Theor Chem Acc (2012) 131:1137
DOI 10.1007/s00214-012-1137-7

REGULAR ARTICLE

# Polarisable multipolar electrostatics from the machine learning method Kriging: an application to alanine

**Matthew J. L. Mills · Paul L. A. Popelier**

Received: 9 May 2011 / Accepted: 29 August 2011 / Published online: 18 February 2012


**Abstract** We present a polarisable multipolar interatomic electrostatic potential energy function for force fields and describe its application to the pilot molecule MeNH-Ala-COMe (AlaD). The total electrostatic energy associated with 1, 4 and higher interactions is partitioned into atomic contributions by application of quantum chemical topology (QCT). The exact atom–atom interaction is expressed in terms of atomic multipole moments. The machine learning method Kriging is used to model the dependence of these multipole moments on the conformation of the entire molecule. The resulting models are able to predict the QCT-partitioned multipole moments for arbitrary chemically relevant molecular geometries. The interaction energies between atoms are predicted for these geometries and compared to their true values. The computational expense of the procedure is compared to that of the point charge formalism.

**Keywords** Quantum chemical topology · Force field · Multipole moment · Polarisation · Atoms in molecules · Machine learning

Published as part of the special collection of articles: From quantum mechanics to force fields: new methodologies for the classical simulation of complex systems.

M. J. L. Mills
Manchester Interdisciplinary Biocentre (MIB),
131 Princess Street, Manchester M1 7DN, UK

P. L. A. Popelier (✉)
School of Chemistry, University of Manchester,
Oxford Road, Manchester M13 9PL, UK
e-mail: pla@manchester.ac.uk

## 1 Introduction

Molecular dynamics and mechanics methods are used to investigate a wide variety of chemical problems, including but not limited to peptide structure [1], semiconductors [2] and the origins of life [3]. There is a competition between the accuracy of the results of such studies and the computational cost of carrying them out. Quantum mechanical (QM) methods may be employed in cases where electronic detail is required [4]. Explicit treatment of the electrons makes the cost of QM simulations very high and as such they are limited to use with small systems over short time periods [5]. Hybrid methods may be employed to reduce this cost, wherein small parts of the system considered chemically important are treated explicitly and the rest of the system is treated approximately [6, 7]. For cases where QM detail is not required, the whole system may be treated using an approximate potential energy function. The most commonly employed class of such functions are referred to as force fields (FF). FF energy expressions make the study of longer simulations or larger systems feasible by sacrificing some of the accuracy of the description of the potential energy surface (PES) of the results for reduced computational loads. The continuing increases in computer power along with the introduction of molecular dynamics codes that can take advantage of GPUs [8] have allowed the recent extension of the domain of application of FFs to much longer simulation times and system sizes than have previously been available [9, 10].

There are many FFs commonly used in biochemical research. Popular potentials include CHARMM [11, 12], AMBER [13, 14], GROMOS [15] and OPLS [16, 17]. Certain aspects differ between them all, but they share a common framework. The central shared assumption is that the potential energy of a chemical system can be written as

a function of the relative positions of its nuclei. This potential energy function is not derived from first principles but is guided by chemical intuition. Typically, it takes the form of a combination of bond-stretch, angle-bend and torsional rotation functions (grouped together as the 'bonded' part) along with an expression for the electrostatic, short-range repulsion and long-range attraction interactions (the 'non-bonded' part). The bond-stretch, angle-bend and torsion terms are usually expressed in the form given in Eq. 1

$$E_{\text{bonded}} = \sum_{n=2}^{n_{\max}} \left( \sum_{B} C_R^n (B - B_0)^n + \sum_{\theta} C_\theta^n (\theta - \theta_0)^n \right) + \sum_{\varphi} \sum_{m=1}^{m_{\max}} C_\varphi^m \cos m\varphi, \quad (1)$$

where $B$, $\theta$ and $\varphi$ refer to bond lengths, angles and dihedrals, respectively, a subscript 0 denotes a reference value and the $C$s are a set of constants to be determined. The values of $m_{\max}$ and $n_{\max}$ refer to the order of the expansion for each term and vary between FF, with higher values expected to represent the energetics of each physical phenomenon more accurately. Work has been carried out on the utility of including coupling terms for a more accurate description of the bonded energy [18]. Various forms for the bond-stretch energy have been proposed and compared [19]. Improper torsional terms may also be added to describe deviation from equilibrium geometries in planar systems [20]. The non-bonded expression typically consists of a Lennard-Jones function paired with a charge–charge representation of the electrostatic interaction, or

$$E_{\text{non-bonded}} = \sum_{i<j} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon r_{ij}} \right). \quad (2)$$

Here, $q_i$ represents the partial charge assigned to the atom $i$, $r_{ij}$ is the distance between atoms $i$ and $j$, $\varepsilon$ is the relative permittivity that allows for screening of the electrostatic interaction and $A_{ij}$, $B_{ij}$ are constants found by application of combining rules to predefined atomic values [21]. These interactions are computed between atoms separated by three or more bonds. Special terms may be added to cover additional physical phenomena such as hydrogen bonding. The parameters present in the bonded equations are typically found by fitting to a set of data. This set may be collected experimentally either [22] by application of ab initio methods to small molecules [23] or by a combination of both [24]. The parameterisation process requires the predefinition of a set of distinct atom types. These are typically chosen to represent atoms within a broad class of functional groups. The transferability of parameters to models outside the fitting set is strongly dependent on which atom types are included.

Force fields as described have achieved success in allowing the investigation into a wide range of chemical problems. However, they have been shown to be deficient in predicting results for important classes of molecules, such as nucleic acid base pairs [25]. In some cases, the electrostatic polarisation has been found to be entirely responsible for binding in biochemical systems [26]. The fitting of parameters to bulk properties can cause the lack of a correct description at the more detailed level of clusters. The FF geometries of small molecular clusters may be in strong disagreement with their ab initio counterparts [27]. Such methods suffer from poor transferability, and methods that predict correct macroscopic behaviour from correct microscopic behaviour should be less affected by this issue. Correct bulk behaviour should emerge from such methods.

Bond-stretch and angle-bend terms are transferable between chemical environments with a low associated error. The non-bonded terms must therefore be responsible for the majority of the lack of transferability found in FF. The electrostatic interaction term in particular is affected [28]. In typical FF, charges are assigned to atom types by fitting the ab initio electrostatic potential around a series of molecules [29]. This method reduces transferability between conformations as in reality the charge distribution is parametrically dependent on the nuclear positions, but this is not recreated when fixed charges are employed.

Capturing the change in electrostatic properties with molecular conformation is the goal of polarisable FF. There are polarisable versions of several major FFs, including AMBER [30] and CHARMM [31]. The polarisation methods employed typically fall into one of three groups: point polarisable dipoles [32], Drude oscillators [33] or fluctuating charges [34]. The first two methods require iteration to self-consistency, which adds computational cost to a force field calculation. All methods suffer from the polarisation catastrophe. This occurs when polarisable electrostatic moments are at a short separation, leading to a divergent evaluation of their interaction energy. This problem is typically avoided by use of an arbitrary damping function [35].

This communication focuses on the replacement of the electrostatic term with a more detailed form that incorporates polarisation by application of a machine learning method. Applications of machine learning methods to the design of potentials are relatively sparse. Applications of neural networks have recently been reviewed by ourselves [36]. Of particular relevance are applications to the simulation of liquids [37, 38] and to the computation of interatomic potentials [39–41]. Kriging (Gaussian Process Regression) has also been used to model potentials for solids [42] and has been applied to QSAR problems [43]. Of the investigated methods, Kriging provides the best

cost-to-accuracy ratio. An application of the current procedure to ethanol was recently described [44] and is here generalised and applied to an amino acid pilot system. *The method proposed here avoids the iterative calculation of the polarisation and does not require a (short-range) damping function.*

Point charges, as employed in FF, are isotropic, whereas the electron distribution of a chemical system is not. Consideration of the ab initio expression for the total molecular energy clarifies this issue. The total molecular energy can be written as a function of reduced density matrices, as specified by Eq. 3, which is valid for closed-shell Hartree–Fock wave functions,

$$E_{\mathrm{tot}} = -\frac{1}{2}\int \mathrm{d}\mathbf{r}_1 \nabla_1^2 \rho(\mathbf{r}_1,\mathbf{r}_2)_{\mathbf{r}_1=\mathbf{r}_2} + \frac{1}{2}\iint \mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \frac{\rho_{\mathrm{tot}}(\mathbf{r}_1)\rho_{\mathrm{tot}}(\mathbf{r}_2)}{r_{12}} - \frac{1}{4}\iint \mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \frac{\rho(\mathbf{r}_1,\mathbf{r}_2)\rho(\mathbf{r}_2,\mathbf{r}_1)}{r_{12}} \quad (3)$$

The first term is the electronic kinetic energy, the second the Coulomb interaction and the third describes the exchange interaction. Partitioning the total energy into intra-atomic and inter-atomic contributions, as detailed in Eq. 4, enables the exploration of a mapping between ab initio energy contributions and the classical FF expressions of Eqs. 2 and 3,

$$E_{\mathrm{tot}} = \sum_A E^A_{\mathrm{kin}} + \frac{1}{2}\sum_A E^{AA}_{\mathrm{coul}} - \frac{1}{4}\sum_A E^{AA}_X + \frac{1}{2}\sum_{\substack{A,B\\B\neq A}} E^{AB}_{\mathrm{coul}} - \frac{1}{4}\sum_{\substack{A,B\\B\neq A}} E^{AB}_X \quad (4)$$

If we restrict the inter-atomic Coulomb interactions in Eq. 4 (the fourth group of terms) to be between atoms separated by more than three bonds, then we obtain the analogue of the electrostatic energy as given by the last term in Eq. 2. An expansion in terms of spherical harmonics of $1/r_{12}$ appearing in the electrostatic energy expression of Eq. 3 (the second term) allows the interaction energy between two atoms $A$ and $B$ to be expressed in terms of multipole moments and a geometric interaction tensor $T^{l_B m_B}_{l_A m_A}$, with the introduction of a convergence criterion, or

$$E_{\mathrm{Coulomb}} = \sum_A \sum_{B\neq A} \sum_{l_A=0}^{l_{\max}} \sum_{l_B=0}^{l_{\max}} \sum_{m_A=-l_A}^{l_A} \sum_{m_B=-l_B}^{l_B} Q_{l_A m_A} T^{l_B m_B}_{l_A m_A} Q_{l_B m_B} \quad (5)$$

