Point Charges Optimally Placed to Represent the Multipole Expansion of Charge Distributions

Ramu Anandakrishnan[1][9], Charles Baker[2][9], Saeed Izadi[3][9], Alexey V. Onufriev[1,4]*

1 Department of Computer Science, Virginia Tech, Blacksburg, Virginia, United States of America, 2 Graduate Program in Biophysics, Harvard University, Cambridge, Massachusetts, United States of America, 3 Department of Engineering Science and Mechanics, Virginia Tech, Blacksburg, Virginia, United States of America, 4 Department of Physics, Virginia Tech, Blacksburg, Virginia, United States of America

Abstract

We propose an approach for approximating electrostatic charge distributions with a small number of point charges to optimally represent the original charge distribution. By construction, the proposed optimal point charge approximation (OPCA) retains many of the useful properties of point multipole expansion, including the same far-field asymptotic behavior of the approximate potential. A general framework for numerically computing OPCA, for any given number of approximating charges, is described. We then derive a 2-charge practical point charge approximation, PPCA, which approximates the 2-charge OPCA via closed form analytical expressions, and test the PPCA on a set of charge distributions relevant to biomolecular modeling. We measure the accuracy of the new approximations as the RMS error in the electrostatic potential relative to that produced by the original charge distribution, at a distance $\hat{2}\times$ the extent of the charge distribution–the mid-field. The error for the 2-charge PPCA is found to be on average 23% smaller than that of optimally placed point dipole approximation, and comparable to that of the point quadrupole approximation. The standard deviation in RMS error for the 2-charge PPCA is 53% lower than that of the optimal point dipole approximation, and comparable to that of the point quadrupole approximation. We also calculate the 3-charge OPCA for representing the gas phase quantum mechanical charge distribution of a water molecule. The electrostatic potential calculated by the 3-charge OPCA for water, in the mid-field (2.8 Å from the oxygen atom), is on average 33.3% more accurate than the potential due to the point multipole expansion up to the octupole order. Compared to a 3 point charge approximation in which the charges are placed on the atom centers, the 3-charge OPCA is seven times more accurate, by RMS error. The maximum error at the oxygen-Na distance (2.23 Å) is half that of the point multipole expansion up to the octupole order.

Competing Interests: The authors have declared that no competing interests exist

* E-mail: alexey@cs.vt.edu

[9] These authors contributed equally to this work.

Introduction

Point multipole expansions are widely used to gain physical insight by providing a simplified expression for a complex distribution of sources of potential fields, such as electrostatic potential due to a charge distribution. Many familiar physics concepts are introduced using the framework of point multipoles because point multipoles provide a means of decoupling the underlying features of a source distribution from the observation point. Furthermore, since each successive term in the multipole expansion decays more rapidly with distance than the previous term, the impact of high order terms becomes small in the far-field, i.e. at distances R such that $R \gg R_0$, where R_0 is the distance of the furthest charge from the expansion center. This property has allowed the point multipole expansion to simplify many practical calculations. For example, algorithms such as the fast multipole [1], local reaction field [2], and fitted point multipole (FPM) [3] methods, use point multipoles to reduce the computational complexity of calculating pairwise interactions between large charge distributions. However, at distances not much larger than R_0, the accuracy of the low order point multipole approximation deteriorates quickly as one approaches the charge distribution, necessitating introduction of higher order terms. This, in turn, may lead to cumbersome algebra and the need to introduce further approximation [4]. Since, in practice, the potential often needs to be calculated in regions where the assumption $R \gg R_0$ does not hold, the point multipole expansion with only one or two lowest order terms may be suboptimal for some practical calculations. For example, in atomistic molecular simulations, amino acids interacting inside a single protein are often only several (1–5) times R_0 apart. For a typical amino acid group within a folded protein such as lysozyme $R_0 \approx 5$ Å, and the distance between amino acid groups ranges from $1R_0$ to $10R_0$. The value of R_0 and the distance at which the potential due to the charge distribution is evaluated, is, of course, problem dependent. Furthermore, compared to point charge approximations, it is generally more difficult to implement point multipole approximations into existing molecular modeling software, especially for commonly used implicit solvent models.

Arguably, one of the most successful point multipole based approximations is the fast multipole method [1]. The fast multipole method partitions the system into a hierarchical set of cubic lattices. Electrostatic interactions between charges within a lattice and in neighboring lattices (the near-field) are treated exactly, while a truncated multipole expansion is used for electrostatic interactions due to atoms in the more distant lattices (the mid- and far-field). The size of the lattices used in the multipole expansion varies, with a larger lattice size being used for more distant lattices [5–7]. This technique reduces the complexity of the computation of pairwise interaction from $O(N^2)$ to less than $O(N\log(N))$, where N is the number of interacting particles [1]. Many improvements to the original technique have been made [6,8–11]. Overall, the fast multipole method has the advantage of lower computational complexity compared to the full pairwise computation, and has a well defined error bound [1]; the method is used in many areas of physics. However the fast multipole method has not been widely adopted in biomolecular simulations, most likely due to its algorithmic complexity, and the discontinuities in calculated potential inherent in the method. [12,13] The local reaction field [2] and fitted point multipole (FPM) [3] methods have also not been widely utilized in the context of biomolecular modeling, again most likely due to their algorithmic complexity.

Here we investigate an alternative to the point multipole expansion for approximating charge distributions, which we call optimal point charge approximation (OPCA). Unlike the fast multipole method, which uses a set of point multipoles to represent the original charge distribution, the OPCA approximates a charge distribution using a given number of point charges. These point charges are chosen so that they optimally reproduce the lowest order multipoles in the expansion of the original distribution. Since OPCAs have a finite size, as opposed to being point-like, they may provide better representation of the original spatially extended charge distribution than a single-center truncated point multipole expansion. In particular, a more accurate representation of the potential in the mid-field may be expected. We prove below that the 1-charge and 2-charge OPCAs are at least as accurate as the equivalent order point multipole approximations, i.e. the point monopole and dipole approximations. Throughout this work we refer to point monopole, dipole, and quadrupole approximations as the truncated point multipole expansions upto the monopole, dipole, and quadrupole order, respectively.

We show that in general it is always possible to numerically determine the OPCA, however, for many practical applications, such as molecular dynamics simulations, analytical expressions are needed. Although it is not always possible to derive a practical analytical expressions for OPCAs, in certain cases we show that reasonable, robust and fairly simple approximations to the OPCAs can be derived, which we refer to as the practical point charge approximation (PPCA). The 2-charge OPCA is one such case for which a practical analytical expression is not readily evident for arbitrary charge distributions. For this case we have derived an approximation to the OPCA, the 2-charge PPCA. In what follows we evaluate the accuracy of our approximations for a set of charge distributions relevant to biomolecular modeling at a distance $R \approx 2R_0$. The accuracy at such distances is most critical for multiscale approximations [14,15] such as the hierarchical charge partitioning method [16]. For smaller distances, multiscale approximations generally use the exact charge distribution in their computations. From a practical standpoint, PPCAs may also be easier than the fast multipole protocol to implement in applications that already utilize point charges, i.e. in many molecular dynamics packages [17,18].

The rest of this work is organized as follows. We first review the multipole expansion concept to orient the reader and provide a convenient notational reference. Next, we describe the theoretical basis for the optimal point charge approximation. We then use this theoretical formalism to derive closed-form expressions for the optimal and the practical point charge approximations for the 1- and 2-charge cases. The accuracy of the 2-charge PPCA was evaluated for a practical application relevant to biomolecular modeling. We also calculated the 3-charge OPCA for approximating a quantum mechanical charge distribution for a water molecule; the resulting OPCA reproduces the electrostatic potential in the mid-field with greater accuracy than the point octupole expansion. Potential uses and future work are discussed in "Conclusions".

Methods

Multipole Expansion

Here we will give a brief overview of the formalism of the point multipole expansion. Since many practical applications, such as molecular dynamics simulations, use point charges, for notational simplicity we will consider discrete charge distributions, but our main results also hold for continuous distributions.

Consider a set of N point charges q_n ($n = 1, 2, ..., N$) located at positions r_n around some chosen origin. Then the potential $\Phi(R)$, of this distribution at a point R from that origin is given by the familiar Coulomb potential

$$\Phi(R) = \frac{1}{4\pi\epsilon_0} \sum_{n=1}^{N} \frac{q_n}{\|R - r_n\|} \tag{1}$$

For distances $R > R_0$ where $R = \|R\|$ and $R_0 = max(\|r_n\|)$, a Taylor series expansion of the potential above gives the classic multipole expansion. In Cartesian coordinates we obtain

$$\Phi(R) = \frac{1}{4\pi\epsilon_0} \left(\frac{1}{R} q + \frac{1}{R^2} \sum_{i=x,y,z} \hat{R}_i p_i + \frac{1}{R^3} \sum_{i,j=x,y,z} \hat{R}_i \hat{R}_j Q_{ij} \right.$$
$$\left. + \frac{1}{6} \frac{1}{R^4} \sum_{i,j,k=x,y,z} \hat{R}_i \hat{R}_j \hat{R}_k O_{ijk} + \dots \right). \tag{2}$$

where

$$q = \sum_{n=1}^{N} q_n \tag{3}$$

$$p_i = \sum_{n=1}^{N} q_n r_{n,i} \tag{4}$$

$$Q_{i,j} = \frac{1}{2} \sum_{n=1}^{N} q_n (3 r_{n,i} r_{n,j} - (r_n)^2 \delta_{ij}) \tag{5}$$

$$O_{i,j,k} = \sum_{n=1}^{N} q_n(15 r_{n,i} r_{n,j} r_{n,k} - 3(r_n)^2 (r_{n,i}\delta_{jk} + r_{n,j}\delta_{ik} + r_{n,k}\delta_{i,j})) \quad (6)$$

q,p,Q,O are known as the monopole, dipole, quadrupole and octupole moments respectively, $\hat{R}_i, \hat{R}_j, \hat{R}_k$, with $i,j,k = x,y,z$, are the unit vectors along the $x,y,$ or z coordinates, and δ_{ij} is the Kronecker delta. The multipole moments are symmetric tensors where the lowest order non-vanishing multipole is origin independent.

