Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

Temperature Dependent Properties of the Aqueous Electron

J. Lan, V. V. Rybkin, A. Pasquarello*

Our methodology combines various advanced simulations, which we sequentially describe in the following sections.

S1. ELECTRONIC STRUCTURE

In the first step, we separately simulate neat water and $e^-_{\rm aq}$ using AIMD under various thermodynamic conditions. We use heavy water to match the conditions in the pulse radiolysis experiments of Bartels *et al.* [1]. We do not include nuclear quantum effects, because they have been shown not to affect the structure of $e^-_{\rm aq}$ in heavy water at room temperature [2] and because they are expected to be even less significant at higher temperatures [3]. The underlying electronic structure is described with the spin-polarized hybrid functional PBEh(0.40), in which the fraction of Fock exchange is set to 0.40. The functional is supplemented with van der Waals interactions through the nonlocal rVV10 functional [4], where the parameter b is set to 5.3 to ensure that the structural and dynamical properties of liquid water are properly reproduced [5]. All the periodic AIMD calculations are performed with the CP2K code [6] employing TZV2P basis sets [7] and Goedecker-Teter-Hutter pseudopotentials [8, 9]. The exchange integrals are calculated with the auxiliary density matrix method [10].

S2. MACHINE LEARNING

In the second step, we use the data obtained in the first step to train two MLPs, denoted MLP(water) for neat water and MLP($e^-_{\rm aq}$) for $e^-_{\rm aq}$. We adopt the neural network scheme introduced by Behler and Parrinello [11], which is implemented in the n2p2 code [12]. The MLP is further improved by running MLP-based molecular dynamics under different thermodynamic conditions. Hybrid-functional energies and forces are calculated for representative configurations selected on the basis of their atomic fingerprints [13] and used for extending the initial data set.

The training set is of great importance to construct an accurate machine learning potential (MLP). The initial data sets are generated by hybrid functional molecular dynamics (MD) employing a timestep of 0.5 fs. The initial conditions for the trajectories of $e^-_{\rm aq}$ are prepared by equilibrating liquid water in a NVT ensemble at various temperatures. To control the temperature, we use a thermostat ensuring canonical sampling through velocity rescaling (CSVR) [14]. The supercells contain 128 water molecules and their densities correspond to the experimental values.

Two MLPs are trained on initial *ab initio* MD simulations. We further improve the MLPs by running MLP-based MD in various thermodynamic conditions and selecting structures for neat water and $e^-_{\rm aq}$ via their atomic fingerprints (CurSel) [13]. For these selected configurations, we recalculate forces and energies at the hybrid functional level and use these results to extend the data set. We list the details of the data sets in Table S1 for neat water and in Table S2 for $e^-_{\rm aq}$.

Table S1: List of datasets used to train the machine learning potential for neat water. The size of the data sets are given in terms of the timestep of 0.5 fs.

Index	Temperature (K)	Ensemble	Size	ρ (g/cm^3)
1	255	NVT	1000	1.1
2	300	NPT	450	$\leq$1.1
3	305	NVT	1000	1.1
4	473	NPT	471	$\leq$1.1
5	573	NPT	478	$\leq$1.1
6	650	NVT	909	0.6
7	650	NVT	842	0.4
8	650	NVT	243	0.1
9	255-650	CurSel	300	0.1-1.1

Table S2: List of datasets used to train the machine learning potential for e^-_{aq}. The size of the data sets are given in terms of the timestep of 0.5 fs. The last column gives the fraction f of configurations in which the electron is localized.

Index	Temperature (K)	Ensemble	Size	ρ (g/cm^3)	f
1	300	NVT	1122	1.115	0%
2	373	NVT	1123	1.073	65%
3	423	NVT	1014	1.029	65%
4	473	NVT	1207	0.977	83%
5	598	NVT	1117	0.761	91%
6	623	NVT	1103	0.684	72%
7	648	NVT	863	0.505	88%
8	673	NVT	519	0.370	100%

