Exploring the rich energy landscape of sulfate-water clusters $\textrm{SO}_{4}^{2-}(\textrm{H}_{2}\textrm{O})_{n=3-7}$: an electronic structure approach

Journal:	*The Journal of Physical Chemistry*
Manuscript ID:	jp-2011-06064n.R1
Manuscript Type:	Special Issue Article
Date Submitted by the Author:	n/a

Complete List of Authors:	Lambrecht, Daniel; University of California, Chemistry Clark, Gary; UC Berkeley, Bioengineering Head-Gordon, Teresa; University of California, Berkeley, Bioengineering Head-Gordon, Martin; University of California, Berkeley, Chemistry

Exploring the rich energy landscape of sulfate-water clusters $SO_4^{2-}(H_2O)_{n=3-7}$: an electronic structure approach

Daniel S. Lambrecht,*,† Gary N. I. Clark,‡ Teresa Head-Gordon,‡ and Martin Head-Gordon†

Department of Chemistry, University of California, Berkeley, Department of Bioengineering, University of California, Berkeley, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, and Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley

E-mail: daniel.lambrecht@berkeley.edu

Abstract

We present a re-investigation of sulfate-water clusters $SO_4^{2-}(H_2O)_{n=3-7}$, which involves several new aspects. Using a joint molecular mechanics / first principles approach, we perform exhaustive searches for stable cluster geometries, showing that the sulfate-water landscape is much richer than anticipated previously. We check the compatibility of the new structures with experiment by comparing vertical detachment energies (VDEs) calculated at the B3LYP/6-311++G** level of theory and determine the energetic ordering of the isomers at the RI-MP2/aug-cc-pVTZ level. Our results are benchmarked carefully against reference energies of estimated CCSD(T)/aug-cc-VTZ quality and VDEs of CCSD(T)/aug-cc-pVDZ quality. Furthermore, we calculate anharmonic vibrational corrections for up to the $n = 6$ clusters, which are shown to be significant for isomer energy ordering. We use energy decomposition analysis (EDA) based on the absolutely localized fragment (ALMO) expansion to gain chemical insight into the binding motifs.

1 Introduction

Sulfates play an important role in industrial and environmental processes ranging from their stabilizing effects on proteins (as part of the famous Hofmeister series[1,2]), as an algaecide,[3] and use in fertilizers.[4] Studies of sulfate-water clusters in particular have direct environmental relevance in formation of aerosols that produce acid rain by-products from fossil fuel consumption, and more recently the same aerosols have been proposed as a controversial geoengineering solution to global warming.[5] Although sulfate aerosols are chemically complex, with likely significant components of organic material, sulfate nano-solvation with water is a fundamental first step in understanding aerosol formation.

Sulfate-water clusters also provides unique insight into the molecular nature of ionic solvation because the sulfate anion charge allows for experimental mass selection of specific sizes, thus allowing the study of formation of the solvation shell by adding one solvent molecule at a time. Infrared[6,7] and photoelectron[8] spectroscopy, as well as theoretical modeling[6–9] studies, have revealed

*To whom correspondence should be addressed

†Department of Chemistry, University of California, Berkeley

‡Department of Bioengineering, University of California, Berkeley

¶Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley

§Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley

that even the smaller sulfate-water clusters exhibit a complex set of low-lying structures.[8]

In this paper we re-investigate sulfate-water clusters $SO_4^{2-}(H_2O)_n$ for $n = 3-7$. This is an interesting size range, as it spans the region where the sulfate's excess electron is just barely stabilized by the solvate environment ($n = 3$), over some intermediate region ($n = 4,5$) and finally the region ($n = 6,7$) where cooperative effects in the water network become important. We note that there has been quite some dispute about the closure of the first solvation shell, with earlier works reporting that sulfate is typically surrounded by $7-13$ waters,[10–18] whereas more recent publications agree that the first solvation shell closes at around $n = 12$.[6,19,20] In particular, we perform an exhaustive search of the energy landscape for the smaller $n = 3-7$ clusters via a combined force field/first principles approach using the polarizable AMOEBA force field[21–23] in combination with density functional theory (DFT) and resolution of the identity second-order Møller-Plesset perturbation theory (RI-MP2) for a subsequent energy ranking and the calculation of vertical detachment energies (VDEs). In order to assess the quality of the DFT and RI-MP2 energies and VDEs we calibrate against coupled cluster theory including singles, doubles and perturbative triples (CCSD(T)). Finally, we check whether the novel structures agree with experiment by comparing the VDEs.

Another aspect of this work is the assessment of anharmonicity. Due to the floppy nature of the clusters a reliable energy ranking requires the calculation of zero-point vibrational energies including anharmonic effects. The results from previous studies[24] show that anharmonic effects can amount to 10-20% of the zero-point energy (ZPE) for the high-frequency modes, which is non-negligible with respect to the energetic ordering of the isomers (being sometimes separated by as little as 0.1 kcal/mol). So far these calculations have only been considered for the smaller clusters ($n = 1-3$).[24] We present estimates of anharmonic effects for up to $n = 6$ and show their influence on the energetic ordering of the isomers. Finally, since sulfate-water gives rise to clusters with very different topologies and binding motifs, it is interesting to look at them from the chemical perspective and to investigate the different contributions to the solvate-solvent and solvent-solvent interactions, especially in view of the hydrogen bonding. To this end we performed energy decomposition analysis (EDA) calculations[25–27] that allow for the separation of the binding energy into components such as frozen-fragment, polarization and charge-transfer contributions.

2 Computational methods

2.1 Exploring the potential energy surface

We use the AMOEBA (Atomic Multipole Optimized Energetics for Biomolecular Applications) force field with the TINKER software program for molecular mechanics and dynamics simulation[28] to explore the potential energy surface. The AMOEBA force field offers high fidelity to *ab initio* calculations but at a computational cost that allows for the fast calculation of a large number of trajectories, and has been shown to be a reliable generator of structures for use in hybrid energy approaches in a recent study of water-sulfate anion clusters.[23] We use the standard AMOEBA model of water[28,29] with a model of sulfate reported in the electronic supplementary material. The force field parameters are extracted from MP2 calculations using the optimized sulfate/water configurations at the level of MP2/aug-cc-pVTZ with a basis set superposition error (BSSE) correction and MP2 density.[30] The multipole moments for each structure are calculated from the optimized wave functions using distributed multipole analysis (DMA)[31,32] and the software of Stone.[33] The remaining van-der-Waals (vdW) and bond parameters are obtained by comparing the AMOEBA optimized structure, energy and frequency to the Gaussian03[34] optimized structures, energies, and frequencies calculated at the MP2/aug-cc-pVTZ level.[30] Details on the statistical performance of the AMOEBA model for sulfate-water as compared to MP2/aug-cc-pVTZ can be found in.[23]

We then explored the energy surface of sulfate-water by performing parallel tempering replica exchange molecular dynamics simulations. For each cluster size we ran 10 different temperature

baths between 140 and 500 K, swapping configurations (replicas) between baths according to the Metropolis criterion. Each simulation was 0.5 ns long and yielded 1000 starting configurations for subsequent geometry optimization, i.e. 10,000 initial geometries for each cluster size. In order to determine the unique isomers of the optimized structures, we tried two different criteria: (1.) a geometric comparison based on the number of hydrogen bonds and (2.) an energy-based criterion. We will report on the number of structures found using this approach in section Section 3.2.1 (see Column 2 of Section 3.2.1).

From previous benchmarks[23] we know that the average error bars in the AMOEBA energies are on the order of ± 2 kcal/mol. We aim at identifying the majority of isomers that are within about 2.5 kcal/mol of the global minimum, i.e. those that have Boltzmann factors larger than $\approx 2\%$ at 300 K. Therefore we initially chose all isomers with relative AMOEBA energies of about 4.5 kcal/mol or less have to be considered in the subsequent first principles calculations. As it turns out, however, the maximum errors in the AMOEBA energies can be several kcal/mol larger (see section Section 3.2.1). The energy window was therefore increased to 10 kcal/mol in order to be safe. We will discuss the number of "survivors" of this prescreening procedure in the results section (Section 3.1, see column 3 of Tab. Section 3.2.1).

2.2 First principles calculations

Although the statistical agreement of AMOEBA structures and energies with high-level calculations was good in previous benchmarks,[23] the energetic ordering can be mixed up in some cases and we also need a method for the reliable calculation of VDEs. Since the number of new isomers that we found is large (see section Section 3.2), we need an approach that is computationally feasible even for larger clusters. For the cluster range that we investigate here, RI-MP2 and DFT are viable compromises between speed and accuracy for the description of the energetic ordering of the closed-shell clusters. DFT is also typically robust enough to describe open-shell species for the calculation of detachment energies and relative energetics at a cost of an SCF calculation. In recent applications to the simulation of the PES of cyanide-water we observed that the popular B3LYP density functional[35–37] gives satisfactory agreement with reference calculations and with experiment, while having only modest computational demand. We use it here for the calculation of VDEs as well as energy decomposition analysis to understand the different contributions to intermolecular binding (i.e. frozen-fragment, polarization and charge-transfer contributions).[25] For the isomer energy ordering we use RI-MP2 theory, which is shown to be very accurate in the following benchmark calculations (it is not adequate for VDEs due to spin contamination in the radical).

In our benchmark calculations we calibrate the B3LYP and RI-MP2 binding energies and relative isomer energies against estimated CCSD(T)/aug-cc-pVTZ results. Altogether we consider 22 isomers in the $n = 3-5$ range. The reference energies are obtained by correcting the CCSD(T)/aug-cc-pVDZ results for basis set effects at the RI-MP2/aug-cc-pVTZ level of theory. To assess the quality of the B3LYP vertical detachment energies, we compare 10 VDEs for low-lying $n = 3,4$ clusters between B3LYP/6-311++G** and CCSD(T)/aug-cc-pVDZ (no MP2 estimates are used here since they are usually deemed unreliable for the calculation of VDEs).

Our calibration calculations are performed as follows. For B3LYP and RI-MP2 we use the dianion geometry optimized within the respective method and basis set. For the B3LYP calculations we tested the performance of the Pople-style 6-31+G*, 6-311++G**[38] and the Dunning aug-cc-pV(D,T)Z[39,40] basis sets and found that the 6-311++G** basis yields sufficiently accurate results at modest cost, such that we use this basis set throughout. All RI-MP2 calculations are performed within an aug-cc-pVTZ basis in combination with the matching fitting basis.[41] For CCSD(T) we use the RI-MP2/aug-cc-pVTZ geometry. Since the sulfate dianion is electronically stable only for $n \geq 3$, we determine all binding energies relative to the global minimum of $SO_4(H_2O)_3$. Relative isomer energies are calculated with respect to the global minimum for each cluster size according to the electronic energy (without zero-point corrections). Since we found anharmonic vibrational corrections to be

quite large, we have devoted separate method and results sections to their description (sections Section 2.3 and Section 3.3). Vertical detachment energies are calculated by taking the energy difference between the monoanion and the dianion energies at the dianion geometry.

