JCTC Journal of Chemical Theory and Computation

Beyond Point Charges: Dynamic Polarization from Neural Net Predicted Multipole Moments

Michael G. Darley, Chris M. Handley, and Paul L. A. Popelier*

Manchester Interdisciplinary Biocentre (MIB), 131 Princess Street, Manchester M1 7DN, United Kingdom

Received May 15, 2008

Abstract: Intramolecular polarization is the change to the electron density of a given atom upon variation in the positions of the neighboring atoms. We express the electron density in terms of multipole moments. Using glycine and *N*-methylacetamide (NMA) as pilot systems, we show that neural networks can capture the change in electron density due to polarization. After training, modestly sized neural networks successfully predict the atomic multipole moments from the nuclear positions of all atoms in the molecule. Accurate electrostatic energies between two atoms can be then obtained via a multipole expansion, inclusive of polarization effects. As a result polarization is successfully modeled at short-range and without an explicit polarizability tensor. This approach puts charge transfer and multipolar polarization on a common footing. The polarization procedure is formulated within the context of quantum chemical topology (QCT). Nonbonded atom−atom interactions in glycine cover an energy range of 948 kJ mol^{-1}, with an average energy difference between true and predicted energy of 0.2 kJ mol^{-1}, the largest difference being just under 1 kJ mol^{-1}. Very similar energy differences are found for NMA, which spans a range of 281 kJ mol^{-1}. The current proof-of-concept enables the construction of a new protein force field that incorporates electron density fragments that dynamically respond to their fluctuating environment.

1. Introduction

For large systems, ab initio calculations quickly become computationally expensive and even prohibitive when simulations need to be carried out. Nonetheless, force fields dramatically accelerate simulations or even enable them. This is often achieved at the expense of accuracy. The associated potentials must be quantitatively accurate,[1,2] however, if one wishes to study problems such as molecular recognition,[3,4] polymorphism,[5] and protein conformation.[6−8]

Force fields suffer from a reliance on parameter fitting and various a priori simplifications, which restrict the transferability of the potentials. A popular simplification in force field design is the assignment of point charges to atoms. Here, the atomic charge density, which is ultimately responsible for the electrostatic interaction, is boldly assumed to be spherical and replaced by a monopole (moment). Another popular simplification occurs in the treatment of polarization where atomic charges are deliberately and permanently enhanced. Alternatively, a dynamic response to a fluctuating external field is modeled by atomic charges acquiring a companion charge, attached to the atom by a fictitious spring. The importance of accurately modeling polarization is widely accepted, as illustrated by a special issue of this journal, published in 2007, dedicated to this critical problem.[9] Furthermore, in 2004 Gresh et al. demonstrated[10] the need for ab initio distributed multipoles and anisotropic distributed polarizabilities in a study on tetrapeptides. Subsequently, a high level MP2 study[11] on conformations of amino acids highlighted the importance of polarizability (as well as multipole moments) as important factors in the design of new force fields. This finding is echoed in recent ab initio conformational work[12] on seven pilot molecules.

Despite dramatic advances in computer hardware, most force fields still adopt the aforementioned simplifications. For example, CHARMM,[13] AMBER,[14] GROMOS,[15] and OPLS[16] treat atoms as nonpolarizable point charges. In

* Corresponding author e-mail: pla@manchester.ac.uk.

10.1021/ct800166r CCC: $40.75
Published on Web 08/07/2008

general, the assignment of atomic charges is not straightforward, and there are many ways that molecular charge distribution can be separated into atomic components. Most modern force fields rely on fitting individual charges to a molecular electrostatic potential (MEP).[17−20] In their careful and deep analysis of this fitting problem Chirlian and Francl showed[21,22] from singular value decompositions of the least-squares matrices that statistically valid charges cannot be assigned to all the atoms in a given molecule. As a result, atoms in similar chemical environments can be given different fitted charges. Ignoring this careful mathematical analysis, arbitrary penalty functions were introduced[23] on the spurious physical grounds that a "buried" atom would somehow contribute less to a perfectly additive electrostatic potential. Another problem arises if we consider different conformations of the same molecule. Different conformations give different MEPs, and hence different charges may be fitted for the same atoms. Furthermore, the fitting procedure does not account for the internal polarization as the charge density changes with conformation. Most MEP fitting procedures take the mean charge for each atom from a number of conformations[24] or fix the charge of a particular atom type to be the same.[23]

The point charge representation is also inadequate for explaining the relative stability of crystal polymorphs of organic molecules or the structural motifs important for molecular recognition.[3,5] These cases are dependent on the strength and the directionality of intermolecular interactions, the directionality being due to the anisotropic distribution of charge. Because of the inability to represent this anisotropic distribution, many point charge models introduce "off-atom" sites. A prime example of this is the placing of off-atom sites at the "lone-pair" locations about oxygen atoms.[25−28]

A more accurate description of the charge distribution uses multipole moments, which can be regarded as the original coefficients of the series expansion that describes the electrostatic potential. Although elaborate compared to point charges, multipole moments can be expressed compactly (and irreducibly) in terms of spherical harmonics.[29] Multipole moments can be determined in a number of ways. Distributed multipole analysis (DMA)[30] determines the multipole moments from the wave function of a molecule by analyzing the overlap between Gaussian functions. DMA moments are employed in force fields such as AMOEBA[31] or the effective fragment potential (EFP) method.[32] Alternatively, the partitioning scheme of Vigné-Maeder and Claverie[33] is used within the "sum of interactions between fragments ab initio" (SIBFA) potential.[34,35] As a further alternative, the Gaussian electrostatic model (GEM)[36,37] approach, which has evolved from SIBFA, uses a full charge distribution based on a density fitting scheme rather than multipole moments. Another way of partitioning the charge density is by means of Wannier functions.[38,39] They were employed to determine multipole moments in simulations, but they were not directly involved in determining the electrostatic interaction.[40] Moments derived by this method have been implemented in simulations of biomolecules.[41]

In this paper we work with the partitioning method of the quantum theory of "Atoms in Molecules",[42,43] which we consider as part of the quantum chemical topology (QCT) approach, a name we justified in ref 44. The QCT approach partitions the electron density into *finite* topological atoms, which exist in real space. The corresponding atomic multipole moments are then obtained by integration of the appropriate density over the atomic volumes. Note that the DMA method allows multipoles to be centered on non-nuclear sites, such as in the middle of a chemical bond. Naively one would think that this flexibility is missing in the QCT approach in which the multipole expansion site coincides with a nucleus, unless there is a non-nuclear attractor with its own multipole moments. However, thanks to a clear distinction between distribution and partitioning[45] one can shift QCT *atomic* multipole moments to a site away from the nucleus, such as the bond midpoint.[45] Alternative chemically intuitive distributed electrostatic moments can be obtained,[46] using the topology of ELF,[47] a prime component of the QCT approach.

QCT moments have been successfully used in the simulation of liquid HF[48] and liquid water.[49,50] These results were preceded by work that corroborated the success of QCT multipole moments in the reproduction of the electrostatic potential[51] and electrostatic or Coulomb atom−atom interaction[52,53] and the prediction of structure of nucleic acid−base pairs.[54] In this systematic work on QCT potentials, the important issue of the convergence of the multipole expansion was featured strongly, leading to solutions that increased the convergence radius[55,56] or accelerated convergence.[45] We also showed that 1−3 and 1−4 interactions can be expressed as a convergent multipole expansion,[57] putting these interactions on a par with nonbonded interactions ($1 - n$, $n > 4$). Even exchange energy can be favorably expanded in terms of exchange moments, but their transferability remains elusive.[58] A separate study[59] determined the lowest rank necessary to achieve a preset error in atom−atom interaction energy.

In the aforementioned work the atomic charges were kept constant with respect to any changes in the orientation or conformation of the molecules. However, this is not representative of reality. The distribution of electron density shifts in response to changes in the intramolecular and intermolecular interactions as orientations and configurations change. Polarization can be modeled by distributed polarizabilities,[60,61] pioneered within the QCT context by Ángyán et al.,[62] and later applied to study[63] field-induced charge transfer along the hydrogen bonding in the water dimer. Since then we have taken a drastically different route to model polarization, which does not introduce an explicit polarizability tensor. Instead, the polarizability is implicit (and still fully anisotropic), in that a machine learning technique is trained to predict atomic multipole moments from the positions of all other atoms in the molecule. The prediction formulas are analytical and nonlinear. They are stored within a trained neural network and can be retrieved, in principle, although this is typically not done in the neural network literature. Of course, in its convoluted analytical shape, the trained neural net can be exported into a molecular dynamics

package, for example. Second, it should be emphasized that the current approach abandons upfront the picture of an external electric field that causes the polarization. Instead, we embrace the more general picture that electron density fragments change in response to a change of nuclear positions, which cause the field. This field, which applies at very short-range and which can be most inhomogeneous (anisotropic), exerts its influence through the ab initio calculations that generate the electron density in the first place. We only need to focus on the result of the field, whether intra- or intermolecular in nature, without approximating it in any way. We believe that this approach automatically takes care of the nonadditivity of polarization. Furthermore, if atoms are taken out of a sufficiently large environment, their electron density has already been adjusted. Hence, in principle there is no need for an iterative scheme that adjusts the atom and its neighbors using an explicit polarizability tensor.

We published the first example and proof-of-concept of the idea of implicit neural net polarization by means of a molecular dynamics simulation[64] of a polarizable hydrogen fluoride dimer. The machine learning method consists of standard backpropagation (artificial) neural networks (NNs). Fluctuating QCT multipole moments, which expressed the intermolecular polarization, were successfully modeled by NNs. The current paper presents the first example of this novel technique applied to intramolecular polarization, illustrated by the molecules glycine and *N*-methylacetamide (NMA).

