A Predictive Model of Strong Hydrogen Bonding in Proteins: The $N^{\delta 1}-H-O^{\delta 1}$ Hydrogen Bond in Low-pH α-Chymotrypsin and α-Lytic Protease

Pablo A. Molina and Jan H. Jensen*

Department of Chemistry, University of Iowa, Iowa City, Iowa 52242

Received: January 10, 2003; In Final Form: April 10, 2003

A computational QM/MM methodology for the systematic study of various structural and spectroscopic properties of strong hydrogen bonds in enzymes is presented. The theoretical model is applied to the $N^{\delta 1}-H-O^{\delta 1}$ hydrogen bond between His57 and Asp102 in the active sites of low-pH α-chymotrypsin and α-lytic protease. The minimum energy structures of both enzymes reproduce the experimental $N^{\delta 1}-O^{\delta 1}$ distance and are used to obtain computational values for the H−D fractionation factor (ϕ), the proton chemical shift (δ_H) of the $H^{\delta 1}$, and changes in δ_H upon isotope substitution ($\Delta\delta_{H-D}$ and $\Delta\delta_{H-T}$). For both enzymes, calculated parameters are in good agreement with available experimental data. Predictions are made for other properties for which experimental values are not available.

I. Introduction

The hydrogen bond (HB) is central to biomolecular structure and function. The spectroscopic characterization of HBs is a vital component of biochemistry, and an increasing amount of detailed information about HBs in biomolecules is becoming available. A good example is the $N^{\delta 1}-H-O^{\delta 1}$ HB in the catalytic triad of serine proteases and esterases for which X-ray structures, proton chemical shifts (δ_H), changes in δ_H upon deuteration ($\Delta\delta_{H-D} = \delta_H - \delta_D$), and H−D fractionation factors (ϕ) have been measured.[1,2] Recently, the first accurate measurements of $\Delta\delta_{H-T}$ in a protein were reported for the serine protease α-chymotrypsin (Cht) complexed with three transition state analogue inhibitors.[3]

The interest in the $N^{\delta 1}-H-O^{\delta 1}$ HB was initially sparked by the unusually low-field δ_H of 18 ppm first observed by Robillard and Shulman[4] in the low-pH form of Cht. On the basis of detailed studies of HBs in small organic molecules, Hibbert and Emsley[5] have suggested that a low-field δ_H is consistent with a strong HB and is accompanied by a positive $\Delta\delta_{H-D}$, a $\phi < 1$, and $1.0 \leq \nu^H_{st}/\nu^D_{st} \leq 1.4$ (where ν_{st} is the vibrational stretch frequency of the hydrogen donor). Subsequently, $\Delta\delta_{H-D} = 1.0 \pm 0.4$ ppm and $\phi = 0.4$ were first measured for low-pH Cht and chymotrypsinogen, respectively, by Frey et al.[6] and Markley and Westler.[7] A $\Delta\delta_{H-D}$ of 1.1 ± 0.5 ppm has also been measured for the Cht-N-Ac-Leu-Phe-CF_3 complex,[8] while low-field δ_H and low ϕ's have been measured for many other serine proteases. Due to the difficulty of assigning vibrational spectra of proteins, experimental measurements of ν^H_{st}/ν^D_{st} for serine proteases have not appeared in the literature.

These measurements provide valuable benchmarks for theoretical models of strong HBs, which in turn would provide valuable tools for the interpretation of these properties. However, the accurate ab initio predictions of these spectroscopic properties for strong HBs present a significant challenge, even for very simple systems. For example, Barich, Nicholas, and Haw[9] have shown that the effect of electron correlation (at the MP2 level) on proton chemical shifts is unusually large for strong HBs. At the highest level of theory considered by Barich et al., the predicted proton chemical shift of hydrogen maleate was still 1.3 ppm larger than the experimental value. Interestingly, Garcia-Viloca et al.[10] have shown that vibrational averaging of the proton-transfer coordinate reduces the proton chemical shift by 1.6 ppm for hydrogen maleate at the RHF level of theory. Because of their use of RHF, δ_H was still significantly overestimated, though the predicted $\Delta\delta_{H-D}$ was in excellent agreement with experiment.

