Polyhedral Water Clusters, II: Correlations of Connectivity Parameters with Electronic Energy and Hydrogen Bond Lengths

ARTICLE *in* JOURNAL OF MOLECULAR STRUCTURE THEOCHEM · JULY 2002

Impact Factor: 1.37 · DOI: 10.1016/S0166-1280(02)00100-8

1 AUTHOR:

David Anick
Tufts University, Medford, MA, United …

78 PUBLICATIONS **2,049** CITATIONS

ELSEVIER

Journal of Molecular Structure (Theochem) 587 (2002) 97–110

THEO CHEM

www.elsevier.com/locate/theochem

Polyhedral water clusters, II: correlations of connectivity parameters with electronic energy and hydrogen bond lengths

David J. Anick

Harvard Medical School, McLean Hospital, Bowditch Building, 115 Mill Street, Belmont, MA 02478, USA

Received 10 January 2002; accepted 27 February 2002

Abstract

Polyhedral water clusters (PWCs) are $(H_2O)_n$ clusters in which every oxygen is three-coordinated. Four polyhedral geometries on which water clusters can be based are the cube ($n = 8$), pentagonal prism ($n = 10$), hexagonal prism ($n = 12$), and 4^45^4 octahedron ($n = 12$). For each of these geometries, a database was obtained of PWCs optimized at the B3LYP/6-311++G** level. The total database contained 1311 hydrogen bonds in 82 optimized PWCs. Using linear regression, correlations were sought linking aspects of the clusters' H-bonding pattern with their electronic energies and with the lengths of their H-bonds. Excluding six very high-energy clusters whose H-bonding pattern included a motif called a cyclic component, 98.9% or more of the variance in clusters' electronic energy could be accounted for by just three connectivity parameters, for each of the geometries. Five families of H-bonds were distinguished based on the presence or absence of a pendent H on the acceptor and donor O and on the layout of adjacent H's. Within each family, the presence or absence of pendent H's on first and second neighbor O's accounted for 86–95% of the variance in H-bond length, and could be used to predict H-bond length with an RMS error $\leq$ 2.4 pm.

Keywords: Water cluster; Polyhedron; Connectivity; Benchmarks; Database; Hydrogen bond length

1. Introduction

Cage-like water clusters have been the subject of much theoretical and experimental work [1–6], and among cage-like structures the polyhedral water clusters (PWCs) form an interesting and important class. In this article 'polyhedral' means that each oxygen atom in the water cluster participates in exactly three H-bonds. Polyhedral clusters include the unit cells of the clathrate hydrates [7]; of these, the dodecahedron ($(H_2O)_{20}$) has been the subject of considerable study [8–11]. Smaller polyhedral clusters such as the cube and the pentagonal prism may also occur as building blocks of larger minimum-energy clusters [12–14]. We offer as further motivation, that ab initio studies of PWCs can provide insight into the H-bond cooperativity that occurs for three-coordinated oxygens in bulk water. Around half of the oxygens in bulk water at 25 °C are three-coordinated at any given moment, based on simulations which find that the average coordination number is around 3.1 [15,16].

Excluding structures that contain triangles, the simplest polyhedral geometries after the cube are the pentagonal prism ($(H_2O)_{10}$), the hexagonal prism ($(H_2O)_{12}$), and an octahedron ($(H_2O)_{12}$) illustrated as Fig. 1. The standard notation for Fig. 1 is 4^45^4, meaning that it has four quadrilateral and four pentagonal faces. For simplicity in this article we call Fig. 1 'the octahedron' even though the hexagonal prism (4^66^2) also has eight faces.

E-mail address: david.anick@gte.net (D.J. Anick).

0166-1280/02/$ - see front matter
PII: S0166-1280(02)00100-8

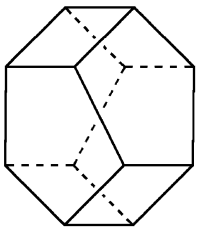

Fig. 1. Octahedron.

For each of these four geometries (i.e. cube or 4^6, pentagonal prism or $4^5 5^2$, octahedron or $4^4 5^4$, hexagonal prism or $4^6 6^2$), we computed a database of $(H_2O)_n$ structures optimized at the B3LYP/6-311++G** level. The number of structures was 14, 27, 20, 21, respectively. This model was chosen because it has been shown in several previous studies to give good results for water clusters [17–22]. We verified this by re-optimizing seven of the water cubes at the MP2/6-311++G** level, to provide polyhedral benchmarks.

For each of the four geometries, we examined correlations between H-bonding patterns and the electronic energy, denoted E^0. Omitting six high-energy clusters of the 'cyclic-component-containing' type, we found a set of three parameters which account for 99% of the variance in E^0, within each polyhedral class. These parameters are denoted b_{FL}, a_{Th}, and sy_m ($m = 5$ or 6) and are defined in Section 2.

Treating the H-bonds in all the optimized structures as comprising a database of 1311 H-bonds, we split it into five subsets and observe that within each subset, 87–95% of the variance in H-bond length can be accounted for by the presence or absence of pendent H's on the first and second order neighbor O's, for an RMS error of about 0.8% of the average O–O distance or less. Although statistical correlations of this type cannot 'explain' PWC properties, we hope they will add to the growing insight into factors that influence the behavior of water clusters and bulk water.

2. Terminology and an illustrative example

PWCs obey numerous interesting mathematical relationships. It is imperative to know some of these relationships and the parameters involved in them, in order to define and interpret the statistical correlations described in this study. Full definitions and proofs are relegated to the companion article [23], but we recount the basics here so as to make this article self-contained.

All PWCs in this study are neutral and obey an analog of the Bernal–Fowler ice rules (details in Ref. [23]). As a consequence in each PWC half of the oxygens carry a pendent or 'free' hydrogen that is not in an H-bond and are called 'F-type' oxygens. The other half have a non-H-bonding lone pair and are called 'L-type' oxygens. Fig. 2(a) shows one cluster in the hexagonal prism database. Solid lines are covalent bonds and dashed lines connect H's that are in H-bonds to their acceptor O's. The 'H-connectivity diagram' for this cluster is shown as Fig. 2(b). The F-type oxygens (respectively, L-type oxygens) are represented as filled (respectively, open) circles, O-to-O H-bonds are solid lines, and asterisks represent bonding hydrogens. The asterisk positions are significant only in that each is closer to one end of the line (H-bond) on which it lies: this end is understood to be the donor O of the H-bond.

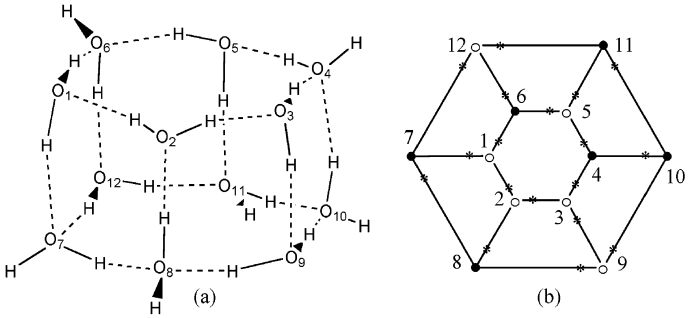

Fig. 2. Representations of a polyhedral water cluster. (a) Plane projection of B3LYP-optimized geometry. (b) H-connectivity diagram.

