Zero Point Energy of Polyhedral Water Clusters

David J. Anick*

Harvard Medical School, McLean Hospital, Centre Building 11, 115 Mill Street, Belmont, Massachusetts 02478

Received: October 11, 2004; In Final Form: May 4, 2005

Polyhedral water clusters (PWCs) are cage-like $(H_2O)_n$ clusters where every O participates in exactly three H bonds. For a database of 83 PWCs, $8 \leq n \leq 20$, geometry was optimized and zero point energy (ZPE) was calculated at the B3LYP/6-311++G** level. ZPE correlates negatively with electronic energy (E^0): each increase of 1 kcal/mol in E^0 corresponds to a decrease of about 0.11 kcal/mol in ZPE. For each n, a set of four connectivity parameters accounts for 98% or more of the variance in ZPE. Linear regression of ZPE against n and this set gives an RMS error of 0.13 kcal/mol. The contributions to ZPE from stretch modes only (ZPE_S) and from torsional modes only (ZPE_T) also correlate strongly with E^0 and with each other.

Introduction

The ab initio study of water clusters is pertinent to experimental systems, atmospheric water, and bulk phenomena such as proton transport and solvation.[1−15] In many instances, one is interested in comparing optimized $(H_2O)_n$ clusters having various geometries or H-bond orientations, to identify the one(s) with the lowest energy or to understand geometry−energy relationships. Zero point energy (ZPE) is a correction term that needs to be added to the electronic energy (E^0) when computing and comparing the total energy of $(H_2O)_n$ clusters. The calculation of ZPE can require a lot of computer time, especially for a model having a large basis set or for larger values of n. Consequently, the ZPE contribution has often been omitted in many studies of water clusters, on the implicit premise that its inclusion would not make a significant difference to the conclusions.

Studies seeking a single lowest-energy $(H_2O)_n$ for various n values[8,16−19] have found that ZPE can affect the $(H_2O)_n$ isomers' ordering by total energy. For $n = 6$, Kim and Kim[16] found that the Pedullo−Kim−Jordan prism geometry[17] has a slightly lower E^0 than the lowest cage geometry ($\Delta E^0 = -0.19$ kcal/mol) but the prism has higher ZPE ($\Delta ZPE = 0.35$ kcal/mol). Because ΔE^0 and $\Delta E^0 + \Delta ZPE$ have opposite signs, the inclusion of ZPE does alter the ordering of these two nearly isoenergetic structures. To achieve this level of precision (tenths of a kcal/mol) with high confidence, explicit individual calculations using very large basis sets will probably always be necessary. The statistical approaches featured in this article are best suited for setting upper bounds on how large ΔZPE can reasonably be expected to get or for refining "big picture" topology−energy correlations that apply broadly to large families of water clusters.

This project used a database approach to explore factors related to ZPE for the class of polyhedral water clusters (PWCs). PWCs are defined as cage-like water clusters in which every oxygen atom is three-coordinated. PWCs are a good class of water clusters on which to initiate a systematic study of ZPE for several reasons. First, at least some examples of PWCs, namely, the $(H_2O)_8$ cubes and the 5^{12} (pentagonal dodecahedron) $(H_2O)_{20}$ clusters, have been observed in experimental systems,[7−11] making the study experimentally relevant. Second, E^0 for PWCs is fairly well understood. A set of connectivity parameters is already established that correlate strongly with E^0, together accounting for 98% or more of the variance in E^0 for a fixed geometry.[20] Third, the class of PWCs is large, for example, McDonald et al.[21] have computed that there are over 30 000 symmetry-distinct PWCs just among those sharing the 5^{12} geometry, so it is reasonable to study PWCs using database methods.

The questions addressed by our project included the following. What is the range of ZPE for a particular family of water clusters (specifically, the $(H_2O)_n$ PWCs)? Is ZPE correlated with E^0? If so, is it positively or negatively correlated, and how good is the correlation (expressed as a correlation coefficient r or an ANOVA value r^2)? Can ZPE be accurately predicted strictly from knowledge of the connectivity pattern (including H-bond orientations) of a PWC? If so, are the same connectivity parameters relevant to ZPE as are relevant for E^0? Last, because there is so much variation in the set of H-bond lengths among PWCs[20,22−25] and because H-bond length correlates with O−H stretch mode frequency, a plausible starting guess or assumption was that most of the variance in ZPE (within the class of PWCs) would be due to the contribution from the O−H stretch modes. This was a testable hypothesis.

To answer these questions, we generated a database of 83 PWC $(H_2O)_n$ clusters, for n values ranging from 8 to 20. Each was optimized via B3LYP/6-311++G**, and ZPE/freq was computed. This model has been used successfully in various water cluster studies, and benchmark comparisons with MP2 have found it to be well suited to the study of PWCs.[26−32] For each value of n, ZPE was regressed against E^0. The set of 83 ZPE values was also regressed against n and a set of connectivity parameters.

Methods

Calculations were done on a Parallel Quantum Solutions (PQS) QuantumCube, using PQS parallel software.[33] Optimization was done in inverse cluster coordinates using the OPTIMIZE algorithm.[34] Setting the optimization "scale" factor to 5.0 gave efficient convergence. Initial geometries were obtained using the approximation algorithm described in ref 35. Statistics and plotting were done with R-project software.[36]

* Corresponding author. David.Anick@rcn.com.