Very recently, Westler, Weinhold, and Markley[11] were finally able to reproduce the proton chemical shift of hydrogen maleate (as well as FHF^-, hydrogen 2,2-dimethylmalonate, and hydrogen phthalate) to within a few tenths of a ppm by including both vibrational averaging and electron correlation effects. Here, the electron correlation was treated at the B3LYP level, with an empirical scaling correction. The same methodology was applied to compute a δ_H value of 18.3 ppm for an 78-atom ab initio model of complexed Cht. Though vibrational averaging could not be performed for this model due to computational expense, the value agrees well with the experimental range of 18.6−18.9 ppm measured for various closely related inhibitors complexed to Cht. The use of a smaller (33-atom) model that allowed for vibrational averaging gave less satisfactory results.

The small model of Westler et al.[11] was also used to calculate $\Delta\delta_{H-T}$ and ϕ values at several constrained $N^{\delta 1}-O^{\delta 1}$ distances. At the $N^{\delta 1}-O^{\delta 1}$ distance (2.6 Å) most closely resembling the measured values (2.57−2.62 Å), the predicted values of $\Delta\delta_{H-T}$ (0.39 ppm) and ϕ (0.50) were in good agreement with experimental values (0.63−0.68 ppm and 0.32−0.43, respectively), although the vibrationally averaged δ_H was significantly overestimated (20.91 vs 18.61−18.95 ppm). The good agreement for ϕ is especially remarkable as only the change in zero-point energy of the stretch frequency was considered (a limitation imposed by the constraint). A previous study of Edison, Weinhold, and Markley[12] has shown that other vibrational modes associated with the HB, such as bends, can contribute significantly to ϕ in model compounds.

In this paper, we use a hybrid QM/MM model (the effective fragment potential method)[13] to construct a predictive model of the $N^{\delta 1}-H-O^{\delta 1}$ HB in the low-pH forms of Cht and α-lytic protease (Alp). The use of a hybrid approach allows for a (static) treatment of a significant portion of the protein (ca. 600 atoms) close to the active site, while retaining an (flexible) ab initio description of the active site. The low-pH form is the simplest

* To whom correspondence should be addressed. E-mail: Jan-Jensen@uiowa.edu.

10.1021/jp0340663 CCC: $25.00
Published on Web 05/29/2003

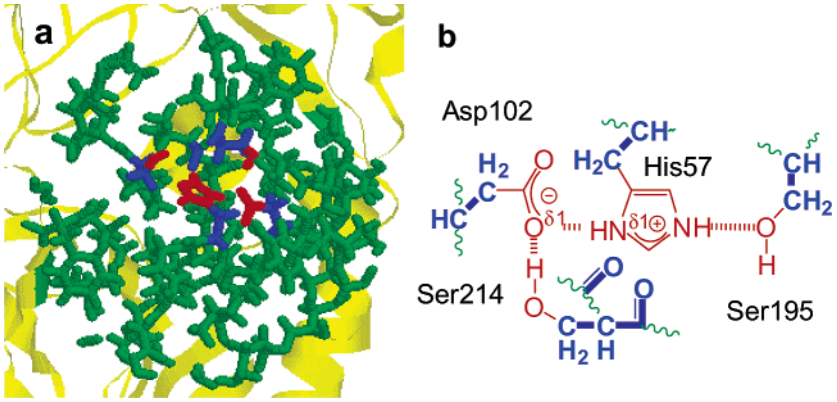

Figure 1. α-Chymotrypsin. (a) Ab initio/buffer/EFP regions (red/blue/green). The EFP describes the protein environment within 13 of the active site. (b) Detail of the ab initio and buffer regions.