If one removes from an H-connectivity diagram all the L-type O's and all lines that connect to L-type O's, the connected pieces that remain are the 'F-components' of the diagram. In Fig. 2(b) there are three F-components, namely {4,10,11}, {6}, and {7,8}. Likewise the L-components are {1,2,3,9}, {5}, and {12}. All of these components are 'acyclic', i.e. contain no closed loops. It is not an accident that there is the same number (three) of F- and of L-components in this example. In Ref. [23], Theorem 3 it is proved that there is a one-to-one match-up among the acyclic F-components, the acyclic L-components, and the 'type FL' bonds, i.e. H-bonds in which the donor has type F and the acceptor has type L. Thus there must also be exactly three type FL bonds in Fig. 2(b). The reader is encouraged to look for these three and to check that there are no others (ans: they are 4–5, 6–12, and 8–2). The notation for the number of type FL bonds is b_{FL}. A PWC in which one or more of the F- or L-components is not acyclic is said to be a cyclic-component-containing (or 'CCC') cluster. The 'contiguity' of a PWC, denoted cg, is the size of the largest F- or L-component. For Fig. 2(b), the component {1,2,3,9} makes cg = 4.

An 'angle' is a set A–B–C of three oxygens in which A and B, and B and C, are H-bonded. It is an 'L-symmetric' angle if B is the donor to both A and C, and 'F-symmetric' if B is the acceptor from both A and C. In Fig. 2(b) the F-symmetric angles are 3–4–10, 5–6–1, 1–7–12, 7–8–9, 9–10–11, and 5–11–12. The L-symmetric angles are 6–1–7, 1–2–3, 4–3–9, 6–5–11, 8–9–10, and 7–12–11. As these lists illustrate, each O in a PWC is the vertex (i.e. the 'B' oxygen) of exactly one symmetric angle. If one selects only the F-symmetric angles that occur in the hexagonal faces, there are three (5–6–1, 7–8–9, 9–10–11). Similarly there are three L-symmetric angles in the hexagonal faces. The count of F- and of L-symmetric angles in hexagonal faces must always agree [23], Eq. (4.4). This number is denoted sy_6. Likewise the count of F- and of L-symmetric angles in all the pentagonal (respectively, four-sided) faces of a PWC agree, and is denoted sy_5 (respectively, sy_4).

Lastly, an angle A–B–C is 'homogeneous' if A, B, C all have the same type (i.e. all F or all L). The homogeneous angles are the angles that lie entirely in one F- or L-component. In Fig. 2(b) the homogeneous angles are 1–2–3, 2–3–9, and 4–10–11. Homogeneous angles may be symmetric or asymmetric; the same is true of inhomogeneous angles. The total number of homogeneous angles is denoted a_{Th}.

3. Methods

Hartree–Fock and B3LYP optimizations were done via Jaguar [24]. Jaguar's default pseudospectral algorithm has poor convergence behavior when basis sets include diffuse functions, so the slower non-pseudospectral method was used for all B3LYP/6-311++G** calculations. MP2 optimizations were done via the GAUSSIAN98 suite of programs [25]. For water clusters it is helpful to add a 'ModRedundant' section to the GAUSSIAN98 input which instructs GAUSSIAN to add all H-bonded O–O distances, all H-acceptor-O distances, and all nearest neighbor O–O–O angles to its set of redundant internal coordinates for optimization.

Benchmarks consisted of fully optimizing seven distinct water cubes, spanning a range of energies, via MP2/6-311++G**, B3LYP/6-311++G**, HF/6-31G**, and the ORIENT program [26]. Taking the MP2 optimum as the gold standard, we computed for each bond ($7 \times 12 = 84$ bonds in all) the difference in O–O length for MP2 vs each of the other methods, and then found the RMS of these differences. We also computed the electronic energy differences among the cubes, for each method.

The clusters chosen for optimization were intended to span a range of energies and properties. When it was recognized from the cubes and an initial group of other clusters that the quantities cg, b_{FL}, sy_m, and a_{Th} would correlate with E^0, additional effort was made to choose subsequent clusters which represented a spread of values for these parameters, and to avoid 'accidental' correlations among them. Statistical analysis was done using the R-project software [27].

4. Results

4.1. Benchmarks

Tables 1 and 2 show the results of benchmark comparisons for seven cube octamers optimized at various levels of theory. We took MP2/6-311++G**

Table 1
RMS values of Δr_{OO} and $\Delta\alpha_{OOO}$ for seven water cubes, for each of three methods vs MP2 Distances in pm, angles in °

Cube	RMS Δr_{OO}			RMS $\Delta\alpha_{OOO}$		
	B3LYP	HF	ORIENT	B3LYP	HF	ORIENT
*D*2*d*	0.13	7.55	4.76	2.13	1.98	3.55
*S*4	0.31	7.66	5.17	2.13	2.06	3.93
Ci	0.68	7.41	4.56	2.48	1.85	4.42
Cs	0.64	7.46	4.69	2.15	2.01	3.75
*C*1-Z	0.80	7.36	4.99	1.97	1.39	3.82
*C*1-Y	1.22	7.24	5.00	2.46	2.17	3.21
*C*4	1.20	6.18	0.97	1.97	0.81	1.90
(All)	0.81	7.28	4.90	2.19	1.81	3.59

calculations as the gold standard against which the other methods were compared. Cubes are identified by their symmetry group (cf. [13]) and are illustrated in Fig. 3. Each cube contains 12 O–O H-bonds and 24 O–O–O angles. In Table 1, the RMS of the 12 differences in O–O H-bond lengths (Δr_{OO}) from their corresponding lengths in the MP2-optimized structure is listed, along with the RMS of the 24 differences for O–O–O angles ($\Delta\alpha_{OOO}$). For the last row of Table 1, the RMS of all 84 length differences from MP2 (respectively, all 168 angle differences) is given. In Table 2, the various methods' electronic energies for the cubes are compared. The cube with *D*2*d* symmetry is taken as the 'zero' point for each method. What is of interest is how closely a method's computed energies approximate the energies computed via MP2. For each method other than MP2, $E^0_{\text{method}} - E^0_{\text{MP2}}$ was computed for the six cubes other than *D*2*d*, and the RMS of these 6 differences is given in the last row of Table 2. Table 2 also lists the contiguity and the values of b_{FL} and a_{Th} for each cube, in anticipation of the correlation studies to be described shortly.

As Table 2 shows, all four methods correctly discern the trend for cubes' energy to increase from the first to the seventh cube listed, but based on the last row of Table 2, B3LYP is on average four times as accurate as either HF or ORIENT, and MP2//B3LYP is more than twice as accurate as B3LYP. As seen in Table 1, B3LYP does considerably better at predicting H-bond lengths than either HF or ORIENT. This explains our reliance on B3LYP for our database calculations. It also suggests that there may be little value in trying to compare or predict B3LYP energies with better than around ±0.2 kcal/mol accuracy (or about 0.3 mH), at least for cubes, since this is the typical error implicit in B3LYP's E^0 values. Similarly, in considering predictions or comparisons for B3LYP's r_{OO} values, it should be kept in mind that these lengths differ from the MP2 lengths by errors on the order of 0.8 pm. The value 0.8 pm represents about 0.3% of the total O–O distance. For MP2 calculations, the O–O distances in this set of 84 H-bonds range from 258.4 to 303.1 pm.