The Optimal Point Charge Approximation (OPCA)

For a given set of N original charges q_n ($n = 1,2,...,N$), we want to determine the position and magnitude of K point charges $\bar{q}_k$ ($k = 1,2,...,K < N$) such that the potential due to these smaller number of point charges, $\bar{\Phi}(R)$ best approximates the potential of the original distribution, $\Phi(R)$. Our criterion for the "best approximation" is as follows: the optimal point charge approximation (OPCA) minimizes the error in the multipole expansion for the K point charges relative to the multipole expansion for the original distribution of N charges. The precise error metric is defined below.

The error metric. Determining the best representative charge distribution is contingent upon the definition of the error metric used. In general, we are concerned with obtaining the best representation of the original potential at any arbitrary point in space outside the distribution. Thus, for the error metric, Δ, one typically chooses the root mean square (RMS) of the error in potential over some volume V (or surface) excluding the volume V_0 containing the charge distribution being approximated, i.e.

$$\Delta^2 = \frac{1}{V \notin V_0} \int_{V \notin V_0} |\Phi(R) - \bar{\Phi}(R)|^2 dV \quad (7)$$

In principle, one can derive the optimal charge placement $\{\bar{q}_k, \bar{r}_k\}$ by minimizing the integral given in Eq. (7) with respect to the values of the new charges, $\{\bar{q}_k\}$ and their positions $\{\bar{r}_k\}$. However, as the number of charges in the representative distribution grows, this equation can be expensive to minimize numerically, let alone to find closed-form analytic expressions for the placement and magnitude of the charges composing the representative distribution. In addition, the choice of the integration volume is somewhat arbitrary. Furthermore – and, perhaps, most importantly – charges chosen in this manner are not guaranteed to have the same multipole moments as the original distribution [19]. This can lead to misinterpretation of the properties of the distribution and, potentially, to unphysical results. At the very least, we would like the new approximate representation to inherit the same asymptotic behavior of the corresponding point multipole expansion of the same order, but with greater accuracy expected from a spatially extended distribution that can better mimic the original charge distribution.

To simplify the problem, we recast Eq. (7) in spherical coordinates and consider the error inside a spherical shell centered on the chosen multipole expansion center, and with arbitrary outer radius $\tilde{R} > R_0$, where R_0 is defined as before, i.e. the distance from the expansion center to the outermost point charge. The error metric now becomes

$$\Delta^2 = \frac{3}{4\pi(\tilde{R}^3 - R_0^3)} \int_{R_0}^{\tilde{R}} \int_0^{2\pi} \int_0^{\pi} |\Phi(R) - \bar{\Phi}(R)|^2 R^2 \sin(\theta) d\theta d\phi dR \quad (8)$$

where θ and ϕ are the usual spherical coordinate inclination and azimuth angles.

In spherical coordinates, the multipole expansion is given by

$$\Phi(R) = \frac{1}{\epsilon_0} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{m=\ell} \frac{1}{2\ell+1} \frac{Y_\ell^m(\theta,\phi)}{R^{\ell+1}} q_\ell^m \quad (9)$$

$$q_\ell^m = \sum_{n=1}^{N} q_n r_n^\ell Y_\ell^m(\theta_n,\phi_n)^* \quad (10)$$

where Y_ℓ^m are the standard spherical harmonics, $*$ denotes the complex conjugate, q_ℓ^m are the spherical multipole moments, and ℓ is the multipole order. Using this expansion as our error metric, Eq. (8), becomes

$$\Delta^2 = \frac{3}{4\pi(\tilde{R}^3 - R_0^3)}$$

$$\int_{R_0}^{\tilde{R}} \int_0^{2\pi} \int_0^{\pi} \left| \frac{1}{\epsilon_0} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{m=\ell} \frac{1}{2\ell+1} \frac{Y_\ell^m(\theta,\phi)}{R^{\ell+1}} (q_\ell^m - \overline{q_\ell^m}) \right|^2 R^2 \sin(\theta) d\theta d\phi dR \quad (11)$$

where q_ℓ^m and $\overline{q_\ell^m}$ are the spherical moments of the original and representative charge distributions respectively. Since the spherical harmonics are orthonormal, Eq. (11), can be further simplified to the following form [20],

$$\Delta^2 = \frac{3}{4\pi\epsilon_0(\tilde{R}^3 - R_0^3)} \int_{R_0}^{\tilde{R}} \sum_{\ell=0}^{\infty} \frac{1}{(2\ell+1)^2 R^{2\ell}} \sum_{m=-\ell}^{m=\ell} |q_\ell^m - \overline{q_\ell^m}|^2 dR \quad (12)$$

Calculating the optimal point charge approximation. The position and magnitude of the representative point charges in OPCA for a given order K are calculated by sequentially minimizing each term in the error expansion Eq. (12), starting with the lowest order (monopole) term. From the structure of Eq. (12) we see that minimizing the difference between the successive multipole moments of the original N-charge and the optimal K-charge distributions is equivalent to minimizing the total error in electrostatic potential. Note that the procedure does not depend on the parameter R, and thus the method does not require explicit integration over a given region. This removes a degree of arbitrariness in defining the "error surface" inherent in several other methods currently used in practice. The use of the multipole expansion as reference allows for the sought after distinct separation of terms by the rate at which they decrease as a function of R, i.e. the monopole term falls off as $\frac{1}{R}$, the dipole falls off as $\frac{1}{R^2}$, etc. A representation that makes terms up to order l in Eq. (12) equal to zero will produce total error whose leading term falls off as $1/R^{l+1}$. For $K = N$, which is the number of charges in the original distribution, the OPCA exactly reproduces the

electrostatic potential due to the original distribution. This is in contrast to the point multipole expansions, which generally require an infinite number of terms to exactly reproduce a given charge distribution.

Note that minimizing the error metric, Eq. (12), minimizes the error in electrostatic potential for the far-field where $R \gg R_0$, but not necessarily in the mid-field. In the mid-field ($R \approx 2R_0$), the contribution due to higher order terms in the multipole expansion can, in principle, be greater than the contribution due to lower order terms. Therefore, minimizing the lowest order terms in the expansion error does not guarantee minimization of the total error in the mid-field: these errors are investigated below for charge distributions most relevant to biomolecules. In the following analysis, for convenience, we have dropped the $1/4\pi_0$ factor in Eq. (2) and (12) and switched from SI to atomic units.

Analytical Expressions for 1- and 2-charge OPCAs

The minimization of the error metric in Eq. (12), which is required to find an OPCA, can be done numerically for any given number of K representative point charges. A numerical procedure for calculating the OPCA representation may be particularly useful in situations where the charge distributions are relatively static and thus the optimal representation does not need to be recalculated. For example, during restrained molecular dynamics simulations, components of the molecule may not move. The OPCA for these components do not need to be recalculated. For applications where the OPCA needs to be recalculated frequently, such as in unrestrained molecular dynamics simulations, one would like to find closed-form analytical expressions that can be used to compute OPCAs at a reduced computational cost and provide derivatives for force calculations.

In the following sections, we apply the general framework developed above to derive simple analytical expressions for the 1-charge OPCA. Note that the 1-charge OPCA is only applicable to a charge distribution with a non-zero net charge, since the monopole moment for a neutral charge distribution is zero. The 2-charge case is more complex: the optimal point charge approximation may result in imaginary charge values for some charge distributions (see Eq. (22) below), and can not be cast in a closed-form formula for some other distributions. Therefore, we derive more practical analytical expressions that approximate the 2-charge OPCA with a reasonable accuracy.

1-charge Optimal Point Charge Approximation for Charged Distributions

By definition, the 1-charge OPCA consists of a single charge. As long as the OPCA charge has magnitude $\bar{q} = \sum_{n=1}^{N} q_n$, i.e. is equal to the total charge of the original distribution, the monopole term of the error expansion, Eq. (12), will be zero. Now, the remaining parameter, namely the position of the charge, is chosen to minimize the dipole term in the error expansion. In this particular case, the dipole term can be made identically zero by solving

$$p_x - \bar{q} \cdot x = 0 \qquad (13)$$

$$p_y - \bar{q} \cdot y = 0 \qquad (14)$$

$$p_z - \bar{q} \cdot z = 0 \qquad (15)$$

for x,y,z where p_x, p_y, p_z are the x,y,z components of the dipole moment p of the original distribution. Solving the above equations we have

$$\bar{q} = q \qquad (16)$$

$$\bar{r} = \frac{p}{q} \qquad (17)$$

So, a charge of magnitude $\bar{q}$ placed at $\bar{r}$ (center of charge) defines the 1-charge OPCA.

2-charge Optimal and Practical Point Charge Approximations

The 1-charge OPCA is the smallest set of point charges required to eliminate the monopole and the dipole term of the error expansion in Eq. (12) for systems with non-zero net charge. However, an error reduction further than the dipole order is often desired for higher accuracy. In such cases, the 2-charge OPCA ($K=2$) is the next step. In deriving an analytical expression for the 2-charge OPCA, the goal is to eliminate the error terms up to the dipole order in Eq. (12), and to minimize the quadrupole and, ideally, the higher order terms.

Due to important differences in the characteristics of charge distributions with zero and non-zero net charges, it is necessary to treat these two cases separately. For charged systems, the monopole and the dipole error terms are eliminated if two point charges with total charge equal to the original net charge are positioned so that their center of charge coincides with the center of charge of the original distribution. For uncharged systems, however, the monopole and dipole terms in the error expansion are eliminated when a pair of charges of equal magnitude but opposite sign are aligned with the direction of the original dipole moment. In other words, to eliminate the error terms up to the dipole in the charged case the location of the center of charge of the 2 charges is constrained, while in the uncharged case the direction of the dipole moment of the 2 charges is constrained. This leads to two different solutions for the two cases.

2-charge approximation for uncharged distributions. For net zero charge distributions, the optimal point charge approximation consists of two charges $\bar{q}_1 = \bar{q}$ and $\bar{q}_2 = -\bar{q}$ located at positions $\bar{r}_1$ and $\bar{r}_2$ respectively. Thus, it takes 7 parameters, q, and the x,y,z components of $\bar{r}_1$, and $\bar{r}_2$, to uniquely define a 2-charge OPCA. By setting

$$\bar{q}(\bar{r}_1 - \bar{r}_2) = \sum_{n=1}^{N} q_n r_n \qquad (18)$$

the dipole term in the error is zero. Now, we will rewrite the positions $\bar{r}_1$ and $\bar{r}_2$ in the following form:

$$\bar{r}_1 = \bar{d} + \frac{\sum q_n r_n}{2\bar{q}}$$

$$\bar{r}_2 = \bar{d} - \frac{\sum q_n r_n}{2\bar{q}} \qquad (19)$$

where $\bar{\boldsymbol{d}}$ represents the geometric center between the two charges of the OPCA. We can see that these positions satisfy relation (18) automatically. By writing the positions of the charges in this manner, we have divided the process of determining the remaining parameters which define the OPCA into two steps, namely, finding the optimal placement of the charges, $\bar{\boldsymbol{d}}$, and finding the optimal magnitude of the charge, $\bar{q}$. Note that finding the optimal charge value fixes the separation between the two charges, since the dipole moment of the representative distribution has been constrained to equal the original dipole moment.

