O−H Stretch Modes of Dodecahedral Water Clusters: A Statistical Ab Initio Study

David J. Anick*

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

Received: October 3, 2005; In Final Form: February 13, 2006

Infrared frequencies calculations were carried out for 20 $(H_2O)_{20}$ water clusters obeying the 5^{12} dodecahedral geometry, optimized at the B3LYP/6-311++G** level. Their combined spectra contained 800 O−H stretch modes, ranging from 2181 to 3867 cm^{-1} (unscaled), which were treated and studied as a database. Of these, 752 modes (94%) could be assigned to a single dominant O*−H* stretch. These 752 were classified into five subdatabases depending on the local H-bond type of the dominant stretch. The frequency (ν) was correlated with the O−H* distance (b_{OH}), with H-bond length (R_{OO}) where applicable, and with other variables. The parameter b_{OH} alone accounted for 96−99% of the variance in ν for stretches in H-bonds. The correlation with R_{OO} is substantially weaker. Normal modes were classified as "high ratio" or "low ratio" depending upon the mode's distribution of kinetic energy among the O−H bonds. High-ratio modes (389 modes, or 49% of our sample) are modeled well as a single oscillator undergoing small perturbations by weak coupling from other oscillators. Low-ratio modes involve strong coupling with at least one other O−H stretch for which b_{OH2} is close to b_{OH}. The IR intensities of modes vary widely but can be explained in terms of a single equation giving dipole moment derivatives as a function of b_{OH}. For the lowest-energy $(H_2O)_{20}$ clusters, their IR stretch spectra contained eight distinguishable absorption bands. An explanation for eight bands in terms of the theory of polyhedral water clusters is offered.

Introduction

Infrared spectroscopy is the principal experimental method through which small- and medium-sized water clusters are probed. Experimental setups in which IR spectra of water clusters have been obtained include free $(H_2O)_n$ for $n \leq 10$,[1−9] benzene−water systems,[10−12] water in helium droplets,[13−14] and water in parahydrogen.[15] Miyazaki et al.[16] and Shin et al.[17] obtained spectra for the protonated species $H^+(H_2O)_n$, $6 \leq n \leq 27$, and Mitsuhiko et al.[18] did likewise for the $[C_6H_6-(H_2O)_n]^+$ system. In all of these, a structural shift from a netlike geometry with some DA (single donor single acceptor) waters to a cagelike geometry with no DA water occurred at the magic number $n = 21$. The lowest-energy geometry for $H^+(H_2O)_{21}$ is believed to be a dodecahedral cage with a single H_2O inside.[19]

The water monomer has a bending mode (ν_2) at 1595 cm^{-1} and symmetric (ν_1) and antisymmetric (ν_3) stretch modes at 3657 cm^{-1} and 3756 cm^{-1}, respectively.[20] Kim et al [21,22] assigned the 2n stretch modes of various $(H_2O)_n$'s ($n = 6,7,8$) to be either ν_1 or ν_3 modes and then examined how the frequencies and intensities for ν_1 and ν_3 modes varied with their H-bonding environment. They admitted that the assignment sometimes became ambiguous when multiple protons shared similar parameters. Lee et al.[23] extended this type of analysis to a set of 25 clusters ranging from the dimer to decamers. Current thinking, however, posits that the concept of ν_1 and ν_3 for stretch modes in $(H_2O)_n$'s is appropriate mainly for H_2O units in which neither proton is H-bonded (e.g., Miyazaki et al.[16] identified ν_1 and ν_3 signals from terminal H_2O's in chainlike $H^+(H_2O)_n$'s for $n < 10$). We view water (focusing on stretches only) as a collection of coupled O−H dipole oscillators that just happen to come two to a molecule.[24,25] Normal modes can involve many O−H dipoles and can defy classification as either "symmetric" or "antisymmetric". Coupling between two dipoles depends in part on their intrinsic frequencies, which depend in turn on their O−H distances. For H-bonds in organic crystals and in water clusters, the intrinsic frequency of a stretch mode is known to be strongly negatively correlated with the O−H distance b_{OH} and positively correlated with the O−O distance R_{OO}.[26]

Buck et al.[9] studied spectra of neutral $(H_2O)_n$, $8 \leq n \leq 10$, both computationally and experimentally. Spectra of some clusters, such as the S_4 and D_{2d} cubes, have just a few absorption bands separated by IR-inactive regions (cf. Figure 1 of ref 9). They called such clusters "crystalline." The absorption bands in the spectra of other clusters appeared spread out across the entire stretch region; such clusters were called "amorphous." Buck et al. implicitly recognized that the correlation between R_{OO} and ν is tight enough that the stretch region of a cluster's IR spectrum can be thought of as displaying its distribution of R_{OO} (i.e., the distribution of H-bond lengths), and this concept was further developed by Sadlej et al.[25,27] They recognized that one major contributor to R_{OO} is the donor's local bonding type (DDA, DAA, DA, etc.). Crystalline and amorphous clusters are those with closely grouped or with spread out R_{OO} values, respectively. In some cases, a crystalline pattern results from a neat divide between H-bonds whose donor has the DDA pattern and those whose donor is DAA. We will build on their work by incorporating advances in the theory of bond lengths for polyhedral water clusters[28−29] that go well beyond consideration of the donor alone and, most importantly, by adding a degree of statistical rigor to various claims by examining them quantitatively for a large database.

The 5^{12} dodecahedral $(H_2O)_{20}$ is of great importance both experimentally and theoretically. Experimentally, it is the principal building block of von Stackelberg's Structure I

* E-mail: David.Anick@rcn.com.

10.1021/jp055632s CCC: $33.50 © 2006 American Chemical Society
Published on Web 03/24/2006

clathrate hydrates,[30] including the commercially important methane hydrate.[31] We have already mentioned the filled protonated 5^{12},[19] and $(H_2O)_n$ cages surrounding cations such as Cs^+ also show a magic number at $n = 20$, presumably due to the 5^{12} geometry.[32]

For this ab initio study, we explored the stretch region (defined as 2000−4000 cm^{-1}) of the predicted IR spectra of 5^{12} dodecahedral $(H_2O)_{20}$ clusters. Over 30 000 symmetry-distinct H-bond arrangements are possible for 5^{12} water clusters.[33] Although some of these are too unstable to represent actual clusters,[28−29,34] the large number of stable $(H_2O)_{20}$ geometries invites a sampling and database approach.

We built upon a database of sixteen 5^{12} $(H_2O)_{20}$ clusters that was described in a previous work.[35] That database contained five clusters whose H-bond arrangements were selected randomly, four whose energies were at or near the lowest energies possible for a 5^{12}, and seven which were chosen to make the spread of cluster energies approximately evenly spaced. For this project, we deleted one cluster that had C_i symmetry (explanation below), and generated five more for a total of 20. The range of cluster electronic energies via B3LYP is 30 kcal/mol. Among the four lowest-energy clusters, the spread is under 0.2 kcal/mol. Our project was to study the database's 800 (20 clusters × 20 waters per cluster × 2 modes per water) O−H stretch modes.

We used B3LYP/6-311++G** for all ab initio calculations. This model has undergone extensive validation for water cluster calculations relative to MP2.[36−42] Infrared spectra computed via DFT typically require scaling; i.e., predicted frequencies need to be multiplied by a factor, usually between 0.9 and 1, to bring them into better agreement with experiments.[43] For B3LYP with a large basis set, 0.961 may be used.[44]

Questions we set out to explore about $(H_2O)_{20}$ IR spectra included the following.

- Do most stretch modes involve large displacement of one particular proton, with other protons moving much less, or are large coordinated movements of two or more protons common? How well do the 40 protons correspond to the 40 stretch modes?
- When a mode is identifiable as primarily associated with one particular O−H* stretch, what percent of the variance in stretch mode frequency can be accounted for by b_{OH} ? By R_{OO}? Are there other significant determinants of stretch mode frequency?
- When a stretch mode does involve large coordinated movement of two or more protons, what factors determine whether two O−H stretches are "coupled" in this sense?
- How can we account for the tremendous variability in IR intensity among stretch modes?
- For the 5^{12} $(H_2O)_{20}$'s of lowest energy, what features might be expected in their IR spectra? Specifically, would these be of a crystalline or of an amorphous nature?