4.2. Electronic energy

This project began when, after computing the B3LYP optima for all 14 inequivalent water cubes, we noticed that their energies fell neatly into five narrow strata. Except for the highest energy pair of cubes, which have *C*4 symmetry and are the only cyclic-component-containing cubes,[1] the strata correspond to the contiguity of the cubes. Fig. 4 shows the result of a least squares fit for E^0 vs. cg for the full data set of 14-points. It also shows the least squares line for the subset of 12-points that omits the two CCC cubes (the two data points at the upper right corner of Fig. 4). Correlation coefficients for E^0 vs. cg are 0.8909 for the 14-point set and 0.9838 for the 12-point subset. RMS ΔE^0 values[2] are 2.57 and 0.74 mH, respectively.

We wondered whether the inclusion of other connectivity parameters could improve the fit. With a little experimentation, we found that a linear combination of b_{FL} and a_{Th} does an excellent job of predicting E^0 for this data set. Correlation coefficients are 0.9973 for the full data set and 0.9983 for the 12-point subset. RMS ΔE^0 values are 0.43 and 0.26 mH, respectively. Taking the *D*2*d* cube as the zero point for energy, the best-fit plane for the 14-point set is given by the formula (in mH)

$$E^0 = 8.96 - 2.23b_{FL} + 1.05a_{Th}, \quad (1)$$

[1] The diagram labeled *C*4 in Fig. 3 is one of the CCC cubes; the other is obtained by reversing the directions of the four H-bonds on the inner square of that diagram.

[2] RMS ΔE^0 in this context is the RMS of the vertical distances from the data points to the best-fit line or surface. It is unrelated to the RMS ΔE^0 of Table 2, which refers to comparing MP2 to other methods.

Table 2
Electronic energy of seven water cubes by various methods Energy in kcal/mol

Cube	cg	b_{FL}	a_{Th}	MP2	MP2//B3LYP	B3LYP	HF	ORIENT
D2d	1	4	0	0.	0.	0.	0.	0.
S4	1	4	0	0.008	−0.011	0.173	0.112	0.070
Ci	2	2	0	3.048	3.034	2.927	2.830	2.596
Cs	3	2	2	4.052	3.958	3.976	3.766	3.474
C1-Z	4	1	4	6.709	6.607	6.436	5.951	5.809
C1-Y	4	1	6	7.891	7.662	7.609	7.089	6.751
C4	4	0	8	11.102	11.083	11.209	9.638	9.967
RMS ΔE^0				–	0.078	0.188	0.764	0.810

and for the 12-point set it is

$$E^0 = 9.44 - 2.32b_{FL} + 0.87a_{Th}. \quad (2)$$

Because several descriptors were tried before settling on b_{FL} and a_{Th}, we considered the possibility of a Bonferroni error. On approach to this is to compute a *p*-value for each descriptor in the model. Paraphrasing Weisberg [28], p. 18, the *p*-value can be defined as the conditional probability that the RMS value of ΔE^0 would increase by as much or more than it does when the index descriptor is deleted from the model, on the condition that the descriptor under consideration is actually irrelevant. *P*-values for both b_{FL} and a_{Th}, for both the 14- and the 12-point data sets, were under 0.0001, making an accidental correlation unlikely.

A statistical approach to the cubes is limited by the fact that the cube's high degree of symmetry restricts the number of distinct cube PWCs to just fourteen. There are 103 distinct pentagonal prism PWCs[3] and several hundred each for the octahedron and the hexagonal prism. We ultimately computed B3LYP optima for 27 pentagonal prisms, of which 23 contain only acyclic components. Given the high correlation of E^0 with b_{FL} and a_{Th} for the cubes, we tried this descriptor set first, for the pentagonal prisms. Correlation coefficients for E^0 vs. the two-descriptor set $\{b_{FL}, a_{Th}\}$ were 0.9845 for the full set and 0.9844 for the non-CCC subset. RMS ΔE^0 values were 1.23 mH for the 27-point set and 1.05 mH for the 23-point non-CCC subset.

By examining pairs of pentagonal prism clusters that had the same values of b_{FL} and a_{Th} but large differences in E^0, one difference stood out consistently: the higher energy cluster in each pair had more of its symmetric angles occurring in the pentagons rather than in the squares. We then included sy_5 as a third descriptor and recomputed the correlations. The correlation coefficients for E^0 vs. $\{b_{FL}, a_{Th}, sy_5\}$ were 0.9965 for the full set and 0.9985 for the non-CCC subset; RMS ΔE^0 values were 0.60 and 0.34 mH, respectively. For the non-CCC subset, the best fitting model is the formula (in mH)

$$E^0 = 7.17 - 1.78b_{FL} + 1.35a_{Th} + 0.80sy_5, \quad (3)$$

where the zero of energy is the energy of the lowest-energy $(H_2O)_{10}$ found, which is shown as Fig. 5(a).

We considered several ways of approaching the problem, apparent for both the cube and the prism, that our modeling efforts fared significantly better when the CCC clusters were omitted. An 'outlier' can be defined, admittedly vaguely, as a data point that does not fit with the rest of the data [28], p. 114. Among the 27 pentagonal prism PWC's we optimized, all four CCC examples have higher E^0 (range 17.83–27.25 mH) than all 23 non-CCC examples (range 0–17.75 mH). When $|\Delta E^0|$ values for the best-fit linear regression for $\{b_{FL}, a_{Th}, sy_5\}$ were listed for all 27 clusters, 3 of the four largest values belonged to CCC clusters, and for all four CCC clusters the predicted E^0 was smaller than the true E^0. For the CCC clusters, the differences between the true E^0 and what Eq. (3) predicts ranged from 0.95 to 2.26 mH, in contrast to the RMS ΔE^0 of 0.34 quoted earlier for the non-CCC clusters. Thus there are several ways in which the CCC clusters fail to fit with the rest of the data.

One approach is to reject the CCC clusters as outliers. Another is to expand the model with

[3] A method for enumerating the PWCs for a given polyhedral geometry is described in section 3 of Ref. [23].