NNs mimic the way a brain functions by using an array of interconnected units that pass information between themselves to recognize complex patterns. Through the modification of the strength of the connections between these units, NNs are able to learn functions. NNs have been applied to a number of systems and are able to represent them without any prior knowledge about the form of the potentials which govern such systems.[65] Instead, NNs are able to learn the true underlying function from a set of pregenerated data. NNs have been used for modeling the Al^{3+} system,[66] a hydrocarbon potential,[67] silicon,[68] the H_3^+ ion,[67] and water.[69,70]

Here we employ NNs to learn the relationship between the atomic multipole moments and the nuclear configuration of its environment (which is the whole molecule for both glycine and NMA). This makes possible a dynamic representation of the electron density, able to react to changes in the local environment. This eliminates the need for ad hoc corrections[32,71] that account for the overlap of electron densities,[72] an inevitable problem when using fixed densities. Additionally, this NN approach allows us to incorporate the effect of charge transfer into the prediction of the multipole moments. Polarization and charge transfer are now treated on a par with traditional electrostatic interactions, as a single dynamic electrostatic term.

2. Background

2.1. Polarization. The effect of polarization is not negligible and is often quoted as accounting for ~15% of the total interaction energy,[73,74] with Yu and van Gunsteren quoting a range of 10−50% of the total interaction energy.[75] Including polarization allows for the fluctuating anisotropic nature of atomic electron densities to be correctly modeled, which is important for the calculation of accurate interaction densities and the accurate determination of the relative stabilities of molecular conformations,[76] polymorphs, and molecular clusters.[77] The importance of polarization can be highlighted by the dipole enhancement of water molecules transferred from the gas to the bulk phase.[78−80]

There are a number of methods for adding explicit polarization to potentials, but the three main methods are polarizable point dipoles, fluctuating charge models, and Drude oscillators.

Polarizable point dipoles are represented by two point charges. These dipoles then interact via a tensor just as static multipole moments do. The magnitude of the dipole is a response to an external field and is calculated iteratively as molecules respond to mutual changes in charge distribution. However, there is a danger of a "polarization catastrophe", which results from the polarization amplifying itself, causing the interaction energy to become infinite. In the AMOEBA model this is prevented by using a Thole-type damping function.[31,81−85] Within the SIBFA model, point dipoles are distributed about the molecules and are placed at lone pair centers and bond barycenters rather than nuclei.[86] This procedure is analogous to the placement of point dipoles in the EFP model.[87] In contrast to AMOEBA, the SIBFA model tackles the polarization catastrophe by introducing a Gaussian screening of the field. Piquemal et al.[88] re-evaluated the SIBFA polarization compared to its counterparts. Overpolarization at short distances can also be overcome by a new method parametrizing a polarizable potential based on Car−Parrinello simulations.[89]

Fluctuating charge models,[90] or charge equalization, iteratively modify atomic charges in response to an external field. The TIP4P-FQ model of water[91,92] is an early application of this method. Its main disadvantage is that it only accounts for isotropic polarization.

Drude oscillators,[93] or charge-on-spring models, represent polarization by two point charges that are tethered by a harmonic spring potential. One of the charges is located at a fixed position while the other is free to move in response to an external field. This method allows for an anisotropic polarization response.[94−96]

Another non-negligible energy component that is not explicitly represented in nonpolarizable models is charge transfer.[97,98] Charge transfer is the partial transfer of charge from a donor to an acceptor molecule, or between atoms, and so affects the electrostatic interactions of the atoms. Thus charge transfer is a more extreme case of polarization.[99] To the best of our knowledge, the only polarization scheme that is currently able to incorporate charge transfer is the FQ approach, which allows the atomic charges to change in response to the chemical environment. Other models account for the effect of charge transfer by including a further component in the potential.[34−36,100] There is an explicit representation of the charge transfer term in the context of SIBFA going back to 1982.[101] Chen and Martínez made the

important observation that the electronegativities that determine the charge transfer are geometry dependent.[99]

2.2. Quantum Chemical Topology and Electrostatic Interaction. Topological atoms are naturally carved out by gradient paths in the electron density. These paths of steepest ascent typically originate at infinity and terminate at a nucleus. They create surfaces that bound the electron density that is then naturally allocated to each atom. An important feature of topological atoms is that they do not overlap and that they exhaust space (i.e., leave no gaps between them). Atomic multipole moments are obtained by integrating the appropriate property density over the atomic volume. The property density is the total charge density multiplied by a regular spherical harmonic,[102] for example, $^1/_2(3z^2 - r^2)$ for one of the five quadrupole moments.

The moments of atom A and B are designated as $Q_{l_A}m_A(\Omega_A)$ and $Q_{l_B}m_B(\Omega_B)$, respectively, where the index l refers to the rank of the multipole moment and m to the component. Each atom has its own local axis system, centered on its nucleus. The Coulomb interaction between two atoms is then given by[52]

$$E^{AB} = \sum_{l_A=0}^{\infty}\sum_{l_B=0}^{\infty}\sum_{m_A=-l_A}^{l_A}\sum_{m_B=-l_B}^{l_B} T_{l_A m_A l_B m_B}(\mathbf{R})\, Q_{l_A m_A}(\Omega_A)\, Q_{l_B m_B}(\Omega_B) \quad (1)$$

where $T(\mathbf{R})$ is the interaction tensor and $\mathbf{R}$ the vector linking the nuclear positions of the respective atoms (i.e., the origins of their local frames). The terms of eq 1 can be collected according to powers of $R = |\mathbf{R}|$ by means of a rank called L, which is defined as $l_A + l_B + 1$. For example, R^{-3} dependence is made up of interactions between two dipole moments and between a monopole moment and a quadrupole moment. The convergence of the multipole expansion can be monitored against a varying rank L. Using Hättig's recurrence formula[103] for the interaction tensor we can generate expansions up to arbitrarily high rank. The exact interaction energy can be obtained via a six-dimensional (6D) integration over the two participating atoms,

$$E^{AB} = \int_{\Omega_A} d\mathbf{r}_A \int_{\Omega_B} d\mathbf{r}_B \frac{\rho_{tot}(\mathbf{r}_A)\,\rho_{tot}(\mathbf{r}_B)}{r_{AB}} \quad (2)$$

where r_{AB} is the distance between two infinitesimally small charge elements and ρ_{tot} is the total charge density.

2.3. Neural Networks. NNs, which have been the subject of interest[104−106] for many years, are a well-researched and popular example of a machine learning technique. We can only give a brief account here and refer to the Appendix and the citations above for further detail. A NN is an array of connected *nodes*, which pass information between themselves. Each node receives a number of inputs and sends an output. The nodes sum their inputs, which are individually multiplied by the relevant weights, and pass this sum through a *transfer function,* which gives the output. It is the alteration of these weights that allows a NN to learn functions.

The architecture of a network is defined by the number of hidden layers and the number of nodes in the input, output, and hidden layers. The hidden layer contains *hidden nodes*, so-called because one does not have direct access to their outputs for the purpose of training. Hence, they must develop their own representation of the input.[105] The hidden layer enables the network to learn complex tasks by extracting progressively more meaningful features from the input patterns. One of the simplest is a single hidden layer feedforward network (see Figure 1). In a feedforward network the nodes only pass information to the next layer and not back to a previous layer. To learn a mapping between input and output patterns, the NN is presented with training data. During the learning process the NN errors in predicting (i.e., reproducing) the values of the training data are used to alter the weights. This is called the *backpropagation of errors* method. More details are given in the Appendix. The process is repeated for every example in the training set before beginning again, with each full pass of the training set being called an *epoch*.

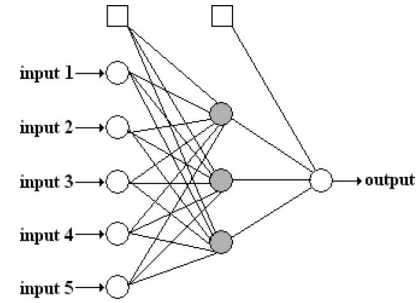

Figure 1. Diagram of a feedforward NN with one hidden layer. The gray circle is a hidden node, a square is a bias, and each connection represents an adjustable weight. In this work, the output is a multipole moment of a given atom and the inputs are the coordinates of neighboring atoms.

Each neuron receives a number (p) of inputs. Each input (signal) is associated with a weight, which can be positive (excitatory) or negative (inhibitory). The activation of neuron k, denoted by a_k, is then defined as the sum of the products of the input and the corresponding weights w_{kj}, or

$$a_k = \sum_{j=0}^{p} w_{kj} x_j \quad (3)$$

Note that, in general, the weight w_{kj} is associated with the connection between neuron j and neuron k. The actual output of a neuron depends on its activation, which has to exceed a given threshold θ for the neuron to fire. Here, the sigmoid function σ acts as a nonlinear transfer function determining how a neuron's output depends on its activation. It is defined in eq 4,

$$y = \sigma(a) = \frac{1}{1 + \exp[-(a - \theta)/\rho]} \quad (4)$$

where ρ controls the shape of the sigmoid. In this work, the output is a multipole moment of a given atom and the inputs are coordinates of neighboring atoms.

To achieve the optimal NN for the prediction multipole moments the architecture of the NN is modified, in this case, only by varying the number of hidden nodes. The net's performance can also be improved by tuning two training parameters, the learning rate and the momentum.[105] Before training the input data must be standardized (see Appendix),

that is, transformed to dimensionless data that have a mean value of zero and a standard deviation of one. The data are then transformed to lie in the interval [0, 1] (i.e., normalized) via eq 5,

$$x_n = \frac{(x - x_{\min})}{(x_{\max} - x_{\min})} \tag{5}$$

where x is a standardized input, x_n is a normalized input, $x_{\min}$ and $x_{\max}$ are the minimum and maximum values, respectively, and the subscript n refers to the normalization.

