Multipolar electrostatics based on the Kriging machine learning method: An application to serine

ARTICLE *in* JOURNAL OF MOLECULAR MODELING · APRIL 2014

Impact Factor: 1.74 · DOI: 10.1007/s00894-014-2172-1 · Source: PubMed

3 AUTHORS, INCLUDING:

Matthew J L Mills
Joint BioEnergy Institute
13 PUBLICATIONS **107** CITATIONS
SEE PROFILE

Paul L A Popelier
The University of Manchester
190 PUBLICATIONS **7,094** CITATIONS
SEE PROFILE

PCCP

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2014, **16**, 4122

The hydration of serine: multipole moments *versus* point charges

Steven Y. Liem[ab] and Paul L. A. Popelier*[ab]

Received 7th November 2013,
Accepted 13th January 2014

DOI: 10.1039/c3cp54723j

www.rsc.org/pccp

Next-generation force fields must incorporate improved electrostatic potentials in order to increase the reliability of their predictions. A crucial decision toward this goal is to abandon point charges in favour of multipole moments centered on nuclear sites. Here we compare the geometries generated by quantum topological multipole moments with those generated by four popular point charge models (TAFF, OPLS-AA, MMFF94x and PFROSST) for a hydrated serine. A main feature of this study is the dual comparison made, both at static level (geometry optimisation *via* energy minimisation) and at dynamic level (*via* molecular dynamics and radial/spatial distribution function analysis). At static level, multipolar electrostatics best reproduces the *ab initio* reference geometry. At dynamic level, multipolar electrostatics produces more structure than point charge electrostatics does, over the whole range. From our previous work on liquid water [*Int. J. Quantum. Chem.*, 2004, **99**, 685], where agreement with experiment only occurs when using multipole moments, we deduce that our predictions for hydrated serine will also be closer to experiment when using multipolar electrostatics. The spatial distribution function shows that only multipolar electrostatics shows pronounced structure at long range. Even at short range there are many regions where waters appear in the system governed by multipolar electrostatics but not in that governed by point charges.

1 Introduction

In spite of their inherent lack of accuracy in representing the electrostatic interaction, atomic partial charges (or simply point charges) continue to dominate molecular simulations. The reason for this status quo is not entirely clear but the evidence in favour of the alternative of multipolar electrostatics is clear as proven by a recent review dedicated to this matter.[1] The majority of systematic work comparing point charge potentials with multipolar potentials focuses on van der Waals complexes in the gas phase. Exploring their potential energy surfaces reveals clear differences in the geometrical predictions made by point charges *versus* those made by multipole moments. This type of work is theoretically useful but of reduced practical value because geometrical predictions of these small systems can now be routinely made *via* high level *ab initio* methods. Moreover, with the condensed matter phase in mind as a target application area, the study of small complexes has limited impact in the understanding of the condensed phase. For example, in order to understand liquid water (the ''solvent of life'') one needs to study complexes beyond the water dimer. Ultimately the comparison between structure and dynamics predicted by point charges and multipole moments needs to be made in the condensed matter phase. Here comparative studies are unfortunately rarer (*e.g.* ref. 2–5).

The Buckingham–Fowler model,[6] proposed in the 1980s, clearly demonstrated[7] the success of multipole moments in reproducing the peculiar geometrical features of van der Waals complexes made up of two polar molecules (also including CO_2, $HC{\equiv}CH$, N_2 and $H_2C{=}CH_2$), aided by the use of analytical second derivatives[8] of the multipolar electrostatic energy. This model involves a type of multipole moment obtained by Distributed Multipole Analysis (DMA).[9] This approach[10] benefits from the spherical tensor formalism, which through its irreducibility avoids the redundancies present in the Cartesian formalism. For example, there are only 7 spherical tensor octopole moments but there are $3^4 = 81$ Cartesian octopole moments. This formalism is also used in the current contribution but with an important difference: the decision on which part of the molecular electron density belongs to a given atom, is topological in nature rather than based on an arbitrary distance criterion as in DMA.

The criterion for partitioning in this work follows that of Quantum Chemical Topology (QCT).[11,12] QCT lets the paths of steepest ascent in the electron density carve out a finite volume in space, which then is the finite volume of the topological atom itself. The electron density within this finite volume is the only contributor to the atomic multipole moment. This protocol is independent of the nature (Gaussian, Slater, plane wave) of the basis functions, and it even works if there are no basis

[a] *Manchester Institute of Biotechnology (MIB), University of Manchester, 131 Princess Street, Manchester M1 7DN, UK. E-mail: pla@manchester.ac.uk*
[b] *School of Chemistry, University of Manchester, Oxford Road, Manchester M13 9PL, UK*

functions (*e.g.* Quantum Monte Carlo). To the contrary, DMA works only with Gaussian primitives and shifts the centre of a contributing Gaussian to the nearest multipole expansion site, which is typically a nucleus. In its original form, the DMA scheme breaks down in the presence of diffuse Gaussians, although this problem has been remedied[13] by borrowing an idea from so-called real space partitioning, which is opposite to its original basis-function space partitioning. Real space partitioning is exactly what QCT prescribes but, curiously, the success of such schemes in defining sensible atomic multipole moments has still not been documented in a recent edition of an authoritative textbook.[14]

In this article we present a comparison between the performance of multipole moments and point charges, for the first time combining both static and dynamic calculations. We now focus on a system, ultimately of relevance to the structure and dynamics of proteins in aqueous solution, which is larger than any of the systems previously studied,[15] which included $(H_2O)_n$ (n = 2 to 9), serine$\cdots(H_2O)_n$ (n = 1 to 5) and tyrosine$\cdots(H_2O)_n$ (n = 1 to 5). That study looked at the geometries of local energy minima of gas phase clusters, whereas here molecular dynamics results complement the picture obtained by *in vacuo* geometry optimisations. In particular, we investigate serine$\cdots(H_2O)_3$ in order to make contact with the previous study mentioned above.[15] So far, QCT multipolar electrostatics has only been examined for rigid molecules, in the context of potential energy explorations or molecular simulations. In order to rigorously examine the difference between QCT multipolar electrostatics and point charge electrostatics, we must set up systems such that the only distinction between them is the electrostatic interaction. This imposes the rigid-body constraint on the point charge calculations. The second important difference between the current study and the one before[15] is the more complete hydration of the serine. Whereas before the serine dipped with its hydroxyl group into an intact water cluster, we now construct a serine with a single hydration layer of 32 water molecules. An even more realistic ''immersed'' serine system puts the serine in a simulation box full of water molecules, focusing on the kinetic effects and the impact of a more realistic representation of hydration.

The current work serves to make the case again in favour of multipolar electrostatics,[16–19] especially in the light of the inherently limited point charge paradigm. For example, in the case of halogen bonding, the latter hits fundamental problems[20] that can be overcome by the multipolar paradigm.[21]

Returning to amino acids (and hence peptides), multipolar electrostatics creates more structure in the surrounding water layers than point charges do, as we will demonstrate. This is an important difference that may influence the experimental interpretation of long-range protein–water dynamics[22] or the elucidation of terahertz and far-infrared spectroscopy of aqueous peptides,[23] for example.

2 Methodology

In this study we will compare two different levels of realism, one *in vacuo* and one in the condensed phase. At the first level we studied a cluster consisting of a single serine molecule with either 3 or 32 water molecules *in vacuo.* The former will be referred to throughout as the *small cluster*, while the latter is called the *full cluster*. For the second level, which is more realistic than the first, we introduced a system consisting of a serine molecule immersed in a cubic box containing 276 water molecules at ambient density, which we discuss below in more detail. We call this system the *immersed system.*

The initial geometry of the *full cluster* was generated by manually distributing water molecules around the serine molecule. This starting configuration was geometry-optimised using the *ab initio* program GAUSSIAN03[24] at HF/6-31G(d) level. This low level of theory [HF/6-31G(d)] is unfortunately dictated by the Gaussian geometry optimization of the serine with 32 waters. Several attempts to optimize with a higher level of theory led to geometries that failed to convergence within the default criteria of GAUSSIAN and within reasonable computation times. As a result, for sake of consistency, we decided to continue with this (admittedly rather low) level of theory, in order to make proper and full comparisons throughout the potential energy surface exploration and the molecular dynamics simulations. We should emphasise that this study focuses on the difference between point charges and multipole moments rather than on the impact of the level of theory, which has been subject to an earlier study[25] from our lab.