The placement of the geometric center $\bar{\boldsymbol{d}}$ of the charges composing the 2-charge OPCA that minimizes the quadrupole term of the error expansion, is given by

$$\bar{d}_k = \frac{2}{3p^2}\left(\sum_{i=x,y,z} Q_{ki}p_i - \left(\frac{\sum_{i,j=x,y,z} Q_{ji}p_ip_j}{4p^2}\right)p_k\right) \qquad (20)$$

where $k=x,y,z$, the $\bar{d}_k$'s are components of $\bar{d}$, p_i,p_j,p_k are the components of the dipole moment (Eq. (4)), and Q_{ki},Q_{kj} are the components of the quadrupole moment (Eq. (5)). This optimal position, known as the center of dipole, was derived previously [20] for a different purpose, namely for matching point multipole expansions between different charge distributions. Now, unlike the point dipole approximation, the 2-charge OPCA has physical size and thus an additional parameter with which to further minimize the error with respect to the given potential. In other words, Eqs. (20) and (18), determine only 6 of the 7 parameters required to define the 2-charge OPCA. Since the quadrupole moment is the lowest order non-zero term remaining in the error expansion, by choosing the optimal charge value we want to further minimize the quadrupole term in the error. However, for any charge value $\bar{q}$ an OPCA placed at the center of dipole has no quadrupole moment as can be seen by setting $N=2$, substituting the center of dipole, Eq. (20), and $q_1=-q_2=\bar{q}$ into Eq. (19), and then substituting these variables into Eq. (6). Thus, the quadrupole term in the error, Eq. (12) is unaffected by the choice of the charge magnitude $\bar{q}$, and the quadrupole term has already been globally minimized. Therefore, to uniquely define the charge $\bar{q}$, we follow the OPCA procedure and globally minimize the next term in the error expansion, namely the octupole term. Specifically, if we consider the $\ell=3$ term of Eq. (12), using the connection formula from spherical multipoles to Cartesian multipoles we can compute

$$\sum_{i,j,k=x,y,z} \frac{\partial}{\partial \bar{q}}\left(O_{ijk} - \overline{O_{ijk}}\right)^2 = 0 \qquad (21)$$

where O_{ijk} and $\overline{O_{ijk}}$ are components of the octupole moments, in Cartesian coordinates (Eq. (6)), of the original distribution and the 2-charge OPCA respectively, for an expansion computed about the center of dipole. By noting that $\overline{O_{ijk}}$ is a function of $\bar{q}$, we find that Eq. (21) is satisfied when $\bar{q}\to\infty$ or if the charge value is given by

$$\bar{q} = \sqrt{\frac{3p^6}{2\sum_{i,j,k=x,y,z} O_{ijk}p_ip_jp_k}} \qquad (22)$$

Thus, Eqs. (19), (20) and (22) define the 2-charge OPCA for the net zero charge case (figure 1), i.e. defines the best placement of charges such that the error metric (Eq. (12)) is minimized.

In some cases, it is possible that

$$\left(\sum_{i,j,k=x,y,z} O_{ijk}p_ip_jp_k\right) \leq 0 \qquad (23)$$

In this case, the charge given by Eq. (22) is imaginary, which is unphysical. This situation occurs when the orientation of the dipole with respect to the octupole moment of the original charge distribution is such that increasing the distance between the charges of the 2-charge OPCA always increases the error. In this case, Eq. (21) is formally satisfied only for $\bar{q}\to\infty$. In a practical calculation, a 2-charge OPCA with inequality (23) is constructed by fixing the separation between the charges $\|\bar{r}_1-\bar{r}_2\|$ to a small value (figure 2(a)), while increasing the OPCA charge accordingly to maintain the dipole moment of the original distribution (Eq. (18)). All of the OPCA charges have real values. If inequality (23) holds, the 2-charge OPCA does not offer an accuracy advantage in the far-field over the optimal point dipole approximation, however, the 2-charge OPCA can always mimic the point dipole approximation to arbitrary precision and thus the two distributions will produce equivalent error. Thus, even if inequality (23) holds, the 2-charge OPCA represents the optimal placement of two point charges and is at least as accurate as the point dipole approximation in the far-field where $R\gg R_0$. However, in the mid-field, such a charge placement may sometimes be slightly less accurate than the optimal point dipole approximation.

For the charge distributions described in the Results section below, we found that setting $\|\bar{r}_1-\bar{r}_2\|=R_0/4$ to determine the value of $\bar{q}$ in Eq. (18), instead of the much more complex Eq. (22), results in the electrostatic potential that is on average within 4% of the optimal $K=2$ OPCA solution (figure 2(b)). Therefore, for practical applications, it may be computationally more efficient to use an empirically determined value for $\|\bar{r}_1-\bar{r}_2\|$ even for cases where the inequality (23) is not satisfied and a true optimal placement of 2-charge can be found via Eqs. (20) and (22). Another important practical consideration is that the contribution of each term in the error expansion, Eq. (12), is smaller than the previous term only if $R\gg R_0$. This is not necessarily true in the all-important mid-field regime where $R\approx 2R_0$. For example, for some charge distributions, the center of dipole $\mathbf{d}$ may be located at $R\geq 2R_0$. For such cases, the error for the optimal point dipole approximation [20,23] (point dipole approximation placed at the center of dipole) and the 2-charge OPCA can become large at $R\approx 2R_0$. To ensure that our 2-charge approximation is reasonably accurate in the mid-field for such cases, we introduce an addition condition: the optimal charge positions are restricted to be within the 1.5 times the maximum extent of the original charge distribution R_0, from the center of geometry.

Thus, for distributions with zero net charge, the 2-charge practical point charge approximation (PPCA), which approximates the 2-charge optimal point charge approximation (OPCA), is determined through the following 4 steps: (i) The two point charges comprising the PPCA are placed such that their center of geometry coincides with the center of dipole (Eq. (20)). (ii) The separation between the charges is fixed at $\|\bar{r}_1-\bar{r}_2\|=R_0/4$. (iii) The position and magnitude of two charges are then determined by Eq. (18) and (19). (iv) In the rare cases when the point charges for the PPCA are at a distance greater than $1.5R_0$ from the center of geometry, the center of dipole for the PPCA is shifted towards the center of geometry for the original charge distribution so that the point charges lie within $1.5R_0$. The constants in conditions (ii) and (iv) above were determined empirically for charge
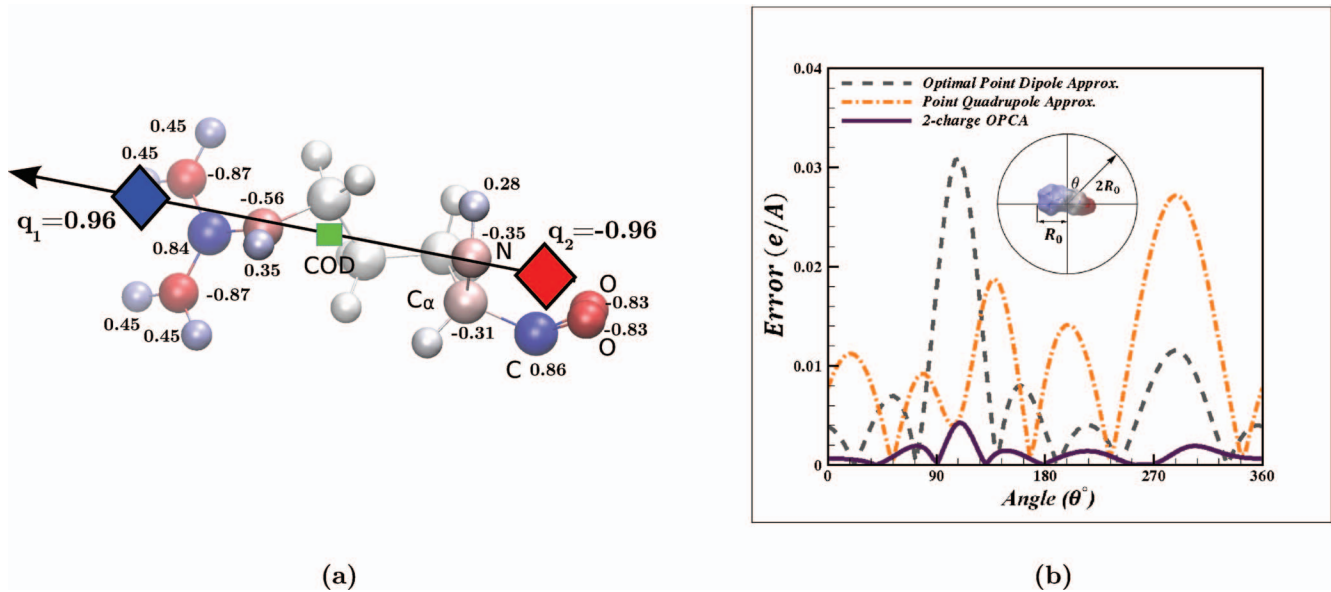
Figure 1. Example of a 2-charge optimal point charge approximation (OPCA). For a sample charge distribution–a neutral amino acid (C-terminal arginine at physiological pH), including the associated NH-CH-COO backbone atom. (**a**) The green square represents the center of dipole (COD), with dipole moment p shown by the arrow. The two diamonds represent the two point charges, q_1 and q_2, of the OPCA. The atomic partial charges are represented as spheres rendered using VMD [21]. The sphere colors range from red to blue representing the charge range of $-1e$ to $+1e$, where e is the atomic unit of charge. The charge values for partial charges $|q| > 0.2e$ are shown next to the atoms. As a visual reference, the backbone heavy atoms are labeled and covalent bonds are included in the figure. (**b**) Error in electrostatic potential for the 2-charge OPCA, point dipole, and point quadrupole with center of dipole as the expansion center. The error is calculated relative to the exact computation, on a circle at a distance $2R_0$, in the plane shown. Here, R_0 is the size of the charge distribution defined as the distance from its center of geometry to the outermost charge. The inset image shows the electrostatic surface potential rendered using GEM [22].
doi:10.1371/journal.pone.0067715.g001

distributions relevant to biomolecular modeling, see the Results section below.

