⁴ while also evaluating the same EPS using Hirshfeld-I²⁰ charges. We trained C-GeM on the EPS of 11 molecules comprised of C, H, O, and/or Cl atoms as shown in Table 1, with the quality of the EPS agreement with the DFT surfaces is analyzed using the mean absolute deviation (MAD)

$$MAD = \frac{1}{n} \sum_{i=1}^n |X_i - Y_i| \quad (9)$$

where n is the number of grid points of the generated surface, X_i is the calculated EPS on grid point i , and Y_i is the corresponding target DFT value on grid point i . In addition, we assess the applicability of C-GeM for reactive chemistry in which we chose a system of hydrogen chloride dissociation in water or in the gas phase. For the aqueous hydrogen chloride systems we used 24 trajectories comprised of Cl^- and H_3O^+ with 4 water molecules as previously reported^[55], to obtain atomic core geometries of the undissociated, transition state, and contact ion pair for CGeM to predict the shell positions. We also compare C-GeM against EEM derived charges obtained from the LAMMPS⁵⁶ ReaxFF⁵⁷ implementation of EEM using the aqueous chloride parameters^[44].

RESULTS

Table 1 provides the atomic and global parameters of the C-GeM model developed on a training set of 11 arbitrarily chosen molecules containing C, H, O, and/or Cl atoms in order to reproduce their *ab initio* EPS, and compares the performance of C-GeM to an EDC based method optimized using EEM, and an EPFC approach using Hirshfeld-I charges. Overall the optimized C-GeM parameters

combined with minimizing the Gaussian electron shell configurations to generate the EPS is nearly as good as the EPFC fitted charges, but using far fewer parameters, while vastly outperforming the EEM approach. We note that the electron shell parameters are “global” in the sense that they are independent of the underlying atom type. We also note that the EEM parameters from ref.44 were trained with Mulliken charges, and the EEM performance might be improved when trained with better partial charge models such as Hirshfeld-I.”

Table 1. Training performance of C-GeM, EEM, and Hirshfeld-I fitted charges for reproducing the *ab initio* electrostatic potential surface. Comparison of MAD (eV units) of the EPS generated with DFT (ω b97bX-V/def2-qzvpp⁵⁴) for 11 training molecules containing H, C, O, and Cl atoms. The resulting C-GeM parameters are fixed when used in the remaining validation tests.

Molecule		EEM	Hirshfeld-I	C-GeM
H ₂ O		0.06	0.10	0.04
HCl		0.12	0.08	0.03
C ₂ H ₄ O ₂ Cl		0.20	0.06	0.16
CH ₃ OH		0.07	0.04	0.05
C ₆ H ₁₀		0.05	0.04	0.05
C ₁₂ H ₇ Cl ₃ O ₂		0.24	0.05	0.10
C ₆ H ₆ O ₃		0.11	0.05	0.14
C ₇ H ₇ ClO		0.23	0.07	0.13
C ₈ H ₆ Cl ₂		0.25	0.06	0.10
C ₄ H ₆ O ₄		0.17	0.02	0.13
C ₆ H ₅ Cl		0.27	0.05	0.06
Average MAD		0.16	0.07	0.09
Global Parameters				
λ	$\omega_{core}(\text{\AA}^{-1})$	$\omega_{shell}(\text{\AA}^{-1})$	$\chi_{shell}(eV)$	$R_{shell}(\text{\AA})$
1.948	0.34	5.49	-14.803	0.784
Atomic Parameters				
Atom type	$\chi(eV)$		$R(\text{\AA})$	
H	14.95		0.556	
C	15.76		0.717	
O	19.35		0.501	
Cl	18.12		0.994	

Having optimized the C-GeM model parameters on a small set of training molecules, we now validate our approach for reproducing the *ab initio* EPS for small organic and halogenated molecules, for hydrocarbons, as well as larger drug or drug-like molecules (Table 2). For small molecules and hydrocarbons C-GeM performs outstanding, predicting the *ab initio* EPS with errors that are comparable to or better than Hirshfeld-I charges. For larger molecules containing three or

all four of the atomic elements developed here, the EPS generated from C-GeM is still in quite good agreement with the DFT electrostatic potential, with an average MAD of 0.11eV compared to 0.06eV for Hirshfeld-I charges and 0.21eV for EEM derived charges.