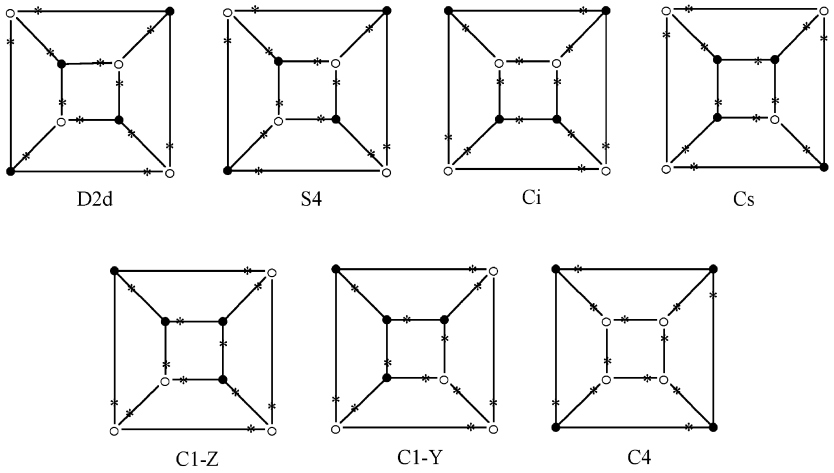

Fig. 3. H-connectivity diagrams for the seven cubes used in the benchmark comparisons.

additional variables, which distinguish CCC clusters. For example, simply including in the model the number of cycle-containing components, denoted n_{cy}, improves the correlation from 0.9965 to 0.9985 and improves the RMS ΔE^0 from 0.60 to 0.40. The formula for the best-fit four-descriptor model is

$$E^0 = 6.92 - 1.74b_{FL} + 1.37a_{Th} + 0.87sy_5 + 0.90n_{cy}. \quad (4)$$

P-values for all descriptors in Formulas (3) and (4) are <0.0001.

An analog of Eq. (4) for cubes can be obtained by including n_{cy} in the model; the resulting formula (cf. (2)) is

$$E^0 = 9.44 - 2.32b_{FL} + 0.87a_{Th} + 0.72n_{cy}. \quad (5)$$

It is not by chance that Formulas (5) and (2) are nearly identical. In going from the 12-point model (2) to the 14-point model (5), two data points are added which

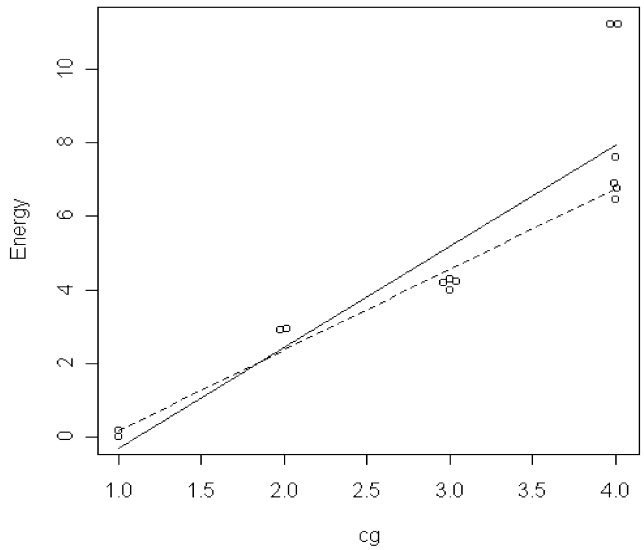

Fig. 4. E^0 vs cg for 14 water cubes with least squares lines for all 14 (solid) and for the 12 non-CCC cubes (dashed).

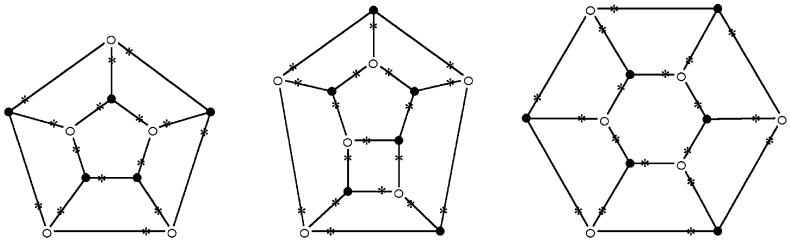

Fig. 5. Lowest energy pentagonal prism, octahedron, and hexagonal prism in database.

have the same b_{FL} and a_{Th} and nearly equal energies (call this common value E^*), and the new descriptor n_{cy} is added which has the same non-zero value n^* for the two new points but has the value 0 for the original 12-points. It is not hard to see that under these conditions, the coefficients in the new model will be the same as in the old for the original descriptors, and the coefficient on the new descriptor will equal $(E^* - E_{pred})/n^*$, where E_{pred} is the value predicted for E^0 for the new points by the old model. Because of this mathematical fact, the coefficient on n_{cy} in Eq. (5) is merely restating the value of the single number $E^* - E_{pred}$ and it should not be viewed as reflecting any statistically meaningful information. A similar but less severe objection can be raised to the inclusion of n_{cy} in Eq. (4), since its coefficient is determined by just four data points.

One other descriptor is worth mentioning because it correlates very strongly with E^0 in the case of the cubes, namely, the square of the dipole moment of the total cluster, denoted D^2_{tot}. For the set of 14 cubes, the correlation of E^0 with the three-descriptor set $\{b_{FL}, a_{Th}, D^2_{tot}\}$ is 0.9995, and RMS ΔE^0 is 0.14 mH, or less than the error implicit in B3LYP. However, D^2_{tot} is not a connectivity parameter, and it cannot be known accurately without first computing the cluster's optimum geometry and wavefunction, so it would not be usable to predict E^0. Still, this observation suggests that the 'problem' with the CCC clusters may be that their energies are raised more than their connectivity parameters can account for, by their exceptionally high dipole moments.

Because of the high energies of CCC clusters, they are among the least likely clusters to be discovered by experimentalists looking at PWCs, and they will occur with low frequency as components of bulk water. In this sense they may be less interesting to water theorists anyway. Combining this observation with the difficulty implicit in trusting conclusions to which only a few points contribute, we decided to proceed to the 12-mers by focusing on clusters with acyclic components only, making no claims that our findings apply to CCC clusters.

For the set of 20 non-CCC octahedra, the correlation coefficient for E^0 vs. $\{b_{FL}, a_{Th}, sy_5\}$ was 0.9987, and RMS ΔE^0 was 0.51 mH. For the 21 hexagonal prisms, which contain hexagons but no pentagons, sy_5 was replaced by sy_6. The correlation coefficient was 0.9946 and the RMS ΔE^0 was 0.86 mH. The respective formulas for E^0 (taking zero energy to be that of Fig. 5(b) and (c), respectively) were

$$E^0 = 3.73 - 1.00b_{FL} + 1.77a_{Th} + 0.79sy_5 \quad (6)$$

and

$$E^0 = 15.33 - 2.44b_{FL} + 1.33a_{Th} + 0.97sy_6. \quad (7)$$

P-values for all descriptors in Formulas (6) and (7) are <0.0002.

The minimum energy octahedron and hexagonal prism in our database, Fig. 5(b) and (c), are two of the structures found by Tsai and Jordan [13] when they used TIP4P and MP2//TIP4P to look for low energy $(H_2O)_{12}$'s. Fig. 5(b) is 'Cage 1' of Ref. [13] and Fig. 5(c) is denoted in Ref. [13] by its symmetry group, which is *D*3. Cage 1 and *D*3 were their lowest energy octahedron and hexagonal prism also, but one of the fused cube structures they studied by this method had slightly lower electronic energy than either PWC.