From all the datasets, we extract a compact dataset consisting of 1753 forces and energies on the basis of their atomic fingerprints. This compact data set is used to train the MLP for the final production runs. As far as the multi-time-step scheme is concerned, it is useful to remark that the criteria for the accuracy of the MLP are less stringent than in a fully MLP-based simulation, because in the former case the MLP only operates as the fast component [15, 16]. The high-dimensional neural network potential for water proposed by Behler and Parrinello is flexible to describe a wide range of thermodynamic properties of water [11]. This method estimates the total energy E by a sum of atomic energies E_n based on a set of atom-centered many-body symmetry functions G_n. Atomic environments defined by cutoff radii of 12.0 bohr are used for hydrogen and oxygen atoms. We used 27 symmetry functions for hydrogen and 30 functions for oxygen, which is the same symmetry function set as used in Ref. [2]. The hydrogen and oxygen neural networks consist of two hidden layers with 25 nodes. The architecture of the neural network is shown in Fig. S2. All related data files including the MLPs, the data sets, the input files, and the molecular dynamics trajectories have been uploaded to the Materials Cloud repository. Detailed comparisons between energies obtained with the MLPs and with the hybrid functional are shown in Fig. S1. In the case of neat water, we obtain a root-mean-square error (RMSE) of 0.93 meV/H_2O for the energies and a RMSE of 124.44 meV/Å for the forces. In the case of e^-_{aq}, we find a RMSE of 2.38 meV/H_2O for the energies and a RMSE of 167.58 meV/Å for the forces.

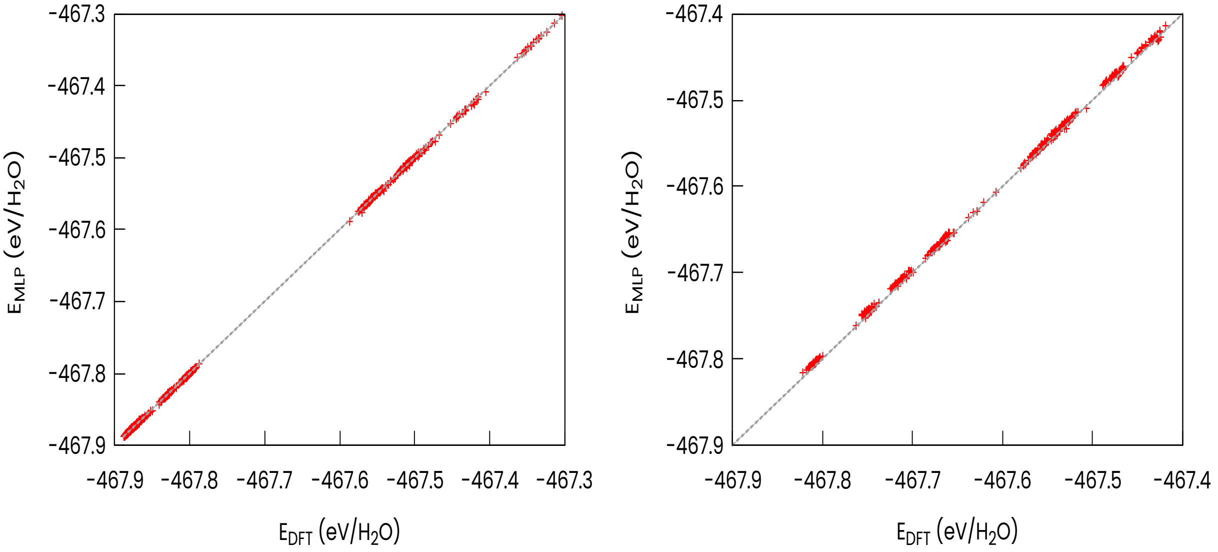

Figure S1: Comparisons between energies obtained with the machine learning potentials and with the hybrid functional for neat water (left) and for e^-_{aq} (right).

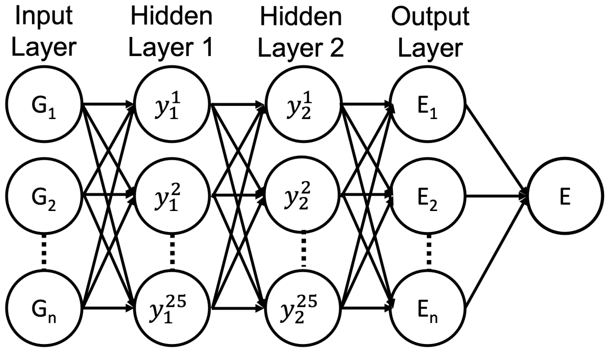

Figure S2: Architecture of neural network used consisting of two hidden layers with 25 nodes. The G_n in the input layer indicate the many-body symmetry functions [11]. The total energy E is obtained as a sum of atomic energies E_n.

