JCTC Journal of Chemical Theory and Computation

Article

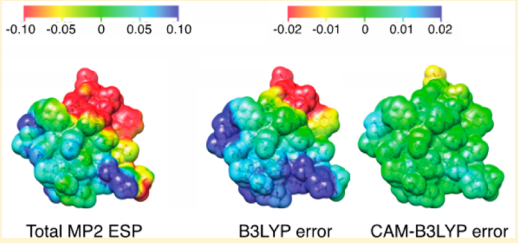
pubs.acs.org/JCTC

Electrostatic Potential of Insulin: Exploring the Limitations of Density Functional Theory and Force Field Methods

Sofie Jakobsen,[†] Kasper Kristensen,[†,‡] and Frank Jensen*[,†]

[†]Department of Chemistry and [‡]The qLEAP Center for Theoretical Chemistry, Aarhus University, DK-8000 Aarhus, Denmark

ABSTRACT: We show that standard density functional theory leads to large errors in the electron density distribution compared to reference second order Møller−Plesset perturbation theory (MP2) calculations for the insulin molecule and zwitterionic peptides, while range-separated versions perform much better. The error is quantified in terms of the electrostatic potential (ESP) on a molecular surface, which shows that standard density functional theory incorrectly predicts partial electron transfer from anionic to cationic sites. In addition, we compare the MP2 calculated ESPs to those predicted by commonly used force fields. Several fixed charge force fields display very similar performances with rather large errors, while polarizable force fields significantly reduce the error. Solvation enhances the molecular ESP, which is partly accounted for by fixed charge force fields, but polarizable force fields again perform significantly better.



INTRODUCTION

The interaction between molecules at large distances is determined exclusively by the electrostatic potential (ESP) arising from the nuclei and the electron distribution.[1] The molecular ESP is important for predicting chemical reactivity and for calculating intermolecular interactions. For a fixed nuclear geometry the ESP depends only on the electron density, and the accuracy of the latter depends on the electronic structure method used for solving the Schrödinger equation. It should be emphasized, though, that the electron density and thereby also the ESP for a given system can be distorted by the surroundings. In a polar solvent, such as water, the electron density becomes polarized and the ESP for a solvated system can differ significant from the corresponding gas phase ESP.

Wave function methods have the distinct advantage that there is a well-defined path for improving the results toward the theoretical limiting value. Hartree−Fock (HF) is a mean-field description while second order Møller−Plesset perturbation theory (MP2), coupled-cluster with singles and doubles (CCSD), and coupled-cluster with singles, doubles, and an approximate treatment of triples (CCSD(T)) provide an accepted progression for improving the description of the electron−electron correlation.[2] Similarly, the cc-pVXZ basis sets[3] provide a well-defined path for reducing the basis set error. Recent technological breakthroughs have enabled the use of correlated wave function methods for treating systems with thousands of atoms,[4] which previously have only been accessible by mean-field theories like density functional theory (DFT).

DFT has become very popular over the latest 20 years for electronic structure calculations in chemistry because of its favorable combination of accuracy and low computational cost. DFT is a mean-field description where the electron exchange and correlation (XC) effects are modeled by a functional of the electron density. A large number of XC functionals have been proposed, many of which contain empirical parameters, and hybrid functionals like B3LYP[5] and BHHLYP[6] in addition include exact (HF) exchange (20% and 50%, respectively). The B3LYP functional has become a de facto standard for many routine applications in chemistry, while pure functionals like BLYP or PBE are commonly used for periodic systems where inclusion of exact exchange is computationally expensive.

More recently range-separated, also called long-range corrected, versions of these functionals have been proposed, where the exchange functional is partitioned into a short- and long-range part by means of a switching function.[7] The short-range part is described by a standard DFT exchange functional, while the long-range part is given by the HF expression. The CAM-B3LYP[8] method is an example of such a range-separated functional which is derived from the B3LYP method. A motivation for developing range-separated functionals is that some molecular properties depend critically on the potential far from the nuclei. Charge-transfer excited states, for example, are calculated to be much too low in energy with standard functionals, while range-separated versions perform significantly better.[9] Electron affinity is another property where standard functionals have serious deficiencies. With the exception of strongly bound anions, standard functionals are only capable of binding a fraction of the excess electron.[10] For small anions and using standard basis sets, this problem is not visible, as the excess electron density is constrained to remain in the molecular vicinity by the basis set. When using (very) extended basis sets, or when the system is large and also contains electron deficient positions, the formal anionic center will only retain a fraction of the excess electron, while the remaining fraction will be transferred to other parts of the system. This is especially problematic for biomolecules like proteins, where

Received: May 31, 2013

   dx.doi.org/10.1021/ct400452f | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