10.1021/jp045359e CCC: $30.25
Published on Web 06/08/2005

Descripton of the Database. The information defining the connectivity pattern of a PWC consists of an "underlying geometry" (UG), which tells which pairs of oxygen atoms are H bonded, and a direction for each H bond, thought of as an arrow from the donor O to the acceptor O. Ice rules, which have been considered by several investigators, impose constraints on which combinations of H-bond directions are permitted.[37] Oxygen atoms that carry a free (also called non-H-bonded, dangling, or pendent) hydrogen have a donor−acceptor−acceptor pattern and are denoted "F" (for "free") or "DAA" or "2AW" (for "2-acceptor-water"). Those having a non-H-bonded lone pair have a donor−donor−acceptor pattern and are labeled "L" (for "lone pair") or "DDA" or "2DW".

Although most PWC arrangements that satisfy the ice rules do correspond to a PES local minimum, some do not. As observed in ref 35, two features that raise a PWC's electronic energy and that sometimes render it unstable are three-sided faces and faces that are all-F or all-L. We therefore excluded from consideration PWCs having either of these features. The project is technically an exploration of ZPE for the class of PWCs having no triangles and no all-F or all-L faces, rather than for all three-coordinated water clusters.

Let us say a little more about the issue of excluded PWCs, which is mostly about predicting which PWC arrangements will be stable. ("Stability" here refers only to the presence of a PES local minimum with the given arrangement.) The excluded set varies with the model used. Working with the model B3LYP/cc-pVDZ for the optimization steps, Singer and co-workers[22] described spontaneous rearrangement of H bonding for various 5^{12} initial setups, including some that we found to be stable and, hence, permissible. For the subtle matter of the stability of 5^{12} PWC arrangements, the basis set can make a big difference, and we strongly recommend basis sets that include diffuse functions. The direction of the difference is that the energy of autoionization consistently rises (i.e., becomes more positive) when switching from the smaller to the larger basis set. For example, Singer and co-workers found that the neutral form of Figure 3 of ref 22 sits on a very flat region of PES for B3LYP/cc-pVDZ and the electronic energy change is −14.3 kcal/mol for conversion to the zwitterion. We computed the same example via B3LYP/ 6-311++G** and determined that the neutral form has a clear PES local minimum whose transition state lies 1.65 kcal/mol higher, and the electronic energy of autoionization is only −7.65 kcal/mol. Thus, the larger basis set can render stable some neutral arrangements that appear unstable or metastable under the smaller basis set. The key points for this article are that if a smaller basis set were used, then we would have less confidence in the results, and because many more structures might appear to be unstable, the excluded set would have to be substantially enlarged. Illustrating the importance of including diffuse basis sets when modeling H bonds, we note that the model B3LYP/cc-pVDZ does not even choose the correct structure for the water dimer: it prefers the C_i-symmetric 2-H-bond dimer over the true minimum that has a single H bond and C_s symmetry.[38,39]

We generated a database of $(H_2O)_n$ PWCs for $n = 8$, 10, 12, 14, 16, and 20. Because of the importance of the 5^{12} dodecahedral arrangement, all of our $(H_2O)_{20}$ clusters had the 5^{12} dodecahedron as their UG. For $n = 8$ and 10, just one (nontriangle-containing) UG is possible, namely, the cube and the pentagonal prism, respectively, and for $n = 12$ there are two (4^45^4 "cage" and 4^66^2 hexagonal prism). For $n = 14$, we considered four geometries (4^35^6, $4^45^46^1$, $4^55^26^2$, and 4^66^3) and for $n = 16$, we considered seven UGs (4^25^8, $4^35^66^1$, and three varieties of $4^45^46^2$, $4^55^26^3$, and 4^66^4). By employing a database containing many UGs, we increase our confidence that any results obtained are applicable to PWCs generally and are not "artifacts" of some particular arrangement.

TABLE 1: Summary of Distribution of PWCs in Database

n	8	10	12	14	16	20
no. random	12[a]	5	7	5	6	5
no. designer		5	5	5	17	11
total	12	10	12	10	23	16

[a] No selection was involved for $n = 8$ because the entire population was used in the database.

In selecting the structures to be included in the database, some issues and tradeoffs arose. One might think that a "random sample" of the full set of (nonexcluded) PWCs would be the best way to choose a database. This turns out to be impractical for several reasons. First, for a strictly random sample one would first have to generate the complete set of permissible structures. Except for the smallest UGs, these sets number in the thousands and the algorithms for generating them are complex.[21,37] Second, it turns out that the properties of greatest interest, such as E^0, are not distributed uniformly. Instead, there is a large "hump" for midrange values of E^0 and total dipole moment,[25,40] meaning that a small random sample would be likely to miss or underrepresent both the higher and lower ends of the E^0 range. Last, certain connectivity features of interest could also be underrepresented or entirely absent from a strictly random database. Specifically, the lowest energy PWCs for a given UG share the property of minimizing the number of FF (or DAA−DAA) H-bonds (we denote this number as B_{FF}), and structures having the minimum B_{FF} rarely occur by chance. Similarly, a motif such as a hexagonal face having a uniform (or homodromic) H-bonding direction occurs rarely by chance. Homodromic faces have been postulated by some to be relevant to structure−energy relationships.[21,24]