Table 2. Testing performance of C-GeM, EEM, and Hirshfeld-I fitted charges for reproducing the *ab initio* electrostatic potential surface for small organic, halogenated, hydrocarbon, and large molecules. Comparison of the MAD (in units of eV) of the EPS generated with DFT (ω b97bX-V/def2-qzvpp⁵⁴) for 22 molecules containing H, C, O, and Cl atoms.

Small Molecules	EEM	Hirshfeld-I	C-GeM
HO ₂	0.07	0.05	0.1
H ₂ O ₂	0.04	0.04	0.04
ClO	0.22	0.08	0.1
Cl ₂ O	0.13	0.05	0.03
ClHO ₂	0.27	0.17	0.18
ClHO ₄	0.06	0.1	0.09
Cl ₂	0.05	0.05	0.02
ClOH	0.19	0.11	0.1
Cl ₂ O ₂	0.12	0.04	0.04
CH ₃ Cl	0.25	0.08	0.03
Average MAD	0.14	0.08	0.07
Hydrocarbons	EEM	Hirshfeld-I	C-GeM
C ₆ H ₆	0.01	0.04	0.03
C ₆ H ₁₂	0.03	0.03	0.01
C ₁₀ H ₂₀	0.06	0.04	0.02
(CH ₃) ₂ C ₄ H ₄	0.03	0.03	0.02
Average MAD	0.03	0.04	0.02
Large Molecules	EEM	Hirshfeld-I	C-GeM
C ₄ H ₆ O ₂	0.06	0.06	0.08
C ₆ H ₁₃ Cl	0.32	0.06	0.03
C ₉ H ₈ O ₂	0.14	0.04	0.12
C ₁₀ H ₁₁ ClO ₃	0.23	0.06	0.20
C ₂₀ H ₂₁ ClO ₃	0.30	0.05	0.15
C ₆ H ₁₄ O	0.09	0.05	0.04
C ₈ H ₆ Cl ₂ O ₃	0.25	0.07	0.10
C ₇ H ₅ ClO ₂	0.31	0.06	0.14
Average MAD	0.21	0.06	0.11

Figure 3 shows the optimized electron shell configuration and the corresponding C-GeM EPS compared to the *ab initio* electrostatic potential surface, in which it is evident that the electronic shells move from atoms with lower electronegativity (H and C) to atoms with higher electronegative (O and Cl). But C-GeM also allows for additional subtle electrostatic effects in which the shells of the Cl₂ molecule are drawn inward along the direction of the bond axis, resulting in a positive sigma hole at the edges of the chlorine molecule. The sigma hole feature cannot be accounted for with atomic centered partial charges, instead requiring multipolar electrostatic models to account for the anisotropy. In addition, the G-GeM generated EPS of the herbicide 2-4-Dichlorophenoxyacetic acid (C₈H₆Cl₂O₃) and cholesterol drug Fenofibrate (C₂₀H₂₁ClO₃) exhibit excellent agreement with the *ab initio* EPS, and are good examples illustrating the strength of C-GeM which can account for sp² and sp³ hybridized carbon, in addition to sp² and sp hybridized oxygens, as well as the presence of more than one chlorine atom.