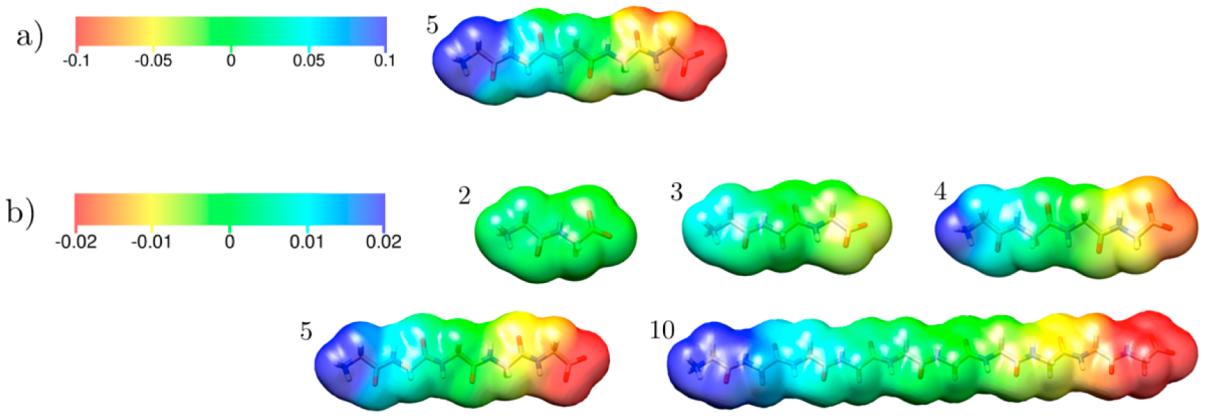


Figure 1. (a) Total MP2/cc-pVDZ ESP for Gly_5 plotted on an HF isodensity surface corresponding to $\rho = 1.0 \times 10^{-4}$ au. (b) The ESP difference between MP2/cc-pVDZ and B3LYP/cc-pVDZ (MP2 − B3LYP) for Gly_2, Gly_3, Gly_4, Gly_5, and Gly_{10}. The distances between the C and N atoms in the two charged groups are 6.2 Å, 9.7 Å, 13.5 Å, 17.0 Å, and 35.3 Å, respectively.

both anionic and cationic side-chains often are present in the same system. It is thus expected that standard XC functionals will provide a significantly different electron distribution, and thus ESP, than range-separated functionals.

In the present paper we show that the use of range-separated XC functionals is crucial for calculating the electron distribution in proteins containing both anionic and cationic sites, and that the use of standard functionals leads to large errors. The effect is illustrated by comparing the DFT and MP2 calculated ESPs on molecular surfaces for a series of poly glycines and for the insulin monomer, which is a 51 amino acid peptide having a total of 787 atoms and 11 charged residues.

For the insulin system the performance of several commonly used force fields (FF) for representing the ESP is quantified against the reference MP2 result.[4c] Since biomolecular force fields are parametrized for the condensed phase, where solvation effects polarize the charge density, we additionally compare the ESP for the solvated and nonsolvated arg-ala-asp tripeptide. Fixed charge force fields lead to large absolute errors, but force fields including higher order multipoles and atomic polarizabilities significantly improve the performance.

COMPUTATIONAL DETAILS

The insulin geometry was obtained from Smith et al. (PDB id 1MSO),[11] where zinc ions and water molecules were removed. The poly glycines geometries were minimized with the OPLS-AA/L FF,[12] while the arg-ala-asp tripeptide was minimized in explicit water using NAMD[13] with the CHARMM27 FF.[14]

Gas phase MP2 and DFT calculations have been done using LSDALTON[15] or Gaussian09.[16] In particular, the insulin MP2 calculation was carried out using the Divide-Expand-Consolidate (DEC) scheme described in ref 4c. Convergence of DFT calculations for large zwitterionic systems is not trivial, and standard methods often converge to saddle points in the parameter space. We have in the present cases verified that the solutions correspond to minima in the parameter space by calculating the lowest Hessian eigenvalue. This issue will be addressed further in a separate paper.

Force field calculations of the ESP have been done using Tinker v. 6.0 for the AMBER99sb,[17] CHARMM22 cmap,[18] OPLS-AA/L,[12] AMOEBA Pro04, and AMOEBA Bio09[19] force fields. The AMOEBA force fields include distributed multipoles and polarizabilities, while the others use a fixed charge model for the electrostatic interaction.

All ESP calculations (electronic structure and force field methods) are performed on a cubic grid with a grid separation of 0.30 au (0.16 Å). The plots visualizing the ESP have been made with the USCF Chimera software.[20]

The calculations of the ESP from the solvated arg-ala-asp tripeptide include explicit water molecules. The solvation shell was created by solvating the peptide in a 40 Å box of water, followed by minimization and a 10 ps molecular dynamics simulation. All water molecules more than 8 Å from the peptide were subsequently removed, resulting in a system containing 223 water molecules. The ESP for the tripeptide including the solvent polarization was calculated with the Effective Fragment Potential (EFP1)[21] QM/MM method using the GAMESS program.[22] The tripeptide constitutes the QM region, whereas the MM region includes the water molecules and are described by a one-electron potential containing three nonbonded energy terms: Coulomb interaction, polarization, and exchange-repulsion. The Coulomb interaction and polarization is modeled by distributed multipoles and polarizabilities, while the exchange-repulsion term is obtained by a fitting procedure.[21] The preparameterized DFT water model has been used in this study.[23]

ELECTROSTATIC POTENTIAL

The ESP ϕ at a point $\boldsymbol{r}$ arising from M nuclei and the electron density ρ is given by eq 1

$$\phi(\boldsymbol{r}) = \sum_{a=1}^{M} \frac{Z_a}{|\boldsymbol{R}_a - \boldsymbol{r}|} - \int \frac{\rho(\boldsymbol{r}')}{|\boldsymbol{r}' - \boldsymbol{r}|} d\boldsymbol{r}' \quad (1)$$

Here $\boldsymbol{R}_a$ and Z_a are the position and charge of nucleus a, respectively. The electron density can be calculated by a variety of electronic structure methods. As the ESP only depends on the electron density distribution, as shown in eq 1, the results are expected to converge rapidly in terms of both electron correlation and basis set. Tsiper and Burke have shown that the ESP on an isodensity surface, which encompasses essentially all the electron density, determines the ESP at all points outside this surface.[24] The quality of the ESP for a given method can thus conveniently be quantified by the deviation from a reference potential on a surface defined by a suitable low isodensity value. We have in the present case used a surface corresponding to an isodensity value of 10^{-4} au (electrons/bohr^3) from an HF/cc-pVDZ calculation, but the results are

dx.doi.org/10.1021/ct400452f | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