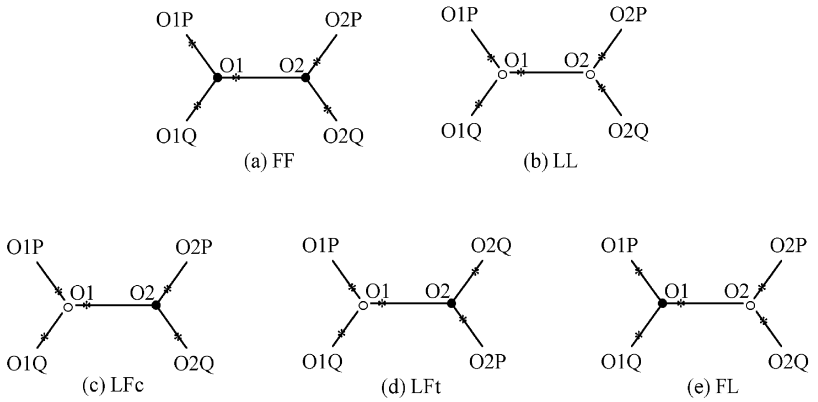

Fig. 6. Templates with secondary oxygen labels for the five bond types.

4.3. Bond lengths

Our total database contained 1311 H-bonds in 82 B3LYP-optimized PWCs. They ranged from 252.7 to 305.1 pm in length. We wondered how much of this variability could be accounted for by connectivity data. No distinction was made in the analysis between H-bonds from CCC and from non-CCC clusters.

An obvious place to start was with the types (i.e. F or L) of the donor and acceptor oxygens. These two variables accounted for 75% of the variance in bond length. The donor's type alone accounted for 49% and the acceptor's type alone, 37%. Before proceeding we split the H-bonds into sub-databases according to type (i.e. FF, FL, LF, or LL). The LF subset was subsequently split further, for reasons that will be explained shortly, into subsets called LFt and LFc. Demographics on each of the five sub-databases are summarized in Table 3.

We examined whether the type of the secondary oxygens, i.e. those oxygens H-bonded to either the donor or the acceptor O, was correlated with the bond length. To do this, we set up a labeled template for each of the five types.

Table 3
Summary of characteristics of the five databases of H-bond lengths from B3LYP-optimized PWCs

Type	# H-bonds	Range (pm)	Mean ± SD (pm)
FF	250	261.4–287.7	276.0 ± 5.48
FL	187	252.7–270.3	263.0 ± 3.90
LL	250	263.8–290.2	278.7 ± 5.54
LFt	375	272.0–305.1	289.7 ± 5.81
LFc	249	275.2–303.6	291.1 ± 5.85

The template for type FF bonds is shown as Fig. 6(a). The donor is labeled O1 and the acceptor is O2. Because O2 has type F, there is a unique O that accepts from O2, call it O2P, and a unique O that donates to O2 (and is not O1), call it O2Q. Both of the secondary oxygens at O1 donate to O1, but they can be distinguished by their positions relative to O2P and O2Q. We designate the one that is *cis* to O2P as O1P, and the one that is *trans* to O2P as O1Q. For programming purposes, *cis* was defined by the torsional angle O1P–O1–O2–O2P having smaller absolute value than a cutoff which was taken to be 60°. We denote the types of O1P, O1Q, O2P, O2Q as t_{1P}, t_{1Q}, t_{2P}, t_{2Q}; these are assigned the numerical values of 1 for F-type, 0 for L-type.

For our database of 250 type FF bonds, their lengths (r_{OO}) were strongly correlated with the 4-descriptor set $\{t_{1P}, t_{1Q}, t_{2P}, t_{2Q}\}$. All four descriptors had p-values <0.0001. The correlation was.9278 and RMS Δr_{OO} was 2.13 pm. This information and the coefficients in the best-fit formula are listed in Table 4.

A similar approach was taken for type LL bonds. Here O1 acts as acceptor from a unique secondary O, which we call O1Q, and as donor to a unique O (not O2), which we call O1P. We label the secondary oxygens at O2 so that O2P is *cis* to O1P and O2Q is *cis* to O1Q. The template is shown as Fig. 6(b). We again found a strong correlation between r_{OO} and the

Table 4
Summary of correlations for each of the five databases between H-bond lengths (r_{OO}) and the types of the secondary oxygens

Type	Δr_{OO} (pm)	r	Coefficients in best-fit formula				
			1	t_{1P}	t_{1Q}	t_{2P}	t_{2Q}
FF	2.13	0.928	274.08	−3.96	−4.56	5.18	4.31
FL	1.91	0.873	261.23	−3.28	−3.28	2.71	2.71
LL	2.40	0.903	278.35	−3.63	−6.72	4.84	2.85
LFt	3.20	0.836	291.30	−3.99	−4.56	3.73	4.72
LFc	2.89	0.872	293.72	−4.75	−6.44	3.68	3.87

set $\{t_{1P}, t_{1Q}, t_{2P}, t_{2Q}\}$, with all p-values <0.0001. See Table 4 for details.

For type LF bonds, we encounter a problem with label assignments. The template is 'overdetermined' in the sense that both O1 and O2 break symmetry and generate an assignment of secondary O labels. We define O1Q to be the unique donor to O1, and define O2P to be the unique acceptor from O2. Some bonds are like Fig. 6(c) and have O1P *cis* to O2P; these are segregated into the database of 'LFc' bonds. Those which are like Fig. 6(d) are classified as *trans* and are put into the set 'LFt'. For each of these sets, strong correlations were found for r_{OO} vs. $\{t_{1P}, t_{1Q}, t_{2P}, t_{2Q}\}$, with all p-values <0.0001. Details are in Table 4.

Type FL bonds pose the reverse problem, an underdetermined template (Fig. 6(e)). One arbitrary choice must be made in assigning, say, the label O1P. Then O2P is defined as *cis* to O1P, and O1Q and O2Q are the other secondary O's. However, a correlation with, say, t_{1P}, is meaningless, since O1P could just as easily have been O1Q. To avoid this, we take our descriptor set to consist of the two symmetric functions $\{t_{1P} + t_{1Q}, t_{2P} + t_{2Q}\}$.[4] The coefficients on t_{1P} and t_{1Q} are then automatically equal, as are the coefficients on t_{2P} and t_{2Q}. See Table 4.

We also examined whether the types of the eight tertiary oxygens correlated with the H-bond lengths. We first extended each of the labeled templates. This is illustrated for type FF bonds as Fig. 7. One problem that arises is that several of the labeled O's may coincide, e.g. if the face containing the angle O1P–O1–O2 is four-sided, then O2PP and O1P are the same O. To distinguish all possible templates reflecting variations in the positioning of 4-, 5-, and 6-sided faces, however, would lead to a proliferation of templates each serving a database too small to be statistically meaningful. We therefore ignored the problem beyond acknowledging its existence.

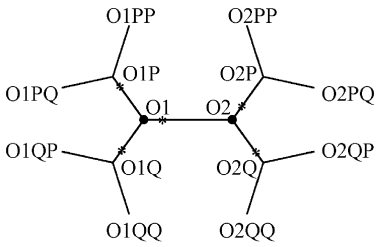

Fig. 7. Template with secondary and tertiary oxygen labels for FF-type bond.