We have to keep in mind that multipolar electrostatics is, in this work at least, imposing the rigid body constraint. Hence, comparison with clusters using the multipolar QCT electrostatics enforce all participating molecules to be rigid, even if described by point charges, which normally do not obey a rigid body constraint. Therefore, the bond lengths and angles of all 32 water geometry-optimised molecules were averaged into a single, symmetrised and fixed water geometry. The original optimised geometry of each of the 32 water molecules was then substituted by this single averaged water geometry. The resulting cluster was then re-optimised at the same level of theory (but with all intra-molecular degrees of freedom fixed). This final cluster is called the *reference cluster* because it was used as the reference geometry for all gas phase comparisons. This reference cluster will also be the starting point for all subsequent geometry optimisations using point charges and multipole moments.

A stripped down version of the reference cluster with only 3 water molecules was extracted from the full cluster. The extracted cluster was optimised again with the same intra-molecular constraint and at the same level of theory. This cluster will be referenced as the *small cluster* in the following sections.

For the immersed system, the geometries of the serine and water molecules were set as identical to those used in the reference cluster. The side of the box is approximately 20 Å long. All molecules are treated as rigid bodies without intra-molecular degrees of freedom.

2.1 Computational details

A number of computational tools were used to facilitate this work. The 2011.10 version of the computer program MOE[26] was used for energy minimisations in conjunction with various point-charge based potentials. We used built-in potentials to

carry out optimisations. Minimisations were executed with the following settings: the stopping criterion for the gradient was set to 0.001 kcal mol^{-1} $Å^{-1}$, the solvation model was set to "gas phase" and the potential's cutoff was disabled in all calculations.

All calculations involving QCT potentials were carried out using the simulation program DLMULTI,[27] which is a modified version of DLPOLY.[28,29] An important feature of DLMULTI is the implementation of Ewald summation, which has been generalised to handle high-rank multipole moments. This implementation is consistent with the spherical tensor formalism that governs molecular interactions inside the simulation box.

There are no Lennard-Jones (LJ) parameters that have been optimised for the multipole moments used in the current work. There are two approaches to overcome this absence of data: one modifies a set of LJ parameters, while the other adopts a different set without alteration. The first approach corresponds to LJ data set 1, simply referred to LJ-1. This set contains a single pair of LJ parameter (σ = 3.15 Å and ε = 0.649 kJ mol^{-1}), which was approximately adjusted to give a reasonable interatomic distance for a pair of hydrogen bonded atoms (*e.g.* oxygen of water and nitrogen of serine). This set is applied to all non-hydrogen elements in the system. This parameter set was constructed for sake of simplicity only. The second LJ set, called LJ-2, consists of only the LJ parameters of the PFROSST potential. This set provided the LJ parameters for all the elements in the small cluster, the full cluster and the immersed system. With these two sets, we are able to examine the impact of LJ interactions on the system, and force comparisons to reveal differences solely caused by electrostatic interactions.

For optimisations using DLMULTI or DLPOLY, the total energy of the system was minimised by performing molecular dynamics using the built-in "zero" option, which loosely speaking corresponds to a zero Kelvin run. These runs were typically executed with the following settings: a timestep of 1.0 fs to integrate the equations of motion, an Ewald precision parameter set to 1×10^{-6}, the potential's cutoff distance set to half the box size (150 Å) and a total calculation length of 20 000 time steps. This large box size is use to mimic a gas phase environment for the cluster. All molecules (in any of the three cases) were treated as rigid bodies in order to conduct a like-for-like comparison between point charges and multipole moments.

For calculations with the immersed system, DLMULTI or DLPOLY was used to perform calculations in the NPT ensemble. The prescribed temperature and pressure was 298 K and 1 atm, respectively. The potential's cutoff distance was set to be 8.0 Å, and 4 nanosecond runs were performed to collect enough data for the Spatial Distribution Function (SDF) analysis. Other settings of these simulations are identical to those used in the optimisation calculations (*i.e.* "zero" option).

2.2 Electrostatic potentials

In this study, we aim to compare two categories of potentials: point-charge based potentials and an electrostatic potential based on QCT multiple moments. The point-charge based potentials include PFROSST,[30] MMFF94x,[31] OPLS-AA[32] and TAFF.[33] We will not discuss the details of each potential as they can be found in the relevant references. However, a very brief account of the QCT electrostatic potential[34] is in place here. The Coulomb energy between two topological atoms Ω_A and Ω_B is given by eqn (1),

$$E_{AB}^{\text{Coulomb}} = \sum_{\ell_1=0}^{\infty}\sum_{\ell_2=0}^{\infty}\sum_{m_A=-\ell_A}^{\ell_A}\sum_{m_B=-\ell_B}^{\ell_B} T_{\ell_A m_A}^{\ell_B m_B}\int_{\Omega_A} d\mathbf{r}_1 \rho_{\text{tot}}(\mathbf{r}_1) R_{\ell_A m_A}(\mathbf{r}_1) \times \int_{\Omega_B} d\mathbf{r}_2 \rho_{\text{tot}}(\mathbf{r}_2) R_{\ell_B m_B}(\mathbf{r}_2) \quad (1)$$

where $R_{\ell m}$ are regular spherical harmonics[34] and ρ_{tot} is the total charge density (electron and nuclear) of either water or serine in the gas phase. The results of the integrals are the respective multipole moments of the topological atoms. The interaction tensor T is a purely geometric object whose scalar components are analytical functions of the mutual orientation of local axis systems centered on nuclei A and B, as well as the internuclear vector.[8,35] The infinite summations (over ℓ_A and ℓ_B) must be truncated to allow practical use of eqn (1). The appropriate number of expansion terms is decided by balancing the higher computational cost of their inclusion against how rapidly the interaction energy converges. The extent of the expansion is monitored by a value termed the maximum rank L, which is defined as

$$L = \ell_A + \ell_B + 1 \quad (2)$$

In all following calculations using QCT electrostatics, L is fixed to a value of 5 because a lower value proved to be insufficient to obtain the experimentally observed undulation at longer range in the $g(r)$ of liquid water,[36] which is a signature of ordered microstructure.

2.3 Analysis of results

2.3.1 Small and full cluster. A straightforward way to compare the optimised clusters (small and full) against the reference cluster is to translate and rotate the entire cluster so that the serine molecules have identical location and orientation. The resulting cluster can be then combined in the same picture. This procedure enables visual examination of differences that may not be obvious when expressed numerically.

For the small cluster, we conducted simple visual inspection of the superimposed clusters as a way of assessing the difference due to the use of different potentials. For the full cluster, we used three approaches to quantify the deviation from the reference cluster. The first approach is the displacement of each atom of the (rigid) water molecule in the cluster away from its corresponding reference location. We have evaluated two different displacements, one using only the oxygen atom, and the other using all three atoms in water. The former displacement captures only translational effects while the latter also includes orientational effects.

The second approach consists of comparisons of angles. Since the optimised serine cluster contains an inter-connected hydrogen bonding network, it is useful to examine the angular characteristics of that network. There are two categories of hydrogen bonds within the cluster: (i) those between the oxygen or the nitrogen of the serine, and its hydrogen-bonded water

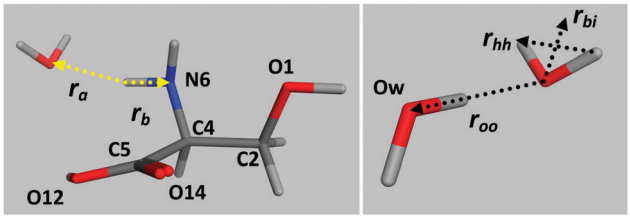

Fig. 1 Selected atomic labels and relevant vectors in (left) the serine···water complex and (right) the water dimer. The left panel shows the vectors used for the evaluation of the "3-atom-angle" (yellow), which is derived from the dot product between vectors $\mathbf{r}_a$ and $\mathbf{r}_b$. The right panel shows the vectors for the calculation of "O–O·H–H" (derived from $\mathbf{r}_{oo}$·$\mathbf{r}_{hh}$) and the "bisector" angle (derived from $\mathbf{r}_{oo}$·$\mathbf{r}_{bi}$).