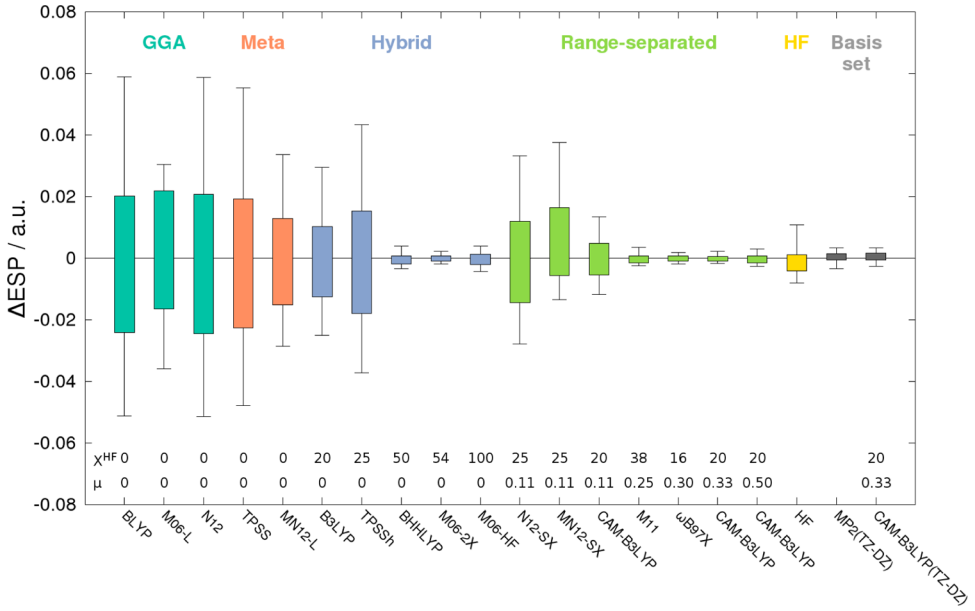


Figure 2. Box-plots showing the ESP difference for various DFT functionals compared to the MP2/cc-pVDZ reference for Gly_5 based on 4,596 points in the density range $[8 \times 10^{-5}-10^{-4}]$ au. The amount of HF exchange (X^{HF}) in percentage and the range separator μ in units of inverse Bohr radii is shown under the bars for each XC functional.

Table 1. Mean Absolute Deviations (MAD) and Maximum Absolute Deviations (MaxAD) in au for the ESP Compared to the MP2/cc-pVDZ Reference for Gly_5[a]

	XC type	X^{HF}	μ	MAD	MaxAD
BLYP	GGA	0	0	0.029	0.059
M06-L	GGA	0	0	0.017	0.036
N12	GGA	0	0	0.024	0.058
MN12-L	meta-GGA	0	0	0.014	0.033
TPSS	meta-GGA	0	0	0.023	0.055
B3LYP	hybrid GGA	20	0	0.012	0.030
TPSSh	hybrid GGA	25	0	0.017	0.043
BHHLYP	hybrid GGA	50	0	0.002	0.004
M06−2X	hybrid GGA	54	0	0.001	0.002
M06-HF	hybrid GGA	100	0	0.002	0.004
N12-SX	RS	25	0.11	0.014	0.033
MN12-SX	RS	25	0.11	0.012	0.038
CAM-B3LYP	RS	20	0.11	0.006	0.013
M11	RS	38	0.25	0.001	0.003
ωB97X	RS	16	0.30	0.001	0.002
CAM-B3LYP	RS	20	0.33	0.001	0.002
CAM-B3LYP	RS	20	0.50	0.001	0.003
HF				0.004	0.008
MP2(TZ-DZ)				0.0012	0.0035
CAM-B3LYP(TZ-DZ)	RS	20	0.33	0.0013	0.0034

[a]GGA = generalized gradient approximation, RS = range-separated. The amount of HF exchange (X^{HF}) is in percentage and the range separator (μ) is in units of a_0^{-1}.

insensitive to the exact value. An isodensity surface corresponding to 10^{-4} au samples typical van der Waals contact distances between molecules.

RESULTS

ESP for Zwitterionic Systems. The incorrect charge separation of zwitterionic systems by standard XC functionals can be illustrated by the ESP for a series of poly glycine peptides (Gly_n, n = 2, 3, 4, 5, and 10) with charged N- and C-terminal groups ($-NH_3^+$ and $-CO_2^-$). Four different classes of DFT functionals have been compared to the MP2 reference: general gradient approximation (GGA) functionals using the density and its gradient; meta functionals, which also include the kinetic energy density; hybrid functionals, which include a fraction of exact HF exchange (X^{HF}); and range-separated (RS) functionals where the exchange functional is interpolated between the DFT expression and X^{HF} by means of an error function depending on a range separator μ as shown in eq 2

$$\frac{1}{r_{12}} = \frac{\mathrm{erf}(\mu r_{12})}{r_{12}} + \frac{\mathrm{erfc}(\mu r_{12})}{r_{12}} \quad (2)$$

dx.doi.org/10.1021/ct400452f | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

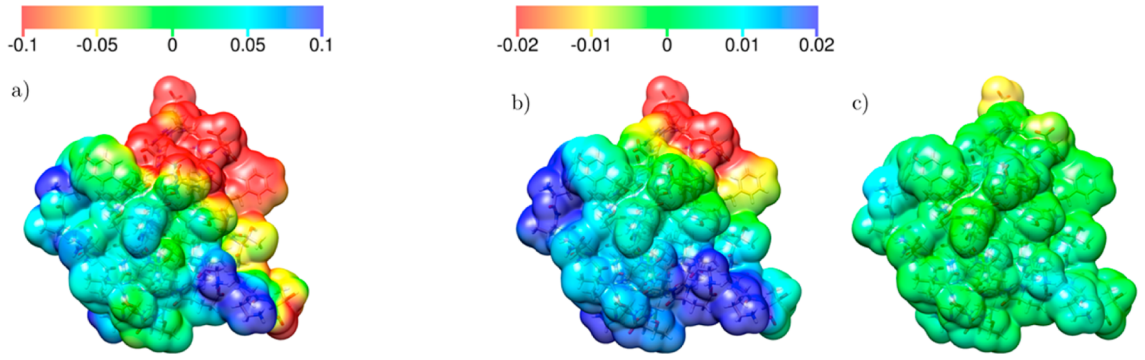


Figure 3. ESP for the insulin monomer plotted on a HF isodensity surface corresponding to $\rho = 1.0 \times 10^{-4}$ au. (a) Total ESP for MP2/cc-pVDZ. (b) ESP difference between MP2 and B3LYP (MP2 − B3LYP). (c) ESP difference between MP2 and CAM-B3LYP (MP2 − CAM-B3LYP).

