JCTC Journal of Chemical Theory and Computation

Dynamically Polarizable Water Potential Based on Multipole Moments Trained by Machine Learning

Chris M. Handley and Paul L. A. Popelier*

Manchester Interdisciplinary Biocentre (MIB), 131 Princess Street, Manchester M1 7DN, Great Britain and School of Chemistry, University of Manchester, Oxford Road, Manchester M13 9PL, Great Britain

Received November 3, 2008

Abstract: It is widely accepted that correctly accounting for polarization within simulations involving water is critical if the structural, dynamic, and thermodynamic properties of such systems are to be accurately reproduced. We propose a novel potential for the water dimer, trimer, tetramer, pentamer, and hexamer that includes polarization *explicitly,* for use in molecular dynamics simulations. Using thousands of dimer, trimer, tetramer, pentamer, and hexamer clusters sampled from a molecular dynamics simulation lacking polarization, we train (artificial) neural networks (NNs) to predict the atomic multipole moments of a central water molecule. The input of the neural nets consists *solely* of the coordinates of the water molecules surrounding the central water. The multipole moments are calculated by the atomic partitioning defined by quantum chemical topology (QCT). This method gives a dynamic multipolar representation of the water electron density without explicit polarizabilities. Instead, the required knowledge is stored in the neural net. Furthermore, there is no need to perform iterative calculations to self-consistency during the simulation nor is there a need include damping terms in order to avoid a polarization catastrophe.

1. Introduction

The existence of life on Earth, and perhaps in the rest of the cosmos, is reliant on the curious and unique properties of water. Liquid water has a number of anomalous properties compared to other similarly sized molecules and displays a great number of solid phases. Understanding water is critical to the understanding of environmental issues,[1–3] green chemistry,[4–7] and biological processes.[8–10] There are a number of excellent reviews on the broad subject of water and its unusual physicochemical features. Water is also an excellent solvent as well as being able to accommodate gases and guest molecules in the form of clathrate hydrates.[11–17]

Water has been the subject of numerous theoretical studies ever since the dawn of computational methods. The first water potential by Bernal and Fowler[18] began what would be a further 40 years of water simulations, with many different potentials designed along the way. However, over time, the focus has shifted from the analysis of small and relatively simple water clusters to larger systems where the number of water molecules is in the hundreds and interacting with other species. As the number and types of atoms involved in the simulation increase, ab initio calculations for these systems become computationally demanding, even exorbitant. Thanks to force fields and water potentials such simulations become tractable, albeit with a loss in accuracy. However, designing an accurate water potential is no easy feat[19] and has been the subject of research for almost 80 years. The design of a potential often becomes a parameter fitting problem aiming at the reproduction of a series of target properties, the number of which has grown as experimental techniques improved. This, along with particular shortcuts and simplifications in the design of a potential, reduces its transferability. Among accurate recent potentials there is the full-dimensional ab initio potential-energy surface for the water dimer of Bowman and co-workers[20] and the first-principles water potential of Bukowski et al.[21]

It is widely recognized that there is a need for polarizability to be included explicitly within water potentials.[19,22–24] It

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

10.1021/ct800468h CCC: $40.75
Published on Web 05/19/2009

has been further suggested that flexibility should not be introduced[25] until polarization is properly modeled. There is also a need for the electron density to be represented more accurately than can be achieved with point charges, better reflecting its anisotropic nature.[26–28] The most popular method is to represent the electron density by multipole moments defined from ab initio wave functions.[29–31]

Recently, artificial learning methods have appeared in potential design. Rather than trying to speculate which functions best model a potential, research groups introduced methods such as neural networks (NNs) to learn the potential from large amounts of ab initio data.[32–35] In this work we present a new method where NNs are trained to learn the relationship between a given water cluster configuration (input) and the multipole moments of an atom in the central water molecule within that configuration (output). After proper training, NNs can predict the multipole moments for an atom (within a water environment) in response to a given water cluster configuration. Hence, the NNs allow the multipole moments to dynamically respond to changes in the local cluster configuration. We developed the above method for dimer, trimer, tetramer, pentamer, and hexamer clusters. In these systems, moments are predicted for the central water molecule based upon the positions of the neighboring water molecules. These dynamic multipole moments combine polarization and charge-transfer effects in a single dynamic correction to the unpolarized Coulomb term. In addition, as we show below, we consider[36] water molecules appearing in water clusters as nonoverlapping. According to QCT (section 2.2), water molecules have finite boundaries and leave no spatial gaps between them. As a result, there is no need to correct for the so-called penetration effect,[37] typically by means of damping functions.

2. Background

2.1. Water Potentials. Water potentials fall into two categories: ab initio potentials and empirical potentials. Ab initio potentials are models where the force field parameters are set to reproduce the potential-energy surface as found by a sample of ab initio calculations.[38–40] Empirical potentials are parametrized to reproduce the bulk phase thermodynamic properties, a well-known example being TIP5P[25] and the more recent TIP4P/2005.[41] Both methods suffer from a lack of transferability. A potential fitted to reproduce the potential-energy surface of small clusters is not an ideal model for bulk conditions. Equally, a potential fitted to reproduce the bulk properties is not the ideal model for a molecule within small cavities and surfaces.

The simplification of using point charges on nuclear positions and at the 'lone pair' positions about the oxygen atom is popular. They are still used in the SPC series of models,[42–44] the TIPS series,[25,45,46] and the model of Nada and van der Eerden[47] who combined TIP4P and TIP5P to create a six-site potential, called NvdE.[48] It is the location and size of these point charges that may be modified to recover the targeted properties of water. However, there are many combinations of charges and water structures (bonds and angles) that will give the correct dipole and quadrupole moments, but this does not mean that any such model will correctly predict further properties.[44] The structure and charge distribution of a water molecule is finely balanced and has an influence on further properties, as expressed by Vega et al. when considering the relative stabilites of ices.[24] This is a view shared by Finney, who finds that classical methods of locating charge, i.e., at 'lone pairs', are not supported by quantum mechanics.[11,12]

It is known that the dipole of water increases from its gas-phase value of 1.85 D[49] to somewhere between 2.3 and 3.1 D[50–54] when moving from the gas phase to the bulk. In fact, water is a very polarizable molecule, able to respond to the electrostatic influence of ions and fields. Water realigns such that it opposes the field. The response to an external field is quickly transmitted through the hydrogen-bonding network. For this reason, liquid water is able to dissolve solids into the component ions. To account for this dipole enhancement, some models have had their charge distribution artificially changed so that the effect is included implicitly, such as in SPC/E[42] and TIP4P.[43] Such models are unreliable for simulations of water in the gas phase, in small cavities, surfaces, or very polar environments.[19,55] It is assumed that a more accurate water model that is transferable to many phases will need to account for polarization correctly.

2.2. Quantum Chemical Topology and Coulomb Interaction. Multipole moments are widely accepted to better represent the electron density of water (and other molecules). Studies by Gresh et al.,[26] Kaminsky and Jensen,[27] and Rasmussen et al.[28] have demonstrated that a multipolar representation of electron density is vital for modeling electrostatic interactions accurately. The multipole moments are coefficients of the series expansion that describes the electrostatic potential generated by an electron density. Multipole moments require more computational resources compared to point charges, even if they are expressed in terms of (irreducible) spherical harmonics as opposed to less compact Cartesian tensors. Multipole moments can be determined from ab initio wave functions by a number of methods, distributed multipole analysis (DMA)[56] being a well-documented and popular one. Multipole moments defined by this method have been successfully employed by Buckingham and Fowler.[57] These moments also turn up in the ASP potential,[29,58] the AMOEBA[31] water model, and the effective fragment potential (EFP) method.[59] Within the "sum of interactions between fragments ab initio" (SIBFA)[60,61] potential, the multipole moments are determined by the partitioning method of Vigné-Maeder and Claverie.[62] A further partitioning method that has grown from SIBFA is the Gaussian electrostatic model (GEM),[63,64] though it relies on density fitting rather than multipole moments.

Within this work, the partitioning of the electron density follows the method of the quantum theory of "atoms in molecules",[65–67] which is part of the quantum chemical topology (QCT) approach. A justification and rationale for the latter name can be found in ref 68. QCT defines topological atoms by the so-called gradient paths in the electron density. Gradient paths originating at infinity follow the direction of steepest ascent in every point of space. They typically (but not necessarily) terminate at nuclei. The three-

dimensional bundle of gradient paths that terminate at a given nucleus defines an atomic volume. A different (two-dimensional) bundle of gradient paths forms an interatomic surface that marks the boundary between two atoms. This bundle terminates at a so-called bond critical point, which lies in between two atoms that share a common boundary. *There are no gaps between topological atoms, and they collectively take up all space.* Atomic multipole moments are calculated by integrating the corresponding property density over the atomic volume. As an integrand of the volume integral, multiplication of the total electron density with regular spherical harmonics gives the required multipole moments.

