PCCP

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Simulating the Solvation Structure of Low- and High-Spin $[Fe(bpy)_3]^{2+}$: Long-Range Dispersion and Many-Body Effects†

Habiburrahman Zulfikri,[a‡] Mátyas Pápai,[b,¶] and Asmus Ougaard Dohn*[a,c]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

When characterizing transition metal complexes and their functionalities, the importance of including the solvent as an active participant is becoming more and more apparent. Whereas many studies have evaluated long-range dispersion effects inside organic molecules and organometallics, less is known about their role in solvation. Here, we have analysed the components within solute-solvent and solvent-solvent interactions of one of the most studied iron-based photoswitch model systems, in two spin states. We find that long-range dispersion effects modulate the coordination significantly, and that this is accurately captured by density functional theory models including dispersion corrections. We furthermore correlate gas-phase relaxed complex-water clusters to thermally averaged molecular densities. This shows how the gas-phase interactions translate to solution structure, quantified through 3D molecular densities, angular distributions, and radial distribution functions. We show that finite-size simulation cells can cause the radial distribution functions to have artificially enlarged amplitudes. Finally, we quantify the effects of many-body interactions within the solvent shells, and find that almost a fifth of the total interaction energy of the solute-shell system in the high-spin state comes from many-body contributions, which cannot be captured by by pair-wise additive force field methods.

1 Introduction

Excited-state transition metal (TM) complexes have received great attention both for their functional capabilities (solar energy conversion, molecular data storage, etc.) and their properties optimal for time-resolved experimental investigations.[1,2] As in most cases the dynamics are probed in solution phase, the already complex excited-state processes, such as electronic transitions and nuclear relaxation, are further complicated by solvation. Time-resolved X-ray scattering is a powerful technique to follow excited-state structural variations, also in solution,[3–7] but the data analysis heavily relies on the simulation of solvation structure; theory thus has an essential role in extracting the structural information from the complex experimental data.

Molecular Dynamics (MD) simulations utilizing explicit solvent molecules are the general computational tools for the calculation of solvation structure. In MD, the nuclei are propagated classically according to Newton's second law. Based on the calculation of forces for the classical equations of motion, there are three levels of MD: i) so-called *ab initio* MD (AIMD), which treats all electrons in the system quantum mechanically (QM). ii) QM/MM MD which only uses a QM treatment for the most important part of the system, e.g. the electrons of the TM complex, while a molecular mechanics (MM) force field (FF), describes the solvent. Lastly, there is iii) fully MM or classical MD, where all forces on all nuclei are solely obtained from evaluations of the FF. AIMD methods include the use of Density Functional Theory (DFT) models to calculate electronic structure, although they are not strictly *ab initio*, and the nuclei are still treated classically, regardless of which potential is used. While AIMD offers the highest accuracy and flexibility for the price of being computationally demanding, FF-driven MD, on the contrary, is very fast but faces limitations connected to the application of a FF. Excited states are especially challenging, as common FFs are developed for ground states, with the state of the art being to take into account the difference in electronic state only in the electrostatic interactions utilizing partial atomic charges[8]Specialized force-fields can be parametrized

[a] *Science Institute and Faculty of Physical Sciences, University of Iceland, VR-III, Reykjavík 107, Iceland; E-mail: asod@hi.is*
[b] *Department of Chemistry, Technical University of Denmark, Kemitorvet 207, 2800 Kgs. Lyngby, Denmark*
[c] *Department of Physics, Technical University of Denmark, Fysikvej 307, 2800 Kgs. Lyngby, Denmark*
¶: *Wigner Research Centre for Physics, P.O. Box 49, H-1525 Budapest, Hungary.*
‡: *Present Address: Department of Chemical Engineering, Faculty of Engineering, Universitas Indonesia, Depok 16424, Indonesia*

† Electronic Supplementary Information (ESI) available: Model system basis set benchmarks, $[Fe(bpy)_3]^{2+}\cdots nH_2O$ (n = 3,4) Interaction Energies, CP2K BSSE Analysis, Water-shell RMSD Comparisons, Further Molecular Mechanics Calculations, Atomic Coordinates. See DOI: 10.1039/cXCP00000x/

for excited states[9,10], but the task quickly becomes prohibitively labour-intensive if many complexes are to be parametrized. In such cases, QM/MM MD often represents a viable tradeoff between accuracy and computational feasibility, but the effect of the approximations within the QM/MM coupling have to be examined in order to assess the accuracy of the simulated solvation structures, especially if the solute and all the solvent is in the QM and MM region, respectively[11].

$[Fe(bpy)_3]^{2+}$ (bpy = 2,2'-bipyridine) is an often studied TM-based prototype for excited-state dynamics: It is known from several investigations utilizing various experimental techniques that this complex undergoes light-induced low-spin (LS, singlet) $\rightarrow$ high-spin (HS, quintet) transition in < 200 fs.[12–16] The LS $\rightarrow$ HS transition is accompanied by$A \sim 0.2$ Å elongation of the Fe-N bonds, as well as changes in the solvation structure. Interestingly, AIMD simulations utilizing density functional theory (DFT) led to contradictory results: in the HS state, two water molecules were expelled from the first solvation shell when utilizing the BLYP exchange-correlation (XC) functional,[17] while two molecules enter the first solvation shell, when using the same functional but including D3 dispersion correction (BLYP-D3[18,19]).[20] These results highlight the large impact of including long-range dispersion. Several studies have shown how including long-range dispersion effects improves the accuracy of the water structure in DFT-based AIMD simulations[21–23], and similarly, many studies have benchmarked the performance of dispersion corrections on organic and organometallic molecules[24–27]. However, to our knowledge, fewer studies concern themselves with dispersion effects on solvation structure directly, and if they do, they are on smaller systems[28].

It was shown for $[Fe(bpy)_3]^{2+}$ that the dispersion-corrected AIMD solvation cages are more consistent with X-ray scattering data than those obtained by classical MD with partial charges on the complex that reproduce the molecular electrostatic potential[29]. Indeed, significant differences are observed between solvation structures simulated by AIMD and those relying on MM models, i.e., QM/MM and classical MD, in particular, the AIMD shells are more structured.[8,20,29] In general, there is a high demand for FF-based simulation techniques that can compete with the accuracy of AIMD at a lower computational cost. In order to reach this goal, it is crucial to gain a thorough understanding of what effects govern solute-solvent and solvent-solvent interactions. In this work, we present a computational study on solvated $[Fe(bpy)_3]^{2+}$ in its LS and HS states utilizing DFT calculations and classical MD simulations. Our aim is twofold: i) to evaluate the role of long-range dispersion and its correction in DFT for noncovalent solute-solvent (water) interactions, and ii) to analyse how the solvation structure of LS and HS $[Fe(bpy)_3]^{2+}$ governs the solute-solvent and solvent-solvent interactions, including through many-body effects. Our strategy for achieving these goals is to start from quantifying the simplest possible solvent-interactions, namely within bimolecular complex-water (hetero)dimer geometries. From there, we incrementally expand our focus, first to larger clusters, and finally to water-networks that encompass the entire complex - water *shells* - which we compare to the liquid-phase averaged solvation shell structure.

The paper is organized as follows: First, we identify the aforementioned central complex-water geometries. We use the dimers to benchmark a series of DFT and long-range dispersion treatments against a more accurate model for electronic correlation. We then include results from MM force fields into the benchmark and increase the size of the test systems. Finally, we correlate the water-shell encompassing the complex to the MD-sampled average solvation structure, and analyse the effects of finite-size simulation cells. Finally, we analyse the magnitude of non-pairwise additive effects on the interaction energies within the solvation shell of the complex.

2 Computational Details

2.1 Construction of solute-solvent systems

Starting from a ground-state $[Fe(bpy)_3]^{2+}$ QM/MM MD trajectory from a previous study[8], we identified two stable $[Fe(bpy)_3]^{2+}\cdots H_2O$ dimers. The first dimer has a water molecule in the vicinity of the π electron cloud of two pyridine moieties of different ligands, and we label its location as pose A (see Fig. 1a). Another dimer was identified by positioning an H_2O such that it interacts via a weak hydrogen bond with a C$-$H bond of one ligand, as indicated by the geometric alignment from the ligand-hydrogen to the lone-pair on the water molecule. We denote this dimer geometry as pose B, as shown in Fig. 1b. Further searches for other dimers were unsuccessful, indicating that there are only two distinct, dimeric poses of an H_2O non-covalently, directly bonded to the $[Fe(bpy)_3]^{2+}$ complex.

Since the complex has the same three bipyridine ligands, one can place in total three and six water molecules in poses A and B, respectively. Since the shortest O$-$O distances between these resulting nine water molecules are around 5 Å, they are not linked with hydrogen bonds. To construct a fully connected inner solvation shell , we add eight additional H_2O that can be classified in two additional poses: (a) H_2O in pose C bridging an H_2O in pose A and an H_2O in pose B as shown in Fig. 1e, and (b) H_2O in pose D that connects three water molecules in pose B as depicted in Fig. 1f. In total, one can insert six and two water molecules in poses C and D, respectively, and together with three H_2O in pose A and six H_2O in pose B. The constructed inner solvation shell thus contains 17 solvent molecules, a configuration that has been also proposed recently by Stark et. al, for the Ruthenium-centered analogue.[30].

To establish other inner solvation shells, one can either eliminate or insert water molecule(s) from the $17H_2O$ configuration. While the elimination mechanism results in an unconnected noncovalent network, the insertion method is found to maintain the network and increase the number of hydrogen-bonds proportionally to the number of added H_2O. In other words, at least a total of 17 water molecules is required for a highly connected hydrogen bond network as shown in Figure 2a. As an illustration of the insertion method, we doubled a single water molecule in pose A of the $17H_2O$ configuration (yielding two water molecules in pose E) to obtain an $18H_2O$ configuration (see the resulting hydrogen bond network Figure 2b). Similarly, we can repeat the process

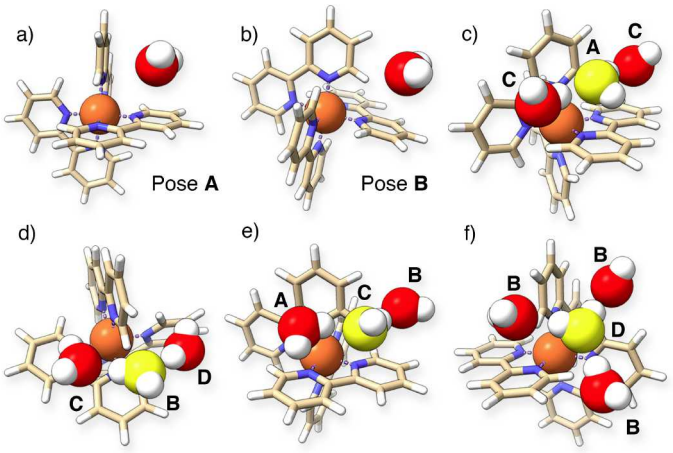

Fig. 1 The $[Fe(bpy)_3]^{2+}$ complex and noncovalently-bonded water molecule(s), the interaction energy curves (IECs) of which are studied in LS and HS states. For the multimer geometries, the oxygen water molecule we scan the Fe-O distance of, is colored yellow.(a,b) shows the two stable dimer geometries, involving 'pose A' and 'pose B' of a single water molecule. Poses C and D (bold labels) are only stable within the multimers shown in (c) through (f).

for other pose A-waters to create $19H_2O$ and $20H_2O$ configurations. The hydrogen bond networks of each of these shells are shown in Figure 2b. These configurations have three sequential hydrogen-bond networks connecting two water molecules in pose D. The four systems consisting of the complex and the inner solvation shell are employed here to understand how the change in the electronic structure of the complex affects the solute-solvent and solvent-solvent interactions at the equilibrium geometries. We note that, even though these four frames will obviously not represent all configurations in the liquid phase, our molecular dynamics simulations indicate that the hydrogen bond network is very well preserved in the liquid phase, and thus, the configurations we have chosen to investigate are representative of configurations that are visited very often during thermal sampling *(vide infra)*.