Here, erf and erfc are the error function and the complementary error function, respectively. A large value of μ corresponds to a large amount of X^{HF} for a given electron−electron distance, r_{12}.

Figure 1a shows the total ESP for MP2 with the cc-pVDZ basis set for Gly_5, while 1b shows the difference in ESP between the MP2 and B3LYP results as a function of peptide length (note the difference in the scale used for the coloring in Figures 1a and 1b). The MP2 and B3LYP results are very similar for Gly_2, but the deviation rapidly increases as the distance between the two charge centers increases. The B3LYP method is incapable of binding a whole electron at the anionic site, and as the distance to the cationic site increases, this leads to a situation where density corresponding to approximately half of an electron is moved to the cationic site. As a consequence, the B3LYP ESP is not sufficiently negative at the $-CO_2^-$ group, and not sufficiently positive at the $-NH_3^+$ group.

The differences in the calculated ESP over the surface points can be quantified by box-plots, where horizontal lines indicate the minimum and maximum values, while the central box represents the central 50% of the data points. The box-plots in Figure 2 present the deviations from the MP2/cc-pVDZ result for Gly_5 using 17 different DFT functionals, based on ∼4,600 points sampled in the density range from 8×10^{-5} to 10^{-4} au, that is, right outside the surface illustrated in Figure 1. The ESP scale is in atomic units, where a value of 0.01 corresponds to an interaction energy of 26.25 kJ/mol with a test particle having a unit charge. The corresponding mean absolute deviations (MAD) and maximum absolute deviations (MaxAD) are shown in Table 1. Figure 2 clearly shows that the error depends on the amount of exact exchange in the method, such that pure GGA and meta-GGA DFT functionals like BLYP,[25] M06-L,[26] N12,[27] MN12-L,[28] and TPSS[29] display the largest deviation, while hybrid functionals improve the results, in particular functionals including a large fraction of HF exchange like BHHLYP[6] (50%), M06-2X[26] (54%), and M06-HF[30] (100%). The range-separated functionals, M11,[31] ωB97X,[32] and CAM-B3LYP[8] with $\mu = 0.33$ a_0^{-1} (standard value) and $\mu = 0.50$ a_0^{-1} (modified value included for comparison), also give accurate descriptions of the ESP, while a smaller range separator of $\mu = 0.11$ a_0^{-1}, which is used in N12-SX, MN12-SX,[33] and a second modified CAM-B3LYP functional, give inferior results. The two rightmost plots in Figure 2 show that changes due to improving the basis set to cc-pVTZ are very minor at both the MP2 and CAM-B3LYP levels compared to errors encountered for DFT functionals with small amounts of HF exchange or without long-range corrections.

BHHLYP is in good agreement with MP2 for all poly glycine peptides (results are only shown for Gly_5). For the present systems, the BHHLYP method contains sufficient HF exchange (50%) to solve the problem of fractional electron transfer, but it has been shown to be problematic for other systems with less strongly bound electrons.[10a] The HF method provides a qualitative correct charge separation for all glycine chains, but the lack of electron correlation leads to a slightly larger deviation relative to the MP2 results.

The incorrect electron density observed for functionals without large amount of HF exchange or without range-separation is a consequence of the self-interaction error in DFT, caused by an incomplete cancellation of the one-electron Coulomb self-repulsion by the exchange functional. Furthermore, the DFT exchange functional has an incorrect distance dependency, which increases the self-interaction errors at long ranges. In HF theory, on the other hand, the exchange energy cancels exactly the Coulomb contribution from a single electron and the potential decays correctly as $-r^{-1}$. Range-separated DFT functionals model the long-range exchange energy by the HF expression, and this effectively solves the problem with fractional electron transfer, provided that the range-separator, μ, is sufficiently large. It should be noted that CAM-B3LYP is not superior to B3LYP for all applications, and for example performs poorly for excited triplet states.[34]

ESP for Insulin. Insulin is a 51 residue peptide containing five cationic and six anionic residues and is used as a representative biomolecule of a realistic size. The distance between the charged residues varies from 5 to 32 Å, and based on the poly glycine results in Figures 1 and 2, it is expected that conventional DFT methods will give an incorrect charge density and ESP for this system as well.

Figure 3a shows the total ESP at the MP2/cc-pVDZ level. Based on the results for the poly glycines we expect that the MP2/cc-pVDZ ESP for the insulin monomer is converged to $\sim 10^{-3}$ au with respect to the basis set. Figures 3b and 3c show MP2 deviations in ESP for B3LYP and CAM-B3LYP (note the difference in the scale used for the coloring in Figures 3a and 3b,3c), and the data is quantified in the box-plots in Figure 4. The B3LYP deviations are correlated with the absolute values of the MP2 ESP, such that regions with negative ESP are not negative enough, and vice versa for positive regions. B3LYP significantly underestimates the charge separations because of artificial charge transfer from anionic to cationic sites, which is completely analogous to the situation for the poly glycines in Figure 1. The CAM-B3LYP method provides a description where the anionic and cationic sites retain a full negative or

dx.doi.org/10.1021/ct400452f | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