2-charge approximation for charged distributions. The 2-charge OPCA consists of two charges $\bar{q}_1$ and $\bar{q}_2$. By setting

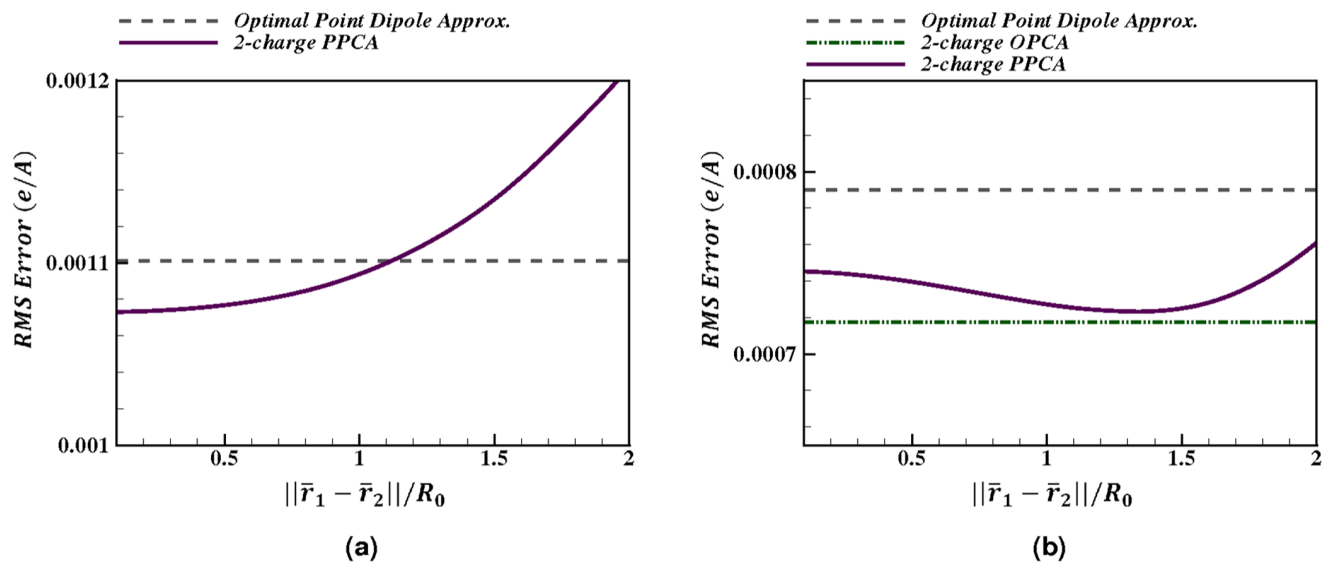$$q = \bar{q}_1 + \bar{q}_2 \qquad (24)$$

Figure 2. Accuracy of the 2-charge practical point charge approximation (PPCA) for charge distributions with a net zero charge. Accuracy is calculated as the RMS error relative to the exact computation, at a distance of 10 Å ($\approx 2R_0$) from the center of geometry. RMS error for the 2-charge PPCA (Eq. (18)) is shown as a function of the distance between the two charges of the PPCA $\|\bar{r}_1 - \bar{r}_2\|$. The RMS error for the 2-charge PPCA is compared to that of the point dipole approximation with an optimal center of expansion. (**a**) Cases where Eq. (23) is true. (**b**) Cases where Eq. (23) is false. This figure also includes the 2-charge optimal point charge approximation (Eq. (22)) for comparison. Connecting lines are shown to guide the eye.
doi:10.1371/journal.pone.0067715.g002

where q is the net charge of the original distribution, the monopole order error term in Eq. (12) becomes zero. If we set the center of charge as the center of expansion for the point dipole approximation, and choose the charges for the 2-charge OPCA such that the center of charge for the OPCA coincides with the center of charge for the original distribution, then

$$\mathrm{p} = \bar{q}_1 \cdot \bar{r}_1 + \bar{q}_2 \cdot \bar{r}_2 = 0 \qquad (25)$$

where $\bar{r}_1$ and $\bar{r}_2$ represent the position vectors for charges 1 and 2 respectively, of the 2-charge OPCA. Note that, with the choice of center of charge as the multipole center of expansion, the $K = 2$ OPCA is guaranteed to be at least as accurate as the point dipole approximation, as measured by the error metric defined by Eq. (12). Thus, to simplify the derivations, we will use the center of charge as the origin for the coordinate system.

The next non-vanishing error term to be minimized is the quadrupole, i.e.

$$min(\sum_{i,j = x,y,z} \left(Q_{ij} - \overline{Q}_{ij} \right)^2) \qquad (26)$$

where $\bar{Q}_{ij}$ and Q_{ij} are the quadrupole moments of the OPCA and the original charge distribution respectively, about the center of charge. The quadrupole tensor defines a unique, orthogonal set of principal axes in three-dimensional space. Since the two point charges of the 2-charge OPCA define a single line, the quadrupole potential can be expected to be best approximated by the 2-charge OPCA with the charges positioned along the principal axis that has the largest absolute principal value (figure 3). Since the quadrupole tensor $\mathbf{Q}$ is a real symmetric matrix, its principal values can be determined by the eigenvalue decomposition

$$\mathbf{Q}\mathbf{v} = \lambda \mathbf{v} \qquad (27)$$

where λ is a principal value (eigenvalue) with the corresponding principal axis (eigenvector) $\mathbf{v}$. Let λ be the largest principal value. Then, by placing the 2-charge OPCA along $\mathbf{v}$, and setting the component of quadrupole moment for the 2-charge OPCA along the principal axis $\mathbf{v}$ equal to the largest principal value λ for the original distribution, we obtain from Eq. (5) for the quadrupole moment

$$\lambda = \bar{q}_1 \bar{r}_1^2 + \bar{q}_2 \bar{r}_2^2 \qquad (28)$$

where $\bar{r}_1$ and $\bar{r}_2$ are the magnitude of the $\bar{r}_1$ and $\bar{r}_2$ vectors with center of charge as the origin.

Substituting the values for $\bar{q}_2$ and $\bar{r}_2$ from Eq. (24) and (25) respectively, we arrive at:

$$\bar{q}_1 = \frac{q\lambda}{\lambda + q\bar{r}_1^2} \qquad (29)$$

The above equation does not provide a unique solution since $\bar{r}_1$ is still unknown. Minimizing the error in the next order multipole term in Eq. (12), the octupole moment, results in quartic equations which may produce imaginary charge values. Therefore, for practical applications, as with the uncharged case, an empirical approximation may be more appropriate. Specifically, we set $\bar{r}_1 = \alpha R_0$ where α is an empirical parameter. Consider for example a typical charge distribution (a glutamic acid) from the sample

charge distributions described in the Results section below. For this charge distribution, figure 4 shows that in the mid field ($R = 2R_0$), with the choice of $\bar{r}_1 \approx 1.6R_0$ this practical point charge approximation (PPCA) is on average more accurate than the point dipole and point quadrupole approximations. For the representative sample charge distributions described in the next section, the PPCA was found to be the most accurate for $\bar{r}_1 = 1.5R_0$.

By placing the 2-charge PPCA along the principal axis with the largest principal value, we eliminate the error due to the largest component of the quadrupole tensor λ. Furthermore, since the quadrupole tensor is traceless, the other two principal values λ_a, λ_b in $\mathbf{Q}$ and $\bar{\lambda}_a, \bar{\lambda}_b$ in $\bar{\mathbf{Q}}$, are of the opposite sign to λ, and $|\lambda_a|, |\lambda_b|, |\bar{\lambda}_a|, |\bar{\lambda}_b| \leq |\lambda|$. Therefore, the error due to the other two components of the quadrupole tensor are reduced as well, i.e. $((\lambda_a - \bar{\lambda}_a)^2 + (\lambda_b - \bar{\lambda}_b)^2) \leq ((\lambda_a)^2 + (\lambda_b)^2)$. As an illustration, consider the example of figure 3(b): the error due to smaller components of the quadrupole tensor shown in figure 3(b) are smaller than the ones in figure 3(c), as $((5.92 - 4.20)^2 + (2.48 - 4.20)^2) < (5.92^2 + 2.48^2)$.

Thus, for a charge distribution with net non-zero charge, the practical point charge approximation is determined by Eq. (29), with $r_1 = 1.5R_0$. The constant of 1.5 was empirically determined for the set of sample charge distributions described in the following section.

Results and Discussion

We consider here two potential applications for the optimal and practical point charge approximations developed above – the approximation of atomic partial charge distributions for amino acid groups within proteins, and the approximation of the charge distribution of water molecule.

Atomic Level Biomolecular Modeling

Molecular modeling is commonly used to study the structure, function and activity of biological systems [24–26]. A common computational bottleneck in biomolecular modeling is the calculation of long-range electrostatic interactions: due to slow decay of these interactions with distance, simply ignoring them beyond a certain cut-off distance may lead to unacceptable accuracy loss [16,27]. Multiscale approximations are one class of methods used to speed up these calculations [5,7,16], where near-field interactions are treated exactly, while an approximation of the charge distribution is used for mid- and far-field computations.

Since the error introduced by such approximations is generally very low in the far-field, understanding the mid-field error of such approximations, including ours, is most relevant. In the context of biomolecular modeling, we consider the lower bound of the mid-field to be no less than 2 times the extent of the charge distribution ($2R_0$); the mid-field for amino acid groups is therefore greater than 5 Å.