Review of Polyhedral Water Clusters and Two Theorems

The 5^{12} dodecahedral $(H_2O)_{20}$ is an example of a polyhedral water cluster (PWC). PWCs are defined as cagelike $(H_2O)_n$ clusters in which every O is 3-coordinated. The theory of PWCs is simplified because there are just two local motifs: the donor−acceptor−acceptor O, denoted DAA or F for its free (or non-H-bonded or dangling) H, and the donor−donor−acceptor O, denoted DDA or L for its non-H-bonded lone pair.

The F and L designations classify the H-bonds of a PWC into four types by "donor type acceptor type"; e.g., FL means donor is F and acceptor is L. For some purposes it helps to separate the LF bonds further into cis and trans subsets,[28−29,34,40] but we will not need this distinction. An O−H* covalent bond in a PWC can therefore occur in any of five environments: H* can be a free H, or it can be in an FF, FL, LL, or LF H-bond. This classification into five O−H types will be the basis of our study of O−H stretch modes.

Several authors have noted that in PWCs, FL H-bonds are the shortest, LF are the longest, and FF and LL have intermediate length.[28−29,34,40] The types of the donor and acceptor account for 75% of the variance in H-bond length (R_{OO}).[40] Accordingly, b_{OH} is longest for FL and shortest for LF, with intermediate lengths for FF and LL.

The types (i.e., F or L) of the donor's and the acceptor's nearest neighbor O's also affect R_{OO}. Specifically, L neighbors at the donor and F neighbors at the acceptor will raise R_{OO}. Reference 40 computed the contribution of each neighbor's type for each bond type, but Singer, Kuo, et al. simplified this by defining a single "topological index", denoted ξ. It is defined (eq 1 of ref 28) by:

$$\xi = (\#\text{ of L neighbors at donor}) + (\#\text{ of F neighbors at acceptor}) \quad (1)$$

Because the donor and the acceptor each have two neighbors, the value of ξ is 0, 1, 2, 3, or 4. Lower values of ξ correspond to shorter H-bonds. Each increase of 1 in the value of ξ adds 3 to 7 pm to R_{OO}.[40] Thus, 20 subtypes of H-bonds have been identified: four donor−acceptor combinations times five possible values of ξ for each.

The number of FF bonds, denoted B_{FF}, is a potent predictor of the electronic energy of a PWC, with each additional FF bond adding about 4 kcal/mol.[28] A related parameter is A_{TH}, the total number of "homogeneous angles", i.e., the number of times that a motif of a chain of three adjacent F's or a chain of three adjacent L's occurs. Configurations having the lowest energy among those with a given polyhedral geometry will have no homogeneous angles,[40] i.e., $A_{TH} = 0$. For the 5^{12} geometry, the minimum possible value of B_{FF} is 3,[45] so the lowest-energy 5^{12} configurations will have $B_{FF} = 3$ and $A_{TH} = 0$.

Because it will be part of our explanation of the predicted spectra of the 5^{12} clusters, we offer two theorems, with proofs, about mathematical relationships among B_{FF}, A_{TH}, and ξ.

Theorem 1. For a PWC having the 5^{12} geometry, if $B_{FF} = 3$, then $A_{TH} = 0$.

Proof. The number of LL bonds, denoted B_{LL}, always equals B_{FF}.[45] Let C_j denote the number of FF or LL bonds in the jth polyhedral face (the index j will run from 1 to 12). The total number of FF or LL bonds is $(\Sigma\, C_j)/2$, because each edge gets counted twice. Thus

$$B_{FF} + B_{LL} = \left(\sum_{j=1}^{12} C_j\right)/2, \qquad \text{or} \quad (2)$$

$$4B_{FF} = \sum_{j=1}^{12} C_j \quad (3)$$

Any polyhedral face with an odd number of sides must contain at least one FF or LL bond.[45] Because all 12 faces are 5-sided, we have $C_j \geq 1$ for each j, $1 \leq j \leq 12$. The right-hand side of eq 3 is therefore at least 12, and it equals 12 if and only if every one of the C_j's is exactly 1.

Because $B_{FF} = 3$, eq 3 tells us that all the C_j's equal 1; i.e., no face has more than one FF or LL bond. But if $A_{TH} > 0$, i.e., if there is a homogeneous angle, then the face that contains

that angle will have two or more FF's (or two or more LL's). It follows that $A_{TH} = 0$.

We remark that the above argument would apply whenever $B_{FF} = f_{odd}/4$, where f_{odd} is the number of polyhedral faces having an odd number of edges. (B_{FF} can never be smaller than $f_{odd}/4$.[45]) In the proof, the key idea would be that $C_j = 1$ if the jth face has an odd number of sides and $C_j = 0$ if the jth face has an even number of sides. In particular, Theorem 1 is also true for the $5^{12}6^2$ geometry that commonly occurs as a clathrate hydrate cell.

Theorem 2. In a PWC having $A_{TH} = 0$, any FF or LL bond has $\xi = 2$; any FL bond has $\xi = 2$, 3, or 4; and any LF bond has $\xi = 0$, 1, or 2.

Proof. When the donor and acceptor of an H-bond in the PWC are both F, then all four nearest neighbor O's must be L. Otherwise, the F neighbor together with the donor and acceptor would form a chain of three F's, contradicting the premise that $A_{TH} = 0$. Then $\xi = 2 + 0 = 2$. Likewise, for an LL bond, all four neighbors must be F, and thus $\xi = 0 + 2 = 2$. For an FL bond, at least one of the donor's neighbors must be L: otherwise the donor O together with the two neighbors would form a chain of three F's. Likewise, at least one of the acceptor's neighbors must be F. Equation 1 gives $\xi = (1 \text{ or } 2) + (1 \text{ or } 2) = 2$, 3, or 4. The proof for LF bonds is similar.

Putting these theorems together, we see that for a low-energy 5^{12} or $5^{12}6^2$ PWC having $B_{FF} = 3$, the number of H-bond subtypes drops from 20 to just 8 (three FL's, one FF, one LL, and three LF's). This will be a key insight when we look at the spectra of low-energy PWCs.

The symmetry groups that are possible for a neutral optimized 5^{12} $(H_2O)_{20}$ are C_1, C_i, C_5, and C_{5i}. There are only 12 configurations that have 5-fold symmetry: eight have $b_{FF} = 10$ and their electronic energy E^0 lies 30 kcal/mol or more above the lowest-energy 5^{12} arrangement, and four have $b_{FF} = 5$ with E^0 being 20 kcal/mol above the lowest-energy 5^{12}. None of these were part of our database. The normal modes for a C_i-symmetric $(H_2O)_{20}$ come in pairs consisting of an IR-inactive symmetric mode and an IR-active antisymmetric mode. Because this description is so different from the case of the typical C_1-symmetric cluster (although the actual spectrum does not look substantially different), and because C_i-symmetric clusters are a tiny minority of all 5^{12} PWCs, we decided to exclude C_i-symmetric clusters from our database. A C_i-symmetric cluster has an even number of FF bonds, and because $B_{FF} \geq 3$, it follows that B_{FF} is at least 4. Thus, none of the excluded C_i-symmetric clusters are among the lowest-energy group for which $B_{FF} = 3$.