S3. MOLECULAR DYNAMICS

In the third step, we simulate liquid water with the MLP(water) under various thermodynamic conditions [1]. We first focus on the temperature dependence of the density of neat water. The simulations are carried out in the NPT ensemble at a constant pressure of 250 bar and at various temperatures. For comparison, we also consider the SPC water model, which is generally used in one-electron pseudopotential schemes. All the simulations start from the experimental density in the given thermodynamic conditions. When the volume reaches a constant value after 10 to 20 ps, the production runs are started and are carried out for a duration of 20 ps.

All simulations of e^-_{aq} start from configurations of neat water as equilibrated within the canonical ensemble at experimental density and temperature (NVT), relying on the fact that MLP(water) gives a satisfying density for neat water. For e^-_{aq}, we applied a multiple time step (MTS) method [15, 16]. The full force based on the PBEh(0.40)-rVV10 functional is evaluated every 2.0 fs, corresponding to four time steps evolved with the reference force based on the MLP(e^-_{aq}). We perform two independent simulations lasting 10 ps for each thermodynamic condition. The MTS-AIMD simulations are performed with the i-PI code [16, 17] interfaced with CP2K for the electronic structure calculations and LAMMPS [18] for the MLP.

S4. TIME DEPENDENT DENSITY FUNCTIONAL THEORY

In the final step, we carry out TDDFT calculations for configurations of e^-_{aq} from the MTS-AIMD simulations using a QM/MM protocol as implemented in the ORCA code [19]. For each adsorption spectrum in the various thermodynamic conditions, we use 40 configurations homogeneously distributed in time between 2 and 10 ps after the addition of the electron. The QM region comprises 70 water molecules closest to the center of e^-_{aq}, corresponding to the cluster with the largest radius fully contained in the simulation cells. The MM region is described through a point charge model and constructed on the basis of 560 surrounding water molecules in the periodic supercells. The rest of the environment is described by the conductor-like polarizable continuum model [20] with the experimental dielectric constant of water at the given thermodynamic condition [21]. In the TDDFT calculations, we use the PBEh(0.40) functional with ma-def2-TZVP basis set [22] and consider the first five excited states. The QM/MM set-up is illustrated in Fig. S3.

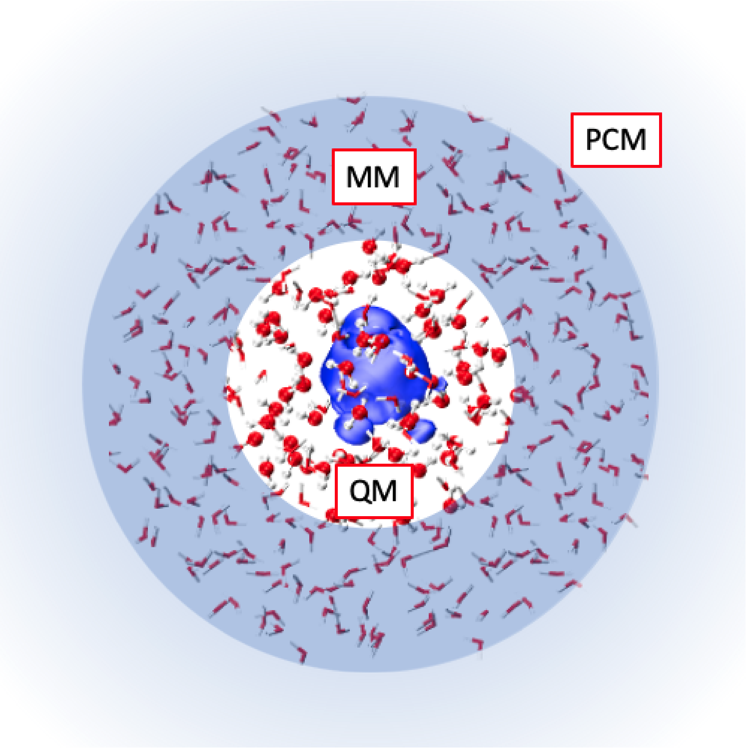

Figure S3: QM/MM set-up for e^-_{eq} embedded in a polarizable continuum model (PCM).