When training the NN we must consider its performance using statistical measures. We make use of the r^2 correlation coefficient, which measures the linear relationship between the predicted output and the desired output, defined in eq 6,

$$r^2 = 1 - \left[\frac{\sum_{j=1}^{N} (a_j - b_j)^2}{\sum_{j=1}^{N} \left(a_j - \left(\frac{1}{N} \sum_{j=1}^{N} a_j \right) \right)^2} \right] \tag{6}$$

where a_j is the target output, b_j the predicted output, and N the number of training examples. The denominator is reminiscent of a standard deviation. An analogous test is performed using a test set of data, unseen by the network during its training. Here the test set has the same size as the training set. The corresponding correlation coefficient is referred to as q^2.

When training a network one should be concerned about generalization, which covers both overfitting and overtraining. Overfitting means that a net has too much flexibility and thereby inappropriately accommodates all the noise and intricacies in the data without regard for the underlying trends.[105] This undesirable effect can be detected if q^2 turns out to be much lower than r^2. Overtraining can occur if the training continues for too many epochs. It is then possible that the net offers no predictive ability for examples that it was not trained with. Rather than monitoring the number of epochs, overfitting can again be detected by q^2 being much lower than r^2. In summary, a properly generalized net (i.e., not suffering from overfitting or overtraining) has an r^2/q^2 ratio close to unity.

3. Computational Details

To properly calculate the interaction energies we must orient the anisotropic atoms correctly. For a given atom we define an atomic local frame (ALF), which is determined by the connectivity of this atom. Next, the multipole moments of this atom are defined with respect to its ALF. This orientation procedure is adopted during the preparation of the training data as well as during the deployment of the NNs as a way to align the moments by the neural nets in an energy minimization or a simulation. The ALF defines the rotation of the predicted moments into the global frame in which the molecule resides. The NNs knowledge of the local chemical environment is stored in the ALF. In summary, the NN both gains information from the ALF and predicts moments within this ALF, in a unified and consistent way.

The method for determining an ALF uses the Cahn−Ingold−Prelog rules for determining the absolute configuration of a chiral center. Note that we do *not* use these rules to determine the chirality of our atoms; we only adopt that part of the rules that ranks groups (attached to a given atom) according to priority. The central atom, whose moments are being predicted, defines the origin of the ALF. The x-axis is determined by the atom with the highest atomic number neighboring the central atom. The xy-plane is then determined by the neighboring atom with the next highest atomic number. In the case of ambiguity, the Cahn−Ingold−Prelog rules are able to decide which atom acquires priority by inspection of the atomic number of the next or more distant atoms. For terminal atoms the process is simpler because the atom has only one connected neighbor, and hence this atom defines the x-axis. To define the xy-plane we inspect the atoms connected to the x-axis atom and, according to the above rules, again determine the atom with the highest atomic number. This atom defines the xy plane. This procedure is illustrated in Figure 2.

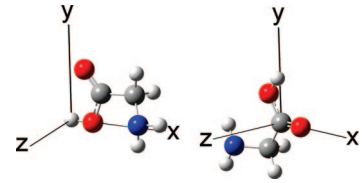

Figure 2. Glycine molecule rotated into two atomic local frames. The left orientation is an example of an atomic local frame for a terminal atom. The right orientation is an example of an atomic local frame for a nonterminal atom.

Once the ALF is defined, the training data are generated. The molecule is geometry- optimized to a local minimum and the corresponding frequencies of the normal modes of vibration are calculated. Then the molecule is distorted randomly along the normal modes of vibration. Each normal mode is distorted in turn by a random amount, according to a randomly assigned sign of displacement, either positive or negative. However, before the displacement is applied we ensure that the total amount of energy does not exceed the imposed limit of 20 kJ mol^{-1}. If so, the distortion is performed again. Figure 3 illustrates a set of geometries for glycine and NMA, and their atomic labeling schemes used throughout this article.

The number of molecular configurations (or geometries) needed for the training of the NN depends upon its architecture, in particular, its number of weights. As a general rule, approximately 10 training examples (i.e., molecular geometries) are required for every weight. Equation 7 provides the number of weights, N_{weight}, for a given architecture with N_{input} input nodes, N_{hidden} hidden nodes, and N_{output} output nodes,

$$N_{\text{weight}} = (N_{\text{input}} + 1)N_{\text{hidden}} + (N_{\text{hidden}} + 1)N_{\text{output}} \tag{7}$$

One adds 1 before multiplication to account for a bias weight. Note that bias nodes only feed connections forward; they do not accept connections. In this study, $N_{\text{input}} = 24$ because glycine has $3N - 6 = 3 \times 10 - 6 = 24$ internal degrees of freedom, and all these degrees are taken to influence the multipole moments of a given atom. Invariably, each multipole moment constitutes the single output for each

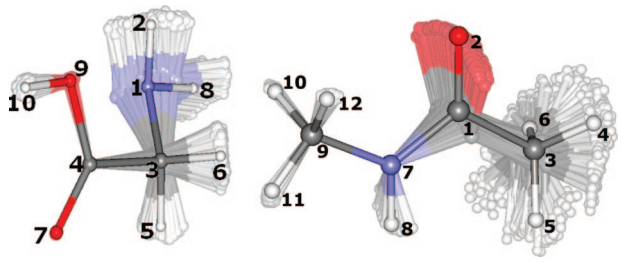

Figure 3. Representative sets of glycine (left) and *N*-methylacetamide (right) molecules distorted along their normal modes of vibration. The reference (local) minimum energy geometry is shown in bold. In the glycine set C_4 defines the origin, the O_7 the *x*-axis, and O_9 the *xy*-plane. For the *N*-methylacetamide set C_9 defines the origin, the N_7 the *x*-axis, and H_{10} the *xy*-plane. In the distorted structures only the atom that defines the origin is in exactly the same position in each structure, while the atoms that define the *x*-axis all have slightly different positions on the *x*-axis due to the random distortion applied to the bond. Similarly, the atoms which define the *xy*-plane all have different positions in the *xy*-plane.

NN, hence, $N_{output}=1$. Equation 7 then specializes to $N_{weight} = 26N_{hidden} + 1$. Because the maximum number of hidden nodes in this work is 10, the maximum number of weights is 261 and hence an adequate number of training examples is 2610. We generated enough training examples to accommodate the largest NN architecture we use.

Multipole moments up to the quadrupole are predicted using the NNs. Higher order moments are not predicted but instead are taken as fixed from the reference geometry (which is local energy minimum). Each moment is predicted by a single NN so there are nine NNs per atom: one for the monopole, three for the dipole, and five for the quadrupole moment.

The NN used to predict the atomic moments is selected using the r^2/q^2 ratio and the NN parameters. The latter refer to the number of hidden nodes, the momentum, and the learning rate. The NN that has an r^2/q^2 ratio closest to unity is selected regardless of that NN's architecture. However, if two architectures have the same r^2/q^2 ratio, then the NN with the smallest number of nodes is selected. If the NNs have the same r^2/q^2 ratio and number of hidden nodes, then the one with the lowest momentum is selected.

The program GAUSSIAN03[107] geometry-optimized glycine and NMA, and calculated the frequencies for the normal modes of vibration. The frequencies are then used to guide the generation of the distorted structures. An energy limit of 20 kJ mol^{-1} is imposed, and the molecule is randomly distorted along the normal modes. This maximum distortion energy is set so that the generated geometries are plausible. The wave functions of each of the distorted geometries are generated and are then used by the program MORPHY[108] to compute the atomic multipole moments.

For this pilot study, all optimizations, frequency, and wave function calculations were performed at B3LYP/6-311+G(2d,p) level. The proposed method is purely based on the electron density and hence does not depend on the details of how this density was generated.

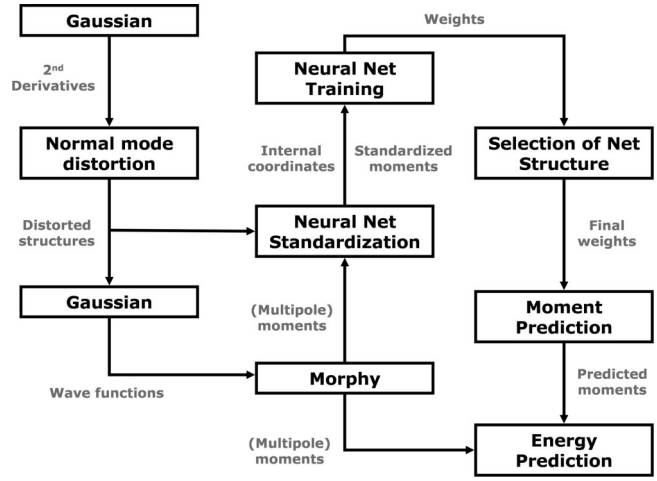

Figure 4. Steps used to create the data for NN training and moment prediction. Black text in boxes represents programs that produce output used by other programs. Grey text represents the output of a program step that is used as input in following programs.

We generated 2610 distorted geometries for each atom with 1305 geometries for NN training and 1305 as a test set. A further 50 distorted molecules were kept in the global frame (i.e., not rotated into an ALF). This set served as a validation set. One could assess the NN's performance by means of a correlation coefficient along the lines of eq 6. However, it is more informative to invoke the interaction energy between a given atom and another atom to assess the NN. For the validation set the *true* moments are calculated from the wave functions and then used to calculate the interaction energy between a subset of all possible atom pairs in glycine. This subset consists of 23 pairs, which is the total number of pairs (45 = 10(10-1)/2) minus all 9 bonded (1-2) interactions and all 13 valence angle (1-3) interactions. In other words, only the remaining 23 interactions of the type 1-4 and 1-*n* ($n \geq 4$) were considered. For glycine, these 23 *true* interaction energies are compared with the interaction energies calculated using the *NN predicted* moments. For NMA, this comparison involved 37 atom pairs.