Methods

Calculations were done on a Parallel Quantum Solutions (PQS) QuantumCube, using PQS parallel software.[46] Optimization was done in inverse cluster coordinates using the OPTIMIZE algorithm.[47] Setting the optimization "scale" factor to 5.0 gave efficient convergence. Initial geometries were obtained using the approximation algorithm described in ref 29. The PQS HESS and FREQ routines compute normal modes and frequencies analytically based on the harmonic approximation. Statistics were done with R-project software.[48]

Results and Discussion

Assignment of Normal Modes To O−H Stretches. A $(H_2O)_{20}$ cluster has 40 stretch modes and 40 protons in 40 O−H covalent bonds. For a general linear system of coupled oscillators with d degrees of freedom and d normal modes, one can picture two extremes. At one extreme, all or most oscillations are strongly coupled and most or all normal modes involve significant displacement in all d degrees of freedom. At the other extreme, coupling is weak and each degree of freedom oscillates essentially independently. At this extreme, each normal mode will be dominated by a single degree of freedom, and each degree of freedom will feature prominently in only one normal mode.

As an interacting system of 40 interacting O−H stretches (setting aside for now bends and other motions), where do dodecahedral $(H_2O)_{20}$ clusters fall between these two extremes? A quick review of the proton displacements for the lists of modes reveals that most modes feature primarily a single "dominant" proton (or O−H* stretch), whereas many had two with comparable displacements, and a few had three or four.

To make the analysis quantitative, we review the classical mechanics concept of the fraction of a mode's energy attributable to each component of an oscillating system. Let masses $m_1, \cdots, m_d$ undergo coupled oscillations, with equilibrium positions $\vec{b}_1, \cdots, \vec{b}_d$, and let $\vec{a}_{ki}$ denote the peak displacement of the ith mass for the kth normal mode whose frequency is $\omega_k/2\pi$. Then the motion of m_i for the kth normal mode is given by

$$\vec{r}_{ki}(t) = \vec{b}_{ki} + C\vec{a}_{ki}\sin(\omega_k t)$$

At $t = 0$, all of the mode's energy is kinetic, and the fraction of kinetic energy carried by m_i is

$$e_{k,i} = m_i|\vec{a}_{ki}|^2 \Big/ \sum_{j=1}^{d} m_j|\vec{a}_{kj}|^2 \qquad (4)$$

Despite the reliance on classical mechanics, eq 4 gives a reasonable way to compare the activity of a proton (or O−H stretch) across multiple modes and to compare how a mode treats various protons. For each $(H_2O)_{20}$ cluster, we computed $\{e_{k,i}\}$, obtaining a 40 × 40 matrix E whose (k,i) entry is $e_{k,i}$. Because $m_1 = \cdots = m_{40}$ for our 40 protons, the kth row of E is (a scalar multiple of) the vector of squared peak displacements of the protons, for the kth mode. The ith column of E approximately describes how the energy of an impulse imparted to the ith proton H_i when the system is at rest would distribute among the modes. This approximation would be exact in a hypothetical system where all proton motions were strictly in line with their O−H bonds and the oxygen nuclei were infinitely heavier than protons (rather than 16 times heavier).

As an illustration with these conventions, each proton of the H_2O monomer can be said to carry 48% of the energy for mode ν_1 and 46% for ν_3. The remaining energy is accounted for by motion of the O. In a normal mode, except for symmetric motions on a single O (like ν_1), any H motion will be balanced by covalently bonded O motion that is approximately 1/16 as great, so the sum of the energy fractions assignable to all protons (i.e., row sums for the matrix E) will typically be around 16/17, or 0.94, rather than 1.

Let $I(k)$ or $I_1(k)$ denote the index for which $e_{k,I(k)}$ is the largest entry in the kth row of E, and put $P_1(k) = e_{k,I(k)}$. $I(k)$ is (the label of) the "dominant" or "primary" proton for mode k, and $P_1(k)$ is the fraction of mode k's energy that is carried by its primary proton. Among the 800 modes in our database, 553 (69%) have $P_1(k) > 0.5$, i.e., more than half of the mode's energy is carried by its primary proton. If the cutoff is lowered to 8/17, meaning that at least half the energy is in the H and O of a single O−H* stretch, the count rises to 594 (74%).

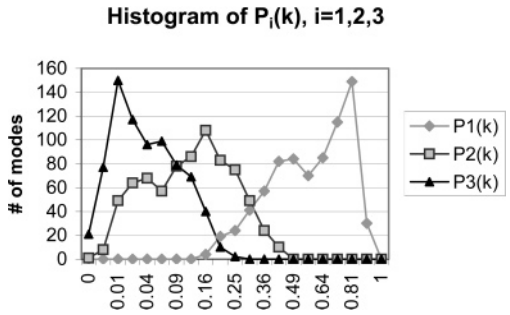

Figure 1. Number of modes for which P_i falls into each interval, $i = 1,2,3$.

Let $I_2(k)$ be the index for which $e_{k,I_2}(k)$ is the second largest entry in the kth row of E, i.e., the proton that carries the next greatest energy for mode k after the dominant proton, and likewise for $I_3(k)$. Put $P_2(k) = e_{k,I_2}(k)$ and $P_3(k) = e_{k,I_3}(k)$. Figure 1 is a histogram showing the distribution of values of $P_1(k)$, $P_2(k)$, and $P_3(k)$. The square root scale adopted for the abscissa permits better visualization of the range near zero and means that we are also showing the distributions of the (normalized) peak displacements of the primary, secondary, and tertiary protons. As Figure 1 shows, the contribution of the tertiary protons and beyond is in general quite small, though there are exceptions.

The median value of $P_1(k)$ for all modes is 0.64. Modes corresponding to free O−H stretches are easily distinguished by the fact that their frequencies lie above 3695 cm^{-1}, whereas modes whose primary H is H-bonded have wavenumbers below 3635 cm^{-1} for this database. The free O−H stretches (here, on DAA only) constitute a separate absorption band that is generally discernible in water cluster spectra.[18−19] The free O−H stretches exhibit especially strong domination by their primary protons: the median value of $P_1(k)$ for the 200 modes above 3695 cm^{-1} is 0.80.

We would like to assign modes to specific protons and vice versa, to correlate the modes' frequencies or intensities with bond types or O−H* distances. We match stretch mode k with proton H_i if both (a) for mode k, the proton that has the largest peak displacement is H_i and (b) the mode for which proton H_i's energy fraction is largest is mode k. Mathematically, let $K(i)$ denote the index for which $e_{K(i),i}$ is largest; the criterion is that $K(i) = k$ and $I(k) = i$. If $K(I(k)) \neq k$ (respectively, $I(K(i)) \neq i$) then mode k (respectively, proton H_i) is unassigned. Under this criterion, 752 of 800 (94%) of modes/protons were unambiguously matched. There were 32 instances where assignment failed because two modes had the same primary proton. Such pairs were invariably close in frequency, with the median difference of their wavenumbers being only 10 cm^{-1}. It never happened in this database that three modes shared the same primary proton. Of the unassigned protons, 34 of 48 (71%) were in LF bonds.

We find it convenient to classify stretch modes as "high ratio" or "low ratio". The relevant ratio is $P_1(k)/P_2(k)$, and we set the admittedly arbitrary cutoff at 4.0. In high-ratio modes, the energy (respectively, displacement) of the primary proton is at least four times (respectively, at least twice) that of any other proton. High-ratio modes can be modeled nicely as small perturbations of a single oscillating dipole. Broadly speaking, low-ratio modes are those that require two or more dipoles to be taken into account. The reciprocal ratio P_2/P_1 tracks the absolute energy fraction P_1 fairly closely: Figure 2 is a scatter diagram plotting both $P_2(k)$ and $P_2(k)/P_1(k)$ against $P_1(k)$. Our database contained 411 (51%) low-ratio modes and 389 (49%) high-ratio modes.

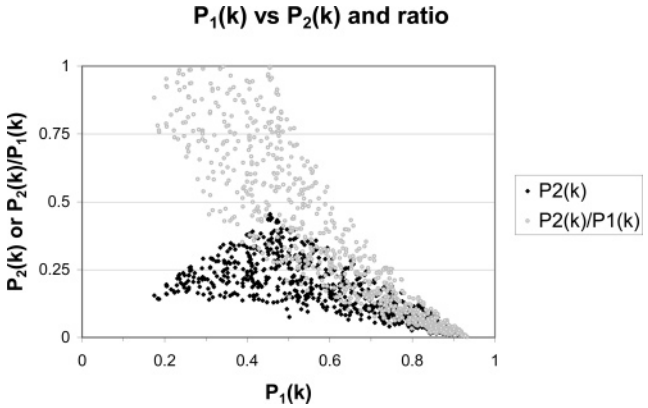

Figure 2. Scatter diagram for P_2 vs P_1, and for P_2/P_1 vs P_1.