We have applied the 2-charge practical point charge approximations (PPCA) developed above to the computation of electrostatic potential for a set of 1188 amino acid groups in five representative biomolecules that span a large range of sizes: a monomer from the virus capsid (Protein Databank (PDB) ID 1A6C) with 513 groups, the villin headpiece protein (PDB ID 1VII) with 36 groups, calcium switch protein (PDB ID 1UWO) with 91 groups, chaperonin GroEL (PDB ID 2EU1) with 524 groups, and myoglobin (PDB ID 1YMB) with 24 groups. The amino acid groups include their associated backbone atoms, NH-CH-CO for non-terminal groups, NH_2-CH-CO for N-terminal groups, and NH-CH-COO for C-terminal groups. Atomic partial charges were taken from the AMBER force field parameters [28].
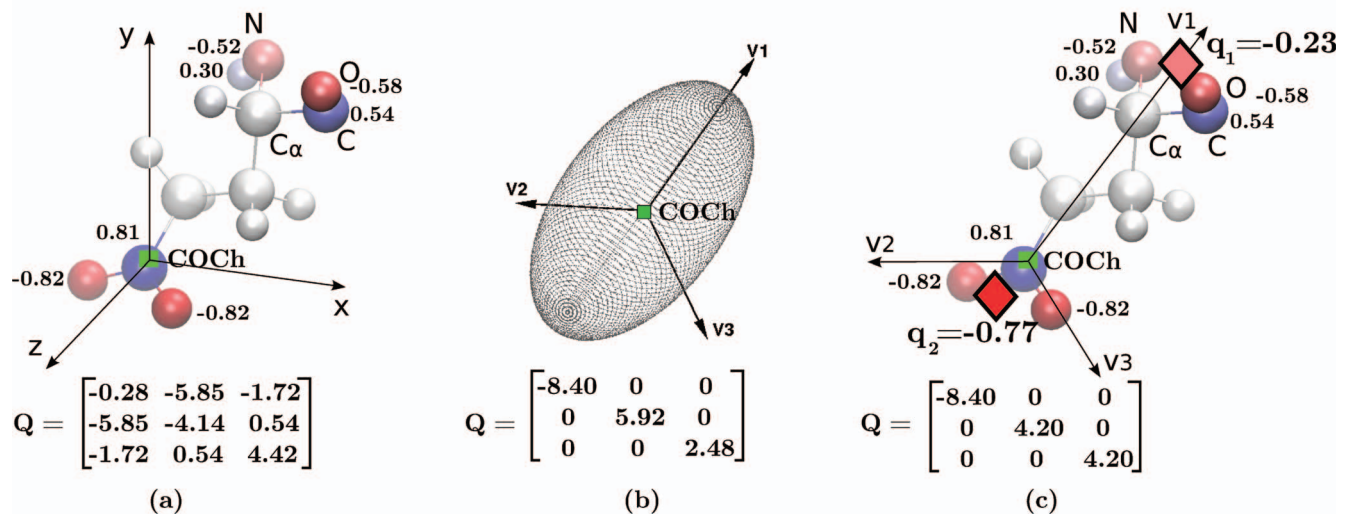
Figure 3. Illustration of a 2-charge practical point charge approximation (PPCA). For a sample charge distribution with non-zero net charge (a glutamic acid group within a protein with net charge $= -1e$, where the group includes the associated NH-CH-CO backbone atoms). (**a**) The original charge distribution with its quadrupole tensor (Eq. (5)) shown below. The atomic partial charges are represented as spheres rendered using VMD [21]. The sphere colors range from red to blue representing the charge range of $-1e$ to $+1e$. The charge values for partial charges $|q| > 0.2e$ are shown next to the atoms. As a visual reference, the backbone heavy atoms are labeled and covalent bonds are included in the figure. The green square shows the center of charge (COCh). (**b**) The principal axes, **v1, v2, v3** of the original charge distribution with the center of charge as origin (green square). Its quadrupole tensor, with the coordinate system aligned to the principal axes (Eq. (27)), is shown below. Here **v1** is the principal axis with the largest principal value. Analogous to the concept of ellipsoid of inertia in Mechanics used to characterize mass distribution, an "ellipsoid of charge" can be imagined here that helps visualize the charge distribution characterized by the quadrupole tensor. (**c**) The 2-charges of the PPCA (red diamonds) are placed such that the quadrupole moment for the PPCA equals the component of the quadrupole moment for the original charge distribution along **v1**. The quadrupole tensor produced by the 2-charge PPCA, with the coordinate system aligned to the principal axes, is shown below. The values of charges are in atomic units (e), and $e.\mathring{A}^2$ is the unit for the quadrupole tensors.
doi:10.1371/journal.pone.0067715.g003

The electrostatic potential was calculated in the mid-field (for two values: $R = 10$ Å $\approx 2R_0$ and $R = 15$ Å $\approx 3R_0$) where the approximation is likely to be least accurate. The electrostatic potential was computed at discrete points on a sphere of radius R, centered at the center of geometry. The spherical surface was discretized into 7200 grid points at which the electrostatic potential was calculated. The RMS error was calculated over all grid points and
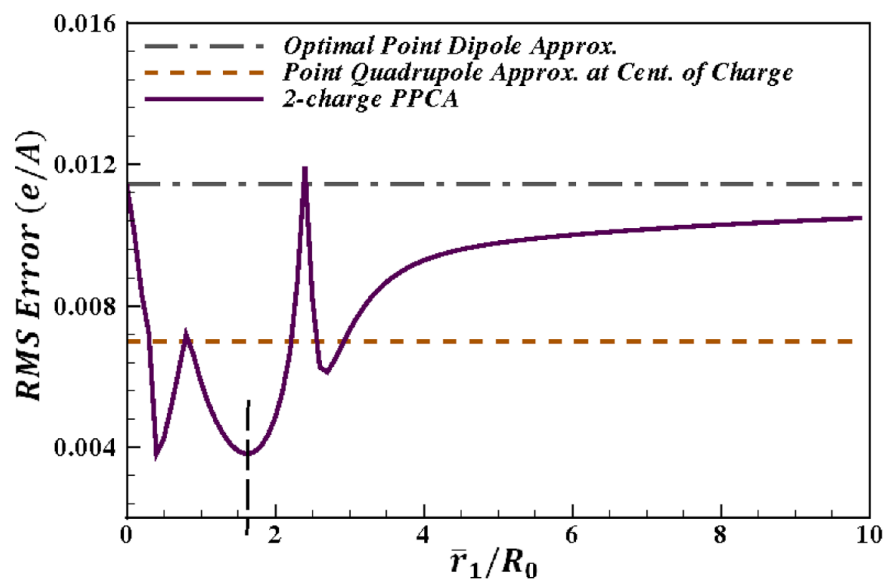

Figure 4. Accuracy of the 2-charge practical point charge approximation (PPCA) as a function of the distance $\bar{r}_1$ from the center of charge. For the sample charge distribution shown in figure 3. Accuracy is calculated as the RMS error, relative to the exact computation, at a distance of $2R_0$, where R_0 is the maximum extent of the charge distribution from the center of geometry. The point dipole and point quadrupole approximations with center of charge as the center of expansion are shown for comparison. The vertical dashed line represents the value $\bar{r}_1 \approx 1.6R_0$ that produces the lowest RMS error for the 2-charge PPCA in this case. Connecting lines are shown to guide the eye.
doi:10.1371/journal.pone.0067715.g004

all the amino acid groups in the sample as $\sqrt{\sum^N |\Phi - \Phi_{ref}|^2 / N}$, where Φ and Φ_{ref} are the electrostatic potential calculated using the approximations and the reference (original) charge distributions, respectively, and N is the number of grid points. The 2-charge PPCA was compared to the optimal point dipole and the point quadrupole approximations. The center of expansion for the point dipole approximation for uncharged and charged distributions are chosen to be the center of dipole and center of charge, respectively, which are known to be optimal [20] for the corresponding point multipole expansions. For the point quadrupole approximation, we found that the choice of center of geometry as the center of expansions for uncharged cases, and the center of charge for charged cases, produced the most accurate result, on average, in the mid-field. Accordingly, we use these points as the expansion centers for the point quadrupole approximation.

In the mid-field ($R = 10$ Å $\approx 2R_0$) the RMS error (0.0053 ± 0.0030 e/Å) for the 2-charge PPCA is comparable to the point quadrupole approximation RMS error (0.0054 ± 0.0027 e/Å), and 23% less than the optimal point dipole approximation RMS error (0.0069 ± 0.0074 e/Å), for the charge distributions considered here, figure 5. On the other hand, when electrostatic potential is calculated at a distance $R = 15$ Å, the RMS error (0.00026 ± 0.00016 e/Å) for the 2-charge PPCA is 34% less than the optimal point dipole approximation RMS error (0.00039 ± 0.00026 e/Å), while being 53% higher than the point quadrupole approximation RMS error (0.00017 ± 0.00011 e/Å). These results reflect the fact that the 2-charge PPCA is always at least as accurate as the optimal point dipole approximation, whereas the PPCA can only try to minimize the error in the

quadrupole term unlike the point quadrupole approximation, which eliminates the error in the quadrupole term. As the distance from the charge distribution increases, the accuracy of the multipole expansion, and, specifically, the accuracy of the point quadrupole approximation improves. This is evident from the errors at a distance $R = 15$ Å(figure 5(b)) which are an order of magnitude lower compared to the errors at a distance $R = 10$ Å (figure 5(a)). Note that the set of amino acid groups used here consist of approximately 20% charged and 80% uncharged distributions.