Figure 4 outlines the data generation procedure. The boxes represent steps that produce output used in the following steps. The first step is the optimization of the molecule using GAUSSIAN, which also computes the second derivatives of the optimized structure. The latter are used to generate the distorted structures in the second step. A single-point calculation is performed for each distorted structure and the corresponding wave function file passed on the program MORPHY. The multipole moments of all atoms for each distorted structure are calculated by MORPHY. The Cartesian coordinates are converted into internal coordinates, which are then standardized and transformed (according to eq 5) to attain a value between 0 and 1. The same transformations are applied to the multipole moments. It is these transformed internal coordinates and moment values that are used as input for the NN training. We train a number of nets, each with a different number of hidden nodes, learning rate, and momentum. The net that gives the r^2/q^2 value closest to 1 is used for the moment prediction (see above for refinements). Along with the NN predicted

Table 1. (a) Optimization of the NN Parameters Used To Predict the Monopole Moments of Each Atom in Glycine[a] and (b) NN Architectures Used To Predict the Multiple Moments of C_3

(a)

atom	N_{hidden}[b]	momentum	learning rate	r^2	q^2
C_3	4	0.6	0.05	0.983	0.976
C_4	4	0.7	0.05	0.996	0.995
H_2	9	0.6	0.05	0.999	0.999
H_5	4	0.6	0.05	0.993	0.993
H_6	5	0.9	0.10	0.993	0.992
H_8	8	0.8	0.10	0.999	0.999
H_{10}	6	0.9	0.10	1.000	0.983
N_1	4	0.6	0.05	0.998	0.997
O_7	9	0.7	0.05	0.999	0.999
O_9	5	0.8	0.10	0.998	0.998

(b)

C_3	N_{hidden}[b]	momentum	learning rate	r^2	q^2
q	4	0.6	0.05	0.983	0.976
μ_x	4	0.6	0.05	0.932	0.903
μ_y	4	0.7	0.05	0.940	0.911
μ_z	4	0.6	0.10	0.921	0.880
Q20[c]	4	0.7	0.05	0.978	0.975
Q21c	4	0.6	0.05	0.776	0.627
Q21s	4	0.8	0.05	0.868	0.801
Q22c	4	0.8	0.05	0.986	0.981
Q22s	4	0.6	0.05	0.726	0.627

[a] The correlation coefficients r^2 and q^2 are explained in the main text. [b] Number of nodes in the hidden layer (see eq 7). [c] Notation of ref 29 is adopted for the five quadrupole moments.

moments (monopole, dipole, quadrupole), the energy prediction also uses the higher moments (octupole, hexadecapole) of the optimized (reference) structure. In principle, all moments could have been predicted by NNs, but their energy contributions to 1-n ($n \geq 4$) interactions are small and hence do not justify the concomitant computational overhead.

Nets with 4 to 10 hidden nodes are trained, with momenta of 0.6, 0.7, 0.8, or 0.9, and a learning rate of 0.05 or 0.1. Hence, for each net a total of $7 \times 4 \times 2 = 56$ nets were trained. The number of epochs remained fixed at 100 000 for all training calculations.

4. Results and Discussion

Table 1a shows the parameters of the best NNs used to predict the monopole moments of each atom in glycine. The equivalent data for NMA are not shown since they are very similar. The variation in any of the three parameters has relatively little effect on the monopole prediction quality of the NNs gauged by the r^2 values. The correlation between the monopole predicted by the NNs and the "true" moments calculated by MORPHY is above 0.97 for each atom. However, it is important to note that one cannot predict the best performing NN parameters in advance, just based on the local environment of each atom. The observed optimal number of hidden nodes in the carbon NNs of glycine is 4 or 5 while the range for NMA is 4 to 9. The range for hydrogen is 4 to 10 for both glycine and NMA. The number of nodes for nitrogen in glycine and NMA is, respectively, from 4 to 8 and from 4 to 6. The range for the glycine atoms is 4 to 10 for NMA. The range of parameters used to predict

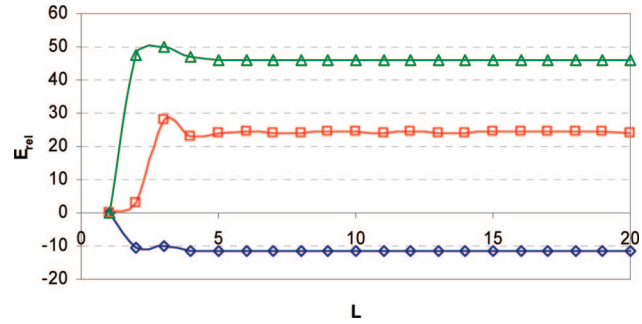

Figure 5. Electrostatic interaction energy (kJ mol^{-1}) of three representative nonbonded interactions in glycine versus rank L; for visual convenience, all energies are relative to the energy at rank $L = 1$ (charge−charge), the largest energy contribution. The N_1−O_9 interaction is marked by red squares, O_7−H_{10} by green triangles, and C_3−H_{10} by blue diamonds.

the moments of each atom in glycine and NMA is similar. Any differences are due to the fact that the local chemical environments of the carbon, nitrogen, and oxygen atoms are not exactly alike.

Table 1b lists the NN parameters used to predict the monopole, dipole, and quadrupole moments of the pivotal α carbon (C_3) in glycine. Each NN has 4 hidden nodes, and 8 out of 9 NNs have a learning rate of 0.05 (the other is 0.1). The monopole moment has the highest r^2 and q^2. Typically, the r^2 or q^2 values decrease as the rank of the moment increases. However, the r^2 and q^2 drops below 0.9 only for some of the quadrupole moments. The Q21c and Q22s moments have the lowest r^2 and q^2 values. Though we do not list the optimal NN parameters of NMA here, their values and the r^2 and q^2 values are similar to those of glycine. In NMA, like in glycine, some quadrupole moments (of C_1) have r^2 and q^2 values below 0.9.

Prediction of multipole moments using NNs and internal coordinates as input is clearly successful, as judged by inspecting the values of the moments themselves. A less direct but physically more valuable way of assessing the performance of the NNs is to look at interatomic interaction energies.[109] The advantage of this assessment is that one can gauge the NN's quality in terms of a single number, namely, energy. The disadvantage is the need to identify other atoms as partners interacting with the central atom for which the NN is set up. Before we analyze the NNs according to errors in energy we must address the technical but paramount issue of convergence control.

A truncated multipole expansion always produces an error compared to the exact interaction energy. Previously we have monitored[110] how energies vary with (expansion) rank L. Factors such as internuclear distance, relative orientation, atomic shape, and the magnitude of the electronic density all determine if the expansion is convergent or not. Comparing the energy profiles calculated from the true multipole moments with the profiles from NN predicted moments reveals how much of the energy differences are due to the prediction error, rather than to an inherent convergence error. To investigate the convergence behavior of the 23 atom−atom interactions in glycine, the energy of each interaction was calculated up to $L = 20$. Figure 5 shows the interaction