All of this serves to explain why we picked the database the way we did. For $n = 8$, there are 12 nonexcluded structures and we used all 12. For each $n > 8$, we generated at least five structures randomly but then supplemented those with "designer" PWCs. The "random" PWCs were obtained using a four-step process: (1) pick a UG; (2) with a random number generator (RNG), assign half of the O's to be "F" and half to be "L"; (3) determine all of the connectivity patterns consistent with the assignments in step 2 and the ice rules via the algorithm of ref 37; (4) using the RNG, select one of the structures generated in step 3. The designer PWCs consisted of some having the lowest B_{FF} values and others that were spaced approximately evenly across the range of E^0 values. We also tried to make sure that a variety of connectivity motifs were represented. In no case did we know a cluster's ZPE when deciding whether to include it. Table 1 summarizes the distribution of random and designer entries in our database. A "random + designer" method was also used implicitly by Singer and co-workers:[40] in modeling energy and heat-capacity dependency on connectivity parameters, they found that a random set of twenty 5^{12} PWCs needed to be supplemented with 10 chosen from among those with lowest energy to replicate a curve derived from the set as a whole.

Results

Correlation of ZPE with E^0. For each value of n, $n = 8$, 10, 12, 14, 16, and 20, we regressed ZPE against E^0. That is,

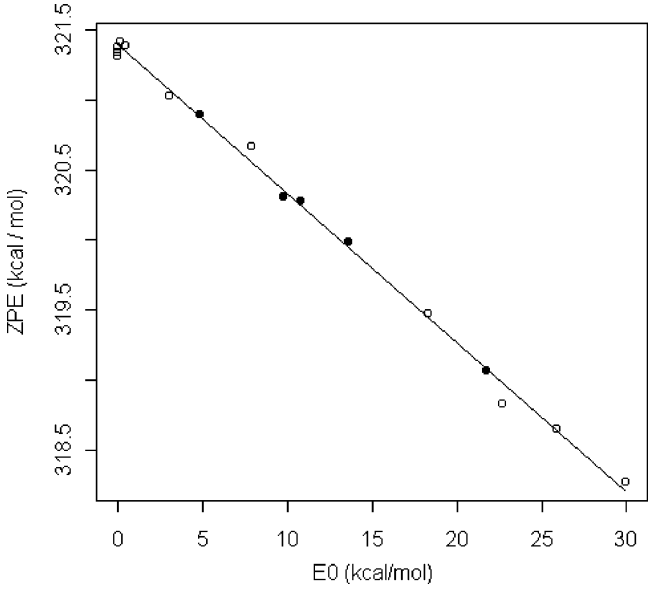

Figure 1. Scatter diagram for ZPE vs E^0 for $n = 20$ ($N = 16$). Lowest E^0 is taken as zero energy. ○ = designer clusters ● = random clusters.

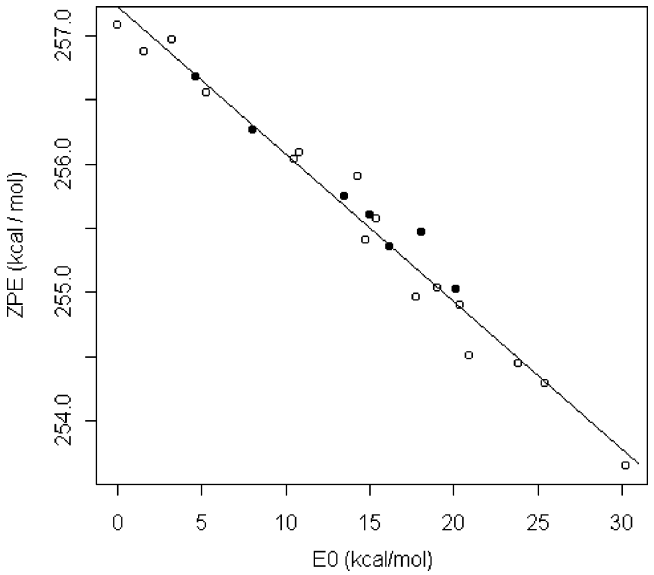

Figure 2. Scatter diagram for ZPE vs E^0 for $n = 16$ ($N = 23$). Lowest E^0 is taken as zero energy. ○ = designer clusters ● = random clusters.

we computed a least-squares best-fit line of the form

$$\mathrm{ZPE} = a + bE^0 \tag{1}$$

Figure 1 shows the scatter diagram of ZPE versus E^0 along with the best-fit line, for $n = 20$ ($N = 16$ points). Designer and random clusters are represented as open and filled circles, respectively, in Figure 1. There is no indication of any difference between the designer and random groups in terms of their relationship between ZPE and E^0. Figure 2 shows the results for $n = 16$ ($N = 23$ points). Results for smaller n values are not depicted graphically, but all of the results are summarized in Table 2. For each n, Table 2 lists the slope (b), intercept (a), root-mean-square deviation of ZPE from the best-fit line (RMSΔ), correlation coefficient (r), and ANOVA (r^2).