$Q_{lm}$ is the $m$th component of a multipole of rank $l$. A moment of rank 1 has $2l + 1$ individual components numbered from $-l$ to $+l$. For example, the dipole moment is of rank 1 and has three components, $m = (-1, 0, 1)$. $T^{l_B m_B}_{l_A m_A}$ is dependent on the internuclear separation and mutual orientation of local coordinate systems centred on the nuclei. We note here that analytical first and second derivatives of this potential are required for geometry optimisation and molecular dynamics calculations. These are known analytically within the rigid body formalism [45], and enable a robust investigation into PESs [46]. A future communication will detail the analytical form of the derivatives for the method described herein. The general formulae for geometry-dependent multipoles have been recently described [47]. It is useful to introduce a rank for an atom–atom interaction, denoted $L$. This rank measures the number of multipole moments used to compute the energy and is equal to $l_A + l_B + 1$. In previous studies, we have shown that $L = 5$ is necessary to reproduce both the broad structural features of a liquid-like system (i.e. water [48]) and the structural details of hydrated biomolecules [27]. Computing the complete $L = 5$, energy would require the building of 25 models, one for each multipole moment of each atom. The building of Kriging models is the most time-consuming part of the process; therefore, we restrict the models to monopole, dipole and quadrupole moments only (i.e. nine components in total). No additional technology is required to treat the higher moments, and they may be included where CPU time is available. The model building time does not increase with the rank of a multipole moment. The decision to restrict the included moments to up to quadrupole means that $l_{\max}$ in Eq. 5 is set to 2. Hence, octupole-dipole, octupole-charge and hexadecapole-charge terms (and their inverse counterparts) are not included in the energy evaluation of Eq. 5 used to test the models. Using the true octupole and hexadecapole moments, we can investigate the contribution of these terms to a total 1–4 and higher interaction energy. The mean of the unsigned values of the sum of these three energy contributions amounts to 8 kJ mol$^{-1}$ for 100 test conformations. This means that, on average, 8 kJ mol$^{-1}$ is un-assessed in terms of Kriging polarisation.

The convergence criterion states that the expansion will converge when the magnitude of a particular vector is smaller than the inter-nuclear distance. This vector is defined as the difference between two position vectors, each marking the location of an electron with respect to a nucleus. In general, this convergence criterion allows the application of Eq. 5 to 1–4 and higher terms; although in some cases, 1–4 interactions may not meet this criterion. In such cases, it is possible to move the predicted multipole moments to different centres by application of a "shift" procedure [49]. The resulting shifted moments allow the calculation of initially non-convergent 1–4 interactions on the same footing as the 1–5 and higher interactions [50]. An algorithm for determining the optimal shift has recently

been described [51]; however, it is not applied here as it requires higher moments to give accurate results. The separation of nuclear and electronic coordinates allows the separate computation of the multipole moments and their subsequent combination to compute energies. Multipole moments provide a more accurate picture of the ab initio electrostatics than atom-centred charges. The values of the multipole moments for a particular system may be determined by several methods. Most common amongst these is the Distributed Multipole Analysis [52], whilst several alternative methods have been defined and compared [53, 54]. FF that include multipolar electrostatics and polarisation have been described [55], but are yet to find use amongst computational biochemists. As such, the number of cases that validate the additional expense of including a more detailed picture of electrostatics is still limited [56].

A method that allows strict atomic partitioning is required in order to keep within the FF framework. The idea central to QCT is the partitioning of a system into subspaces by means of a gradient vector field. The methods of QCT thus include the partitioning of a molecular charge density into regions that can be described as atoms by consideration of the gradient vector field of the charge density [57, 58]. QCT is rooted in quantum mechanics [59], using atomic theorems to define atomic properties. The use of the term QCT has been justified previously [60, 61]. Atomic properties are described by integrals over the resulting atomic *basins*, where the integrand is a point property. For example, the integration over the basin of 1 returns the atomic volume, $N$. When 1 is replaced by the electron density, the result is the atomic monopole moment, or charge. The partitioning is applicable to chemical wave functions in general. Figure 1 shows the QCT atoms of an extended conformation of AlaD. The boundaries between atoms are clear. When an interatomic surface is not present, atoms extend to infinity and in practical applications have to be capped by an isosurface in $\rho(\mathbf{r})$.

The atomic multipole moments (AMM) can be expressed as integrals over the atomic basins.

$$Q_{lm}(\Omega) = \int_{\Omega} \mathrm{d}\mathbf{r}\rho_{\mathrm{tot}}(\mathbf{r})R_{lm}(\mathbf{r}) \tag{6}$$

$\rho_{\mathrm{tot}}$ is the total ground-state electron density and $R_{lm}$ is a regular spherical harmonic function, which can be complex. $\Omega$ denotes integration over an atomic basin. We work with real spherical harmonics; therefore, Eq. 6 has to be modified according to the procedure described in Stone's monograph [62]. The AMMs have a parametric dependence on the nuclear positions and a direct dependence on the electron density, i.e. AMMs are conformationally dependent [63]. For a given conformation, there is a

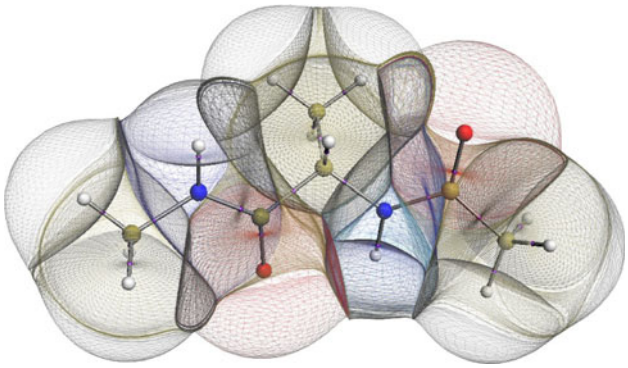

**Fig. 1** The QCT atomic partitioning of the AlaD molecule. Atoms are capped by the $\rho = 0.001$ a.u. isosurface

single value for each AMM in a topologically partitioned molecule. We can therefore describe polarisation as the change in the value of a multipole moment of a QCT atom when the molecular nuclear positions (and hence the internal coordinates) are altered. This approach replaces a previously explored avenue, where polarisation was treated in the context of *long-range* Rayleigh–Schrödinger perturbation theory [64]. There the change in an AMM (typically only the dipole moment) is expressed in terms of an electric field through a polarisability (tensor). In the current method, we do not express polarisability explicitly. Instead, *we construct a direct mapping between an AMM of interest to nuclear coordinates of its environment*. This method is valid at short range because it draws its data from (super)molecular wave functions that obey the Pauli principle and whose electron densities lead to well-defined non-overlapping atoms. An additional advantage of the use of QCT is that any system can be decomposed into discrete atoms, and thus the distinction between inter- and intramolecular polarisation can be removed. That is, an atom is polarised by all other atoms in the system, whether this is composed of a single molecule (as in the current case) or a cluster of molecules. Previous work [37] has described the application of the current method to clusters of rigid water molecules, where the AMMs of a central water molecule were modelled as dependent on the position and orientation of a number of neighbouring water molecules (from 1 to 4). Provided a consistent set of coordinates can be defined for every structure in a training set, Kriging models can be built for the AMMs that model inter- and intramolecular polarisation on the same footing.

Figure 2 illustrates the concept of a moment directly varying as a function of varying nuclear positions. This figure displays the value of the monopole moment of $C_\alpha$ of AlaD for a series of values of the protein backbone dihedral angles $\varphi$ and $\psi$. $C_\alpha$ is chosen as its AMMs are expected to vary most significantly with these central angles. It should be noted that the monopole moment $Q_{00}$ coincides with the

*net* atomic charge because Eq. 6 involves the *total* charge density, which contains the nuclear charge as well. Rotations were carried out around these angles starting from the extended structure in Fig. 1 without optimisation of the resulting structures. An angular step of 10° was employed in both dihedral angles. The charge remains positive in all cases, but varies over the torsion space between almost neutral ($Q_{00} = 0.001$e) and nearly a full electronic charge ($Q_{00} = 0.884$e). In general, the relationship between $Q_{00}$ and $\varphi$, $\psi$ is particularly complicated, but also smooth. We cannot confidently map this plot onto a Ramachandran map because internal coordinates other than $\varphi$, $\psi$ are not altered (or relaxed). In fact, a small number of conformations were so strained that a wave function could not be obtained for them with GAUSSIAN's default parameters. Rather than commenting on local extrema, we just point out the broad periodicity over 180° in $\varphi$ and $\psi$.

Replacement of the isotropic charge–charge electrostatic interaction energy by Eq. 5 is not immediately feasible due to the expense involved in evaluating a wave function for each conformation of interest and integrating the electron density over atomic basins. It is first necessary to find an alternative analytical form for the AMMs that avoids such an evaluation at every step. In this work, this is achieved by modelling the dependence of the AMMs on the internal coordinates of a molecule. A prescription for defining the internal coordinates of a molecule is required. We use an atomic local frame (ALF) to express the Cartesian coordinates of a molecule from an atomic point of view. Each atom of the molecule has a different ALF. The frame is centred on the atom in question. The *x*-axis is chosen to point along the line from the origin to the heaviest atom bonded to it in the molecular graph. The *xy*-plane is defined so as to contain both the heaviest and second heaviest atoms bonded to the origin atom. In the case of terminal atoms, the *x*-axis is defined by the only atom bonded to the origin atom, and the *xy*-plane is defined to contain the *x*-axis atom and its own heaviest bonding partner. Where two or more duplicate elements are bonded to a central atom, the Cahn-Ingold-Prelog rules are employed to decide priority. If all atoms are equal according to these rules, then the lowest numbered amongst them is used. For example, the atom H18 has the frame H18–N17–C6. Being a terminal hydrogen atom, its *x*-axis is defined by its bonding partner N17. This atom has two other bonding partners, C6 and C19. As C6 is directly bonded to O10 and C3, with C19 bonded to H20, H21 and H22, the former is of higher priority. The AMMs are computed in these frames.