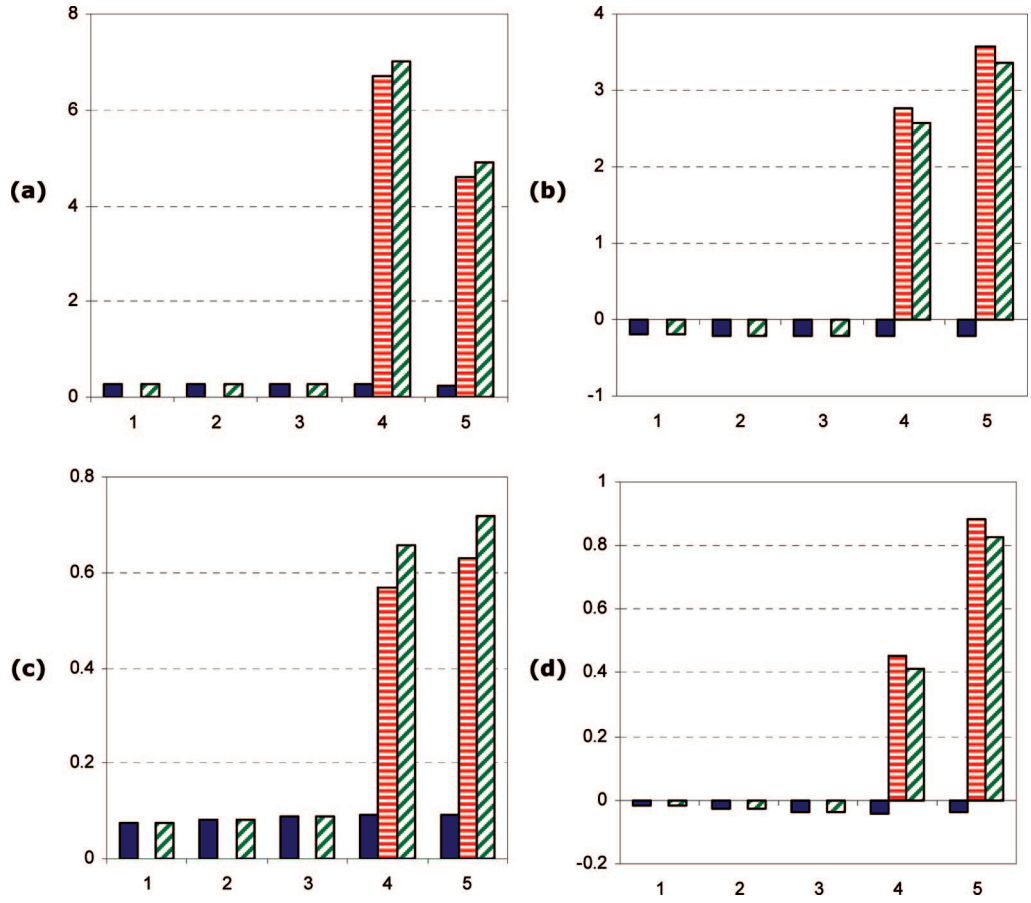

Figure 6. Interaction energy (kJ mol^{-1}) as a function of expansion rank L for four 1-4 atom−atom interactions. (a) N_1-O_9 in glycine, (b) O_7-O_{10} in glycine, (c) O_2-H_8 in NMA, (d) C_3-C_9 in NMA. All energies are relative to the expansion energy calculated using the true moments (monopole, dipole, quadrupole, octupole, and hexadecapole). The blue bar represents the energy calculated from NN-predicted moments (monopole, dipole, and quadrupole) and (fixed) higher moments from the reference geometry. The red bar represents the energy calculated from the *true* monopole, dipole, and quadrupole moments, without higher moments. The green bar is the energy calculated from the NN-predicted moments only (monopole, dipole, and quadrupole), without higher moments.

energy as L is increased for three representative atom−atom interactions. All interactions studied here are convergent including N_1-O_9, the highest energy (501 kJ mol^{-1}) interaction, and O_7-H_{10}, the lowest (−447 kJ mol^{-1}). Of all the interactions studied, the O_7-H_{10} interaction also has the shortest internuclear distance in the reference geometry (2.27 Å).

At rank $L = 1$ the energy represents the charge−charge interaction, $L = 2$ also includes the energy of the charge−dipole interaction, and L=3 includes the charge−quadrupole and dipole−dipole interaction energy. Figure 5 shows how the interaction energy behaves asymptotically beyond $L = 5$, where the change in interaction energy is exceedingly small, less than 0.02 kJ mol^{-1}. This indicates that the interaction has practically converged. Although Figure 5 only shows three energy profiles, these are representative of 1-4 interactions in glycine and NMA. By visual inspection *all* interactions converged in glycine and NMA.

Figure 6 shows the interaction energy as a function of expansion rank L for four 1-4 atom−atom interactions, namely, N_1-O_9 (2.73 Å, termini) in glycine, O_7-H_{10} (2.27 Å, carboxy group) in glycine, and O_2-H_8 (3.11 Å, peptide bond) and C_3-C_9 (3.79 Å, C_α-C'_α) in NMA. The blue bars in the histograms demonstrate how well the NNs predict the monopole, dipole, and quadrupole moments. The bars indicate the energy difference between the true energy (obtained with all true moments up to hexadecupole) and the energy calculated from the NN predicted moments (with true octupole and hexadecupole). All energy differences are of the order of tenths of a kJ mol^{-1} or less. They vary very little with the expansion rank L. This means that already at $L = 1$ (charge−charge) most of the energy difference is present.

Figure 6 also shows the effect on the energy of removing the higher moments (octupole and hexadecupole), as marked by the red bars (horizontal stripes). They show how the energy deviates from the true one, when using the true monopole, dipole, and quadrupole moment. This deviation can be as high as over 6 kJ mol^{-1} for the N_1-O_9 interaction in glycine, which is reaffirmed in the O_2-C_9 interaction in NMA (not shown). Fortunately, the deviation is about an order of magnitude smaller in the $C_\alpha-C'_\alpha$ interaction in NMA, as well as in the peptide bond (O_2-H_8). The green bars are the equivalent of the red ones, where the true monopole, dipole, and quadrupole moments are replaced by the NN predicted ones. The green bars echo what was concluded based on the red ones.

Now that we have shown that the nonbonded interactions are convergent we can use predicted moments in the multipolar expansion to calculate the interaction energies. We assess the performance of the NN by measuring the discrepancy between true and predicted energy. Each of the 23 (nonbonded) interactions in glycine is monitored by an energy error distribution over the external validation set of 50 distorted molecular geometries. The quality of the 50 predictions for each interaction can be measured by the minimum and maximum error, the error range, and the error averaged over the 50 predictions. This information is given in Table 2 for glycine and for the 37 nonbonded interactions in NMA. The glycine interactions cover an energy range of 948 kJ mol^{-1}. The most attractive interaction is O_7-H_{10} (carboxylic oxygen and hydroxyl hydrogen) with an energy of -447 kJ mol^{-1}, while the most repulsive interaction is between N_1 and O_9 (501 kJ mol^{-1}). The N_1-O_9 interaction has the largest average absolute energy difference (E_{diff}) of 2.7 kJ mol^{-1}. However, this represents a difference of just 0.5% between the true energy (E_{true}) and predicted energy ($E_{predicted}$). Although this interaction has the greatest E_{diff} it is a small fraction of the total interaction energy (E_{int}). The interactions of glycine have an average E_{diff} of 0.2 kJ mol^{-1}, and the largest E_{diff} is just under 1 kJ mol^{-1}. Three interactions have an E_{diff} of 0.5 kJ mol^{-1} or greater, and four have an E_{diff} greater than 0.2 kJ mol^{-1}. A general observation is that as E_{int} increases so does E_{diff}. The range of energy differences ("E_{diff} range" in Table 2) gives an indication of how consistent the energy prediction is. Ideally E_{diff} and the E_{diff} range would both be zero but this is not the case. Again, as E_{diff} increases so does the "E_{diff} range".

The nonbonded interactions of NMA have an energy range of 281 kJ mol^{-1}, which is more than three times smaller than the range for glycine. The average E_{diff} is 0.2 kJ mol^{-1}, which is remarkably similar to glycine's value. The most attractive interaction is O_2-H_8 (located in the peptide bridge), which has an E_{int} of -217 kJ mol^{-1}. The most repulsive interaction is C_1-H_{12} (64 kJ mol^{-1}). Here the interaction with the second highest E_{int} value, O_2-C_9, has the highest E_{diff} value, that is, 1.1 kJ mol^{-1}. The C_3-H_8 interaction has the second highest E_{diff} of 0.5 kJ mol^{-1}. As with glycine NMAs, E_{diff} and "E_{diff} range" increase as the absolute value of E_{int} increases.

Overall, Table 2 shows that the predicted energies are in good agreement with the true energy. Also, the predicted moments can reproduce interactions over a large energy range. For example, the O_7-H_{10} of glycine has an E_{int} of -447 kJ mol^{-1}, but O_7 also interacts with N_1 and this interaction has an energy of 437 kJ mol^{-1}. This corresponds to an energy range of 884 kJ mol^{-1}.

Now we look at *the sum* of all nonbonded interaction energies of glycine and NMA for each of the 50 distorted molecules. Figure 7 shows the absolute energy difference (E_{diff}) between the total expansion and the total predicted nonbonded energy as well as the percentage difference. The total nonbonded interaction energy ("E_{true} total") of the 50 distorted glycine molecules has a range of 143 kJ mol^{-1} (from 229 to 371 kJ mol^{-1}). The total predicted nonbonded interaction energy ("$E_{predicted}$ total") has a range of 143 kJ mol^{-1} (from 227 to 370 kJ mol^{-1}). The average energy difference (E_{diff}) is -0.3 kJ mol^{-1}, and the minimum and maximum E_{diff} values are -4.7 and 4.0 kJ mol^{-1}, respectively. NMA has an average E_{true} of -511 kJ mol^{-1} and a range of 156 kJ mol^{-1} (from -591 to -435 kJ mol^{-1}). The average E_{diff} for each interaction of NMA is 0.2 kJ mol^{-1}. The minimum and maximum E_{diff} values are 0.0 and 3.5 kJ mol^{-1}, respectively.

Table 2. Interaction Energy (E_{int}), Absolute Energy Difference between True and Predicted Interaction Energy (E_{diff}), Averaged over a Set of 50 Distorted Geometries, Minimum and Maximum Absolute Energy Difference, and the Range of Interaction Energy Differences of (a) Glycine and (b) *N*-Methylacetamide Molecules[a]