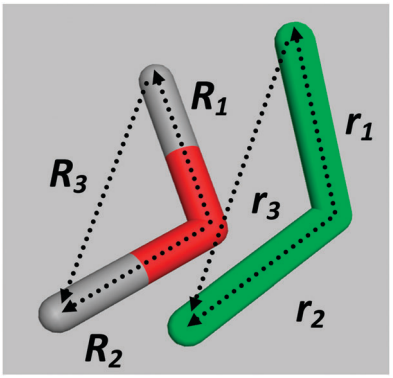

Fig. 2 Evaluation of alignment between 3 pairs of vectors derived from two corresponding water molecules. The alignment angle is derived from the dot product between two vectors, $\mathbf{R}_\alpha$·$\mathbf{r}_\alpha$ (α = 1, 2, or 3). These 3 alignment angles are designated in the main text and other figures as H1, H2 and HH, respectively. The two coloured molecule is the reference from the *ab initio* geometry optimised cluster and the green coloured molecule is its counterpart in the cluster using potentials.

and (ii) those formed between hydrogen-bonded water molecules. For either category we defined a "3-atom-angle", $X_{serine}\cdots H_{serine/water}\cdots O_{water}$ (where X can be oxygen or nitrogen from serine) (for an example see yellow fragment in Fig. 1, left). In the reference cluster, we have identified 7 such angles, which will be used later to quantify the similarity between geometries generated by various potentials. For the latter category, we monitor the angle between the vector joining the 2 oxygens and the vector between the two hydrogens of the acceptor molecule (see Fig. 1, right). For an optimised gas phase water dimer, this angle should be 90°, on symmetry grounds. In addition, the angle between the intermolecular OO axis and the ∠HOH bisector of the acceptor molecule was also monitored (see Fig. 1 for more detail).

The third and final approach focused on molecular alignment. For each water molecule, we examined its alignment with respect to its corresponding water molecule in the reference cluster. This alignment was measured by the angle between the vector in the water molecule in the reference cluster and the equivalent vector in the water molecule in the optimised cluster. For this study, we used three pairs of vectors to monitor the quality of alignment: the two OH bond vectors marked by $\mathbf{r}_1$ and $\mathbf{r}_2$, in one water molecule (see Fig. 2), and $\mathbf{R}_1$ and $\mathbf{R}_2$ in another water molecule. A third vector is used in this alignment measurement, that is, $\mathbf{r}_3$ and $\mathbf{R}_3$ in Fig. 2, which is formed by linking the two hydrogen atoms in the molecule. Note that the latter vector is identical to $\mathbf{r}_{hh}$ (in the right panel of Fig. 1). The alignment angle is defined as the arccosine of the scalar product between two unit vectors, $\frac{\mathbf{R}_\alpha}{|\mathbf{R}_\alpha|}\cdot\frac{\mathbf{r}_\alpha}{|\mathbf{r}_\alpha|}$ (α = 1, 2, or 3). These three alignment angles are designated as H1 (for α = 1), H2 (for α = 2), and HH (or α = 3), respectively.

2.3.2 Immersed serine. For the immersed system, we focused on examining the structural characteristics of the system by initially comparing the radial distribution functions (RDF) of various atoms. The RDF was evaluated using the build-in functionality of both DLPOLY and DLMULTI. We have also used Spatial Distribution Functions (SDFs) to examine the local environment around a central atom. The SDF presents a detailed picture showing both radial and angular dependency. The SDF gives an approximate indication of the number of neighbouring atoms next to the central atom. The SDF is represented by iso-surfaces of specific values (iso-values). A higher iso-value means the presence of more oxygen atoms and hence a higher water concentration. The detailed procedure to evaluate a SDF has been described in an earlier study.[5] For this study, we evaluated two kinds of SDFs: an *atomic SDF* and a *molecular SDF*. The former kind is calculated between serine's oxygen (or nitrogen) acting as a central atom and a neighbouring water's oxygen atom. The latter kind simultaneously inspects the distribution of water oxygens surrounding the nitrogen and the three oxygens of the serine.

3 Results and discussion

3.1 Small and full cluster

Fig. 3 compares the optimised geometries for the small cluster obtained by MOE, DLPOLY and DLMULTI, as well as GAUSSIAN (*ab initio*, reference cluster). Fig. 4 shows the same as in Fig. 3 but then for the full cluster. For the small cluster, using the location and orientation of the three labelled water molecules as a guide, we found that OPLS-AA compared poorly with all three waters having moved away from their reference positions. In the left panel of Fig. 3 molecule 1 seemingly ends up (as 1′) in a location close to one of the reference sites (labelled 3) but in reality it is situated far behind the molecule (as can be seen in the middle panel of Fig. 3, which is a 90° rotation of the left panel). The prediction made by MMFF94x (magenta) is equally unsatisfactory with all three waters having moved away from their reference locations and not close to any of the correct reference sites. The optimised TAFF (cyan) cluster is slightly better because the final locations of two of the waters are reasonably close to the reference sites. However, the orientations of all three waters are notably different from the reference.

For the remaining potentials (QCT using LJ-1 parameter set and PFROSST), the positions of all three waters are close to the reference but their location and orientation deviate from the reference ones to various extents. It is perhaps difficult to determine which one is superior because there is no clear winner in terms of location and orientation. For protein hydrations with multiple

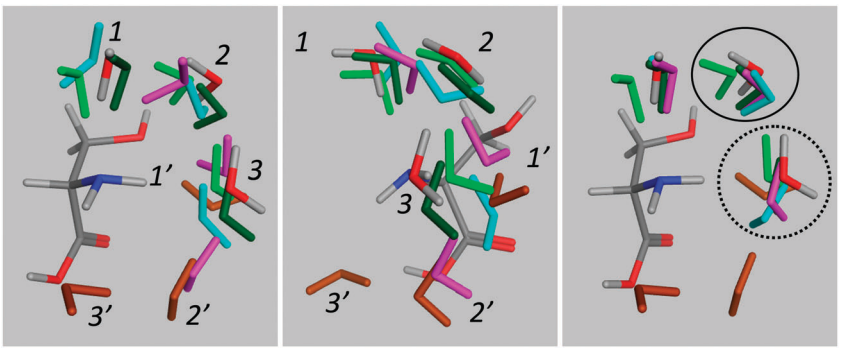

Fig. 3 Comparison of MOE, GAUSSIAN and DLPOLY optimisations for the small cluster. (Left) MOE, (middle) same as the left panel but the system is rotated to the left by 90° around the vertical axis and (right) DLPOLY, where LJ-1 parameters are used for all potentials. The reference cluster (generated by GAUSSIAN) is depicted in the usual atomic colours (red, blue, light and dark grey), QCT (green), MMFF94x (magenta), OPLS-AA (brown), PFROSST (dark green) and TAFF (cyan). The labels 1, 2 and 3 indicate the locations of the water molecules before and after (1′, 2′ and 3′) optimisation with OPLS-AA. The QCT cluster was obtained by using DLMULTI, always in conjunction with the "LJ-1" parameter set. For the right panel, the solid circled area contains waters from different potentials that are close together and possess similar orientation. The dotted circled area contains waters that are further apart and with notably different orientation.

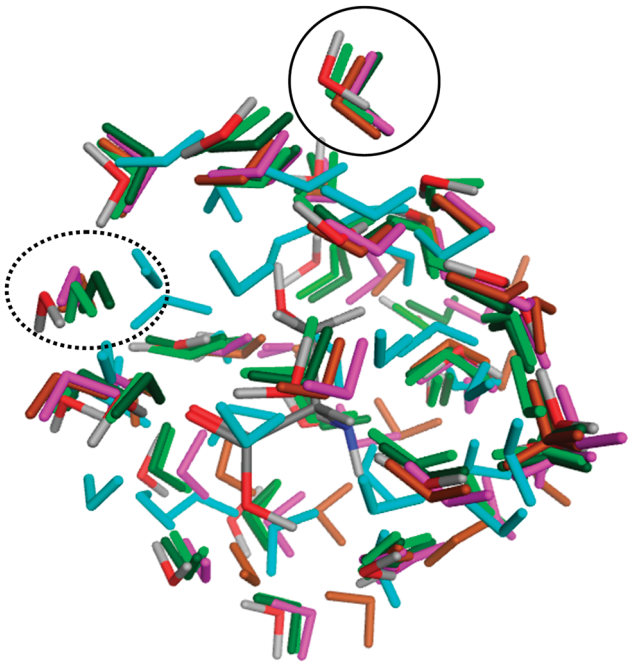

Fig. 4 Comparison of MOE, GAUSSIAN and DLPOLY optimisations for the full cluster. The colour scheme is identical to that of Fig. 3. The solid circled area contains waters from different potentials that are close together and possess similar orientation. The dotted circled area contains waters that are further apart and with notably different orientation. The QCT cluster was obtained using the "LJ-1" set.

water molecules, the fact that the QCT potential produces results that are similar to point-charge based potentials had been reported previously.[15]

The differences observed in these clusters are the combined result of two independent driving forces: the LJ potential and the electrostatic potential. In order to eliminate the effects of the former and solely focus on the influences of the latter, we made a slightly different comparison in which the contributions due to the LJ potential are made identical. Indeed, in the former comparisons of Fig. 3, the LJ parameters are defined as specified by the concomitant potentials (that have their own electrostatics). In the current new comparison, we made the LJ interaction within the cluster identical for all potentials by using the "LJ-1" parameter set. The right panel of Fig. 3 shows the outcome of such comparison as obtained by DLPOLY. The picture shows that the waters nearest to the serine's hydroxyl oxygen (top) are now gathered into two tightly packed groups (oval and left of oval) (not counting the QCT water, which always uses LJ-1) with limited variation in orientation. In contrast, the group of water molecules at the right hand side of the serine's nitrogen (dotted oval) have a very different orientation compared to the reference water, and are further apart from it. The difference in behaviour between the former two water groups on one hand, and the latter group on the other hand, seems to correlate well with whether the water molecule acts as a hydrogen bond acceptor or donor. The waters in the former groups (oval and left of oval, Fig. 3 right panel) are acting as donors while those in the group at the right of the nitrogen (dotted oval) are acceptors.