In each case, we found a negative correlation of ZPE with E^0. The slopes were consistently around −0.11, except for the 12 cubes, which gave a slope of −0.137. Implicit in this result was the finding that the range of ZPE values was approximately $^1/_9$ of the range of E^0 values, for each n. Correlation of ZPE with E^0 was excellent, with E^0 accounting for 89−99+% of the variance in ZPE. RMSΔ values ranged from 0.045 to 0.20 kcal/mol. RMSΔ was considerably smaller for the three subdatabases that shared a single UG than for the three subdatabases that combined multiple UGs.

TABLE 2: Summary of Best-Fit Lines for ZPE vs E^0

n	N	a^a	b^b	RMSΔ[a]	r (corr)	r^2 (ANOVA)	no. geoms
8	12	128.151	−0.1372	0.045	.9924	.9849	1
10	10	160.317	−0.1130	0.076	.9838	.9679	1
12	12	192.670	−0.1113	0.200	.9448	.8926	2
14	10	224.867	−0.1127	0.142	.9825	.9653	4
16	23	257.226	−0.1148	0.154	.9865	.9732	7
20	16	321.393	−0.1062	0.066	.9983	.9966	1

[a] y intercept and RMS deviation, units are kcal/mol. [b] Slope, dimensionless.

Regression of ZPE against Connectivity Parameters. Various connectivity parameters have been proposed as contributing to the energy of PWCs. We have already mentioned B_{FF}, the number of FF H-bonds. For LF H-bonds (i.e., those whose donor is L and whose acceptor is F), their total number, denoted B_{LF}, always equals $(^n/_2) - B_{FF}$, so B_{LF} is not an independent variable once we include B_{FF} in the model. However, the number of LF bonds that locate the two nearest H's in a trans orientation,[20,25] also called "strong" LF bonds by Kirov,[23,24] is an independent variable, and we denote it as B_{LFT}. The parameter A_{TH} denotes the total number of times that a chain of three adjacent F's or three adjacent L's occurs. The parameters B_{FF} and A_{TH} are positively correlated with each other, but they are independent. We let H_0 denote the number of homodromic faces, and let F_k be the number of k-sided faces for $k = 4$, 5, or 6. Because of the formulas [35] $F_4 + F_5 + F_6 = 2 + (^n/_2)$ and $F_5 = 12 - 2F_4$, only one of $\{F_4, F_5, F_6\}$ is independent, and we use F_4. The variables S_5 and S_6 contribute significantly to determining E^0, where S_k counts the number of times that the H−O−H angle of a DDA (or "L") lies in a k-sided face,[35] for $k = 4$, 5, or 6. Because $S_4 + S_5 + S_6 = {}^n/_2$, only two of these (we use S_5 and S_6) are independent. Thus, a plausible set of PWC connectivity parameters to consider is

$$\mathbb{S} = \{1, n, B_{FF}, B_{LFT}, A_{TH}, H_0, F_4, S_5, S_6\}$$

Regression of ZPE against the set $\mathbb{S}$ yielded a correlation coefficient of 1.0000 and an RMSΔ of 0.127 kcal/mol. The high correlation is due mainly to the fact that the principal determinant of ZPE is n, the number of H_2O units. Regression of ZPE against $\{1, n\}$ alone gave $r = 0.9999$.

We used backward elimination[41] starting from set $\mathbb{S}$ to remove variables whose contribution to the model were not significant. Significance is measured by a p value, and we set the significance cutoff at $p < 0.01$. The backward elimination algorithm computes the p value for each variable in a regression model and removes the least significant variable, repeating this step until all of the remaining variables have p values below the cutoff. Applied to $\mathbb{S}$, the variables kept were $\mathbb{S}_1 = \{n, B_{FF}, B_{LFT}, A_{TH}, F_4\}$. All have p values < 0.0005. The last variable to be eliminated was S_5, with a p value of 0.045, so it is possible that this variable would achieve significance if a larger database were available. The least-squares fit for the set $\mathbb{S}_1$ is

$$\mathrm{ZPE} \approx (16.0667)n - (.0781)B_{FF} + (.0255)B_{LFT} - (.1357)A_{TH} - (.0912)F_4 \tag{2}$$

with RMSΔ = 0.127. For comparison, backward elimination

TABLE 3: Best-Fit Lines for ZPE Components vs E^0 for $n = 20$

component	a^a	b^b	RMSΔ[a]	r (corr)	r^2 (ANOVA)
ZPE-S	203.1879	.1120	0.163	.9908	.9817
ZPE-B	48.0732	.0006	0.043	.1514	.0229
ZPE-T	70.1320	−.2188	0.139	.9982	.9964
ZPE	321.3931	−.1062	0.066	.9983	.9966

[a] y intercept and RMS deviation, units are kcal/mol. [b] Slope, dimensionless.

with E^0 instead of ZPE as the dependent variable for same database yielded a different set, $\mathbb{S}_2 = \{n, B_{FF}, A_{TH}, F_4, S_5, S_6\}$

$$E^0 \approx (-47989.8664)n + (.7756)B_{FF} + (1.2977)A_{TH} + (.9544)S_5 + (.6821)S_6 + (2.1009)F_4 \quad (3)$$

with RMSΔ = 1.241. Last, for each value of n we repeated the regression of ZPE against $\mathbb{S}_1$ for the subdatabases. We removed F_4 from $\mathbb{S}_1$ for these calculations in the situations where there was a single UG. Correlations ranged from $r = 0.9876$ for $n = 12$ to $r = 0.9987$ for $n = 20$, supporting the predictive power of this set of variables.