For the two dimer poses and the 17-20 shells, we carried out geometry optimizations in the gas phase using the B3LYP functional[31,32], D3(BJ) dispersion correction[19,33] and def2-TZVP basis set[34] . We performed frequency calculations in the ground state to investigate the nature of the obtained stationary points, and found no imaginary frequencies. Unless otherwise stated, the results of quantum chemical calculations were obtained using ORCA package version 4.2.1[35,36].

2.2 Benchmarking density functionals and dispersion treatments

Interaction energy scans of the dimer systems were performed as a function of the Fe−O distance, radially along the Fe-O direction spanned from the relaxed geometry. At least 32 points were evaluated to plot smooth interaction energy curves in the range of 3.6 to 15.0 Å. To allow identifying the depth of the curve, many points at an incremental of 0.1 Å were evaluated in the vicinity of the minimum. The generated geometries were not further optimized.

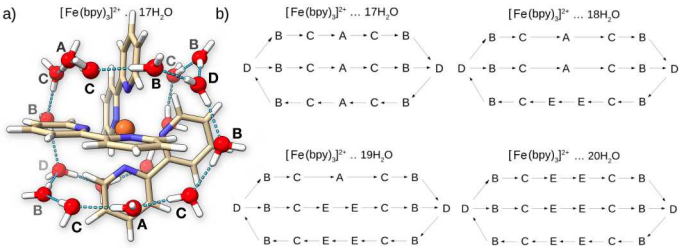

Fig. 2 Structural representation of the $[Fe(bpy)_3]^{2+}$ complex and its first solvation shell. a) Three dimensional structure of a complex drawn with the 'licorice' representation and the $17H_2O$ network drawn such as to identify the hydrogen-bonding network. The individual poses are labeled A, B, C, and D, respectively. b) Hydrogen bond networks of different first solvation shells containing 17/18/19/20 water molecules. The arrow links two water molecules in different poses from the donor to the acceptor of hydrogen bond.

Three density functionals from different rungs on the Jacob's ladder of DFT[37] were selected for the benchmark. They are the second-rung BLYP[38,39] GGA functional, which have been used in AIMD studies of the same complex before[17,20], the fourth-rung B3LYP hybrid functional, and the fifth-rung B2PLYP[18] double-hybrid functional. While far from an exhaustive benchmark of the density functional-jungle, this selection systematically increases the detail with which first exchange and then correlation terms are included in the functional describing the electronic density.

The default settings for the GGA and hybrid calculations as regards the handling of the two-electron Coulomb repulsion integrals were employed. The resolution of identity approximation for the Coulomb energy (RI-J)[40] was used for the BLYP calculations in combination with the def2/J auxiliary basis set[41]. Computation of the second-order perturbation term in B2PLYP calculations were sped up using the resolution of identity approximation[42] and the corresponding auxiliary basis set generated by the `AutoAux` procedure[43].

To overcome the basis set superposition error (BSSE), all the interaction energies presented in this paper are counterpoise-corrected according to the formulation of Boys and Bernardi[44]. The basis set employed for the scan is the minimally augmented Karlsruhe basis set: ma-def2-TZVP[45]. Our basis-set convergence study of the interaction energy of a bimolecular system, that consists of a water molecule noncovalently bonded to a model complex of Fe^{2+} and two pyridine ligands ($[Fe(py)_2]^{2+}$), indicates that augmenting the def2-TZVP basis set with only additional s and p functions yields accurate counterpoise-corrected DFT interaction energy within 0.01 eV (see section S1.1) For the $[Fe(bpy)_3]^{2+}nH_2O\cdots H_2O$ (n = 2, 3) multimer calculations, the system is fragmented into (i) the scanned water molecule ($\cdots H_2O$), and (ii) the rest of the system ($[Fe(bpy)_3]^{2+}nH_2O$).

We examine three approaches to account for the missing dispersion energy on the studied functionals. They are the atom-pairwise dispersion correction DFT-D3(BJ), the charge-dependent atom-pairwise dispersion correction DFT-D4[46,47] and the density-dependent dispersion correction scheme, DFT-NL[48,49]. The three-

body dispersion energy in DFT-D3(BJ) and DFT-D4 approaches was accounted for by the third-order Axilrod-Teller-Muto (ATM) term[50,51]. In DFT-D4, partial atomic charges were determined using the classical electronegativity equilibration procedure[52]. We note that, the rational BJ damping function was applied in all DFT-D4 calculations even though, as suggested by the author of DFT-D4 paper, the BJ abbreviation is not included in the name of the method.

DFT-NL computations can be performed either (a) self-consistently, i.e, orbitals and density are optimized in the presence the full DFT-NL exchange-correlation potential, or (b) non-self-consistently where the usual self-consistent calculation was first performed, and the resulting electron density is taken as input for the evaluation of the non-local energy term. We employed here the latter, less costly procedure as our tests on the model system indicated that both procedures obtain comparable interaction energies within chemical accuracy (see section S1.1, table S3). The short-range attenuation parameter b of the studied functionals was taken from previous works benchmarking the S22 and S66 datasets[49,53]. On the other hand, as the optimization of the long-range parameter C yield only slight improvement on the aforementioned datasets, its original value ($C = 0.0093$) is hence used here.

The obtained DFT interaction energy curves were benchmarked against the more accurate DLPNO-CCSD(T) method[54,55], which has been shown to accurately reproduce the reference values of closed-and open-shell systems in the GMTKN55 superset[56]. Differently from DFT calculations employing the ma-def2-TZVP basis set, DLPNO-CCSD(T) calculations here employed the def2-TZVP basis set without augmentation as the augmented basis function was found to not only increase the computational cost but also yield larger basis-set superposition error as shown in Table S4. Investigation of the interaction energy on the model system reveals that the reference interaction energy, obtained with the def2-QZVP basis and TightPNO settings, is within 0.01 eV of the calculation using a def2-TZVP basis set and NormalPNO settings (see Section S1.2) Therefore, the DLPNO-CCSD(T) calculations with the def2-TZVP basis set and the `NormalPNO` settings represent an effective choice at reasonable computational cost. Furthermore, to speed up calculations, RIJCOSX[57,58] approximation was used together with higher number of grids, i.e. `Grid6 Gridx6`.

2.3 Many-Body Expansion Simulations

The LS and HS geometry-optimized clusters consisting of the complexes surrounded by 17-20 water clusters were used to calculate all 3-body interaction energy terms that encompasses the cluster and 2 waters, using the same level of theory as the geometry optimizations. The combinatorial task of obtaining single point energies of the right cluster-combinations was automated with a python class available at `https://gitlab.com/asod/febpy3_solvation`, using ASEs ORCA calculator-interface. BSSEs were addressed by employing the full counterpoise-correction, meaning that electron-less basis functions were placed on *all* nuclear positions of atoms that do not explicitly enter into the single-point calculation at hand[59], such that all calculations are carried out with the basis functions of the full system. For instance, to get the interaction energy of the dimer comprised of molecule 1 and 2, $E(r_1, r_2)$ of the cluster comprised of the complex and 17 waters, ghost atoms were placed on $r_3, \ldots, r_{18}$.

2.4 Classical Molecular Dynamics Simulations

The $[Fe(bpy)_3]^{2+}$ in the LS state was parametrized within the General AMBER (GAFF) potential, using MCPB.py, as described elsewhere[60]. Partial charges for each electronic state were generated using the CHELPG model[61], shown previously to be one of the best-performing methods for these systems[8]. Changing the charges is the only modification we make to the FF when simulating the HS state in the MM model. The complex was then solvated in a box of 7329 water molecules modelled with the TIP4PEW potential[62], using the 'leap' program from the 'Ambertools' toolbox[63], which also added two Cl^- counterions, to charge-neutralize the system. The density and temperature were equilibrated using OpenMM for 500 ps in the NPT ensemble , using a Monte Carlo Thermostat to keep the pressure at 1 bar, and a Langevin propagator to keep the temperature at 300 K[64]. The solvation shell was then sampled for 50 ns in the NPT ensemble, saving frames to disk every 500th fs. All analysis was done after discarding the first nanosecond to equilibration. Previous studies using MM MD for solvation have circumvented FF-parametrization of complexes by using positional restraints on the complex-atoms to their DFT-optimized geometries[3,4,15,65,66], as explained elsewhere[8]. We simulated both free- and restrained complexes within our AMBER parametrization, as well as performing restrained simulations using the AMBER, CHARMM[67], and OPLS FFs. The FF/TIP4PEW LJ parameters were combined using the Lorentz-Berthelot combining rules[68,69]. All positional restraining potentials used a 500 kJ/(mol·Å^2) For the 3D molecular density plot, all frames were aligned to the first frame using the Kabsch algorithm[70,71] on the complex alone, to obtain the rotation matrices that align the complex with its reference. These matrices were then used to rotate all atomic positions in each frame, before 3D-histogramming all water-oxygen positions, using cubic 0.15 Å bins.

3 Results and discussion

Prior to analysing the potential energy curves of $[Fe(bpy)_3]^{2+}\cdots H_2O$, we first ensure that the starting optimized geometries of the studied bimolecular systems belong to a minimum. B3LYP-D3(BJ) frequency calculations using harmonic approximation found no negative eigenvalues of the Hessian matrix, showing that poses A and B are the two stable positions of a water molecule noncovalently interacting with the transition metal complex in its LS state. It is important to note that B3LYP optimizations of $[Fe(bpy)_3]^{2+}\cdots H_2O$ without any dispersion correction treatments only yields a stable pose B geometry, but finds no pose A minimum geometry. We examined this qualitative difference between the B3LYP and B3LYP-D3(BJ) solvation structure by performing further RI-MP2 geometry optimizations, which obtained both poses A and B of H_2O, in agreement with the B3LYP-D3(BJ) results (See the SI or Zenodo

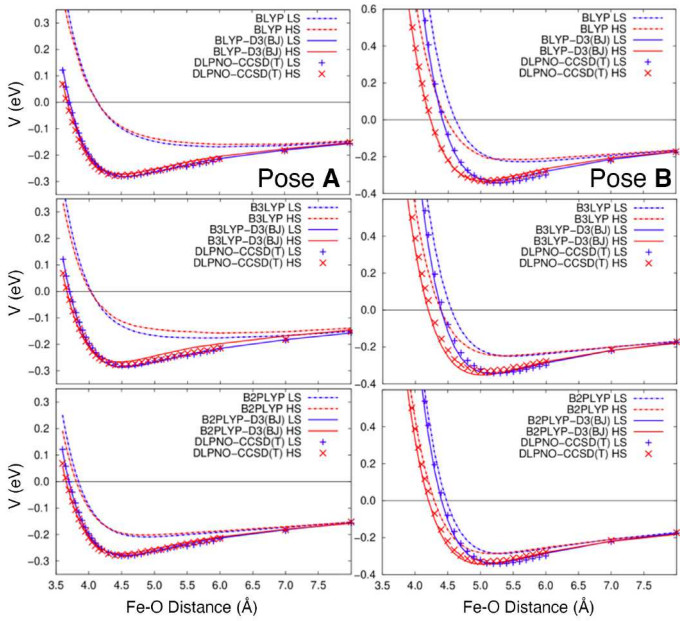

Fig. 3 Potential interaction energy curves of $[Fe(bpy)_3]^{2+}\cdots H_2O$ as a function of Fe−O distance for a water molecule in pose A (left) and pose B (right). Three DFT functionals of different rung, i.e., BLYP (top), B3LYP (middle) and B2PLYP (bottom), with and without the D3(BJ) dispersion correction are compared against the DLPNO-CCSD(T) reference.

repository[72] for structures). This demonstrates that a D3(BJ) correction to the B3LYP functional is needed to correctly obtain all stable geometries of the studied noncovalently bound dimer systems.