For each of the five data sets, inclusion of the tertiary along with the secondary O types improved the correlation and Δr_{OO} considerably. Note that this raised the number of descriptors to 12, except for the FL bonds, where the symmetry requirement meant that we considered just six symmetric descriptors. *P*-values were computed for all descriptors, and in five instances the p-value exceeded 0.005; these variables were removed and the linear regression was done without them. Of the 49 descriptors kept in the five models, only three had p-values > 0.001, and 39 had p-values <0.0001, so a Bonferroni error is unlikely. Table 5 summarizes the results.

For the energy calculations, the databases were too small to test the validity of energy predictions based on formulas (1)–(7), but the H-bond databases were considerably larger. For each of the five bond types we split the database randomly, assigning one fifth of the bonds to a test set and the remainder to a training set. Descriptors designated in Table 5 as insignificant were deleted from the models. Best-fit formulas derived from the training set were applied to the test set. Table 6 shows the results. The results support the usefulness of these linear regressions for the prediction of H-bond lengths in PWCs, based on connectivity data alone. The largest test set Δr_{OO} value, which was 2.36 pm for the LFt bonds, represents 0.8% of the average LFt bond length.

Lastly, we wondered whether the five types of bonds actually represent distinct populations, or whether the categorization was merely a convenience to facilitate our analysis. The five sets were compared

[4] Symmetric means that their value does not depend upon how the assignment was made.

Table 5
Summary of correlations for each of the five databases between H-bond lengths (r_{OO}) and the types of the secondary and tertiary oxygens (* = descriptor deleted because $p > 0.005$)

Type	Δr_{OO} (pm)	r	Coefficients in best-fit formula											
			1	t_{1P}	t_{1Q}	t_{2P}	t_{2Q}	t_{1PP}	t_{1PQ}	t_{1QP}	t_{1QQ}	t_{2PP}	t_{2PQ}	t_{2QP}	t_{2QQ}
FF	1.26	0.975	274.5	−5.1	−5.3	5.6	5.5	−1.1	−1.6	−1.0	−2.1	1.0	1.2	1.4	1.4
FL	1.35	0.939	261.0	−3.9	−3.9	3.4	3.4	*	−1.1	−1.1	*	*	1.0	1.0	*
LL	1.55	0.962	278.8	−4.6	−6.6	4.9	3.7	−0.9	−1.2	−1.7	−1.1	1.2	1.4	1.3	*
LFt	2.14	0.932	291.4	−6.3	−5.8	5.3	6.9	−2.2	−1.2	−2.0	−0.9	*	1.2	2.2	1.9
LFc	1.96	0.945	291.4	−6.5	−7.5	5.8	6.7	−2.0	*	−2.1	−2.5	2.5	2.2	1.2	1.8

in pairs using Welch's *t*-test, to see whether the differences in mean and SD listed in Table 3 could be due to chance. Only the pairs FF and LL, and LFt and LFc, came close to having similar characteristics. For FF vs LL, $p < 10^{-6}$ and the 95% confidence interval for the difference in means was (1.65, 3.58). For LFt vs LFc, $p < 0.005$ and the 95% CI was (0.44, 2.31).

5. Discussion

5.1. Benchmarks

Our results support the existing literature on the usefulness of B3LYP and MP2//B3LYP, when used with a large basis set, for studies of small and medium-sized water clusters [17–19].

Bond lengths via HF differed by an RMS of over 7 pm from lengths via MP2, but what Table 2 does not show, is that there was little variability among the errors. For these 84 H-bonds, HF consistently gave r_{OO} values between 2 and 3% greater than MP2. Combined with the unexpected finding (Table 2) that O–O–O angles came closer to MP2 via HF than via B3LYP, this suggests that reasonable cube geometries might be obtained by shrinking or scaling the HF-optimized geometries by a fixed scaling factor of approximately 0.974. We have no reason at this time to expect that HF-optimization followed by scaling would work for non-cube water clusters, but it may be an approach worth investigating.

Table 6
Summary of correlations and Δr_{OO} (in pm) for training and test sets, for the five bond types

Type	Training set			Test set		
	#	r	Δr_{OO}	#	r	Δr_{OO}
FF	200	0.974	1.30	50	0.978	1.13
FL	150	0.942	1.33	37	0.924	1.48
LL	200	0.964	1.55	50	0.950	1.60
LFt	300	0.937	2.08	75	0.911	2.36
LFc	199	0.940	1.92	50	0.954	2.17

The ORIENT program calculates energy by combining electrostatic, induction, dispersion, exchange-repulsion and charge transfer terms, which are computed using parameters derived from pairwise interaction of water molecules. ORIENT did better than HF at predicting r_{OO} for water cubes, but fared worse with O–O–O angles and energy. We believe this illustrates the limitations of an approach based solely on the pairwise potential, when it comes to PWCs.

5.2. Electronic energies

For the four polyhedral databases, ANOVA calculations show that between 98.9 and 99.7% of the variance in E^0 for non-CCC clusters is accounted for by the connectivity parameters b_{FL}, a_{Th}, and sy_5 or sy_6 as appropriate. Our databases are admittedly small, and the correlations could decrease as more clusters are computed. Still, the very low *p*-values mean that the correlations are almost certainly not coincidental. In the best fit formulas (2), (3), (6), (7), the coefficient of b_{FL} is always negative and the coefficients of a_{Th} and sy_5 or sy_6 are always positive. Increasing b_{FL} lowers the energy of a cluster, while increasing a_{Th} or sy_5 or sy_6 raises it. Let us see why it might make sense for each of our parameters to correlate with E^0.

There is an inverse relationship between an H-bond's length and its strength. As Table 3 shows,

bonds of type FL are shorter and presumably stronger than the other types. The total number of H-bonds in a PWC $(H_2O)_n$ is fixed at $3n/2$ [23], Eq. (1.1), so as b_{FL} increases the number of other bond types decreases and the total E^0 goes down. To formalize this argument, let b_{FF} and E_{FF} denote the number and average energy of type FF bonds, and likewise for the notations b_{LL}, b_{LF}, E_{FL}, E_{LL}, E_{LF}. A crude model that omits cooperativity is

$$E^0 \approx (n)E_{H_2O} + E_{FF}b_{FF} + E_{FL}b_{FL} + E_{LL}b_{LL} + E_{LF}b_{LF}. \quad (8)$$

Using the relationships [23], Eq. (3.4) $b_{FF} = b_{LL} = (n/2) - b_{FL}$ and $b_{LF} = (n/2) + b_{FL}$, this becomes

$$E^0 \approx C_0 + C_1 b_{FL}, \quad (9)$$

where C_0 and C_1 are constants. This 'explains' why an approximately linear dependence of E^0 on b_{FL} might be anticipated.