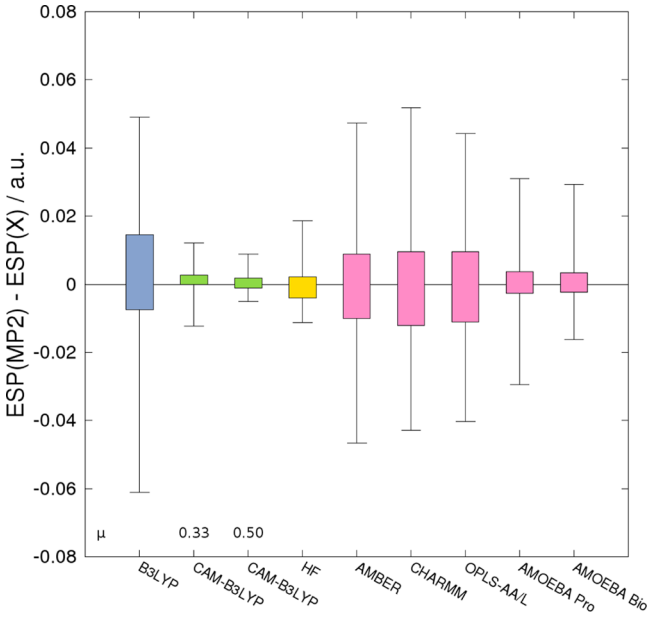


Figure 4. Box-plots showing the ESP difference relative to the MP2/cc-pVDZ reference for insulin based on 43,415 points in the density range $[8 \times 10^{-5}-10^{-4}]$ au.

positive charge, and thus gives a significantly more accurate ESP, as seen in Figure 3c, although there are still differences between the MP2 and CAM-B3LYP ESPs near some of the charged groups. These differences are small compared to the MP2-B3LYP differences, and can be further reduced by increasing the range-separation parameter from the default value of 0.33 a_0^{-1} to 0.50 a_0^{-1} (see Figure 4), where the latter value introduces more HF exchange for a given electron−electron distance. The HF result is closer to the MP2 reference than B3LYP, with errors in the same order as those observed for the poly glycine peptides.

ESP for Insulin Using Force Fields Methods. The simulation of large biomolecular systems relies on a FF description for the energy surface and thus the intra- and intermolecular interactions. In classical force fields, like AMBER,[17] CHARMM,[14] and OPLS,[12] the long-range electrostatic interaction is described by Coulomb interaction between partial atomic charges, while modern methods like AMOEBA[19] include higher order permanent multipole moments as well as atomic dipole polarization. Other methods used for implementing polarization include charge-on-a-spring (e.g., PIPF-CHARMM[35] and the polarizable CHARMM water model SWM-4NDP[36]), and the fluctuating charge model (e.g., UFF[37] and an OPLS-AA based FF developed in Friesner's group[38]). These models describe the permanent ESP using point charges only, and the polarization is unlikely to be significantly better than the induced dipole model in AMOEBA. The electrostatic parameters are often derived by fitting to a molecular ESP for small representative systems and assuming transferability. Having available a high quality reference ESP for realistic sized biomolecules allows a calibration of the FF results, and thus assessing the accuracy of the electrostatic interaction in simulations.

The ESP generated from a fixed charge FF has two shortcomings. First, only one point charge per atom is used, and regardless of how these charges are chosen, they are incapable of reproducing the ESP accurately.[39] Second, a fixed set of point charges cannot account for the geometry dependence of the ESP, and thus for differences between the system for which the charges are derived and the actual system. The AMOEBA FF includes both higher order electric moments and polarization and is capable of compensating for both deficiencies.

From the box-plot in Figure 4 it is evident that there is little difference in the accuracy of the fixed partial charge FFs (OPLS, AMBER, CHARMM), while the two versions of the polarizable AMOEBA FF perform significantly better. Using the quartile levels (upper and lower box limits) as representative errors, the fixed partial charge force fields have typical errors of ∼0.01 au at a van der Waals distance, which translates into errors in the interaction energies of ∼13 kJ/mol for a (typical) partial charge of 0.5 e. The AMOEBA FFs reduce this error to ∼4 kJ/mol.

Fixed charge FFs are parametrized for the condensed phase and mimic the interactions under such conditions. The partial charges, consequently, contain an integrated mean polarization corresponding to aqueous solvation. This polarization results in increased partial charges compared to the gas phase,[40] and therefore also increased values of the ESP.

Figure 5 shows how the ESP error relative to the MP2 result depends on the actual value of the potential for a fixed charge and a polarizable force field (AMBER and AMOEBA Bio09). The corresponding plots for OPLS and CHARMM are very similar to the one for AMBER. A correlation between the value of the potential and the error (MP2-FF) is observed for the

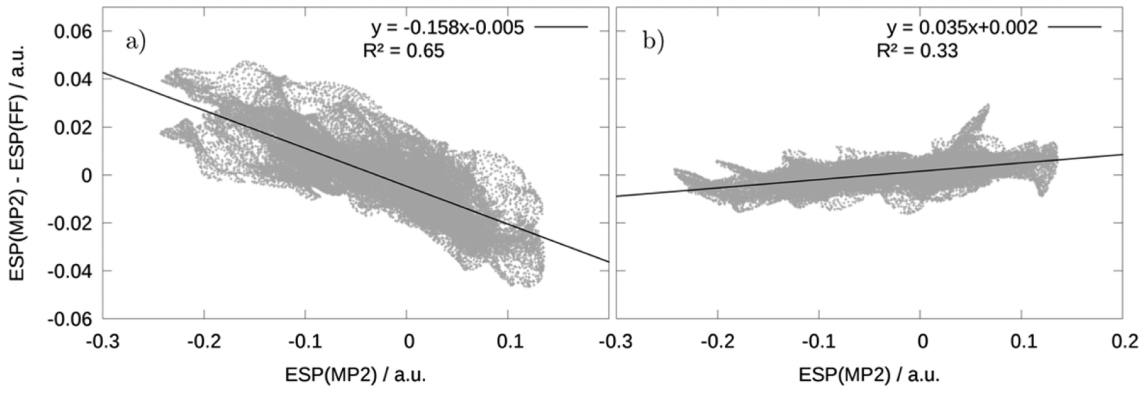


Figure 5. Correlation between the ESP error ($ESP_{MP2} - ESP_{FF}$) and the MP2/cc-pVDZ reference ESP for insulin based on 43,415 points sampled in the density range $[8 \times 10^{-5}-10^{-4}]$ au. (a) AMBER99sb. (b) AMOEBA Bio09.