TABLE 1: Summary of the 752 Assigned Stretch Modes/ Protons by H-bond Type

		b_{OH} (pm)				freq (cm^{-1}) (scaled)			
type	N	min	mean	s.d.	max	min	mean	s.d.	max
FL	98	98.34	100.65	1.50	105.69	2096	2906	255.8	3314
FF	97	97.34	98.60	0.79	102.05	2704	3283	140.3	3504
LL	95	97.33	98.50	0.81	101.68	2741	3295	145.9	3519
LF	266	96.70	97.34	0.31	98.45	3302	3512	62.8	3635
free H	196	96.22	96.28	0.03	96.39	3697	3716	5.6	3725

As expected, all of the 48 unassigned modes were low ratio, because multiple protons are significantly involved whenever the assignment is unclear.

Another issue we checked is whether stretch modes include some component of displacement perpendicular to the O−H bond direction. For each mode, we computed the angle between the primary proton's vector displacement (i.e. $\vec{a}_{k,I(k)}$) and the direction of its O−H bond at the energy minimum. This angle never exceeded 9°, and it exceeded 4° for only four of the 800 modes. Thus, it is very accurate to think of PWC stretches as occurring strictly along the direction of the O−H bonds. The outliers occurred among the higher-energy clusters that were the most polarized.

Correlation of Frequency with Covalent Bond Length. Given the expected close relationship between ν and b_{OH}, we explored this relationship for the 752 matched modes. It seemed reasonable to omit the unmatched modes (respectively, H-bonds) because they represented just 6% of the total and because each one was arguably very similar to a mode (respectively, H-bond) that was retained in the analysis. We segregated them by the bond type of their dominant proton. Demographics for the five subsets are listed in Table 1. The greatest variability in both ν and b_{OH} occurs for the FL bonds and the least occurs for the free H's. Briefly, the FL bonds generate an absorption band centered around 2900 cm^{-1}, FF and LL virtually coincide and generate a band centered near 3300, and the LF band is centered around 3500. The free H modes generate a very narrow band near 3700: note that the range is just 28 cm^{-1}. Experimentally, the stretch region of bulk water begins around 2900 cm^{-1}; thus, the prediction is that modes for FL bonds in finite clusters can absorb somewhat lower than bulk water. We shall see later why the lower portion of the FL frequency range is excluded for the lowest-energy clusters. The FF and LL subsets appear statistically indistinguishable; sometimes they have been treated as a single-bond class,[28] but for the database of ref 40, which combined several underlying geometries, they were found to be distinct using the t-test.

For each of the five types, we did a least-squares best fit for ν as a function of b_{OH}. Results are given in Table 2, where a

TABLE 2: Correlation of Frequency in cm^{-1} with b_{OH} by H-bond Type

type	a^a	b^b	ΔRMS^a	r (corr)	r^2 (anova)	other var.	ΔRMS^a
FL	20031	−170.14	14.1	.9985	.9970	$(b_{OH})^2$	11.5
FF	20724	−176.90	12.8	.9959	.9918	b_{OH2}	11.4
LL	20986	−179.61	9.2	.9981	.9961	b_{OH2}	8.6
LF	23017	−200.38	13.0	.9784	.9572	b_{OH2}	12.7
free H	18162	−150.04	2.9	.8570	.7345	—	2.9

[a] Units are cm^{-1}. [b] Units are cm^{-1}/pm.

and b are the coefficients in the best-fit line, $\nu = a+(b)b_{OH}$. The single parameter b_{OH} accounts for 96% of the variance in ν for LF bonds and over 99% for the other H-bonds. The numbers 96 and 99% are "analysis of variation" (anova) figures, defined as the square of the correlation coefficient between ν and b_{OH}. These high anova values mean that we can expect a cluster's spectrum to mirror very closely its distribution of b_{OH} values. A cluster's spectrum will be crystalline or amorphous according to whether its b_{OH} values tend to clump or to spread out. The correlation drops to 0.86 for the free H's, but the range is so small that the effect of b_{OH} on ν for free H stretches would be difficult to observe experimentally in any event.

We looked for other parameters that might explain some of the remaining variance in ν. Variables we tried including along with b_{OH} were $(b_{OH})^2$, the H-to-acceptor O distance b_{H-O} (meaningful only for H-bonded protons), R_{OO} (which is very close to $b_{OH} + b_{H-O}$), and the O−H bond length for the mode's secondary proton, denoted b_{OH2}. A p-value cutoff of 0.01 was used to decide significance. Inclusion of the parameter $(b_{OH})^2$ was significant for the FL modes only. This makes sense in that the FL bonds had the largest range, and the correlation of ν with b_{OH} is nonlinear when viewed over a wide range.[26] Inclusion of $(b_{OH})^2$ for FL was also the addition that effected the greatest reduction in mean-squared error (ΔRMS drops from 14.1 to 11.5 cm^{-1}, Table 2). The parameter b_{OH2} was significant for FF, LL, and LF bonds only. Although it makes sense for this parameter to pale for free H stretches, where the secondary proton typically accounted for about one tenth of a mode's energy, it was unexpected that FL modes are unaffected by it. The difference was dramatic, with p-values < 0.001 for FF, LL, and LF, but $p > 0.2$ for FL.

For all four H-bond types, R_{OO} was a poorer predictor of ν than b_{OH}. The value of ΔRMS for ν vs R_{OO} was 18.5 cm^{-1} for LF and ranged from 38.1 to 53.5 cm^{-1} for the other types. The parameters R_{OO} and b_{H-O} are (negatively) correlated with b_{OH}, so they are "significant" when used by themselves. However, for FF, LL, and LF, they do not improve the correlation more than chance would predict when used in combination with b_{OH}. In other words, R_{OO} has no "new" information with which to predict ν that is not already in b_{OH}. For FL bonds, adding R_{OO} does significantly improve the correlation, but this is because R_{OO} is nonlinear in b_{OH}, and including R_{OO} in effect adds $(b_{OH})^2$ to the model. When b_{OH} and $(b_{OH})^2$ are both present as independent variables, adding R_{OO} does not make a statistically significant further contribution. The irrelevance of b_{H-O} to ν supports the view that the relevant oscillating unit consists of the O* and H* only and does not involve the rest of the H-bond.

Relationship between Primary and Secondary Protons. Only 98 of 800 modes (12%) had their primary and secondary protons sharing a donor O (i.e., on the same H_2O unit). Therefore, we wondered if we could explain how a mode "chooses" its secondary proton.

Our operating model of coupled stretch oscillators means that we think of our 40 stretch modes per cluster as if they were the normal modes of a 40 × 40 matrix whose diagonal entries $-\lambda_i$ reflect the intrinsic force constants for individual stretches and whose off-diagonal entries c_{ij} are coupling coefficients. A perturbation argument [49] based on assuming the $\{c_{ij}\}$ are small (cf. eq 29 of ref 50) shows that the secondary proton for the mode dominated by H_m will typically be H_j, where j is the index that maximizes (or one of the indices that comes close to maximizing) $|\tau_{jm}|$, where $\tau_{jm} = c_{jm}/(\lambda_j - \lambda_m)$, for the given value of m. We saw earlier that the intrinsic frequency of an O−H stretch is strongly correlated with b_{OH}, and as a result, $\lambda_j-\lambda_m$ can be expected to be approximately proportional to $b_{OHj}-b_{OHm}$, a quantity we denote as Δb_{OH}.