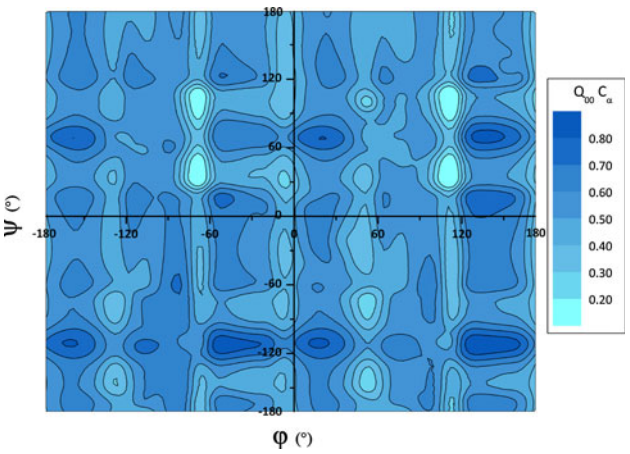


**Fig. 2** Contour map of the relationship between the protein dihedral angles $\varphi$, $\psi$ and the monopole moment ($Q_{00}$) of the AlaD α-carbon

Internal coordinates are used to build the Kriging models as for a system of *N* atoms, there are 3 *N*-6 such coordinates. The dimensionality of the feature space is important when building models. Higher-dimensional relationships are more difficult to model, so *N* should be minimised where possible. However, Kriging is able to cope well with high values of *d*. For the AlaD system the dimensionality is 60. The internal coordinates used in this work are the magnitudes of the position vectors of the atoms defining the *x*-axis and *xy*-plane and the angle between these two vectors. The position of any atom *not* used to define the ALF is described using spherical polar coordinates. Machine learning methods can be used to build models for the relationship between AMMs and nuclear geometry. The Kriging method [65, 66] (also termed Gaussian Process Regression) is employed here for this purpose. Internal coordinate vectors for a set of appropriate examples and the corresponding AMMs constitute the training data of the models. The result of the model building is an analytical form for the AMMs as a function of nuclear position,

$$y(\mathbf{x}) = \mu + \sum_{i=1}^{N_{\text{train}}} w_i \exp\left[-\sum_{j=1}^{d} \theta_j \left|x_j^i - x_j\right|^{p_j}\right] \tag{7}$$

Here, $\mu$ is the global trend in the data and $w_i$ is the Kriging weight for a particular training example *i*, computed from the training data. $\theta_j$ and $p_j$ are the *j*th components of the $3N - 6$ dimensional Kriging vector parameters $\boldsymbol{\theta}$ and $\mathbf{p}$. These are vectors with $3N - 6$ elements that describe the weight each feature is given in making a prediction ($\boldsymbol{\theta}$) and the power to which the difference between *test* and *training* values of features are raised. There are therefore two times the number of coordinates parameters ($2(3N - 6)$) to be determined. In Eq. 7, $x_j^i$ is an internal coordinate for a *training* set geometry, and $x_j$ is the corresponding value of that coordinate for a *test* geometry. For such a *test* geometry, the Kriging model assumes an AMM to be composed of the global trend, added to a weighted sum over all the training examples of the correlation between the structures of the

*test* and each *training* conformation. The correlation is here measured by the exponential power correlation function, although alternatives may be used in its place. It should be noted that for conformations far outside the training set, the magnitude of the feature differences is large. As this value increases, the exponential factor becomes smaller and the predicted moments tend towards the global trend, which is an estimate of the average over all conformations. Conformationally averaged AMMs have been previously suggested as a method for the inclusion of a better electrostatic description than the point charge method [67]. The fact that the Kriging models reduce to this representation for unseen conformations is an additional advantage of using Eq. 7 to model the polarisation of AMMs. The building of a Kriging model involves maximisation of the logarithm of the likelihood function $L$ of the training data with respect to these parameters,

$$\begin{aligned}\log L(\boldsymbol{\theta}, \mathbf{p}, \mu, \sigma | y^i; i = 1, 2, \ldots, N_{\text{train}}) \\ = -\frac{N_{\text{train}}}{2}\log(\sigma^2) - \frac{1}{2}\log(|\mathbf{K}|) \\ - \frac{(\mathbf{y} - \mathbf{1}\mu)^T \mathbf{K}^{-1}(\mathbf{y} - \mathbf{1}\mu)}{2\sigma^2}\end{aligned} \tag{8}$$

$\mathbf{K}$ is the correlation matrix whose elements are the values of the exponential power correlation function for pairs of examples in the training set, $\mathbf{y}$ is the vector of training outputs and $\mu$, $\sigma^2$ are the mean and variance of the training data. The vector $\mathbf{1}$ is has $N_{train}$ components that are all equal to 1. The elements of $\mathbf{K}^{-1}(\mathbf{y} - \mathbf{1}\mu)$ are the Kriging weights referred to in Eq. 7. The computational expense of the training lies in inverting $\mathbf{K}$ for each step of the optimisation; therefore, search methods that require small numbers of steps are generally preferred. A particle swarm optimisation was used to find the optimal parameters $\boldsymbol{\theta}$ and $\mathbf{p}$ for each AMM. The results of this optimisation were then subjected to a Nelder–Mead downhill search to refine their values. A more detailed account of Kriging as applied to AMMs can be found in [37].

The quality of the final models depends on the number of training examples, $N_{\text{train}}$, and how well the AMM hypersurfaces can be modelled as Gaussian processes. The ultimate test of the method lies in the accuracy of geometry optimisation and molecular dynamics calculations. Carrying out such a test requires expressions for the electrostatic forces and parameterisation of Eq. 1 against the electrostatic method defined here. In addition, a mathematical form for the long-range attractive and short-range repulsive interactions must be chosen. Whether this will prevent a close agreement with the underlying QM method for such tests is an open question as of yet. We work under the assumption that improved electrostatics will allow a closer agreement with the QM method than the point charge formalism. The accuracy of the isolated method can be investigated by creating a series of test cases, generated by the same method as the training cases. Comparison of the predicted AMMs (for each atom) to the true values found by the evaluation of Eq. 6 by numerical integration provides a direct measure of the predictive power of the models. The force field electrostatic energy can be computed using the predicted and true moments, measuring the feasibility of the replacement of the charge–charge interaction in FF with the machine learning–based polarisable multipole method described herein.

In future work, we will investigate the transferability of Kriging models by a feature selection process. The prediction formula in Eq. 7 requires that the full set of internal coordinate features be present. In order to use the model to predict the AMMs of atoms in environments other than those of the training molecules, the list of features needs to be truncated. For example, if we wish to predict AMMs for an alanine α-carbon in a protein using the models built with AlaD, we can only include the features present in both systems, i.e. the coordinates of the side chain and peptide bond atoms. The values of $\boldsymbol{\theta}$ can be interpreted as weights for each feature, giving their relative importance to a predicted AMM. The features can be ranked in the order of importance by consideration of their $\boldsymbol{\theta}$ values, and models can be rebuilt using a set of features that describe the environment in decreasing detail as features are removed in this rank order. As atom types emerge naturally from this analysis, they do not have to be imposed onto systems by arbitrary considerations. Measuring the effect of this feature selection on the accuracy of the moment predictions allows the quantification of the transferability of the QCT AMM models. Such a prescription allows the models built here for the AlaD molecule to be used to predict the electrostatics of larger molecules, for example longer polyalanine chains. These models will provide accurate local polarisation due to the features that are included in the transferred Kriging models. Any polarisation effect from the presence of atoms not in the originally modelled molecule will not be captured by the transferred models. For this reason, investigations are underway into the length of amino acid chain that can be feasibly modelled, as well as the extent of the protein environment that must be included in order for the AMMs to become constant with respect to including more environment.

## 2 Computational details

All ab initio calculations were performed using GAUSSIAN03 [68]. Structures were created and inspected with GaussView3.0 [69]. Atomic integrations were carried out with the programs AIMall [70] and MORPHY01 [71, 72], with the GUI [73] of the latter used for visualisation of

atomic basins. Kriging models were built and tested using the in-house programs EREBUS and NYX, respectively. Marvin [74] was used in the generation of force field minimum energy structures. Molecular images were produced with VMD [75, 76].

## 3 Generation of training data

Testing the method requires a pilot system of appropriate relevance to the intended final application. Alanine acts as such a system, providing the structural elements common to all amino acids. We cap the single amino acid structure with –NHMe and –COMe to produce AlaD. The connectivity and numbering system for this molecule is shown in Fig. 3, along with the $\varphi$ and $\psi$ angles that are used in Figs. 2 and 4. It should be noted that the proposed methodology is not dependent on any protein-specific factors. It may therefore be considered generic and can be applied, for example, to biomolecules (e.g. proteins, DNA/RNA), liquids and solvated ions, and can be used to compute the interactions between any combination of these species.

The quality of the final Kriging models is dependent on the sampling method used to select the training data. It is necessary to include nuclear arrangements similar to those that will later occur in the context of prediction. Here, we test the ability of Kriging to predict nuclear arrangements generated by the same process as the training data, so this similarity is guaranteed. As future applications lie in FFs, the method of structure generation must be relevant to molecular dynamics or geometry optimisation. Low-frequency normal modes can often be used to describe large-scale movements in proteins [77] and structural changes in smaller molecules. We therefore first perform a search to find as many of the minima on the PES as possible. The AlaD molecule was drawn and then optimised at the HF/6-31G** level of theory. The resulting geometry was used to seed a conformational search on the Dreiding force field PES. 500 structures were produced via this method. These structures were then re-optimised at the HF/6-31G** level. This low level was employed for three reasons, at the

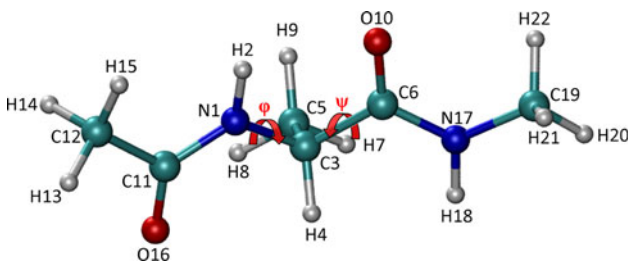


**Fig. 3** Structure of the AlaD molecule and the numbering system employed

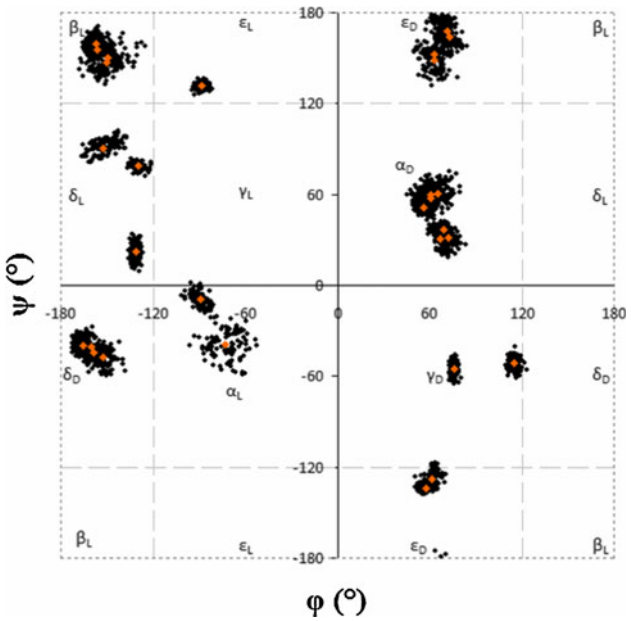


**Fig. 4** $\varphi$, $\psi$-Projection plot of the minimum energy structures of AlaD (shown in *orange*) and the structures distorted along the normal modes of vibration (shown in *black*)