The electrostatic interaction energy between two atoms is given by eq 1[69]

$$E^{\mathrm{AB}} = \sum_{l_\mathrm{A}=0}^{\infty}\sum_{l_\mathrm{B}=0}^{\infty}\sum_{m_\mathrm{A}=-l_\mathrm{A}}^{l_\mathrm{A}}\sum_{m_\mathrm{B}=-l_\mathrm{B}}^{l_\mathrm{B}} T_{l_\mathrm{A}m_\mathrm{A}l_\mathrm{B}m_\mathrm{B}}(\mathbf{R})Q_{l_\mathrm{A}m_\mathrm{A}}(\Omega_\mathrm{A})Q_{l_\mathrm{B}m_\mathrm{B}}(\Omega_\mathrm{B}) \quad (1)$$

The multipole moments of atom A, $Q_{l_\mathrm{A}m_\mathrm{A}}(\Omega_\mathrm{A})$, and atom B, $Q_{l_\mathrm{B}m_\mathrm{B}}(\Omega_\mathrm{B})$, interact through the tensor $T(\mathbf{R})$. $\mathbf{R}$ is the vector from nucleus A to nucleus B, the origins of the local frames for each atom. Collecting together the terms of eq 1 by their power of $R = \mathbf{R}$, we gather together terms of the same rank, L, defined as $l_\mathrm{A} + l_\mathrm{B} + 1$, where l is the rank of the multipole moment. For example, R^{-3} dependence consists of interactions between two dipole moments ($l_\mathrm{A} = l_\mathrm{B} = 1$) and between a monopole moment ($l = 0$) and a quadrupole moment ($l = 2$). By varying L, the convergence of the multipole expansion can be monitored. Hättig's recurrence formula[70] for the interaction tensor generates expansions up to arbitrarily high rank. The exact interaction energy can be obtained via a six-dimensional integration over the two participating atoms Ω_A and Ω_B

$$E^{\mathrm{AB}} = \int_{\Omega_\mathrm{A}} \mathrm{d}\mathbf{r}_\mathrm{A} \int_{\Omega_\mathrm{B}} \mathrm{d}\mathbf{r}_\mathrm{B} \frac{\rho_{\mathrm{tot}}(\mathbf{r}_\mathrm{A})\rho_{\mathrm{tot}}(\mathbf{r}_\mathrm{B})}{r_{\mathrm{AB}}} \quad (2)$$

where r_{AB} is the distance between two infinitesimally small charge elements and ρ_{tot} is the total charge density (which includes the nuclear charge). Before[71] we made a distinction between the terms "electrostatic" and "Coulomb". The former term is only well defined in the context of (long-range) intermolecular perturbation theory, while the latter applies to the interaction of any charge densities, whether in an intra- or intermolecular context. Since we will sample the electron density from atoms in supermolecules (i.e., water clusters) we are not working in a perturbation context, and hence, the term Coulomb is more appropriate. However, some texts use the two terms interchangeably.

QCT multipole moments are successful in MD simulations of liquid hydrogen fluoride and water[30,72,73] and aqueous solutions of imidazole as well as neat liquid imidazole.[74]

2.3. Polarization. Polarization causes up to a 70% increase in the dipole moment of water, and polarization is often quoted as accounting for ~15% of the total interaction energy[75,76] or as high as 50%.[77] The easiest way to account for polarization is implicitly, fitting the model parameters so that the experimental bulk phase properties are recovered. However, this does not allow for a dynamic anisotropic response of the electron density to an external field and changes in the local chemical environment.

The effect of an electric field upon a molecule can appear in three ways, as outlined by Yu and van Gunsteren.[77] A molecule can respond to an external field by a combination of reorientation, geometrical changes, and electronic redistribution. All models, whether or not they include a geometric or electronic response to polarization, will induce spatial reorientation of a molecule in response to an external field. However, this reorientation will of course be affected by any geometrical and electronic polarization responses that are accounted for by the model. Flexible models with static charge distributions allocated to atoms do account for polarization as a change in geometry in turn changes the molecular electron distribution. However, most water models assume a rigid geometry and concern themselves with the inclusion of electronic polarization.

Polarization can be accounted for explicitly in a number of ways. Three popular methods are (i) polarizable point dipoles,[78] (ii) fluctuating atomic charges,[79] and (iii) attaching a fictitious negative charge[80] to the molecule by a harmonic spring. The danger of the point dipole method is the "polarization catastrophe", where the dipoles respond in such a way that the interaction energy becomes infinite. In order to prevent a polarization catastrophe, where the dipole moments become infinite, a Thole damping function limits the response of the dipole moments.[81–84] The point dipole method appears in the AMOEBA water potential,[31] where polarizable point dipoles are located on atomic centers. Within the SIBFA model, polarizable point dipoles are also situated at off-nuclear positions.[60,85–87] This is analogous to the method in the EFP force fields.[88] The charge-on-spring method refined by MacKerell Jr. and Roux[89] is a simple concept adhered to in the past and more recently in the charge-on-spring class of water potentials by the van Gunsteren group.[90–93] Here polarization is introduced by a negative point charge tethered to the oxygen of the water molecule by a harmonic spring. Finally, the fluctuating charge method allows for the charge at atomic sites to change in response to the external field. This means charge can be redistributed about the molecule or transferred between two molecules. Hence the fluctuating charge method can model both polarization and charge transfer, where there is a partial transfer of charge between the donor and acceptor molecules that are interacting.[94,95] In this view charge transfer appears as a more extreme case of polarization.[96] This method is seen in combination with the TIP4P[79] model and the POL5 model, which is a modification of the TIP5P/ST2 model.[97] Unlike other polarization models, the fluctuating charge approach models polarization and charge transfer together, without additional terms to represent charge transfer.[59,60,85,86] Within the SIBFA model, charge transfer is explicitly represented by further terms.[98]

Recently, Houlding et al. proposed a novel method for incorporating polarization into a simulation of a hydrogen fluoride dimer[99] via dynamic QCT multipole moments. Drawing on large amounts of ab initio data, NNs were trained

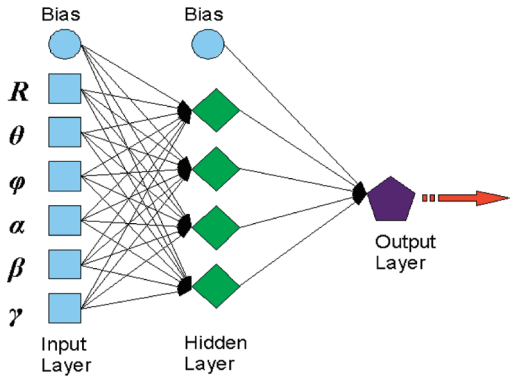

Figure 1. Diagram of a feedforward neural network with one hidden layer. The blue square is an input node, the blue circles are bias nodes, the green diamonds are hidden nodes, and the purple pentagon is an output node. In this work, the output is a multipole moment of a given atom and the inputs are the polar and Euler coordinates of the neighboring water molecules, shown here for the dimer cluster.

to predict the atomic multipole moments of hydrogen fluoride from thousands of configurations of hydrogen fluoride dimers. Following this work, Darley et al. presented a method[100] to tackle *intramolecular* polarization in a similar manner. In that work, polarization was a response to changes in the conformations of *N*-methylacetamide and glycine. In this work we follow the polarization method first introduced by Houlding et al. but expand the method to account for polarization caused by further nearest neighbors.

2.4. Neural Networks. For many years, NNs have been the subject of interest[101–103] as they are useful in discovering mappings between input and output data. In essence they accomplish this by classifying data (possibly large amounts) in arbitrarily high dimensional spaces. A NN is an array of connected nodes, processing units called threshold logic units (TLU), which pass information between themselves. A number of inputs are received by the node, which then sends an output. Each individual input to a node is multiplied by relevant weights, the products subsequently being summed. This sum is then passed through an activation function, whose function value is the final output. Alteration of these weights allows the NN to learn functions and relationships.

The main feature of a NN is the architecture, defined by the number of hidden layers and the number of nodes in the input, output, and hidden layers. Figure 1 shows a NN with inputs for the dimer, a single hidden layer, and an output. There are four hidden nodes and a bias node for both the input and hidden layer. Note that the bias nodes only send signals to the next layer; they do not receive any inputs. The hidden nodes are so named because the user does not have direct access to their outputs and because the hidden neurons must develop their own representation of the inputs.[102] The hidden layer allows the network to learn complex relationships by finding meaningful features from the inputs. The simplest NN is a feedforward network where there is only a single hidden layer of nodes. In a feedforward network, the nodes only pass information to the next layer and not back to a previous layer or to nodes in the same layer. NNs learn the mappings between inputs and outputs from a training set of examples. This *supervised learning* involves the reproduction of a *given* output from the associate input pattern, in order to alter the weights. This is the *backpropagation of errors* method. The process is repeated for every example in the training set before beginning again, with each full pass of the training set called an *epoch*.

Each neuron in layer k sums p inputs x_j from the previous layer j, which are each multiplied by their relevant weight w_{kj}, resulting in activation a_k, as shown in eq 3

$$a_k = \sum_{j=0}^{p} w_{kj} x_j \tag{3}$$

Each weight w_{kj} expresses the relationship between neuron k and neuron j. A weight can be positive or negative for a, respectively, excitatory or inhibitory connection. The output of a neuron must exceed a given threshold, θ, in order to be activating. Typically, the nonlinear sigmoid transfer function determines if a neuron's output is activating, given by eq 4 where ρ defines the shape of the sigmoid and y is the output.

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

In this work we are mapping our inputs, the internal coordinates that describe the water clusters, to our outputs, the multipole moments of the atoms in the central water molecule of the cluster.