Sadlej et al.[25] postulated that the coupling coefficient c_{ij} would have the form

$$c_{ij} = q_i q_j w_{ij} \tag{5}$$

where q_i and q_j are dipole moment derivatives associated with the two stretches, and

$$w_{ij} = (\hat{e}_i \cdot \hat{e}_j - 3(\hat{e}_i \cdot \hat{r}_{ij})(\hat{e}_j \cdot \hat{r}_{ij}))/r_{ij}^3 \tag{6}$$

In eq 6, $\hat{e}_i$ and $\hat{e}_j$ are unit vectors parallel to the O−H bonds, and if $\vec{r}_{ij}$ denotes a vector from a point Q_i on the ith dipole to a point Q_j on the jth dipole, then $r_{ij} = |\vec{r}_{ij}|$ and $\hat{r}_{ij} = \vec{r}_{ij}/r_{ij}$ is a unit vector parallel to $\vec{r}_{ij}$. On the basis of matching this model to experimental ice spectra, Buch and Devlin[51] recommended that Q_i and Q_j be taken to lie along their O−H bonds and 0.65 Å from the O, but for simplicity, we will locate Q_i at H_i and Q_j at H_j. Sadlej et al also added a momentum transfer term to the right-hand side of eq 5 when H_i and H_j share a donor O.

Two factors clearly influence $|\tau_{jm}|$ and, hence, the choice of secondary proton. First, other things being equal, a larger $|w_{jm}|$ means a larger coupling, and $|w_{jm}|$ varies approximately as $O(r_{ij}^{-3})$. As a generalization, protons that are closer to H_m have better chances of becoming H_m's secondary proton. Second, the denominator $\lambda_j-\lambda_m$ means that the largest values for $|\tau_{jm}|$ can be expected when Δb_{OH} is minimized, regardless of how far H_j is from H_m. Thus the choice of H_m's secondary proton reflects two competing trends: a "proximity" trend favoring protons nearer to H_m, and a "similarity" or "resonance" trend that gives preference to protons for which b_{OHj} is numerically closest to b_{OHm}.

We were particularly interested in exploring the tradeoff between proximity and similarity effects. For this reason we computed, for each mode k, the distance r_{12} between $H_{I_1}(k)$ and $H_{I_2}(k)$ (henceforth for simplicity denoted H1 and H2) and $\Delta b_{OH} = b_{OH1}-b_{OH2}$. Figure 3 is a scatter diagram plotting r_{12} against Δb_{OH}. We distinguish in Figure 3 the low-ratio H-bonding modes, the high-ratio H-bonding modes, and the free H modes.

Figure 3 has some interesting features. First, notice the fact that the much of the data appears in strata or bands, especially in the lower right region. The lowest three strata consist respectively of H1−H2 pairs that share a donor (we say dd), that have the acceptor of one H as the donor of the other (called da), and that share an acceptor (denoted aa). The fact that their r_{12} values are so uniform within strata is mainly an artifact of the dodecahedral geometry and its close-to-108° angles. Free H modes feature much smaller $|\Delta b_{OH}|$ values (median 0.0117 pm) than H-bonding modes (median 0.07885), but this is mainly because the range of b_{OH} is so much smaller for free H's to start with (cf. Table 1). However, the difference between low-ratio H-bonding modes (median 0.0388) and high-ratio H-bonding modes (median 0.2767) is highly significant. We added to Figure 3 a vertical line at 0.27 pm, which signals a cutoff

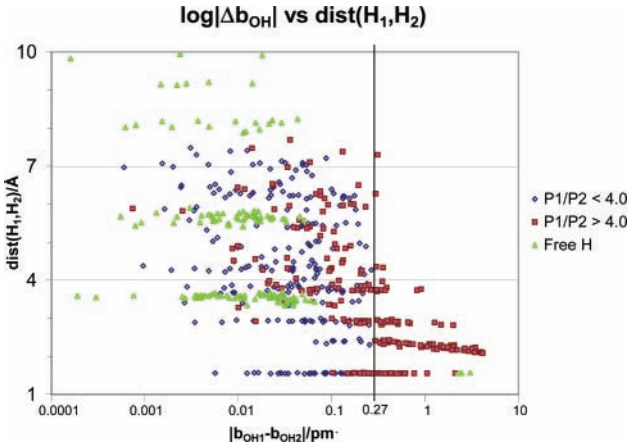

Figure 3. Scatter diagram for r_{12} vs $\log|\Delta b_{OH}|$, for high- and low-ratio H-bonded modes and for free H modes.

above which there are no low-ratio modes. All but two modes having $|\Delta b_{OH}| > 0.39$ pm share an O (i.e., dd, da, or aa).

Among the modes whose primary proton H_m is free, just three choose the other H in the H_2O containing H_m as their secondary proton. These are the only modes with a free primary H whose secondary proton is not free. Nearly half of the others (93 of 197 modes) have their secondary proton attached to an O that is a nearest neighbor to the O of the primary proton: these comprise the band of free H data centered on $r_{12} = 3.5$. The drop-off in the number of modes as a function of r_{12} supports the importance of the proximity effect for free H's. By contrast, among the H-bonded data, for the 356 modes where H1 and H2 do not share an O (i.e., not dd, da, or aa), there is little dependence on r_{12}, suggesting that similarity plays a bigger role than proximity. The fact that 242 (68%) of these H-bonded modes are low ratio also indicates that similarity is the principal selection factor.

We expected to see a trend toward decreasing $\log|\Delta b_{OH}|$ values with increasing r_{12}, reflecting the idea that an increasingly powerful effect due to $|\Delta b_{OH}|$ would be needed to overcome the effects of increasing H1−H2 separation. The trend exists, but it is weak and disappears beyond approximately $r_{12} = 4$ Å. We suspect that the reason for this is that $|\Delta b_{OH}|$ is only an approximate marker for the true denominator $|\lambda_1-\lambda_2|$, and this "noise" effectively renders $\log|\Delta b_{OH}|$ meaningless below a certain point, i.e., below that point, $\log|\Delta b_{OH}|$ bears little correlation with $\log|\lambda_1-\lambda_2|$.

We did not split the free H modes into high- and low-ratio subsets in Figure 3, because we did not find that doing so added any insight. Of the 200 free H modes, 138 (69%) are high ratio, contrasted with only 251 of 600 H-bonding modes (42%). We noted earlier that the median energy fraction carried by the primary proton is highest for the free H modes. Only 4 of 200 (2.0%) free H modes go unassigned, compared with 44 of 600 H-bonding modes (7.3%). Taken together, these observations show that free H modes as a class fit the "perturbation of a single oscillator" paradigm much better than the H-bonding modes do as a class, which tells us that, with rare exceptions, they undergo weak coupling. The dominance of proximity over similarity for free H modes also points to weak coupling. One might have guessed that the free H's, whose intrinsic frequencies are much more closely bunched together (measured as the range of either b_{OH} values or of ν_k's), would have an excellent opportunity to find other H's that give small values of $|\lambda_j-\lambda_m|$ and would therefore couple more strongly on average. Instead, free H's experience weaker coupling across the board than bonding H's. The explanation lies in the dipole derivatives $\{q_i\}$