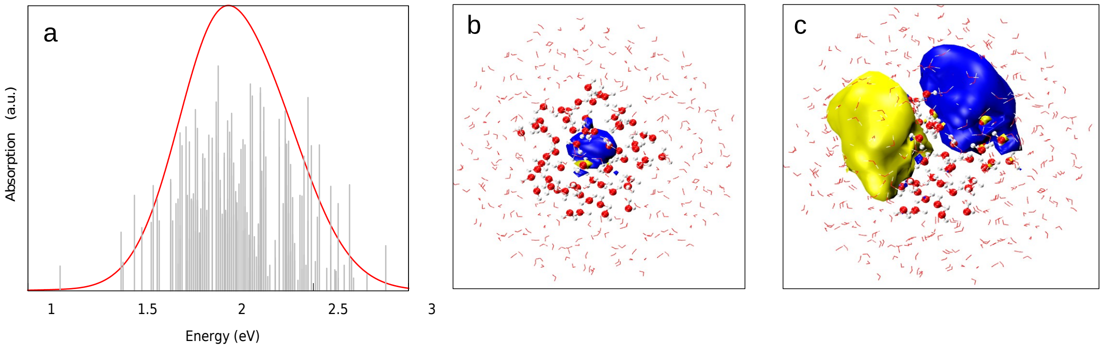

Figure S4: (a) Absorption spectra calculated from TDDFT at 300 K, where the maximum results from *s*-*p* transitions. The TDDFT calculations include five excited states for each considered configuration. (b) *s* state orbital of e^-_{eq}. (c) *p*-state orbital of e^-_{eq}.

S5. TEMPERATURE DEPENDENT PROPERTIES OF e^-_{eq}

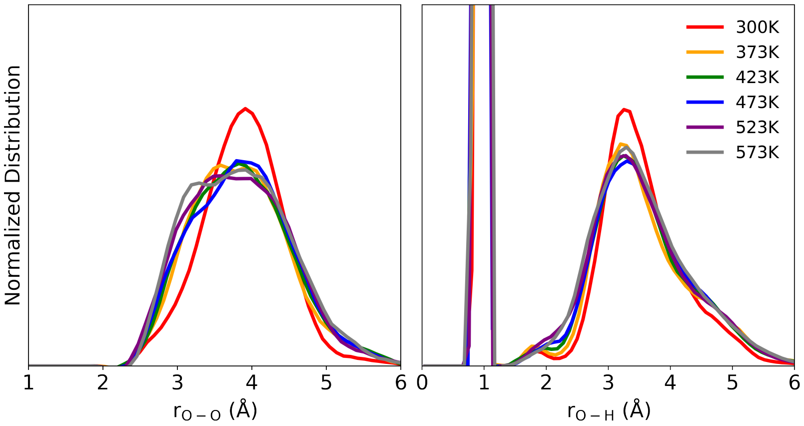

Figure S5: Normalized distributions of O-O (left panel) and O-H (right panel) correlations in the first shell of the aqueous electron at various temperatures.