Surprisingly, the influence of the LJ potential seems to be very small for OPLS-AA (brown) because the DLPOLY result (right panel) is almost identical to that obtained by MOE (left panel). In contrast, both the MMFF94x and TAFF cluster significantly improved their similarity with the reference because all three water molecules are close to their respective reference locations. However, the orientation of one of the waters in the TAFF cluster still deviates markedly from the reference. In this comparison, the result for MMFF94x is remarkably similar to that for PFROSST as their respective waters are always found near each other. The resultant clusters from both potentials remain relatively close to the reference sites with some deviation in position and orientation. In fact, the agreement of these two clusters can be considered to be equal or even marginally better than that achieved by the QCT potential because the variation in location is comparatively smaller.

For point-charge based potentials, the order in decreasing similarity is: PFROSST ≈ MMFF94x > TAFF > OPLS-AA.

Table 1 List of atomic charges (in units of *e*) for QCT and various point-charge based potentials. The Δ_α (where α = serine + water, serine or water) represents the mean absolute difference from PFROSST charges for the specified molecule

	Atom	QCT	PFROSST	MMFF94x	OPLS-AA	TAFF
Serine	O1	−1.229	−0.680	−0.680	−0.683	−0.680
	C2	0.718	0.120	0.280	0.145	0.280
	H3	0.594	0.400	0.400	0.418	0.400
	C4	0.536	0.251	0.331	0.063	0.510
	C5	1.854	0.659	0.659	0.723	0.570
	N6	−1.171	−0.990	−0.990	−0.542	−1.080
	H7	−0.023	0.080	0.000	0.060	0.000
	H8	0.422	0.360	0.360	0.210	0.360
	H9	0.366	0.360	0.360	0.210	0.360
	H10	−0.019	0.080	0.000	0.060	0.000
	H11	−0.002	0.080	0.000	0.060	0.000
	O12	−1.299	−0.650	−0.650	−0.533	−0.570
	H13	0.670	0.500	0.500	0.310	0.400
	O14	−1.356	−0.570	−0.570	−0.500	−0.550
Water	O_w	−1.181	−0.860	−0.860	−0.834	−0.620
	H16	0.591	0.430	0.430	0.417	0.310
	H17	0.590	0.430	0.430	0.417	0.310
			$\Delta_{serine+water}$	0.028	0.089	0.090
			Δ_{serine}	0.034	0.074	0.105
			Δ_{water}	0.000	0.160	0.017

The similarity between PFROSST and MMFF94x is perhaps not surprising because the atomic charges of MMFF94x have the smallest difference when compared with PFROSST (see Table 1) and the difference is solely due to the serine molecule. Moreover, the charges of the serine atoms (O1 and N6 in Fig. 1) that formed hydrogen bonds with the water molecules are identical for both potentials. The charges of the two carbon atoms (C2 and C4) do have sizeable differences but their distance from the hydrogen bonded waters may mean that their influence on the final structure is limited.

For TAFF and OPLS-AA, in spite of the charge differences for the water molecule (Δ_{water} = 0.017*e* and 0.160*e*), the overall atomic charge differences with respect to PFROSST are comparable ($\Delta_{serine+water}$ = 0.089*e* and 0.090*e*). Nevertheless, the resultant clusters are found to be markedly different. The TAFF cluster has two waters close to their corresponding reference's location and orientation but the third water (right of nitrogen) is quite far away from the reference and adopted a rather different orientation. This could be attributed to the fact that the charges for serine atoms in this part of the molecule are roughly comparable to those in PFROSST. The unique behaviour of the OPLS-AA cluster is presumably due to the large variation in charges for N6 and C4, which resulted in the large displacement of all three waters from their reference positions.

The comparison in which the same LJ parameters (LJ-1) are used for all potentials (Fig. 3, right) also gives us the opportunity to examine the influences of the LJ potential on the final configuration of the cluster. This comparison demonstrated that the combined effects of LJ and electrostatic potentials cannot be predicted easily. The outcome depends on the intricate interplay between the two contributions to each potential, that is, electrostatic and LJ. The impact of the LJ potential to the final organisation of the cluster is dependent on the specific potential, which can vary from small (for OPLS-AA) to large (MMFF94x and TAFF). Even though LJ interaction is traditionally considered to be much weaker than electrostatic interaction (except at short range), our findings show that given the right circumstances, the LJ potential can have a decisive influence on the final geometry of the system. In particular, a comparison between the left and right panel in Fig. 3 shows a marked improvement for the TAFF and MMFF94x clusters.

The small cluster is a simplistic approximation for a hydrated serine molecule because it omitted many important interactions that exist in a real hydrated system. To enhance realism we carried out a detailed comparison between QCT (using DLMULTI) and all point-charge based potentials (using MOE) for the full cluster. For this comparison, we used both sets of LJ parameters, "LJ-1" and "LJ-2", for the QCT potential. The conditions for all calculations have been detailed in Section 2.1.

Fig. 4 shows the optimised geometries from all potentials rotated to the same reference frame and superimposed. The TAFF (cyan) cluster deviates most from the reference. The OPLS-AA (brown) cluster is slightly better but still has a considerable number of water molecules that are some distance away from their reference locations. For the other potentials (*i.e.* QCT, PFROSST and MMFF94x), it is difficult to decide visually which one is superior. The solid circle in Fig. 4 highlights closely packed water molecules with similar orientation as the reference, while the dotted circle marks water molecules with quite different relative orientations. Both water clusters (solid and dotted) have been observed in Fig. 3.

In order to quantify the deviations from the reference cluster, we have devised a selection of criteria to characterise the various clusters (details have been described in Section 2.3). Fig. 5 shows the mean deviation in atomic position between an atom in the reference cluster and the corresponding atom in a cluster generated by each of the six potentials tested. One mean deviation (blue) is obtained as an average over all atoms in the cluster, while the other (red) averages only over the oxygens. As expected, the deviation for all atoms is always greater than the values for oxygen atoms only. This is not surprising because the "all atoms" deviation also detects the deviation in molecular orientation, by the inclusion of the hydrogen atoms. The values in Fig. 5 confirm the order of similarity with the reference, deduced visually from Fig. 4. Moreover, Fig. 5 produces a more precise order for the potentials used in this work, starting with the largest deviation (and hence worse performance): TAFF > OPLS-AA > MMFF94x > PFROSST ≈ QCT LJ-1 ≈ QCT LJ-2. Looking at the last two models we can deduce that, even with different LJ parameters, the deviations are very similar. Hence, the impact of LJ parameters is limited. It is perhaps unexpected to see that the PFROSST results are so close to the QCT ones.

Fig. 6 shows the mean deviation (in degrees) between the three angles (the bisector angle, O–O·H–H angle and the "3-atom-angle" defined in the caption of Fig. 1) occurring in the full reference cluster and occurring in the corresponding cluster generated by each of the six potentials tested. These three angles are specifically chosen to monitor the angular

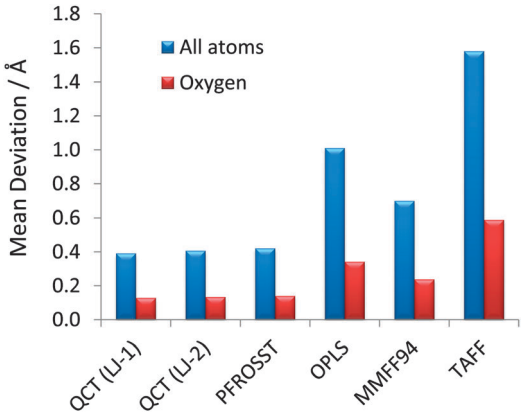

Fig. 5 Mean deviation (in Å) between the atomic positions occurring in the full reference cluster and occurring in the corresponding cluster generated by each of the six potentials tested. One mean deviation (blue) is obtained as an average over all atoms in the cluster, while the other (red) over only the oxygens.