Figure 5(a) also shows that the 2-charge PPCA is on average significantly more accurate than the point dipole and quadrupole approximations for net non-zero charge distributions, compared to net zero charge distributions. Thus, the 2-charge PPCA should be significantly more accurate than the point dipole and quadrupole approximations for molecular structures that contain a significant portion of charged amino acids. And this is precisely the type of structures where the use of long-range cut-offs may lead to large errors. Also note that the standard deviation in RMS error for the 2-charge PPCA (0.0030 e/Å) is comparable to that of the point quadrupole approximation (0.0027 e/Å) and is less than half of that of the optimal point dipole approximation (0.0074 e/Å). Thus, the 2-charge PPCA is a considerably "tighter" approximation than the equivalent order optimal point dipole approximation. The higher standard deviation in RMS error for the point dipole approximation is primarily due to the cases where the center of dipole, for charge distributions with zero net charge, falls outside the extent of the original charge distribution. In these cases the "optimal" center of expansion for the point dipole approximation can be very close to the point at which the electrostatic
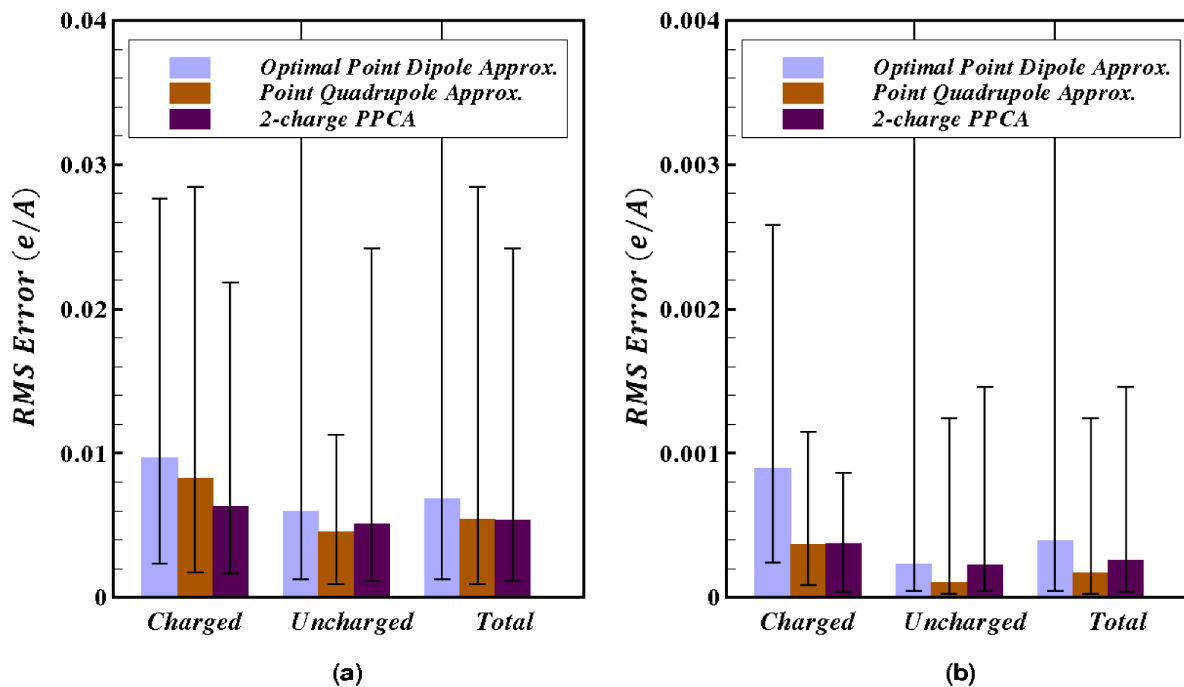

Figure 5. Accuracy of the 2-charge practical point charge approximation (PPCA). For a sample set of charge distributions relevant to biomolecular modeling. Accuracy of the point dipole and point quadrupole approximations are shown for comparison. Accuracy is calculated as the RMS error relative to the exact computation. (**a**) Error calculated at a distance of 10 Å $\approx 2R_0$ where R_0 is the maximum extent of the charge distribution from the center of geometry. (**b**) Error calculated at a distance of 15 Å $\approx 3R_0$ from the center of geometry. Error bars show the maximum and minimum absolute error. The upper values for the error bars that are cut off at the top are 0.14 and 0.006 in the left and right panels, respectively.

doi:10.1371/journal.pone.0067715.g005

potential is approximated, resulting in large errors. This source of error is explicitly removed from the PPCA.

Optimal Point Charge Approximation for Water Molecule

Water is critical for life [29,30], and is one of the most extensively studied molecules[31–34]. Accurate yet computationally efficient description of the solvent environment is essential for realistic biomolecular modeling. Commonly used simple fixed point charge models of water have achieved a reasonable compromise between accuracy and speed, but these are by no means perfect [35,36]; the search for more accurate yet computationally facile models continues [4,37,38]. The ability of a given model to reproduce electrostatic properties of the highly polar water molecule is critical to success of the model [4]. Obviously, any reasonable model needs to account for the large dipole moment of water molecule in order to reproduce dielectric properties of the liquid state. But higher moments are important too: for example, one of the components of the water quadrupole tensor is large, and was shown to have strong effect on the liquid water structure seen in simulations [39]. The octupole order terms have also been shown to be of importance: for example, these affect water structure around ions [40]. An intricate interplay between the dipole, quadrupole and octupole moments gives rise [41] to the experimentally observed charge hydration asymmetry of aqueous solvation – strong dependence of hydration free energy on the sign of the solute charge. Thus, accurate yet computationally facile representations of the complex charge distribution of water molecule should be of interest.

As an illustration of the OPCA approach, we show here that the 3-charge OPCA can accurately reproduce a quantum mechanical charge distribution of the water molecule up to the octupole moment. The specific charge density for the electron distribution of the water molecule used here (Fig. 6(a)) was determined by the CCSD method with aug-cc-pCVTZ basis set [43–45] at experimental equilibrium geometry in the gas phase. The electron charge density distribution was calculated for a box with side length of 4 Å and resolution of 0.05 Å. The resulting multipole moments of water molecule in the gas phase are comparable to available experimental values [46] (Table 1). We stress, however, that the specific charge distribution is used here only to illustrate the OPCA method and its capabilities; in what follows, no claims

Table 1. Multipole moments of a water molecule in the gas phase.

		QM (this work)	3-charge OPCA	Experimental
Dipole(D)	p_z	1.81	1.81	1.86
Quadrupole($D\text{Å}$)	Q_{zz}	0.08	0.08	0.11
	Q_{xx}	−2.53	−2.53	−2.625
	Q_{yy}	2.45	2.45	2.515
Octupole($D\text{Å}^2$)	O_{zzz}	−1.35	−1.17	NA
	O_{xxz}	−1.25	−1.44	NA
	O_{yyz}	2.61	2.61	NA

The computed values for the quantum mechanical (QM) charge distribution and the 3-charge OPCA are compared to the corresponding experimental values [46]. The coordinate system is that of Figure 6(a). Due to the symmetry, for the octupole tensor $O_{xxz} = O_{xzx} = O_{zxx}$ and $O_{yyz} = O_{yzy} = O_{zyy}$. Components of multipole moments with a value of zero are not shown. 1 debye (D) = 0.2082 eÅ.
doi:10.1371/journal.pone.0067715.t001

regarding suitability of this distribution for simulation of liquid phase water [39] are made.

Since the water molecule is neutral, it can not be represented by a 1-charge OPCA. A 2-charge OPCA can accurately represent the dipole moment but not the quadrupole and octupole moments of the distribution, which are important for an accurate representation of water [36,39–41]. Therefore, we calculate the 3-charge OPCA, as follows. In general, the 3-charge OPCA consists of three charges $\bar{q}_1, \bar{q}_2$, and $\bar{q}_3$, located at r_1, r_2, and r_3, representing 12 independent variables. But in the case of water any solution must respect the C_{2v} symmetry of the molecule, which reduces the number of independent variables to 10 (by assuming $\bar{q}_1 = \bar{q}_2$ and $\bar{z}_1 = \bar{z}_2$). Following the general procedure outlined in the "Calculating the optimal point charge approximation" section above, we first eliminate the monopole term in the error expansion Eq. (12), by setting

$$q = \bar{q}_1 + \bar{q}_2 + \bar{q}_3 = 0 \qquad (30)$$

where $q = 0$ is the monopole moment of the original charge distribution for water. Then, we eliminate the dipole term in the error expansion via

$$p_i = r_{1i}\bar{q}_1 + r_{2i}\bar{q}_2 + r_{3i}\bar{q}_3, \ i = x,y,z \qquad (31)$$

where p_i are the x,y, or z component of the dipole moment and r_{1i}, r_{2i}, r_{3i} are the x,y, or z components of r_1, r_2 and r_3. Note that in the coordinate system standard for water molecules (figure 6(a)), p_z is the only non-zero component of the dipole moment. Finally, we eliminate the quadrupole term in the error expansion, Eq. (12), by setting

$$Q_{i,j} = \qquad (32)$$
$$\frac{1}{2}\left[\bar{q}_1(3r_{1i}r_{1j} - (r_1)^2\delta_{ij}) + \bar{q}_2(3r_{2i}r_{2j} - (r_2)^2\delta_{ij}) + \bar{q}_3(3r_{3i}r_{3j} - (r_3)^2\delta_{ij})\right]$$
$$i,j = x,y,z$$

where $Q_{i,j}$ are the terms in the quadrupole tensor of the original charge distribution. Note that in the coordinate system chosen (figure 6(a)), all off-diagonal terms in the quadrupole tensor are zero, i.e. $Q_{ij} = 0, i \neq j$. Thus, we are left with a total of 9 independent equations –1 for the monopole q, 3 for the dipole terms p_x, p_y, and p_z, and 5 for the terms in the symmetric traceless quadrupole tensor, $Q_{xx}, Q_{yy}, Q_{xy}, Q_{xz}$, and Q_{xz} – to solve for 10 variables. This is an under-determined system of equations, leaving one additional variable. Solving the above set of equations results in the following solution, with $\bar{q}_1$ as the remaining variable.