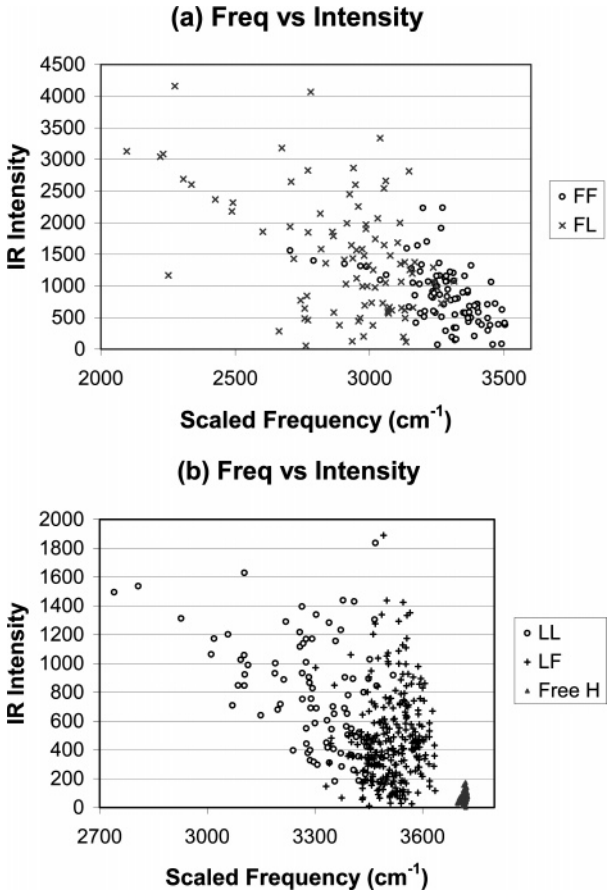

Figure 4. Scatter diagrams for IR intensity vs wavenumber: (a) FF, FL; (b) LL, LF, free-H.

in eq 5, which we have not emphasized in our analysis. The $\{q_i\}$ are consistently smaller for free H's than for bonding H's: Sadlej et al.,[25] citing an earlier work by Rowland et al.,[52] indicate that q_i for a free H is about 6 times smaller than q_i for an H-bonded stretch resonating at 3220 cm^{-1}. This difference outweighs the effect of smaller $|\lambda_j-\lambda_m|$ values.

Stretch Mode Intensity. The IR intensity of O−H stretch modes exhibits an extreme degree of variability. Computed intensities in this database ranged from 0.69 to 4160 km/mol (median: 415). Figure 4a is a scatter diagram of intensity vs frequency for modes assigned to FF and FL protons. Figure 4b does likewise for LL, LF, and free H. Themes that are visually evident from Figure 4a,b are the dramatic concentration of free H modes into a narrow frequency and intensity range, and a weak inverse correlation of intensity with frequency for FF, FL, and LL types. Both effects are known to be consequences of the fact[24−25,52−53] that q_i is negatively correlated with ν, but we hoped to go further than this and to explain the "scatter" that occurs among modes of similar frequency.

We reasoned that the intensity of an O−H stretch mode should be proportional to the squared magnitude of its transition dipole moment (tdm), and that the tdm should equal the weighted vector sum of the dipole derivatives for all the O−H stretches that contribute to it. For the kth mode, this is saying

$$\mathrm{Int}_k = C\,|\mathrm{tdm}_k|^2 \tag{7}$$

$$\mathrm{tdm}_k = \sum_{i=1}^{d} q_i \vec{a}_{ki} \tag{8}$$

where C is independent of k and

Equation 8 assumes that the $\{\vec{a}_{ki}\}$ have been normalized in mass-weighted coordinates, i.e.

$$|\vec{a}_{ki}|^2 = e_{k,i} \tag{9}$$

(cf. Equation 4). Sadlej et al.[25] argued that the relation between a stretch's dipole derivative q and its intrinsic frequency ν would be given by a formula of the form

$$q = (s_0 + s_1\nu)^{1/2} \tag{10}$$

where s_0 and s_1 are constants. They used experimental values for vapor and for ice surface to peg s_0 and s_1. Because ν closely tracks b_{OH}, we postulated that a likely formula for q_i would be

$$q_i = (c_0 + c_1 b_{OH_i})^{1/2} \tag{11}$$

Treating the coefficients c_0 and c_1 as parameters, we correlated $\sqrt{\text{Int}_k}$ with $|\sum_{i=1}^{d}(c_0 + c_1 b_{OH_i})^{1/2}\vec{a}_{ki}|$, obtaining a correlation coefficient r as a function of c_0 and c_1. We adjusted c_0 and c_1 so as to maximize r. Equation 11 worked very well for modes that did not include a significant contribution from free H's. For optimal results across all modes, we used eq 11 when q_i was describing a bonded H but we needed an additional degree of freedom for the free H's. With the formula

$$q_i = \begin{cases} (1267+283(b_{OH_i}-100))^{1/2} & \text{if } b_{OH_i} > 96.5 \text{ (i.e., H−bonded)} \\ 9.5 & \text{if } b_{OH_i} < 96.5 \text{ (i.e., free)} \end{cases} \tag{12}$$

we achieved a correlation of $r = 0.9972$, and the RMS value of $\sqrt{\text{Int}_k} - |\sum_{i=1}^{d} q_i\vec{a}_{ki}|$ was around 0.87. Equation 12 has also been normalized to make $C = 1$ in eq 7.

The extremely good correlation and the small RMS (0.87 where $\sqrt{\text{Int}_k}$ has a range of 0.83 to 64.5), obtained using a model with just three degrees of freedom, tell us that eqs 7, 8, and 12 are an adequate explanation for the large variability in Int_k. The fact that we needed to make an "exception" to eq 11 for the free H's could mean that Sadlej's eq 10 might also need to be modified for free H's, or it could signal an inaccuracy in B3LYP intensity calculations. We will address this issue in the next subsection.

How does Figure 4a,b derive from Eqs 7, 8, and 12? First, consider the high-ratio case. The right-hand side of eq 8 will be dominated by the single term $q_i\vec{a}_{ki}$, where $i = I(k)$, and $|\vec{a}_{ki}|^2$ will be between 0.50 and 0.93 (cf. Figure 2), so tdm_k will (very approximately) vary with q_i^2, which is linear in b_{OH*} by eq 12. Generally, the higher a mode's ratio, the closer its B3LYP intensity will adhere to the straight line given by $I = 0.93(1267 + 283(b_{OHi} - 100))$. For low-ratio modes, the secondary and tertiary protons have a greater impact on the sum in eq 8. Their motion can combine either constructively or destructively with that of the primary proton H_i, resulting in considerably greater departures from the best-fit line. The extremes of destructive and constructive interference are realized with the C_i-symmetric clusters, where a symmetric mode has zero intensity and an antisymmetric mode has double the tdm that a single O−H stretch would generate. (It could be argued that the sum in eq 8 might be dominated by terms having the largest q_i^2 rather than the largest $|\vec{a}_{ki}|^2$, but this does not occur in practice because of the fact, observed above, that the secondary and tertiary protons have b_{OH} values (and hence q_i values) that are quite close to b_{OH*}.) The modes associated with FL bonds show the best correlation ($r = 0.66$) of Int_k with b_{OH*} (and hence with ν), but this is only because the 11 data points below 2650 cm^{-1}, comprising the "tail" in Figure 4a, are all high ratio and most of them fall quite close to a straight line. If these data points are removed, the correlation drops to 0.26 and becomes statistically insignificant. A similar analysis applies to FF and to LL bonds. Among LF bonds, neither the high- nor the low-ratio subset has a significant correlation of intensity with b_{OH*}.

B3LYP Stretch Intensity Benchmarks. If the formula $(1267 + 283(b_{OH_i} - 100))^{(1/2)}$ was applied to a free H, q_i would be in the range 13−15, rather than 9.5 as in eq 12. Do free H stretches behave fundamentally differently from H-bonded stretches, or is the problem in the model B3LYP? How reliable are B3LYP stretch intensity calculations for water clusters?

Figure 4 of Lee et al.[23] juxtaposes spectra computed via B3LYP/6-311++G** and via MP2/DZP for 14 clusters (monomer through decamers). Visually, it is apparent that the intensities computed via B3LYP track quite closely those computed via MP2. Tables 4 and 5 of ref 23 give their numerical raw data for frequencies and intensities. There are 12 clusters that are hexamers or larger for which Lee et al. provide data for benchmarking B3LYP intensities against MP2 intensities. These 12 clusters encompass 117 H-bonded and 55 free H stretches.