All wave-function based calculations were performed within the frozen core approximation. A development version of the Q-Chem program package was used throughout.[42]

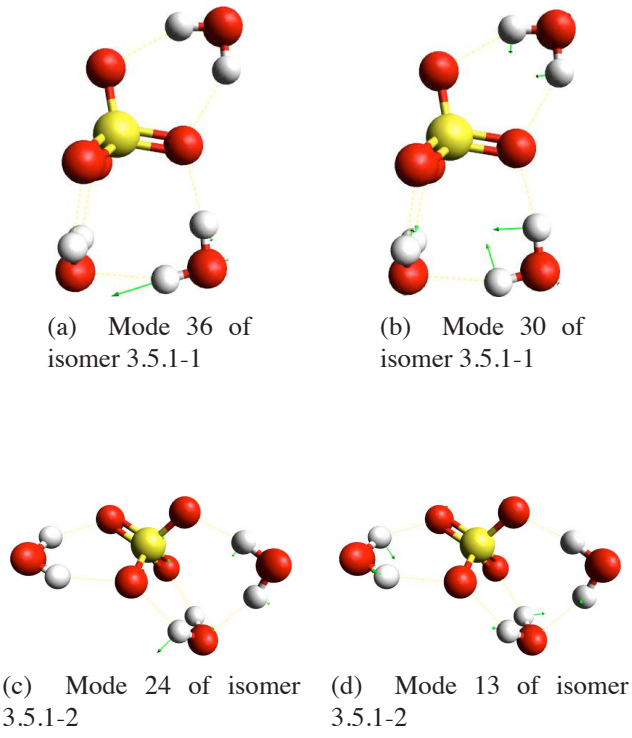
2.3 Zero-point vibrational energy

Anharmonicity and mode coupling can play an important role for weakly bound systems like solvation clusters. Unfortunately a full treatment is very expensive for the larger clusters, as it requires the use of high-level vibrational correlation methods and high-quality potential energy surfaces, which come at great computational cost. In vibrational configuration interaction (VCI) theory,[43] for example, quadruple and higher excitations are typically required to obtain reasonably converged results, which comes at the cost of an unfavorable scaling of $O(N^8)$ or higher and large memory requirements. The necessity for including such high excitations stems from the nature of the vibrational many-body interactions as well as size inconsistency of the truncated VCI expansion. Second-order vibrational perturbation theory (VPT2),[44,45] as a cheaper and size-consistent alternative, unfortunately faces problems with Fermi resonances, i.e. systems in which degeneracies in the vibrational energy levels lead to near-zero energy denominators.

We therefore resorted to the transition optimized spherical harmonics (TOSH) model[46] for the treatment of anharmonicities, as implemented in the Q-Chem program package.[42] TOSH uses an inexpensive model for the anharmonic wave function, which is based on first-order perturbation theory and is therefore not plagued by the zero-denominator problem. The TOSH model anharmonic wave function is obtained by shifting the harmonic wave function, which is normally centered about the equilibrium geometry, by an amount σ in order to model the asymmetry of the anharmonic wave function. The shift parameter σ is obtained by comparing the first-order perturbative energy expression based on the TOSH wave function with the conventional VPT2 energy expression. This incorporates the anharmonic correction in the wave function, such that TOSH arguably "gives second-order accuracy at only first-order cost"[47] and "does not suffer from the problems associated with Fermi resonances, which plague the VPT2 approach".[46]

Furthermore, resorting to the TOSH model reduces the number of third- and fourth-order derivatives of the energy with respect to nuclear displacements that have to be calculated, since it requires only diagonal fourth-order derivatives η_{iijj} at worst.[46] Nevertheless, the numerical calculation of these derivatives still requires many single-point calculations (only up to second derivatives are available analytically in Q-Chem[42]). In order to make the calculations feasible even for the larger clusters, we calculate the anharmonic corrections at the B3LYP level in combination with a small 6-31+G* basis set. To give the reader an impression of the computational cost, we note that even at this modest level of theory a calculation for a single $n = 6$ isomer took 23.2 days on a single-core 3.0 GHz Intel Xeon linux machine, of which 23 days are spent in the derivative calculation and 3 hours in the actual calculation of the TOSH energy.

To illustrate the breakdown of VPT2 and the performance of TOSH in the case of Fermi resonance, we consider two isomers ("3.5.1-1" and "3.5.1-2", explanation of nomenclature in section Section 3.2) of the $n = 3$ cluster for which Fermi resonance occurs. Table Table 1 shows the harmonic, VPT2 and TOSH frequencies for these clusters, and Figure Figure 1 shows the structures of the two isomers as well as normal-mode vectors for the Fermi-resonant modes. The VPT2 frequency of 4164 cm^{-1} for the water-water donor OH stretch (mode 36) of isomer 3.5.1-1 is unreasonably high. A typical range for the water OH stretch would be 3300–3650 cm^{-1}. We note that the harmonic frequency for mode 36 is 3611 cm^{-1}, which is within 0.6 cm^{-1} of the first overtone of mode 30, the concerted three-water bend at 1805.2 cm^{-1} (see Figure 1 (a) and (b)). The unreasonable VPT2 result is thus likely to be due to Fermi resonance between modes 36 and 30. The TOSH frequency of 3303 cm^{-1} for mode 36, on the other hand, is very rea-



(a) Mode 36 of isomer 3.5.1-1

(b) Mode 30 of isomer 3.5.1-1

(c) Mode 24 of isomer 3.5.1-2

(d) Mode 13 of isomer 3.5.1-2

Figure 1: Isomers 3.5.1-1 and 3.5.1-2 of the $n = 3$ cluster as well as illustrations of the Fermi-resonant mode pairs 36/30 and 24/13, respectively.

sonable.

Next we observe that the VPT2 frequency of 2402 cm^{-1} for mode 24 (water libration) of isomer 3.6 is unreasonably large. This is again due to Fermi resonance, as seen by comparing the harmonic frequencies of mode 24 (933.2 cm^{-1}) and 13 (a concerted intermolecular twisting motion of the water molecules at 466.5 cm^{-1}). The TOSH frequency for mode 24 is again at a reasonable 867.9 cm^{-1}.

In order to further assess the quality of the TOSH anharmonic corrections, we compare our TOSH/B3LYP/6-31+G* results with the CC-VSCF/MP2/TZP vibrational calculations of Miller et al.[24] for the global $n = 3$ minimum (see section Section 3.2 for a description of the isomers). We note that this is an example where VPT2 is expected to perform well, since no Fermi resonances occur. Table Table 2 shows the harmonic frequencies and anharmonic corrections obtained at the CC-VSCF, TOSH, VPT2 and VCI(2) levels of theory. We remark that this comparison of TOSH and CC-VSCF is not completely fair, since we are using the B3LYP/6-31+G* potential energy surface, whereas Miller et al. used an MP2/TZP potential. It is known that anharmonic frequencies from B3LYP potential energy surfaces are typically too low by on the order of 50-100 cm^{-1} in the high-frequency region compared to experiment, whereas MP2 frequencies show better agreement over the whole frequency range.[48] We therefore expect that our results overestimate anharmonicity in the high-frequency region, i.e. 3000 cm^{-1} and above. However, we have little alternative since anharmonic calculations on the MP2/TZP potential energy surface would be intractably expensive for the larger clusters (both because of the larger basis set and the unavailability of higher than first order analytic derivatives). We also note that CC-VSCF has limited accuracy itself, although the errors are usually sufficiently small to serve as a benchmark for our purposes.[49]

The anharmonic corrections $\Delta\omega$ are calculated as the difference between the harmonic and anharmonic frequencies. Since the absolute magnitude of anharmonic effects should be smaller than for the harmonic ZPE, adding anharmonicity as a correction at a rather modest basis set level should be viable. Overall we note that TOSH gets all trends in the anharmonic corrections right, i.e. all signs and the order of magnitude are correct. VPT2 on the other hand predicts the wrong sign in 2 cases, although the order of magnitude is generally correct. VCI(2) fails to predict the right signs in 10 cases as well as over- or underestimates the magnitude of the anharmonic correction significantly in the majority of cases, which clearly demonstrates that VCI(2) is too low a truncation of the VCI expansion to be useful for this cluster size. The root mean square deviation (RMSD) of the TOSH anharmonic corrections with respect to the CC-VSCF results is 41 cm^{-1}, which is slightly smaller than for VPT2 (51 cm^{-1}) and much better than VCI(2) (239 cm^{-1}).

As expected, we observe that the anharmonically corrected frequencies calculated on the B3LYP potential energy surface are systematically too small for the high-frequency modes (i.e. > 3000 cm^{-1}) by on the order of 100 cm^{-1} compared to CC-VSCF/MP2, whereas the lower frequencies are typically reproduced faithfully at the TOSH/B3LYP level.[50] The anharmonic corrections, however, are very similar at the TOSH/B3LYP and the CC-VSCF/MP2 levels. For

Table 1: Selected harmonic, VPT2 and TOSH frequencies for isomers no. 2 and 6 of the $n = 3$ cluster at the B3LYP/6-31+G* level (all in cm^{-1}). Modes 31-36 correspond to the symmetric and asymmetric OH stretches, mode 30 corresponds to concerted water bend, mode 24 is water libration, and mode 13 corresponds to an intermolecular water twisting motion. Fermi resonances are shown in bold.

	Harmonic		VPT2		TOSH	
Mode	3.5.1-1	3.5.1-2	3.5.1-2	3.5.1-2	3.5.1-1	3.5.1-2
36	**3611**	3606	**4164**	3413	**3303**	3358
35	3542	3529	3299	3308	3243	3286
34	3499	3514	3148	3196	3139	3199
33	3461	3473	3182	3139	3235	3148
32	3438	3445	3098	3091	3098	3092
31	3379	3379	3124	3079	3103	3060
30	**1805**	1780	1804	1782	1741	1749
	⋮		⋮		⋮	
24	937.8	**933.2**	1027	**2402**	975.5	**867.9**
	⋮		⋮		⋮	
13	456.0	**466.5**	794.1	658.6	794.1	585.9
	⋮		⋮		⋮	

some low-frequency vibrations we find deviations of up to 66 cm^{-1}, which may be due to differences in the potential energy surface, the vibrational correlation method, or combinations thereof.