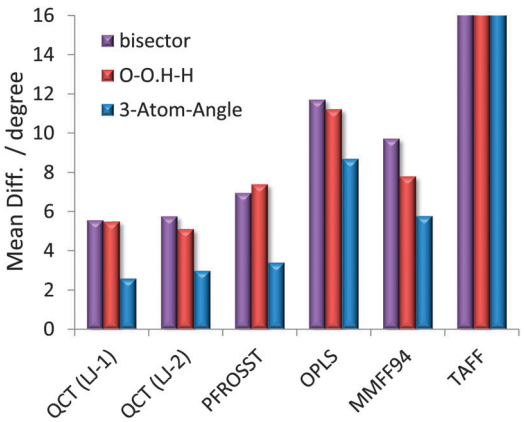

Fig. 6 Mean deviation (in degrees) between the three angles (see Fig. 1) occurring in the full reference cluster and occurring in the corresponding cluster generated by each of the six potentials tested. Note that the value for the TAFF potential has been truncated (with values originally between 17 and 44 degrees).

characteristics of two hydrogen bonded molecules. The first two angles occur between water molecules while the third occurs between a water molecule and serine molecule. The results confirm that QCT potential is consistently better than its point-charge based counterpart for all three angles and the difference between "LJ-1" and "LJ-2" is fairly small. However, PFROSST is again very close to QCT results. The amount of deviation in increasing order, starting with the smallest, is QCT LJ-1 $\approx$ QCT LJ-2 $<$ PFROSST $<$ MMFF94x $<$ OPLS-AA $<$ TAFF. The closeness of PFROSST to QCT is surprising because these angles are controlled by the directionality of hydrogen bond, which is normally lacking for point-charge based potentials while it is inherent in QCT electrostatics. The small difference between PFROSST and QCT could be attributed to the "networking" effect – whereby the waters surrounding the serine molecule are part of an intricately inter-connected hydrogen bonding network. Such a network orients a given molecule in its correct

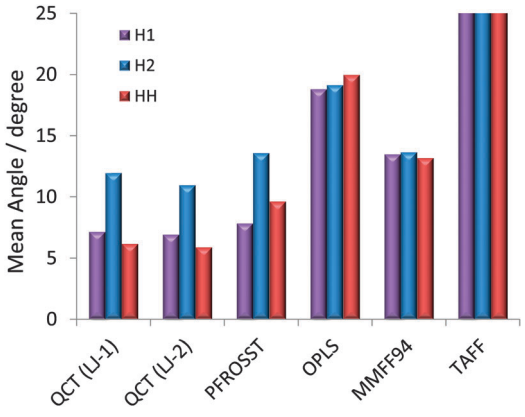

Fig. 7 Mean deviation (in degrees) of the three alignment angles (see Fig. 2) between water molecule occurring in the full reference cluster and those occurring in the corresponding cluster generated by each of the six potentials tested. Note that the value for the TAFF potential has been truncated (with values originally between 43 and 67 degrees).

position without having to rely on the directing effect of the oxygen lone pairs themselves. To make this explanation more concrete, let us consider the global minimum of the water dimer (see Fig. 1, right panel), where this network is lacking. Therefore the hydrogen-bond-accepting water is uniquely guided by its own lone pairs. In other words, the bisector angle of this system will be too large when using point charges only. If the hydrogen bond accepting water molecule in this dimer were surrounded by waters (as part of a network) then these surrounding waters start directing this water molecule, without needing a proper electrostatic description of its oxygen lone pairs. This effect has been observed before[15] in microhydration study with 1 to 5 water molecules.

Fig. 7 shows the mean deviation (in degrees) of the three alignment angles (see Fig. 2) between water molecule occurring in the full reference cluster and those occurring in the corresponding cluster generated by each of the six potentials tested. The potentials can be arranged in terms of quality of alignment, starting with the best alignment: QCT LJ-2 > QCT LJ-1 > PFROSST > MMFF94x > OPLS-AA > TAFF. This ordering is consistent for all three types of alignment angles (which are interrelated). Compared to other measures of similarity this is the only occasion where QCT LJ-2 demonstrates notable but albeit small differences from QCT LJ-1. For the point-charge potentials, the order is the same as in the two previous comparisons. The current comparison shows that the QCT potential can better reproduce the orientational character (including directionality) of hydrogen bonded water molecules in the cluster.

The results for the full cluster are consistent with those obtained for the small cluster. Even though the PFROSST potential seems to be slightly better than QCT for the small cluster, the more realistic full cluster comparison re-affirms that QCT potential has greater ability to reproduce the reference cluster. Nevertheless, PFROSST is the best potential among the point charge based potentials.

3.2 Immersed serine system

We have used the immersed system to conduct a more precise comparison in an environment that is closer to real physical conditions.

Table 2 Average electrostatic, LJ energies and system volume from simulations of the immersed system using the QCT and PFROSST potentials

	PFROSST	QCT
Electrostatic/kJ mol^{-1}	−12914.0	−13778.0
LJ/kJ mol^{-1}	2651.4	4624.2
Volume/Å^3	8470.7	8395.9
Density/g cm^{-3}	0.9945	1.003

By using the "LJ-2" parameter set, this comparison allows us to examine the differences between the point charges and multipole moments directly by minimising interference from other factors. More importantly, the immersed system enables us to examine the hydration of serine in a more realistic environment with finite temperature.

The results for the system energies and density (Table 2) were obtained from production runs of 4 ns. The difference in the mean system density was less than 1% but the energetics were considerably different: the QCT system had a lower electrostatic energy (6%) but a higher LJ energy (43%). These differences compensate each other because the difference in the total energy (not listed) is only 11%. The small difference in electrostatic energy is perhaps surprising because the atomic charges in the QCT system are generally larger than those of the PFROSST potential (see Table 1). We presume that the contribution from higher multipole moments counterbalanced the contribution from charge–charge interactions. Secondly, it is likely that the steepness of the r^{-12} part in the LJ potential makes a very small change in volume (and hence density) cause a very large change in repulsive (and hence LJ) energy. Overall, the fact that the densities of the two systems are so similar makes a comparison between the two systems meaningful, and hence useful insights on the effects of point-charges and multipole moments can be learnt.

The immersed system is capable of providing information on the immediate environment of atoms in the system. An RDF is a useful means to present this kind of information. Fig. 8 compares the oxygen–oxygen RDFs, denoted $g_{O_w\text{–}O_w}(r)$, of water obtained from experiment[37] and three computational methods (QCT, PFROSST and "High Level QCT"[25]). The latter and experiment refer to pure liquid water. "High level QCT" also refers to MD simulations using rigid-body multipolar electrostatics at CCSD/aug-cc-pVTZ level (taken from Fig. 5 in ref. 25). The profiles of PFROSST (dashed red) and QCT (solid green) are notably different for each other. Indeed, the QCT water molecules reveal much more structure at medium and long range, as indicated by the clear peaks and valleys that appear in the entire range (up to 8 Å). We note that multipole moments from such a low level theory wave function still reproduce the essential qualitative features that exist in liquid water (solvating the serine) at room temperature. On the other hand, the PFROSST water behaves more like a gas with one peak in the RDF. Clearly, experiment (solid brown) shows more local structure than PFROSST but not as pronounced as QCT. However, when QCT is used with a combination of a high level of theory CCSD/aug-cc-pVTZ and LJ parameters optimisation then a strong agreement with experiment is obtained.

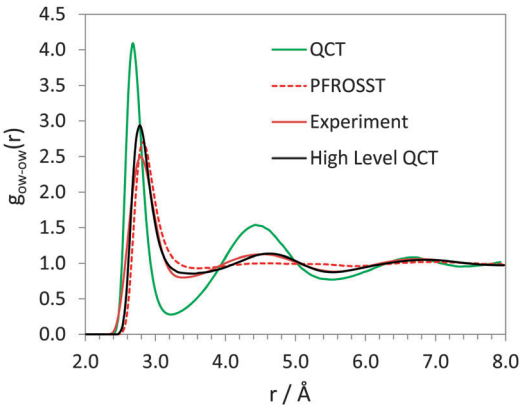

Fig. 8 Comparison of water–water radial distribution functions (RDF). The label "Experiment" refers to an interrogation of scattering data by modern data analysis and computer simulations tools (by A. K. Soper in ref. 37). The label "high level QCT" refers to MD simulations of pure water using rigid-body multipolar electrostatics at CCSD/aug-cc-pVTZ level (taken from Fig. 5 in ref. 25).