For the 117 H-bonded stretches, B3LYP means and medians fall within 30 km/mol of MP2, and the RMS value of $(\text{Int}_{B3LYP} - \text{Int}_{MP2})$ is 65 km/mol (range for Int_{MP2} is 0−3530, median is 410). Among the 26 modes with intensity 800 km/mol or higher, which account for the dominant features of these clusters' spectra, $\text{RMS}\{\text{Int}_{B3LYP} - \text{Int}_{MP2}\}$ is just 78 km/mol; i.e., for the most important signals, the difference is comfortably below 10%.

Comparing Int_{B3LYP} and Int_{MP2} for the 55 free H stretches, the mean value of $(\text{Int}_{B3LYP} - \text{Int}_{MP2})$ is −10 km/mol, and its RMS 15 km/mol. In absolute terms, this is better than for the H-bonded stretches, but because the median free H Int_{MP2} is 80 km/mol, it represents a larger relative error and supports the idea that B3LYP is consistently undercomputing q_i for free H's. In Table 3, we have assembled benchmarks for H_2O, $(H_2O)_2$, and $(H_2O)_4$ intensities computed via B3LYP, MP2/aug-cc-pVDZ, and MP2/aug-cc-pVTZ. We also list the high-level calculations of Schofield and Kjaergaard[54] for $(H_2O)_2$ relative intensities. (The difficulties of obtaining and interpreting experimental absolute intensity figures for the dimer are addressed by Zilles and Person[55] and Low and Kjaergaard[56]). On the basis of all these data, we tentatively conclude that B3LYP does a good or very good job of predicting IR intensity for H-bonded stretches, but for free H stretches it tends to give values that are too low (relative to MP2) by as much as 25%.

Predicted Spectra of Low-Energy PWCs. We observed above (Theorem 2) that a PWC having $A_{TH} = 0$ will typically have 8 H-bond types: FL_x, $x = 2,3,4$; FF_2; LL_2; and LF_x, $x = 0,1,2$, where the subscript is the topological index. The mean ± sd b_{OH} values associated with these 8 categories are, respectively: 101.6 ± 0.2, 100.3 ± 0.1, 99.7 ± 0.2; 98.6 ± 0.2; 98.6 ± 0.2; 97.8 ± 0.1, 97.5 ± 0.1, 97.2 ± 0.05. Plugging these values into Table 2, we find that the associated IR frequency ranges for FF_2 and LL_2 essentially coincide, but otherwise these ranges are distinct or minimally overlap. We can expect a crystalline spectrum with 7 distinct or nearly distinct H-bonding bands and an eighth band for the free H's. Figure 5 shows a spectrum obtained by combining predicted spectra for the four lowest energy 5^{12} PWCs in our database. As predicted, the eight bands are clearly distinguishable and

TABLE 3: Benchmark Comparisons of Some Computed IR Intensities in km Mol^{-1}

	monomer			dimer				cyclic S_4 tetramer	
	ν_1	ν_2	ν_3	H-bonded	donor free	acc ν_1	acc ν_3	H-bonded (total)	free (total)
MP2/aug-cc-pVDZ	13.2	89.0	76.1	303	119	22	101	2812	325
MP2/aug-cc-pVTZ	9.6	66.7	57.2	314	117	21	105	2855	384
B3LYP/6-311++G**	9.2	66.7	56.8	332	80	16	86	2950	279
QCISD/6-311++G(2d,2p)[a]				310	109	15	83		

[a] Relative intensities only, from ref 54, scaled to make first entry = 310.

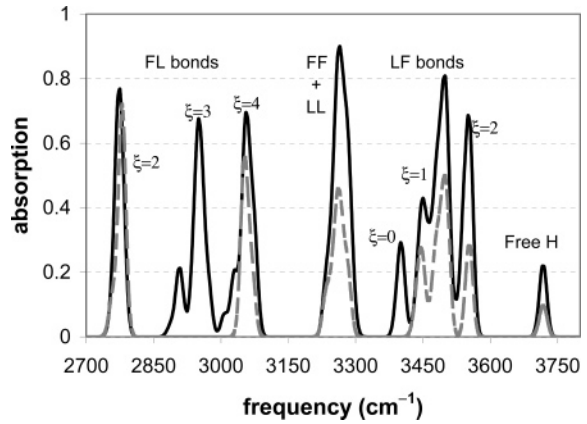

Figure 5. Predicted IR spectrum for Dod-1 (dashed line) and sum of 4 lowest energy 5^{12} clusters (solid line).

have been labeled. The dashed line in Figure 5 is the spectrum of Dod-1. Dod-1 is our lowest energy 5^{12} PWC, which is pictured as Figure 8a in ref 29. The spectrum for Dod-1 lacks a band for FL_3 because it happens not to have any of this bond type, and the signals from its two LF_0 bonds merge into the red end of the LF_1 band.

The FL_3 and LF_1 bands exhibit substructures consisting of a splitting into a smaller subband toward the red and a larger subband toward the blue side. We have no explanation for these splittings. They do not reflect gaps in the distribution of b_{OH} values within the FL_3 and LF_1 bond classes. Examination of individual modes for Dod-1 showed that modes associated to LF bonds with b_{OH} longer than 97.58 pm averaged 10 cm^{-1} redder than the formula $23017 - (200.38)b_{OH}$ (cf. Table 2) would predict, whereas LF bonds with b_{OH} below this cutoff averaged 10 cm^{-1} bluer than the formula. The other bands do not exhibit substructure.

By Theorem 2, bond types FL_0, FL_1, LF_3, and LF_4 do not occur in a PWC with $A_{TH} = 0$. These bond types represent the shortest (FL_0 and FL_1) and the longest (LF_3 and LF_4) H-bonds in water clusters. The effect of excluding them is to reduce the spread of the stretch region for low-energy clusters. Exclusion of the shortest bonds is probably one of several factors operating in bulk water that cause absorption to drop off rapidly below 2950 cm^{-1}.

Can we explain the gap between Figure 5, whose lowest peak occurs at 2775 cm^{-1}, and the cutoff of the bulk water stretch signal below 2900 cm^{-1}? One obvious issue is the validity of predicted DFT frequencies. The minimum-energy clusters for $(H_2O)_n$, $n = 7$, 8, 9, and 10, each have an FL bond whose IR signal has been measured. Their signals occur at 2940, 3090 (D_{2d}) and 3060 (S_4), 3070, and 3055 cm^{-1}, respectively.[8−9,25] Scaled DFT frequencies (scaling factor 0.961) for these modes are 2987, 3170 (D_{2d}) and 3132 (S_4), 3160, and 3153 cm^{-1}, respectively. Better agreement would result from setting the scaling factor for FL bonds to 0.937. For $n = 7$, the bond has type FL_3, and for $n = 8$, 9, and 10, they are FL_4. The FL_4 band of Figure 5 (drawn using 0.961) matches quite well these experimental frequencies for FL_4 bonds, and the FL_3 band of Figure 5 matches the 2940 cm^{-1} signal for $(H_2O)_7$. Thus, we do not believe the gap results from too-low scaled DFT frequencies. A partial explanation probably lies in the fact that melting itself can blue-shift an FL peak by 100 cm^{-1} or more, as Brudermann et al.[57] demonstrated for $(H_2O)_9$; and FL_2 bonds may be scarcer in liquid water than in 5^{12} clusters.