current stage of FF development. First, Hartree–Fock lacks dispersion energy, which allows us to separate out the Coulomb part of the ab initio energy [78]. This is important for the parameters of the empirical part of the potential, subject to a future publication in progress, where the Lennard-Jones function is employed to compute the 1–4 and more distant 1–*n* dispersion interactions. Secondly, HF/6-31G** also allows for fast computation of the required training and test data, where the small basis set compensates for the systematic error known to be present in HF. The accuracy of the method described herein is not strongly dependent on the underlying ab initio method. Therefore, we can extend conclusions drawn at this level to higher levels. Thirdly, the interatomic Coulomb energies are larger in magnitude at HF level compared to a method that incorporates electron correlation, using the same basis set. Indeed, the HF method is known to produce much more polar wave functions than post-HF or DFT methods. The energy prediction errors that will be discussed below are therefore worst upper limits in terms of absolute value. For the worst predicted test conformation, the monopole moment for N1, for example, changes from −1.57e to −1.05e (or ∼30%) on moving from HF/6-31G** to B3LYP/6-311+G(2d,p) (the latter evaluated at the HF/6-31G** geometry). From Eq. 9a, one can deduce that smaller absolute values of an AMM lead to smaller errors in interaction energy. Indeed, the first two terms in Eq. 9a show that an AMM prediction error is amplified (i.e. multiplied) by the magnitude of an AMM on a probe atom. Fourthly and finally, given that we can make general

comments about the method based on HF wave functions, we can also make comparisons to HF results for long polyalanine chains in the very near future, connecting to the results in this paper. At higher levels of theory, comparison to chains of significant length is prohibited by the scaling of ab initio geometry optimisation. We note here that, in total, 29 minima were found for the molecule on the HF/6-31G** PES.

Population of the basins around these minimum energy structures is achieved by employing the normal modes of vibration. Normal modes were computed for each of the minima. The dataset was created by randomly distorting each minimum along a normal mode by a random factor, 100 times for each minimum. This produced 2,900 distorted structures. The distorted structures are shown projected onto the $\varphi$, $\psi$ plane as black points in Fig. 4, along with the minima shown as orange points. This projection shows the extent of the distortions in $\varphi$, $\psi$ space, but in reality, each conformation differs from another in all internal coordinates. The combination of minima and normal mode distortions means that the training set contains information related to conformational change. For example, rotation about an amide bond would be well predicted in terms of AMMs as minima corresponding to the starting and end structures are present in the training set, as well as a set of intermediate structures generated by the normal mode distortion procedure. For FF applications, it is possible to restrict the number of minima used in the model building process to minima belonging to a set of chemically prevalent structures. Certain regions of the Ramachandran map may not appear with a significant probability during a molecular dynamics simulation. In addition to this, minima may be structurally similar in terms of their $\varphi$ and $\psi$ angles, as shown by the clusters of orange points in Fig. 4. Such structures may differ in other significant coordinates (particularly torsion angles of the type H18–N17–C6–O10), but some are related by rotations of the capping methyl groups which may bias the models towards particular backbone structures. Reduction in the number of minima used to populate the training set is possible by consideration of the RMSD between pairs of structures in a chosen coordinate space and should result in more accurate Kriging models. This is because for a given training set size, there will be more distorted structures per minimum and the interpolation effort can be focused on interpolating between the distortions about a particular minimum rather than between the different minima themselves. A careful study of which of the full set of minima should be included will be carried out prior to the application of the currently proposed method to proteins.

The 2,900 distorted structures were randomly numbered, and wave functions for each of the first 1,000 examples were computed. It will turn out, by experimentation, that 1,000 examples were sufficient to train and test the models. The final 1,900 structures were hence discarded. The AMMs were then determined with respect to the principle axes of inertia. Prior to model building, the AMMs must be rotated to the ALF of the atom they are centred on [79]. Determining the AMMs in a common reference frame and then rotating them to their ALFs saves $N_{\text{train}} * (N_{\text{atom}} - 1)$ wave function evaluations compared to the alternative of rotating the molecule to each ALF prior to carrying out the ab initio calculations and subsequent atomic integration. A total of 22,000 atomic integrations were carried out. The average time for a single point computation was 127 s, and integrations averaged 16 s per atom for H, 47 s for C, 99 s for N and 58 s for O. The approximate total CPU time to produce the data necessary to build the Kriging models was 254 h. A local Linux computer cluster composed of approximately 200 compute nodes was used to generate the results. Therefore, approximately 2 h 30 min was required to prepare the model builds in real time. The majority of the total CPU time for the entire process was expended on building the Kriging models. The values of the Kriging vector parameter $\mathbf{p}$ (Eq. 7) may be optimised in the maximisation of the likelihood to take any real number value in the range $0 < p \leq 2$, but may also be fixed at either 1 or 2. The reduction in flexibility of the correlation function caused by fixing $\mathbf{p}$ results in a poorer fit, but significant savings in the computation time needed to build the models can be achieved as the number of parameters is halved. For the sequentially built models, those with optimised values of $\mathbf{p}$ took an average of 15 h per AMM. The average times with $\mathbf{p}$ fixed at 1 and 2 were 1 h 20 min and 55 m, respectively. Overall, if $\mathbf{p}$ is optimised, then the Kriging CPU time is 3,564 CPU h, or 93% of the total time to build the models. Using the Linux cluster, this represents 18 h in reality. For $p = 1$, the percentage spent building models becomes 50%, and for $p = 2$, it is 44%. Work towards an efficient FORTRAN implementation of the Kriging model building process is ongoing. A significant speedup of the building of models for individual AMMs will make feasible increased values of $l_{\max}$ in Eq. 5. The current program (EREBUS) is written in the JAVA language. Of the data production CPU time, 13% represents ab initio calculation and the remaining 87% is spent on atomic integration. These relative percentages will fluctuate as higher levels of theory are employed and algorithmic efficiency improvements are made.

## 4 Results and discussion

Testing the final models requires us to determine their ability to reproduce the values of Eq. 5 for arbitrary chemically relevant conformations of AlaD. We can write

the difference between the true and predicted energy, for a given multipole–multipole interaction, in terms of the true AMMs and their prediction errors as follows,

$$\Delta E_{l_A m_A}^{l_B m_B} = T_{l_A m_A}^{l_B m_B}(Q_{l_A m_A}\Delta Q_{l_B m_B} + Q_{l_B m_B}\Delta Q_{l_A m_A} - \Delta Q_{l_A m_A}\Delta Q_{l_B m_B}) \quad (9a)$$

where

$$\Delta Q_{lm} = Q_{lm}^{\text{true}} - Q_{lm}^{\text{predicted}} \quad (9b)$$

noting that

$$T_{l_A m_A}^{l_B m_B} \propto 1/r^{l_A + l_B + 1} \quad (10)$$

where $r$ is the internuclear distance.

There are three issues to consider when examining the interaction energy error for a pair of atoms. First is the errors in the predicted moments, second the magnitude of the true moments and third the magnitude of the interaction tensor. The AMM prediction errors are foremost as they can be controlled by the improvement of the models. It is clear that large errors in the prediction of the AMMs will lead to large errors in the prediction of $E_{AB}$. Larger-valued AMMs will cause the inflation of the first two terms inside the brackets. This may happen, for example, in the case of oxygen atoms, which carry a high net charge or for very polar carbon atoms. The value of the interaction tensor also influences the error. For atoms that are well separated in space, the magnitude of $T_{l_B m_B}^{l_A m_A}$ is diminished compared to atoms that are in close proximity. In the latter case, $T_{l_B m_B}^{l_A m_A}$ may be very large and interaction energies are subsequently large. In addition, the inverse power dependence of $T_{l_B m_B}^{l_A m_A}$ on the values of $l_A$ and $l_B$ implies that the contribution to the total error decreases with increasing $L = l_A + l_B + 1$. *Attempts to improve the accuracy of the method should therefore be focused on the modelling of the lower rank AMMs.*

The optimal number of training examples is unknown a priori. Previous tests have shown that 500 examples are sufficient for systems with numbers of atoms similar to AlaD. To investigate this, Kriging models were built with sequentially increasing training set sizes, beginning at 100 and increasing in steps of 50, up to a maximum of 500. This method allows the inspection of the developing accuracy of the models as more examples are added to the training set. Training examples were filtered based on the magnitude of their integration errors, measured by the values of $L(\Omega)$. Ideally, this quantity should be zero for all atoms; deviation from this value implies inaccurate integration of AMMs. The cut-off was set at 0.001 au. The remaining data of the original 1,000 examples were used as an *external* test set, amounting to 100 test geometries.

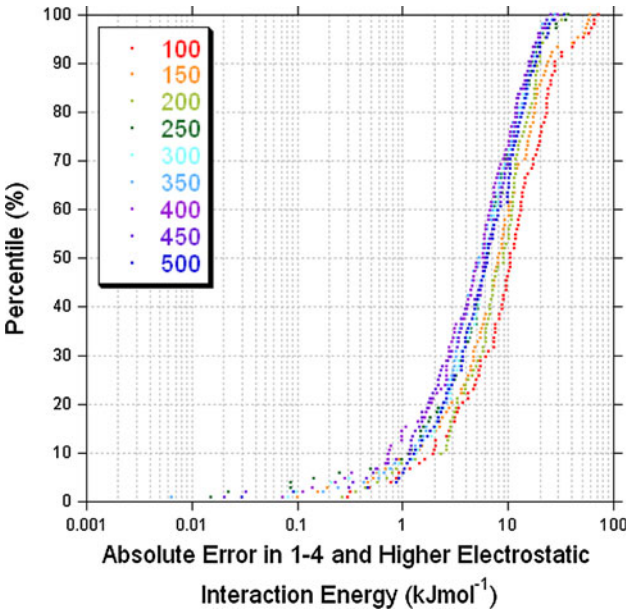


**Fig. 5** Development of absolute total 1–4 and higher electrostatic interaction energy errors as training set size is sequentially increased. The average absolute interaction energy is 83 kJ mol$^{-1}$ and has range from 0.04 to 248.5 kJ mol$^{-1}$