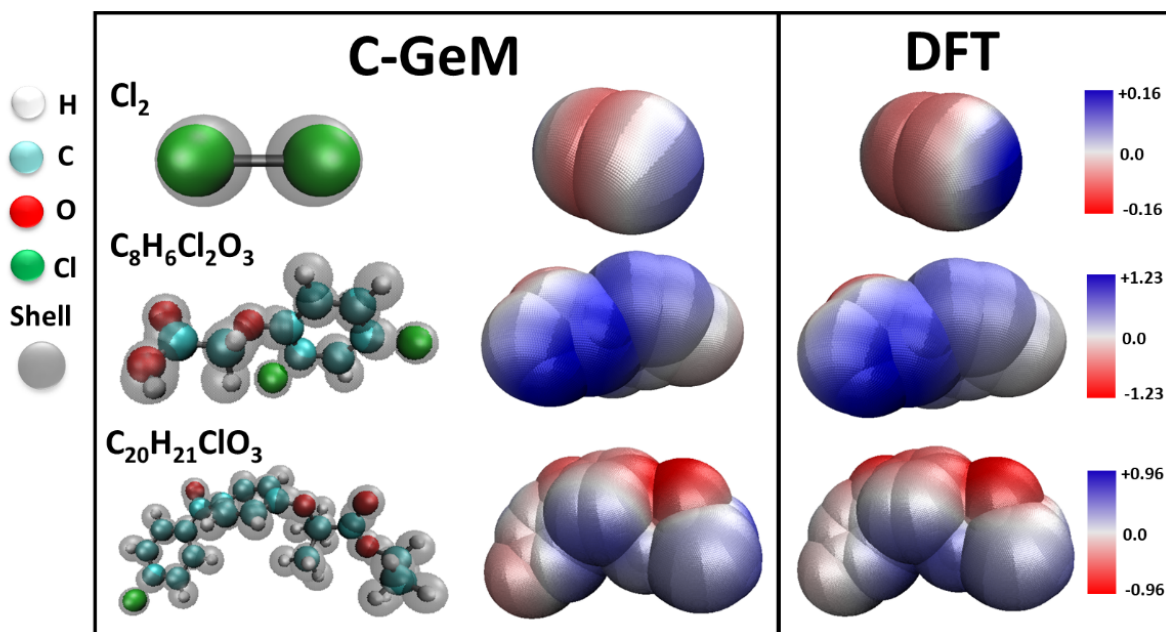


Figure 3. Representative examples of the optimized electronic shells and resulting electrostatic potential for various molecules using C-GeM and compared to *ab initio* (wB97X-V/def2-qzvpp). The EPS generated with C-GeM for Cl₂ shows the presence of the sigma hole. The herbicide 2-4-Dichlorophenoxyacetic acid (C₈H₆Cl₂O₃) and cholesterol drug Fenofibrate (C₂₀H₂₁ClO₃) containing different types of hybridized carbon, oxygen, and one or two chlorine atoms.

In order to test the applicability of C-GeM to molecular dynamics for reactive force fields, we obtained DFT (B3LYP/aug-cc-pVTZ) reactive path trajectories of hydrogen chloride dissociation, in which it has previously been determined that four water molecules are the smallest

water cluster required to stabilize the dissociation by forming the solvent shared ion pair $\text{H}_3\text{O}^+(\text{H}_2\text{O})_3\text{Cl}^-$.⁵⁵ Using the DFT reaction path trajectories going from the initial undissociated (UD) configuration through the transition state (TS) and to the final contact ion pair (CIP), we assessed whether C-GeM can predict the dissociation of HCl into ionic fragments in the presence of the 4 water molecules. Figure 4a illustrates the DFT reaction path trajectories of the three stationary points and the corresponding core configurations (top) and the corresponding optimization showing the predicted shell configuration using C-GeM (bottom). It is found that in the UD configuration the HCl molecule is comprised of two shells - one on centered on the chlorine atom and the other shifted off the hydrogen atom towards the more electronegative chlorine. Along the dissociation pathway to the TS configuration, the shell of the hydrogen atom detaches completely and binds to the chlorine forming the Cl^- anion. This additional electronic shell on the Cl^- anion can now pair with the hydronium cation H_3O^+ to stabilize the CIP configuration.