In order to achieve the optimal NN for the prediction multipole moments the architecture of the NN is modified by varying the number of hidden nodes. NNs predictions can also be improved by altering two training parameters, the learning rate and the momentum.[102] Before training the input data must be standardized, i.e., transformed to dimensionless data that have a mean value of zero and a standard deviation of one. Subsequently, the data are transformed to lie in the interval [0,1] via eq 5

$$x_{i,n} = \frac{(x_i - x_{\min})}{(x_{\max} - x_{\min})} \tag{5}$$

where x_i are standardized input data, $x_{i,n}$ are normalized input data, and $x_{\min}$ and $x_{\max}$ are minimum and maximum values of the standardized data, respectively.

When training, the NN's performance is determined by 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.

A properly trained NN is one that is well generalized. This means that the NN is not overtrained nor overfitted. Overfitting means that the NN suffers from an overly flexible architecture. As a result, it inappropriately absorbs all the

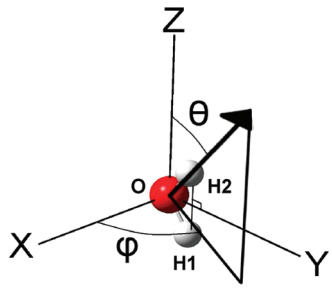

Figure 2. Molecular local frame (MLF) of a central water molecule. The *yz* plane is defined by the HOH plane.

noise in the training set and does not focus on the prime underlying trends in the data. Overtraining refers to when the NN has been trained for too many epochs and offers no predictive ability.

3. Computational Details

Multipole moments allow for an anisotropic description of the electron density of atoms. Multipole moments must be correctly aligned in space relative to each other because the Coulomb interaction between moments is orientation dependent. In our previous work[100] on glycine and *N*-methlyacetamide we introduced an atomic local frame (ALF). In this work we use a molecular local frame (MLF). This means that the atomic multipole moments for a particular water molecule all have the same local frame. This frame, shown in Figure 2, is simply defined by aligning the positive *y* axis along the HOH bisector, with both hydrogen atoms in the positive *y* direction. The *yz* plane is then defined by the HOH plane, with the first hydrogen atom lying at the negative side of the *z* axis.

This orientation convention determines *both* the generation of the training data for the NNs *and* the prediction of the moments by the NNs. The MLF defines the rotation of the moments from the MLF into the global frame. The training data, the multipole moments, are generated with respect to the MLF for every training example. Consequently, the NNs will predict moments with respect to the MLF. This means that the orientation of the MLF is embedded in the training data. The number of cluster configurations (or geometries) needed for the training is determined by the NN's architecture, more specifically, the number of weights. A rule of thumb is to have approximately 10 training examples (i.e., cluster configurations) for every weight being trained. Configurations are taken from molecular dynamics simulations[74] of pure liquid water at ambient conditions performed with nonpolarized (gas phase) QCT multipole moments. For each water molecule in the simulation we find the nearest neighbors that would form the dimer, trimer, tetramer, pentamer, or hexamer clusters. This signifies that we build our clusters in a hierarchical manner about a central water molecule. In other words, the dimer clusters lie within their trimer cluster, the trimer clusters lie within the tetramer clusters, and so forth. This is how approximately 5000 configurations of each cluster size are generated.

The program GAUSSIAN03[104] generated wave functions at the B3LYP/aug-cc-pVTZ level, for each cluster configuration, without geometry optimization. The internal geometry of each water molecule is fixed at the gas-phase-optimized values at the B3LYP/aug-cc-pVTZ level. Due to the modular nature of our method, the electron density may be obtained from wave functions generated at other levels of theory, possibly more advanced, future computing power allowing. For the central water molecule (which lies in the MLF as described above) the program MORPHY[105–107] generated the (atomic) QCT multipole moments for each atom.

The inputs for the NNs are generated from the Cartesian coordinates of the cluster. The Cartesian coordinates are transformed into a set of nonredundant coordinates following the method laid out by Stone.[108] For systems of rigid and nonlinear molecules we have $6(N-1)$ coordinates, where N is the number of molecules. For example, a system of 3 molecules has 12 coordinates, $6(3-1)=12$. With the central molecule at the origin of the MLF and aligned as described above, the position of each neighboring water molecule is described by three polar coordinates and three Euler angles. The polar coordinates are the distance R_{OO} between the central water oxygen atom and the oxygen atom of the neighboring water. The angle θ spans the vector $\boldsymbol{R}_{OO}$ and the z axis, and the angle φ spans the (positive) x axis and the projection of the vector $\boldsymbol{R}_{OO}$ on the xy plane. The three Euler angles, α, β, and γ are measured with respect to the reference water, where the HOH plane lies in the xz plane, with the hydrogen atoms in the negative z direction.

After defining the coordinates of each configuration (for a given cluster size), the data are standardized and normalized for subsequent exploitation by the NNs. For a given atom in the central water molecule, both the coordinates for a given configuration and the multipole moments for the atom in that configuration are transformed to lie between 0 and 1. Training starts operating on these transformed input and output data. Using different network architectures, momenta, and learning rates, NNs are trained for each multipole moment, up to and including the hexadecapole moment ($l=4$), of each atom in the water molecule. To start training, the input data set for a given moment is split into 10 unique validation sets. Each 10% of the data set serves as a validation set in turn. For each assigned validation data the reminder of the data (90%) set is divided up, such that two-thirds are used for training (60%) and the remaining third (30%) for testing in early stopping. In other words, for a given set of training parameters and architecture, we tested for early stopping and validated on 10 different sets.

Training is performed to maximize the training set correlation coefficient r^2. We aim to ensure that the NN is capable of making predictions for examples that were *not* seen in the training set. Hence, we require that the correlation coefficient for the validation set, v^2, is also maximized and close to 1. The statistic v^2 is calculated by the same formula as r^2 (eq 6) but by inserting data of the validation set only. For completeness we also determine the same correlation coefficient for the early stopping set, which we call q^2. The latter statistic is not to be confused with the q^2, the familiar cross-validation correlation coefficient (leave one/many out). However, it is possible that NN is overtrained and offers no predictive ability. To monitor proper NN generalization and

to select the best NN we demand that the ratio r^2/v^2 is close to 1. To certify that the training does not suffer from overtraining or overfitting we monitor the performance of the NN as it is trained. To do this we make use of the early stopping data set. Using the early stopping set we combine two methods for monitoring the performance, which have been described by Prechelt.[109] We test for the loss of generalization of the NN via the generalization loss function, defined in eq 7

$$GL_{(t)} = 100\left(\frac{E_{\mathrm{ea}(t)}}{E_{\mathrm{opt}(t)}} - 1\right) \tag{7}$$

where $E_{\mathrm{opt}(t)}$ is the lowest root-mean-square error (rmse), for the early stopping set, that has ever occurred by epoch t. This corresponds to the best performance ever seen. $E_{\mathrm{ea}(t)}$ is the performance found at a given epoch t. $GL_{(t)}$ is a measure of the loss of generalization at a time t. It is desirable that during the training increasingly better models are generated. If so, the error $E_{ea(t)}$ will always be less than $E_{opt(t)}$, in which case $GL_{(t)}$ is negative. However, if $GL_{(t)}$ is positive then training stops. This testing is initially performed for a given interval of t (here 25 epochs) rather than for every epoch. However, there may be a chance that generalization recovers and improves if training continues. For this reason we test the *progress*. This means that we monitor how many times $GL_{(t)}$ exceeds a specified threshold. If $GL_{(t)}$ exceeds the threshold 10 times then training is ended. When training is signaled to be stopped it returns to the best weights and location on the error surface achieved so far. Generalization loss is then tested, with a lower threshold, at each epoch before training is ended. The initial $GL_{(t)}$ is set to 0.01, which turned out to be adequate for our training purposes. This threshold allows the training to escape local minima in the fitness landscape. Lowering the threshold to 0.005 and progressing epoch by epoch (rather than every 25 epochs) is more appropriate to explore a local minimum. The latter threshold aids in finding the best possible solution without leaping away from it.

This combination of methods described by Prechelt[109] guarantees that training minimizes the errors while avoiding that it is trapped in local minima on the error surface. By training 10 times and testing on 10 different early stopping sets and validation sets we make sure that the training and testing data sets have not introduced a bias and that generalization is maintained without overfitting or overtraining.

The training of single-layer networks occurred with a hidden layer of 4−20 nodes, with momenta of 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9, and learning rates of 0.1, 0.25, 0.5, and 0.75.

For *external* testing of the NN's performance a further 1000 configurations were generated. For the dimer clusters, the moments for each of the molecules were generated. These are the *true* moments. Using these moments we can judge the ability of the NNs to predict moments for the water molecules for a given dimer configuration and also the Coloumb interaction energy.