3.1 Potential interaction energy curves of the complex-water dimer

We use the standard definition of the total interaction energy V as the energy of the entire cluster comprised of n (hetero)monomers, minus the sum of the energies of the individual i molecules, each positioned at a set of atomic coordinates r_i:

$$V = E(r_1,\ldots,r_N) - \sum_{i=1}^{N} E(r_i). \quad (1)$$

Potential interaction energy curves of the $[Fe(bpy)_3]^{2+}\cdots H_2O$ bimolecular system in LS and HS states were computed individually for an H_2O in poses A and B. Figure 3 shows the interaction energy results. The first thing we observe when looking at the figure is that all curves produced without accounting for long-range dispersion effects altogether give the largest errors from the DLPNO-CCSD(T) results. Secondly, these errors are more pronounced in pose A than in pose B, indicating that dispersion interactions play a larger role for the pose A geometry. In pose A, the lone pairs on the water oxygen occupy the same regions of space as parts of the π-electron systems from the two pyridine moieties (see figure 1), indicating that a significant amount of the interaction must be governed by these interactions. On the other hand, there is only one π-electron system in the vicinity of water in pose B, but the

Table 1 Fe−O minimum energy (V_{min}) in eV of the CP-corrected interaction energy curves of $[Fe(bpy)_3]^{2+}\cdots H_2O$. The Fe-O distance giving the most binding interaction energy is denoted R_{min}, and its values are in Å.

Method	Pose A			
	LS		HS	
	R_{min}	V_{min}	R_{min}	V_{min}
BLYP	6.0	−0.169	6.0	−0.160
B3LYP	5.8	−0.176	6.0	−0.157
B2PLYP	4.9	−0.209	4.8	−0.202
BLYP-D3(BJ)	4.5	−0.281	4.5	−0.272
B3LYP-D3(BJ)	4.5	−0.286	4.5	−0.267
B2PLYP-D3(BJ)	4.5	−0.278	4.5	−0.273
BLYP-D4	4.5	−0.304	4.5	−0.292
B3LYP-D4	4.5	−0.303	4.5	−0.281
B2PLYP-D4	4.5	−0.286	4.5	−0.280
BLYP-NL	4.4	−0.310	4.4	−0.298
B3LYP-NL	4.5	−0.303	4.4	−0.282
B2PLYP-NL	4.5	−0.296	4.5	−0.289
AMBER	4.7	−0.307	4.7	−0.289
CHARMM	4.8	−0.293	4.8	−0.276
OPLS	4.8	−0.304	4.8	−0.286
DLPNO-CCSD(T)	4.6	−0.281	4.4	−0.278
	Pose B			
BLYP	5.6	−0.227	5.6	−0.215
B3LYP	5.5	−0.249	5.4	−0.248
B2PLYP	5.3	−0.286	5.2	−0.286
BLYP-D3(BJ)	5.2	−0.330	5.1	−0.336
B3LYP-D3(BJ)	5.2	−0.343	5.0	−0.352
B2PLYP-D3(BJ)	5.2	−0.337	5.0	−0.343
BLYP-D4	5.2	−0.342	5.1	−0.346
B3LYP-D4	5.2	−0.352	5.0	−0.360
B2PLYP-D4	5.2	−0.342	5.0	−0.348
BLYP-NL	5.1	−0.355	5.0	−0.354
B3LYP-NL	5.1	−0.358	5.0	−0.364
B2PLYP-NL	5.2	−0.351	5.0	−0.357
AMBER	5.4	−0.327	5.3	−0.318
CHARMM	5.3	−0.332	5.2	−0.319
OPLS	5.5	−0.326	5.3	−0.318
DLPNO-CCSD(T)	5.3	−0.342	5.0	−0.331

interaction between $[Fe(bpy)_3]^{2+}$ and H_2O is stabilized by weak C−H$\cdots$O hydrogen bonding, of which the attractive components are often mostly electrostatic in nature[73]. Consequently, although the Fe-O distance is larger in pose B than in pose A, the complex-water binding energy is more attractive in pose B.

The Fe−O distances that give the most binding interaction energies, R_{min}, are also displayed in table 1. The table shows that the reference DLPNO-CCSD(T) potential energy curve R_{min} distances contract from LS to HS state by 0.2 Å and 0.3 Å for poses A and B, and is accompanied by changes in the interaction energy that are below the chemical accuracy limit of 1 kcal/mol or roughly 0.04 eV. The LS and HS IECs are practically the same beyond R_{min}. On the other hand, the (HS−LS) interaction energy difference becomes more significant as the Fe−O distance decreases (see Figure 3). For instance, at very short Fe−O distances of 3.65 Å, the (HS−LS) interaction energy difference decreases by 0.043

eV and 1.556 eV for pose A and B, respectively.

Previous BLYP-D3 MD simulations have shown that, when sampled with this model, the shortest thermally accessible Fe-O distances in solution at room temperature is around 4 Å[20,29], a region where especially pose B experiences large decreases in interaction energies when going from LS to HS.

Without any long-range dispersion correction, the agreement of the DFT potential energy curve to that of the DLPNO-CCSD(T) reference improves as one climbs the Jacob's ladder. For instance, the LS state calculations in pose A reveal that from BLYP to B3LYP to B2PLYP, the R_{min} reduces from 6.0 Å to 5.8 Å to 4.9 Å (the reference value is 4.6 Å), which corresponds to the reduction of V_{min} from −0.169 eV to −0.176 eV to −0.209 eV (the reference value is −0.281 eV). The agreement worsens at shorter Fe−O distances of, again, 3.65 Å where BLYP, B3LYP, and B2PLYP calculations result in repulsive interaction energy of 0.392 eV, 0.327 eV, and 0.186 eV, respectively (the reference value is 0.058 eV). The hardest barrier is found in BLYP without long range dispersion corrections.

When any of the dispersion corrections investigated here is included in the DFT calculations, the agreement of the resulting IECs to that of the reference improves considerably, irrespective of which correction is employed. The effect completely overshadows the differences in energy from switching functionals. The most ubiquitous correction, D3 (with the BJ damping function), was not developed specifically with solvation in mind[19], and is not dependent of the electronic density, which means that two calculations of different electronic states on the same nuclear geometry would yield the same dispersion correction. Thus, benchmarks such as these are important to perform. In this specific case, the results show that the least expensive BLYP-D3(BJ) method should perform well for more computationally demanding simulations, such as sampling thermally averaged structures of the transition metal complex and its solvation structure at the DFT level.

Unless otherwise stated, all geometries used in this work were relaxed using B3LYP-D3(BJ). We note that including or excluding Hartree-Fock exchange has little to no effect on R_{min} and thus, we choose continue using the B3LYP-D3(BJ) method as the reference in the following cluster studies, in order to be consistent with the method used in obtaining the optimized geometries. We additionally note that, although the results of the three dispersion corrected calculations are comparable within chemical accuracy, one can deduce that the general trend of V_{min} is DFT-NL $\leq$ DFT-D4 $<$ DFT-D3(BJ) $<$ DFT.

Since classical molecular dynamics employing non-specialized force fields have also been used extensively to support experimental investigations of the solute-solvent interaction of solutions of transition metal complexes[3,7,8,65,74–77], we also examine the IECs obtained with CHELPG partial charges of the complex, using the TIP4PEW potential for water, and LJ parameters from three major general FFs: AMBER, CHARMM, and OPLS. The OPLS FF have been prevalent when using classical potentials to model the Time-Resolved X-ray Solution Scattering (TR-XSS) signal from solvation shell-responses to excitations[3,4,15,65,66], so we shall focus mainly on this FF. As shown in Figure 4 and table 1, the OPLS/TIP4PEW method slightly overestimates R_{min} by 0.2 Å with respect to the DLPNO-CCSD(T) value. Moreover, although the trend of the HS−LS energy difference is in accordance with the reference, the OPLS/TIP4PEW potential is very repulsive at short Fe−O distance. For instance, for the LS state, OPLS/TIP4PEW calculations show that the repulsive interaction already dominates at distances of 3.9 Å and 4.8 Å for a water molecule in pose A and B, respectively, distances for which the DLPNO-CCSD(T) interactions are still binding, i.e., −0.147 eV and −0.273 eV. In pose A, the differences in binding energies obtained with the three FFs are almost indiscernible. However, for pose B, the AMBER- and especially the CHARMM LJ parameters seem to shift the repulsive 'wall' of the potential to shorter Fe-O distances. This is because in these FFs, hydrogen atoms on carbons adjacent to nitrogens are of a different atom type (i.e. have a different set of parameters). Pose B involves exactly one of these hydrogens (See fig. 1, b), which have slightly reduced vdw radii in the parametrisations, explaining the differences from OPLS. All LJ parameters used can be found in S5.6 of the S.I.

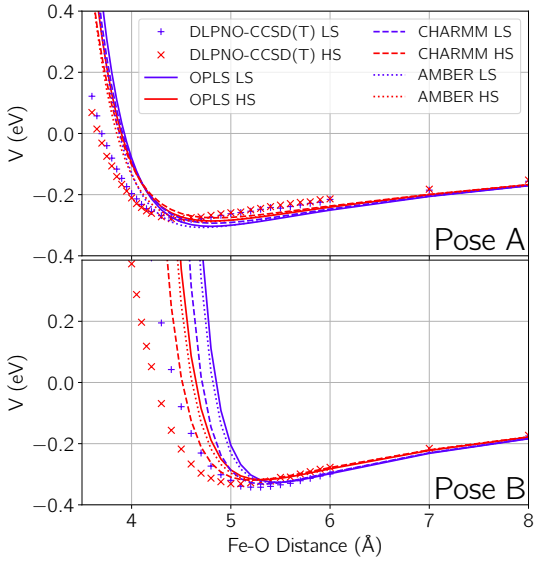

Fig. 4 Interaction energy curves of $[Fe(bpy)_3]^{2+}\cdots H_2O$ as a function of Fe−O distance for a water molecule in pose A (top) and pose B (bottom) computed at the DLPNO-CCSD(T) level, and compared to MM results calculated combining the LJ parameters from OPLS, CHARMM, and GAFF, respectively, with TIP4PEW. The MM partial charges for each state were calculated using the CHELPG method[61], which has been shown to be most accurate for these types of complexes previously[8].