As a final comment we mention that the location of the first peak is slightly different as well, appearing at 2.7 Å and 2.8 Å, for QCT and PFROSST, respectively. This smaller distance for the first peak does not seem to have any significant implication on the system density.

The RDFs for the heavy atoms of serine (O1, O14, N6 and C5) and water oxygen (O_w) are shown in Fig. 9. These RDFs are not as smooth as $g_{o_w\text{–}o_w}(r)$ because of the presence of a single serine molecule in the system. However, we believe all key features have been captured in the RDFs presented here. On the whole, the RDFs from QCT have characteristics that are distinct from the PFROSST system, for each panel in Fig. 9. The QCT system has more fine structure in the RDFs (*e.g.* the number and position of both peaks and valleys). Similarly to the $g_{o_w\text{–}o_w}(r)$ profile in Fig. 8, the first peak in the QCT system is consistently located at a shorter distance and the difference is between 0.2 Å (for $g_{o14\text{–}o_w}(r)$ and $g_{o1\text{–}o_w}(r)$) and 0.3 Å (for $g_{n6\text{–}o_w}(r)$). Unusually, the comparison for $g_{c5\text{–}o_w}(r)$ shows that the QCT system has a small shoulder before the main peak (at 3.5 Å). Secondly, it is unusual that the position of the QCT main peak is slightly further (by 0.1 Å) than the main peak in the PFROSST system. In addition, the first peak in the PFROSST system has a shoulder that does not exist in the QCT system.

The findings from the RDFs point to considerable differences in the immediate environment around the serine molecule between QCT and point-charge based potentials. Similarly, the local environment of a given water molecule also demonstrates significant differences between these two types of potentials. We believe that the presence of additional peaks and valleys indicates that the immediate surroundings of the serine (and water) in the QCT system is more "structured". However, the nature of RDF prevents us from gaining detailed understandings relating to the spatial distribution of molecules within the system. To this end, we first examined O_w–O_w SDFs, which only involve water molecules. Fig. 10 shows these SDFs for a cut-off radius of 4.0 Å, focusing on the first hydration

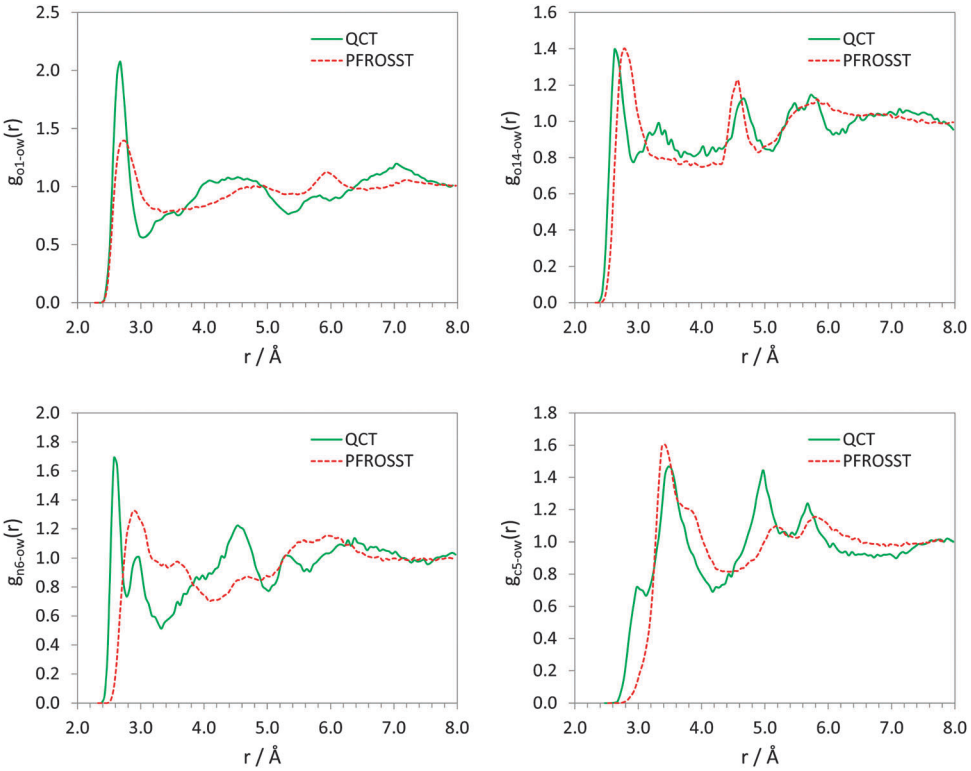

Fig. 9 Radial distribution functions (RDF) for selected atoms (from top left clockwise: O1, O14, N6 and C5, see Fig. 1) of the serine molecule and the oxygen atom of any water molecule.

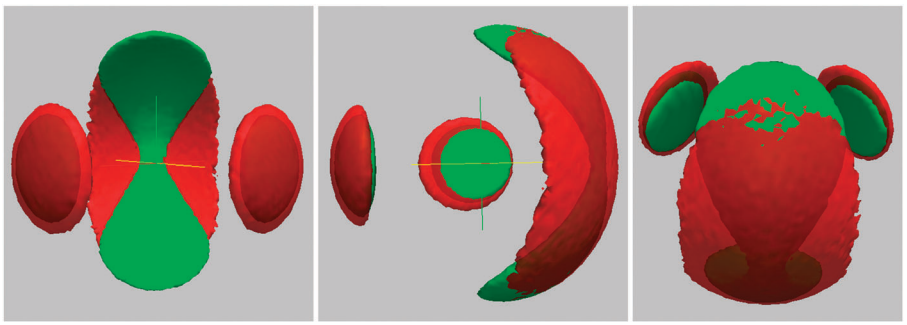

Fig. 10 Comparison between the QCT (green) O_w–O_w SDF and the PFROSST (red) O_w–O_w SDF (r_{cut} = 4.0 Å and surface iso-value = 4.0), shown from three different points of view.

shell, while Fig. 11 shows them for a cutoff of 6.0 Å, highlighting both the first and second hydration shell. Note that the iso-value for Fig. 11 is half that of Fig. 10, which therefore reveals sparser and fuzzier structures that are also larger (the latter not being obvious because of the different scales used between both figures).

In the centre of the axis system of Fig. 10 one places the oxygen (in a water of interest). The two lobes left and right of this axis system show the oxygens of waters that are hydrogen bonded to the central water, each accepting a hydrogen of the central water. Inside each red lobe is a slightly smaller green lobe (with non-coinciding centres), indicating that QCT (green) and PFROSST (red) are very similar. However, the green and red

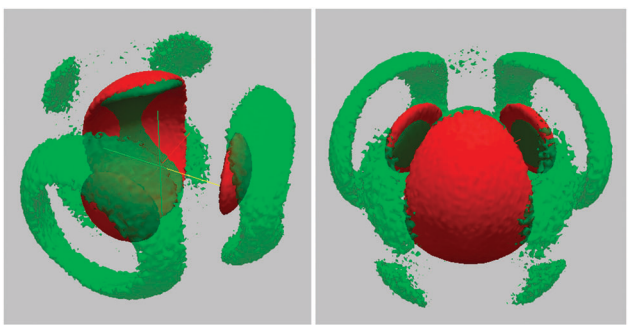

Fig. 11 Comparison between the QCT (green) O_w–O_w SDF and the PFROSST (red) O_w–O_w SDF (r_{cut} = 6.0 Å and surface iso-value = 2.0), shown from two different points of view.

lobes in the centre of Fig. 10 differ greatly. These lobes represent oxygens in waters that are hydrogen bonded to the central oxygen atom as hydrogen donors. The QCT distribution (green) displays a unique gradual narrowing towards the molecular plane (of the central molecule), which has also been observed in our previous study of liquid water using QCT potentials derived from higher level of theory.[4,25] This narrowing implies that the QCT potential reproduces the tetrahedral character in a more pronounced way than the PFROSST potential (red).

At longer range (cutoff of 6.0 Å, Fig. 11), the difference between QCT and PFROSST is even more striking because features that existed in the QCT system are entirely missing in the PFROSST system. In the QCT system, we observe a (green) torus-like structure around each of the two lobes. This torus is made up of waters that are hydrogen bonded to waters in the lobes. Behind the large and compact concave disk behind the lone pairs of the central water, there are four additional fuzzy distributions (in green). We believe that these four distributions correspond to waters that are hydrogen bonded to waters in the concave disk, which contains hydrogen donor waters.