Components of ZPE. The formula by which ZPE is computed is

$$\mathrm{ZPE} = \frac{hc}{2}\sum_{i=1}^{9n-6}\mathrm{freq}_i \quad (4)$$

where freq_i are the frequencies in cm^{-1} of the cluster's normal modes, listed in ascending order. An $(H_2O)_n$ cluster has $9n-6$ normal modes, of which the highest $2n$ are O−H stretch modes, the next n are H−O−H bending modes, and the lowest $6n-6$ are torsional modes. We can split ZPE into its contributions from stretch, bend, and torsional modes

$$\mathrm{ZPE} = \mathrm{ZPE_S} + \mathrm{ZPE_B} + \mathrm{ZPE_T} \quad (5)$$

In eq 5, $\mathrm{ZPE_S}$ is defined as the contribution from the stretch modes only, that is

$$\mathrm{ZPE_S} = \frac{hc}{2}\sum_{i=7n-5}^{9n-6}\mathrm{freq}_i \quad (6)$$

and likewise for $\mathrm{ZPE_B}$ and $\mathrm{ZPE_T}$.

We regressed $\mathrm{ZPE_S}$, $\mathrm{ZPE_B}$, and $\mathrm{ZPE_T}$ against E^0 for each n. The results for $n = 20$ were typical and are shown in Table 3. In Table 3, a and b are the intercept and slope as in eq 1. $\mathrm{ZPE_S}$ is positively correlated with E^0, $\mathrm{ZPE_T}$ is negatively correlated, and $\mathrm{ZPE_B}$ is uncorrelated (p = 0.58). Moreover, the RMSΔ values for $\mathrm{ZPE_S}$ and $\mathrm{ZPE_T}$ are considerably poorer than for (total) ZPE. The only way this can happen is if the deviations from the least squares lines for $\mathrm{ZPE_S}$ and for $\mathrm{ZPE_T}$ are inversely correlated so that they approximately cancel out. Figure 3 shows the three components for each $(H_2O)_{20}$, along with the total ZPE, plotted against E^0. Vertical positions have been adjusted for better visual inspection, that is, a fixed offset was added to all $\mathrm{ZPE_S}$ values, a different fixed offset was added to all $\mathrm{ZPE_T}$ values, and so on. As Figure 3 makes clear, when a cluster's $\mathrm{ZPE_S}$ (open circle) lies above the $\mathrm{ZPE_S}$ line, its $\mathrm{ZPE_T}$ (open square) generally lies below the $\mathrm{ZPE_T}$ line and vice versa. The deviations approximately cancel out when they are added, leading to a small RMSΔ for ZPE (filled circles). Numerically, the correlation coefficient for $\mathrm{ZPE_S}$ versus $\mathrm{ZPE_T}$ is 0.9957.

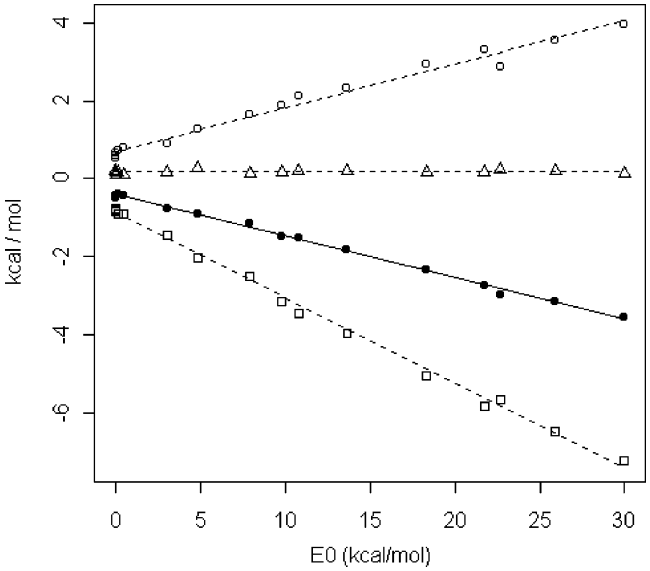

Figure 3. Scatter diagram for components of ZPE vs E^0 for $n = 20$. Key: □ = $\mathrm{ZPE_T}$ △ = $\mathrm{ZPE_B}$ ○ = $\mathrm{ZPE_S}$ ● = ZPE.