The maximum deviation from the CC-VSCF results is 89 cm^{-1} for TOSH and 113 cm^{-1} for VPT2. We observe that VCI(2) fails completely for the highest-energy mode and predicts an anharmonic correction that is in error by 650 cm^{-1}. We conclude that TOSH yields anharmonic corrections that are of similar quality as VPT2 without the problems associated with Fermi resonances.

2.4 Benchmarks

In our benchmark calculations we find a decent agreement between the B3LYP/6-311++G** and the CCSD(T)/aug-cc-pVTZ(est.) binding energies (graph in supplementary material). Linear regression analysis with $f(x) = a + bx$ shows that the correlation between the B3LYP and the CCSD(T) binding energies is, overall, quite good: The slope b of 0.964 is close to 1 with a correlation coefficient of $r^2 = 0.998$. We are therefore confident that B3LYP/6-311++G** can reproduce most of the qualitative trends in the binding energies and is thus sufficient for analyzing the binding trends. However, the deviations are too large for a reliable determination of the isomer energetic ordering, since on an absolute scale the mean (maximum) deviations are on the order of 1.01 (2.56) kcal/mol, as shown in Table 3. We also note that the B3LYP energies deteriorate with increasing cluster size, since $b < 1$. Size-dependent binding trends should, however, be reproduced qualitatively correct.

For a more reliable energetic ordering we therefore resort to the RI-MP2/aug-cc-pVTZ results, which exhibit a good agreement with the CCSD(T)/aug-cc-pVTZ(est.) reference energies (Table 3). Linear regression analysis yields a slope of $b = 0.999$ and a correlation coefficient of $r^2 = 1.000$ for the RI-MP2/CCSD(T)(est.) plot (not shown here). The RI-MP2 results thus reproduce the coupled cluster trends almost perfectly. This shows that higher-order correlation effects are not significant in the clusters included in the benchmark set. We note, however, that it is known that some clusters require the inclusion of connected triples.[51] Furthermore, our reference values were obtained only at the aug-cc-pVTZ basis set level; it is therefore possible that larger basis sets would be required to ensure basis set convergence of the observed properties.

Table 2: Comparison of CC-VSCF/MP2/TZP calculations from Ref.[24] with the results of our TOSH/B3LYP/6-31+G* anharmonic correction scheme for the global minimum of the $n = 3$ isomers. ω ($\Delta\omega$) denotes the vibrational frequency (anharmonic correction) at the respective level of theory (all in cm^{-1}). RMSD and maximum errors are given with respect to the CC-VSCF results.

	Ref.[24]			This work			
Mode	$\omega_{harm.}$	$\omega_{CC\text{-}VSCF}$	$\Delta\omega$	$\omega_{harm.}$	$\Delta\omega_{TOSH}$	$\Delta\omega_{VPT2}$	$\Delta\omega_{VCI(2)}$
OH symmetric stretching	3698	3414	-284	3550	-195	-230	366
OH asymmetric stretching	3691	3396	-295	3544	-262	-262	-93
OH symmetric stretching	3687	3454	-233	3538	-232	-139	-7
OH symmetric stretching	3683	3463	-220	3533	-253	-253	-70
OH asymmetric stretching	3679	3380	-299	3517	-282	-353	-117
OH asymmetric stretching	3675	3410	-265	3514	-342	-342	-120
Bend of water	1770	1727	-43	1793	-25	70	68
Bend of water	1746	1711	-35	1768	-31	-31	128
Bend of water	1738	1699	-39	1759	-22	20	116
SO2 asymmetric stretching	1094	1076	-18	1052	-19	-19	150
SO2 symmetric stretching	1058	1043	-15	1021	-8	-15	151
SO2 asymmetric stretching	1016	1000	-16	985	-11	-11	152
SO4 stretching	913	901	-12	873	-2	-9	157
Libration of water	869	936	67	855	130	130	323
Libration of water	852	913	61	844	115	103	290
Libration of water	833	905	72	824	132	132	283
SO2 bending	589	585	-4	578	-8	-8	151
SO2 bending	580	576	-4	568	-7	-7	151
SO2 bending	567	565	-2	555	-4	-4	155
Intermolecular	470	550	80	486	146	146	413
Intermolecular	464	555	91	476	106	106	346
Intermolecular	460	490	30	464	104	100	316
RMSD error in $\Delta\omega$					**41**	**51**	**239**
Max. error in $\Delta\omega$					**89**	**113**	**650**

Table 3: Mean deviation (MD), root mean square deviation (RMSD) and maximum deviation from the CCSD(T)/aug-cc-pVTZ(est.) binding energies for 22 sulfate-water clusters with $n = 3-5$ (all in kcal/mol).

Method	MD	RMSD	MAX
B3LYP/6-311++G**	1.058	1.368	2.557
RI-MP2/aug-cc-pVTZ	0.000	0.097	0.309

The analogous benchmark for the B3LYP/6-311++G** vertical detachment energies is shown in Figure 2. A linear regression analysis shows that, overall, the correlation between the B3LYP and the CCSD(T) results is good ($r^2 = 0.937$). On an absolute scale the MD and RMSD of 0.015 and 0.070 eV are small. We note, however, that the biggest deviation is 0.135 eV, which is relatively large compared to the differential effects between the isomers for a given cluster size. Furthermore, the linear regression slope of $b = 0.931$ shows that the fidelity deteriorates with increasing cluster size. We therefore conclude that B3LYP/6-311++G** is suitable to describe the overall trends in the VDEs and yields statistically decent results, although individual outliers are possible.

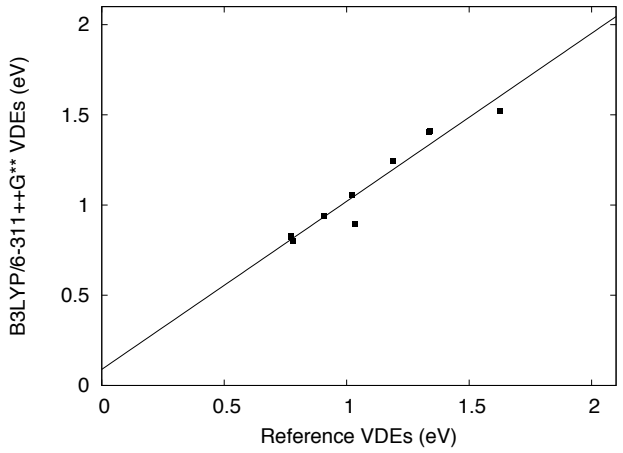


Figure 2: Comparison of vertical detachment energies at the B3LYP/6-311++G** versus CCSD(T)/aug-cc-pVDZ reference. Linear regression to $f(x) = a + bx$ yields $\{a = 0.088, b = 0.931\}$ with a correlation coefficient of $r^2 = 0.937$.

3 Results

3.1 Fidelity of the AMOEBA energy surface

Based on an average AMOEBA-vs.-MP2 error of 2 kcal/mol,[23] we initially applied an energy window of 4.5 kcal/mol to select candidates with ab initio energies of 2.5 kcal/mol and less. While this worked well in many cases and produced several new structures, we found that the AMOEBA energies for some structures differ significantly from the ab initio values. For example, isomer 6-II (6.12.0-1) has an AMOEBA energy of 6.2 kcal/mol, while B3LYP and CCSD(T) predict energies of 3.8 kcal/mol (B3LYP) or 3.4 kcal/mol (CCSD(T)). This overestimation caused 6-II not to show up in our original energy window. The average deviation for the $n = 7$ isomers was again good, but here the outliers were even more pronounced: Several structures had relative energies of 8 kcal/mol and more according to AMOEBA, while the B3LYP energies were on the order of 2 kcal/mol or less. We found that these outliers occurred mostly for structures with many sulfate-water bonds and a small water network. This suggests an imbalance in the description of sulfate-water binding versus water-water binding in the current AMOEBA parametrization. To make sure that we did not miss important low-lying structures, we increased the AMOEBA energy window to 10 kcal/mol.

3.2 Isomer gallery and VDE comparison

We start by first making general observations and investigating the general trends, before having a closer look at some selected structures. Figures Figure 3-Figure 7 show the energetically low-lying isomers that we found using our AMOEBA search procedure. Our nomenclature for the isomers is as follows: Every structure is sorted into a category "*n.s.w*", where *n* is the number of waters, *s* the number of sulfate-water hydrogen bonds and *w* the number of water-water bonds. Within each category the structures are sorted according to increasing electronic energies relative to the global minimum (see sec. Section 3.3 for remarks on vi-

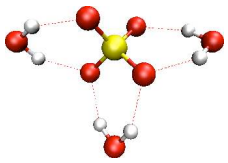
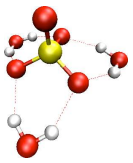
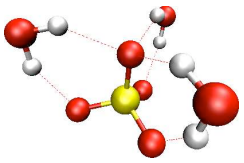
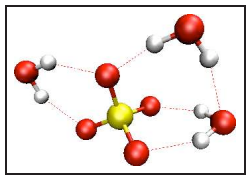
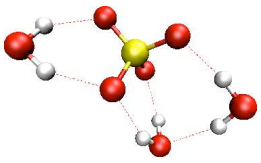
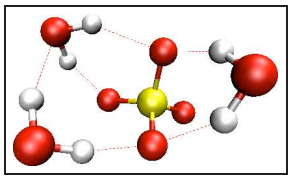
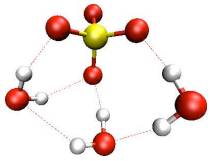
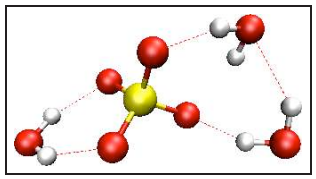
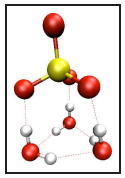
3.6.0			
E_{rel}	0.00	0.07	0.58
VDE	1.058	0.941	0.897
3.5.1			
E_{rel}	1.08	1.58	3.10
VDE	0.826	0.636	0.805
3.4.2			
E_{rel}	3.53		
VDE	0.697		
3.4.1			
E_{rel}	2.86		
VDE	0.814		
3.3.3			
E_{rel}	1.05		
VDE	0.800		

Figure 3: Gallery of low-energy $n = 3$ clusters. In our cluster nomenclature "$n.s.w$" n signifies total number of water molecules, s number of water-sulfate bonds, and w denotes the number of water-water bonds. The number of non-bonded hydrogens can be calculated as $2n - s - w$. Energies below each picture are the relative non-ZPE corrected cluster energies calculated at the B3LYP/6-311++G** level of theory and are given in kcal/mol. VDEs reported in eV.

brational effects), such that we can assign a unique and descriptive name to each isomer. For example, 3.6.0-1 denotes an $n = 3$ cluster, more particular the lowest-lying structure in the class with three water molecules, 6 sulfate-water bonds and no water-water bonds. The maximum value that $s + w$ can assume is $2n$. Since we sorted the structures according to their w value from smallest (top) to largest (bottom), looking at the table of isomers from top to bottom gives an impression of how frequently we find favorable sulfate-water interactions as compared to water-water networks. To make it easier for the reader to identify newly found structures, we have marked them with boxes.