interaction	E_{int}	E_{diff}	Min E_{diff}	max E_{diff}	E_{diff} range
		(a)			
C3−H10	85.92	0.62	0.07	1.86	1.79
C4−H8	219.27	0.35	0.01	1.44	1.42
C4−H2	273.27	0.54	0.00	1.59	1.59
H2−H10	85.94	0.04	0.00	0.36	0.36
H2−H5	14.86	0.05	0.00	0.15	0.14
H2−H6	9.50	0.07	0.00	0.19	0.19
H2−O7	−157.23	0.06	0.00	0.69	0.69
H2−O9	−208.27	0.22	0.02	1.10	1.08
H5−H10	15.77	0.05	0.00	0.20	0.19
H5−H8	15.83	0.08	0.00	0.24	0.24
H5−O7	−45.93	0.16	0.00	0.60	0.60
H5−O9	−38.03	0.13	0.00	0.48	0.48
H6−H10	7.55	0.06	0.00	0.17	0.17
H6−H8	9.26	0.08	0.00	0.19	0.19
H6−O7	−22.62	0.19	0.01	0.48	0.47
H6−O9	−19.97	0.17	0.00	0.53	0.53
H8−H10	64.34	0.03	0.00	0.16	0.16
H8−O9	−147.92	0.10	0.00	0.38	0.38
N1−H10	−212.53	0.17	0.00	0.59	0.59
N1−O7	437.17	0.36	0.03	1.15	1.12
N1−O9	501.26	0.97	0.01	2.70	2.69
O7−H10	−446.56	0.31	0.02	0.57	0.55
O7−H8	−137.97	0.06	0.00	0.23	0.23
average	13.17	0.21	0.01	0.70	0.69
min	−446.56	0.03	0.00	0.15	0.14
max	501.26	0.97	0.07	2.70	2.69
range	947.82	0.94	0.07	2.56	2.55
		(b)			
C1−H10	43.60	0.34	0.01	1.56	1.54
C1−H11	32.65	0.21	0.00	2.40	2.40
C1−H12	63.93	0.34	0.00	1.51	1.50
C3−C9	−7.60	0.26	0.03	0.93	0.90
C3−H10	−1.04	0.03	0.00	0.21	0.21
C3−H11	−0.78	0.03	0.00	0.15	0.15
C3−H12	−1.60	0.04	0.00	0.25	0.25
C3−H8	−15.23	0.53	0.00	1.64	1.63
H4−C9	8.66	0.05	0.00	0.19	0.19
H4−H10	1.14	0.01	0.00	0.06	0.06
H4−H11	0.86	0.01	0.00	0.10	0.10
H4−H12	1.72	0.01	0.00	0.06	0.06
H4−H8	13.37	0.07	0.00	0.23	0.23
H4−N7	−39.48	0.27	0.01	1.09	1.08
H5−C9	3.94	0.06	0.00	0.22	0.22
H5−H10	0.49	0.01	0.00	0.04	0.04
H5−H11	0.39	0.01	0.00	0.07	0.07
H5−H12	0.74	0.01	0.00	0.07	0.07
H5−H8	5.83	0.13	0.00	0.52	0.52
H5−N7	−18.99	0.38	0.03	1.51	1.49
H6−C9	6.98	0.06	0.00	0.22	0.22
H6−H10	0.91	0.01	0.00	0.05	0.05
H6−H11	0.69	0.01	0.00	0.06	0.06
H6−H12	1.36	0.02	0.00	0.10	0.10
H6−H8	11.19	0.11	0.00	0.42	0.42
H6−N7	−32.52	0.34	0.02	1.20	1.18
H8−H10	13.25	0.09	0.00	0.49	0.49
H8−H11	10.84	0.09	0.00	0.98	0.98
H8−H12	17.63	0.07	0.00	0.23	0.22
O2−C9	−201.58	1.12	0.01	3.49	3.49
O2−H10	−29.48	0.24	0.00	1.21	1.20
O2−H11	−22.30	0.13	0.01	1.75	1.74
O2−H12	−45.29	0.27	0.00	0.94	0.93
O2−H4	−53.31	0.30	0.00	1.13	1.13
O2−H5	−24.25	0.28	0.00	0.97	0.96
O2−H6	−40.84	0.43	0.02	1.58	1.56
O2−H8	−217.36	0.25	0.01	0.82	0.82
average	−13.82	0.18	0.00	0.77	0.76
min	−217.36	0.01	0.00	0.04	0.04
max	63.93	1.12	0.03	3.49	3.49
range	281.29	1.12	0.03	3.45	3.45

[a] All energies in kJ mol^{-1}.

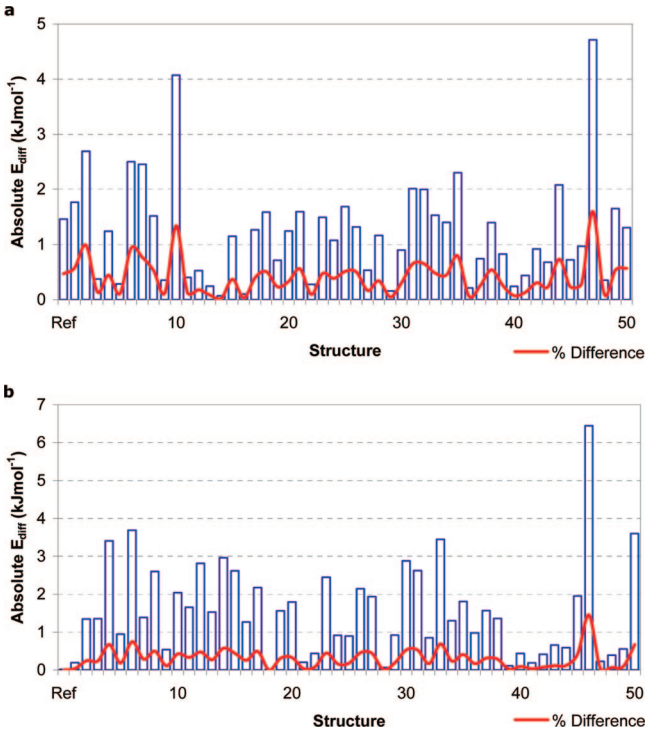

Figure 7. *Total* absolute interaction energy difference of each of the 50 distorted glycine molecules for 1-*n* interactions ($n \geq 4$). The absolute energy difference is $|E_{true} - E_{predicted}|$, where E_{true} is the expansion energy calculated using the true moments and $E_{predicted}$ is the energy calculated using the NN predicted moments. The percentage difference refers to absolute differences.

We initially carried out this work at the HF/6-31G(d) level of theory. The HF method is known to overestimate charge transfer in molecules, witnessed by the exaggerated atomic charges compared to those calculated with Møller−Plesset perturbation theory.[111] We have also observed an increase in QCT charges in going from a B3LYP wave function to a HF one. These increased charges mostly lead to increased interaction energies. For example, the total nonbonded interaction energy for the local reference structure of glycine increases from 310 kJ mol^{-1} using B3LYP moments to 658 kJ mol^{-1} using HF moments, and the energy range increases from 143 kJ mol^{-1} to 250 kJ mol^{-1}. This increase is most noticeable when the individual interactions are considered. The average C_3-H_{10} interaction energy of glycine increases from 86 kJ mol^{-1} to 155 kJ mol^{-1}, a 80% increase, which is the largest change observed. However, not all of the interactions increase in energy. If the interaction has an absolute value less than 50 kJ mol^{-1}, then the interaction increases in magnitude; that is, repulsive interactions become more repulsive and attractive interactions more attractive. However, as these interactions are smaller in magnitude the cumulative effect of using B3LYP moments rather than HF moments is to lower the overall interaction energies for glycine and NMA.

The current method uses NNs to capture the relationship between the local chemical environment of an atom and the polarization of that atom. We now look at how this internal polarization affects the interaction energy. We calculated the energy *including* polarization using the original (i.e., "true") moments calculated for each of the 50 distorted geometries of glycine. These energies are our reference. To calculate the energy *excluding* polarization we use the moments from the local reference structure in place of the true moments of the distorted molecules. The reference moments are rotated into the correct local frame in each molecule before the energy is calculated. The polarization contribution to the total nonbonded energy and to each of the 23 interactions in glycine is shown in Figure 8.

The polarization energy (E_{pol}) shows great variation. In 12 out of 23 interactions E_{pol} is negative; that is, the energy excluding polarization is higher. In 11 interactions E_{pol} is less than 1 kJ mol^{-1}, and only three interactions (four when including the total interactions) have an E_{pol} greater than 5 kJ mol^{-1}. The E_{pol} for the total nonbonded interaction energy is 12.7 kJ mol^{-1} or 4.2%. The origin of 90% of E_{pol} is due to the difference in charge−charge energy between the two atoms. This is seen for each interaction regardless of the types of atoms that are interacting. Hence, interactions between atoms with large charge differences will have the largest E_{pol}. The N_1-O_9 interaction has an E_{pol} of 17.4 kJ mol^{-1} which is the largest E_{pol} for any interaction (including the total energy). For atoms which have similar charges, such as C and H, E_{pol} is small.

The current method has advantages and limitations. An advantage is that the approach is not limited to a single method for predicting the moments. Alternative machine learning methods can be used in place of the backpropagation NNs, which we have started to investigate. Second, the atoms and their polarization are independent of each other. So, adding new atoms or updating existing ones can be done without having to modify the existing atoms. For example, there is no need to retrain the NNs for a carbonyl carbon when retraining the oxygen because the training sets are independent of each other. Moreover, different moments can be predicted by different machine learning techniques. Third, charge transfer, which normally receives a separate treatment compared to dipolar polarization, is now a unified and streamlined part of general multipolar polarization. A current limitation is the CPU cost of the data preparation and training. The exact amount of time needed depends on the number of inputs required to maintain a ratio of 10 examples for each NN weight. Generating the wave functions accounts for ∼25% of the total time needed, calculation of the moments ∼65%, and the NN training the remaining ∼10%. Another current limitation is the lack of optimization of the polarization procedure. Just to demonstrate the proof-of-concept of this novel technique we use all internal coordinates as input for the moment prediction. In some cases this may mean that each atom is given more information than is necessary to correctly predict that atom's moments.