Last, we regressed each of the components against $\{1, B_{FF}, B_{LFT}, A_{TH}, H_0\}$ for $n = 20$. We found that H_0 had no significance for any component, and B_{FF} and B_{LFT} also had p values > 0.2 for $\mathrm{ZPE_S}$ and $\mathrm{ZPE_T}$. The parameter A_{TH} was highly correlated ($p < 10^{-6}$) with both $\mathrm{ZPE_S}$ and $\mathrm{ZPE_T}$: positively correlated with $\mathrm{ZPE_S}$ and negatively correlated with $\mathrm{ZPE_T}$. For $\mathrm{ZPE_B}$, the parameters B_{FF}, B_{LFT}, and A_{TH} had p values $< 10^{-4}$, but because the RMSΔ of $\mathrm{ZPE_B}$ is so small, $\mathrm{ZPE_B}$ makes very little contribution to explaining the variance in ZPE values.

Discussion

The tight and consistent correlation between ZPE and E^0 for PWCs (Table 2) is our most robust finding. According to the slopes in Table 2, each 1 kcal/mol increase in E^0 corresponds to a decrease of about 0.11 kcal/mol in ZPE. Inclusion of the ZPE correction makes almost no difference to the ranking of PWCs by energy, but it "softens" the total energy differences by about 11%. Two PWCs whose E^0 values differ by 10 kcal/mol can be expected to have their total energy (i.e., $E = E^0 +$ ZPE) values differ by about 8.9 kcal/mol.

The correlation is partially "explained" by eqs 2 and 3. The parameters B_{FF} and A_{TH} are arguably the principal determinants of both ZPE and E^0. Their coefficients in eq 2 are both about −0.10 times the corresponding coefficients in eq 3. This observation by itself would suggest that ZPE should equal approximately (a constant) $-(0.1) \times E^0$, thus explaining the ZPE−E^0 correlation.

The parameter F_4 also appears in both regression equations, but the coefficient ratio is −0.43 instead of −0.1. This weakens the correlation between ZPE and E^0 for any database where F_4 can assume more than one value, that is, in databases containing multiple UGs. Indeed, we have already noted that RMSΔ values for ZPE vs E^0 are considerably worse for the subdatabases that combine multiple UGs (Table 2).

The dependence of E^0 on parameters S_5 and S_6 was tentatively explained in ref 35 as due to how a double donor H−O−H molecule sits relative to the plane of the polyhedral face that contains it. ZPE appears to be unaffected or minimally affected by these parameters. Conversely, B_{LFT} is important to ZPE but does not meet the significance cutoff for E^0. (Actually, $p = 0.03$ for inclusion of B_{LFT} in the regression model for E^0, so B_{LFT} might prove significant if a larger database were available.)

B_{LFT} has to do with the positioning of neighbor H_2O molecules in the LF H-bonds.

Our results correlating the components of ZPE, with E^0 and with connectivity parameters, raise more questions than they answer. Upon first seeing the negative correlation of ZPE with E^0, we imagined this correlation might be mediated by differing O−H distance distributions being reflected in different stretch mode frequencies and leading to different ZPE_S values, whereas ZPE_T might have little correlation with E^0. This guess was completely wrong. Instead, ZPE_S correlates positively with E^0, and the dominant component of ZPE is ZPE_T, whose negative correlation with E^0 more than reverses the positive contribution of ZPE_S when the components are added. Moreover, the fact that the deviations of ZPE_S and ZPE_T from their least-squares lines nearly cancel out suggests a tight relationship between the stretch and torsional mode frequencies. So far, such a relationship defies explanation.

More connectivity parameters may exist in addition to the list we considered here that correlate well with E^0, ZPE, or ZPE components. Inclusion of the right parameters might lead to a more complete and convincing explanation for the statistical trends we have observed. Singer and co-workers[40] noted that any such parameters would have to be invariant under symmetries of the UG ("graph invariants"). They used this idea to generate a hierarchy of connectivity parameters, with the "level" of a parameter being its degree as a polynomial in the $(3n/2)$ variables that stand for individual H-bond directions. Their method is generalizable to periodic ice lattices. Unfortunately, the method is not practical for UGs that lack substantial symmetry or for a database like ours that contains multiple UGs. For instance, for Figure 3, item 9 of ref 35, which was included as one of our 16-mers, its UG has symmetry group C_2. For this UG, the graph invariants method generates 11 first-order invariants and 144 second-order invariants. One would need at least 155 PWCs based on this UG alone in order to begin to study them by this method: otherwise one would have an underdetermined system of equations that would yield no information. Also, the set of invariants is meaningful only for the specific UG, for example, one could not apply the 155 invariants of our previous example to a different UG. For the very special case of the 5^{12} UG, however, there are no first-order invariants and just 7 invariants at the second level, and these include two invariants that equate with linear combinations of our variables B_{FF} and B_{LFT}. Hopefully, if a large enough database of optimized 5^{12} clusters and their ZPEs can be obtained, the questions we have studied can be revisited using a graph-invariant-derived set of parameters.

Conclusions

Our principal results are as follows:

1. ZPE of PWCs is negatively correlated with electronic energy (E^0). Each 1 kcal/mol increase in E^0 corresponds to a 0.11 kcal/mol drop in ZPE. From 89% to 99+% (depending on the particulars of the database) of the variance in ZPE is accounted for by E^0 alone. Comparisons of PWC's that use E^0 alone are likely to remain valid after the ZPE correction is included, though the energy differences will shrink by about 11%.