In summary, the distribution pattern observed in the SDFs is consistent with that observed in the RDFs (Fig. 8). Indeed, the SDF at shorter range (Fig. 10), at 4 Å, corresponds to the first peak in the RDF while the longer ranged SDF distribution (Fig. 11), 6 Å is related to the second peak of the RDF. Fig. 8 shows the absence of a second peak in the PFROSST RDF, which is consistent with the lack of any distribution in the PFROSST SDF (red) at longer range. The opposite is true for QCT where the second peak in Fig. 8 (solid) is clearly matched by substantial distributions (green) in Fig. 11. Hence, *the QCT system contains long range hydrogen bonding networks with tetrahedral characteristics while the PFROSST system is non-tetrahedral, even at short range.*

The spatial distribution of waters around the serine molecule was examined by looking at the SDFs of the oxygen (O1, O12 and O14) and nitrogen (N6) atoms with a cutoff distance of 4.0 Å. Fig. 12–15 shows the results for O1, O12, O14 and N6, respectively. We now discuss these four figures in turn. For O1,

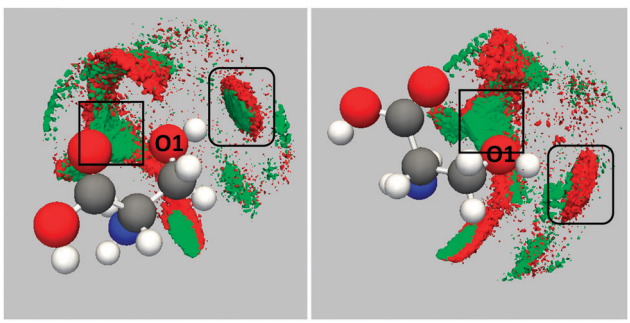

Fig. 12 The O1···O_w SDF (r_{cut} = 4.0 Å and surface iso-value = 6.0): a comparison between QCT (green) and PFROSST (red) seen from two different points of view (left and right panel). The black rectangle identifies the region of SDF containing water molecules that are hydrogen bonded to O1 as donors. Similarly, the round-corner rectangle marks the SDF region that contains acceptor water molecules.

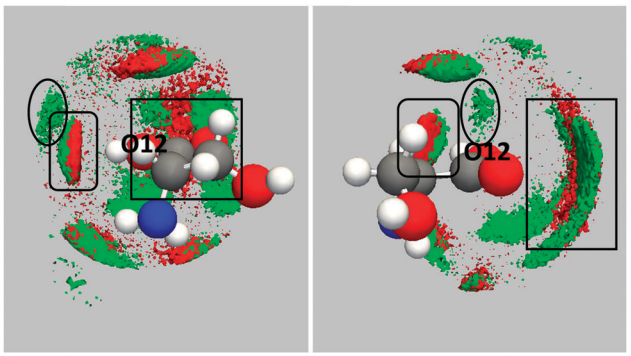

Fig. 13 The O12···O_w SDF (r_{cut} = 4.0 Å and surface iso-value = 6.0): a comparison between QCT (green) and PFROSST (red) seen from two different points of view (left and right panel). The oval marks the distribution of waters that exist in the QCT system only. The rectangle and rounded rectangle have the same meaning as in Fig. 12.

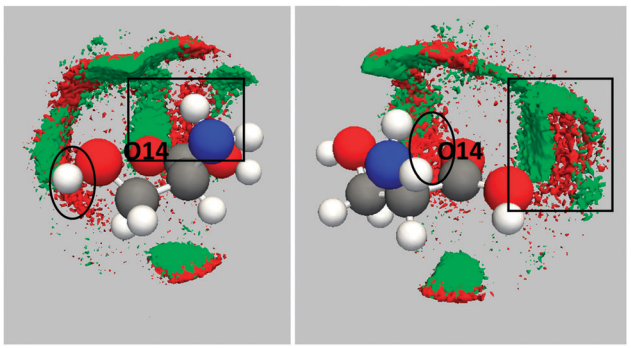

Fig. 14 The O14···O_w SDF (r_{cut} = 4.0 Å and surface iso-value = 6.0): a comparison between QCT (green) and PFROSST (red) seen from two different points of view (left and right panel). The black rectangle marks the same distribution features observed in Fig. 13. The oval identifies the distribution unique to the PFROSST potential.

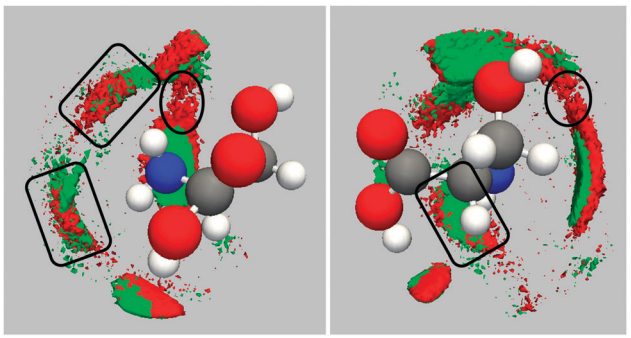

Fig. 15 The N6···O_w SDF (r_{cut} = 4.0 Å and surface iso-value = 6.0): a comparison between QCT (green) and PFROSST (red) seen from two different points of view (left and right panel). The oval marks the region where water molecules occur only in the PFROSST system. The rounded rectangle has the same meaning as in Fig. 12.

there are notable differences between QCT and PFROSST system, but similarities can also be identified. The most noticeable similarity is the lobe marked with a round-corner black

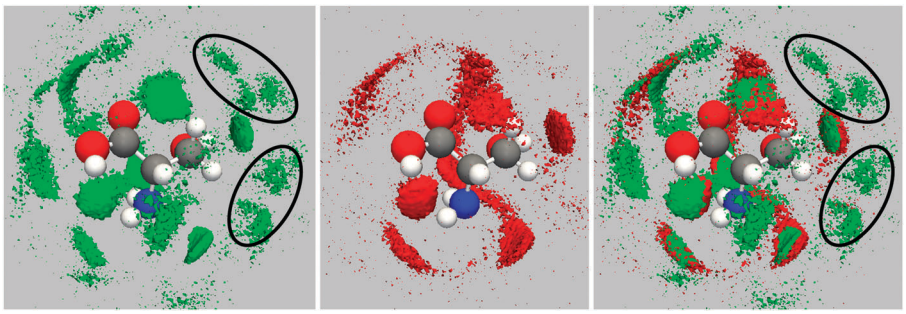

Fig. 16 Comparison of the molecular SDFs (r_{cut} = 6.0 Å and surface iso-value = 7.0). The molecular SDF for QCT (left) and that for PFROSST (middle) are coloured in green and red, respectively. The picture on the right is produced by overlapping these two molecular SDFs.

rectangle in Fig. 12 (left and right panel). This lobe corresponds to water molecules that are hydrogen bonded to the hydrogen atom attached to O1. This lobe-shaped distribution is typical for waters that are hydrogen bonded to an OH group. This type of distribution was observed before in $O_w\cdots O_w$ SDFs (see Fig. 10). However, the radius of the lobe for PFROSST is slightly bigger than for QCT (see Fig. 12), which suggests that the hydrogen bond in the QCT system is more directional and localised. The shape of the collection of donor waters that are hydrogen bonded to O1 (black rectangle, Fig. 12) is very different from that in the water–water case (Fig. 10, $O_w\cdots O_w$ SDF). Perhaps this is caused by these donor waters also interacting with O14 (carbonyl oxygen), which is close enough. The comparison of Fig. 12 shows that the QCT and PFROSST SDFs are broadly similar but there are regions where the PFROSST system has considerably more waters present. Also, when the two SDFs (green and red) do coincide with each other, the distribution of water in the QCT system is more localised.

Looking at Fig. 13, the most noticeable feature from the SDFs for O12 (in the hydroxyl group) is that the QCT and PFROSST distributions of donor waters (within black rectangle) barely overlap. In addition, the oval identifies a region where water molecules appear only in the QCT system. The proximity between O14 (carbonyl) and O12 (which both appear in the carboxylic acid group) means that the features (marked by the rectangle) in the SDFs are common to both oxygen atoms. Each panel of Fig. 13 shows two distinctive strips of waters in the QCT system, as marked by the rectangle. In contrast, the (red) waters of the PFROSST system appear in one single strip, which is largely situated in between the two green QCT strips. This contrast can again be attributed to the directionality of the QCT potential.

Fig. 14 compares the QCT and PFROSST water SDFs from the point of view of O14 (carbonyl). Apart from the region marked by the black rectangle, the remaining distributions (red and green) in this SDF seem to overlap. However, there is at least one region where only PFROSST shows waters (red oxygens), indicated by the oval.