3.2 Interaction energy curves of $[Fe(bpy)_3]^{2+}nH_2O\cdots H_2O$ (n = 2, 3)

The previous discussion deals with the sole interaction between $[Fe(bpy)_3]^{2+}$ and a single H_2O molecule, whereas in the solution phase, there will be a host of noncovalent interactions among the complex and solvent molecules, as well as between the solvent molecules themselves. To inspect the extent of the effect of dispersion correction to DFT calculations on systems containing multiple noncovalent interactions, we scan the electronic energy

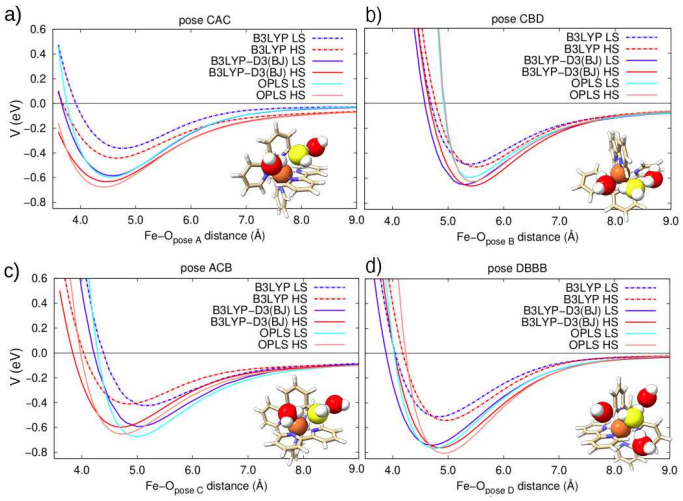

Fig. 5 B3LYP, B3LYP-D3(BJ) and OPLS interaction energy curves of $[Fe(bpy)_3]^{2+}nH_2O\cdots H_2O$ (n = 2, 3) in LS and HS states as a function of a single Fe−O distance. The translated molecule is shown in yellow on the insets. See the S.I. section S5.1 for curves calculated using AMBER and CHARMM. The results are nearly identical to the OPLS results.

of systems that consist of $[Fe(bpy)_3]^{2+}$ noncovalently interacting with three or four hydrogen-bonded water molecules, by scanning the Fe−O distance of a single water molecule for each cluster, and considering the remaining water molecules as a single fragment in the evaluation of the interaction energy. As shown in Figure 5 and SI Table S6, we notice the similarity of the results to those of the $[Fe(bpy)_3]^{2+}\cdots H_2O$ systems discussed earlier. First, the results of DFT calculations using different functionals and any of the tested dispersion corrections are very close to each other, as was also the case for the dimers. Secondly, the application of any dispersion treatment to a particular functional once again improves the agreement to the reference B3LYP-D3(BJ) values. In general, the minimum energy, V_{min}, of the DFT-D3(BJ) and DFT-D4 are compatible, and that of DFT-NL is slightly more negative. Finally, in DFT calculations without a dispersion correction, V_{min} improves as one ascends the Jacob's ladder. In contrast to the $[Fe(bpy)_3]^{2+}\cdots H_2O$ IECs, R_{min} does not vary much. This is because the hydrogen bonds among the water molecules help maintain the location of the minimum, even for the dispersion-uncorrected DFT methods. The R_{min}-values from the OPLS method are also very close to those of the B3LYP-D3(BJ) reference. Importantly, the OPLS IECs are closer to the B3LYP-D3(BJ) IECs than to the dispersion-uncorrected B3LYP curves. Additionally, We note that OPLS interaction energies are more repulsive than those of B3LYP-D3(BJ) at shorter Fe−O distances, inheriting the behavior of IECs of $[Fe(bpy)_3]^{2+}\cdots H_2O$ (Figure 4). Finally, the differences we observed in the pose B results across the 3 used FFs for the dimer system are diminished for the multimers (see the S.I. S5.1), indicating that the subtle changes in the LJ potentials from different hydrogens are overshadowed when more interactions (Coulomb and LJ) are added to the system via the additional waters.

3.3 Relationship between the vacuum-relaxed clusters and liquid structure

We now return to the 17-20 water shells described previously. To evaluate how the vacuum geometries probe interactions that translate to the liquid phase, we have performed molecular dynamics simulations of the complex in an aqueous environment, and obtained the molecular density distribution of water-oxygens as described in the computational details. As also previously mentioned, to be consistent with previous work, we continue this section mainly focusing on results from using the OPLS force field to sample solvation structure around the complex restrained to its B3LYP-D3(BJ) geometry. See the S.I. section S5 for further comparisons between FFs and restrained/free simulations. In figure 6, we compare the average, 3-dimensional solvent shell structure sampled using MM MD with the B3LYP(D3)/def2-TZVP gas phase-relaxed 17-water shell geometry. The top images show the

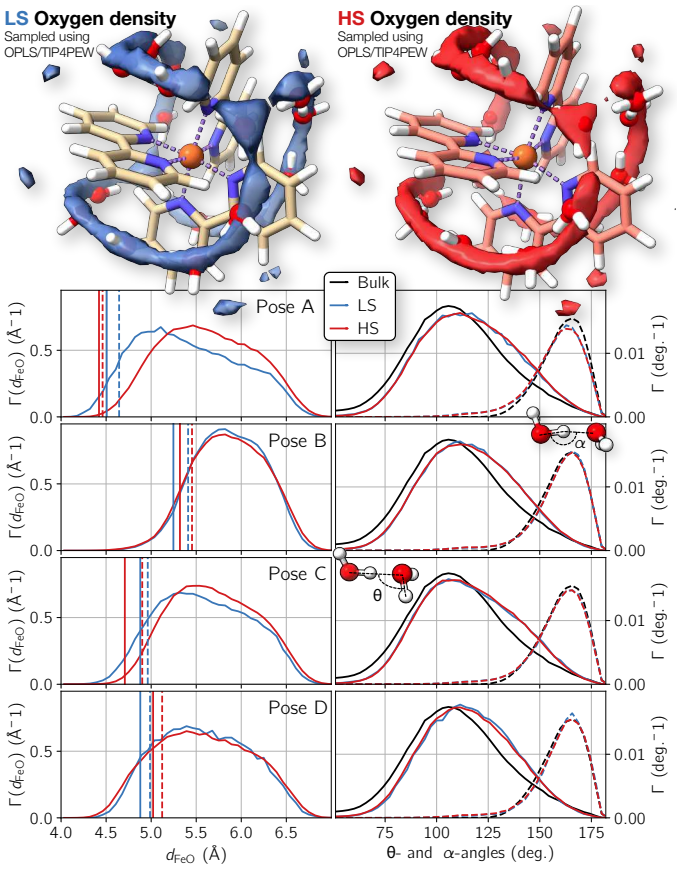

Fig. 6 Top: The molecular density distribution of water-oxygens sampled from MM MD simulations using OPLS/TIP4PEW parameters and partial charges on the complex generated from its LS and HS electronic states. The densities are plotted together with the B3LYP-D3(BJ)-optimized geometries of the 17-water shells. Bottom left: Histograms of the Fe-O distances of the 4 poses in the inner shell, from the LS run (blue) as well as the HS run (red). The fully drawn lines represent the vacuum-relaxed B3LYP-D3(BJ) distances, whereas the dashed lines represent the vacuum-optimized OPLS/TIP4PEW distances. Bottom right: Comparison of the water-water acceptor angles (θ, lines) and donor angles (α, dashed lines) in the two solvation shells to their bulk-water values. The donor angles have been scaled with 0.4 to fit in the same axis as the θ-angles.

molecular density distribution of the water-oxygens, plotted together with the B3LYP-D3(BJ)-optimized solvent shells in their LS (left) and HS (right) states, respectively. The triangular shape on the top left side coincides nicely with the pose D water, and in general, there is good overlap between the vacuum optimized water positions and the thermally sampled density. The bottom left plots in the figure relates the Fe-O distances from the various poses in the MM thermal sampling to the same distances obtained from the B3LYP-D3(BJ) and OPLS/TIP4PEW geometry optimizations (fully drawn lines and dash-dotted lines, respectively). The water-poses within the MM-shells of each frame of the trajectory were determined by aligning each MM frame to the B3LYP-D3(BJ) geometry, and finding the map that rearranges the indexes of the innermost 17 MM oxygens in such a way that the overall distance from the oxygens of the B3LYP-D3(BJ) optimization is minimized, using the Linear Sum Assignment (LSA) method[78]. Based on the Hydrogen bonding criterion from Wernet *et al.*[79], we do not label any MM water that has moved more than 3.4 Å during the LSA process. The plots on the right side of the figure show the distributions of two angles that describe the nature of the hydrogen-bonded network in water, and compares these angular distributions of the shell to those of bulk water. The donor-angle, α, determines how directly the hydrogen-donating water is pointing at the hydrogen-accepting water. The acceptor angle, θ, shows how much the hydrogen-accepting water is skewed with respect to the oxygen-oxygen axis of the hydrogen bond. We observe no changes in the donor angles in the shells around the complex sampled with the LS and HS partial charges, respectively. With respect to the bulk water, only the pose A waters seem to exhibit a slightly more flat donor angle, consistent with the strain from their position in the network depicted in figure 2. Again, for the acceptor angle, there is no real differences between the two electronic states, but all poses in the shell show varying degrees of a shift towards larger angles, on average. In the gas-phase, the average OPLS/TIP4PEW Fe-O distances are 0.11 Å, 0.15 Å, 0.07 Å, and 0.08 Å longer than for the DFT model, for poses A-D respectively. Thermally averaged at 300 K, all average Fe-O distances increase by $\sim$ 0.5 Å - 0.8 Å, consistent with the anharmonic shapes of the interaction curves in figures 3 and 5 . When thermally sampled, Pose A and C increase their average Fe-O distance around the complex in its HS state, whereas both the dashed and fully drawn vertical lines show that both the vacuum-optimized OPLS/TIP4EW and B3LYP-D3(BJ) geometries do the opposite, although to a lesser extent. Both poses are found nearest the pyridine rings, with no direct hydrogen bonding connection to the complex. When the complex expands during HS state optimization, the pose C and A waters can intercalate more tightly as the net positive bipyridine ligands attract the partially negative oxygens. However, when thermal vibrations of the complex will counteract this intercalation, even when the complex is restrained using Hookean potentials. We generally found only small effects on the solvation shell from restraining the complexes. The main effects we observe for restraining the complex is a $\sim$ 0.2 Å contraction in the average Fe-O distance of pose A waters, and a $\sim$ 0.1 Å contraction in the average Fe-O distance of pose C waters, compared to the non-restrained simulation. These differences are most likely caused by the added rigidity of the restrained ligands allowing for a slightly tighter water intercalation, which would affect these poses the most (see S.I. section S5.2 for details).

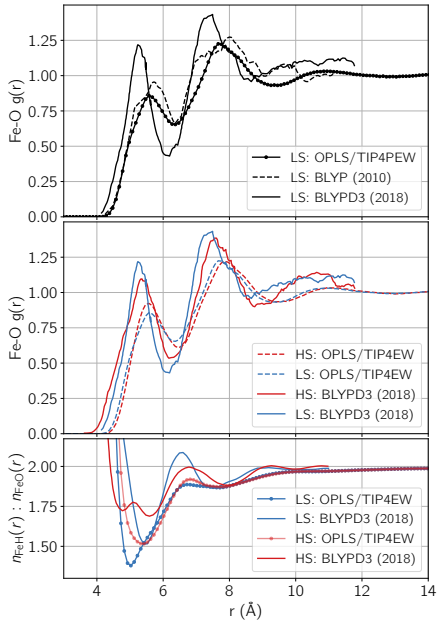

Fig. 7 Analysis of the LS Fe-water correlations in the liquid phase, expressed through RDFs (top and middle plot) and coordination number-ratios (bottom plot). Top: The dotted line shows the results from the MM OPLS/TIP4PEW simulation used in the connection to the vacuum-relaxed geometries from figure 6. The MM RDFs, here sampled in NPT, are in agreement with the NVT-sampled MM RDFs in a previous study[8]. The fully drawn and dashed lines have been digitized from the 2018 work by L. M. Lawson Daku[20] and the 2010 work of L. M. Lawson Daku & A. Hauser[17], respectively, simulating fully QM systems using BLYP, with and without the D3 long-range dispersion correction. Middle: Comparisons of the solvation shell changes from LS-HS excitation. The BLYP-D3 results show a decrease in amplitude of the first peak, and both the first and second solvent shell average distance increases. For the OPLS/TIP4PEW results, the changes are less pronounced. Bottom: Ratios between the Fe-H and Fe-O running coordination numbers, indicating the preferred orientation of the waters toward the Fe center. Since water has two hydrogens per oxygen, the water is randomly oriented when this ratio is 2. The further this ratio goes below 2, the more the oxygen-end of the water has a preferential orientation towards the Fe center, and vice versa. The BLYP-D3 results have been digitized from the same work as in the other subplots.[20]