2. ZPE for PWCs can be predicted very accurately from n and the connectivity parameters B_{FF}, B_{LFT}, A_{TH}, and (if applicable) F_4. For the full database of 83 PWCs from cubes to dodecahedra, the RMS error was 0.13 kcal/mol. The RMS error will be smaller if the database is restricted to a single underlying geometry.

3. ZPE can be represented as the sum of three components that are the contributions to ZPE from O−H stretch modes (ZPE_S), from H−O−H bending modes (ZPE_B), and from torsional modes (ZPE_T). ZPE_B is uncorrelated with E^0, whereas ZPE_S (resp. ZPE_T) correlates positively (resp. negatively). ZPE_S and ZPE_T are very strongly correlated with each other in a manner that causes their deviations from least-squares lines to nearly cancel out when added.

Could our results help with "close-call" situations such as the cage and prism hexamer example cited in the Introduction? The "predicted" ΔZPE would be about $(-0.11) \times (-0.19) = 0.02$ kcal/mol, but since the cage and hexamer have different UGs, an RMSΔ of around 0.2 kcal/mol might be expected. Two standard deviations would therefore be 0.4 kcal/mol, making our predicted range for a 2σ confidence level 0.02 ± 0.4 or $(-0.38, 0.42)$. This example would be "too close to call", that is, we would not be able to predict whether the actual ΔZPE would exceed the $-\Delta E^0$ value of 0.19. In fact, as we saw ΔZPE here is 0.35, within but near the high end of this range.

For the cage and prism, it is hard to say whether our results should apply at all because the cage is not a PWC and the prism would technically be "excluded" because it contains triangular faces. This leads to a good question: how broadly will the ZPE−E^0 correlation apply to nonpolyhedral water clusters? We have not yet conducted any studies of this question, but we do not expect the correlation to be as good. The fact that the correlation weakens as multiple UGs are included suggests that the correlation will become weaker still when non-PWCs are included. Studying this question for other water cluster families may provide insight into what features of PWCs contribute to making the correlation so strong.

The correlations of ZPE_S and ZPE_T with each other and with E^0 remain a mystery for which we have no good hypotheses at this time. Perhaps many or most torsional modes can be associated closely with a single H_2O unit or H bond, and then one could split up ZPE differently, that is, one could examine the contribution to ZPE from each H_2O unit or from each H bond. If the torsional and stretch contributions were correlated on each subunit, this could explain how their sums over the entire cluster become correlated. Buch and Devlin[42] may have taken the first steps toward a subunit analysis of IR spectra when they interpreted O−H stretch modes in ice via their disordered tetrahedral model of ice. Conceivably, different H-bond environments (i.e., FF, chain of three L's, etc.) might influence both the local torsional and the local stretch mode frequencies in a manner that generates the observed correlations.

Glossary

A_{TH}	total number of chains of three F's or three L's in a PWC
B_{FF}	total number of H bonds that have type FF
B_{LFT}	total number of H bonds that have type LFT
DAA	donor−acceptor−acceptor pattern for an oxygen atom (also "F")
DDA	donor−donor−acceptor pattern for an oxygen atom (also "L")
E^0	ground-state electronic energy
F	describes an oxygen atom in a PWC that has a single non-H-bonded (or "free") H
FF	describes an H bond in which donor and acceptor are both "F"
F_k	number of faces of a PWC having k sides ($k = 4$, 5, or 6)
H_0	number of homodromic faces of a PWC
L	describes an oxygen atom in a PWC that has a non-H-bonded lone pair

LFT	describes an H bond in a PWC for which the donor is "L", the acceptor is "F", and the additional hydrogens on the donor and acceptor are situated trans
n	always refers to the number of H_2O units in a cluster
PWC	polyhedral water cluster, that is, cage-like $(H_2O)_n$ in which every O is in three H bonds
r	correlation coefficient
RMSΔ	root-mean-squared deviation of data from a best-fit line or plane
S_k	total number of H−O−H angles in a PWC that lie in faces of *k* sides ($k = 4$, 5, or 6)
UG	underlying geometry, that is, the pattern of H bonding among the oxygen atoms of a water cluster without regard for the H-bond directions or lengths

References and Notes

(1) Bryce, R. A.; Vincent, M. A.; Hillier, I. H. *J. Phys. Chem. A* **1999**, *103*, 4094−4100.

(2) Naik, J.; Kim, J.; Majumdar, D.; Kim, K. S. *J. Chem. Phys*. **1999**, *110*, 9116−9127.

(3) Caldwell, J. W.; Kollman, P. A. *J. Phys. Chem*. **1992**, *96*, 8249−8251.

(4) Rempe, S. B.; Pratt, L. R. *Fluid Phase Equilib*. **2001**, *183*, 121−132.

(5) Carrillo-Tripp, M.; Saint-Martin, H.; Ortega-Blake, I. *J. Chem. Phys.* **2003**, *118*, 7062−7073.

(6) Nauta, K.; Miller, R. E. *Science* **2000**, *287*, 293−295.