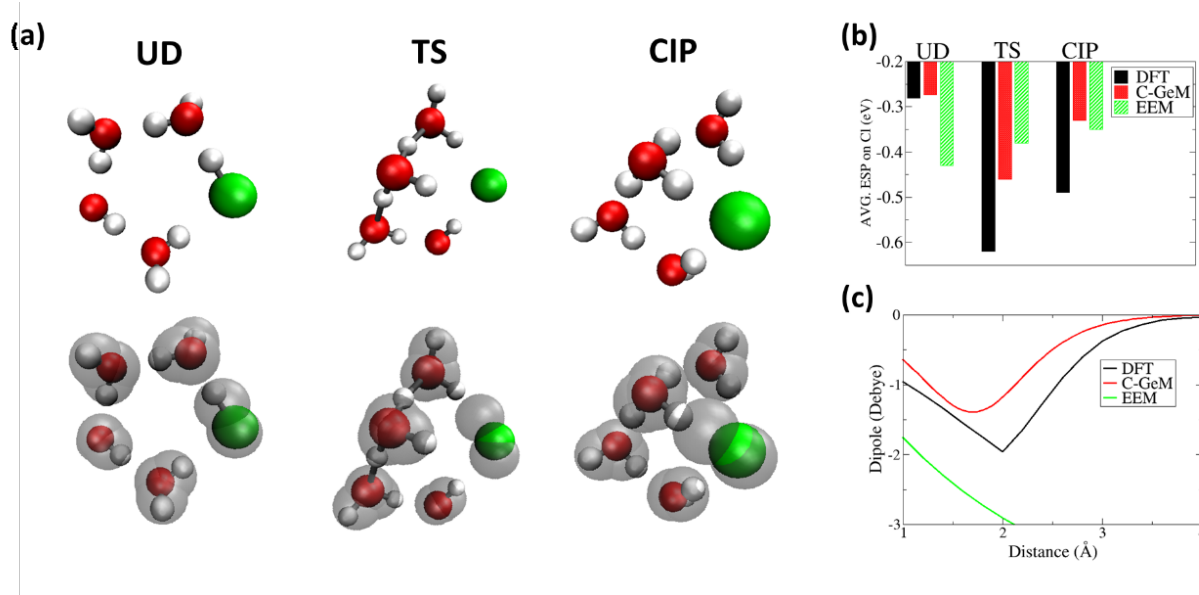


Figure 4. *Hydrogen chloride dissociation in aqueous solvent using the C-GeM model.* (a) Atomic core position of the undissociated, transition state and contact ion pair configurations derived from DFT (B3LYP/aug-cc-pVTZ) trajectories (top) and C-GeM prediction of the electronic shell positions (bottom). (b) Average electrostatic potential on the chlorine atom in the undissociated, transition state, and contact ion pair configurations comparing DFT (black), C-GeM (red) and EEM (green). (c) Dipole moment of isolated HCl as a function of H-Cl distance comparing DFT (black), C-GeM (red) and EEM (green).

Figure 4b plots the average EPS on the chlorine atom at the three stationary points using EEM, DFT and C-GeM in which C-GeM is in qualitative agreement with the DFT result, characterized by an increased negative charge on the chlorine atom in the TS followed by a decrease

of charge when forming the CIP. By contrast, the EEM result shows a qualitatively incorrect decrease in charge on the chlorine at the TS and therefore fails to describe the charge transfer from hydrogen to chlorine upon dissociation in water. This behavior is a general flaw in charge equilibration methods, and is expected to exist even in the more advanced ACKS2 and split charges methods, which rely on a reference charge state to which the atoms are constrained to upon dissociation regardless of the environment. C-GeM on the other hand can account for both dissociation in solution and gas phase of the hydrogen chloride. This is supported in Figure 4c which plots the dipole moment of the isolated HCl as it dissociates to neutral atoms in which C-GeM shows reasonable agreement with DFT (which has an unphysical Coulson–Fischer point), whereas the EEM dipole moment goes to infinity upon dissociation due to the unphysical charge transfer at long distances.

DISCUSSION AND CONCLUSION

C-GeM is a new method for predicting the electrostatic properties in molecules by coarse graining the electrons through a core-shell description of the atom, and which allows the shell positions to adjust according to the molecular configuration in a complex molecular environment. The basic principle behind C-GeM is that the magnitude of the core-shell interaction reflects the electronegativity of a given atom type, enabling the shells to flow from atoms with lower to higher electronegativities, while also giving rise to complex electrostatic behavior of sigma holes present in many systems but illustrated here with a homonuclear halogen diatomic. [We note that the molecular origin of the sigma hole remains controversial.⁵⁸⁻⁶⁰ The C-GeM model determines the strength of the sigma hole based on both the electronegativity of the atom and arising from a general induction effect, which treats charge transfer and polarization on the same footing. Future work will consider how and if C-GeM can produce a sigma hole in more complex molecular environments than that of the Cl₂ molecule presented here.](#)

While future work will require the formulation of parameters for other atomic elements, the C-GeM model has been illustrated for H, C, O and Cl atoms to demonstrate its development and performance. C-GeM has been shown to predict the EPS of 22 molecules containing with very good accuracy, as well as accounting for the correct dissociation of hydrogen chloride into ionic fragments in solution and to neutral atoms in the gas phase. Furthermore, the C-GeM approach offers several advantages over other EDC/EEM type methods: (1) using the optimized position of the shells, the electrostatic potential of a given system is easily obtained; (2) the use of explicit electrons with an

integer charge value allows for a straightforward description of bond breaking into ionic fragments; (3) out-of-plane polarization for planar or aromatic molecules is well-described; (4) the anisotropy of shell positions with respect to the atomic cores can give rise to electrostatic features such as the sigma hole, which usually require a multipole expansion; (5) C-GeM can be incorporated in molecular dynamics (MD) simulations in such a way that at in every MD time step the positions of the coarse-grained electrons are minimized in a Born Oppenheimer type approach to describe polarization and charge transfer.