$$\bar{q}_2 = \bar{q}_1 \qquad (33)$$

$$\bar{q}_3 = -2\bar{q}_1 \qquad (34)$$

$$\bar{x}_1 = \bar{x}_2 = \bar{x}_3 = \bar{y}_3 = 0 \qquad (35)$$

$$\bar{z}_1 = \bar{z}_2 = \frac{(Q_{zz} - Q_{xx})}{3p_z} + \frac{p_z}{4\bar{q}_1} \qquad (36)$$
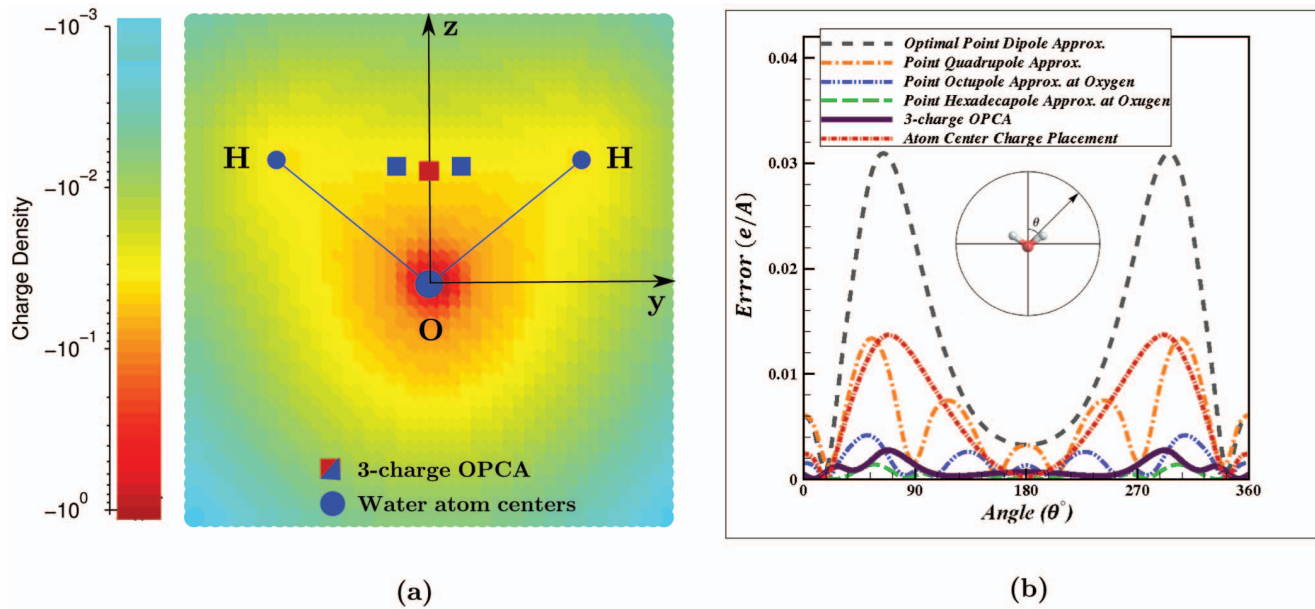
Figure 6. 3-charge optimal point charge approximation (OPCA) for water molecule. (a) The quantum mechanical electron charge density is visualized by a light blue to red colormap representing the charge density range of 0 to -1 e per $0.05 \times 0.05 \times 0.05$ Å^3. The figure shows a 3 Å $\times$ 3 Å slice of the charge distribution in the y-z plane of the water atom centers. The origin is located at the center of the oxygen atom, the water atoms lay in the y-z plane, and the z-axis bisects the hydrogen atoms. The blue dots represent the water atom centers and the red and blue squares represent the 3 OPCA charges. The central OPCA charge has a value of $-26e$ and the other two are $13e$ each. (b) The error in electrostatic potential relative to the exact computation, calculated at $2 \times R_0 = 2.8$ Å from the oxygen atom, in the y-z plane. In this case R_0 is chosen to be 1.4 Å, the mean van der Waals radius of water [42], and $2 \times R_0$ approximates the distance between the oxygen atoms in two closest water molecules. For comparison, we show the error for the 4 lowest point multipole approximations as well as for a commonly used approximation which places point charges on atom centers. To match the dipole moment of the original charge distribution, the charge placed at the oxygen position equals $-0.64e$, while the charges on the hydrogen centers are $0.32e$ each. The same relative ordering of errors is seen in the x-z and x-y planes (not shown).
doi:10.1371/journal.pone.0067715.g006

$$\bar{z}_3 = \frac{(Q_{zz} - Q_{xx})}{3p_z} - \frac{p_z}{4\bar{q}_1} \quad (37)$$

$$\bar{y}_1 = -\sqrt{\frac{(Q_{yy} - Q_{xx})}{3\bar{q}_1}} \quad (38)$$

$$\bar{y}_2 = +\sqrt{\frac{(Q_{yy} - Q_{xx})}{3\bar{q}_1}} \quad (39)$$

The value for $\bar{q}_1$ is determined numerically to minimize the octupole term in the error expansion, Eq. (12). The resulting OPCA charges are $\bar{q}_1 = 13e, \bar{q}_2 = 13e$, and $\bar{q}_3 = -26e$, located at $(0, -0.16, 0.49)$, $(0, 0.16, 0.49)$, and $(0, 0, 0.47)$ Å, respectively, (Fig. 6). (The conversion factor of 0.2082 eÅ/debye is used to convert the multipole moments in table 1 to atomic units (eÅ, eÅ^2, etc.) used in Eq. (33) – (39).)By construction, the multipole moments from the 3-charge OPCA and the quantum mechanical charge distribution are identical up to quadrupole order, as shown in (Table 1). The tight clustering of the 3 OPCA charges away from the oxygen nucleus is unexpected – in many commonly used water models the charges are placed on atom centers. Mathematically, the clustering results from minimizing the error at the octupole level, used to determine the value of q_1 in Eq. (33)–(39). For fixed dipole and quadrupole moments, the extent of the

OPCA charge distribution is controlled by the octupole moment of the original charge distribution. The small distance between the opposite OPCA charges necessitates their large magnitude to ensure correct dipole moment of the charge distribution. Although the position and magnitude of the OPCA charges may appear unusual, the OPCA representation may be more accurate than the atom-centered alternative. For example, when we place the point charges on atom centers, and adjust the corresponding charge magnitudes so that the error at the dipole order is eliminated, the RMS error in electrostatic potential at 2.8 Å around the oxygen atom center is 0.0073 e/Å, which is many times larger than the RMS error for the 3-charge OPCA (0.0010 e/Å), as shown in Fig. 6(b). Note that the OPCA representation is designed to best approximate multipole moments of the original charge distribution; it remains to be seen whether this strategy leads to accurate reproduction of other physical properties of water. A water model based on the OPCA representation will be presented in a separate study.

Figure 6(b) compares the error in electrostatic potential calculated by the 3-charge OPCA with the error produced by the point multipole approximations, relative to the exact computation using the original charge distribution. (Error calculations exclude any points that fall within the extent of the original charge distribution.) The error shown in figure 6(b) is calculated on a circle, in the plane of the water atoms (y-z plane), at $R = 2.8$ Å from the oxygen atom, which approximates the oxygen-oxygen (contact) distance between two closest water molecules. The overall RMS error in electrostatic potential calculated on a $R = 2.8$ Å spherical surface centered at the oxygen atom is 0.0010 e/Å with maximum error of 0.0027 e/Å for the 3-

charge OPCA, compared to 0.0015 e/Å with maximum error of 0.0041 e/Å for the point octupole expansion. The overall RMS error in electrostatic potential at $R = 2.23$ Å (experimental oxygen-Na^+ distance [47]) is 0.0036 e/Å (1.20 kcal/mol) with maximum error of 0.0042 e/Å (1.39 kcal/mol) for the 3-charge OPCA compared to 0.0065 e/Å (2.16 kcal/mol) with maximum error of 0.0074 e/Å (2.46 kcal/mol) for the point octupole expansion.

Conclusions

Truncated point multipole expansions are a widely used approach to approximate potentials produced by complex charge distributions. However, if only the lowest order terms in the multipole expansion are kept, as is often done in practical calculations, the point multipole expansion can produce considerable error in the mid-field. Furthermore, implementation of such approximations into existing electrostatic models that were originally developed for point charge distributions, *e.g.* pairwise implicit solvent models, presents many challenges. In this work, we have introduced an alternative to the point multipole expansion– the optimal point charge approximation (OPCA). An OPCA consists of a given number of point charges which are optimally placed to best reproduce the electrostatic potential due to the original charge distribution. By construction, OPCAs retain many of the useful properties of point multipole expansions, in particular they retain the asymptotic behavior of the point multipole expansion. At the same time, an expansion based on OPCAs can be more accurate than the point multipole expansion of the same order.

We have provided a general framework for calculating OPCAs to any order. We have also derived closed-form analytical expressions for the 1-charge OPCA, and closed-form analytical expressions that approximate the 2-charge OPCA with reasonable accuracy – the 2-charge practical point charge approximation (PPCA). We note that higher order closed-form, analytical OPCAs may be challenging to derive, but for some applications, lower order OPCAs may be sufficient. The analytical expressions derived here for the 1-charge and 2-charge OPCAs, are guaranteed to be at least as accurate as the corresponding point multipole expansion of the same order. These analytic expressions not only provide physical insight but are more computationally efficient than the numerical minimization procedures that are in general required to obtain the optimal point charge approximation. Thus, these analytic expressions may be particularly useful in applications such as molecular dynamics where computational speed is critical.

For a set of sample charge distributions relevant to biomolecular modeling, the 2-charge PPCA was found to be on average 23% more accurate than the point dipole approximation, and comparable in accuracy to the point quadrupole approximation in the mid-field (electrostatic potential evaluated at 2 times the extent of the charge distribution). The standard deviation in RMS error for the 2-charge PPCA was also 59% lower than that of the point dipole approximation and comparable to that of the point quadrupole approximation, suggesting that the approximation offered by the PPCA is "tighter" than that of the point dipole.

We also calculated the 3-charge optimal point charge approximation to represent a (quantum mechanical) gas phase charge distribution of water molecule. The electrostatic potential approximated by the 3-charge OPCA in the mid-field (2.8 Å from the oxygen atom) is on average 33.3% more accurate than that of the point octupole approximation. Interestingly, the positions of the 3 OPCA charges are quite different from atom center charge

placements based on simple point charge models such as SPC or TIP3P. Further investigation is necessary to determine if and how such a 3-charge approximation can be used in practical applications.

Representing complex charge distributions by a small number of point charges is not, by itself, a novel idea. There are a number of methods, such as RESP [48], CHELP [49], CHELPG [50], CHELMO [19], Finite Point Charge (FPC) [3], coarse graining [16,51,52] and others [53] that empirically fit a set of point charges to a given charge distribution by minimizing various error metrics in electrostatic potential over some volume or surface surrounding the charge distribution. However, a key difference between the above methods and the optimal point charge approximation introduced here, is that the OPCAs (and their practical approximations, PPCAs) inherit the physically appealing asymptotic properties of the point multipole approximation, i.e. the error in potential is guaranteed to fall off at least as fast as $1/R^{k+1}$, where R is the distance from the origin and k is the highest order of the multipole terms retained in the expansion. In contrast, fitting the representative charges to minimize electrostatic error over some arbitrary volume or surface (e.g. molecular surface) does not guarantee the above asymptotic behavior, and can potentially lead to relatively large errors outside the volume or surface used for fitting.

Furthermore, in comparison to point multipoles, expansions based on PPCAs have many desirable properties that may be useful in practical computations; in particular, their mathematical form – the sum of Coulombic contributions from point sources – is simpler than that of the conventional point multipole expansion and is amenable to common speed-up schemes such as the generalized Born implicit solvent model [54]. Thus, PPCAs may be easier to implement into existing molecular dynamics protocols.