Figure 5 shows the S-curves for each training set size when used to predict the total 1–4 and higher electrostatic interaction energies of these test conformations. The curves show the $\log_{10}$ of the *absolute* interaction energy error against their *percentile values*. For example, the 50th absolute energy error percentile is the error that 50% of the predicted geometries are within. The interaction energy error can be defined as the difference between the energy evaluated in Eq. 5 using the true AMMs and using the predicted AMMs, where $A$ and $B$ are separated by three or more bonds. Alternatively, the same interaction energy error can be obtained by summing the energy difference evaluated in Eq. 9a, over all $l_A$, $l_B$, $m_A$ and $m_B$ values, and over all atoms $A$ and $B$ separated by three or more bonds. It should be emphasised again that the absolute value of this interaction energy error is plotted on the $x$-axis of Figs. 5, 6, 7 and 8. *Improvement of the models is manifested in leftward movement of the S-curve*. The randomly selected example geometries forming the initial training set of 100 distorted geometries (red curve) result in significant errors compared to the optimal models, with a maximum of 73 kJ mol$^{-1}$ and an average of 15 kJ mol$^{-1}$. Increasing the training set size by adding 50 examples each time produces a visually distinguishable improvement of the models up to a training set size of 250 examples. Thereafter, the progression is small and does not represent an improvement in every case. This suggests that 250 examples are required to produce the optimal models for the atoms of AlaD. Smaller training set sizes reduce the time needed to build all of the

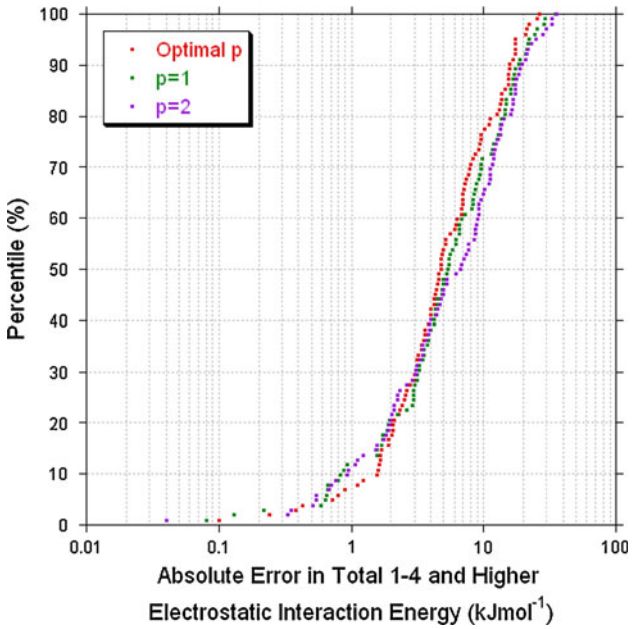


**Fig. 6** Absolute total 1–4 and higher electrostatic interaction energy errors for optimal training set size models where **p** is optimised (*red*), fixed at 1 (*green*) and fixed at 2 (*purple*). The average absolute interaction energy is 83 kJ $mol^{-1}$ and has range from 0.04 to 248.5 kJ $mol^{-1}$

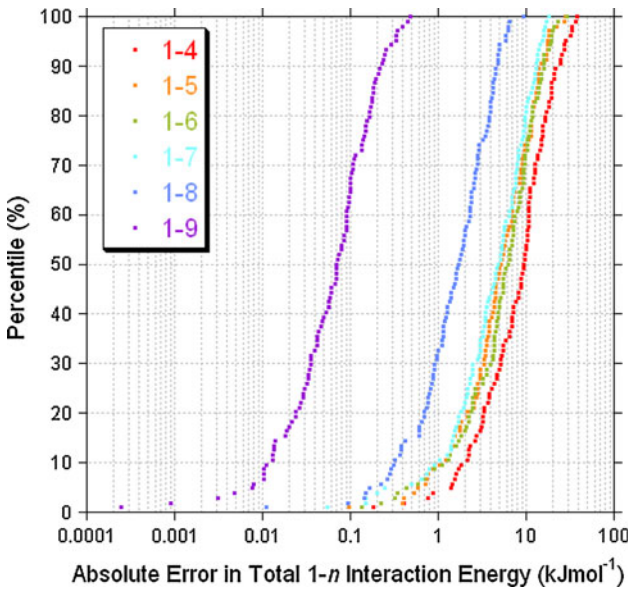


**Fig. 7** Absolute total 1–*n* electrostatic interaction energy errors for optimal training set size models for $n = 4$–9

AMM models for a system, reducing the number of wave functions required, integrations over the electron density along with both the number of steps in the sequential building of Kriging models and the dimensionality of the correlation matrix, **K**. In future applications, the number of calculations need not be as many as those carried out here.

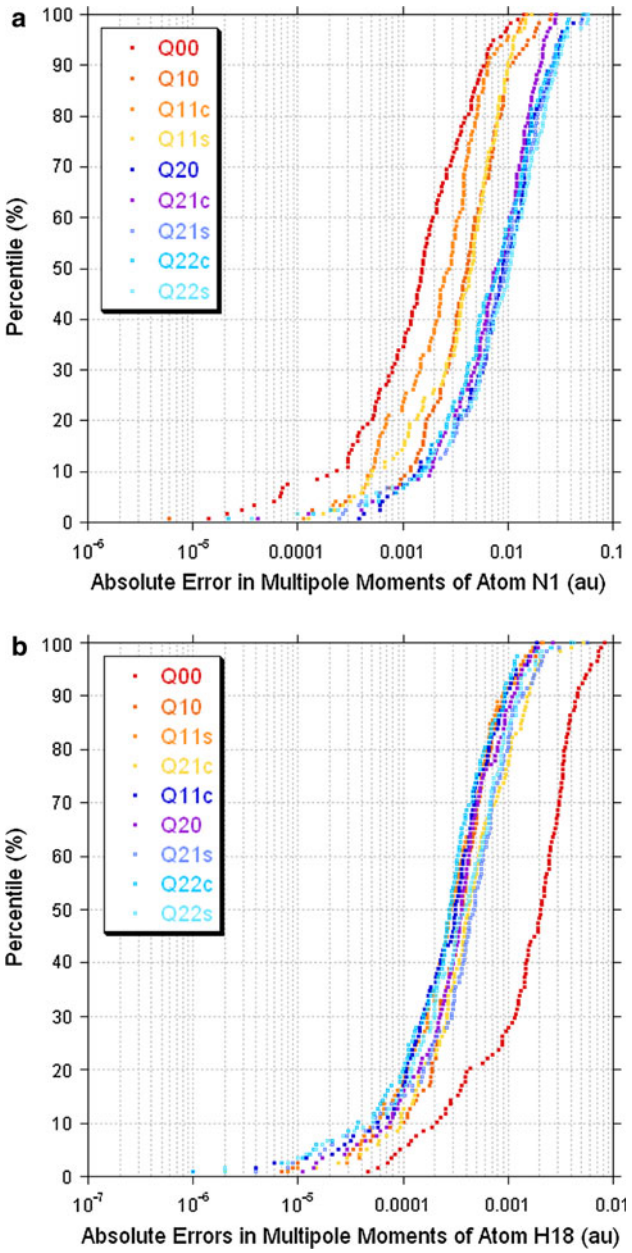


**Fig. 8** S-curves of the absolute error in prediction of the nine multipole moments of atoms **a** N1 and **b** H18 for each structure in the test set. MM errors are given in appropriate atomic units (see caption for Table 2)

We are not limited to using the same number of training examples for each model, thus the final AMM models were chosen by consideration of the $r^2$ correlation coefficient of each when used to predict the training set. This procedure was carried out by EREBUS. The optimal models for each AMM can then be combined to predict optimally accurate values for AMMs, and from these interaction energies can be obtained.

Figure 6 shows the absolute total interaction energy error S-curves for the optimal training set size models

where **p** is optimised (red) and fixed at 1 (green) and 2 (purple). Fixing the values of **p** at 1 improves the lowest 20% of the prediction errors compared to the optimal **p**. Specifically, Fig. 6 shows that the green curve ($p = 1$) intersects the red curve (optimal **p**) at the point with coordinates (2 kJ mol$^{-1}$, 20%); moving beyond this point, from the left to right, one finds that the red curve is superior to the green one. Similarly, fixing the values of **p** at 2 (purple curve) improves the lowest 30% of the prediction errors compared to the optimal **p**. However, the upper part of the curve deteriorates for both $\mathbf{p} = (1, 1, \ldots, 1)$ and $\mathbf{p} = (2, 2, \ldots, 0.2)$ with the maximum error (not easily perceptible from the graph) increasing by 8.7 kJ mol$^{-1}$ for $\mathbf{p} = (1, 1, \ldots, 0.1)$ and 8.3 kJ mol$^{-1}$ for $\mathbf{p} = (2, 2, \ldots, 0.2)$. As the low percentile errors remain within the generally quoted chemical accuracy of 4 kJ mol$^{-1}$ in all cases, we choose to proceed with the models wherein **p** is an optimised parameter, gaining accuracy in the higher percentiles but increasing total CPU time, as discussed above. In future applications that require significantly greater numbers of models to be built, fixing **p** may become necessary in order to make the building of those models computationally feasible.

The average absolute error in the total 1–4 and higher (1–$n$, $n \geq 4$) electrostatic interaction energy when predicting using the **p**-optimal models is 7 kJ mol$^{-1}$, with a maximum of 26 kJ mol$^{-1}$ and a minimum of 0.1 kJ mol$^{-1}$. The standard deviation over the test cases is 6 kJ mol$^{-1}$. The range of absolute interaction energies corresponding to these errors is from 0.04 to 248 kJ mol$^{-1}$, with an average value of 83.3 kJ mol$^{-1}$. These total energy errors are "low-detail" tests of the method, involving summation over all atom pairs $A$, $B$ and AMMs. An understanding of how the method can be improved can be gained by inspecting the results at a greater level of detail.

To further break down the total energy errors into their components, we can remove the summation over all atom pairs and consider groups of individual atom–atom interactions. Table 1 shows the average results for the 1–$n$ interactions between pairs of elements for values of $n$ found in the AlaD system, along with the average values of the interaction energies for each combination. Figure 7 shows the S-curves for the total $1 - n$ interaction energies summed over atom pairs.

There are 174 1–4 and higher interactions in the full set, divided by element-element criteria into 10 subsets: 57 H–H interactions, 50 C–H, 24 O–H, 18 N–H, 9 C–C, 8 O–C, 4 N–C, 2 N–O, 1 O–O and 1 N–N. Dividing the interactions according to $n$ produces 35 subsets in total. In general, the average interaction errors decrease with the value of $n$, as can be seen from Fig. 7. The only significant exception to this rule is for the N–C interactions. The general pattern occurs because the dependence on $n$ is a consequence of Eqs. 9a, 9b and 10. In general, higher values of $n$ imply greater average values of $R$, reducing the magnitude of $T^{l_A m_A}_{l_B m_B}$ in the calculation of the prediction errors. In some cases, the molecule may fold in such a way as to nullify this by making certain 1–$n$ distances (where $n > 4$) shorter than 1–4 distances. Deviation can also be caused by the fact that the AMMs of two atoms of the same element may not be predicted equally as accurately as each other over the test set. The 1–5N–C distances in the test set are on average 15% longer than the 1–4 distances. Therefore, the second explanation must be true, and this is borne out by an examination of the AMM contributions to the energy error.