3.2.1 General observations regarding the energy landscape

Already at first glance it becomes apparent that our number of structures exceeds the previously published numbers[8,9,52] by far, which alludes to the fact that the potential energy landscape is much richer than anticipated or explored so far. For fairness we remark that the largest numbers of new structures were found for $n = 7$, which may have played a marginal role in previous work. We nevertheless find many new structures for the other cluster sizes ($n = 3 - 6$) as well.

It is remarkable that, while the previously reported structures comprise the global minima for $n = 3 - 6$, we found at least one additional structure for each n that is populated with a probability of 15% or more at 300K. For $n = 6$ we found one structure (6.7.5-2) that is virtually isoenergetic to the established global minimum (6.7.5-1, or 6-I according to[8]). Even more excitingly, for $n = 7$ we found a new candidate (7.9.5-1) for the global minimum (although we do not claim that the level of theory chosen here generates the correct energetic ordering). Moreover, two more of the new structures (7.9.5-2 and 7.9.5-3) are virtually isoenergetic to the global minimum.

Note that for Figs. Figure 6 and Figure 7 we only show isomers in an energy range of 0-1.1 kcal/mol in order to keep the gallery size manageable. Thus the previously found structures 6-II–6-IV and 7-II–7-V, which are also sampled by our AMOEBA simulations, had to be moved to the supplementary material. For Figs. Figure 3 to Figure 5 we show all minima in an 0-2.5 kcal/mol B3LYP energy window, corresponding to Boltzmann weights of 15.5% and more.

Our calculations show quite impressively how the number of minima grows as the cluster size increases (Section 3.2.1). Already for the 5-water cluster we found on the order of 100 distinct minima and up to 2,100 for the seven-water cluster. Of these, only a small fraction is below our first AMOEBA selection threshold of 4.5 kcal/mol, with only 10 structures for $n = 3$, on the order of 60 structures for $n = 6$, and on the order of 180 for $n = 7$. Compared to the previously published minima, we were able to increase the number of known low-lying minima ($\leq$ 1.1 kcal/mol) by up to one order of magnitude (from 1 to 16). This illustrates the usefulness of our multi-step approach, i.e. starting with a parallel tempering MD approach based on an economical, but on average energetically faithful method such as the AMOEBA force field to preselect a limited number of energetically low-lying starting structures.

Furthermore, the number of very low-lying structures is small compared to the number of initial structures. This alludes to the fact that it may be possible to rationalize the favorable structures to a certain degree using chemically intuitive criteria, for example using the notion of competing effects such as binding via electrostatic versus charge transfer effects, as has been done before[8] and will be discussed below. Nevertheless, this knowledge can only applied *a posteriori* and does avoid the local minima search problem, which remains an NP-hard problem for solvation clusters. Our results illustrate the importance of statistical effects in the potential energy landscape exploration even for such small clusters.

3.2.2 Chemical binding motifs

General trends. Due to the composition of the system one can trivially distinguish between two basic types of bonds: sulfate-water and water-water bonds. Among these, a finer distinction can be made on the basis of whether they are exclusive, or shared between two or even three binding partners. For example, we observe that every sulfate oxygen can bind (i) one, (ii) two, or (iii)

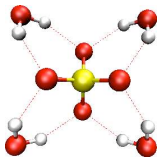
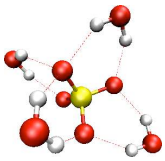
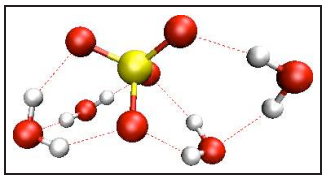
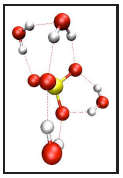
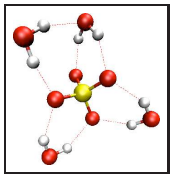
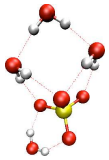
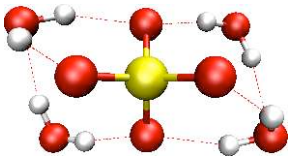
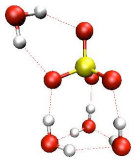
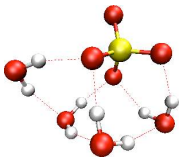
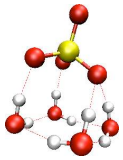
4.8.0			
E_{rel}	0.15	0.33	2.06
VDE	1.587	1.523	1.324
4.7.1			
E_{rel}	0.33	0.38	
VDE	1.407	1.409	
4.6.2			
E_{rel}	1.16	1.76	
VDE	1.335	1.320	
4.5.3			
E_{rel}	0.00	2.54	
VDE	1.245	1.228	
4.4.4			
E_{rel}	2.09		
VDE	1.059		

Figure 4: Gallery of low-energy $n = 4$ clusters. framebox figures show new, previously undiscovered structures. Energies below each picture are the relative non-ZPE corrected cluster energies calculated at the B3LYP/6-311++G** level of theory. Relative energies are reported in kcal/mol, VDEs in eV.

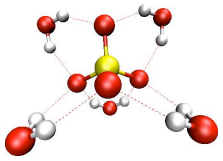
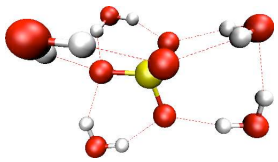
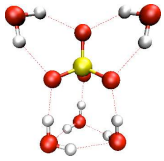
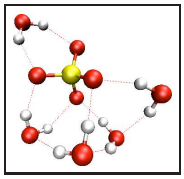
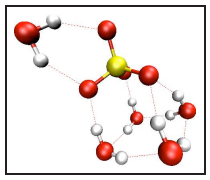
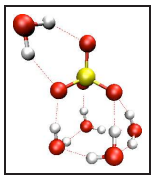
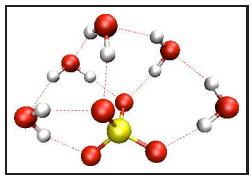
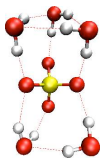
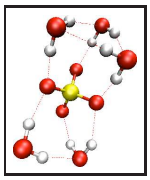
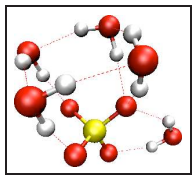
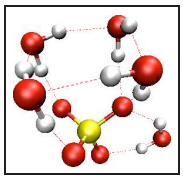
5.10.0			
E_{rel}	1.56		
VDE	1.608		
5.9.1			
E_{rel}	2.11		
VDE	1.693		
5.7.3			
E_{rel}	0.00	2.09	
VDE	1.802	1.573	
5.6.4			
E_{rel}	1.13	1.16	1.96
VDE	1.963	2.133	1.704
			
E_{rel}	1.75	1.76	2.03
VDE	1.716	1.802	1.586
			
E_{rel}	1.62	1.63	
VDE	1.606	1.720	

Figure 5: Gallery of low-energy $n = 5$ clusters. framebox figures show new, previously undiscovered structures. Energies below each picture are the relative non-ZPE corrected cluster energies calculated at the B3LYP/6-311++G** level of theory. Relative energies are reported in kcal/mol, VDEs in eV.

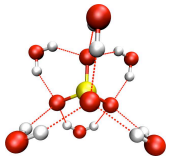
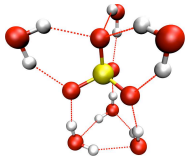
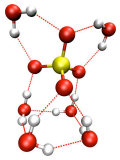
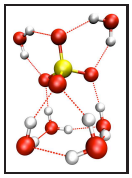
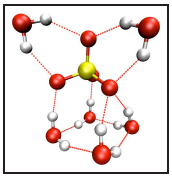
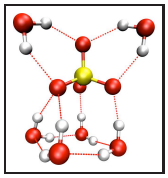
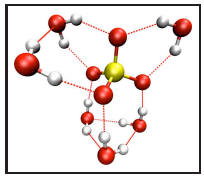
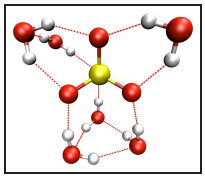
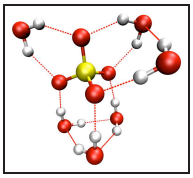
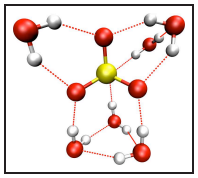
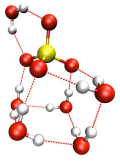
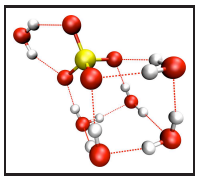
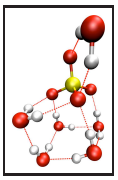
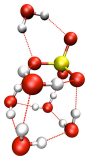
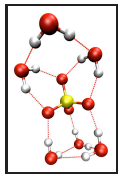
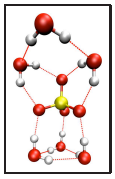
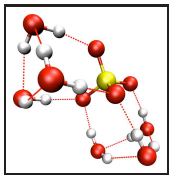
6.12.0				
E_{rel}	3.83			
VDE	2.867			
6.9.3				
E_{rel}	0.24			
VDE	2.292			
6.8.4				
E_{rel}	0.26	0.29	0.76	0.79
VDE	2.174	2.195	2.147	2.155
				
E_{rel}	0.97	0.99	1.05	1.05
VDE	2.147		2.253	2.242
6.7.5				
E_{rel}	0.00	0.07	0.48	0.49
VDE	2.017	2.111	2.008	2.026
				
E_{rel}	0.64	0.73		
VDE	2.151	2.149		
6.6.6				
E_{rel}	0.51	0.53		
VDE	2.060	2.068		

Figure 6: Gallery of low-energy $n = 6$ clusters. framebox figures show new, previously undiscovered structures. For space reasons we list only structures below 1.05 kcal/mol as well 6.12.0-1 for comparability with the literature. Energies below each picture are the relative non-ZPE corrected cluster energies calculated at the B3LYP/6-311++G** level of theory. Relative energies are reported in kcal/mol, VDEs in eV.