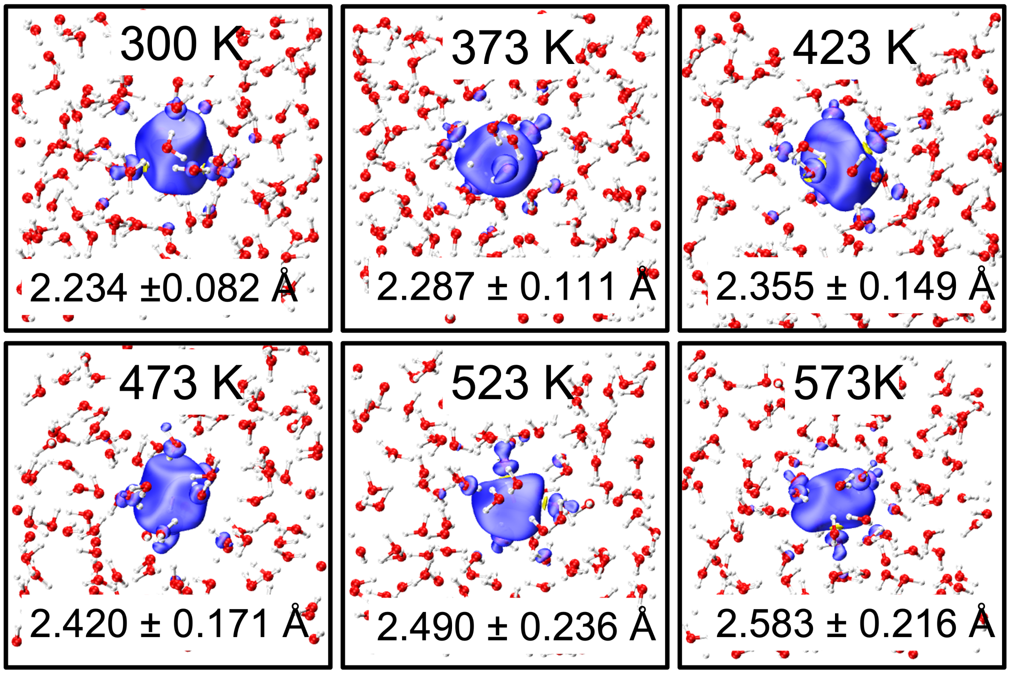

Figure S6: Spin densities of the aqueous electron from *ab-initio* molecular dynamics at various temperatures. The average gyration radii and their standard deviations are indicated.

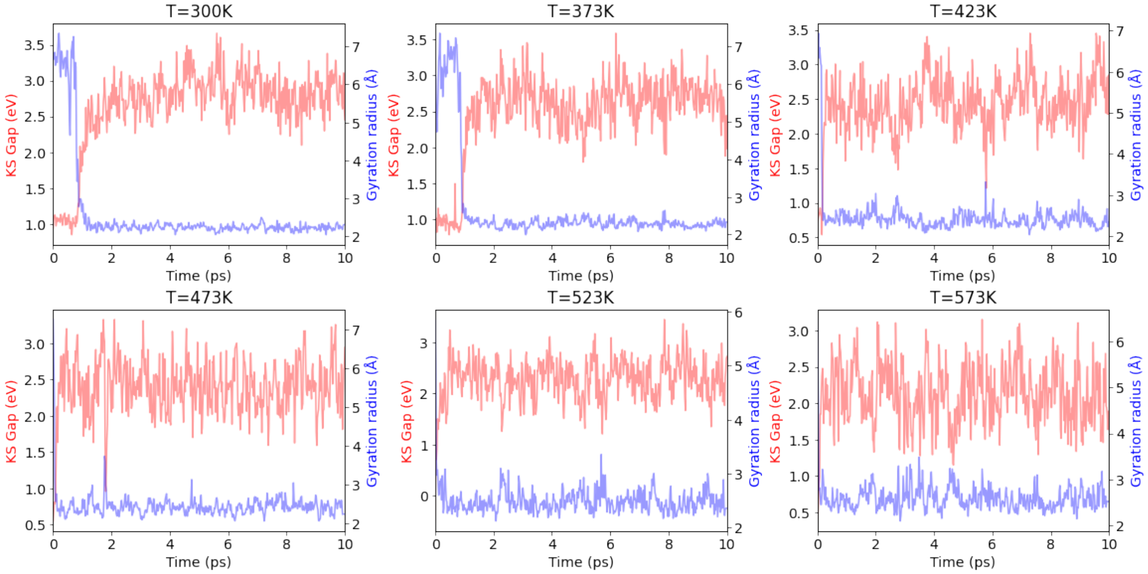

Figure S7: Kohn-Sham gap and gyration radius as a function of time at various temperatures.