Equation 9a suggests that the most fundamental elements in the total energy errors are the errors in the individual AMMs as these can be decreased in magnitude by employment of better models. Figure 8 shows the S-curves for the absolute errors in prediction of the AMMs for the example atoms N1 and H18, whilst Table 2 collects the prediction errors on each AMM of each atom averaged over the test set.

For N1 (Fig. 8a), the monopole is the AMM with the smallest magnitude prediction errors, followed by the three dipole components and then the five quadrupole components. This pattern is repeated for all non-hydrogen atoms whose AMMs vary little with distortion of the nuclear skeleton. For instance, in some cases, the charge may have errors close to the dipole components, e.g. $Q_{10}$ is better predicted on average than $Q_{00}$ for atom C5. In general for these atoms, an inverse relationship exists between prediction accuracy and moment rank, $l$.

This relationship is not observed for H18 (Fig. 8b), wherein the absolute errors in the prediction of the monopole moment are significantly worse than the rest of the moments. This pattern is repeated across all hydrogen atoms in AlaD. The charge is predicted with the same order of magnitude absolute average error for all atoms, with the maximum at 0.0039e (C18) and minimum at 0.0018e (H8). The higher AMMs have an order of magnitude greater average error in prediction for non-hydrogen atoms compared to hydrogen. To place these errors in context, the average true values of each moment are included in Table 2.

Whilst the ultimate genesis of the total energy errors is in the moment predictions, it is interesting to separate their influence into atomic contributions. This is achieved by setting all moments to their true values except for those of one atom, which is predicted using its optimal models. The test set total energies were computed 22 times, each atom being treated as described. That is, the true atoms of 21 atoms interacting with the predicted moments on one atom compared to the completely true energies. Table 3 collects the resulting errors in the total energy for each atom. The

**Table 1** Average absolute error in total 1–$n$ ($n \geq 4$) electrostatic interaction energy $\overline{|\Delta E|}$ for element–element interactions grouped by $n$, along with average magnitudes of each group of interaction energies $\overline{|E|}$

| Elements | 1–$n$ | $\overline{\lvert E \rvert}$ | $\overline{\lvert \Delta E \rvert}$ | $\lvert \Delta E \rvert_{\min}$ | $\lvert \Delta E \rvert_{\max}$ | $\sigma$ |
|---|---|---|---|---|---|---|
| H–H | 1–4 | 14.66 | 1.31 | 0.0060 | 7.43 | 1.18 |
| | 1–5 | 33.34 | 0.97 | 0.0043 | 3.81 | 0.81 |
| | 1–6 | 66.82 | 0.81 | 0.00085 | 4.14 | 0.68 |
| | 1–7 | 6.55 | 0.62 | 0.018 | 4.24 | 0.59 |
| | 1–8 | 6.88 | 0.53 | 0.010 | 2.27 | 0.40 |
| | 1–9 | 1.10 | 0.10 | 0.00024 | 0.48 | 0.098 |
| H–C | 1–4 | 484.17 | 7.07 | 0.20 | 33.42 | 5.74 |
| | 1–5 | 71.49 | 3.13 | 0.023 | 15.98 | 2.95 |
| | 1–6 | 345.07 | 3.07 | 0.063 | 15.88 | 3.06 |
| | 1–7 | 27.88 | 2.71 | 0.058 | 19.03 | 2.64 |
| | 1–8 | 12.14 | 0.71 | 0.0092 | 3.84 | 0.65 |
| H–N | 1–4 | 106.67 | 3.75 | 0.064 | 15.63 | 3.31 |
| | 1–5 | 538.37 | 3.14 | 0.0038 | 12.54 | 2.51 |
| | 1–6 | 37.77 | 2.62 | 0.018 | 19.45 | 2.65 |
| | 1–7 | 23.98 | 1.86 | 0.064 | 8.57 | 1.44 |
| H–O | 1–4 | 702.46 | 4.25 | 0.076 | 13.60 | 3.37 |
| | 1–5 | 180.07 | 4.61 | 0.0089 | 22.87 | 4.03 |
| | 1–6 | 650.10 | 3.87 | 0.064 | 20.76 | 3.65 |
| | 1–7 | 195.78 | 2.23 | 0.045 | 10.77 | 2.07 |
| | 1–8 | 29.89 | 1.93 | 0.020 | 12.46 | 1.86 |
| C–C | 1–4 | 1,709.98 | 4.00 | 0.12 | 14.10 | 3.41 |
| | 1–5 | 162.07 | 2.04 | 0.011 | 8.81 | 1.64 |
| | 1–6 | 385.89 | 2.06 | 0.042 | 7.78 | 1.67 |
| | 1–7 | 28.96 | 0.58 | 0.011 | 2.56 | 0.46 |
| N–C | 1–4 | 147.69 | 1.80 | 0.012 | 7.26 | 1.55 |
| | 1–5 | 1,413.28 | 3.52 | 0.046 | 11.59 | 2.73 |
| | 1–6 | 75.92 | 1.42 | 0.012 | 6.02 | 1.11 |
| O–C | 1–4 | 1,090.56 | 4.12 | 0.0038 | 16.20 | 3.42 |
| | 1–5 | 1,055.32 | 2.30 | 0.016 | 7.68 | 1.63 |
| | 1–6 | 70.09 | 1.33 | 0.011 | 7.07 | 1.23 |
| | 1–7 | 273.35 | 1.72 | 0.033 | 8.36 | 1.52 |
| N–O | 1–4 | 978.73 | 2.12 | 0.041 | 8.11 | 1.89 |
| | 1–6 | 712.95 | 2.09 | 0.034 | 16.38 | 2.79 |
| N–N | 1–4 | 1,071.55 | 2.51 | 0.033 | 11.69 | 2.36 |
| O–O | 1–6 | 634.34 | 1.65 | 0.014 | 14.10 | 2.47 |

All energies are given in kJ mol$^{-1}$. Each line corresponds to a set

quantities these errors are measured against are the same as the energies plotted in Fig. 6, where the average absolute total interaction energy is 80 kJ mol$^{-1}$ and the range is approximately 250 kJ mol$^{-1}$.

Then, N–C ordering for the 1–n interactions as discussed above is explained by the contributions of atoms N1 and N17. The 1–4 N–C interaction is between atoms N17 and C5, whilst the two 1–5 interactions occur between atoms N1–C19 and N17–C11. N1 makes a significantly higher contribution to the total energy error than does N17, the former 2.31 kJ mol$^{-1}$ to the latter 0.61 kJ mol$^{-1}$ on average. In addition, C19 makes the highest contribution of all C atoms with 2.20 kJ mol$^{-1}$. The combination of the errors on the atoms N1 and C19 therefore outweigh the increase in the interaction distance in influencing the energy errors.

The errors in Table 3 can be somewhat correlated to the chemical environment of the molecule. For example, the methyl group formed by atoms H7, H8, H9 and C5 shows a trend where H9 is worse predicted than H7 and H8. The molecular distances between these H atoms and O16 are smaller for H9, suggesting a stabilising interaction is formed between this pair of atoms. This results in a more polarised H9 with larger higher moments than H7 and H8. The greater magnitude of the moments means that their prediction errors will be larger for models that are equally as accurate in percentage terms. This example is not

**Table 2** Average absolute atomic multipole moment values $\overline{|Q|}$ and corresponding average absolute prediction errors $\overline{|\Delta Q|}$ over the test set examples

| | $Q_{00}$ | | $Q_{10}$ | | $Q_{11c}$ | | $Q_{11s}$ | |
|---|---|---|---|---|---|---|---|---|
| | $\overline{\lvert Q\rvert}$ | $\overline{\lvert\Delta Q\rvert}$ | $\overline{\lvert Q\rvert}$ | $\overline{\lvert\Delta Q\rvert}$ | $\overline{\lvert Q\rvert}$ | $\overline{\lvert\Delta Q\rvert}$ | $\overline{\lvert Q\rvert}$ | $\overline{\lvert\Delta Q\rvert}$ |
| Nitrogen | | | | | | | | |
| N1 | 1.54 | 0.0025 | 0.19 | 0.0055 | 0.14 | 0.0034 | 0.11 | 0.0050 |
| N17 | 1.55 | 0.0031 | 0.12 | 0.0042 | 0.14 | 0.0042 | 0.13 | 0.0044 |
| Oxygen | | | | | | | | |
| O10 | 1.40 | 0.0020 | 0.31 | 0.0020 | 0.44 | 0.0025 | 0.24 | 0.0022 |
| O16 | 1.41 | 0.0027 | 0.36 | 0.0017 | 0.42 | 0.0056 | 0.23 | 0.0021 |
| Carbon | | | | | | | | |
| C3 | 0.58 | 0.0029 | 0.02 | 0.0037 | 0.56 | 0.0037 | 0.09 | 0.0039 |
| C5 | 0.22 | 0.0031 | 0.02 | 0.0028 | 0.02 | 0.0032 | 0.03 | 0.0042 |
| C6 | 1.84 | 0.0029 | 0.63 | 0.0034 | 0.34 | 0.0027 | 0.43 | 0.0032 |
| C11 | 1.84 | 0.0024 | 0.33 | 0.0023 | 0.68 | 0.0025 | 0.27 | 0.0025 |
| C12 | 0.16 | 0.0032 | 0.03 | 0.0036 | 0.05 | 0.0038 | 0.02 | 0.0027 |
| C19 | 0.67 | 0.0039 | 0.40 | 0.0050 | 0.25 | 0.0053 | 0.29 | 0.0052 |
| C–Hydrogen | | | | | | | | |
| H4 | 0.04 | 0.0030 | 0.08 | 0.0009 | 0.03 | 0.0012 | 0.05 | 0.0011 |
| H7 | 0.05 | 0.0022 | 0.01 | 0.0007 | 0.11 | 0.0005 | 0.01 | 0.0007 |
| H8 | 0.05 | 0.0018 | 0.08 | 0.0006 | 0.03 | 0.0010 | 0.06 | 0.0007 |
| H9 | 0.05 | 0.0024 | 0.08 | 0.0012 | 0.02 | 0.0013 | 0.06 | 0.0010 |
| H13 | 0.03 | 0.0028 | 0.04 | 0.0007 | 0.03 | 0.0010 | 0.08 | 0.0009 |
| H14 | 0.02 | 0.0022 | 0.07 | 0.0008 | 0.05 | 0.0008 | 0.04 | 0.0009 |
| H15 | 0.02 | 0.0024 | 0.06 | 0.0007 | 0.06 | 0.0007 | 0.05 | 0.0007 |
| H20 | 0.04 | 0.0025 | 0.05 | 0.0008 | 0.06 | 0.0011 | 0.05 | 0.0011 |
| H21 | 0.03 | 0.0028 | 0.05 | 0.0011 | 0.05 | 0.0012 | 0.05 | 0.0013 |
| H22 | 0.03 | 0.0026 | 0.04 | 0.0011 | 0.05 | 0.0012 | 0.06 | 0.0010 |
| N–Hydrogen | | | | | | | | |
| H2 | 0.45 | 0.0025 | 0.11 | 0.0005 | 0.07 | 0.0005 | 0.09 | 0.0004 |
| H18 | 0.46 | 0.0023 | 0.09 | 0.0005 | 0.08 | 0.0004 | 0.08 | 0.0004 |