We can also sample and compare Radial Distribution Functions (RDFs) of solute-solvent spatial correlations[80–82]. The RDF is the ratio between the (spherically symmetric) local number density of particles, sampled in spherical shells along r from the each i'th particle, to every other j'th particle, $\delta(|\mathbf{r}_i - \mathbf{r}_j| - r)$, and the isotropic density ρ:

$$g(r) = \frac{1}{\rho}\left\langle \sum_{i}^{N_i}\sum_{j\neq i}^{N_j} \delta(|\mathbf{r}_i - \mathbf{r}_j| - r)\right\rangle. \quad (2)$$

For MD simulations, the local density is simply numerically histogrammed using finite-sized r-bins, averaged over many frames, and $\rho = N/V$, is the number of particles over the simulation

volume. Thus, one can choose e.g. to sample the Fe-O RDF by letting i only run over the (single) Fe atom in the simulation, and j over all water-oxygens in the box. The black dotted line in figure 7 shows the Fe-O_{water} radial distribution function (RDF) sampled from the OPLS/TIP4PEW dynamics simulation, compared to the two AIMD simulations from literature [17,20]. A previous study performed a similar comparison, focusing on the effect on which partial charge-generation method is used [8]. As none of the charge-generation methods could fully reproduce the dispersion-corrected AIMD RDF, we have limited ourselves to only using the CHELPG-charge generation method here, focusing instead on comparing the differences in the liquid structure to the differences found in the interaction curves of the previous section, to analyse the seemingly nonelectrostatic origin of the differences in the liquid-phase descriptions. The MM RDF resembles the non-dispersion-corrected AIMD RDF most closely, with almost equivalent peak distances. The thermally averaged BLYP-D3 solvent shell is more structured, with shorter characteristic Fe-O distances, indicating that the long-range dispersion effect on solvation shown to be important in the previous sections are not captured completely by the MM model. Comparing to figure 6, we observe that all the identified characteristic poses lie within the first RDF peak, as expected. We note that the second RDF peak, signifying the second solvation shell centered around Fe is thus only indirectly affected by Fe-water interactions. The large LS-HS changes in solvation structure predicted by the AIMD simulations are less pronounced in the MM model.

Remembering the observations from the interaction energy curves of pose A and B in figure 3 the lower, shifted repulsive region in the HS state corresponds to the less abrupt RDF increase in the AIMD simulations [20,29]. However, all of the changes in the solvation structure of the complex cannot be explained from these two simple IECs: In pose A, we observed very small differences between the minimum well depth and position, as well as between the two repulsive walls, going from LS to HS. For pose B, the LS minimum was found at slightly larger distances in the LS state, yet the AIMD LS RDFs is shifted to a *shorter* distance. For the OPLS/TIP4PEW model, the interaction energy curves of which were presented in figure 4, is consistent between interaction energy curves and RDFs. For the tri- and tetramer geometries shown in figure 5, we observed no concerted elongation or shortening of minimum distances across all geometries, but an overall net increase in binding in the HS state, also consistent with the slightly increased amplitude of the first Fe-O RDF peak.

The bottom plot in figure 7 shows the ratio of running coordination numbers $n_{\mathrm{Fe-H}} : n_{\mathrm{Fe-O}}$. In the LSstate, the overall trends from the BLYP-D3 simulation are reproduced by the MM simulation, but the MM average orientation has a stronger preference of oxygen towards the Fe center around the Fe-O distances that make up the first peak in the Fe-O RDF. At the end of the first solvation shell, after 6 Å, the BLYP-D3 curve has a peak which is not reproduced in the MM curve, and overall, as from the RDFs, the BLYP-D3 results are more structured than for the MM results. For the HS state, both types of simulations reduce the strength of the preferred orientation of oxygen towards the Fe center in the first solvation shell, but the reduction is much more drastic for BLYP-D3, which also shows a double-minimum not reproduced by the MM model. The chosen MM model [8] does not take into account that the metal-ligand bonds are weakened in the HS [20], which would allow for a higher degree of re-orientation within the intercalated water. One way of testing this hypothesis in future studies would involve using specialized force fields for excited states of these types of molecules [9,10].

Apart from the obvious differences in level of sophistication between electronic structure models and classical potentials, other simulation parameters can play a role on the final solvation structure. The AIMD simulations include explicit treatment of the electronic density of the entire system in their model, which of course comes at much higher computational cost. This limits the size of the simulation boxes that can reasonably be computed. As previously mentioned, the isotropic density, ρ, used to produce the RDF in equation 2 for periodic- but finite systems (i.e. sampled in NVE, NVT, or NPT) is calculated as the number of particles over the total simulation box volume, V. However, this calculation of the isotropic density ignores the fact that some volume of the simulation box will never be accessible to the solvent molecules, since it is taken up by the solute. Thus the 'effective' volume available to the solvent is less than the total volume used in the normalization to the isotropic density. This affects the normalization of the RDF, and its long-r limit [82–85]. The effect is increased as simulation box size decreases, as the 'excluded' volume stays the same (or is reduced a lot less), while the total box volume decreases. Figure 8 compares RDFs sampled from OPLS/TIP4PEW MD simulations with box size and number of water molecules identical with the literature AIMD simulations [20], to the simulation using the larger box, afforded by the much simpler MM potential form used in this work. We notice how the the small-system RDFs (magenta lines) are shifted towards higher amplitudes, and do not converge to the theoretical limit of 1, when the local- and bulk density becomes equal. It is evident from the big-system C-O RDF (bottom plot, black line) that the majority of the correlations have died out already at 14 Å. Thus, the absence of convergence to 1 for the small-system RDFs is thus not due to the solvent shell-structure extending far beyond the simulation box size of the small system, but rather, due to the issue of normalizing to the density using the full box volume instead of the effective volume, as discussed above. We note that finding the correct effective volume for finite systems is not trivial [85], and that if the effective volume stays the same for the molecules in the two electronic states. This implies that, for instance, in x-ray diffuse scattering studies [29] the volume-based error should cancel out in the calculated x-ray difference-scattering signal from these RDFs [86].

In conclusion, figure 8 shows that the effective-volume effect explains long distance-discrepancies, but cannot explain all of the discrepancies at shorter values of r. Before continuing to investigate the interactions within the solvent shell themselves, we also show the effects on the solvent shell of the counterions in the system, by repeating the small-system simulation after removing the counterions (dashed magenta lines in figure 8). Since the small system size gives a high effective concentration (roughly

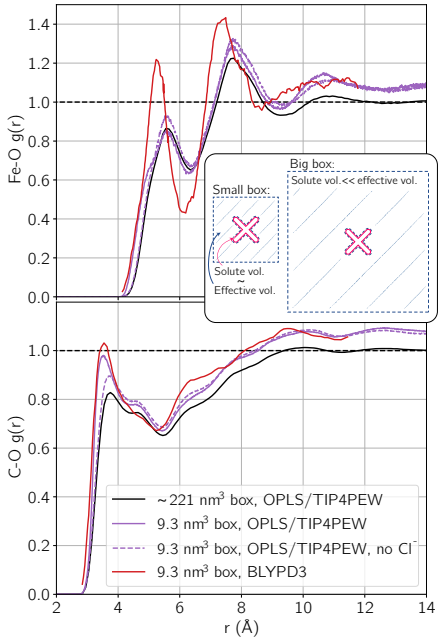

Fig. 8 Investigating the effects of box-sizes on the solvent shell RDFs around the complex in the LS electronic state. Top: Fe-O correlations. Bottom: C-O. The dashed black line shows the long-r limit for infinite systems. The red, fully drawn lines have been digitized from L. M. Lawson Daku[20] and Khakhulin *et al*[29], for Fe-O and C-O, respectively, although the underlying dataset was originally created for work in the first reference. The magenta lines show RDFs calculated from OPLS/TIP4P simulations using the same number of molecules and cubic 21 Å-length box size as in the AIMD simulation. The dashed lines show RDFs from an OPLS/TIP4PEW without the Cl^- counterions.

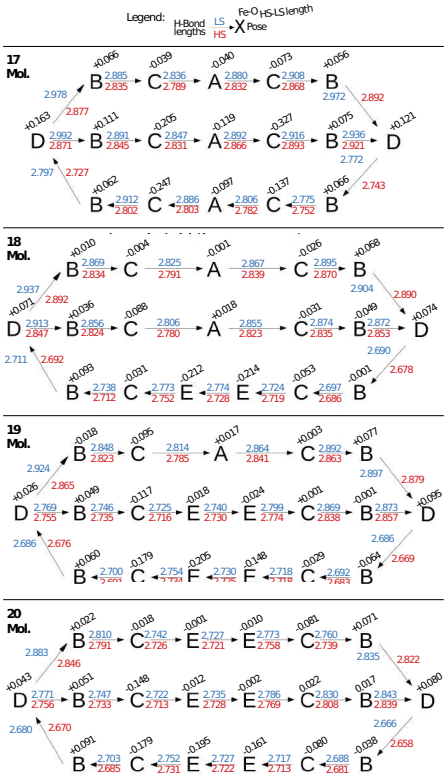

Fig. 9 Overview of the structural changes in the hydrogen-bonded water networks of the first solvation shell consisting of $n = 18 \ldots 20$ water molecules in total. All distances are in Å. The blue and red numbers are O-O distances across the hydrogen bonds for the LS and HS state, respectively. The numbers above each water show the contraction/expansion of the Fe-O distances, $r_{\mathrm{FeO}}(\mathrm{HS}) - r_{\mathrm{FeO}}(\mathrm{LS})$.

180 mM, or $\sim$ 18 times higher than what is used experimentally[29]), it forces the counterions closer to the complex. Thus, they may affect the structure of its solvent shell. Comparing the solid and dashed magenta lines, we observe that simulations including the counterion produces a C-O RDF with an considerably larger first peak, which has also shifted towards shorter C-O distances, compared to the simulation without the counterions. This indicates that the presence of the counterions can indeed affect the shell structure. In this case, the presence of counterions in the vicinity of the complex seem to increase the C-O coordination, because the waters are actually coordinated to Cl^-, which are now only kept at a fixed distance of roughly 7.5 Å from the Fe atom, due to the smaller box size.

3.4 Structure and Interactions within the $[Fe(bpy)_3]^{2+} \cdots nH_2O$ (n = 17...20)-clusters

Since the LS to HS Fe-N elongation drive an overall small increase in the size of the complex, the geometry of the first solvation shell will find a new minimum. Figure 9 shows the changes in O-O distances along the hydrogen bond network of the shells going from LS to HS, as well as the changes Fe-O distance. Looking at the figure, we first see no simple, general expansion or contraction of either hydrogen bonds or Fe-O distances. However, upon further analysis, some general trends do reveal themselves: Focusing first on the 17-shell, water molecules in poses B, which interact with the ligands via weakly attractive C–H$\cdots$O hydrogen bonds all elongate, and become longer by 0.073 Å on average. Since three H_2O in poses B are non-covalently connected by an H_2O in pose D, the elongation is propagated such that the water molecules in pose D also move further away by 0.142 Å on average. In contrast, water molecules in poses A and C move closer to the metal center by 0.086 Å and 0.171 Å, respectively. The pose A water molecules are oriented with the oxygen facing inwards toward the center, in the same regions of space as the $3d_{xy}$, $3d_{yz}$ and $3d_{xz}$ orbitals centered around the iron atom. As the occupation of these t_{2g} orbitals decreases from LS to HS, the repulsion between the partially negative water-oxygen and these orbitals are lessened, which explains that the water molecules in pose A will move closer to the metal-ion center upon relaxation on the HS surface. The overall change in Fe–O distances amounts to an average decrease by 0.033 Å. All shells exhibit similar, small overall decreases in Fe-O distances upon excitation, with values of 0.030 Å, 0.030 Å, and 0.014 Å, for the 18-20 shell, respectively.