dx.doi.org/10.1021/ct400452f | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

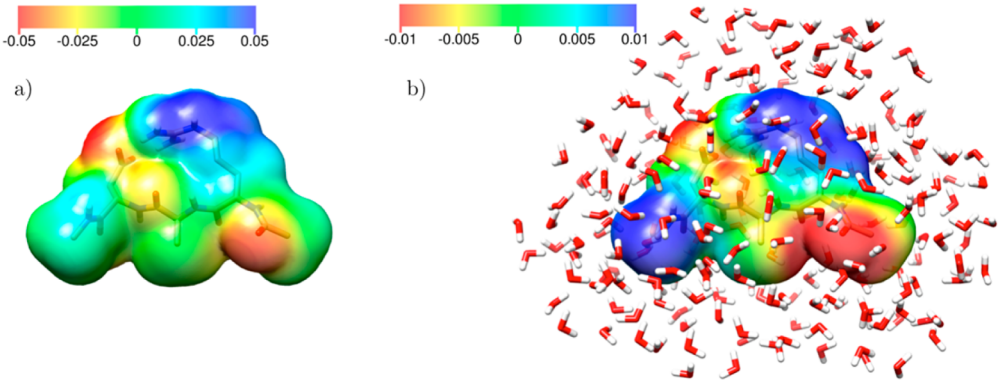


Figure 6. ESP for arg-ala-asp plotted on a HF gas phase isodensity surface corresponding to $\rho = 1.0 \times 10^{-4}$ au. (a) CAM-B3LYP/cc-pVDZ gas phase ESP. (b) CAM-B3LYP/cc-pVDZ ESP difference between solvent and gas phase (solvent − gas).

fixed charge FFs; fitting a linear equation to the data points results in a line with slope −0.16, showing that the FF overestimates the absolute value of the ESP. This correlation between the ESP error and the reference value is a direct consequence of the overpolarized partial charges. The result is that fixed charge force fields cause largest errors in regions of charged or partial charged functional groups. The error for the AMOEBA FF, on the other hand, does not appear to have any systematic trend, and a linear fit gives a slightly positive slope of +0.04.

Effects of Solvation. The fixed charge FFs attempt to model the solvent polarization by enhancing the gas phase polarity, while polarizable FFs should account for the solvation effects explicitly. In this section we discuss the ESP for the arg-ala-asp tripeptide solvated by water. This system contains two charged residues and has an appropriate size for explicit solvation. The role of the water is to polarize the peptide and does not contribute directly to the calculated ESP.

For this system the reference ESP is calculated using a QM/MM method, as discussed in the Computational Details section. For the insulin monomer and the poly glycines the CAM-B3LYP ESP is very close to the MP2 result (Figures 2, 3, and 4), and CAM-B3LYP with the cc-pVDZ basis will consequently be used for the QM region containing the peptide. The water molecules constitute the MM region and will be described by a standard DFT-parametrized water model.[23]

To understand the effect of solvation, the ESP from the solvated peptide based on the EFP/CAM-B3LYP calculation is compared to the ESP from the CAM-B3LYP gas phase calculation. Figure 6a shows the total potential in the gas phase, with the difference between the solvated and gas phase systems shown in Figure 6b (note the difference in the color scale in Figures 6a and 6b). Based on sampling in ∼7,000 points in the density range from 8×10^{-5} to 10^{-4} au, the absolute value of the potential increases on average by 32% in solvation compared to the gas phase, and the corresponding MAD is 0.006 au This is significantly larger than the difference between the MP2 and CAM-B3LYP calculated gas phase potentials, having an MAD of 0.0008 au for the tripeptide, and the CAM-B3LYP ESP should thus provide an adequate reference for solvated systems.

The FF ESP for the solvated peptide is compared to the EFP/CAM-B3LYP ESP in the box-plots in Figure 7 (blue bars), along with the gas phase results (red bars). The fixed charge FFs (AMBER, CHARMM, and OPLS) ignore many-body

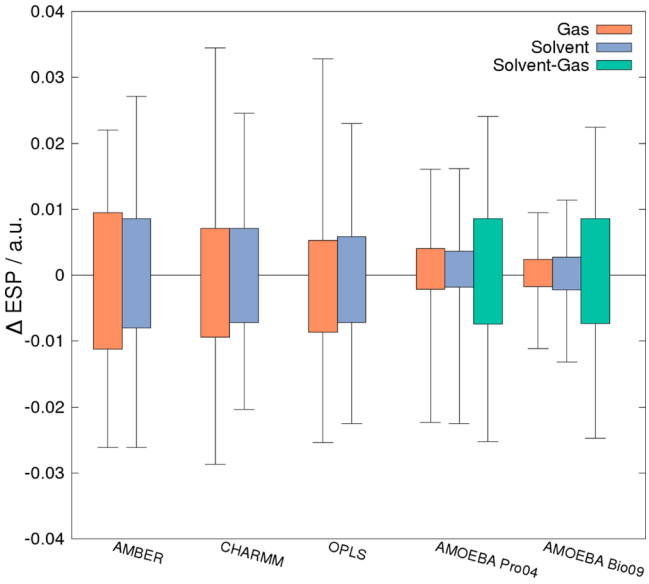


Figure 7. ESP differences for the arg-ala-asp tripeptide in solvent and gas phase based on 6,968 points in the density range $[8 \times 10^{-5}–10^{-4}]$ au. "Gas" shows the ESP difference between CAMB3LYP/cc-pVDZ and FF in the gas phase. "Solvent" shows the ESP difference between EFP/CAMB3LYP/cc-pVDZ and FF for the solvent polarized peptide. "Solvent-Gas" shows the ESP difference between FF calculations of the solvated and the gas phase peptide.

electrostatic effects, and as a result, the ESP is independent of the surrounding water molecules. For the polarizable AMOEBA FF, on the other hand, water contributes indirectly by polarization. All induced dipoles (for solvent and peptide atoms) are determined iteratively, but the potential is calculated as a sum over permanent and induced *peptide* multipoles only.