| | $Q_{20}$ | | $Q_{21c}$ | | $Q_{21s}$ | | $Q_{22c}$ | | $Q_{22s}$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\overline{\lvert Q\rvert}$ | $\overline{\lvert\Delta Q\rvert}$ | $\overline{\lvert Q\rvert}$ | $\overline{\lvert\Delta Q\rvert}$ | $\overline{\lvert Q\rvert}$ | $\overline{\lvert\Delta Q\rvert}$ | $\overline{\lvert Q\rvert}$ | $\overline{\lvert\Delta Q\rvert}$ | $\overline{\lvert Q\rvert}$ | $\overline{\lvert\Delta Q\rvert}$ |
| Nitrogen | | | | | | | | | | |
| N1 | 0.34 | 0.0120 | 0.15 | 0.0097 | 0.58 | 0.0120 | 0.59 | 0.0110 | 0.27 | 0.0140 |
| N17 | 0.36 | 0.0086 | 0.36 | 0.0081 | 0.41 | 0.0093 | 0.37 | 0.0092 | 0.51 | 0.0087 |
| Oxygen | | | | | | | | | | |
| O10 | 0.06 | 0.0040 | 0.08 | 0.0039 | 0.08 | 0.0041 | 0.08 | 0.0049 | 0.10 | 0.0044 |
| O16 | 0.07 | 0.0082 | 0.03 | 0.0038 | 0.10 | 0.0028 | 0.12 | 0.0097 | 0.05 | 0.0035 |
| Carbon | | | | | | | | | | |
| C3 | 0.21 | 0.0065 | 0.04 | 0.0052 | 0.07 | 0.0065 | 0.48 | 0.0075 | 0.08 | 0.0064 |
| C5 | 0.06 | 0.0061 | 0.04 | 0.0043 | 0.05 | 0.0062 | 0.04 | 0.0062 | 0.04 | 0.0059 |
| C6 | 0.16 | 0.0041 | 0.37 | 0.0037 | 0.19 | 0.0029 | 0.32 | 0.0031 | 0.33 | 0.0030 |
| C11 | 0.34 | 0.0020 | 0.22 | 0.0032 | 0.34 | 0.0024 | 0.26 | 0.0027 | 0.11 | 0.0034 |
| C12 | 0.06 | 0.0070 | 0.07 | 0.0062 | 0.07 | 0.0068 | 0.10 | 0.0062 | 0.06 | 0.0065 |
| C19 | 0.17 | 0.0078 | 0.26 | 0.0080 | 0.31 | 0.0085 | 0.10 | 0.0088 | 0.17 | 0.0094 |
| C–Hydrogen | | | | | | | | | | |
| H4 | 0.10 | 0.0012 | 0.06 | 0.0015 | 0.13 | 0.0014 | 0.02 | 0.0016 | 0.05 | 0.0019 |
| H7 | 0.12 | 0.0012 | 0.03 | 0.0008 | 0.01 | 0.0010 | 0.18 | 0.0010 | 0.03 | 0.0011 |

**Table 2** continued

| | $Q_{20}$ | | $Q_{21c}$ | | $Q_{21s}$ | | $Q_{22c}$ | | $Q_{22s}$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\overline{\lvert Q\rvert}$ | $\overline{\lvert \Delta Q\rvert}$ | $\overline{\lvert Q\rvert}$ | $\overline{\lvert \Delta Q\rvert}$ | $\overline{\lvert Q\rvert}$ | $\overline{\lvert \Delta Q\rvert}$ | $\overline{\lvert Q\rvert}$ | $\overline{\lvert \Delta Q\rvert}$ | $\overline{\lvert Q\rvert}$ | $\overline{\lvert \Delta Q\rvert}$ |
| H8 | 0.10 | 0.0009 | 0.12 | 0.0011 | 0.15 | 0.0006 | 0.02 | 0.0008 | 0.05 | 0.0007 |
| H9 | 0.12 | 0.0019 | 0.10 | 0.0018 | 0.14 | 0.0014 | 0.03 | 0.0016 | 0.04 | 0.0013 |
| H13 | 0.09 | 0.0010 | 0.05 | 0.0018 | 0.10 | 0.0009 | 0.12 | 0.0017 | 0.07 | 0.0011 |
| H14 | 0.11 | 0.0008 | 0.10 | 0.0010 | 0.08 | 0.0008 | 0.07 | 0.0012 | 0.09 | 0.0012 |
| H15 | 0.08 | 0.0008 | 0.11 | 0.0010 | 0.10 | 0.0007 | 0.06 | 0.0010 | 0.09 | 0.0011 |
| H20 | 0.08 | 0.0014 | 0.09 | 0.0013 | 0.08 | 0.0011 | 0.09 | 0.0012 | 0.10 | 0.0013 |
| H21 | 0.08 | 0.0015 | 0.08 | 0.0016 | 0.09 | 0.0015 | 0.87 | 0.0015 | 0.85 | 0.0014 |
| H22 | 0.08 | 0.0013 | 0.08 | 0.0010 | 0.07 | 0.0012 | 0.09 | 0.0013 | 0.10 | 0.0012 |
| N–Hydrogen | | | | | | | | | | |
| H2 | 0.02 | 0.0007 | 0.01 | 0.0007 | 0.02 | 0.0006 | 0.01 | 0.0008 | 0.01 | 0.0006 |
| H18 | 0.01 | 0.0005 | 0.01 | 0.0007 | 0.01 | 0.0007 | 0.01 | 0.0004 | 0.01 | 0.0006 |

All values are in the appropriate atomic units (i.e. e for $Q_{00}$, eBohr for $Q_{1m}$ and eBohr$^2$ for $Q_{2m}$)

**Table 3** Contributions of all multipole moment models of each atom individually to the absolute error in the total 1–4 and higher electrostatic energy prediction $\overline{|\Delta E|}$

| | $\overline{\lvert \Delta E\rvert}$ | $\lvert \Delta E\rvert_{\min}$ | $\lvert \Delta E\rvert_{\max}$ | $\sigma$ |
|---|---|---|---|---|
| N1 | 2.31 | 0.054 | 11.27 | 2.25 |
| H2 | 1.55 | 0.0030 | 6.78 | 1.35 |
| C3 | 0.69 | 0.010 | 2.25 | 0.54 |
| H4 | 1.51 | 0.10 | 4.99 | 1.16 |
| C5 | 1.13 | 0.024 | 4.25 | 1.06 |
| C6 | 1.38 | 0.0046 | 7.01 | 1.28 |
| H7 | 0.54 | 0.0024 | 1.97 | 0.42 |
| H8 | 0.59 | 0.019 | 2.00 | 0.49 |
| H9 | 0.82 | 0.022 | 8.42 | 1.22 |
| O10 | 0.64 | 0.0077 | 2.96 | 0.60 |
| C11 | 0.11 | 0.0024 | 0.50 | 0.10 |
| C12 | 1.66 | 0.00070 | 7.08 | 1.31 |
| H13 | 2.51 | 0.025 | 12.18 | 2.11 |
| H14 | 2.07 | 0.0072 | 9.38 | 1.86 |
| H15 | 2.27 | 0.042 | 9.07 | 1.92 |
| O16 | 1.70 | 0.0042 | 19.38 | 3.37 |
| N17 | 0.61 | 0.013 | 3.41 | 0.58 |
| H18 | 1.57 | 0.014 | 5.02 | 1.21 |
| C19 | 2.20 | 0.0090 | 8.32 | 1.83 |
| H20 | 1.37 | 0.029 | 6.01 | 1.21 |
| H21 | 1.52 | 0.0061 | 6.11 | 1.26 |
| H22 | 1.43 | 0.024 | 12.17 | 1.76 |

All energies are reported in kJ mol$^{-1}$. The average absolute interaction energy is 83 kJ mol$^{-1}$ and has range from 0.04 to 248.5 kJ mol$^{-1}$

affected by the second influential observation, which is that the number of interactions is not constant for each element–element interaction pair at a specific value of $n$. For example, there are 7 H–H 1–5 interactions, compared to 8 O–H 1–5 interactions in this table. The capping methyl groups are involved in the greatest number of interactions in the molecule, so their absolute error summed over these interactions is much higher. This explains the greater errors observed for H13, H14, H15, H20, H21 and H22 in the AlaD system.

We close this section by some comments on the use of CPU time in the evaluation of the energy with the current approach of high-rank polarisable AMMs. The size of systems treatable by this method depends on the time taken to evaluate the energy. Comparing this to the time for the corresponding charge–charge energy calculation provides a

quantitative measure of the computational expense of including Eq. 5 in a force field. Evaluation of Eq. 5 requires prediction of the AMMs and subsequent evaluation of the interaction tensor for each term in the summation. The latter can be carried out using either explicit formulae [80] or a recurrence relation [81]. Predicting the AMMs requires a set of Cartesian coordinates that are then transformed into internal coordinates in an atomic local axis system. These coordinates are normalised and used in Eq. 7 to produce normalised values for the AMMs, which are subsequently unnormalised. The time-dependence lies in the system size (number of internal coordinates) and the number of training examples used to build a particular model. For the calculations described herein, the average time to produce ($22 \times 9 = 198$) AMMs was 140 ms. Of this, the time to compute internal coordinates was negligible (<0.01%). Eighty percent of the AMM prediction time corresponds to the reading of Kriging model files prior to making predictions. If these files were read and stored in the RAM once, instead of reading them from disk when needed, the cost would be significantly reduced.

Summation of the energy expression with the interaction tensor expressed explicitly, required on average 0.5 ms, whilst the recurrence relation gave an average of 0.2 ms. This is in comparison with calculations carried out using only the monopole moments and the Coulomb law, which average 0.02 ms. The ratio of these values should be machine-independent, and therefore we state inclusion of this method in a force field framework will incur approximately an order of magnitude increase in time required for the evaluation of the electrostatics. However, one should not forget that the energy evaluation is dominated by the monopole–monopole interactions for the systems we eventually apply the proposed method to (i.e. those with thousands of atoms). High-rank multipolar interactions only need to be evaluated at short-range. There is further scope for optimisation of the method, both in moment prediction and in energy evaluation.