For larger clusters we require a different procedure. For the trimer clusters and larger, 1000 clusters of 50 molecules were generated to act as the test configurations. The arbitrary number 50 were large enough to ensure that each water molecule that is a member of the central clusters we investigate (trimer, tetramer, etc.) can see its own first solvation shell. For the central water, the $n - 1$ nearest neighbors are identified ($n \geq 3$, where n is the total number of molecules in the cluster). From the coordinates of the central molecule and these neighbors the *true* moments are found for the central water molecule. The process is then repeated for each of the $n - 1$ neighbors of the central cluster, where for these neighbors we find their own nearest neighbors and predict moments for these clusters. This ensures that we have the true moments for each of the water molecules based upon their own $n - 1$ nearest neighbors as these are the configurations that are seen when the NNs predict multipole moments for each water molecule in the central cluster of n molecules. NNs are trained to predict moments for a molecule that lies at the center of a cluster of the molecule and its own nearest neighbors. For example, if we were to take the configuration shown in Figure 3a and generate the multipole moments from it, the only water molecule for which these true moments would match the predicted moments is for molecule 2. That is because the true moments and the predicted moments can only be compared if they are in the very same position, that is, where the *molecule is considered at the center of its own nearest neighbors.* It is false to predict moments for molecule 3 based upon the positions of 2 and 1 because molecule 1 is the nearest neighbor of 2 (in the MD simulation from which the cluster is sampled) but not of 3. Instead, the true moments for molecule 3 must be generated based upon *its own nearest neighbors*. Figure 3b shows the actual situation in a small region of the MD simulation from which we sample our test configurations. We predict moments for molecule 1 based upon its own nearest neighbors, 2 and 3. We must then take the true moments for molecule 1 from the wave function of the cluster of molecules 1, 2, and 3. For molecule 2 we predict the moments for this molecule based upon the positions of its own nearest neighbors, namely, 1 and 5. The true moments for molecule 2 are taken from the wave function of the cluster 2, 1, and 5. For molecule 3 its nearest neighbors are in fact 7 and 4. The moments for molecule 3 are predicted based upon the positions of molecules 7 and 4. The true moments are also generated from the wave function of the cluster 3, 4, and 7. Ultimately this procedure must be followed because the *NNs are trained on moments taken from molecules at the center of their own clusters.*

Figure 4 summarizes the sequence of processes of model building and validation. It includes the sampling of the training clusters, the generation of wave functions, the calculation of multipole moments, the training of NNs, the assessment of their prediction performance in terms of the moments themselves, and the Coulomb atom−atom interactions. The upper left corner of Figure 4 starts with the training and test configurations being sampled from the same MD simulation. The wave functions are then calculated for the training configurations ("Gaussian of central cluster") and for the test configurations ("Gaussian of central + neighbour clusters"). From the electron densities correspond-

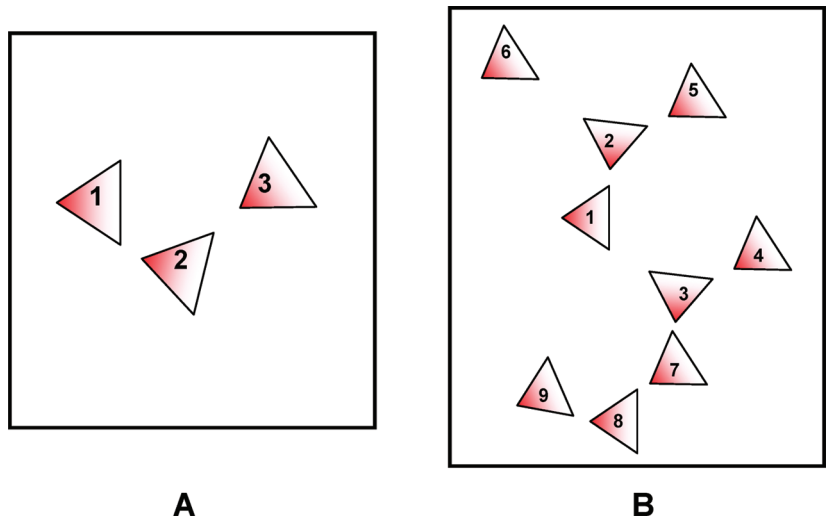

Figure 3. (a) Cartoon of the trimer cluster where molecule 2 is the central molecule of the cluster and molecules 1 and 3 are the nearest neighbors of molecule 2. The red shaded ends of the triangles represent the oxygen atom ends, while the white shaded corners of the triangles represent the hydrogen atom ends. (b) Cartoon of a cluster of 9 water molecules.

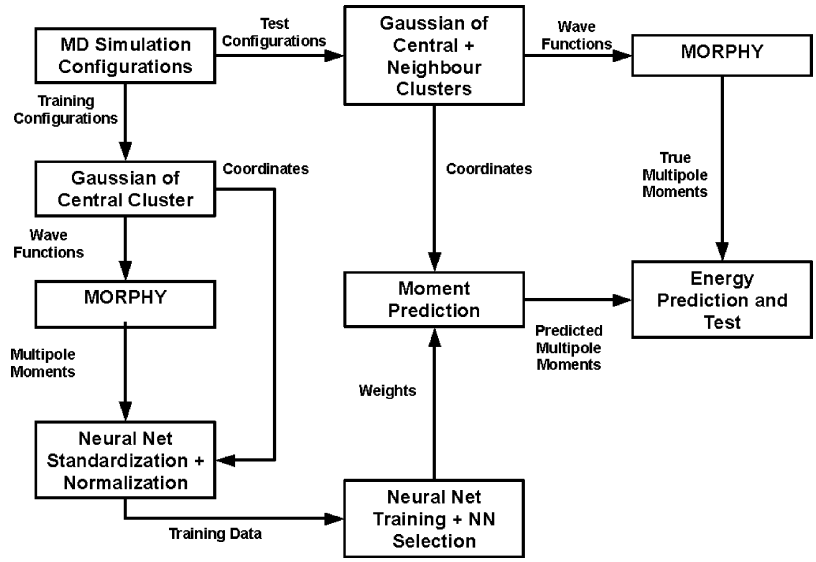

Figure 4. Schematic of the sequence of processes followed to generate and test the models.

ing to all wave functions the multipole moments are computed ("MORPHY"). The training data for the NNs (both input and output) are then prepared prior to actual training ("neural net standardization + normalization"). These training data are then used to train NNs for each moment of each atom. Note that the NNs predict the multipole moments solely based upon the coordinates of the nearest neighbors of a given water molecule. The trained knowledge of the NNs is stored as weights and network architecture. The predicted moments are then confronted with the true moments for a given test configuration (middle right of diagram in Figure 4). The quality of the training is then assessed by the correlation between the predicted and true moments and atom−atom Coulomb interaction energies.

We compared the CPU time required to evaluate all atom−atom interactions by our polarizable multipole model and by the TIP3P potential. On the basis of an average of 1000 water clusters, this overhead is about 61% for dimers and 59% for tetramers.

4. Results and Discussion

Table 1 shows the Pearson correlation coefficient values measured for the training (r^2), early stopping (q^2), and validation (v^2) data sets for the best performing NNs for each moment of the oxygen atom in water dimers. It is clear that generalization is maintained across all the data sets for each of the moments. For each moment the correlation coefficients obtained for the training, early stopping, and validation sets are very similar. Table S1 in the Supporting Information shows the equivalent data for one of the hydrogen atoms, since the data for the other hydrogen are very similar. The correlation coefficients are similar to those found for the oxygen atom. Overall, there is a trend for the performance

Table 1. Statistical Performance of the NN Training of Oxygen in Water Dimers

moment	no. nodes in hidden layer	r^2	q^2	v^2
Q00	10	0.973	0.969	0.973
Q10	9	0.818	0.819	0.816
Q11c	12	0.785	0.761	0.784
Q11s	9	0.916	0.896	0.916
Q20	9	0.992	0.990	0.993
Q21c	10	0.985	0.985	0.986
Q21s	10	0.977	0.976	0.979
Q22c	8	0.984	0.982	0.984
Q22s	11	0.993	0.992	0.992
Q30	11	0.977	0.972	0.977
Q31c	11	0.983	0.982	0.983
Q31s	7	0.949	0.941	0.953
Q32c	11	0.978	0.976	0.978
Q32s	13	0.974	0.971	0.974
Q33c	14	0.980	0.973	0.981
Q33s	11	0.976	0.972	0.978
Q40	12	0.961	0.957	0.961
Q41c	13	0.958	0.951	0.959
Q41s	10	0.990	0.988	0.990
Q42c	15	0.976	0.974	0.976
Q42s	12	0.934	0.924	0.936
Q43c	13	0.648	0.565	0.648
Q43s	14	0.977	0.971	0.977
Q44c	12	0.878	0.873	0.878
Q44s	10	0.862	0.853	0.863

of the NNs to decrease as the rank of the moment increases. However, there are further trends within the set of components of a given moment (i.e., for fixed l in Qlm, where component m varies). In Table 1 we see that for the oxygen atom the y component (Q11s) of the dipole moment is more easily predicted than the two other components, x and z. According to the MLF of water (Figure 2), the x (Q11c) and z (Q10) components of the dipole moment describe the out-of-plane and in-plane deflections of the dipole moment, respectively. These deflections are small. For the dimer cluster test configurations, the average absolute value of the x dipole component is 0.04 and 0.18 D for the z component. These are very small compared to the average value of y dipole component of 2.01 D and with respect to the magnitude of the dipole moment.[54,36] The dipole moment of water is most affected if a nearest neighbor is a hydrogen-bond donor or acceptor. This has a large effect on the magnitude of the dipole moment and thus the magnitude of the y component of the dipole moment. This means that the y component is more easily predicted than the other components because the location of the neighboring molecule has a large influence on it.