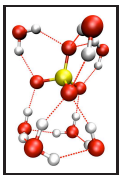
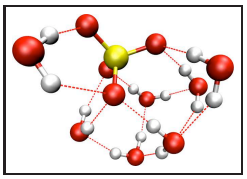
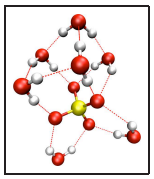
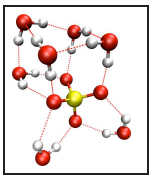
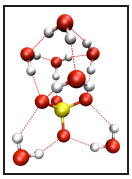
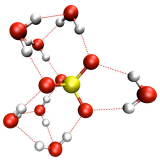
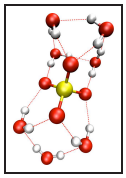
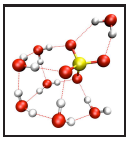
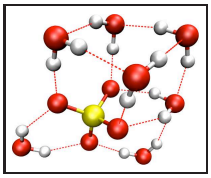
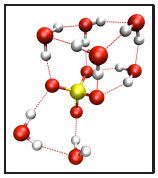
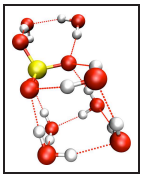
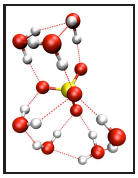
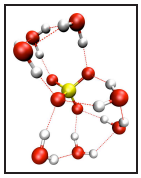
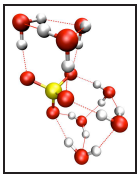
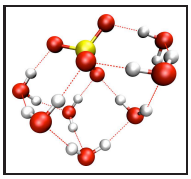
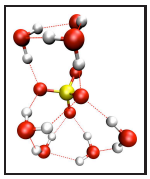
7.10.4				
E_{rel}	1.02			
VDE				
7.9.5				
E_{rel}	0.00	0.03	0.07	0.40
VDE	2.763	2.737	2.723	2.720
7.8.6				
E_{rel}	0.03	0.48	0.65	
VDE	2.665	2.614	2.518	
				
E_{rel}	0.66	0.96	1.03	
VDE	2.623	2.588	2.585	
7.7.7				
E_{rel}	0.23	0.29	0.30	
VDE	2.769	2.776	2.556	
7.6.7				
E_{rel}	0.21	0.25		
VDE	2.446	2.444		

Figure 7: Gallery of low-energy $n = 7$ clusters. framebox figures show new, previously undiscovered structures. Due to the multitude of new structures, we can only show the ones with energies below 1 kcal/mol here. Other structures are given in the supplementary material. Energies below each picture are the relative non-ZPE corrected cluster energies calculated at the B3LYP/6-311++G** level of theory. Relative energies are reported in kcal/mol, VDEs in eV.

Table 4: Number of distinct local minima: (1.) in total, as generated by our AMOEBA simulations, (2.) within a 4 kcal/mol energy window from the global minimum, as ranked by AMOEBA, and (3.) lower than 1.1 kcal/mol, as determined at the B3LYP/6-311++G** level (this work). For comparison we also show (4.) the number of low-lying minima from the literature[8,9,52] as ranked using our B3LYP energies. Note that for $n = 5$ we report both, the ZPE uncorrected and the ZPE-corrected numbers. For all other isomers, ZPE-uncorrected values are reported.

	Number of isomers			
	AMOEBA		B3LYP	Lit.
n	total	$\leq$ 4.5 kcal/mol	$\leq$ 1.1 kcal/mol	$\leq$ 1.1 kcal/mol
3	10	10	5	3
4	40	10	5	3
5	100	12	2 (3)	2 (3)
6	800	60	17	5
7	2,100	180	16	1

three water molecules. The observed maximum of three bonds per sulfate oxygen arises probably because of sterical reasons – the vdW spheres for the hydrogens in (iii) are virtually touching each other. Examples for types (i) and (ii) can be found in structure 3.6.0-1, where the "left" and "right" waters have one exclusive and one shared bond each, and the "central" water has two shared bond. An example for binding to three waters (iii) can be found in structure 3.6.0-2.

In our simulations we found water acting as a hydrogen donor (electron acceptor) in up to two bonds and as a hydrogen acceptor (electron donor) typically in up to one bond, such that the highest coordination number we observe is three. Thus we can characterize the water-water bonds by the number of hydrogen bonds donated (D) or accepted (A) as D, DD, AD and ADD (other motifs were not observed here). As n increases, we find an increasing tendency for water-water networks to form and a trend from lower to higher coordination numbers for the water, as is also observed for pure water clusters.[53]

The interplay of binding motifs (i)-(iii) and A-ADD gives rise to a multitude of structures. However, it is possible to isolate certain typical motifs. We first note that there are structures that are (1) exclusively sulfate-water bound, such as in the 3.6.0, 4.8.0, 5.10.0 and 6.12.0 classes. For $n = 7$ the highest sulfate coordination number is 12 because of the sterical crowding of the water ligands described above. Another recurring motif is that of a (2) water dimer that is bound to the sulfate, as shown e.g. in 3.5.1-1. Here one of the water molecules forms two hydrogen bonds to sulfate, whereas the other water is connected to the sulfate via one hydrogen end and to the first water with the other hydrogen. We note that the geometry of the "sulfate-bound water dimer" is changed in comparison to the global minimum of the free dimer in that the hydrogen atoms stand *cis* rather than *trans*, which is necessary to allow formation of three sulfate-water bonds. Furthermore, we find (3c) "cyclic" water trimers, i.e. monodromic three-membered water rings as in 3.3.3-1 and (3o) "open" water trimers as in 3.4.2-1. In the cyclic trimer the water molecules are arranged in a ring whith one hydrogen on each water bound to the sulfate. The next binding motif is a (4c) cyclic water tetramer, i.e. monodromic four-membered water rings such as in 4.4.4-1. Similarly to the three-membered rings one set of hydrogen atoms is arranged in a ring and the other four hydrogens are bound to the sulfate. However, due to the larger number of water ligands now two water molecules have to share a bond to a sulfate oxygen. Also (4o) open water tetramers are found, such as is 5.7.3-1. Similarly, cyclic (5c) and open (5o) pentamers as well as hexamers (6c and 6o) are found. Especially for the larger water networks several other variations are possible, where parts of the network show a cyclic and other an open structure, and by merging motifs (1)-(6). For example, 7.8.6-4 exhibits a (1) + (2)$\cup$(4c) structure, where (2)$\cup$(4c) denotes a

cyclic tetramer merged with a dimer. In the larger clusters ($n \geq 6$) low-lying structures arise with water in the second solvation shell, which gives rise to additional binding motifs.

For the small $n = 3,4$ clusters the gallery is "top-heavy", i.e. we find many energetically favorable structures with almost exclusive sulfate-water binding and very few favorable water-water binding motifs. Starting with $n = 5$, the largest class of structures is the 5.6.4 group of 8 relatively low-lying structures with 4 water-water and 6 sulfate-water bonds, whereas the class with exclusive sulfate-water binding comprises only one structure (5.10.0-1), albeit the global minimum. The relative stability of 5.10.0-1 stems from the fact that the number of waters does not yet allow the formation of a water-water network that can compete with the 2×(ii) + 2×(iii) sulfate-water binding motif, which will be discussed in detail below. Our results thus support previous findings[8] that the structures are mainly due to two competing effects, namely the competition between sulfate-water and water-water binding. A more detailed discussion will follow in section Section 3.2.2, where we discuss individual binding motifs and their stabilities.

Starting from $n = 6$, we find only one structure (6.12.0-1) with exclusive sulfate-water binding (4× (iii)) which is already outside (3.8 kcal/mol) our energy window, while the majority of low-lying structures have more extended water networks. This can be rationalized by noting that sharing the sulfate-water bond between two or more waters seems to destabilize the bond such that alternative water-water interactions become more favorable, and accommodating more than three waters is probably impossible because of sterical reasons. Not surprisingly, we find very stable structures amongst the structure classes in which the hydrogen bonds are distributed such that there is sufficient sulfate-water bonding and the rest of the bonds are formed in the water network such as in e.g. the 6.9.3, 6.8.4 and 6.7.5 classes.

For the $n = 7$ clusters the largest number of sulfate-water bonds is 12, which is found only in one energetically unfavorable structure (7.12.2-1), that again is outside our energy window (4.7 kcal/mol). The majority of low-lying structures again have more extended water networks such as in the 7.9.5, 7.8.6, 7.7.7 and 7.6.7 classes. $n = 7$ is special in that some of the low-lying structures (7.6.7) have uncoordinated hydrogens and in the large fraction of low-lying structures with water in the second solvation shell.

In the earlier literature it has been advocated that the first solvation shell closes at $n = 6$,[10–18] whereas more recent investigations all support a solvation shell size of $n = 12$.[6,19,20] For $n = 6$ our data shows an interesting feature: Many of the lowest-lying structures (6.7.5-1 – 6.7.5-3 and 6.7.5-5 – 6.7.5-6 at 0.0 – 0.7 kcal/mol) already have a water in a second shell. Also for $n = 7$ one finds many low-lying structures with a second-shell water, e.g. 7.9.5-1 – 7.9.5-4 (0.0 – 0.4 kcal/mol). However, one can still find some very low-lying isomers in which all water molecules are arranged in a first solvation shell, e.g. 7.8.6-1 (0.0 kcal/mol) or 7.7.7-1 – 7.7.7-3 (0.2 – 0.3 kcal/mol). This suggests that the first solvation shell is not fully closed around $n = 6-7$, but in order to determine the size of the solvation shell one would have to go to larger n. Our data also shows that the nature of the solvation shell changes from predominantly sulfate-bound to a *coexistence* of sulfate- and water-bound structures in the $n = 6-7$ range.

We next compare our observations of the cluster structures with the distribution of relative energies (Figure 8). Several trends can be observed: (1.) The density of isomers per energy interval is much larger for the high-lying structures than in the low-lying regime. (2.) For the larger clusters the relative energies are spread out relatively evenly, such as in $n = 6-7$, whereas for the smaller sizes ($n = 3-5$) relatively large gaps are found. (3.) The cluster energies as functions of size seem to follow a shell or band structure, where the energies for the high-lying isomers decrease in the $n = 3-6$ series. Interestingly, this trend seems to stagnate for $n = 6-7$.