In terms of future work, we have just begun to address these limitations. The ultimate goal is to describe polarizable electrostatics in arbitrary amino acids and peptide bonds. Using all internal coordinates, in light of our ultimate goal, means that the NNs are not transferable to other similar atoms in different molecules. Transferability of atom types is a key feature of force fields, and using transferable NNs is one of our aims. Creation of transferable NNs will require a method

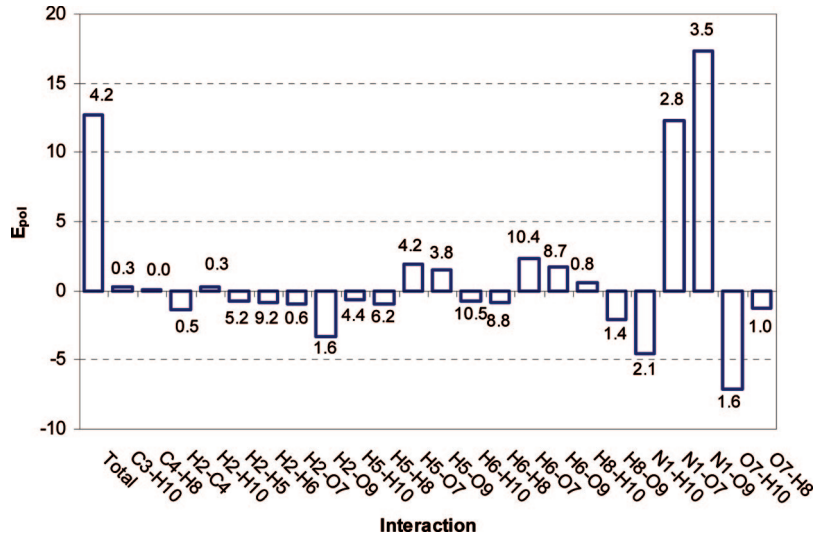

Figure 8. Average energy difference (E_{pol} in kJ mol^{-1}) between the true and the nonpolarized interaction energies of 50 distorted glycine molecules. The numbers above the bars mark the average percentage difference.

for feature selection before the NN training. This is currently being investigated. Also, we have not yet investigated the performance of the predicted moments over the torsional space covered by known protein structures. We are in the process of testing the performance of our method using larger structures. We should also mention that we have obtained successful preliminary results of the current methodology being tested on water clusters. It is important that this methodology will be able to tackle amino acids and water in a unified manner to deliver a reliable protein solvation model.

5. Conclusions

Intramolecular polarization changes an atomic electron density in response to a variation in the positions of the neighboring atoms. We propose a novel method that captures this change directly, by following the change in the atomic multipole moments that express the atomic electron density. Quantum chemical topology provides the partitioning scheme to define the atoms. On the basis of the pilot systems glycine and NMA, we prove that NNs are able to predict with good accuracy the change in atomic monopole, dipole, and quadrupole moments upon changes in molecular configuration. This approach puts charge transfer and dipolar (and high rank multipolar) polarization on a common footing. The current proof-of-concept opens the avenue for a realistic and integrated methodology for a peptide/protein force field, with "informed" atoms that can rapidly and correctly adjust their electronic features to a given environment.

Acknowledgment. We thank the EPSRC for financial support (partially through EP/C015231/1) and GlaxoSmith-Kline (Stevenage, GB) for additional funding.

Appendix

It is convenient to place the adaptation of the threshold on the same footing as that of the weights. This is simply achieved by thinking of the threshold as an extra weight that is driven by an (artificial) input constantly tied to the value −1. This leads to the negative of the threshold being called the *bias*. The transfer function limits the output of the node to a set range, typically between 0 and 1 or between −1 and +1. To modify the weights of the NN is presented a training data set, with known inputs and outputs. The weights of the NN are modified based upon the error between the prediction made for the examples in the training set and the actual desired output. This type of training is known as *supervised learning*. Training continues for a number epochs as the weights are changed and the NN learns from the training set.

The error based upon the output of the final layer of the NN is fed back through the NN, determining the local gradients of the network as we step back through the network. We have to make use of this gradient as we have no knowledge about the true output for each of the hidden layers. This method is known as *back-propagation of errors*. This process is continued for every example in the data set in an effort to produce results converging upon the function that replicates the relationship between the inputs and the outputs. To begin training the network the weights are randomly initialized. The first step is the network performing a forward pass on a training example and finding a final prediction y. From this the error is calculated via eq A1,

$$e_k(n) = d_k(n) - y_k(n) \tag{A1}$$

where $e_k(n)$ is the output error and $d_k(n)$ is the desired output. This error is then sent back through the network,and the local gradient is calculated, if ρ is set equal to 1 in eq 4,

$$\frac{\partial y_k(n)}{\partial a_k(n)} = \frac{\exp(-a_j(n))}{[1+\exp(-a_j(n))]^2} \tag{A2}$$

$$\frac{\partial y_k(n)}{\partial v_k(n)} = y_k(n)[1 - y_k(n)] \tag{A3}$$

$$\delta_k^{(L)}(n) = e_k^{(L)}(n)\frac{\partial y_k(n)}{\partial a_k(n)} \tag{A4}$$

The above is only applicable to the final layer of neurons as $e_k^{(L)}$ cannot be defined for a hidden layer. Instead the local gradient is found via eq A5,

$$\delta_j(n) = \frac{\partial y_j(n)}{\partial a_j(n)} \sum_k \delta_k(n)\, w_{kj}(n) \tag{A5}$$

or it can be expressed as the following equation:

$$\delta_j(n) = y_j(n)[1 - y_j(n)] \sum_k \delta_k(n)\, w_{kj}(n) \tag{A6}$$

The calculation of the local gradients can then be used to determine the changes to the weights.

$$w_{kj}^{(l)}(n+1) = w_{kj}^{(l)}(n) + \alpha[w_{kj}^{(l)}(n) - w_{kj}^{(l)}(n-1)] + \eta \delta_j^{(l)}(n)\, y_i^{(l-1)}(n) \tag{A7}$$

where α is the momentum and η is the learning rate. The learning rate controls the magnitude of the changes made to the weights and has a range between 0 and 1. Large changes to the weights mean that the network is not trapped in local minima on the error surface, though the minima may be missed. However, if smaller changes are used then there is less chance that the true minima is missed, though training is slower. The momentum links changes to the weights to the change that took place previously. This ensures that the weights are changed by the necessary magnitude. For instance, if the two previous weights were of the same sign then a large change is expected; however, if the signs were different then the changes made are small. The momentum term accelerates the gradient descent or stabilizes the learning in regions where the sign oscillates.

Standardization is a transformation of the data to dimensionless data, shown in eq A8,

$$x_{\mathrm{st},i} = \frac{(x_i - \bar{x})}{\sigma_n} \tag{A8}$$

where the standard deviation is $\sigma_n = \sqrt{[(1/N)\sum_{i=1}^{N}(x_i - \bar{x})^2]}$ and $\bar{x}$ is the mean. It is easy to prove that the mean of the standardized data is zero and their standard deviation one.

References

(1) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. A* **2004**, *108*, 621.

(2) Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27.

(3) Price, S. L.; Harrison, R. J.; Guest, M. F. *J. Comput. Chem.* **1989**, *10*, 552.

(4) Lehn, J. M. *Science* **2002**, *295*, 2400.

(5) Price, S. L. *CrystEngComm* **2004**, *6*, 344.

(6) Sokalski, W. A.; Keller, D. A.; Ornstein, R. L.; Rein, R. *J. Comput. Chem.* **1993**, *14*, 970.

(7) Dobson, C. M. *Nature* **2003**, *426*, 884.

(8) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235.

(9) Jorgensen, W. L. *J. Chem. Theory Comput.* **2007**, *3*, 1877.

(10) Gresh, N.; Kafafi, S. A.; Truchon, J.-F.; Salahub, D. R. *J. Comput. Chem.* **2004**, *25*, 823.

(11) Kaminsky, J.; Jensen, F. *J. Chem. Theory Comput.* **2007**, *3*, 1774.

(12) Rasmussen, T. D.; Ren, P.; Ponder, J. W.; Jensen, F. *Int. J. Quantum Chem.* **2006**, *107*, 1390.

(13) Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; Slater, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

(14) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Profetajr, S.; Wiener, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.

(15) van Gunsteren, W. F.; Berendsen, H. J. C. *GROMOS*; Groningen, The Netherlands, 1987.

(16) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.

(17) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129.

(18) Cox, S. R.; Williams, D. E. *J. Comput. Chem.* **1981**, *2*, 304.

(19) Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431.

(20) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.

(21) Chirlian, L. E.; Francl, M. M. *J. Comput. Chem.* **1987**, *8*, 894.

(22) Francl, M. M.; Chirlian, L. E. The pluses and minuses of mapping atomic charges to electrostatic potentials. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley: New York, 2000; Vol. 14; p 1.

(23) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269.

(24) Reynolds, C. A.; Essex, J. W.; Richards, W. G. *J. Am. Chem. Soc.* **1992**, *114*, 9075.

(25) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(26) Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2000**, *112*, 8910.

(27) Nada, H.; van der Eerden, J. P. J. M. *J. Chem. Phys.* **2003**, *118*, 7401.

(28) Vinter, J. G. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 653.

(29) Stone, A. J. *The Theory of Intermolecular Forces*; Clarendon: Oxford, U.K., 1996.

(30) Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233.

(31) Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933.

(32) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J. Phys. Chem. A* **2001**, *105*, 293.

(33) Vigne-Maeder, F.; Claverie, P. *J. Chem. Phys.* **1988**, *88*, 4934.

(34) Gresh, N. *J. Comput. Chem.* **1995**, *16*, 856.

(35) Piquemal, J.-P.; Williams-Hubbard, B.; Fey, N.; Deeth, R.; Gresh, N.; Giessner-Prettre, C. *J. Comput. Chem.* **2003**, *24*, 1963.

(36) Piquemal, J.-P.; Cisneros, G. A.; Reinhardt, P.; Gresh, N.; Darden, T. A. *J. Chem. Phys.* **2006**, *124*, 104101.

(37) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. *J. Comput. Chem.* **2007**, *3*, 1960.

(38) Wannier, G. *Phys. Rev.* **1937**, *52*, 191.

(39) Iannuzzi, M.; Parrinello, M. *Phys. Rev. B* **2002**, *66*, 155209.