The trends seen in the dipole moments in the oxygen atom, Table 1, are not seen in the hydrogen atoms, Table S1 in the Supporting Information. Hydrogen displays almost equal correlation coefficients for all three components of the dipole moment, in contrast to oxygen. The hydrogen atom dipole moments are more dependent upon the precise location of the neighboring water molecule. Hence, the x and z components of a hydrogen dipole moment are easier to predict. For the oxygen atom dipole moment, the main factor determining the dipole moment is which end of the central molecule the nearest neighbor resides at. In other words, the nearest neighbor is either at the oxygen end of the water molecule (negative y semiaxis in Figure 2) or at the hydrogen atom end (positive y semiaxis).

Table 2 shows the decrease in the correlation coefficients for oxygen multipole moments as the cluster size increases from the dimer to the hexamer. As the correlation coefficients decrease, the root mean squared error (rmse) of moment prediction increases, with increasing cluster size. Table S2 of the Supporting Information shows a similar comparison for the hydrogen atoms. For both oxygen and hydrogen, the ability of the NNs to correctly predict the charge and dipole moments diminishes with increasing cluster size. A second feature shared by oxygen and hydrogen is the similarity in r^2 values for their monopole moments. However, the NNs' performance differs dramatically when predicting dipole moments of the hydrogen compared to the oxygen. For example, r^2 values larger than 0.8 still occur for hydrogen dipole components in the hexamer, while r^2 values for oxygen's components can be as low as 0.2. This suggests that the hydrogen atoms are more sensitive to the location of the neighboring molecules. The NNs can cope better with a strong local variation in the dipole moment due to varying positions of the neighboring waters. Conversely, the NNs would be challenged by the more diffuse variation in their dipole moments. In other words, the causal relationship between the position of the water neighbors (input) and oxygen dipole moments (output) is more intricate and buried in the data set. Finally, Table S3 of the Supporting Information shows the correlation coefficients for the NNs predicting the oxygen quadrupole moments with increasing cluster size. It is clear that the ability of the NNs to predict the quadrupole moments is not as adversely affected by the increasing cluster, unlike for the oxygen dipole moments. The majority of the quadrupole moments NNs still have correlation coefficients greater than 0.7 with a fair number even above 0.85. This suggests that the quadrupole moments are more sensitive to the local configuration of the cluster for larger, more homogeneous, bulk-like, clusters.

Table 3 shows the ability of the models to predict the total charge of the central water cluster (in each cluster size). Table 4 shows the ability to predict the (total) dipole moment of the central water (in each cluster). Using the NNs for each model, an in-house code predicted the multipole moments for water molecules and calculated the Coulomb energy from these moments. To calculate the Coulomb interaction between multipole moments of any rank we employ the recursive formula of Hättig.[70] The total dipole moment is described with respect to the MLF. As the cluster size increases, the rise in the average absolute errors for predicting the total charge of the whole cluster and total dipole moment of the central water molecule (within the cluster) is small. We are still able to predict both properties for the larger clusters (tetramer, pentamer, and hexamer) with some accuracy because once we reach these larger cluster sizes the variation of the total charge and total dipole moment is small and less dependent on the local arrangement of the neighboring molecules. This means that in heterogeneous environments, such as small clusters and interfaces, the local arrangement of the neighbors is a critical influence on the total charge and total dipole moment of a water molecule.

Table 2. Statistical Performance of the NN Training of Oxygen in Water with Increasing Cluster Size (rmse in au)

cluster	moment	hidden layer nodes	r^2	r^2 rmse	q^2	q^2 rmse	v^2	v^2 rmse
dimer	Q00	10	0.973	0.040	0.969	0.042	0.973	0.040
	Q10	9	0.818	0.067	0.819	0.068	0.816	0.064
	Q11c	12	0.785	0.070	0.761	0.075	0.784	0.068
	Q11s	9	0.916	0.049	0.896	0.053	0.916	0.048
trimer	Q00	10	0.937	0.041	0.921	0.045	0.937	0.042
	Q10	11	0.621	0.084	0.626	0.088	0.622	0.087
	Q11c	11	0.685	0.078	0.628	0.084	0.683	0.077
	Q11s	16	0.821	0.058	0.764	0.065	0.821	0.059
tetramer	Q00	11	0.885	0.060	0.848	0.068	0.885	0.062
	Q10	10	0.425	0.105	0.284	0.116	0.387	0.108
	Q11c	11	0.501	0.091	0.411	0.098	0.498	0.089
	Q11s	10	0.581	0.086	0.461	0.097	0.518	0.095
pentamer	Q00	22	0.531	0.084	0.433	0.097	0.520	0.089
	Q10	10	0.280	0.107	0.148	0.121	0.278	0.112
	Q11c	12	0.323	0.115	0.171	0.125	0.298	0.121
	Q11s	13	0.271	0.113	0.118	0.122	0.124	0.110
hexamer	Q00	18	0.680	0.080	0.567	0.096	0.666	0.087
	Q10	19	0.314	0.120	0.083	0.136	0.221	0.132
	Q11c	18	0.186	0.140	0.071	0.149	0.159	0.142
	Q11s	17	0.197	0.123	0.085	0.129	0.120	0.129

Table 3. Comparison of the Total Charge Errors (au) of the Central Water Cluster in Each Cluster Size[a]

cluster	average	min	max
dimer	0.0007	−0.0412	0.0464
trimer	–0.0003	–0.0442	0.0527
tetramer	–0.0024	–0.0895	0.0794
pentamer	0.0085	–0.0964	0.1069
hexamer	0.0091	–0.0853	0.1959

[a] The average is taken over all the test configurations for each cluster size.

Table 4. Comparison of Average Absolute Dipole Moment Errors (au) of the Central Water Molecule and the Maximum Absolute Dipole Moment Error for Each Cluster Size[a]

cluster	average abs	max abs
dimer	0.020	0.200
trimer	0.039	0.250
tetramer	0.061	0.312
pentamer	0.072	0.366
hexamer	0.077	0.362

[a] The average is taken over all test set configurations for each cluster size.

The models based upon larger clusters show that the charge and dipole moment of the central water molecule of the cluster is less dependent upon the local arrangement of the nearest neighbors. This is reassuring because the models developed for small clusters are suitable for modeling water in heterogeneous conditions, while our models based on larger clusters are more representative of the polarization response of a water molecule in the bulk phase. This conclusion is supported by the dipole moments predicted for the central water molecule as the cluster size increases.

Figure 5 shows that the models correctly recover the expected dipole moment enhancement of water as we move from the gas phase (1.85 D) to the bulk phase (3.07 D from the work of Bastista et al.[110] and 2.34 D from our previous study[54]). We can also see that the dipole moment of water spans a range of around 0.3 D and that the distributions for each cluster size show signature distribution profiles. The dipole moment for the dimer shows two peaks at 1.93 and 2.11 D, respectively. This is because the dipole moment of a water molecule depends upon whether the nearest neighbor is located at the hydrogen or the oxygen end of the water molecule. This observation highlights that the dipole moment of water is dependent on the hydrogen bonds that the central water molecule is involved in. There is also structure to the distribution for the trimer cluster. In the trimer set there is a main peak at 2.11 D and two lesser peaks at 1.97 and 2.17 D. This can be explained if we consider the arrangements possible of the two neighbors about the central water molecule in the trimer cluster. The two neighbors can both be located at the oxygen end of the water molecule or, alternatively, both at the hydrogen end. The third possibility is that one neighbor is at one end of the central water molecule while the other neighbor resides at the opposite end of this water molecule. As we increase the cluster size further, the distribution of the dipole moments adopts the shape of a bell curve. The peak is now shifted to a higher dipole moment: 2.14 D for the tetramer cluster and 2.27 D for the hexamer. Again, this evolution in the distributions and the shift of the peaks to higher dipole moments suggests that, in building hierarchical water models, we can create a

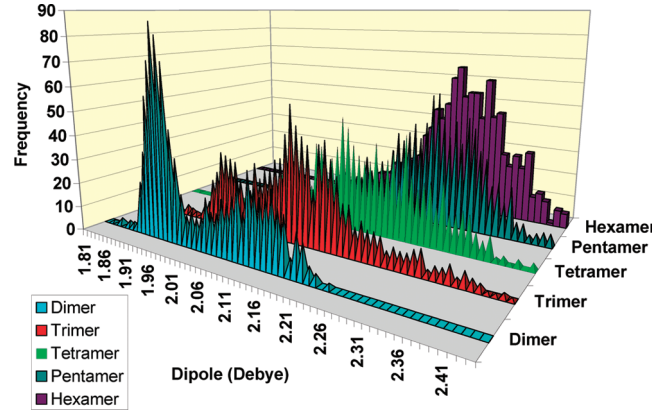

Figure 5. Dipole enhancement effect of water molecules for dimer, trimer, tetramer, pentamer, and hexamer clusters, as predicted by (single layer) NNs.

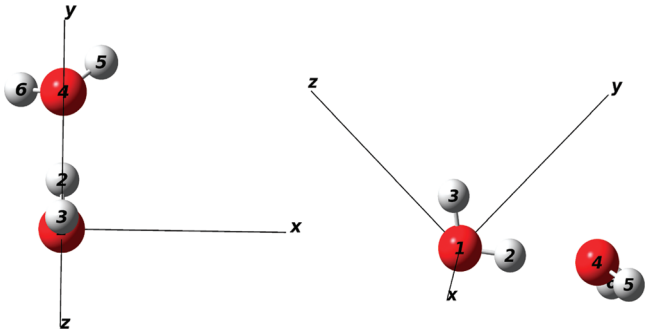

Figure 6. Two perspectives of the dimer configuration selected to test the convergence of intermolecular interactions. This configuration was chosen in view of its similarity to the global minimum in the gas phase. The water molecule H3−O1−H2 is situated at the origin. The global frame (labeled *x*, *y*, and *z*) of the dimer coincides with the MLF of H3−O1−H2, as shown in Figure 2.

range of water models that together are able to describe water moving from the gas phase to the interface and into the bulk phase.