The new nuclear arrangements and electronic density of the complex also affect distances between the water molecules themselves. On average, the O–O distance decreases by 0.051 Å from the LS to the HS state, for the 17-shell. Calculating the RMSD between the relaxed geometries of the water molecules in the two states, the 17-shell value of 0.320 Å is reduced to

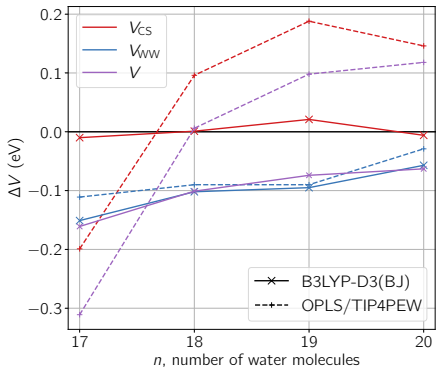

Fig. 10 HS−LS Differences complex-shell (V_{CS}, red), water-water (V_{WW}, blue), and total (V, purple) interaction energies in the $[Fe(bpy)_3]^{2+}\cdots nH_2O$ (n = 17, 18, 19, 20) at the B3LYP-D3(BJ)/def2-TZVP geometries. The fully drawn lines represent the B3LYP-D3(BJ) results, whereas the dashed lines show the OPLS/TIP4PEW results. The main discrepancies are found to arise from the complex-shell interaction energy.

0.200 Å for the 20-shell. Since the 17-shell contains the least amount of water molecules that still maintain the full hydrogen bond network around the complex, it makes sense that the individual waters need to rearrange themselves more to accommodate the changes in the solute structure, compared to shells with more water molecules. This tighter, slightly closer network of water molecules around the excited complex is consistent with the changes in the first peak of the OPLS/TIP4PEW RDFs shown in figure 7.

Looking at the overall interactions within these close solvation shells of the complex, we can divide up the interactions in two main categories: (1) complex-water, and (2) water-water. Up until now, we have been focusing on the first type of interactions. Next, in order to investigate the relationship between these two types of interactions, we consider each of the complex-water clusters as two supermolecular entities, i.e., the complex and its solvation shell. Thus, V, the total interaction energy introduced in equation 1, can be broken down into the interaction energy between the complex and its solvation shell, V_{CS}, and the interaction energy of the n water molecules excluding the complex, V_{WW}. The following definitions are used to define these interactions:

$$V_{CS} = E([Fe(bpy)_3]^{2+}nH_2O) - E([Fe(bpy)_3]^{2+}) - E(nH_2O) \quad (3)$$

$$V_{WW} = E(nH_2O) - \textstyle\sum_i^n E([H_2O]_i). \quad (4)$$

Note that with these definitions, $V_{CS} + V_{WW} = V$. The energies of each fragments have been calculated with the basis functions of the full system, to account for BSSE (see the S.I. section S6.1 for details). Note that due to the $E(nH_2O)$ term in V_{CS}, the complex-shell energy excludes many-body effects on the interaction energy among the waters *alone*, e.g. polarization of water by water itself, but includes them in between the complex and the solvent. Such many-body effects are not included in the MM model. Figure 10 shows that, for both models, V starts out more negative in the HS state than in the LS state, but as n increases, the total difference in interaction energies of the two states decreases. Interestingly, the difference in B3LYP-D3(BJ) complex−shell interaction energies, V_{CS} (red line), are indiscernible within chemical accuracy between the two states, meaning that if only the direct interactions between solute and solvent existed, there should be no change in the solvation shell structure. The water−water interaction energies, V_{WW}, are more stable in the HS states, however, although this difference is reduced with increasing numbers of water in the shell. This reduction is consistent with the aforementioned finding that the waters would need to rearrange themselves less for larger shells. Altogether, this shows that the cause for the changes in the shell geometries do not originate from changes in the direct interactions between solute and solvent, but rather in changes in interactions between the water molecules themselves, induced by the geometric and electronic changes within the complex. This, again, increases the challenge for purely pairwise additive models such as OPLS/TIP4PEW, which does not include changes in water-water interactions due to e.g. polarization induced by the electronic density of the complex in its potential. Looking at V_{WW}, and recalling that it can only include polarization effects between the waters themselves, we see that the TIP4PEW model, not including any non-pairwise additive contributions at all, still manage to reproduce this term with a standard deviation from B3LYP-D3(BJ) of only 13.6 meV. However, the complex-shell interactions of B3LYP-D3(BJ) are less accurately captioned by the OPLS/TIP4PEW force field, even though we have seen in the previous sections that the MM model performs relatively well for the smaller dimer-tetramer systems, where many-body effects matter less. Here, the model predicts a qualitatively different response to electronic excitation for 18+ water shells, becoming *less* binding, whereas the B3LYP-D3(BJ) model becomes *more* binding. It would be reasonable to expect that the complex, being positively charged, would be the main cause for polarization between the waters. We will analyse this effect in more detail in the next section.

3.5 Many-body expansion of interaction energy of $[Fe(bpy)_3]^{2+}\cdots 17H_2O$

The many-body expansion of the total interaction energy from equation 1 of a system comprising N molecules can be formulated as a sum of n-body interaction terms (1≤n≤N) [87–89],

$$V(r_1,\cdots,r_N) = \sum_{i=1}^{N} V_{1B}(r_i) + \sum_{i<j}^{N} V_{2B}(r_i,r_j)$$

$$+ \sum_{i<j<k}^{N} V_{3B}(r_i,r_j,r_k) + \cdots + V_{NB}(r_i,r_j,r_k,\cdots,r_N) \quad (5)$$

$$= E(r_1,\ldots,r_N) - \sum_{i=1}^{N} E(r_i), \quad (6)$$

where r_i is a collection of coordinates of all atoms of molecule i. V_{1B}-term is zero. Since every calculation of an energy of a fragment is carried out with the basis functions of the full system as described in the computational details, we omit explicitly indicating which basis is used in the evaluation of the energies. The V_{2B}

term is the pairwise interaction energy,

$$V_{2B}(r_i,r_j) = E(r_i,r_j) - E(r_i) - E(r_j). \quad (7)$$

The definition of higher n-body terms V_{nB} is given recursively by subtracting all fewer-body interaction terms:

$$V_{nB}(r_1,\cdots,r_n) = V(r_1,\cdots,r_n) - \sum_{i<j}^{N} V_{2B}(r_i,r_j) - \cdots$$

$$- \sum_{i<j<\cdots<n-1}^{N} V_{(n-1)B}(r_i,r_j,\cdots,r_{n-1}). \quad (8)$$

Thus, a single 3-body interaction energy term of 3 molecules at r_1, r_2, r_3, is:

$$V_{3B}(r_1,r_2,r_3) = V(r_1,r_2,r_3) - \sum_{i<j}^{3} V_{2B}(r_i,r_j) \quad (9)$$

$$= E(r_1,r_2,r_3) - \sum_{i<j}^{3} E(r_i,r_j) + \sum_{i}^{3} E(r_i). \quad (10)$$

For the 17-water cluster, with the complex always being included as one of the monomers in the calculations, we can evaluate $17!/(2!(17-2)!) = 136$ V_{3B}-terms. Lastly, by using the recursive relation in equation 8 we can obtain the sum of all n-body interactions of the 18-body (17 waters + 1 complex) system with $n \geq 3$ as:

$$V_{\mathrm{MB}} = \sum_{i<j<k}^{18} V_{3B}(r_i,r_j,r_k) + \cdots + V_{18B}(r_1,\ldots,r_{18}) \quad (11)$$

$$= V(r_1,\ldots,r_{18}) - \sum_{i<j}^{18} V_{2B}(r_i,r_j) \quad (12)$$

In this way, V_{MB} defines the total many-body interaction energy *not* explicitly captured by pairwise-additive potentials, such as the static-charge OPLS.

Analysis of the components of the total interaction energy of $[Fe(bpy)_3]^{2+}\cdots 17H_2O$ in LS and HS states are shown in Table 2. r_c denotes the coordinates of the complex. The first entry shows the sum of all 2-body interaction energies where one of the monomers is the complex (17 in total). The next entry shows the sum of the rest of the 2-body energies from pure water-pairs, 136 in total. Adding these terms up gives us the third row, the sum of all 2-body interaction energies. The fourth row contains the sum of all 3-body energies where one of the monomers is the complex, while the fifth row shows the total interaction energy of all $\geq$3-body terms, as given by equation 12. The water-water interactions are of course less strong individually, but summed up, we note that they contribute *more* to the total 2-body interactions than the complex-water interactions. From LS to HS state of $[Fe(bpy)_3]^{2+}\cdots 17H_2O$, the sum of the water-water two-body terms increases by 0.075 eV, and by 0.505 eV for the complex−water interaction, translating to an overall less bound shell, in the HS state, when only including pairwise additive interactions. This contrasts with the findings from the dimer curves of poses A and B in the previous section, showing the limitations of the capabilities of such models in describing the concerted interactions within a solvation shell.

Analysing the higher-order many-body terms first, interestingly, the sum of the three-body terms involving the complex, $\sum_{i<j}^{17} V_{3B}(r_c,r_i,r_j)$ *increases* the total interaction energy in all cases (an anticooperative effect). Secondly, despite the metal-centered character of both spin states of $[Fe(bpy)_3]^{2+}$, we found that the sum of all higher body terms (n>2) contribute differently to the total interaction energy: 10.0% and 18.8% for LS and HS states, respectively. If one solely accounts for the two-body terms of the total interaction energy, one would find that the interaction in the LS state of$[Fe(bpy)_3]^{2+}\cdots 17H_2O$ is more stable by 0.580 eV. However, the picture changes dramatically when one adds higher-body terms to the expression of the total interaction energy: The HS state becomes the more stable state by 0.161 eV. The two main factors that can cause this increase in many-body interactions are (i) changes in the electronic density of the complex which causes changes in the polarization of the water network, and (ii) changes in the nuclear geometry of the water network itself. By investigating the partial atomic charges on the complex in the two states (see table S5.4 in the S.I.), we observe that the CHELPG-method predicts that just under half an electronic elementary charge is moved from Fe to the ligands upon excitation to the HS state. In the last section we observed that the HS 17-shell is tightened by an average O−O distance decrease of 0.051 Å. Both changes in charge density and geometry are subtle, but we note that V_{CS} is a lot less binding than V_{WW}, as shown in figure 10. This points toward the changes in the water-water geometry, rather than changes in electronic density of the complex causing this change in the higher-body interaction term. To test this, we calculated the many-body interaction energy using the LS electronic configuration on the HS geometry, $V_{\mathrm{MB}}^{\mathrm{LS}}(r_{\mathrm{HS}})$, and vice versa, $V_{\mathrm{MB}}^{\mathrm{HS}}(r_{\mathrm{LS}})$ (see the S.I. section S7 for details). $V_{\mathrm{MB}}^{\mathrm{LS}}(r_{\mathrm{HS}})$ made up 12.2 % of the total interaction energy, whereas $V_{\mathrm{MB}}^{\mathrm{HS}}(r_{\mathrm{LS}})$ made up only 10.1%. We thus conclude that the HS geometry is required for the increase in many-body interactions, but without the change in multiplicity as well, the increase is a lot less dramatic. To summarize, the drastic increase in the many-body interaction strength is only allowed for the right water shell geometries, but when the geometric conditions are fulfilled, a change in spin configuration can increase the many-body interactions significantly. Future studies could test this mechanism further, e.g. by repeating the presented analysis on systems with larger transfers of charge upon excitation, and/or smaller changes in nuclear geometry upon excitation.

Experimentally, Miller *et al.* found that the HS lifetime of the complex was modulated by choice of solvent, and attributes this to changes in solvent-solvent and/or solvent-solute interactions, brought about by the change in the complex geometry due to the Fe-N elongation[90]. Our findings in this, and the previous sections suggest that especially the solvent-solvent interactions are affected by the LS→HS transition.