ACKNOWLEDGMENTS. We thank the National Science Foundation for support under Grant No. CHE-1665315 for the development of C-GeM for electrostatic potentials and non-reactive force fields, and for support by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Scientific Discovery through Advanced Computing (SciDAC) program for its reactive force field capabilities as illustrated on HCl. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We would like to thank Abdulrahman Aldossary for providing the reaction path trajectories of the (H₂O)₄HCl system, and Abdulrahman Aldossary, Luke Bertels, Akshaya Das, Matthias Loipersberger, Christopher Stein, and Martin Head-Gordon for discussions.

REFERENCES

1. Jing, Z. *et al.* Polarizable force fields for biomolecular simulations: recent advances and applications. *Annu. Rev. Biophys.* **48**, 371–394 (2019).
2. Brock, K., Talley, K., Coley, K., Kundrotas, P. & Alexov, E. Optimization of electrostatic interactions in protein-protein complexes. *Biophys. J.* **93**, 3340–3352 (2007).
3. Cisneros, G. A., Karttunen, M., Ren, P. & Sagui, C. Classical electrostatics for biomolecular simulations. *Chem. Rev.* **114**, 779–814 (2014).
4. Welborn, V. V., Ruiz Pestana, L. & Head-Gordon, T. Computational optimization of electric fields for better catalysis design. *Nature Catal.* **1**, 649–655 (2018).
5. Furukawa, H., Cordova, K. E., O’Keeffe, M. & Yaghi, O. M. The chemistry and applications of metal-organic frameworks. *Science* **341**, 1230444 (2013).
6. Zhou, H. C., Long, J. R. & Yaghi, O. M. Introduction to metal-organic frameworks. *Chem. Rev.* **112**, 673–674 (2012).
7. Weckhuysen, B. M. & Yu, J. Recent advances in zeolite chemistry and catalysis. *Chem. Soc. Rev.* **44**, 7022–7024 (2015).
8. He, X., Zhu, Y. & Mo, Y. Origin of fast ion diffusion in super-ionic conductors. *Nature Comm.* **8**, 15893 (2017).
9. Wang, A., Kadam, S., Li, H., Shi, S. & Qi, Y. Review on modeling of the anode solid electrolyte interphase (SEI) for lithium-ion batteries. *Comp. Mat.* **4**, 15 (2018).