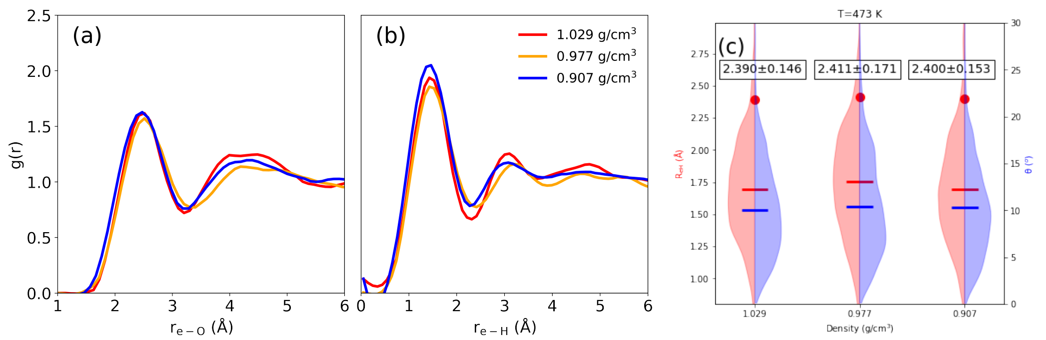

Figure S8: Radial distribution functions of (a) O and (b) H atoms with respect to the electron at 473 K and at various densities. (c) Distribution of R_{eH} (red) and angle between e-H and e-O (blue), where the hydrogen is coordinated to the electron. The means of the distributions are given by horizontal dashes. The average gyration radii of the electron are indicated by red dots and their values are illustrated in the boxes.

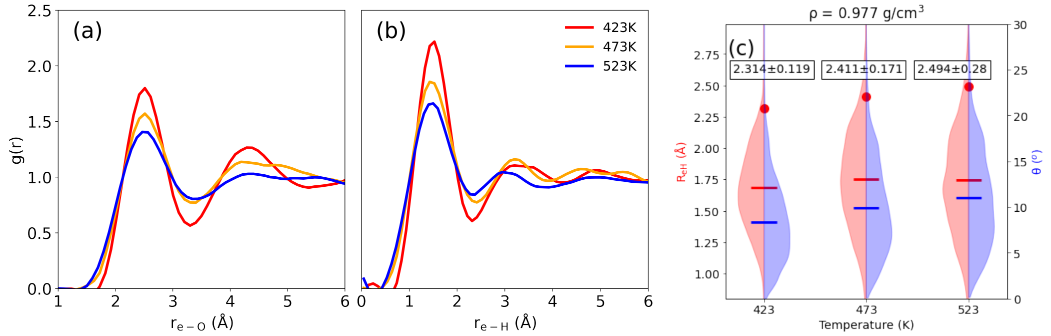

Figure S9: Radial distribution functions of (a) O and (b) H atoms with respect to the electron at $\rho = 0.977\text{g/cm}^3$ and at various temperatures. (c) Distribution of R_{eH} (red) and angle between e-H and e-O (blue), where the hydrogen is coordinated to the electron. The means of the distributions are given by horizontal dashes. The average gyration radii of the electron are indicated by red dots and their values are illustrated in the boxes.

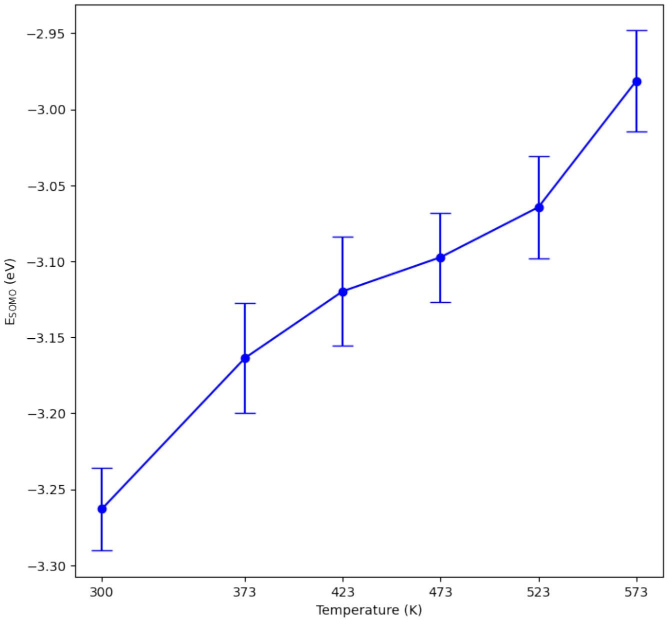

Figure S10: Single occupied molecular orbital level (or negative vertical binding energy) vs. the vacuum level as a function of temperature.