Based on the coefficients of a_{Th} in Formulas (2), (3), (6), (7), each homogeneous angle costs a PWC anywhere from 0.87 to 1.77 mH. Since this represents an energy cost in excess of the cost of two FF or two LL bonds (which would be included in the b_{FL} term by the earlier argument), it can be viewed as measuring a cooperativity effect. With this interpretation, the key point is that 3 F-type or 3 L-type O's 'cooperate' differently from 3 O's in an inhomogeneous angle. However, a_{Th} is highly correlated with both b_{FL} and with other energy-relevant parameters such as D^2_{tot}. These correlations are mathematically unavoidable: as the number of homogeneous angles increases, individual F- and L-components become larger and there are fewer of them, and as components grow so does the total cluster polarization. Thus a_{Th} might be merely a marker for other energy-associated parameters.

Each symmetric angle in a pentagonal (respectively, hexagonal) face costs about 0.8 mH (respectively, 0.97 mH). Because of the relationships [23], Eq. (4.5) $sy_4 = 5 - sy_5$ for pentagonal prisms, $sy_4 = 6 - sy_5$ for octahedra, and $sy_4 = 6 - sy_6$ for hexagonal prisms, our correlations could just as easily have been done using sy_4 instead of sy_5 or sy_6. The energy cost should be viewed as the cost of the tradeoff of a symmetric angle in a 5- or 6-sided face for one in a 4-sided face. Interestingly, sy_m ($m = 5,6$) is not significantly correlated with b_{FL} or a_{Th} in our database.

Because the number of F-symmetric and of L-symmetric angles in m-sided faces ($m = 4,5$, or 6) is the same, we cannot tell from our data whether the energy-relevant motifs are the L-symmetric angles, the F-symmetric angles, or both. One observation supports the notion that the L-symmetric angles may be the key. Focusing on the H–O–H angles at L-type O's, there is little distortion from the monomer angle (which B3LYP computes as 105.0°): in this database of 437 such angles, the range is 103.1–108.9°. In four-sided faces these H's flare outward from the polyhedral surface, whereas in H–O–H angles that are part of five- or six-sided faces the H's move toward the interior of the face (cf. Fig. 2(a)). Conceivably, the latter arrangement has higher energy.

Extrapolating our results to polyhedral $(H_2O)_n$'s for $n > 12$, the lowest energy PWCs for a given polyhedral geometry are likely to among those which maximize b_{FL} while minimizing a_{Th} and sy_m ($m \geq 5$). These goals may require tradeoffs. For example, the PWC of Fig. 5(b) has $b_{FL} = 5$, which is the the maximum possible by Eq. (3.3) of Ref. [23], but it has $sy_5 = 2$; the next lowest energy octahedron among the 20 we optimized has $sy_5 = 0$ but $b_{FL} = 4$. It can easily be proved to be impossible for an octahedral PWC to have both $b_{FL} = 5$ and $sy_5 < 2$. Theorem 6 of Ref. [23] means that, for any polyhedral geometry, it is possible to distribute the pendent H's so that $a_{Th} = 0$. Because b_{FL} also counts the number of acyclic F-components, the combination of making $a_{Th} = 0$ while maximizing b_{Fl} can be thought of heuristically as spreading out or spacing apart the pendent H's as much as possible.

Among polyhedra with faces of 4–7 sides[5], there are four 'magic numbers' which allow for an arrangement that simultaneously optimizes all four parameters, i.e. in which $b_{FL} = n/2$, $a_{Th} = 0$, and $sy_5 = sy_6 = sy_7 = 0$. They are $n = 8$, 12, 16 and 24. For $n = 8$ these are the $D2d$ and $S4$ cubes, and for $n = 12$, the PWC of Fig. 6(c). For $n = 16$ and $n = 24$ the optimal PWCs by this criterion are shown in Fig. 8. We offer the conjecture that these will turn out to be the lowest energy PWCs for $n = 12$,

[5] With one known exception, unit cells of clathrate hydrates have faces with 4–7 sides [7].

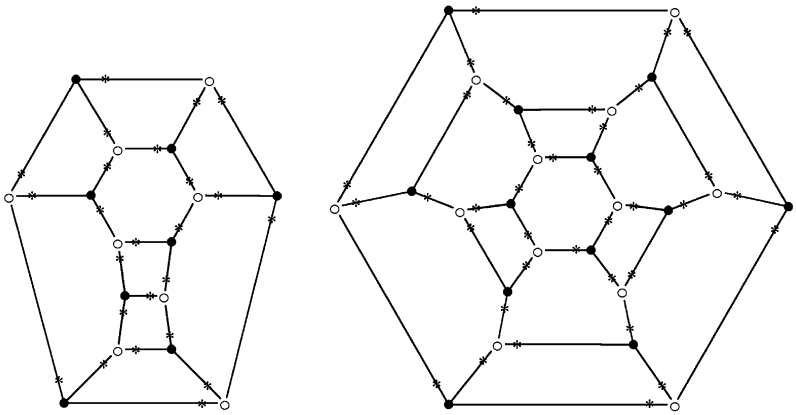

Fig. 8. Candidates for lowest energy polyhedral 16-mer and 24-mer.

16, and 24. The 24-mer has Th symmetry and its polyhedral geometry ($4^6 6^8$) is the unit cell of the clathrate hydrate of HPF_6 [7].

5.3. Bond lengths

In their studies of $(H_2O)_6$'s and $(H_2O)_7$'s Kim et al. [2,3] distinguished four H-bonding environments for a water molecule in a cluster, based its being donor in either one or two H-bonds, and acceptor in either one or two H-bonds. Two of their environments, 'dda' and 'daa', describe three-coordinated O's and correspond to our L-type and F-type. They showed that the H-bonding environment has a significant influence on the frequency and intensity of the IR vibrational modes associated with each monomer in a cluster, but did not focus on the relationship between H-bonding environment and H-bond length. Jordan [29] recognized three classes of H-bonds between three-coordinated oxygens, which in our terminology would be FL, FF + LL, and LF. Jordan observed that on average FL bonds are the shortest, followed by FF and LL, and then LF bonds. Table 3 shows that Jordan's classification extends nicely to the H-bonds in our database. Our database is large enough for the t-test to distinguish FF from LL and LFt from LFc, at least for H-bond lengths as computed by B3LYP. In their experimental and theoretical studies of C_6H_6–$(H_2O)_9$, Gruenloh et al., expanded Jordan's observation to include some H-bonds in which one of the O's is two-coordinated [30]. If 'type B' (for 'both') denotes a two-coordinated O that is a single donor and single acceptor, Gruenloh et al., found that H-bonds of types FL, FB, and BL were strongest and shortest; type LF were longest and weakest; and types LB and BF in between. This illustrates that our results are likely extendable, with modification, to cage-like water clusters that are not strictly polyhedral.

For all five bond types, the type of each secondary oxygen has a significant ($p < 0.0001$) influence on the bond length. From Table 4 a consistent feature emerges: H-bond length is decreased when the donor has more F-type rather than L-type nearest neighbors, and increased when the acceptor has more F-type neighbors. Table 5 shows that, omitting a few whose contribution is not statistically significant, the tertiary O's follow the same pattern. Having more F-type tertiary O's at the donor (respectively, acceptor) decreases (respectively, increases) r_{OO}. The best-fit formula coefficients in Table 5 show that the effect on r_{OO} of exchanging an F-type tertiary O for an L-type is on the order of 0.9–2.5 pm, whereas for a secondary O the effect size ranges from 3.4 to 7.5 pm.