10. L., C., A., F. K. & U., S. Fuel cells – fundamentals and applications. *Fuel Cells* **1**, 5–39 (2001).
11. Suen, N. T. *et al.* Electrocatalysis for the oxygen evolution reaction: Recent development and future perspectives. *Chem. Soc. Rev.* **46**, 337–365 (2017).
12. Chen, D. L., Stern, A. C., Space, B. & Johnson, J. K. Atomic charges derived from electrostatic potentials for molecular and periodic systems. *J. Phys. Chem. A* **114**, 10225–10233 (2010).
13. Shi, Y. *et al.* Polarizable atomic multipole-based AMOEBA force field for proteins. *J. Chem. Theo. Comput.* **9**, 4046–4063 (2013).
14. Singh, U. C. & Kollman, P. A. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **5**, 129–145 (1984).
15. Mortier, W. J., Ghosh, S. K. & Shankar, S. Electronegativity-equalization method for the calculation of atomic charges in molecules. *J. Am. Chem. Soc.* **108**, 4315–4320 (1986).
16. Verstraelen, T., Ayers, P. W., Van Speybroeck, V. & Waroquier, M. ACKS2: Atom-condensed Kohn-Sham DFT approximated to second order. *J. Chem. Phys.* **138**, 074108 (2013).
17. Geidl, S. *et al.* High-quality and universal empirical atomic charges for chemoinformatics applications. *J. Cheminform.* **7**, 59 (2015).
18. Nistor, R. A., Polihronov, J. G., Müser, M. H. & Mosey, N. J. A generalization of the charge equilibration method for nonmetallic materials. *J. Chem. Phys.* **125**, 094108 (2006).
19. Mulliken, R. S. Electronic population analysis on LCAO-MO molecular wave functions. I. *J. Chem. Phys.* **23**, 1833–1840 (1955).
20. Bultinck, P., Van Alsenoy, C., Ayers, P. W. & Carbó-Dorca, R. Critical analysis and extension of the Hirshfeld atoms in molecules. *J. Chem. Phys.* **126**, (2007).
21. Heidar-Zadeh, F. *et al.* Information-theoretic approaches to atoms-in-molecules: Hirshfeld family of partitioning schemes. *J. Phys. Chem. A* **122**, 4219–4245 (2018).
22. Manz, T. A. & Sholl, D. S. Improved atoms-in-molecule charge partitioning functional for simultaneously reproducing the electrostatic potential and chemical states in periodic and nonperiodic materials. *J. Chem. Theo. Comput.* **8**, 2844–2867 (2012).
23. Reed, A. E., Weinstock, R. B. & Weinhold, F. Natural population analysis. *J. Chem. Phys.* **83**, 735–746 (1985).
24. Salomon-Ferrer, R., Case, D. A. & Walker, R. C. An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **3**, 198–210 (2013).
25. Brooks, B. R. *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 (1983).
26. Lotan, I. & Head-Gordon, T. An analytical electrostatic model for salt screened interactions between multiple proteins. *J. Chem. Theo. Comput.* **2**, 541–555 (2006).
27. Yap, E. H. & Head-Gordon, T. New and efficient Poisson-Boltzmann solver for interaction of multiple proteins. *J. Chem. Theo. Comput.* **6**, 2214–2224 (2010).
28. Jurrus, E. *et al.* Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **27**, 112–128 (2018).
29. Demerdash, O., Wang, L. & Head-Gordon, T. Advanced models for water simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **8**, e1355 (2018).
30. Mortier, W. J., Ghosh, S. K. & Shankar, S. Electronegativity equalization method for the calculation of atomic charges in molecules. *J. Am. Chem. Soc.* **108**, 4315–4320 (1986).
31. Kadantsev, E. S., Boyd, P. G., Daff, T. D. & Woo, T. K. Fast and accurate electrostatics in metal organic frameworks with a robust charge equilibration parameterization for high-