S6. GYRATION RADIUS

The e^-_{aq} is characterized by the spin-density distributions, the center of which is calculated from the second moment tensor. Since the position operator $\mathbf{r}$ is not defined under periodic boundary conditions, the gyration radius is not defined either. However, the simulation cell is large enough to achieve localized spin density distributions. The Gaussian cube files used to store the spin density distributions have been centered in the supercell to ensure that the spin density vanishes at the supercell boundaries. The centre of the spin density distribution $\rho^s(\mathbf{r})$ is then given as follows:

$$\mathbf{r}_c = \int \rho^s(\mathbf{r})\mathbf{r}d\mathbf{r}. \tag{S1}$$

Using the spin density distribution on the real-space grid, we obtain:

$$\mathbf{r}_c = \sum_{i=N}^{N} \rho^s(\mathbf{r})\mathbf{r}, \tag{S2}$$

The second moment tensor of this distribution reads

$$\mathbf{S} = \int (\mathbf{r} - \mathbf{r}_c)(\mathbf{r} - \mathbf{r}_c)\rho^s(\mathbf{r})d\mathbf{r}, \tag{S3}$$

and the gyration radius is calculated as

$$r_{\mathrm{gyr}} = \sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}, \tag{S4}$$

where λ_1, λ_2, and λ_3 are the eigenvalues of $\mathbf{S}$.

[1] D. M. Bartels, K. Takahashi, J. A. Cline, T. W. Marin, and C. D. Jonah, J. Phys. Chem. A **109**, 1299 (2005).
[2] J. Lan, V. Kapil, P. Gasparotto, M. Ceriotti, M. Iannuzzi, and V. V. Rybkin, Nat. Commun. **12**, 766 (2021).
[3] R. Hart, C. Benmore, J. Neuefeind, S. Kohara, B. Tomberli, and P. Egelstaff, Phys. Rev. Lett. **94**, 047801 (2005).
[4] R. Sabatini, T. Gorni, and S. de Gironcoli, Phys. Rev. B **87**, 041108 (2013).
[5] F. Ambrosio, G. Miceli, and A. Pasquarello, J. Phys. Chem. B **120**, 7456 (2016).
[6] T. D. Kühne, M. Iannuzzi, M. Del Ben, V. V. Rybkin, P. Seewald, F. Stein, T. Laino, R. Z. Khaliullin, O. Schütt, F. Schiffmann, *et al.*, J. Chem. Phys **152**, 194103 (2020).
[7] J. VandeVondele and J. Hutter, J. Chem. Phys **127**, 114105 (2007).
[8] S. Goedecker, M. Teter, and J. Hutter, Phys. Rev. B **54**, 1703 (1996).
[9] C. Hartwigsen, S. Gœdecker, and J. Hutter, Phys. Rev. B **58**, 3641 (1998).
[10] M. Guidon, J. Hutter, and J. VandeVondele, J. Chem. Theory Comput. **6**, 2348 (2010).
[11] J. Behler and M. Parrinello, Phys. Rev. Lett. **98**, 146401 (2007).
[12] A. Singraber, J. Behler, and C. Dellago, J. Chem. Theory Comput. **15**, 1827 (2019).
[13] G. Imbalzano, A. Anelli, D. Giofré, S. Klees, J. Behler, and M. Ceriotti, J. Chem. Phys **148**, 241730 (2018).
[14] G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys **126**, 014101 (2007).
[15] M. Tuckerman, B. J. Berne, and G. J. Martyna, J. Chem. Phys **97**, 1990 (1992).
[16] V. Kapil, J. VandeVondele, and M. Ceriotti, J. Chem. Phys **144**, 054111 (2016).
[17] V. Kapil, M. Rossi, O. Marsalek, R. Petraglia, Y. Litman, T. Spura, B. Cheng, A. Cuzzocrea, R. H. Meißner, D. M. Wilkins, *et al.*, Comput. Phys. Commun. **236**, 214 (2019).
[18] S. Plimpton, J. Comput. Phys. **117**, 1 (1995).
[19] F. Neese, Wiley Interdiscip. Rev. Comput. Mol. Sci. **2**, 73 (2012).
[20] V. Barone and M. Cossi, J. Phys. Chem. A **102**, 1995 (1998).
[21] D. Fernandez, A. Goodwin, E. W. Lemmon, J. Levelt Sengers, and R. Williams, J. Phys. Chem. Ref. Data **26**, 1125 (1997).
[22] J. Zheng, X. Xu, and D. G. Truhlar, Theoretical Chemistry Accounts **128**, 295 (2011).