In general, the sizes of the boxes in Figure 7 decrease for all three fixed charge FFs in solvent compared to gas phase, indicating that the FF ESP represents the potential from the solvated system better than the gas phase. The error reductions, however, are relatively small compared to the large absolute errors, and the errors are furthermore unsystematic and independent of the ESP, which is opposite to the gas phase results shown in Figure 5. Table 2 shows the corresponding MAD values, which are consistent with the box-plots. For the AMOEBA FF there is no significant difference between the errors in the gas and solvated phases.

dx.doi.org/10.1021/ct400452f | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 2. MAD Values for ESP Calculated by FF Methods Compared to the CAMB3LYP/cc-pVDZ (Gas Phase) and EFP/CAMB3LYP/cc-pVDZ (Solvent) Results, Respectively[a]

	AMBER	CHARMM	OPLS	AMOEBA Pro04	AMOEBA Bio09
gas phase	0.010	0.010	0.009	0.004	0.003
solvent	0.008	0.008	0.008	0.004	0.003

[a]All units are au.

The description of solvation by the AMOEBA FF can be quantified by comparing the gas phase and solvent results, and this is shown in Figure 7 as green bars. The MAD value for this difference is 0.008 au for both versions of the FF, which is close to the CAM-B3LYP gas phase vs solvent MAD (0.006 au). The change in potential due to solvation is of the same order as the MAD for fixed charge FFs in solvent phase, and significantly larger than the errors for the AMOEBA FF. This result—combined with the fact that the AMOEBA MADs for gas and solvated phases are very similar—demonstrates that the AMOEBA FF is able to account for the polarizability due to solvation in an accurate fashion, and thus provides a balanced description of both gas- and solvation phases.

In summary, the polarizable AMOEBA FF is able to reduce the gas phase ESP errors significantly compared to fixed charge FFs (Figure 4 for insulin and the red bars of Figure 7 for the tripeptide), and this result is reproduced for the solvated tripeptide, although the advantage over fixed charge FFs decreases compared to the gas phase. While fixed charge FFs partly account for solvent polarization they have inferior performances relative to FFs incorporating solvent effects by explicit polarization. Considering the large changes induced by solvation (green bars in Figure 7), the solvation results for the fixed charge FFs (blue bars in Figure 7) indicate that these methods only in a very qualitative sense account for solvation effects.

CONCLUSION

The MP2 ESP for model peptides and the insulin molecule has been used for assessing the accuracy of different density functional and force field methods. Commonly used DFT functionals, such as BLYP and B3LYP, incorrectly predict partial electron transfer from anionic to cationic sites such that regions with negative ESP are not negative enough and vice versa for positive regions. The largest errors are observed for pure DFT functionals, while increasing amounts of exact exchange systematically reduce the error. Range-separated density functionals, such as CAM-B3LYP, perform much better because of the elimination of self-interaction errors at long distances, as long as the range-separator value is not chosen too small. These results have large implications when using density functional methods for zwitterionic systems when the anionic and cationic sites are separated by more than ∼10 Å. This problem is even more severe for proteins containing multiple charged sites. The use of nonrange-separated density functionals is *strongly discouraged*, as the fundamentally incorrect description of the charge distribution invariably will lead to derived errors in energies and structures. For periodic systems and QM/MM studies with large QM regions the use of standard DFT functionals, especially pure functionals without exact exchange, may lead to erroneous results.

Fixed charge force fields display large errors while force fields employing atomic multipole moments and polarizabilities perform significantly better, both in the gas and solvated phases. Part of the gas phase error is caused by the fact that fixed charge force fields model the mean polarization. This, however, does not account for the complete error, and also for solvated systems it was found, that polarizable force fields reduce the error of the ESP significantly compared to fixed charge force fields.

AUTHOR INFORMATION

Corresponding Author

*E-mail: frj@chem.au.dk.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement no. 291371. Support from the Danish Center for Scientific Computation, Danish Natural Science Research Council, is gratefully acknowledged.

REFERENCES

(1) Stone, A. J. *The Theory of Intermolecular Forces*; Oxford University Press: Oxofrd, U.K., 2002; pp 36−48.

(2) (a) Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*; Wiley: Chichester, U.K., 2000; pp 142−191; (b) Jensen, F. *Introduction to Computational Chemistry*, 2nd ed.; Wiley: Chichester, U.K., 2007; pp 133−189.

(3) Dunning, J. T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007−1023.

(4) (a) Ziolkowski, M.; Jansik, B.; Kjaergaard, T.; Jorgensen, P. Linear scaling coupled cluster method with correlation energy based error control. *J. Chem. Phys.* **2010**, *133*, 014107. (b) Kristensen, K.; Ziółkowski, M.; Jansík, B.; Kjærgaard, T.; Jørgensen, P. A Locality Analysis of the Divide−Expand−Consolidate Coupled Cluster Amplitude Equations. *J. Chem. Theory Comput.* **2011**, *7*, 1677−1694. (c) Kristensen, K.; Hoyvik, I.-M.; Jansik, B.; Jorgensen, P.; Kjaergaard, T.; Reine, S.; Jakowski, J. MP2 energy and density for large molecular systems with internal error control using the Divide-Expand-Consolidate scheme. *Phys. Chem. Chem. Phys.* **2012**, *14*, 15706−15714.

(5) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(6) Becke, A. D. A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys.* **1993**, *98*, 1372−1377.

(7) (a) Iikura, H.; Tsuneda, T.; Yanai, T.; Hirao, K. A long-range correction scheme for generalized-gradient-approximation exchange functionals. *J. Chem. Phys.* **2001**, *115*, 3540−3544. (b) Leininger, T.; Stoll, H.; Werner, H.-J.; Savin, A. Combining long-range configuration interaction with short-range density functionals. *Chem. Phys. Lett.* **1997**, *275*, 151−160.