- throughput virtual screening of gas adsorption. *J. Phys. Chem. Lett.* **4**, 3056–3061 (2013).
32. Ionescu, C.-M., Geidl, S., Svobodovávař, R., Eková, S. & Koč, J. Rapid calculation of accurate atomic charges for proteins via the electronegativity equalization method. *J. Chem. Inf. Model.* **53**, 2548–2558 (2013).
 33. Van Duin, A. C. T., Dasgupta, S., Lorant, F. & Goddard, W. A. ReaxFF: A reactive force field for hydrocarbons. *J. Phys. Chem. A* **105**, 9396–9409 (2001).
 34. Liang, T., Devine, B., Phillpot, S. R. & Sinnott, S. B. Variable charge reactive potential for hydrocarbons to simulate organic-copper interactions. *J. Phys. Chem. A* **116**, 7976–7991 (2012).
 35. Leven, I. & Head-Gordon, T. Inertial extended-Lagrangian scheme for solving charge equilibration models. *Phys. Chem. Chem. Phys.* **21**, 18652–18659 (2019).
 36. Warren, G. L., Davis, J. E. & Patel, S. Origin and control of superlinear polarizability scaling in chemical potential equalization methods. *J. Chem. Phys.* **128**, 144110. (2008).
 37. Nistor, R. A. & Müser, M. H. Dielectric properties of solids in the regular and split-charge equilibration formalisms. *Phys. Rev. B* **79**, 104303 (2009).
 38. Mathieu, D. Split charge equilibration method with correct dissociation limits. *J. Chem. Phys.* **127**, 224103 (2007).
 39. Nistor, R. A., Polihronov, J. G., Müser, M. H. & Mosey, N. J. A generalization of the charge equilibration method for nonmetallic materials. *J. Chem. Phys.* **125**, 094108 (2006).
 40. Islam, M. M., Kolesov, G., Verstraelen, T., Kaxiras, E. & Van Duin, A. C. T. EReaxFF: A pseudoclassical treatment of explicit electrons within reactive force field simulations. *J. Chem. Theo. Comput.* **12**, 3463–3472 (2016).
 41. Dapp, W. B. & Müser, M. H. Redox reactions with empirical potentials: Atomistic battery discharge simulations. *J. Chem. Phys.* **139**, 064106 (2013).
 42. Ekesan, S., Kale, S. & Herzfeld, J. Transferable pseudoclassical electrons for aufbau of atomic ions. *J. Comput. Chem.* **35**, 1159–1164 (2014).
 43. Su, J. T. & Goddard, W. A. Excited electron dynamics modeling of warm dense matter. *Phys. Rev. Lett.* **99**, 185003 (2007).
 44. Rahaman, O. *et al.* Development of a ReaxFF reactive force field for aqueous chloride and copper chloride. *J. Phys. Chem. A* **114**, 3556–3568 (2010).
 45. Naserifar, S., Brooks, D. J., Goddard, W. A. & Cvicek, V. Polarizable charge equilibration model for predicting accurate electrostatic interactions in molecules and solids. *J. Chem. Phys.* **146**, 124117 (2017).
 46. Chialvo, A. A., Moucka, F., Vlcek, L. & Nezbeda, I. Vapor-liquid equilibrium and polarization behavior of the GCP water model: Gaussian charge-on-spring versus dipole self-consistent field approaches to induced polarization. *J. Phys. Chem. B* **119**, 5010–5019 (2015).
 47. Kiss, P. T., Sega, M. & Baranyai, A. Efficient handling of gaussian charge distributions: An application to polarizable molecular models. *J. Chem. Theo. Comput.* **10**, 5513–5519 (2014).
 48. Cisneros, G. A. Application of gaussian electrostatic model (GEM) distributed multipoles in the AMOEBA force field. *J. Chem. Theo. Comput.* **8**, 5072–5080 (2012).
 49. Paricaud, P., Předota, M., Chialvo, A. A. & Cummings, P. T. From dimer to condensed phases at extreme conditions: Accurate predictions of the properties of water by a Gaussian charge polarizable model. *J. Chem. Phys.* **122**, 244511 (2005).
 50. Chialvo, A. A. & Cummings, P. T. Simple transferable intermolecular potential for the molecular simulation of water over wide ranges of state conditions. *Fluid Phase Equilib.* **150–151**, 73–81 (1998).

51. Elking, D., Darden, T. O. M. & Woods, R. J. Gaussian induced dipole polarization model. *J. Comput. Chem.* **28**, 1261–1274 (2007).
52. Bitzek, E., Koskinen, P., Gähler, F., Moseler, M. & Gumbsch, P. Structural relaxation made simple. *Phys. Rev. Lett.* **97**, 1–4 (2006).
53. Shao, Y. *et al.* Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* **113**, 184–215 (2015).
54. Mardirossian, N. & Head-Gordon, M. ω B97X-V: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy. *Phys. Chem. Chem. Phys.* **16**, 9904–9924 (2014).
55. Forbert, H., Masia, M., Kaczmarek-Kedziera, A., Nair, N. N. & Marx, D. Aggregation-induced chemical reactions: Acid dissociation in growing water clusters. *J. Am. Chem. Soc.* **133**, 4062–4072 (2011).
56. Plimpton, S. Fast Parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).
57. Aktulga, H. M., Fogarty, J. C., Pandit, S. A. & Grama, A. Y. Parallel reactive molecular dynamics: Numerical methods and algorithmic techniques. *Parallel Comput.* **38**, 245–259 (2012).
58. Gonthier, J. F.; Thirman, J.; Head-Gordon, M. Understanding non-covalent interactions: correlated energy decomposition analysis and applications to halogen bonding. *Chimia (Aarau)* **72**, 193–198 (2018).
59. Politzer, P.; Murray, J. S.; Clark, T. Halogen bonding and other σ -hole interactions: a perspective. *Phys. Chem. Chem. Phys.* **15**, 11178–11189 (2013).
60. Thirman, J.; Engelage, E.; Huber, S. M.; Head-Gordon, M. Characterizing the interplay of Pauli repulsion, electrostatics, dispersion and charge transfer in halogen bonding with energy decomposition analysis. *Phys. Chem. Chem. Phys.* **20**, 905–915 (2018).