(40) Silvestrelli, P. L.; Parrinello, M. *J. Chem. Phys.* **1999**, *111*, 3572.

(41) Sagui, C.; Pomorski, P.; Darden, T. A.; Roland, C. *J. Chem. Phys.* **2004**, *120*, 4530.

(42) Bader, R. F. W. *Atoms in Molecules; A Quantum Theory*; Clarendon Press: Oxford, 1990.

(43) Popelier, P. L. A. *Atoms in Molecules. An Introduction*; Pearson Education: London, U.K., 2000.

(44) Popelier, P. L. A.; Aicken, F. M. *ChemPhysChem* **2003**, *4*, 824.

(45) Joubert, L.; Popelier, P. L. A. *Mol. Phys.* **2002**, *100*, 3357.

(46) Pilme, J.; Piquemal, J.-P. *J. Comput. Chem.* **2008**, *29*, 1440–1449.

(47) Silvi, B.; Savin, A. *Nature (London)* **1994**, *371*, 683.

(48) Liem, S.; Popelier, P. L. A. *J. Chem. Phys.* **2003**, *119*, 4560.

(49) Liem, S.; Popelier, P. L. A.; Leslie, M. *Int. J. Quantum Chem.* **2004**, *99*, 685.

(50) Liem, S. Y.; Popelier, P. L. A. *J. Chem. Theory Comput.* **2008**, *3*, 353.

(51) Kosov, D. S.; Popelier, P. L. A. *J. Phys. Chem. A* **2000**, *104*, 7339.

(52) Popelier, P. L. A.; Joubert, L.; Kosov, D. S. *J. Phys. Chem. A* **2001**, *105*, 8254.

(53) Popelier, P. L. A.; Kosov, D. S. *J. Chem. Phys.* **2001**, *114*, 6539.

(54) Joubert, L.; Popelier, P. L. A. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4353.

(55) Popelier, P. L. A.; Rafat, M. *Chem. Phys. Lett.* **2003**, *376*, 148.

(56) Rafat, M.; Popelier, P. L. A. *J. Chem. Phys.* **2005**, *123*, 204103.

(57) Rafat, M.; Popelier, P. L. A. *J. Chem. Phys.* **2006**, *124*, 144102.

(58) Rafat, M.; Popelier, P. L. A. Topological atom-atom partitioning of molecular exchange energy and its multipolar convergence. In *Quantum Theory of Atoms in Molecules*; Matta, C. F., Boyd, R. J., Eds.; Wiley-VCH: Weinheim, Germany, 2007; Vol. 5; p 121.

(59) Rafat, M.; Popelier, P. L. A. *J. Comput. Chem.* **2007**, *28*, 832.

(60) Garmer, D. R.; Stevens, W. J. *J. Phys. Chem.* **1989**, *93*, 8263.

(61) Stone, A. J. *Mol. Phys.* **1985**, *56*, 1065.

(62) Angyan, J. G.; Jansen, G.; Loos, M.; Haettig, C.; Hess, B. A. *Chem. Phys. Lett.* **1994**, *219*, 267.

(63) in het Panhuis, M.; Popelier, P. L. A.; Munn, R. W.; Angyan, J. G. *J. Chem. Phys.* **2001**, *114*, 7951.

(64) Houlding, S.; Liem, S. Y.; Popelier, P. L. A. *Int. J. Quantum Chem.* **2007**, *107*, 2817.

(65) Lorenz, S.; Gross, A.; Scheffer, M. *Chem. Phys. Lett.* **2004**, *395*, 210.

(66) Gassner, H.; Probst, M.; Lauenstein, A.; Hermansson, K. *J. Phys. Chem. A* **1998**, *102*, 4596.

(67) Prudente, F. V.; Acioli, P. H.; Soares Neto, J. J. *J. Chem. Phys.* **1998**, *109*, 8801.

(68) Behler, J.; Parrinello, M. *Phys. Rev. Lett.* **2007**, *98*, 146401.

(69) No, K. T.; Chang, B. H.; Kim, S. Y.; Jhon, M. S.; Scheraga, H. A. *Chem. Phys. Lett.* **1997**, *271*, 152.

(70) Cho, K.-H.; No, K. T.; Scheraga, H. A. *J. Mol. Struct.* **2002**, *641*, 77.

(71) Piquemal, J.-P.; Gresh, N.; Giessner-Prettre, C. *J. Phys. Chem. A* **2003**, *107*, 10353.

(72) Spackman, M. A. *Chem. Phys. Lett.* **2006**, *418*, 158.

(73) Friesner, R. A. *Adv. Protein Chem.* **2006**, *72*, 79.

(74) Hodges, M. P.; Stone, A. J. *J. Phys. Chem. A.* **1998**, *102*, 2455.

(75) Yu, H.; van Gunsteren, W. F. *Comput. Phys. Commun.* **2005**, *172*, 69.

(76) Rasmussen, T. D.; Ren, P.; Ponder, J. W. *Int. J. Quantum Chem.* **2007**, *107*, 1390.

(77) Halgren, T. A.; Damm, W. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236.

(78) Batista, E. R.; Xantheas, S. S.; Jónsson, H. *J. Chem. Phys.* **1998**, *109*, 4546.

(79) Coulson, C. A.; Eisenberg, D. *Proc. R. Soc. London, Ser. A* **1966**, *291*, 445.

(80) Handley, C. M.; Popelier, P. L. A. *Synth. React. Inorg. Met.-Org. Nano-Met. Chem.* **2008**, *38*, 91.

(81) Thole, B. T. *Chem. Phys.* **1981**, *59*, 341.

(82) Caldwell, J. W.; Kollman, P. A. *J. Phys. Chem.* **1995**, *99*, 6208.

(83) Gao, J.; Habibollazadeh, D.; Shao, L. *J. Phys. Chem.* **1995**, *99*, 16460.

(84) Soteras, I.; Curutchet, C.; Bidon-Chanal, A.; Dehez, F.; Ángyán, J. G.; Orozco, M.; Chipot, C.; Luque, F. J. *J. Chem. Theory Comput.* **2007**, *3*, 1901.

(85) Cisneros, G. A.; Tholander, S. N.; Parisel, O.; Darden, T. A.; Elking, D.; Perera, L.; Piquemal, J.-P. *Int. J. Quantum Chem.* **2008**, *108*, 1905.

(86) Ledecq, M.; Lebon, F.; Durant, F.; Giessner-Prettre, C.; Marquez, A.; Gresh, N. *J. Phys. Chem. B* **2003**, *107*, 10640.

(87) Chen, W.; Gordon, M. S. *J. Chem. Phys.* **1996**, 105.

(88) Piquemal, J.-P.; Chelli, R.; Procacci, P.; Gresh, N. *J. Phys. Chem.* **2007**, *111*, 8170.

(89) Masia, M. *J. Chem. Phys.* **2008**, *128*, 184107.

(90) Patel, S.; Brooks, C. L. *Mol. Simul.* **2006**, *32*, 231.

(91) Bhat, T. N.; Bentley, G. A.; Boulot, G.; Greene, M. I.; Tello, D.; Dall'Acqua, W.; Souchon, H.; Schwarz, F. P.; Mariuzza, R. A.; Poljak, R. J. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 1089.

(92) Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, *101*, 6141.

(93) Harder, E. M. A. V.; Vorobyov, I. V.; Lopes, P. E. M.; Noskov, S. Y.; MacKerell, A. D., Jr.; Roux, B. *J. Chem. Theory Comput.* **2006**, *2*, 1587.

(94) Yu, H.; Hansson, T.; van Gunsteren, W. F. *J. Chem. Phys.* **2003**, *118*, 221.

(95) Yu, H.; van Gunsteren, W. F. *J. Chem. Phys.* **2004**, *121*, 9549.

(96) Yu, H.; Geerke, D. P.; Liu, H.; van Gunsteren, W. F. *J. Comput. Chem.* **2006**, *27*, 1494.

(97) Hemmingsen, L.; Amara, P.; Ansoborlo, E.; Field, M. J. *J. Phys. Chem. A* **2000**, *104*, 4095.

(98) Hagberg, D.; Karlstrom, G.; Roos, B. O.; Gagliardi, L. *J. Am. Chem. Soc.* **2005**, *127*, 14250.

(99) Chen, J.; Martínez, T. J. *Chem. Phys. Lett.* **2007**, *438*, 315.

(100) Gordon, M. S.; Slipchenko, L.; Li, H.; Jensen, J. H. *Annu. Rep. Comput. Chem.* **2007**, *3*, 177.

(101) Gresh, N.; Claverie, P.; Pullman, A. *Int. J. Quantum Chem.* **1982**, *22*, 199.

(102) Popelier, P. L. A. *Mol. Phys.* **1996**, *87*, 1169.

(103) Haettig, C. *Chem. Phys. Lett.* **1996**, *260*, 341.

(104) Vapnik, V. N. *Statistical Learning Theory*; John Wiley: New York, 1998.

(105) Gurney, K. *An Introduction to Neural Networks*; Routledge: London, U.K., 1997.

(106) Haykin, S. *Neural Networks: A Comprehensive Foundation*; Macmillan College Publishing Company: New York, 1994.

(107) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; 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, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, Revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.

(108) Popelier, P. L. A. *MORPHY98*; Manchester, U.K., 1998.

(109) Rafat, M.; Shaik, M.; Popelier, P. L. A. *J. Phys. Chem. A* **2006**, *110*, 13578.

(110) Rafat, M.; Popelier, P. L. A. *J. Chem. Phys.* **2006**, *124*, 144102.

(111) Velders, G. J. M.; Feil, D. *J. Phys. Chem.* **1992**, *96*, 10725.

CT800166R