The comparison of water SDFs from the point of view of N6 (Fig. 15) shows that the QCT and PFROSST distributions coincides in many places. However, we can also identify regions where only PFROSST puts waters (marked by the oval in either panel). The most unusual feature in the comparison is the shape of the water distribution marked by the round-corner rectangle. This distribution corresponds to waters (acceptor) that are hydrogen bonded to a hydrogen atom attached to N6. For O1 and O12, these waters coalesce into a circular lobe. However, for the nitrogen atom, the shape is distinctively non-circular for either QCT or PFROSST system.

So far, we have examined the SDFs from a specific atom's perspective and used a relatively short cutoff (4.0 Å) to evaluate the SDF, which encompasses the first peak in the atom's RDF. This short-range SDF provides some insight into the first hydration shell. A longer range molecular SDF, now combining the points of view of all oxygens and nitrogens in serine (cutoff of 6.0 Å), is shown in Fig. 16. The molecular SDF enables an overall understanding of the complete environment of the serine molecule. Many features that appear in Fig. 16 are identical to those observed in the first hydration shell. However, there are extra features that are unique to the QCT system (see regions marked by an oval). These regions indicate the presence of considerable amount of water appearing at distances between 4.0 and 6.0 Å from the nearest heavy atom (O or N). At this range, many small green or red dots appear elsewhere, away from the oval-marked regions, and surround the entire serine. The pattern of the red dots differs from that of the green dots in that the green dots coalesce more, in small clusters, whereas the red dots end up in a more diffuse arrangement.

The observations drawn from the SDFs are in line with those obtained from the RDFs. The discrepancies between the RDF of QCT and PFROSST show up as a variation in the location and number of waters in the SDFs. At distances below 4.0 Å, the SDFs of QCT and PFROSST do overlap to some extent but more importantly, display notable differences. At longer range, beyond 4.0 Å, only the QCT system exhibits the existence of substantial distributions at specific locations around the serine molecule.

4 Conclusion

We have contrasted the multipolar electrostatics *versus* point charge electrostatics by examining a hydrated serine system. In particular, we have used QCT multipole moments and four types of well-known point charges (PFROSST, MMFF94x, TAFF

and OPLS-AA) on a system containing a single hydrated serine molecule. At one level, we studied a serine with either 3 (small cluster) or 32 (full cluster) water molecules *in vacuo.* All along, a reference cluster consisting of a local energy minimum obtained at *ab initio* level was used for comparison. At the second level, which is more realistic we immersed the serine in a cubic box containing 276 water molecules (immersed system).

The results of the two *in vacuo* clusters were consistent with each other. Based on sensitive geometrical criteria devised in this study, QCT multipole moments best reproduce the reference cluster geometry while PFROSST charges come at a very close second. We believe that the similarity between QCT moments and PFROSST charges is partially due to the highly interlocking hydrogen bonding network surrounding serine, which prevents multipole moments to show their superior directionality. This phenomenon has been observed in our earlier study of protein hydration. The OPLS-AA and TAFF charges are consistently the worst in reproducing the reference geometry. Interestingly, Lennard-Jones interaction can be a decisive factor in determining the final geometry of the hydrated cluster even though it is traditionally considered to be much weaker than electrostatic interaction.

The results of the immersed cluster (as obtained by molecular dynamics simulation) demonstrate that the *in vacuo* cluster is rather limited in its ability to mimick a hydrated serine. The difference between QCT multipole moments and PFROSST charges are revealed in a different light: the immersed system illustrated that the subtle variations in the full cluster are now replaced by more pronounced differences in the distribution of water molecules around serine. There are significant disparities in the first hydration shell, which we attributed to the ability of QCT multipole moments to reproduce the directionality of the hydrogen bond. More importantly, the QCT system also possesses long-range structure in the second hydration shell, which is completely absent in the PFROSST system.

These results demonstrate the importance of dynamic effects in that differences in the distribution of water molecules surrounding a serine may significantly alter the energetics and ultimately the thermodynamic properties of the hydrated system. Consequently, geometry optimisation alone is inadequate in revealing the complicated nature of such systems.

References

1 S. Cardamone, T. Hughes and P. L. A. Popelier, 2013, in preparation.
2 G. M. Day, J. Chisholm, N. Sham, W. D. S. Motherwell and W. Jones, *Cryst. Growth Des.*, 2004, **4**, 1327.
3 S. Y. Liem, M. S. Shaik and P. L. A. Popelier, *J. Phys. Chem. B*, 2011, **115**, 11389.
4 S. Y. Liem and P. L. A. Popelier, *J. Chem. Theory Comput.*, 2008, **4**, 353.
5 M. S. Shaik, S. Y. Liem, Y. Yuan and P. L. A. Popelier, *Phys. Chem. Chem. Phys.*, 2010, **12**, 15040.
6 A. D. Buckingham and P. W. Fowler, *Can. J. Chem.*, 1985, **63**, 2018.
7 P. L. A. Popelier, A. J. Stone and D. J. Wales, *Faraday Discuss.*, 1994, **97**, 243.
8 P. L. A. Popelier and A. J. Stone, *Mol. Phys.*, 1994, **82**, 411.
9 A. J. Stone, *Chem. Phys. Lett.*, 1981, **83**, 233.
10 U. Koch, P. Popelier and A. Stone, *Chem. Phys. Lett.*, 1995, **238**, 253.
11 R. F. W. Bader, *Atoms in Molecules. A Quantum Theory.*, Oxford Univ. Press, Oxford, Great Britain, 1990.
12 P. L. A. Popelier, *Atoms in Molecules. An Introduction.*, Pearson Education, London, Great Britain, 2000.
13 A. J. Stone, *J. Chem. Theory Comput.*, 2005, **1**, 1128.
14 A. J. Stone, *The Theory of Intermolecular Forces. Second Edition.*, Clarendon Press, Oxford, 1st edn 2013, vol. 32.
15 M. S. Shaik, M. Devereux and P. L. A. Popelier, *Mol. Phys.*, 2008, **106**, 1495.
16 N. Gresh, S. A. Kafafi, J.-F. Truchon and D. R. Salahub, *J. Comput. Chem.*, 2004, **25**, 823.
17 S. L. Price, in *Rev.in Comp.Chem.*, ed. K. B. Lipkowitz and D. B. Boyd, Wiley, New York, USA, 2000, vol. 14, p. 225.
18 S. L. Price, *Acc. Chem. Res.*, 2009, **42**, 117.
19 P. L. A. Popelier, *Curr. Top. Med. Chem.*, 2012, **12**, 1924.
20 W. L. Jorgensen and P. Schyman, *J. Chem. Theory Comput.*, 2012, **8**, 3895.
21 A. J. Stone, *J. Am. Chem. Soc.*, 2013, **135**, 7005.
22 K. Meister, S. Ebbinghaus, Y. Xu, J. G. Duman, A. DeVries, M. Gruebele, D. M. Leitner and M. Havenith, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 1617.
23 T. Ding, T. Huber, A. P. J. Middelberg and R. J. Falconer, *J. Phys. Chem. A*, 2011, **115**, 11559.
24 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. J. Montgomery, J. T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *GAUSSIAN03*, In Gaussian, Inc., Pittsburgh PA, 2003.
25 M. S. Shaik, S. Y. Liem and P. L. A. Popelier, *J. Chem. Phys.*, 2010, **132**, 174504.
26 MOE; 2005.06 edn, Molecular Operating Environment (MOE), Chemical Computing Group Inc., Montreal, Canada, http://www.chemcomp.com, 2005.
27 M. Leslie, *Mol. Phys.*, 2008, **106**, 1567.
28 W. Smith, I. T. Todorov and M. Leslie, *Z. Kristallogr.*, 2005, **220**, 563.

29 I. T. Todorov, W. Smith, K. Trachenko and M. T. Dove, *J. Mater. Chem.*, 2006, **16**, 1911.
30 PFROSST, Molecular Operating Environment (MOE), 2012.10 edn, Chemical Computing Group Inc., Montreal, Quebec, Canada, 2012.
31 T. A. Halgren, *J. Comput. Chem.*, 1996, **17**, 490.
32 G. A. Kaminsky, R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen, *J. Phys. Chem. B*, 2001, **105**, 6474.
33 M. Clark, R. D. Cramer III and N. Van Opdenbosh, *J. Comput. Chem.*, 1989, **10**, 982.
34 P. L. A. Popelier, L. Joubert and D. S. Kosov, *J. Phys. Chem. A*, 2001, **105**, 8254.
35 C. Haettig, *Chem. Phys. Lett.*, 1996, **260**, 341.
36 S. Y. Liem, P. L. A. Popelier and M. Leslie, *Int. J. Quantum Chem.*, 2004, **99**, 685.
37 A. K. Soper, *ISRN Phys. Chem.*, 2013, 279463.