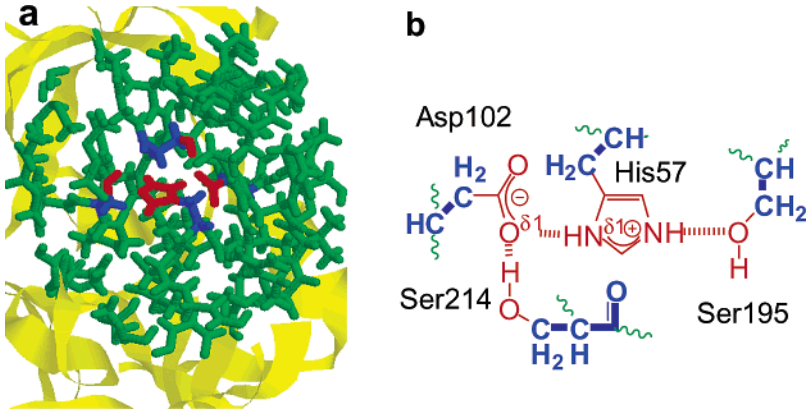

Figure 2. α-Lytic protease. Same as Figure 1.

possible model of the strong HB in serine protease, resulting in a relatively small QM region, thereby making a relatively rigorous computational treatment feasible. Cht and Alp are chosen because several spectroscopic measurements have been made for each system, which indicate a significant difference in HB strength.

The paper is organized as follows. First, we describe the QM/MM model in detail as well as three methodological innovations that allow for the accurate prediction of ϕ and δ_H: (1) The combined use of the partial Hessian vibrational analysis[14] and the vibrational-SCF method,[15] which allows for the efficient calculation of anharmonic vibrational wave functions and energies for a QM/MM model; (2) The combined use of the partial pressures of H_2O and D_2O to scale a ϕ calculated relative to $H_2O_{(gas)}/D_2O_{(gas)}$ and an anharmonic treatment of select energy levels to give accurate ϕ's; (3) The combined use of Rablen et al.'s[16] proton chemical shielding scaling factor and anharmonic corrections to give accurate low-field δ_H's. Second, we discuss the predictions of δ_H, ϕ, $\Delta\delta_{H-D}$, $\Delta\delta_{H-T}$, ν^H_{st}/ν^D_{st}, as well as the $N^{\delta1}-H$ and $N^{\delta1}-O^{\delta1}$ distances for Cht and Alp, obtained using our computational methodology. By direct comparison to experimental data, we demonstrate that our model is the first consistently predictive model of strong hydrogen bonding in an enzyme active site. Finally, we summarize our results and discuss future directions.

II. Computational Methodology

II.a. QM/MM Model. The QM/MM model of the active site has been discussed in considerable detail elsewhere[17] for α-chymotrypsin (Cht) and is only briefly summarized here. The method consists of three computational regions (Figures 1 and 2): ab initio/buffer/EFP. The EFP describes the electrostatic potential of the enzyme within 13 Å of the active site by a multipole expansion (charges through octupoles at all atoms and bond midpoints and dipole polarizability tensors for each valence localized molecular orbital). The EFP is obtained by several RHF/6-31G(d) ab initio calculations,[17,18] as described in section II.b. The EFP and ab initio regions are separated by a buffer[19] region of localized molecular orbitals, obtained at the RHF/6-31G(d) level of theory, that are kept frozen during the SCF optimization of the ab initio region. The geometry of the EFP and buffer regions are kept fixed at the coordinates obtained by Blevins and Tulinski[20] for Cht and Fujinaga et al.[21] for α-lytic protease (Alp). The geometry of the ab initio region is optimized at the B3LYP/6-31G(d) level of theory, in the presence of the buffer and EFP regions.