The fact (1.) that the number of high-lying isomers grows more rapidly than for the low-lying ones can be understood based on purely statistical arguments: While there are relatively few energetically favorable binding motifs, there are many possible arrangements of the solvent shell that lead to relatively unfavorable binding energies. We will revisit this point when we discuss individual binding motifs (section Section 3.2.2). For the larger clusters there are more possible combinations of

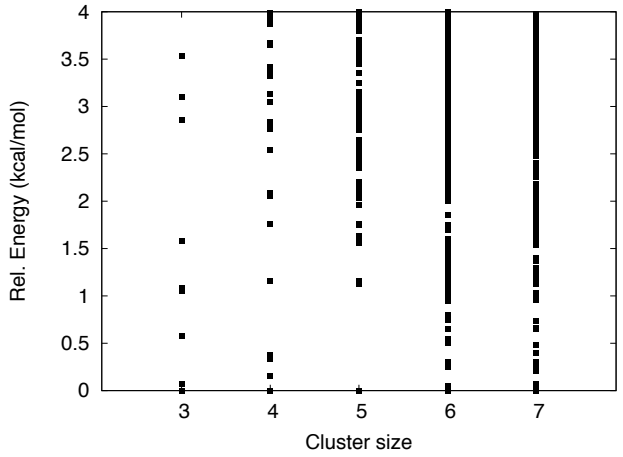


Figure 8: Distribution of relative energies for the $n = 3-7$ clusters.

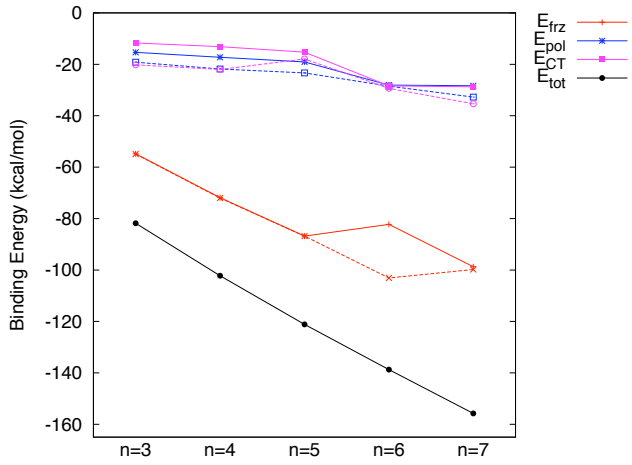
binding motifs, which explains (2.) that in these the energetic distribution can be more uniform. The band structure (3.) of the energies can be understood by a very similar argument, i.e. recalling that there is only a limited number of energetically favorable binding motifs which have a typical energy range. This gives rise to an energetic band structure. Perturbing these motifs slightly, e.g. by changing the position of just one water ligand, a small energy perturbation is introduced that gives the energy band a certain width. From the observation that the band structure trends in the $n = 3-5$ and $n = 6-7$ subseries are different we conclude that a qualitative change in the structure of the solvation shell occurs between $n = 5$ and $n = 6$. This is in accordance with the earlier observation that direct sulfate-water bonding is saturated once 12 sulfate-water bonds have been formed and thus there is a shift from a dominance of sulfate-water structures towards larger water networks from $n = 6$ on.

Interesting trends can also be extracted from the plot of vertical detachment energies versus cluster size (see supplementary material). We observe that the VDEs increase almost linearly in the $n = 3-5$ series, whereas the slope changes around $n = 6,7$. It is not surprising that the slope is steeper for the smaller clusters and smaller for the larger clusters: As we saw previously, the number of sulfate-water bonds reaches saturation around $n = 5-6$, thus also the direct stabilization of the sulfate excess electron reaches a maximum around $n = 6$. The additional water-water bonds may stabilize the excess electron on the sulfate indirectly (e.g. through charge transfer within the water network), but this effect is only secondary since the charge transfer decays quickly with distance.

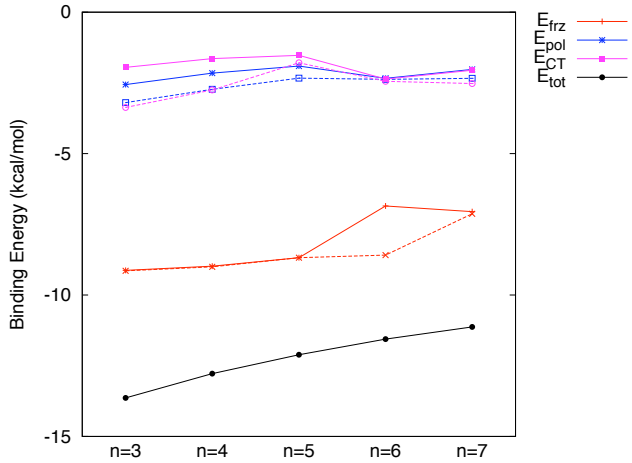
We now turn to the question of *why* the binding motifs change when going from small to large cluster size. One reason that we have already discussed is sterical crowding of the water ligands. Other qualitative arguments were given previously[8] based mostly on electrostatic effects. Using energy decomposition analysis (EDA) we can gain more refined insight by analyzing the trends quantitatively in terms of frozen-fragment (E_{frz}), polarization (E_{pol}) and charge-transfer (E_{CT}) effects, which add up to the total interaction energy (E_{tot}). Figure 9 (a) shows the different components of the binding energy as a function of cluster size for the lowest-lying isomer (solid lines) and for the cluster with the largest energy component in each category (dashed lines).

We observe that the trends in the frozen-fragment contributions change around $n = 6$. While for $n = 3-5$ the E_{frz} component is identical for the lowest-lying isomer and the isomer with the biggest frozen-fragment component (because they are, in fact, identical), the two lines split up for $n = 6$. While the $n = 6$ isomer with the largest frozen-fragment contribution (6.12.0-1) essentially follows the trend of the $n = 3-5$ clusters, the lowest-lying isomer (6.7.5-1) has significantly smaller frozen-fragment contributions. For $n = 6-7$ the maximum frozen-fragment contribution stagnates, because the sulfate-water coordination sphere is saturated and thus only 12 direct sulfate-water bonds can be formed (7.12.2-1). Interestingly, the E_{frz} component for the lowest-lying $n = 7$ isomer (7.9.5-1) is only 1.1 kcal/mol higher, such that the dotted and solid lines get closer again. At the same time we observe a monotonic increase in the polarization and CT contributions in the $n = 3-7$ series: While for $n = 3,4$ the more favorable structures show stronger electrostatic binding, the larger clusters exhibit structures with stronger polarization and charge transfer components. This is in accordance with the finding that in the larger clusters the formation of water-water networks is favored, which contain less electrostatic interactions with the sulfate and allow for more polarization and charge

transfer. The reader is referred to the supplement for some selected EDA results, which show that the polarization and charge transfer component is especially large in the clusters with extensive water-water networks.



(a) Absolute binding energies.



(b) Binding energies per hydrogen bond.

Figure 9: Binding energies vs. cluster size (kcal/mol), (a) on an absolute scale, (b) divided by the number of water units. For each cluster size we show the binding energies of the lowest-lying isomer (solid lines) and the biggest binding energy components (dotted lines).

Another interesting trend to investigate is the onset and the consequences of cooperative effects. To this end we look at the binding energy averaged over the number of hydrogen bonds and plotted as a function of cluster size, as shown in Figure 9 (b). The magnitude of the total binding energy per hydrogen bond decreases as the cluster size increases from -13.6 kcal/mol ($n = 3$) to -11.1 kcal/mol per hydrogen bond ($n = 7$). Apparently, cooperative effects decrease the binding energy per solvent molecule as the cluster grows. This is remarkable insofar as in pure water clusters the binding energy per unit grows, i.e. the cooperative effects are stabilizing. For example, on average each hydrogen bond within the water dimer contributes 4.6 kcal/mol, in the trimer 4.8 kcal/mol, tetramer 6.5 kcal/mol, and 6.9 kcal/mol in the pentamer (all numbers at the B3LYP/aug-cc-pVQZ level).[53]

Where does this decrease in per-bond binding energy come from? Clearly, the biggest effects are seen in the frozen-fragment electrostatic interactions, whose stabilizing effect becomes weaker as the cluster size increases. Polarization and charge-transfer energies per hydrogen bond become less favorable in the row from $n = 3$ to $n = 5$, then jump back to stronger stabilization for $n = 6$ and $n = 7$. We can thus conclude that the main reason for the destabilizing collective effects lies in less favorable frozen-fragment interactions. This is similar to the findings for pure water clusters, where the binding for larger clusters is dominated by polarization and, to a lesser extent, charge transfer, whereas the frozen-fragment interaction becomes even repulsive. Now for the sulfate-water clusters we clearly see a superposition of two effects: For the small clusters there is very favorable frozen-fragment interaction due to the negative charge on sulfate, which makes the E_{frz} much more favorable than in pure water. As the water network becomes larger, the polarization and charge-transfer interactions become more favorable as in pure water at the cost of less favorable, although still dominating, electrostatic interactions. The reason we see a decrease in per-bond binding energy is thus because the initially strongly stabilizing frozen-fragment interaction between sulfate and the water ligands cannot "scale" with cluster size. To understand this scaling effect better, we will investigate it further in the next section.

Specific binding motifs. In order to better understand the overall trends in the binding motifs, it is helpful to investigate individual motifs and their energetics. For that purpose the five binding motifs depicted in Figure 10 are most informative. In the top table we show the binding energy

E_{frz}	-54.9	-50.0	-50.2	-50.9	-65.8
E_{pol}	-15.1	-15.7	-15.8	-24.0	-25.8
E_{CT}	-11.6	-14.1	-13.9	-24.2	-26.6
E_{tot}	-81.6	-79.8	-79.9	-99.1	-118.3

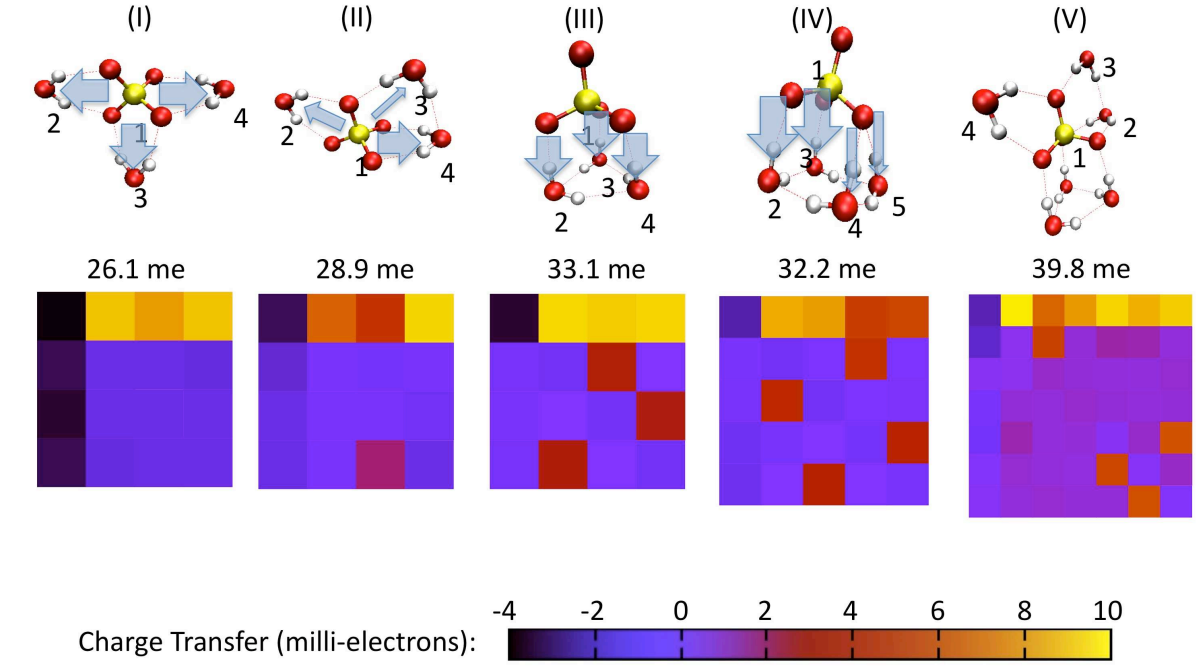


Figure 10: Energy decomposition analysis for five representative binding motifs. The table shows the EDA components at the B3LYP/6-311++G** level (in kcal/mol). Below we show the structures of the selected binding motifs as well as the numbering for the ALMO fragments. The color-coded matrices in the bottom indicate the charge transfer (measured in milli-electrons, me) between the fragments, i.e. the element in row i and column j denotes the charge donation from fragment i to j. The fragment numbers are given in the molecular graphs, and the thickness of the arrows is proportional to the charge transfer from the sulfate to the water ligands.