In order to test the validity of the predicted multipole moments the convergence of the multipolar interactions is monitored. A single test configuration sampled from the same MD simulation that is the source of the training data is adequate provided it is common and possessing short-range interaction (i.e., a strong hydrogen bond). This configuration has not been seen by the NNs during training. The configuration, shown in Figure 6, has been selected to resemble the global minimum of the (gas phase) water dimer, differing from it mainly by an asymmetry-inducing tilt of the water molecule H6−O4−H5.

Convergence of multipole moment interaction energies is important if we wish to accurately calculate Coulomb interactions. We dedicated much attention to convergence issues.[69,71,111–116] We can assess the convergence of an interaction by comparing the Coulomb interaction between two atoms for a given rank L to the "exact" Coulomb interaction. The latter is found by a six-dimensional (6D) integration over the volumes of two interacting topological atoms (see eq 2). Since this calculation does not invoke a multipole expansion it can never suffer from convergence problems. We have explicitly shown[100,112] that a given set of multipole moments for two interacting atoms has a particular convergence profile, that is, a plot of the Coulomb interaction energy with respect to rank L. We can use the convergence profile as a way to assess the prediction of the multipole moments. If the NNs were perfect then the predicted multipole moments would be exactly the same as the true moments for the test configuration. In that case the convergence profiles for both sets of multipole moments would be exactly the same. Upon the basis of this principle we can use the difference in convergence behavior as a measure of how well the predicted moments match the true moments. Figure 7 shows the convergence profiles between H2, on one hand, and O4, H5, and H6, on the other. The profiles are calculated by subtracting the exact 6D atom−atom interaction energy from the interaction energy obtained from multipole moments (true or NN predicted). Further convergence profiles for the remaining interactions are shown in the Supporting Information (Figures S1 and S2).

By comparing the convergence profiles from the predicted and true moments for a particular atom−atom interaction it is clear that the main difference in interaction energies is due to the error in the prediction of the atomic monopoles. This can be explained against the apparently contradictory background of their very high value correlation coefficients of prediction. A small relative error for a moment, such as the monopole, translates into a large energy error, compared to the energy errors due to the performance of the other moments. This indicates that correlation coefficients are not the sole judge of the prediction quality. The difference between the convergence profiles for a particular interaction remains almost constant. This means that if we were to further improve the prediction of the interaction energies we must focus on improving the prediction of the monopole moments first. This makes sense since the lower order moments are involved in more moment−moment interactions for a given rank L.

The worst convergence is seen in Figure 7 for the interaction between H2 and O4. This is not surprising considering the short range of this interaction (1.66 Å), while the typical length of a hydrogen bond is about 1.97 Å in liquid water. Compared to the actual magnitude of the atom−atom interactions the errors due to differences in convergence are small (∼0.5−5%) depending upon the type of interaction being computed. Looking at Figures 7 and S1 and S2 in the Supporting Information it is clear that the difference between true and predicted energies is dominated by the charge−charge term. This is because the corresponding energy curves (true and predicted) are roughly parallel with increasing L.

Table S4 of the Supporting Information exhaustively shows how the average $L = 5$ atom−atom interaction energy changes with increasing cluster size up to the pentamer. It lists percentages, which are calculated as 100 E_{true}(A,B) − $E_{\text{predicted}}$(A,B)/E_{true}(A,B), where A and B represent any possible atom pair and "predicted" refers to the NN. The percentages are averaged over all test configurations for a given cluster size. The percentages change very little with increasing cluster size and never rise above 2.5%. This is not unexpected as the ability of the NNs to accurately predict the monopole moments decreases with increasing cluster. The average atom−atom energy only changes in magnitude by up to ∼5% between the dimer and the pentamer (or hexamer).

If we consider the average absolute errors, we see that the errors are smaller for interactions between water molecules that on average become more distantly separated. The average absolute interaction energy errors (for each A,B pair) is on the order of a few kJ/mol, while the average interaction energies are on the order of hundreds of kJ/mol. We do not repeat this table for the nonpolarizable models because the percentages would be embarrassingly high, including for TIP3P. Indeed, in terms of a table such as Table S4 of the Supporting Information, the TIP3P potential performs poorly. We already know that multipolar Coulomb interactions give

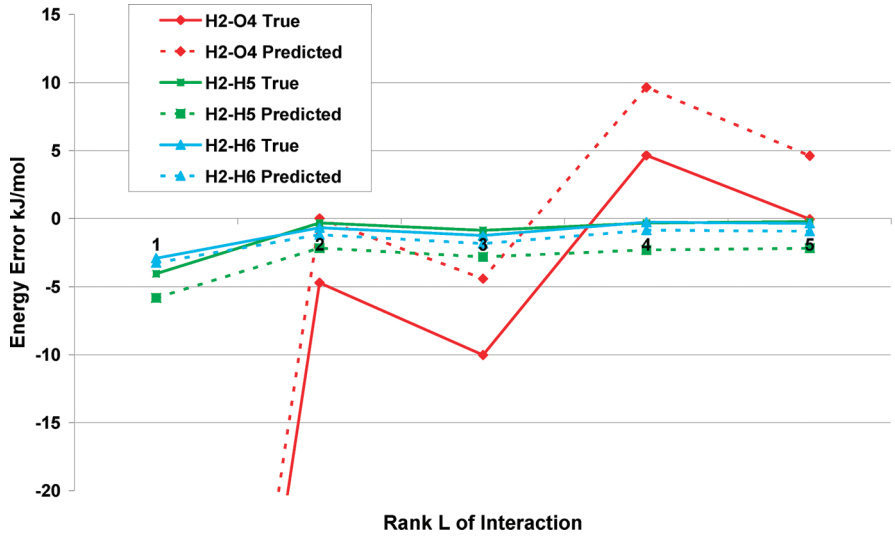

Figure 7. Plot of the convergence of the intermolecular interactions between H2 and O4, H5, and H6, in the water dimer, as the interaction rank *L* increases.

energies close to the energies obtained by 6D integration. Taking the interaction energies calculated using the true multipole moments as *exact energies* we find that TIP3P has average percentage errors of at least 50%, in all interactions, for any cluster size. The same percentage errors occur when unpolarized moments are taken from the monomer wave function of a single water molecule in the gas phase.

Next we compare the total Coulomb interactions, for each cluster size, obtained using the polarized multipole moments, the unpolarized gas phase monomer moments, and the TIP3P point charge model. Figure 8 shows the cumulative plots of the absolute total Coulomb interaction energy errors for the dimer and pentamer clusters, respectively. Further plots for the trimer, tetramer, and hexamer clusters are provided in Figure S3 of the Supporting Information. The "QCT curve" is the cumulative energy error curve found for the static water gas-phase monomer multipole moments. These plots investigate if polarization improves the energy predictions with respect to interaction energies found for the test cluster using the true multipole moments for these clusters. Curiously, "QCT" and TIP3P perform well for the pentamer test configurations. However, inspection of individual atom−atom interaction energies proves that these two models are very wrong and that the accuracy of the total Coulomb interaction is actually a fortuitous cancellation of large errors. The NN-based model is in fact more accurate because the individual atom−atom interactions are also accurate. These curves demonstrate that models cannot just be judged by a single number.

The polarizable model performs better than all other methods except in the case of the pentamer and hexamer clusters. However, from our previous analysis of the atom−atom interaction energies we conclude that for TIP3P and the unpolarized QCT models the interaction energy errors are fortuitous (in the case of the QCT gas-phase monomer moments) or a result of how the model has been fitted. In the case of TIP3P the point charges have been fitted to reproduce the thermodynamic properties of the bulk phase. We know from our previous work[54] that as water cluster size is increased and the number of nearest neighbors about a central water molecule increases, the dipole moment of the water molecule at the center of the cluster increases asymptotically to the bulk phase value. The pentamer cluster size is important for clusters in our hierarchical construction (section 3) as it is at this size that the central water molecule has four nearest neighbors around it. We could consider the pentamer cluster to be the cluster where the central water molecule has obtained its first solvation shell of neighbors. It is also at this cluster size that a large proportion of the dipole enhancement has occurred. It is then possible that the pentamer configurations provide a local environment similar to that of the bulk phase. Since TIP3P is a water model aiming at modeling bulk water it is expected to predict better Coulomb interaction energies than the polarizable model. This in spite of TIP3P's very large errors for the individual atom−atom Coulomb interaction energies. The TIP3P point charges have been parametrized for this type of situation and not for heterogeneous configurations such as the smaller clusters. We are already aware that the point charges can be fitted in a number of ways to reproduce the same properties. However, the differences are clear when we consider the atom−atom interactions. Though TIP3P is a better model for the pentamer clusters, it is right for the wrong reasons. Table 5 summarizes the performance of the models for each cluster size.