Previous solvation studies have used MM and electrostatic QM/MM models[8]. The former of these models neglects many-body effects such as induced polarization everywhere, and the latter only in the MM region. these models produce solvation

structures that are generally less strongly bound than what is predicted by fully QM AIMD simulations[8,20], which includes many-body interaction. Based on the analysis above, we propose that some of this discrepancy is caused by many-body effects within the solvation shell of the complex. Interestingly, as the QM/MM MD simulations did also not fully reproduce the AIMD results, further testing this hypothesis using polarizable embedding QM/MM models[91–94] could prove insightful, as they would allow for simulations including more solvent than what is feasible with AIMD, while still including solvent polarization. For this specific complex, the aforementioned role of the pose-A water molecules in their close vicinity to the iron center could mean that simply including these three molecules in the QM subsystem might already improve the model significantly. This, however comes with the added complexity of maintaining the solvent molecules in the QM region[95], sacrificing the modeling of diffusion-based processes, or employing an adaptive QM/MM scheme that allows for the molecules to switch description, which comes at a considerable increase in computational cost[96].

Table 2 Counterpoise-corrected B3LYP-D3(BJ)/def2-TZVP many-body expansions (and their percentages) of the total interaction energy in eV of $[Fe(bpy)_3]^{2+}\cdots 17H_2O$ in LS and HS states. r_c denotes the coordinates of the complex.

Energy	$[Fe(bpy)_3]^{2+}\cdots 17H_2O$			
	LS		HS	
$\sum_i^{17} V_{2B}\ (r_c, r_i)$	−3.274	(40.2%)	−2.769	(33.3%)
$\sum_{i<j}^{17} V_{2B}\ (r_i \neq r_c, r_j \neq r_c)$,	−4.059	(49.8%)	−3.984	(47.9%)
$\sum_{i<j,}^{18} V_{2B}(r_i, r_j)$	−7.333	(90.0%)	−6.753	(81.2%)
$\sum_{i<j}^{17} V_{3B}\ (r_c,\ r_i,\ r_j)$	0.271	(−3.3%)	0.232	(−2.8%)
V_{MB}	−0.818	(10.0%)	−1.559	(18.8%)
V	−8.151	(100.0%)	−8.312	(100.0%)

Conclusions

In this work, we have benchmarked computational models for describing the noncovalent interactions between the photoswitch-workhorse complex $[Fe(bpy)_3]^{2+}$ in two electronic states, and the neighbouring water molecules in its closest solvation shell. Our benchmark showed that long-range dispersion effects play a significant role in governing the solute-solvent noncovalent interactions which cannot be ignored, but also that these effects are accurately approximated by computationally inexpensive, parametrized models such as the D3(BJ) correction, even though these have not been specifically created with solvation in mind. When compared to DLPNO-CCSD(T) binding energies, each of the tested dispersion correction models gave results within the limit of chemical accuracy within one another. Thus, for future solvation-shell sampling studies of similar systems, we would recommend including of long-range correlation via one of the dispersion-corrections studied here, if using GGA functionals or even if also including exact exchange. We then continued to analyse the solute-solvent and solvent-solvent binding energy interactions and geometries of water shells consisting of 17 to 20 waters, correlating the gas-phase relaxed geometry to the thermally averaged molecular densities sampled using classical MD with CHELPG-derived charges and LJ parameters from the OPLS and TIP4PEW force fields for the complex and water, respectively. Comparing to AIMD-structure from literature, we first identified that part of the cause for the more structured AIMD RDF is due to simulation-size limitations of the more costly AIMD model. To account for the rest of the AIMD/MMMD differences, we then analysed the effects of non-pairwise additive interactions within the solvation shells. Here, we found that the cause for the LS-HS changes in shell geometries did not originate from changes in the direct solute-shell interactions, but rather in changes in interactions between the water molecules themselves, induced by the geometric and electronic changes within the complex. This, again, increases the challenge for purely pairwise additive models such as OPLS/TIP4PEW, which cannot describe changes in water-water interactions due to e.g. polarization induced by the electronic density of the complex. Diving deeper into the many-body interactions of the shell, we found that almost 19% of the total interaction energy of the HS state is from non-pairwise additive interactions, almost twice as much as for the LS state, indicating that even relatively small electronic and geometric perturbations of the system can rather drastically change the ratio with which the terms in the many body expansion convergence to zero. Such changes of interactions within the solvent shell can have large impact on the desired functionality of the complex[90]. Future work on similar systems should then strive to strike the right balance between choosing models that include these kinds of effects, while keeping in mind what other effects on the sampled structure can be caused by the limits demanded by increases in computational cost, such as finite-size simulation effects, basis set superposition errors, and shorter sampling times.

Author Contributions

H. Zulfriki: Simulations, methodology, analysis, visualization, writing - original draft. M. Pápai: Simulations, methodology, analysis, writing - original draft. A. O. Dohn: Conceptualization, simulations, methodology, analysis, visualization, supervision, project administration, writing - original draft and review & editing.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors thank Dr. G. Levi for discussions and feedback.

Funding was provided by the Icelandic Research Fund, Grant 196279-051. M. Pápai acknowledges funding from the following institutions: The Independent Research Fund Denmark, Grant No. 8021-00347B; the Hungarian National Research, Development and Innovation Fund, Grant No. NKFIH PD 134976; the Government of Hungary and the European Regional Development Fund under Grant No. VEKOP-2.3.2-16-2017-00015; the János Bolyai Scholarship of the Hungarian Academy of Sciences. The calculations were carried out at the Icelandic High-Performance Computing Center , and at the DTU High Performance Computing Center.

Notes and references

1 M. Chergui, *Acc. Chem. Res.*, 2015, **48**, 801–808.

2 K. J. Gaffney, *Chem. Sci.*, 2021, **12**, 8010–8025.

3 K. Haldrup, G. Vankó, W. Gawelda, A. Galler, G. Doumy, A. M. March, E. P. Kanter, A. Bordage, A. Dohn, T. B. van Driel, K. S. Kjær, H. T. Lemke, S. E. Canton, J. Uhlig, V. Sundström, L. Young, S. H. Southworth, M. M. Nielsen and C. Bressler, *The Journal of Physical Chemistry A*, 2012, **116**, 9878–9887.

4 E. Biasin, T. B. van Driel, K. S. Kjær, A. O. Dohn, M. Christensen, T. Harlang, P. Vester, P. Chabera, Y. Liu, J. Uhlig, M. Pápai, Z. Németh, R. Hartsock, W. Liang, J. Zhang, R. Alonso-Mori, M. Chollet, J. M. Glownia, S. Nelson, D. Sokaras, T. A. Assefa, A. Britz, A. Galler, W. Gawelda, C. Bressler, K. J. Gaffney, H. T. Lemke, K. B. Møller, M. M. Nielsen, V. Sundström, G. Vankó, K. Wärnmark, S. E. Canton and K. Haldrup, *Phys. Rev. Lett.*, 2016, **117**, 013002.

5 T. B. van Driel, K. S. Kjær, R. W. Hartsock, A. O. Dohn, T. Harlang, M. Chollet, M. Christensen, W. Gawelda, N. E. Henriksen, J. G. Kim, K. Haldrup, K. H. Kim, H. Ihee, J. Kim, H. Lemke, Z. Sun, V. Sundström, W. Zhang, D. Zhu, K. B. Møller, M. M. Nielsen and K. J. Gaffney, *Nat. Commun.*, 2016, **7**, 13678.

6 K. Haldrup, G. Levi, E. Biasin, P. Vester, M. G. Laursen, F. Beyer, K. S. Kjær, T. Brandt van Driel, T. Harlang, A. O. Dohn, R. J. Hartsock, S. Nelson, J. M. Glownia, H. T. Lemke, M. Christensen, K. J. Gaffney, N. E. Henriksen, K. B. Møller and M. M. Nielsen, *Phys. Rev. Lett.*, 2019, **122**, 063001.

7 E. Biasin, Z. W. Fox, A. Andersen, K. Ledbetter, K. S. Kjær, R. Alonso-Mori, J. M. Carlstad, M. Chollet, J. D. Gaynor, J. M. Glownia, K. Hong, T. Kroll, J. H. Lee, C. Liekhus-Schmaltz, M. Reinhard, D. Sokaras, Y. Zhang, G. Doumy, A. M. March, S. H. Southworth, S. Mukamel, K. J. Gaffney, R. W. Schoenlein, N. Govind, A. A. Cordones and M. Khalil, *Nature Chemistry*, 2021, **13**, 343–349.

8 M. Abedi, G. Levi, D. B. Zederkof, N. E. Henriksen, M. Pápai and K. B. Møller, *Phys. Chem. Chem. Phys.*, 2019, **21**, 4082–4095.

9 G. Prampolini, F. Ingrosso, A. Segalina, S. Caramori, P. Foggi and M. Pastore, *Journal of Chemical Theory and Computation*, 2018, **15**, 529–545.

10 V. Diez-Cabanes, G. Prampolini, A. Francés-Monerris, A. Monari and M. Pastore, *Molecules*, 2020, **25**, 3084.

11 A. O. Dohn, *International Journal of Quantum Chemistry*, 2020, **120**, 1–22.

12 W. Gawelda, V.-T. Pham, M. Benfatto, Y. Zaushitsyn, M. Kaiser, D. Grolimund, S. L. Johnson, R. Abela, A. Hauser, C. Bressler and M. Chergui, *Phys. Rev. Lett.*, 2007, **98**, 057401.

13 W. Zhang, R. Alonso-Mori, U. Bergmann, C. Bressler, M. Chollet, A. Galler, W. Gawelda, R. Hadt, R. Hartsock, T. Kroll, K. Kjær, K. Kubiček, H. Lemke, H. Liang, D. Meyer, M. Nielsen, C. Purser, J. Robinson, E. Solomon, Z. Sun, D. Sokaras, T. Van Driel, G. Vankó, T.-C. Weng, D. Zhu and K. Gaffney, *Nature*, 2014, **509**, 345–348.

14 G. Auböck and M. Chergui, *Nat. Chem.*, 2015, **7**, 629–633.

15 K. S. Kjær, T. B. Van Driel, T. C. B. Harlang, K. Kunnus, E. Biasin, K. Ledbetter, R. W. Hartsock, M. E. Reinhard, S. Koroidov, L. Li, M. G. Laursen, F. B. Hansen, P. Vester, M. Christensen, K. Haldrup, M. M. Nielsen, A. O. Dohn, M. I. Pápai, K. B. Møller, P. Chabera, Y. Liu, H. Tatsuno, C. Timm, M. Jarenmark, J. Uhlig, V. Sundström, K. Wärnmark, P. Persson, Z. Németh, D. S. Szemes, E. Bajnóczi, G. Vankó, R. Alonso-Mori, J. M. Glownia, S. Nelson, M. Sikorski, D. Sokaras, S. E. Canton, H. T. Lemke and K. J. Gaffney, *Chem. Sci.*, 2019, **10**, 5749–5760.

16 K. S. Kj, 2017, **8**, 8.

17 L. M. Lawson Daku and A. Hauser, *J. Phys. Chem. Lett.*, 2010, **1**, 1830–1835.

18 S. Grimme, *J. Chem. Phys.*, 2006, **124**, 034108.

19 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

20 L. M. Lawson Daku, *Phys. Chem. Chem. Phys.*, 2018, **20**, 6236–6253.

21 I.-C. Lin, A. P. Seitsonen, I. Tavernelli and U. Rothlisberger, *Journal of Chemical Theory and Computation*, 2012, **8**, 3902–3910.

22 K. Forster-Tonigold and A. Groß, *The Journal of Chemical Physics*, 2014, **141**, 064501.

23 M. J. Gillan, D. Alfè and A. Michaelides, *The Journal of Chemical Physics*, 2016, **144**, 130901.

24 S. Grimme and J.-P. Djukic, *Inorganic Chemistry*, 2011, **50**, 2619–2628.

25 S. Grimme, A. Hansen, J. G. Brandenburg and C. Bannwarth, *Chemical Reviews*, 2016, **116**, 5105–5154.

26 D. J. Liptrot and P. P. Power, *Nature Reviews Chemistry*, 2017, **1**, 0004.

27 M. Bursch, E. Caldeweyher, A. Hansen, H. Neugebauer, S. Ehlert and S. Grimme, *Accounts of Chemical Research*, 2018, **52**, 258–266.