components according to ALMO/EDA,[25] and in the matrices below we display the charge-transfer contributions from fragment i (row) to fragment j (column). Thus row 1 is associated with sulfate $\rightarrow$ water charge donation, column 1 with water $\rightarrow$ sulfate back-donation, and $(i = 2 \ldots n,\ j = 2 \ldots n)$ with the water-water network. Diagonal elements can be interpreted as intramolecular polarization accompanying the charge transfer.

The most simple binding motif is shown on the left (I): All water molecules are bound to the sulfate via two hydrogen bonds. As the EDA shows, this motif is optimal with respect to frozen-fragment interactions (-9.2 kcal/mol per H bond), whereas the polarization and CT contributions are only -2.5 and -1.9 kcal/mol per bond. Next (II) we consider a structure where two waters form a hydrogen bond; note that here the CT component becomes more stabilizing by 0.5 kcal/mol per H bond. In the central structure (III) we show a three-membered monodromic ring, where the CT component is similarly large. We then note that four-membered monodromic rings (IV) are not very favorable according to the frozen-fragment interaction (-6.4 kcal/mol per bond on average), but that the polarization and charge transfer become more favorable at -3.0 and -3.0 kcal/mol per H bond. This more detailed comparison thus supports earlier findings that larger water-water networks are favorable for polarization and charge transfer, and frozen-fragment interactions are favored when direct sulfate-water bonds can be formed.

A further inspection of the EDA components reveals that the binding in sulfate-water is quite different from that in pure water clusters (see comment above about collective effects). While in pure water the main contribution to the stabilization of water-water network formation comes from polarization with a contribution from charge transfer, the binding in sulfate-water is donimated by E_{frz}. Also, charge transfer and polarization components are typically equally large in sulfate-water. For the clusters investigated here the electrostatic frozen-fragment component is typically largest, but E_{pol} and E_{CT} grow faster as the cluster size increases. For very large clusters we therefore expect to asymptotically approach the pure water findings, except for a solvation region around the sulfate, which distorts the water-water network locally.

The different role of charge transfer in sulfate-water as compared to pure water is probably influenced by the excess charge on the sulfate. As denoted by the arrows in Figure 10, sulfate transfers charge to the water ligands. Interestingly, the extent of charge transfer depends on the binding motif. For exclusive sulfate-water bonds the transfer amounts to an average of 4.4 me per H bond (I), and the donation into each water molecule is roughly equally strong. When water-water bonds are formed, however, the situation changes drastically (II). The charge transfer into the fragment 4 remains strong, but for the purely sulfate-bound fragment 2 it is roughly halved. Also the charge transfer into the singly bound fragment 3 is reduced. We thus conclude that sulfate prefers to donate electron density into water ligands that act as electron donors in a water-water network. This can be understood by considering that fragment 4 donates some charge to fragment 3, i.e. the water-water H bond indirectly "pulls" electron density from the sulfate. It is furthermore interesting that the presence of water-water networks reduces the charge donated into exclusively sulfate-bound waters, i.e. the different water ligands enter a competition for a fraction of the sulfate excess electron density.

For the monodromic 3-ring (III) we again find strong charge donation from the sulfate into the ring, where each water ligand receives on the order of 11 me. It is interesting to compare this to the monodromic 4-ring (IV), where two ligands (4 and 5) have to share a sulfate oxygen for steric reasons. The charge transfer into IV's waters 4 and 5 is approximately halved compared to that of waters 2 and 3 or the waters in (III). We thus conclude that shared H bonds are unfavorable with respect to charge transfer, which is again in agreement with earlier findings.

In structure (V) we show how three binding motifs compete for charge transfer and find that electron donation into ligands like 2 or 5-7 that are themselves donating charge into a water network is again increased, whereas the donation into the single water ligand is again decreased. This is a trend we observed for all our structures: Water networks can apparently stabilize the additional charge very well, whereas single water molecules can do this to

a much lesser extent. Of the simple binding motifs, three-membered rings are apparently best for the charge transfer, followed by four-membered rings (due to the sharing of hydrogen bonds). It is furthermore interesting to observe how the sulfate increases the charge transfer within the water-water networks compared to that in the water dimer (2.4-2.8 me[26]) and that this effect seems to become stronger as the size of the water network increases.

3.3 Anharmonic effects

The effects of including (1.) harmonic and (2.) anharmonic ZPEs in the relative isomer energies for $n = 3,4,6$ are shown in Figure 11 (a)-(c). As already discussed in the literature, ZPEs can change the relative energetic ordering of the isomers and are thus important to include in accurate calculations. Our results demonstrate this clearly for all cluster sizes investigated. We even find some more surprising results: The influence of anharmonic effects on the relative energies is often bigger than that of the harmonic corrections. This means that the differential effects when calculating relative energies at the anharmonic level of theory are bigger than at the harmonic level. Since our anharmonic calculations take up to fourth-order nuclear derivatives into account, they are more sensitive to the shape of the potential energy surface and thus to the individual cluster geometries. Harmonic calculations, on the other hand, take only up to second derivatives into account and thus depend more on the local properties of the potential energy surface, which leads to stronger cancellation of vibration effects when relative energies are calculated.

At this point it is striking that including harmonic ZPEs may be as adequate, or inadequate, as sticking to purely electronic energies. In fact, in several cases the lowest-lying isomer as predicted by the purely electronic energy is equal to the lowest-lying isomer after anharmonic correction, while the harmonic values differ. We do not claim here to have obtained the correct ranking, as this would require high-level electronic structure theory calculations and a more sophisticated treatment of vibrational effects. Our results, however, demonstrate how important the inclusion of anharmonic effects is in obtaining a reliable energetic

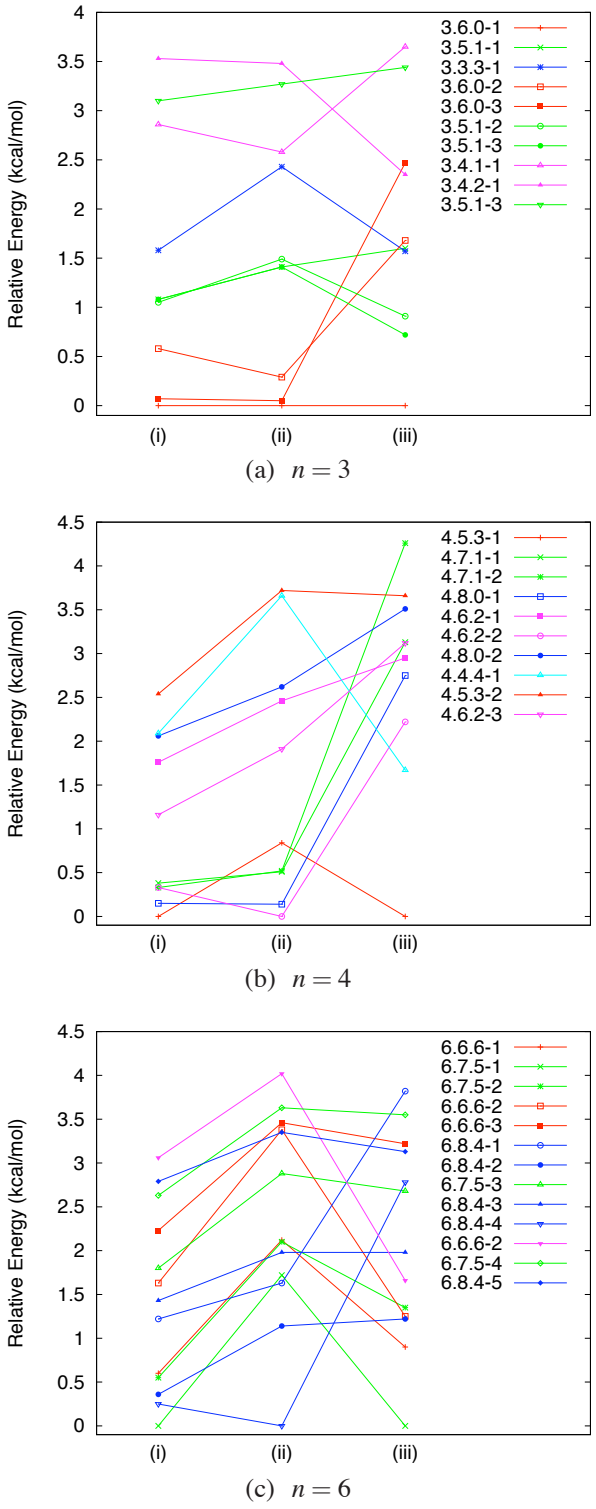


Figure 11: Relative isomer energies for $n = 3,4,6$: (i) = electronic energies, (ii) = including harmonic ZPE, (iii) = including anharmonic corrections (all in kcal/mol).

ranking.

It is interesting to investigate the effects that harmonic and anharmonic corrections have on different types of clusters. For easier comparison we have chose the same colors for structures of the same class within each subfigure (Figure 11). Interestingly, the structures with more sulfate-water bonds are typically favored by harmonically corrected binding energies. In contrast, the anharmonic corrections typically favor structures with extended water-water networks.