(7) Blanton, W. B.; Gordon-Wylie, S. W.; Clark, G. R.; Jordan, K. D.; Wood, J. T.; Geiser, U.; Collins, T. J. *J. Am. Chem. Soc*. **1999**, *121*, 3551−3552.

(8) Buck, U.; Ettischer, I.; Melzer, M.; Buch, V.; Sadlej, J. *Phys. Rev. Lett*. **1998**, *80*, 2578−2581.

(9) Wei, S.; Shi, Z.; Castleman, A. W., Jr. *J. Chem. Phys*. **1991**, *94*, 3268−3270.

(10) Shin, J.-W.; Hammer, N. I.; Diken, E. G.; Johnson, M. A.; Walters, R. S.; Jaeger, T. D.; Duncan, M. A.; Christie, R. A.; Jordan, K. D. *Science* **2004**, *304*, 1137−1140.

(11) Miyazaki, M.; Fujii, A.; Ebata, T.; Mikami, N. *Science* **2004**, *304*, 1134−1137.

(12) Plummer, P. L. M. *J. Phys. Chem. B* **1997**, *101*, 6251−6253.

(13) Mei, H. S.; Tuckerman, M. E.; Sagnella, D. E.; Klein, M. L. *J. Phys. Chem. B* **1998**, *102*, 10446−10458.

(14) Sadeghi, R. R.; Cheng, H.-P. *J. Chem. Phys*. **1999**, *111*, 2086−2094.

(15) Jensen, J. O.; Samuels, A. C.; Krishnan, P. N.; Burke, L. A. *Chem. Phys. Lett*. **1997**, *276*, 145−151.

(16) Kim, J.; Kim, K. S. *J. Chem. Phys*. **1998**, *109*, 5886−5895.

(17) Pedulla, J. M.; Kim, K.; Jordan, K. D. *Chem. Phys. Lett*. **1998**, *291*, 78−84.

(18) Brudermann, J.; Melzer, M.; Buck, U.; Kazimirski, J. K.; Sadlej, J.; Buch, V. *J. Chem. Phys*. **1999**, *110*, 10649−10652.

(19) Lee, H. M.; Suh, S. B.; Kim, K. S. *J. Chem. Phys*. **2001**, *114*, 10749−10756.

(20) Anick, D. J. *J. Mol. Struct*.: *THEOCHEM* **2002**, *587*, 97−110.

(21) McDonald, S.; Ojamäe, L.; Singer, S. J. *J. Phys. Chem. A* **1998**, *102*, 2824−2832.

(22) Kuo, J.-L.; Ciobanu, C. V.; Ojamäe, L.; Shavitt, I.; Singer, S. J. *J. Phys. Chem. A* **2003**, *118*, 3583−3588.

(23) Kirov, M. V. *J. Struct. Chem*. **2002**, *43*, 266−273.

(24) Kirov, M. V. *J. Struct. Chem*. **2002**, *43*, 274−283.

(25) Chihaia, V.; Adams, S.; Kuhs, W. F. *Chem. Phys*. **2004**, *297*, 271−287.

(26) Kim, K.; Jordan, K. D. *J. Phys. Chem*. **1994**, *98*, 10089−10094.

(27) Nova, J. J.; Sosa, C. *J. Phys. Chem*. **1995**, *99*, 15837−15845.

(28) Loerting, T.; Liedl, K. R.; Rode, B. M. *J. Chem. Phys*. **1998**, *108*, 2672−2679.

(29) Smith, A.; Vincent, M. A.; Hillier, I. H. *J. Phys. Chem. A* **1999**, *103*, 1132−1139.

(30) Anick, D. J. *J. Mol. Struct*. **2001**, *574*, 109−115.

(31) Christie, R. A.; Jordan, K. D. *J. Phys. Chem. A* **2001**, *105*, 7551−7558.

(32) Anick, D. J. *J. Phys. Chem. A* **2003**, *107*, 1348−1358.

(33) *PQS*, version 3.1, Parallel Quantum Solutions, 2013 Green Acres Road, Fayetteville, AR 72703.

(34) Baker, J.; Kessi, A.; Delley, B. *J. Chem. Phys*. **1996**, *105*, 192−212.

(35) Anick, D. J. *J. Chem. Phys*. **2003**, *119*, 12442−12456.

(36) R Project for Statistical Computing, http://www.r-project.org

(37) Anick, D. J. *J. Mol. Struct*.: *THEOCHEM* **2002**, *587*, 87−96.

(38) Smith, B. J.; Swanton, D. J.; Pople, J. A.; Schaefer, H. F.; Radom, L. *J. Chem. Phys*. **1990**, *92*, 1240−1247.

(39) Fellers, R. S.; Leforestier, C.; Braly, L. B.; Brown, M. G.; Saykally, R. J. *Science* **1999**, *284*, 945−948.

(40) Kuo, J.-L.; Coe, J. V.; Singer, S. J.; Band, Y. B.; Ojamäe, L. *J. Chem. Phys.* **2001**, *114*, 2527−2540.

(41) Weisberg, S. *Applied Linear Regression*; John Wiley & Sons: New York, 1980.

(42) Buch, V.; Devlin, J. P. *J. Chem. Phys*. **1999**, *110*, 3437−3443.