The optimal point charge approximation presented here is a new concept; thus its many applications and potentially useful properties remain unexplored in this proof-of-concept work. We expect OPCA/PPCA to have utility in coarse-grained [51,55] and multi-scale methods [16], especially in dynamics [27] where analytic expressions and the simplicity of the algorithms is key. The approximations we have introduced provide a systematic way of deriving approximate charge distributions that have the potential to be both computationally efficient and produce an accurate representation of the original electrostatic potential. To further improve the representation of the original potential via OPCAs, future work may consider partitioning the original charge distribution into several domains, and finding OPCA/PPCA for each of them separately, similar to the distributed multipoles approach [56–58]. Further exploration of the mathematical and physical properties of OPCAs is also desirable. Finally, the 3-charge OPCA for a charge distribution representing water molecule is quite accurate to the octupole order. This accuracy, combined with the simplicity of a 3-charge OPCA, is noteworthy.

Author Contributions

Conceived and designed the experiments: CB RA SI AO. Performed the experiments: CB SI. Analyzed the data: CB RA SI. Contributed reagents/materials/analysis tools: CB RA SI. Wrote the paper: CB RA SI AO.

References

1. Greengard L, Rokhlin V (1987) A fast algorithm for particle simulations. Journal of Computational Physics 73: 325–348.
2. Lee F, Warshel A (1992) A local reaction field method for fast evaluation of long-range electrostatic interactions in molecular simulations. Journal of Chemical Physics 97: 3100–3107.
3. Colonna F, Evleth E, Ángyán JG (1992) Critical analysis of electric field modeling: Formamide. J Comput Chem 13: 1234–1245.
4. Ichiye T, Tan ML (2006) Soft sticky dipole-quadrupole-octupole potential energy function for liquid water: An approximate moment expansion. The Journal of Chemical Physics 124: 134504+.
5. Carrier J, Greengard L, Rokhlin V (1988) A fast adaptive multipole algorithm for particle simulations. SIAM Journal on Scientific and Statistical Computing 9: 669–686.
6. Cai W, Deng S, Jacobs D (2007) Extending the fast multipole method to charges inside or outside a dielectric sphere. Journal of Computational Physics 223: 846–864.
7. Lambert CG, Darden TA, Board JA Jr (1996) A multipole-based algorithm for efficient calculation of forces and potentials in macroscopic periodic assemblies of particles. Journal of Computational Physics 126: 274–285.
8. Greengard L, Rokhlin V (1997) A new version of the fast multipole method for the Laplace equation in three dimensions. Acta Numerica 6: 229–269.
9. Schmidt KE, Lee MA (1991) Implementing the fast multipole method in three dimensions. Journal of Statistical Physics 63: 1223–1235.
10. Cheng H, Greengard L, Rokhlin V (1999) A Fast Adaptive Multipole Algorithm in Three Dimensions. Journal of Computational Physics 155: 468–498.
11. Ying L (2006) A kernel independent fast multipole algorithm for radial basis functions. Journal of Computational Physics 213: 451–457.
12. Bishop TC, Skeel RD, Schulten K (1997) Difficulties with multiple time stepping and fast multipole algorithm in molecular dynamics. Journal of Computational Chemistry 18: 1785–1791.
13. Pollock EL, Glosli J (1995) Comments on PPPM, FMM, and the Ewald Method for Large Periodic Coulombic Systems. Available: http://arxiv.org/abs/cond-mat/9511134. arXiv:cond-mat/9511134.
14. Cruz FA, Barba LA (2009) Characterization of the accuracy of the fast multipole method in particle simulations. International Journal for Numerical Methods in Engineering 79: 1577–1604.
15. Koumoutsakos P (1995) Fast multipole methods for three dimensional N-body problems. Technical Report 96N25344, NASA, Ames Research Center.
16. Anandakrishnan R, Onufriev AV (2010) An N log N approximation based on the natural organization of biomolecules for speeding up the computation of long range interactions. Journal of Computational Chemistry 31: 691–706.
17. Pearlman D (1995) AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. Computer Physics Communications 91: 1–41.
18. Schlick T (2010) Molecular Modeling and Simulation: An Interdisciplinary Guide. Interdisciplinary Applied Mathematics. Springer.
19. Sigfridsson E, Ryde U (1998) Comparison of methods for deriving atomic charges from the electrostatic potential and moments. Journal of Computational Chemistry 19: 377–395.
20. Platt D, Silverman B (1996) Registration, orientation, and similarity of molecular electrostatic potentials through multipole matching. Journal of Computational Chemistry 17: 358–366.
21. Humphrey W, Dalke A, Schulten K (1996) VMD – Visual Molecular Dynamics. Journal of Molecular Graphics 14: 33–38.
22. Gordon JC, Fenley AT, Onufriev A (2008) An analytical approach to computing biomolecular electrostatic potential. II. Validation and applications. Journal of Chemical Physics 129: 075102.
23. Gramada A, Bourne P (2008) Resolving a distribution of charge into intrinsic multipole moments: A rankwise distributed multipole analysis. Physical Review E 78: 1–7.
24. Freddolino P, Arkhipov A, Larson S, McPherson A, Schulten K (2006) Molecular dynamics simulations of the complete satellite tobacco mosaic virus. Structure 14: 437–449.
25. Karplus M, McCammon J (2002) Molecular dynamics simulations of biomolecules. Nature Structural and Molecular Biology 9: 646–652.
26. Salsbury FR (2010) Molecular dynamics simulations of protein dynamics and their relevance to drug discovery. Current Opinion in Pharmacology 10: 738–744.
27. Anandakrishnan R, Daga M, Onufriev AV (2011) An n log n generalized born approximation. Journal of Chemical Theory and Computation 7: 544–559.
28. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, et al. (2005) The Amber biomolecular simulation programs. Journal of Computational Chemistry 26: 1668–1688.
29. Tait MJ, Franks F (1971) Water in Biological Systems. Nature 230: 91–94.
30. Ball P (1999) Life's Matrix: A Biography of Water. Farrar, Straus, and Giroux, New York.
31. Kale S, Herzfeld J (2012) Natural polarizability and flexibility via explicit valency: The case of water. The Journal of Chemical Physics 136: 084109+.
32. Kell GS (1975) Density, thermal expansivity, and compressibility of liquid water from 0.deg. to 150.deg. Correlations and tables for atmospheric pressure and saturation reviewed and expressed on 1968 temperature scale. Journal of Chemical & Engineering Data 20: 97–105.
33. Tu Y, Laaksonen A (2000) The electronic properties of water molecules in water clusters and liquid water. Chemical Physics Letters 329: 283–288.
34. Dill KA, Truskett TM, Vlachy V, Hribar-Lee B (2005) Modeling water, the hydrophobic effect, and ion solvation. Annual Review of Biophysics and Biomolecular Structure 34: 173–199.
35. Guillot B (2002) A reappraisal of what we have learnt during three decades of computer simulations on water. Journal of Molecular Liquids In Molecular Liquids 101: 219–260.
36. Mark P, Nilsson L (2001) Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. J Phys Chem A 105: 9954–9960.
37. Wu Y, Tepper HL, Voth GA (2006) Flexible simple point-charge water model with improved liquid-state properties. The Journal of chemical physics 124: 024503+.
38. Walsh TR, Liang T (2009) A multipole-based water potential with implicit polarization for biomolecular simulations. J Comput Chem 30: 893–899.
39. Niu S, Tan ML, Ichiye T (2011) The large quadrupole of water molecules. The Journal of Chemical Physics 134: 134501+.
40. Te JA, Ichiye T (2010) Understanding structural effects of multipole moments on aqueous salvation of ions using the soft-sticky dipolequadrupoleoctupole water model. Chemical Physics Letters 499: 219–225.
41. Mukhopadhyay A, Fenley AT, Tolokh IS, Onufriev AV (2012) Charge hydration asymmetry: the basic principle and how to use it to test and improve water models. Journal of Physical Chemistry B 116: 9776–9783.
42. Franks F (2000) Water: a matrix of life. Cambridge: Royal Society of Chemistry, 11 pp.
43. Dunning TH (1989) Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. Journal of Chemical Physics 90: 1007.
44. Kendall RA, Dunning TH, Harrison RJ (1992) Electron-affinities of the 1st-row atoms revisited–systematic basis-sets and wave-functions. Journal of Chemical Physics 96: 6796.
45. Woon DE, Dunning TH (1995) Gaussian basis sets for use in correlated molecular calculations. V. Core-valence basis sets for boron through neon. Journal of Chemical Physics 103: 4572.
46. Clough SA, Beers Y, Klein GP, Rothman LS (1973) Dipole moment of water from Stark measurements of H2O, HDO, and D2O. The Journal of Chemical Physics 59: 2254–2259.
47. Joung IS, Cheatham TE (2008) Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. J Phys Chem B 112: 9020–9041.
48. Bayly CI, Cieplak P, Cornell W, Kollman PA (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. Journal of Physical Chemistry 97: 10269–10280.
49. Chirlian LE, Francl MM (1987) Atomic charges derived from electrostatic potentials: A detailed study. Journal of Computational Chemistry 8: 894–905.
50. Breneman CM, Wiberg KB (1990) Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. Journal of Computational Chemistry 11: 361–373.
51. Basdevant N, Borgis D, Ha-Duong T (2007) A coarse-grained protein-protein potential derived from an all-atom force field. Journal of Physical Chemistry B 111: 9390–9.
52. Izvekov S, Voth GA (2005) A multiscale coarse-graining method for biomolecular systems. Journal of Physical Chemistry B 109: 2469–73.
53. Swart M, van Duijnen PT, Snijders JG (2001) A charge analysis derived from an atomic multipole expansion. Journal of Computational Chemistry 22: 79–88.
54. Chen J, Brooks CL, Khandogin J (2008) Recent advances in implicit solvent-based methods for biomolecular simulations. Current Opinion in Structural Biology 18: 140–148.
55. Izvekov S, Voth GA (2005) Multiscale coarse graining of liquid-state systems. Journal of Chemical Physics 123: 134105.
56. Gramada A, Bourne PE (2011) Coarse-graining the electrostatic potential via distributed multipole expansions. Computer Physics Communications 182: 1455–1462.
57. Stone AJ (1981) Distributed multipole analysis, or how to describe a molecular charge distribution. Chemical Physics Letters 83: 233–239.
58. Stone AJ, Alderton M (1985) Distributed multipole analysis. Molecular Physics 56: 1047–1064.