28 L. Zhou, J. Xu, L. Xu and X. Wu, *The Journal of Chemical Physics*, 2019, **150**, 124505.

29 D. Khakhulin, L. M. Lawson Daku, D. Leshchev, G. Newby, M. Jarenmark, C. Bressler, M. Wulff and S. Canton, *Phys. Chem. Chem. Phys.*, 2019, **21**, 9277–9284.

30 C. W. Stark, A. Trummal, M. Uudsemaa, J. Pahapill, M. Rammo, K. Petritsenko, M.-M. Sildoja and A. Rebane, *Commun. Chem.*, 2019, **2**, 1–6.

31 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.

32 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.

33 S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456–1465.

34 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.

35 F. Neese, *WIREs: Comput. Mol. Sci.*, 2012, **2**, 73–78.

36 F. Neese, *WIREs: Comput. Mol. Sci.*, 2018, **8**, e1327.

37 J. P. Perdew and K. Schmidt, AIP Conf. Proc., 2001, pp. 1–20.

38 A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098.

39 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785.

40 K. Eichkorn, O. Treutler, H. Öhm, M. Häser and R. Ahlrichs, *Chemical physics letters*, 1995, **240**, 283–290.

41 F. Weigend, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1057–1065.

42 M. Feyereisen, G. Fitzgerald and A. Komornicki, *Chem. Phys. Lett.*, 1993, **208**, 359–363.

43 G. L. Stoychev, A. A. Auer and F. Neese, *J. Chem. Theory Comput.*, 2017, **13**, 554–562.

44 S. F. Boys and F. Bernardi, *Mol. Phys.*, 1970, **19**, 553–566.

45 J. Zheng, X. Xu and D. G. Truhlar, *Theor. Chem. Acc.*, 2011, **128**, 295–305.

46 E. Caldeweyher, C. Bannwarth and S. Grimme, *J. Chem. Phys.*, 2017, **147**, 034112.

47 E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth and S. Grimme, *J. Chem. Phys.*, 2019, **150**, 154122.

48 O. A. Vydrov and T. Van Voorhis, *The Journal of chemical physics*, 2010, **133**, 244103.

49 W. Hujo and S. Grimme, *J. Chem. Theory Comput.*, 2011, **7**, 3866–3871.

50 B. Axilrod and E. Teller, *J. Chem. Phys.*, 1943, **11**, 299–300.

51 Y. Muto, *J. Phys. Math. Soc. Japan*, 1943, **17**, 629.

52 S. A. Ghasemi, A. Hofstetter, S. Saha and S. Goedecker, *Phys. Rev. B*, 2015, **92**, 045131.

53 J. Arago, E. Orti and J. C. Sancho-Garcia, *J. Chem. Theory Comput.*, 2013, **9**, 3437–3443.

54 C. Riplinger, P. Pinski, U. Becker, E. F. Valeev and F. Neese, *J. Chem. Phys.*, 2016, **144**, 024109.

55 M. Saitow, U. Becker, C. Riplinger, E. F. Valeev and F. Neese, *J. Chem. Phys.*, 2017, **146**, 164105.

56 D. G. Liakos, Y. Guo and F. Neese, *J. Phys. Chem. A*, 2020, **124**, 90–100.

57 F. Neese, F. Wennmohs, A. Hansen and U. Becker, *Chem. Phys.*, 2009, **356**, 98–109.

58 R. Izsák and F. Neese, *J. Chem. Phys.*, 2011, **135**, 144105.

59 J. F. Ouyang, M. W. Cvitkovic and R. P. Bettens, *J. Chem. Theory Comput.*, 2014, **10**, 3699–3707.

60 P. Li and K. M. Merz, *Journal of Chemical Information and Modeling*, 2016, **56**, 599–604.

61 C. M. Breneman and K. B. Wiberg, *Journal of Computational Chemistry*, 1990, **11**, 361–373.

62 H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura and T. Head-Gordon, *The Journal of Chemical Physics*, 2004, **120**, 9665–9678.

63 D. Case, H. Aktulga, K. Belfon, I. Ben-Shalom, S. Brozell, D. Cerutti, I. T.E. Cheatham, G. Cisneros, V. Cruzeiro, T. Darden, R. Duke, G. Giambasu, M. Gilson, H. Gohlke, A. Goetz, R. Harris, S. Izadi, S. Izmailov, C. Jin, K. Kasavajhala, M. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. Simmerling, N. Skrynnikov, J. Smith, J. Swails, R. Walker, J. Wang, H. Wei, R. Wolf, X. Wu, Y. Xue, D. York, S. Zhao, and P. Kollman, *Amber 2021*, 2021.

64 P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, R. P. Wiewiora, B. R. Brooks and V. S. Pande, *PLOS Computational Biology*, 2017, **13**, e1005659.

65 K. Haldrup, A. O. Dohn, M. L. Shelby, M. W. Mara, A. B. Stickrath, M. R. Harpham, J. Huang, X. Zhang, K. B. Møller, A. Chakraborty, F. N. Castellano, D. M. Tiede and L. X. Chen, *The Journal of Physical Chemistry A*, 2016, **120**, 7475–7483.

66 K. Haldrup, W. Gawelda, R. Abela, R. Alonso-Mori, U. Bergmann, A. Bordage, M. Cammarata, S. E. Canton, A. O. Dohn, T. B. van Driel, D. M. Fritz, A. Galler, P. Glatzel, T. Harlang, K. S. Kjær, H. T. Lemke, K. B. Møller, Z. Németh, M. Pápai, N. Sas, J. Uhlig, D. Zhu, G. Vankó, V. Sundström, M. M. Nielsen and C. Bressler, *The Journal of Physical Chemistry B*, 2016, **120**, 1158–1168.

67 K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov and A. D. Mackerell, *Journal of Computational Chemistry*, 2009, NA–NA.

68 H. A. Lorentz, *Annalen der Physik*, 1881, **248**, 127–136.

69 D. Berthelot, *Comptes. Rendus. Acad. Sci.*, 1898, **126**, 1703—-1855.

70 W. Kabsch, *Acta Crystallogr. A*, 1976, **32**, 922–923.

71 J. C. Kromann, *Calculate Root-mean-square deviation (RMSD) of Two Molecules Using Rotation, GitHub, version 1.2.7*, 2018, `http://github.com/charnley/rmsd`.

72 H. Zulfriki, M. Pápai and A. O. Dohn, 2022.

73 A. Stone, *The Theory of Intermolecular Forces*, Oxford University Press, 2013.

74 M. Pápai, M. Abedi, G. Levi, E. Biasin, M. M. Nielsen and K. B. Møller, *The Journal of Physical Chemistry C*, 2019, **123**, 2056–2065.

75 K. S. Kjær, T. B. V. Driel, T. C. B. Harlang, K. Kunnus, E. Biasin, K. Ledbetter, R. W. Hartsock, M. E. Reinhard, S. Koroidov, L. Li, M. G. Laursen, F. B. Hansen, P. Vester, M. Christensen, K. Haldrup, M. M. Nielsen, A. O. Dohn, M. I. Pápai, K. B. Møller, P. Chabera, Y. Liu, H. Tatsuno, C. Timm, M. Jarenmark, J. Uhlig, V. Sundstöm, K. Wärnmark, P. Persson, Z. Németh, D. S. Szemes, É. Bajnóczi, G. Vankó, R. Alonso-Mori, J. M. Glownia, S. Nelson, M. Sikorski, D. Sokaras, S. E. Canton, H. T. Lemke and K. J. Gaffney, *Chemical Science*, 2019, **10**, 5749–5760.

76 A. K. Das, R. V. Solomon, F. Hofmann and M. Meuwly, *The Journal of Physical Chemistry B*, 2016, **120**, 206–216.

77 G. Vankó, A. Bordage, M. Pápai, K. Haldrup, P. Glatzel, A. M. March, G. Doumy, A. Britz, A. Galler, T. Assefa, D. Cabaret, A. Juhin, T. B. van Driel, K. S. Kjær, A. Dohn, K. B. Møller, H. T. Lemke, E. Gallo, M. Rovezzi, Z. Németh, E. Rozsályi, T. Rozgonyi, J. Uhlig, V. Sundström, M. M. Nielsen, L. Young, S. H. Southworth, C. Bressler and W. Gawelda, *The Journal of Physical Chemistry C*, 2015, **119**, 5888–5902.

78 D. F. Crouse, *IEEE Transactions on Aerospace and Electronic*

Systems, 2016, **52**, 1679–1696.

79 P. Wernet, D. Nordlund, U. Bergmann, M. Cavalleri, M. Odelius, H. Ogasawara, L. A. Naslund, T. K. Hirsch, L. Ojamae, P. Glatzel, L. G. M. Pettersson and A. Nilsson, *Science*, 2004, **304**, 995–999.

80 D. McQuarrie and M. A, *Statistical Mechanics*, Harper & Row, 1975.

81 M. E. Tuckerman, *Statistical mechanics theory and molecular simulation*, Oxford University Press, New York, 2010.

82 J.-P. Hansen and I. MacDonald, *Theory of Simple Liquids*, Elsevier, 4th edn, 2013.

83 P. A. Egelstaff, *An Introduction to the Liquid State*, Academic Press, 1967.

84 J. J. Salacuse, A. R. Denton and P. A. Egelstaff, *Physical Review E*, 1996, **53**, 2382–2389.

85 *Fluctuation Theory of Solutions*, ed. P. E. Smith, E. Matteoli and J. P. O. Connell, CRC Press, 2016.

86 A. O. Dohn, E. Biasin, K. Haldrup, M. M. Nielsen, N. E. Henriksen and K. B. Møller, *Journal of Physics B: Atomic, Molecular and Optical Physics*, 2015, **48**, 244010.

87 D. Hankins, J. Moskowitz and F. Stillinger, *J. Chem. Phys.*, 1970, **53**, 4544–4554.

88 G. R. Medders, V. Babin and F. Paesani, *Journal of Chemical Theory and Computation*, 2013, **9**, 1103–1114.

89 G. A. Cisneros, K. T. Wikfeldt, L. Ojamäe, J. Lu, Y. Xu, H. Torabifard, A. P. Bartók, G. Csányi, V. Molinero and F. Paesani, *Chem. Rev.*, 2016, **116**, 7501–7528.

90 J. N. Miller and J. K. McCusker, *Chemical Science*, 2020, **11**, 5191–5204.

91 D. Loco, L. Lagardère, S. Caprasecca, F. Lipparini, B. Mennucci and J.-P. Piquemal, *Journal of Chemical Theory and Computation*, 2017, **13**, 4025–4033.

92 E. Örn Jónsson, A. O. Dohn and H. Jónsson, *Journal of Chemical Theory and Computation*, 2019, **15**, 6562–6577.

93 A. O. Dohn, E. Örn Jónsson and H. Jónsson, *Journal of Chemical Theory and Computation*, 2019, **15**, 6578–6587.

94 T. Giovannini, A. Puglisi, M. Ambrosetti and C. Cappelli, *Journal of Chemical Theory and Computation*, 2019, **15**, 2233–2245.

95 B. Kirchhoff, E. Örn Jónsson, A. O. Dohn, T. Jacob and H. Jónsson, *Journal of Chemical Theory and Computation*, 2021, **17**, 5863–5875.

96 A. W. Duster, C.-H. Wang, C. M. Garza, D. E. Miller and H. Lin, *WIREs Computational Molecular Science*, 2017, **7**, e1310.