## 5 Conclusions

The machine learning method Kriging has been employed in the modelling of the dependence of the QCT atomic multipole moments of the system usually referred to in the literature as alanine dipeptide or AlaD. The Kriging models can be used to compute atom–atom electrostatic interaction energies instead of the point charge models of traditional FF. The inclusion of multipole moments gives a more detailed description of the electrostatics than the point charge model. The use of Kriging allows a natural and direct description of polarisation as the response of an electron distribution to conformational change, avoiding the short-range polarisation catastrophe. The multipole moments are completely determined by the gradient vector field of the electron density, and the use of QCT provides smooth multipole moment surfaces without discontinuities. The method as described is general. It may be applied to any system of interest given a set of internal coordinates that describe the geometry. The underlying ab initio level of theory may be replaced with any other level of theory that allows computation of an accurate wavefunction, with the added cost being purely in CPU time required to generate the data. More multipole moments can be added to each atom to improve the electrostatic description without alteration in the method. For AlaD, the maximum error in the total electrostatic energy is 26 kJ $mol^{-1}$ and the average error is 7 kJ $mol^{-1}$. These values may be improved by consideration of the sampling method used to create the training set and should improve with the use of post-HF and DFT methods to generate the underlying data set.

The function may be employed inside the force field framework with subsequent optimisation of the remaining empirical parameters (e.g. force constants for bond-stretch and angle-bend terms). The computational cost of the method in application is approximately one order of magnitude greater than the point charge method. First derivatives of the function will soon be available, allowing application to molecular dynamics and mechanics experiments.

**Acknowledgments** The authors would like to thank the EPSRC for financial support. Part of the computational element of this research was achieved using the High Throughput Computing (CONDOR) facility of the Faculty of Engineering and Physical Sciences, the University of Manchester.

## References

1. Hegefeld WA, Chen SE, DeLeon KY, Kuczera K, Jas GS (2010) J Phys Chem A 114(47):12391–12402
2. Perez-Angel CE, Seminario JM (2011) J Phys Chem C 115(14):6467–6477
3. Swadling JB, Coveney PV, Greenwell CH (2010) J Am Chem Soc 132(39):13750–13764
4. Car R, Parrinello M (1985) PhysRevLett 55(22):2471–2474
5. Remler DK, Madden PA (1990) Mol Phys 70(6):921–966
6. Warshel A, Levitt M (1976) J Mol Biol 103:227–249
7. Hu H, Yang W (2009) J Mol Struct THEOCHEM 898(1–3):17–30
8. Stone JE, Phillips JC, Freddolino PI, Hardy DJ, Trabuco LG, Schulten K (2007) J Comput Chem 28(16):2618–2640
9. Voelz VA, Bowman GR, Beauchamp K, Pande VS (2010) J Am Chem Soc 132(5):1526–1528
10. Khalili-Araghi F, Jogini V, Yarov-Yarovoy V, Tajkhorshid E, Roux B, Schulten K (2010) Biophys J 98(10):2189–2198
11. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) J Comp Chem 4:187–217
12. Foloppe N, MacKerell AD Jr (2000) J Comp Chem 21(2):86–104
13. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Profetajr S, Wiener P (1984) J Am Chem Soc 106:765–784

14. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) J Am Chem Soc 117:5179–5197
15. Oostenbrink C, Villa A, Mark AE, van Gunsteren WF (2004) J Comput Chem 25(13):1656–1676
16. Jorgensen WL, Swenson CJ (1985) J Am Chem Soc 107:569–578
17. Kaminsky GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) J Phys Chem B 105(28):6474–6487
18. Halgren TA (1996) J Comp Chem 17:490–519
19. Dinur U, Hagler AT (1994) J Comput Chem 15(9):919–924
20. Tuzun RE, Noid DW, Sumpter BG (1997) J Comput Chem 18(14):1804–1811
21. Al-Matar AK, Rockstraw DA (2004) J Comput Chem 25(5):660–668
22. Lifson S, Warshel A (1968) J Chem Phys 49(11):5116–5129
23. Ewig CS, Berry R, Dinur R, Hill JR, Hwang MJ, Li H, Liang C, Maple J, Peng Z, Stockfisch TP, Thacher TS, Yan L, Xiangshan N, Hagler AT (2001) J Comput Chem 22(15):1782–1800
24. Halgren TA (1995) J Comput Chem 17(5–6):490–519
25. Banas P, Hollas D, Zgarbova M, Jurecka P, Orozco M, Cheatham TE III, Sponer J, Otyepka M (2010) J Chem Theory Comput 6(12):3836–3849
26. Tong Y, Mei Y, Ji CG, Li YL, Chang GJ, Zhang JZH (2010) J Am Chem Soc 122(14):5137–5142
27. Shaik MS, Devereux M, Popelier PLA (2008) Mol Phys 106:1495–1510
28. Ponder JW, Case DA (2003) Adv Protein Chem 66:27–85
29. Bayly CI, Cieplak P, Cornell WD, Kollman PA (1993) J Phys Chem 97(40):10269–10280
30. Cieplak P, Caldwell J, Kollman P (2001) J Comput Chem 22(10):1048–1057
31. Patel S, Brooks CL III (2004) J Comput Chem 25(1):1–15
32. Rick SW, Stuart SJ (2002) Potential and algorithms for incorporating polarizability in computer simulations. In: Lipkowitz KB, Boyd DB (eds) Reviews in computational chemistry, vol 18. Wiley-VCH, New York, pp 89–146
33. Harder E, Anisimov VN, Vorobyov IV, Lopes PEM, Noskov SY, MacKerell Jr AD, Roux B (2006) J Chem Theory Comput 2:1587–1597
34. Hemmingsen L, Amara P, Ansoborlo E, Field MJ (2000) J Phys Chem A 104:4095–4101
35. Thole BT (1981) Chem Phys 59:341–350
36. Handley CM, Popelier PLA (2010) J Phys Chem A 114:3371–3383. doi:10.1039/b905748j
37. Handley CM, Hawe GI, Kell DB, Popelier PLA (2009) Phys Chem Chem Phys 11:6365–6376. doi:10.1039/b905748j
38. Houlding S, Liem SY, Popelier PLA (2007) Int J Quantum Chem 107(14):2817–2827
39. Behler J, Parrinello M (2007) Phys Rev Letts 98:146401–146404
40. Hobday S, Smith R, Belbruno J (1999) Model Simul Mater Sci Eng 7:397–412
41. Sanville E, Bholoa A, Smith R, Kenny SD (2008) J Phys Condens Matter 20:285219
42. Bartok AP, Payne MC, Kondor R, Csanyi G (2010). Phys Rev Lett 104(13):136403–136406
43. Hawe GI, Alkorta I, Popelier PLA (2010) J Chem Inf Model 50:87–96
44. Mills MJL, Popelier PLA (2011) Comput Theor Chem "Special issue: ESPA 2010": in pages
45. Popelier PLA, Stone AJ (1994) Mol Phys 82:411–425
46. Popelier PLA, Stone AJ, Wales DJ (1994) Farad Discuss 97:243–264
47. Elking DM, Perera L, Duke R, Darden T, Pedersen LG (2010) J Comput Chem 31(15):2702–2713
48. Liem SY, Popelier PLA, Leslie M (2004) Int J Quantum Chem 99:685–694
49. Joubert L, Popelier PLA (2002) Mol Phys 100:3357–3365
50. Rafat M, Popelier PLA (2006) J Chem Phys 124:144102–144108
51. Solano CJF, Pendás AM, Francisco E, Blanco MA, Popelier PLA (2010) J Chem Phys 132:194110
52. Stone AJ (1981) Chem Phys Lett 83(2):233–239
53. Pilme J, Piquemal J-P (2008) J Comput Chem 29:1440–1449
54. Volkov A, Coppens P (2004) J Comput Chem 25:921–934
55. Devereux M, Plattner N, Meuwly M (2009) J Phys Chem A 113(47):13199–13209
56. Ponder JW, Wu C, Pande VS, Chodera JD, Schnieders MJ, Haque I, Mobley DL, Lambrecht DS, DiStasio RAJ, Head-Gordon M, Clark GNI, Johnson ME, Head-Gordon T (2010) J Phys Chem B 114:2549–2564
57. Bader RFW (1990) Atoms in molecules. A quantum theory. Oxford University Press, Oxford
58. Popelier PLA (2000) Atoms in molecules. An introduction. Pearson Education, London
59. Bader RFW, Popelier PLA (1993) Int J Quantum Chem 45(2):189–207
60. Popelier PLA, Bremond EAG (2009) Int J Quantum Chem 109:2542–2553
61. Popelier PLA, Aicken FM (2003) Chem Phys Chem 4:824–829
62. Stone AJ (1996) The theory of intermolecular forces. Clarendon, Oxford
63. Koch U, Popelier PLA, Stone AJ (1995) Chem Phys Lett 228:253–260
64. in het Panhuis M, Popelier PLA, Munn RW, Angyan JG (2001) J Chem Phys 114:7951–7961
65. Krige DG (1951) J Chem Metal Min Soc S Afr 52:119–139
66. Cressie N (1993) Statistics for spatial data. Wiley, New York
67. Plattner N, Meuwly M (2009) J Mol Model 15(6):687–694
68. GAUSSIAN03, Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JAJ, Vreven JT, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2003) Gaussian Inc, Pittsburgh
69. GaussView3.0. (2003) Semichem Inc, Gaussian Inc, Pittsburgh
70. Keith TA (2011) AIMAll. 11.04.03 edn. http://aim.tkgristmill.com
71. Popelier PLA (1996) Comput Phys Commun 93(2–3):212–240
72. Popelier PLA (1994) Chem Phys Lett 228(1–3):160–164
73. Rafat M, Devereux M, Popelier PLA (2005) J Mol Graph Model 24:111–120
74. Marvin (2010) 5.3.1 edn. ChemAxon (http://www.chemaxon.com)
75. Humphrey W, Dalke A, Schulten K (1996) J Mol Graph 14:33–38
76. Stone J (1998) An efficient library for parallel ray tracing and animation. University of Missouri, Rolla
77. Jensen F, Palmer DS (2011) J Chem Theory Comput 7(1):223–230
78. Jensen F (2007) Introduction of computational chemistry, 2nd edn. Wiley, Chichester
79. Su ZW, Coppens P (1994) Acta Cryst A50:636–643
80. Haettig C, Hess BA (1994) Mol Phys 81:813–824
81. Haettig C (1996) Chem Phys Lett 260:341