(8) Yanai, T.; Tew, D. P.; Handy, N. C. A new hybrid exchange−correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51−57.

(9) Rohrdanz, M. A.; Martins, K. M.; Herbert, J. M. A long-range-corrected density functional that performs well for both ground-state properties and time-dependent density functional theory excitation energies, including charge-transfer excited states. *J. Chem. Phys.* **2009**, *130*, 054112−054118.

(10) (a) Jensen, F. Describing Anions by Density Functional Theory: Fractional Electron Affinity. *J. Chem. Theory Comput.* **2010**, *6*, 2726−2735. (b) Vydrov, O. A.; Scuseria, G. E. Assessment of a long-range corrected hybrid functional. *J. Chem. Phys.* **2006**, *125*, 234109. (c) Ruzsinszky, A.; Perdew, J. P.; Csonka, G. I.; Vydrov, O. A.; Scuseria, G. E. Spurious fractional charge on dissociated atoms:

Pervasive and resilient self-interaction error of common density functionals. *J. Chem. Phys.* **2006**, *125*, 194112.
(11) Smith, G. D.; Pangborn, W. A.; Blessing, R. H. The structure of TSB-6 human insulin at 1.0Å resolution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2003**, *59*, 474−482.
(12) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(13) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(14) MacKerell, A. D.; Banavali, N.; Foloppe, N. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* **2000**, *56*, 257−265.
(15) LSDALTON a linear scaling molecular electronic structure program, Release Dalton2011, see http://daltonprogram.org (accessed 31.05.2013).
(16) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision D.01; Gaussian, Inc.: Wallingford, CT, 2009.
(17) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712−725.
(18) Mackerell, A. D.; Feig, M.; Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400−1415.
(19) Ponder, J. W.; Ren, P.; Wu, C. Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *J. Chem. Theory Comput.* **2011**, *7*, 3143−3161.
(20) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605−1612.
(21) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. An effective fragment method for modeling solvent effects in quantum mechanical calculations. *J. Chem. Phys.* **1996**, *105*, 1968−1986.
(22) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. General atomic and molecular electronic structure system. *J. Comput. Chem.* **1993**, *14*, 1347−1363.
(23) Adamovic, I.; Freitag, M. A.; Gordon, M. S. Density functional theory based effective fragment potential method. *J. Chem. Phys.* **2003**, *118*, 6725−6732.
(24) Tsiper, E. V.; Burke, K. Rules for minimal atomic multipole expansion of molecular fields. *J. Chem. Phys.* **2004**, *120*, 1153−1156.
(25) (a) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098−3100. (b) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785−789.
(26) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2007**, *120*, 215−241.
(27) Peverati, R.; Truhlar, D. G. Exchange-Correlation Functional with Good Accuracy for Both Structural and Energetic Properties while Depending Only on the Density and Its Gradient. *J. Chem. Theory Comput.* **2012**, *8*, 2310−2319.
(28) Peverati, R.; Truhlar, D. G. An improved and broadly accurate local approximation to the exchange-correlation density functional: The MN12-L functional for electronic structure calculations in chemistry and physics. *Phys. Chem. Chem. Phys.* **2012**, *14*, 13171−13174.
(29) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.
(30) Zhao, Y.; Truhlar, D. G. Density Functional for Spectroscopy: No Long-Range Self-Interaction Error, Good Performance for Rydberg and Charge-Transfer States, and Better Performance on Average than B3LYP for Ground States. *J. Phys. Chem. A* **2006**, *110*, 13126−13130.
(31) Peverati, R.; Truhlar, D. G. Improving the Accuracy of Hybrid Meta-GGA Density Functionals by Range Separation. *J. Phys. Chem. Lett.* **2011**, *2*, 2810−2817.
(32) Chai, J.-D.; Head-Gordon, M. Systematic optimization of long-range corrected hybrid density functionals. *J. Chem. Phys.* **2008**, *128*, 84106.
(33) Peverati, R.; Truhlar, D. G. Screened-exchange density functionals with broad accuracy for chemistry and solid-state physics. *Phys. Chem. Chem. Phys.* **2012**, *14*, 16187−16191.
(34) Leang, S. S.; Zahariev, F.; Gordon, M. S. Benchmarking the performance of time-dependent density functional methods. *J. Chem. Phys.* **2012**, *136*, 104101−104112.
(35) Xie, W.; Pu, J.; MacKerell, A. D.; Gao, J. Development of a Polarizable Intermolecular Potential Function (PIPF) for Liquid Amides and Alkanes. *J. Chem. Theory Comput.* **2007**, *3*, 1878−1889.
(36) (a) MacKerell, A. D.; Lamoureux, G.; Roux, B. A simple polarizable model of water based on classical Drude oscillators. *J. Chem. Phys.* **2003**, *119*, 5185. (b) Lamoureux, G.; Harder, E.; Vorobyov, I. V.; Roux, B.t.; A.D.M., Jr. A polarizable model of water for molecular dynamics simulations of biomolecules. *Chem. Phys. Lett.* **2006**, *418*, 245−249.
(37) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.
(38) Banks, J. L.; Kaminski, G. A.; Zhou, R.; Mainz, D. T.; Berne, B. J.; Friesner, R. A. Parametrizing a polarizable force field from ab initio data. I. The fluctuating point charge model. *J. Chem. Phys.* **1999**, *110*, 741−754.
(39) (a) Francl, M. M.; Carey, C.; Chirlian, L. E.; Gange, D. M. Charges fit to electrostatic potentials. II. Can atomic charges be unambiguously fit to electrostatic potentials? *J. Comput. Chem.* **1996**, *17*, 367−383. (b) Tan, J. S.; Boerrigter, S. X. M.; Scaringe, R. P.; Morris, K. R. Core-shell potential-derived point charges. *J. Comput. Chem.* **2012**, *33*, 950−957.
(40) Halgren, T. A.; Damm, W. Polarizable force fields. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236−242.