Conclusion

We have used a database of 20 5^{12} $(H_2O)_{20}$ water clusters to examine statistical trends among the O−H stretch modes, computed at a high level of ab initio theory. We believe this is the first study to look at a large number of normal modes in this way. Principal findings include:

- A large majority (94%) of modes can be assigned unambiguously to a single O−H stretch, using a strict assignment criterion.
- For nearly 75% of modes, half or more of their energy is carried by the primary O−H* stretch. Very few modes involve significant motion of four or more protons.
- When an assignment to a specific proton can be made, the O−H* distance b_{OH*} correlates extremely well with mode frequency. For those H* that have an acceptor, frequency is much more tightly correlated with b_{OH*} than with the O−O distance.
- It is convenient to classify modes as "high ratio" or "low ratio", the "ratio" being the energy carried by the primary proton divided by that of the secondary proton. High-ratio modes are describable as a single oscillator undergoing only small coupling perturbations from other oscillators, but this description becomes less and less appropriate as the ratio drops toward 1. In this database, modes were nearly evenly split between high- (>4.0) and low- (<4.0) ratio modes. The distribution of ratios varies with the bond type, with free O−H stretches having the highest ratios on average.
- The "choice" of a stretch mode's secondary proton reflects a complex tradeoff of proximity (measured as dist(H_1,H_2)) and similarity (measured as $|b_{OH1} - b_{OH2}|$), both of which influence the coupling between stretch oscillators. In 88% of modes, the secondary proton occurred on a different H_2O unit than the primary proton.
- IR intensities vary by nearly 4 orders of magnitude in this database. Intensities are explained very accurately via a weighted vector sum of all protons' displacements, weighted by dipole derivatives (q_i) that are a function of b_{OH*}. In particular, a square-root formula for q_i (for H-bonded protons) is dramatically supported by our data.
- The lowest-energy 5^{12} clusters, and low-energy PWCs in general, can be expected to have (up to) eight bands in the stretch region of their IR spectra. This can be derived from the theory of PWC bond lengths, which predicts eight nearly nonoverlapping b_{OH} intervals.

Will these conclusions be transferable to clusters other than PWCs or to bulk water? Performing this type of study initially on a 5^{12} database conferred certain advantages: a well-

understood theory of b_{OH} values with a wide spread for b_{OH}; uniformity of some geometric features such as O−O−O angles and certain H−H distances; and simplification of the number of H-bonding environments to just five (FF, FL, LL, LF, free). These advantages may not apply to other cluster sets. For bulk water and for ice surface, we suspect that the spread of b_{OH} values and its tradeoff with proximity are similar enough to 5^{12}'s so that the big picture of a high assignability rate (of modes to protons) with many high-ratio modes will still apply. For ice interior and nanoparticle interiors,[24] however, the majority of b_{OH} values clump into a narrow interval. The result will be low assignability rates and low-ratio modes as more "coherent" modes arise.

References and Notes

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

(2) Goldman, N.; Fellers, R. S.; Brown, M. G.; Braly, L. B.; Keoshian, C. J.; Leforestier, C.; Saykally, R. J. *J. Chem. Phys*. **2002**, *115*, 10148−10163.

(3) Gregory, J. K.; Clary, D. C. *J. Chem. Phys.* **1996**, *105*, 6626−6633.

(4) Viant, M. R.; Cruzan, J. D.; Lucas, D. D.; Brown, M. G.; Liu, K.; Saykally, R. J. *J. Phys. Chem. A* **1997**, *101*, 9032−9041.

(5) Cruzan, J. D.; Viant, M. R.; Brown, M. G.; Saykally, R. J. *J. Phys. Chem. A* **1997**, *101*, 9022−9031.

(6) Liu, K.; Brown, M. G.; Cruzan, J. D.; Saykally, R. J. *J. Phys. Chem. A* **1997**, *101*, 9011−9021.

(7) Liu, K.; Brown, M. G.; Saykally, R. J. *J. Phys. Chem. A* **1997**, *101*, 8995−9010.

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

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

(10) Gruenioh, C. J.; Carney, J. R.; Hagemeister, F. C.; Arrington, C. A.; Zwier, T. S. *J. Chem. Phys*. **1998**, *109*, 6601−6614.

(11) Gruenioh, C. J.; Carney, J. R.; Hagemeister, F. C.; Zwier, T. S. *J. Chem. Phys*. **2000**, *113*, 2290−2303.

(12) Lee, J. Y.; Kim, J.; Lee, H. M.; Tarakeshwar, P.; Kim, K. S. *J. Chem. Phys*. **2000**, *113*, 6160−6168.

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

(14) Burnham, C. J.; Xantheas, S. S.; Miller, M. A.; Applegate, B. E.; Miller, R. E. *J. Chem. Phys*. **2002**, *117*, 1109−1122.

(15) Fajardo, M. E.; Tam, S. *J. Chem. Phys*. **2001**, *115*, 6807−6810.

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

(17) 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.

(18) Miyazaki, M.; Fujii, A.; Ebata, T.; Mikami, N. *J. Phys. Chem. A* **2004**, *108*, 10656−10660.

(19) Zwier, T. S. *Science* **2004**, *304*, 1119−1120.

(20) Bernath, P. F. *Spectra of Atoms and Molecules*; Oxford University Press: New York, Oxford, 1995; p. 241.

(21) Kim, J.; Kim, K. S. *J. Chem. Phys*. **1998**, *109*, 5886.

(22) Kim, J.; Majumdar, D.; Lee, H. M.; Kim, K. S. *J. Chem. Phys*. **1999**, *110*, 9128.

(23) Lee, H. M.; Suh, S. B.; Lee, J. Y.; Tarakeshwar, P.; Kim, K. S. *J. Chem. Phys.* **2000**, *112*, 9759.

(24) Buch, V.; Bauerecker, S.; Devlin, J. P.; Buck, U.; Kazimirski, J. K. *Int. Rev. Phys. Chem*. **2004**, *23*, 375.

(25) Sadlej, J.; Buch, V.; Kazmirski, J. K.; Buck, U. *J. Phys. Chem. A* **1999**, *103*, 4933−4947.

(26) Jeffrey, G. A. *An Introduction to Hydrogen Bonding*; Oxford University Press: New York, Oxford, 1997; Chapter 11.2.

(27) Sadlej, J., *Chem. Phys. Lett*. **2001**, *333*, 485−492.

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

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

(30) Davidson, D. W. Clathrate Hydrates. *Water: A Comprehensive Treatise*, 2nd ed.; Franks, F., Ed.; Plenum Press: New York, 1976; Vol. 2, Chapter 3.

(31) Koh, C. A. *Chem. Soc. Rev*. **2002**, *31*, 157−167.

(32) Selinger, A.; Castleman, A. W. *J. Phys. Chem*. **1991**, *95*, 8442.

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

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

(35) Anick, D. J. *J. Phys. Chem. A* **2005**, *109*, 5596−5601.

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

(37) Nova, J. J.; Sosa, C. *J. Phys. Chem*. **1995**, *99*, 15837.

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

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

(40) Anick, D. J. *J. Mol. Struct*. (*Theochem*) **2002**, *587*, 97−110.

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

(42) Anick, D. J. *J. Phys. Chem. A* **2003**, *107*, 1348.

(43) Scott, A. P.; Radom, L. *J. Phys. Chem*. **1996**, *100*, 16502−16513.

(44) Spanget-Larsen, J. *Chem. Phys.* **1999**, *240*, 51−61. See also http://srdata.nist.gov/cccbdb/

(45) Anick, D. J. *J. Mol. Struct*. (*Theochem*) **2002**, *587*, 87−96.

(46) *PQS* v. 3.1; Parallel Quantum Solutions: Fayetteville, AR.

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

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

(49) Brown, R. D.; Bassett, I. M. *Proc. Phys. Soc.* **1958**, *71*, 724.

(50) Tschumper, G. http://zopyros.ccqc.uga.edu/lec_top/pt/ptlec.html (1997).

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

(52) Rowland, B.; Kadagathur, N. S.; Devlin, J. P.; Buch, V.; Feldman, T.; Wojcik, M. J. *J. Chem. Phys*. **1995**, *102*, 8328.

(53) Hermansson, K.; Lindgren, J.; Probst, M. M. *Chem. Phys. Lett*. **1995**, *233* (4), 371.

(54) Schofield, D. P.; Kjaergaard, H. G. *Phys. Chem. Chem. Phys*. **2003**, *5*, 3100.

(55) Zilles, B. A.; Person, W. B. *J. Chem. Phys*. **1983**, *79* (1), 65.

(56) Low, G. R.; Kjaergaard, H. G. *J. Chem. Phys*. **1999**, *110* (18), 9104.

(57) Brudermann, J.; Buck, U.; Buch, V. *J. Phys. Chem. A* **2002**, *106* (3), 453.