Table 5 compares the total Coulomb interaction energies for each cluster size by means of the average absolute errors and the 50th, 90th, and 99th percentile absolute energy errors. This information can be read from Figure 8 and Figure S3 of the Supporting Information. As the cluster size increases, the average absolute total Coulomb interaction error also increases. This is not surprising since the number of atoms for which moments are being predicted is also increasing with cluster size. The latter, in turn, increases the number of prediction errors. However, as the cluster size increases the average total Coulomb interaction energy increases from ∼25 kJ/mol for the dimer to ∼100 kJ/mol for the pentamer clusters. Although the predicted errors are increasing, the

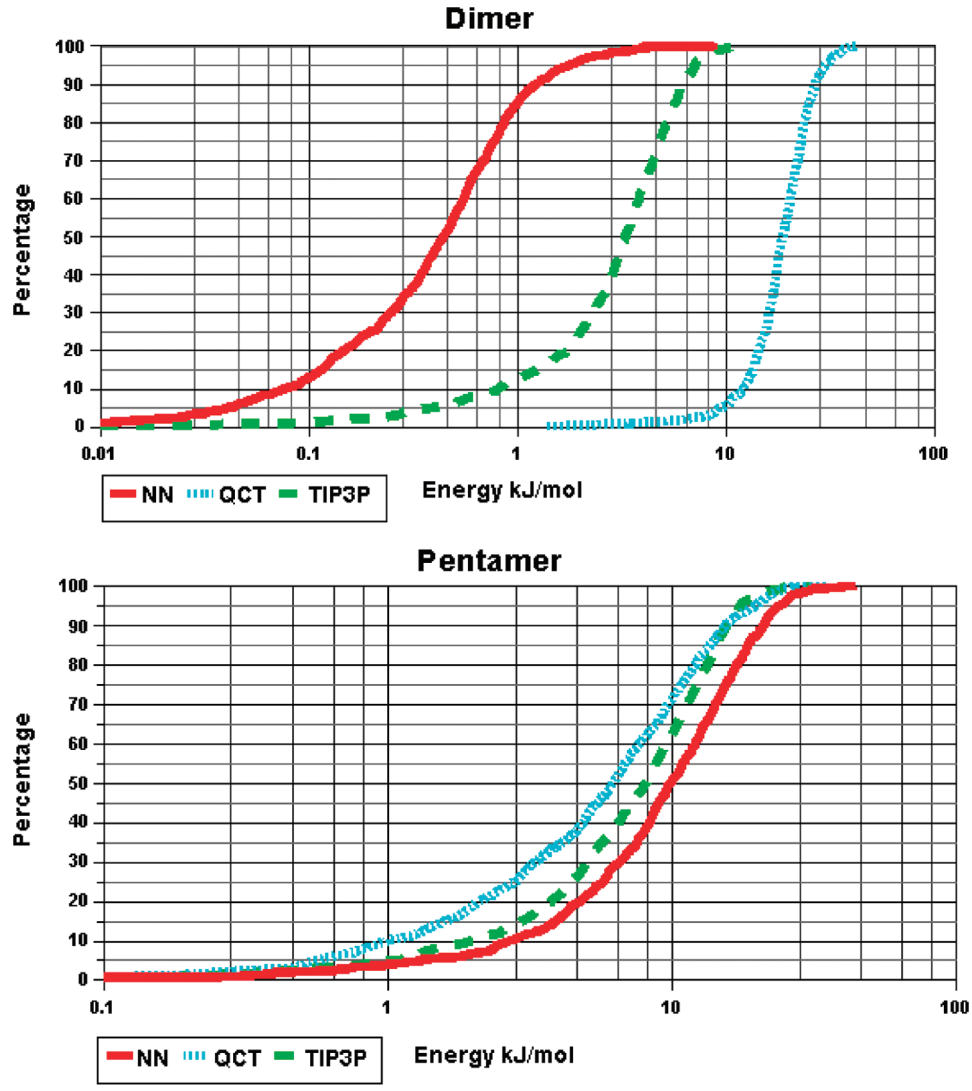

Figure 8. Semilogarithmic cumulative plot of the absolute total Coulomb interaction energy errors for the dimer and pentamer water clusters.

Table 5. Comparison of the Average Absolute Total Coulomb Interaction Energy Errors (in kJ/mol): The 50th, 90th, and 99th Percentile Absolute Total Coulomb Interaction Energy Error

cluster size	model	average	50th percentile	90th percentile	99th percentile
2	NN	0.62	0.44	1.23	3.66
	QCT unpolarized	19.20	18.67	26.86	36.45
	TIP3P	3.52	3.29	6.19	9.61
3	NN	2.59	2.01	5.27	11.44
	QCT unpolarized	10.06	9.62	18.34	24.81
	TIP3P	6.11	4.46	14.07	22.41
4	NN	5.40	4.36	11.37	21.29
	QCT unpolarized	8.86	8.04	16.96	24.55
	TIP3P	6.50	5.11	13.63	25.25
5	NN	11.14	9.89	21.05	30.60
	QCT unpolarized	7.46	6.24	15.43	23.95
	TIP3P	8.61	8.14	15.60	23.12
6	NN	16.37	15.01	27.72	41.89
	QCT unpolarized	8.22	6.41	16.56	29.60
	TIP3P	11.54	10.96	19.27	31.04

percentage errors (calculated as $100\sum_{A,B}^{\text{all pairs}}|E_{\text{true}}(A,B) - E_{\text{predicted}}(A,B)|/E_{\text{true}}(A,B)$) remain fairly constant, at about 10% at worst. More importantly, on average, the NN model for any cluster size is never in error by more than the sum energy of all the possible hydrogen bonds the central water molecule can make (~5 kJ/mol for each hydrogen bond).

At the end of this section we briefly explain how the proposed method avoids the "polarization catastrophe". One

should keep in mind that QCT atoms do not overlap. Instead, the atoms exist as malleable and finite regions in space. Their multipole moments are always well defined, even if the molecules in the clusters come very close to each other. Second, one should remember that the current method does not invoke polarizability tensors. The polarization is implicitly embedded in the trained neural networks. The nets can only produce *finite* multipole moments, and the *effect* of polarization is already fully taken into account when the net generates its multipolar output. On the basis of these important features a polarization catastrophe can never occur.

The proposed polarization approach can be incorporated in molecular dynamics simulations. Inside the MD program DLMULTI (which contains multipolar Ewald summation[117]), we already implemented a subroutine that stores the weights of a trained neural net. The net then predicts, on the fly, the multipole moments of all water molecules in the simulation box, when given their respective environments. In principle, the weights of neural nets obtained by training in one laboratory can be passed on to another lab that performs the simulation. Working within a rigid body framework, forces and torques can be obtained by differentiating eq 1. Note that the differentiation product rule then yields three terms, but only the term with the differentiation of the interaction tensor T is currently included. The nonlinear behavior and anisotropy of polarization is in principle captured by the net and made available to the simulation engine.

5. Conclusions

We proposed a novel, polarizable multipolar water potential that uses neural networks (NNs) to predict atomic multipole moments for any given water cluster configuration, in principle. This method eliminates the need to perform iterations during an MD simulation, unlike for many polarizable models. Second, this method allows for the Coulomb interaction to include polarization and charge transfer, treated on a par, as part of a dynamic Coulomb interaction term. The new water models have been critically analyzed at all stages during the prediction of cluster Coulomb interaction energies in order to assess the ability of the NNs to predict multipole moments on the atoms of a water molecule at the center of a particular cluster. The performance of NNs diminishes with increasing cluster size, which is related to two problems. The first is that the space in which the NNs are trying to fit functions is increasing in dimensionality. The second is that, as the cluster size increases, certain multipole moments become less sensitive to the location of the nearest neighbor water molecules about the central water molecule.

In summary, we critically assessed the NNs' ability to predict the moments of the atom. Additionally, we can further test the NNs by using the predicted multipole moments to compute the charge of the atoms, the net charge of the central water molecule, the dipole moment of the central water molecule, and the atom−atom and total cluster Coulomb interaction energies. We conclude that the NN-based model allows for a more complete description of the Coulomb interactions between water molecules, compared to the nonpolarizable QCT model and the TIP3P water potential. Using the gas-phase monomer moments as a reference, we find that, on average, the polarization of the multipole moments accounts for 50% of the atom−atom Coulomb interaction energies. We showed that, due to the fitting method behind TIP3P, it is able to predict total Coulomb interaction energies that are more accurate than the polarizable model. However, this accuracy is due to a fortuitous cancelation of very inaccurate atom−atom interaction energies. This suggests that other empirical point charge models, fitted in a fashion similar to TIP3P, may again yield accurate total interaction energies for large clusters of water molecules but not for the correct reason.

Fixed charge density models do not accurately predict Coulomb interaction energies for clusters that are representative of the type of local configurations that a water molecule would experience in heterogeneous conditions, such at the water−air interface. Our pentamer model is probably appropriate for the simulation of water in the bulk phase, while the models developed based on smaller water clusters are suited to predicting the Coulomb interaction energies between water molecules in the gas phase and at an interface. Using this hierarchy of models, it would be possible to perform simulations of the ice surface melting where the Coulomb interactions are represented accurately using dynamic polarizable multipole moments.

We are already exploring more elaborate NNs to predict the multipole moments. Two hidden layers in a NN introduce more flexibility to fit a function to the training data. This added flexibility, to be reported on in a future article, has already improved the accuracy of the prediction of the monopole moments of oxygen and improved the Coulomb interaction energies for the dimer, trimer, and tetramer clusters. We are also exploring the use of other statistical learning machines called radial basis function networks, and Kriging models, which already show a better ability to predict accurate polarizable multipole moments compared to the NNs.