4 Conclusion

We re-visited sulfate-water clusters, where we investigated new aspects or extended previous investigations. In particular, we presented a joint molecular mechanics/first-principles procedure, which was used to perform exhaustive searches of the potential energy surface for the $SO_4^{2-}(H_2O)_{n=3-7}$ clusters. We found a multitude of new structures many of which are significantly populated at room temperature. For $n = 6$ as well as $n = 7$ we found several new structures that are virtually isoenergetic with the global minimum. Furthermore, for $n = 7$ we found a new candidate for the global minimum. This illustrates the usefulness of our search strategy. At the same time the sheer overwhelming number of 2,100 and more possible structures for a still rather small cluster like $n = 7$ clearly demonstrates that even at this size a molecular dynamics/statistical description of the system may be both, more viable and appropriate. Some of us have recently presented a promising MD technique for photoelectron spectra as well as IR spectra simulation[54,55] based on quasiclassical trajectory molecular dynamics (QCT-MD).[56–58] A possible extension of this work would therefore be the QCT-MD simulation of sulfate-water clusters, for example to obtain complete photoelectron or IR spectra. Nevertheless, examination of the numerous low-energy structures yields insights into different energetically competitive binding motifs.

Using energy decomposition analysis we investigated the different energy components that contribute to binding in sulfate-water and water-water hydrogen bonds and found that, as discussed qualitatively previously,[8] binding to sulfate is dominated by electrostatics in the smaller clusters. For growing cluster sizes we increasingly find structures with extensive water-water networks, i.e. the competition between sulfate-water and water-water networks is shifted in favor of water-water bonds. We found a rather remarkable effect to be responsible for this: EDA results clearly show that collective effects for pure water become more favorable as the water network grows,[53] but that sulfate-water bonds become weaker. The favorable water network is thus not only due to stabilizing collective effects in the water network, but is also due to destabilizing collective effects in the sulfate-water interaction. There are probably several reasons for the destabilization of the sulfate-water bond. One is that for larger clusters more and more water molecules have to share hydrogen bonds to the sulfate, which, on the average, become weaker.

In the water network attached to the sulfate we also noted an important difference with respect to pure water: The charge transfer among the water molecules can be increased compared to pure water. This is probably due to the charge the sulfate injects into the water network. For future research it would be interesting to test over which distances the sulfate can influence the water network or how other types of anions influence the charge transfer in water, especially in connection with earlier findings that the "sulfate ion patterns water at long distance".[59] The EDA results were obtained with the B3LYP functionals, and, as discussed elsewhere[25,26,55] it may slightly overestimate charge transfer.

Finally we gave an assessment of anharmonic effects for the clusters up to $n = 6$ and showed that harmonic ZPE corrections are not sufficient for a reliable energetic ranking of the clusters. This, along with the findings regarding the exploration of the potential energy surface, demonstrates that an MD treatment may be more appropriate even for small clusters, since it incorporates anharmonic effects naturally and deals appropriately with statistics.

References

(1) Hofmeister, F. *Arch. Exp. Pathol. Pharmacol.* **1888**, *24*, 247–260.

(2) O'Brien, J. T.; Prell, J.; Bush, M.; Williams, E. R. *J. Am. Chem. Soc.* **2010**, *132*, 8248–8249.

(3) Moore, G.; Kellerman, K. *Bull. Bur. Plant Ind. USDA* **1905**, *76*, 19–55.

(4) Mehring, A. L. In *Fertilizers*; Blanck, F. C., Ed., 2nd ed.; Reinhold, New York, 1955; pp 89–140.

(5) G.Y. Stokes and A.M. Buchbinder and J.M. Gibbs-Davis and K.A. Scheidt and F.M. Geiger, *Vibrational Spectroscopy* **2009**, *50*, 86 – 98.

(6) Moore, D.; Wöste, L.; Meijer, G.; Neumark, D. *J. Chem. Phys.* **2006**, *125*, 111102.

(7) Bush, M. F.; Saykally, R.; Williams, E. R. *J. Am. Chem. Soc* **2007**, *129*, 2220–2221.

(8) Wang, X.; Sergeeva, A.; Yang, J.; Xing, X.; Boldyrev, A.; Wang, L. *J. Phys. Chem. A* **2009**, *113*, 5567.

(9) Gao, B.; Liu, Z. *J. Chem. Phys.* **2005**, *123*, 224302.

(10) Ohtaki, H.; Radnai, T. *Chem. Rev.* **1993**, *93*, 1157–1204.

(11) Caminiti, R. *Chem. Phys. Lett.* **1982**, *88*, 103.

(12) Musinu, A.; Paschina, G.; Piccaluga, G.; Magini, M. J, *Appl. Crystallogr.* **1982**, *15*, 621.

(13) Caminiti, R.; Paschina, G.; Pinna, G.; Magini, M. *Chem. Phys. Lett.* **1979**, *64*, 39.

(14) Caminiti, R. *Chem. Phys. Lett.* **1982**, *86*, 214.

(15) Caminiti, R.; Paschina, G. *Chem. Phys. Lett.* **1981**, *82*, 487.

(16) Licheri, G.; Piccaluga, G.; Pinna, G. *Appl. Crystallogr.* **1973**, *6*, 392.

(17) Caminiti, R.; Johansson, G. *Acta Chem. Scand.* **1981**, *A35*, 373.

(18) Cannon, W. R.; Pettitt, B. M.; McCammon, J. A. *J. Phys. Chem* **1994**, *98*, 6225–6230.

(19) Wang, X.-B.; Yang, X.; Nicholas, J. B.; Wang, L.-S. *Science* **2001**, *294*, 1322–1325.

(20) Jungwirth, P.; Curtis, J. E.; Tobias, D. J. *Chem. Phys. Lett.* **2003**, *367*, 704–710.

(21) Ren, P.; Ponder, J. *J. Comput. Chem.* **2002**, *23*, 1497–1506.

(22) Ponder, J.; Case, D. *Adv. Prot. Chem.* **2003**, *66*, 27–86.

(23) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; Distasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.

(24) Miller, Y.; Chaban, G.; Zhou, J.; Asmis, K.; Neumark, D.; Gerber, R. *J. Chem. Phys.* **2007**, *127*, 094305.

(25) Khaliullin, R. Z.; Bell, A. T.; Head-Gordon, M. *J. Chem. Phys.* **2008**, *128*, 184112.

(26) Khaliullin, R. Z.; Bell, A. T.; Head-Gordon, M. *Chem.-Eur. J.* **1009**, *15*, 851–855.

(27) Khaliullin, R.; Head-Gordon, M.; Bell, A. *J. Chem. Phys.* **2006**, *124*, 204105.

(28) Ren, P.; Ponder, J. *J. Phys. Chem. B* **2003**, *107*, 5933–5947.

(29) Rasmussen, T.; Ren, P.; Ponder, J.; Jensen, F. *Int. J. Quant. Chem.* **2007**, *107*, 1390–1395.

(30) Wu, C.; Ponder, J. W. Personal communication.

(31) Stone, A. *Chem. Phys. Lett.* **1981**, *83*, 233–239.

(32) Stone, A.; Alderton, M. *Mol. Phys.* **1985**, *56*, 1047–1064.

(33) Stone, A. J.; Dullweber, A.; Hodges, M. P.; Popelier, P. L. A.; Wales, D. J. *ORIENT 3.2: A program for studying interactions between molecules*; University of Cambridge, 1995-6.

(34) Frisch, M. J. et al. *Gaussian 03, Revision C.02*, Gaussian, Inc., Wallingford, CT, 2004.

(35) Becke, A. D. *J. Chem. Phys.* **1999**, *98*, 5648.

(36) Lee, C.; Yang, W.; Parr, R. *Phys. Rev. B* **1988**, *37*, 785–789.

(37) Vosko, S.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200–1211.

(38) Krishnan, R.; Binkley, J.; Seeger, R.; Pople, J. *J. Chem. Phys.* **1980**, *72*, 650.

(39) Dunning, T. *J. Chem. Phys.* **1989**, *90*, 1007.

(40) Kendall, R.; Dunning, T.; Harrison, R. *J. Chem. Phys.* **1992**, *96*, 6796.

(41) Weigend, F.; Köhn, A.; Hättig, C. *J. Chem. Phys.* **2002**, *116*, 3175.

(42) Shao, Y. et al. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172–3191.

(43) Whitehead, R.; Handy, N. C. *J. Mol. Spec.* **1975**, *55*, 356–373.

(44) Adel, A.; Dennison, D. *Phys. Rev.* **1933**, *43*, 716–723.

(45) Wilson, E.; Howard, J. *J. Chem. Phys.* **1936**, *4*, 260.

(46) Lin, C. Y.; Gilbert, A.; Gill, P. *Theor. Chem. Account* **2008**, *120*, 23–35.

(47) *Q-Chem User's Manual*, version 3.2; Q-Chem, Inc.: 5001 Baum Blvd, Suite 690, Pittsburgh, PA 15213, 2009.

(48) Chaban, G. M.; Gerber, R. B. *Theor. Chem. Account* **2008**, *120*, 273–279.

(49) R. B. Gerber and G. M. Chaban and B. Brauer and Y. Miller, In *Theory and applications of computational chemistry: the first 40 years (a volume of technical and historical perspectives)*; C. E. Dykstra,, G. Frenking,, K. Kim,, G. Scuseria,, Eds.; Elsevier, Amsterdam, 2005; Chapter First-principles calculations of anharmonic vibrational spectroscopy of large molecules, pp 165–194.

(50) Chaban, G. M.; Gerber, R. B. *Theoretical Chemistry Accounts* **2007**, *120*, 273–279.

(51) Wang, X.; Werhahn, J.; Wang, L.; Wang, L.; Kowalski, K.; Lauberau, A.; Xantheas, S. *J. Phys. Chem. A* **2010**, *113*, 9579.

(52) Pye, C. C.; Rudolph, W. W. *J. Phys. Chem. A* **2001**, *105*, 905–912.

(53) Cobar, E. A.; Head-Gordon, M.; Bergman, R. in preparation.

(54) Lambrecht, D. S.; Clark, G. N. I.; Head-Gordon, T.; Head-Gordon, M. *J. Phys. Chem. A* **2011**, 110315121920022.

(55) Ramos Cordoba, E.; Lambrecht, D. S.; Head-Gordon, M. *Faraday Discuss.* **2011**, *150*, 345.

(56) Karplus, M.; Porter, R.; Sharma, R. *J. Chem. Phys.* **1965**, *43*, 3259.

(57) Porter, R.; Raff, L.; Miller, W. *J. Chem. Phys.* **1975**, *63*, 2214.

(58) Porter, R. *Ann. Rev. Phys. Chem.* **1974**, *25*, 317–355.

(59) Prell, J.; Bush, M.; Williams, E. *J. Am. Chem. Soc* **2010**.

Table Of Contents Graphic