For the five bond types, the secondary and tertiary O types account for between 87 and 95% of the variance in r_{OO}. Many other structural parameters are correlated with r_{OO} in this database and can help to account for the remaining variance. These include local angles such as O1–O2–O2P, the distances from

O1 to O1P and O1Q and from O2 to O2P and O2Q, and the length of the projection of the cluster's total dipole moment (vector) onto a unit vector in the direction of O1–O2. Several of these may also correlate with each other and with the secondary and tertiary O types. As none of these are connectivity parameters, an analysis of their relevance to r_{OO} lies beyond the scope of this article.

6. Conclusions

1. Benchmark calculations at the MP2 level support the usefulness of B3LYP/6-311++G** and MP2//B3LYP for the study of PWCs. Specifically, in 7 water cubes, B3LYP nearest neighbor O–O distances differ by an RMS of 0.8 pm from MP2, and B3LYP electronic energy differs by an RMS of 0.3 mH from MP2.
2. Based on six examples studied, PWCs that contain cyclic components (CCC) have exceptionally high energies and can behave as outliers with respect to energy-connectivity correlations obeyed by PWCs that contain only acyclic components.
3. Electronic energy (E^0) as computed by B3LYP correlates strongly with certain connectivity parameters, i.e. properties of a PWC that depend solely on the H-bonding pattern. For the 12 non-CCC water cubes, b_{FL} and a_{Th} account for 99.7% of the variance in E^0. For 23 non-CCC pentagonal prisms and for 20 non-CCC octahedra studied, the set $\{b_{FL}, a_{Th}, sy_5\}$ likewise accounts for 99.7% of the variance in E^0. For 21 non-CCC hexagonal prisms, $\{b_{FL}, a_{Th}, sy_6\}$ accounts for 98.9% of the variance in E^0. Within each of these four families of PWCs, E^0 is negatively correlated with b_{FL} and positively correlated with a_{Th} and with sy_5 or sy_6. *P*-values for these energy descriptors are <0.0002.
4. Based on 1311 H-bonds in B3LYP-optimized PWCs, five H-bond types, defined by the types (i.e. presence or absence of pendent H's) of the donor and acceptor and the positioning of certain H's, may be identified as comprising distinct populations in PWCs. Within each bond type, H-bond length (r_{OO}) is strongly correlated ($p < 0.0001$) with the types of the four secondary O's and, less strongly, with the types of the eight tertiary O's. Having F-type (i.e. carries a pendent H) O's within one or two H-bonds of the donor O decreases r_{OO}, while having F-type O's one or two H-bonds from the acceptor increases r_{OO}. Linear regression applied to randomly selected training subsets permits prediction of r_{OO} values for the complementary test subsets with an accuracy ranging from 1.1 to 2.4 pm, or 0.4–0.8% of r_{OO}.

References

[1] J. Rodriguez, D. Laria, E.J. Marceca, D.A. Estrin, J. Chem. Phys. 110 (18) (1999) 9039.
[2] J. Kim, D. Majumdar, H.M. Lee, K.S. Kim, J. Chem. Phys. 110 (18) (1999) 9128.
[3] J. Kim, K.S. Kim, J. Chem. Phys. 109 (14) (1998) 5886.
[4] J.M. Pedulla, K. Kim, K.D. Jordan, Chem. Phys. Lett. 291 (1998) 78.
[5] J.M. Pedulla, K.D. Jordan, Chem. Phys. 239 (1998) 593.
[6] C.J. Tsai, K.D. Jordan, J. Chem. Phys. 95 (5) (1991) 3850.
[7] D.W. Davidson, Clathrate hydrates, in: F. Franks (Ed.), 2nd ed, Water, A Comprehensive Treatise, vol. 2, Plenum Press, New York, 1976 Chapter 3.
[8] A. Khan, J. Chem. Phys. 110 (24) (1999) 11884.
[9] K. Laasonen, M.L. Klein, J. Phys. Chem. 98 (1994) 10079.
[10] A. Khan, Chem. Phys. Lett. 21 (1994) 443.
[11] S. Wei, Z. Shi, A.W. Castleman Jr., J. Chem. Phys. 94 (4) (1991) 3268.
[12] L.S. Sremaniak, L. Perera, M. Berkowitz, J. Chem. Phys. 105 (9) (1996) 3715.
[13] C.J. Tsai, K.D. Jordan, J. Phys. Chem. 97 (1993) 5208.
[14] D.J. Wales, I. Ohmine, J. Chem. Phys. 98 (9) (1993) 7245–7257.
[15] N. Yoshii, H. Yoshie, S. Miura, S. Okasaki, J. Chem. Phys. 109 (12) (1998) 4873.
[16] A.G. Kalinchev, J.D. Bass, Chem. Phys. Lett. 231 (1994) 301.
[17] A. Smith, M.A. Vincent, I.H. Hillier, J. Phys. Chem. A 103 (1999) 1132.
[18] M. Sprik, J. Hutter, M. Parrinello, J. Chem. Phys. 105 (3) (1996) 1142.
[19] S.S. Xantheas, J. Chem. Phys. 102 (11) (1995) 4505.
[20] J.J. Nova, C. Sosa, J. Chem. Phys. 99 (1995) 15837 Dimer benchmark.
[21] K. Kim, K.D. Jordan, J. Phys. Chem. 98 (1994) 10089 Dimer Benchmark.
[22] J. Kim, J.Y. Lee, S. Lee, B.J. Mhin, K.S. Kim, J. Chem. Phys. 102 (1995) 310.
[23] D.J. Anick, Polyhedral Water Clusters, I: Formal Consequences of the Ice Rules, (this issue).
[24] Jaguar 4.0, Schrodinger, Inc., Portland, OR, 2000.
[25] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, V.G. Zakrzewski, J.A. Montgomery, Jr., R.E. Stratmann, J.C. Burant, S. Dapprich, J.M. Millam, A.D. Daniels, K.N. Kudin, M.C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C.

Adamo, S. Clifford, J. Ochterski, G.A. Petersson, P.Y. Ayala, Q. Cui, K. Morokuma, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J. Cioslowski, J.V. Ortiz, A.G. Baboul, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, J.L. Andres, C. Gonzalez, M. Head-Gordon, E.S. Replogle, J.A. Pople, GAUSSIAN98, Revision A.9, Gaussian, Inc., Pittsburgh, PA, 1998.

[26] A.J. Stone, A. Dullweber, M.P. Hodges, P.L.A. Popelier, D.J. Wales, ORIENT: A program for studying interactions between molecules, Version 3.2, University of Cambridge, 1995, Available at http://fandango.ch.cam.ac.uk.

[27] R Poject for Statistical Computing, http://www.r-project.org.

[28] S. Weisberg, Applied Linear Regression, Wiley, New York, 1980.

[29] K.D. Jordan, Oral communication.

[30] C.J. Gruenloh, J.R. Carney, F.C. Hagemeister, T.S. Zwier, J.T. Wood III, K.D. Jordan, J. Chem. Phys. 113 (6) (2000) 2290.