Acknowledgment. We thank the EPSRC for financial support through grant EP/C015231 and Dr. G. I. Hawe for valuable comments.

Supporting Information Available: Convergence of the intermolecular interactions between O1 and O4, H5, and H6, in the water dimer, as the interaction rank L increases; convergence of the intermolecular interactions between H3 and O4, H5, and H6, in the water dimer, as the interaction rank L increases; cumulative plot of the absolute total electrostatic interaction energy for the trimer, tetramer, and hexamer systems; statistical performance of the NN training of a representative hydrogen atom in water dimers; statistical performance of the NN training of hydrogen in water with increasing cluster size (rmse in au); statistical performance of the NN training of oxygen in water with increasing cluster size (rmse in au); comparison of atom−atom total interaction energy errors (percentage) with increasing cluster size. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Dash, J. G.; Fu, H.; Wettlaufer, J. S. *Rep. Prog. Phys.* **1995**, *58*, 115.

(2) Hervig, M.; Thompson, R. E.; McHugh, M.; Gordley, L. L.; Russell, J. M., III.; Summers, M. E. *Geophys. Res. Lett.* **2001**, *28*, 971.

(3) Prenni, A. J.; DeMott, P. J.; Kreidenweis, S. M. *Atmos. Environ.* **2003**, *37*, 4243.

(4) Savage, P. E. *Chem. Rev.* **1999**, *99*, 603.

(5) Dabiri, M.; Baghbanzadeh, M.; Nikcheh, M. S.; Arzroomchilar, E. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 436.

(6) Gao, R.; Dai, W.-L.; Le, Y.; Yang, X.; Cao, Y.; Li, H.; Fan, K. *Green Chem.* **2007**, *9*, 878.

(7) Chen, L.; Li, C.-J. *Adv. Synth. Catal.* **2006**, *348*, 1459.

(8) 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.* **1994**, *91*, 1089.

(9) Okada, T.; Fujiyoshi, Y.; Silow, M.; Navarro, J.; Landau, E. M.; Shichida, Y. *Proc. Natl. Acad. Sci.* **2002**, *99*, 5982.

(10) Killian, J. A.; von Heijne, G. *Trends Biochem. Sci.* **2000**, *25*, 429.

(11) Finney, J. L. *J. Mol. Liq.* **2001**, *90*, 303.

(12) Finney, J. L. *Philos. Trans. R. Soc. London, Ser. B: Biol. Sci.* **2004**, *359*, 1145.

(13) Ludwig, R. *Angew. Chem., Int. Ed* **2001**, *40*, 1808.

(14) Robinson, G. W.; Zhu, S.-B.; Singh, S.; Evans, M. W. *Water in Biology, Chemistry and Physics*; World Scientific Publishing: Singapore, 1996.

(15) Franks, F. *Water: a comprehensive treatise*; Plenum Press: New York, 1972.

(16) Franks, F. *Water: a matrix of life*; Royal Society of Chemistry: Cambridge, 2000.

(17) Stillinger, F. H. *Science* **1980**, *209*, 451.

(18) Bernal, J. D.; Fowler, R. H. *J. Chem. Phys.* **1933**, *1*, 515.

(19) Guillot, B. *J. Mol. Liquids* **2002**, *101*, 219.

(20) Huang, X.; Braams, B. J.; Bowman, J. M. *J. Phys. Chem. A* **2006**, *110*, 445.

(21) Bukowski, R.; Szalewicz, K.; Groenenboom, G. C.; van der Avoird, A. *Science* **2007**, *315*, 1249.

(22) Paricaud, P.; Predota, M.; Chialvo, A. A.; Cummings, P. T. *J. Chem. Phys.* **2005**, *122*, 244511.

(23) Chen, B.; Xing, J.; Siepmann, J. I. *J. Phys. Chem. B* **2000**, *104*, 2391.

(24) Vega, C.; Sanz, E.; Abascal, J. L. F. *J. Chem. Phys.* **2005**, *122*, 114507.

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

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

(27) Kaminsky, J.; Jensen, F. *J. Chem. Theor. Comput.* **2007**, *3*, 1774.

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

(29) Millot, C.; Stone, A. J. *Mol. Phys.* **1992**, *77*, 439.

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

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

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

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

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

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

(36) Devereux, M.; Popelier, P. L. A. *J. Phys. Chem. A* **2007**, *111*, 1536.

(37) Freitag, M. A.; Gordon, M. S.; Jensen, J. H.; Stevens, W. J. *J. Chem. Phys.* **2000**, *112*, 7300.

(38) Matsuoka, O.; Clementi, E.; Yoshimine, M. *J. Chem. Phys.* **1975**, *64*, 1351.

(39) Niesar, U.; Corongiu, G.; Clementi, E.; Kneller, G. R.; Bhattacharya, D. K. *J. Phys. Chem.* **1990**, *94*, 7949.

(40) Lie, G. C.; Clementi, E. *Phys. Rev. A* **1985**, *33*, 2679.

(41) Vega, C.; Abascal, J. L. F.; Conde, M. M.; Aragones, J. L. *Faraday Discuss.* **2008**, *141*, 1.

(42) Berendsen, H.; Postma, J.; Van Gunsteren, W.; Hermans, J. *Interaction models for water in relation to protein hydration*; Reidel: The Netherlands, 1981.

(43) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269.

(44) Glättli, A.; van Gunsteren, W. F.; Daura, X. *J. Chem. Phys.* **2002**, *116*, 9811.

(45) Jorgensen, W. L. *J. Am. Chem. Soc.* **1981**, *103*, 335.

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

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

(48) Bishop, C. L.; Pan, D.; Liu, L. M.; Tribello, G. A.; Michaelides, A.; Wang, E. G.; Slater, B. *Faraday Discuss.* **2008**, *141*, 1.

(49) Clough, S. A.; Beers, Y.; Klein, G. P.; Rothman, L. S. *J. Chem. Phys.* **1973**, *59*, 2254.

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

(51) Silvestrelli, P. L.; Parrinello, M. *Phys. Rev. Lett.* **1999**, *82*, 3308.

(52) Gregory, J. K.; Clary, D. C.; Liu, K.; Brown, M. G.; Saykally, R. J. *Science* **1997**, *275*, 814.

(53) Gubskaya, A. V.; Kusalik, P. G. *J. Chem. Phys.* **2002**, *117*, 5290.

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

(55) Kollman, P. A. *Acc. Chem. Res.* **1996**, *29*, 461.

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

(57) Buckingham, A. D.; Fowler, P. W. *Can. J. Chem.* **1985**, *63*, 2018.

(58) Millot, C.; Soetens, J.-C.; Martins-Costa, M. T. C.; Hodges, M. P.; Stone, A. J. *J. Phys. Chem.* **1998**, *102*, 754.

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

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

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

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

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

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

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

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

(67) Matta, C. F.; Boyd, R. J. *The Quantum Theory of Atoms in Molecules*; Wiley-VCH: Weinheim, Germany, 2007.

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

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

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

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

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

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

(74) Shaik, M. S. PhD thesis, Design of a Multipolar Potential leading to a Description of the Self-Assembly of Imidazole in Aqueous Solution, The University of Manchester, 2008.

(75) Friesner, R. A. Modeling Polarization in Proteins and Protein−Ligand Complexes: Methods and Preliminary Results. In *Advances in Protein Chemistry*; Baldwin, R.; Baker, D., Eds.; Academic Press: New York, 2005; Vol. 72, p 79.

(76) Hodges, M.; Stone, A. J.; Cabaleiro Lago, E. *J. Phys. Chem.* **1998**, *102*, 2455.

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

(78) Stern, H. A.; Rittner, F.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **2001**, *115*, 2237.

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

(80) Sprik, M.; Klein, M. L. *J. Chem. Phys.* **1988**, *89*, 7556.

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

(82) 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.

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

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

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

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

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

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

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

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

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

(92) Yang, M.; Senet, P.; Alsenoy, C. V. *Int. J. Quantum Chem.* **2005**, *101*, 535.

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

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

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

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

(97) Stern, H. A.; Rittner, F.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **2001**, *115*, 5.

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

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

(100) Darley, M. G.; Handley, C. M.; Popelier, P. L. A. *J. Chem. Theory Comput.* **2008**, *4*, 1435.

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

(102) Gurney, K. *An Introduction to Neural Networks*; Taylor and Francis: London, 1997.

(103) Haykin, S. *Neural Networks: A Comprehensive Foundation*, 2nd ed.; Prentice-Hall, New Jersey, 1999.

(104) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J., 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.

(105) Popelier, P. L. A.; Bone, R. G. A. *MORPHY98*, UMIST: Manchester, 1998.

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

(107) Popelier, P. L. A. *Chem.Phys.Lett.* **1994**, *228*, 160.

(108) Stone, A. J. *The Theory of Intermolecular Forces*, 1st ed; Clarendon Press:, 1996; Vol. 32.

(109) Prechelt, L. *Neural Networks* **1998**, *11*, 761.

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

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

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

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

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

(115) Kosov, D. S.; Popelier, P. L. A. *J. Chem. Phys.* **2000**, *113*, 3969.

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

(117) Leslie, M. *Mol. Phys.* **2008**, *106*, 1567.

CT800468H