The harmonic vibrational frequencies are calculated by the partial Hessian vibrational analysis of Li and Jensen.[14] In this method, a partial Hessian (PH) is computed numerically (by double difference) by displacing only the atoms in the ab initio region. Overall translational and vibrational degrees of freedom are projected out of the PH, and subsequent diagonalization yields $3N_{AI}$-3 vibrational frequencies corresponding to vibrational motion of the N_{AI} ab initio atoms.[14]

II.b. Construction of the EFP Region. The EFP region is too large to be computed by a single ab initio calculation. Instead, the protein EFP is constructed from several pieces using a divide-and-conquer approach. For Cht, the "remove-combine-and-scale" method of Minikis, Kayirs, and Jensen[18] was used to construct the EFP. For Alp, we used a subsequently implemented divide-and-conquer approach based on the neutralization method of Bellido et al.[22] which avoids charge scaling (the "expand-collect-and-correct" method[23]). While the latter approach is more computationally convenient, both methods yield similar accuracy.[23]

The fragments used to construct the EFP region of Cht (Figure 1) have been described in detail elsewhere.[17] For Alp (Figure 2), the EFP fragments are chosen as follows: Two crystallographic waters (53 and 64) and 43 residues (roughly 21% of the total number of residues in the enzyme) are used to simulate the molecular environment within 13 Å of the active site. The 13 Å sphere is divided into six spatially distinct fragments consisting of the following residues: 190−199, 211−218, 225−230, 88−104, and 40−43_53−59A (where _ indicates a disulfide linkage). The EFP parameters for the fragments are obtained by six separate RHF/6-31G(d) single-point calculations and then assembled into a single EFP.[23]

II.c. Calculation of ϕ. The H/D fractionation factor (ϕ) is the equilibrium constant for the transfer of D from the bulk solvent (X_2O_{liq}) into the protein (PX; X = H or D)

$$PH + {}^1/_2D_2O \rightleftharpoons PD + {}^1/_2H_2O \tag{1}$$

$$\phi = \frac{[PD][H_2O_{liq}]^{1/2}}{[PH][D_2O_{liq}]^{1/2}} \tag{2}$$

An ab initio description of liquid water is computationally demanding but it can be avoided by following Graul, Brickhouse, and Squires[24] and rewriting eq 2 as

$$\phi = \frac{[PD][H_2O_{gas}]^{1/2}[D_2O_{gas}]^{1/2}[H_2O_{liq}]^{1/2}}{[PH][D_2O_{gas}]^{1/2}[H_2O_{gas}]^{1/2}[D_2O_{liq}]^{1/2}} \tag{3}$$

$$\phi = \frac{[PD][H_2O_{gas}]^{1/2}}{[PH][D_2O_{gas}]^{1/2}} \cdot \left(\frac{[D_2O_{gas}]/[D_2O_{liq}]}{[H_2O_{gas}]/[H_2O_{liq}]}\right)^{1/2} \tag{4}$$

$$\phi = \frac{[PD][H_2O_{gas}]^{1/2}}{[PH][D_2O_{gas}]^{1/2}} \cdot \left(\frac{P_{D_2O}}{P_{H_2O}}\right)^{1/2} \tag{5}$$

Here, P_{H_2O} and P_{D_2O} are the partial pressures of H_2O and D_2O, respectively. The H/D fractionation factor is then calculated by the following equation:[24]

$$\phi = \sqrt{\frac{P_{D_2O}}{P_{H_2O}}} \cdot e^{-\left[\left(G_{PD}+\frac{1}{2}G_{H_2O}\right)-\left(G_{PH}+\frac{1}{2}G_{D_2O}\right)\right]/RT} = \sqrt{\frac{P_{D_2O}}{P_{H_2O}}} \cdot e^{-\left(\frac{1}{2}\Delta G_{H_2O/D_2O}-\Delta G_{PH/PD}\right)/RT} \tag{6}$$

The free energies of the proteins and water are initially calculated (at 277 K) using the usual rigid rotor-harmonic oscillator approximations.

$$G = G^{trans} + G^{rotat} + G^{vib} \tag{7}$$

For the proteins, the $3N_{AI} - 3$ frequencies (where N_{AI} is the number of ab initio atoms) are used for the vibrational contribution, while the translational and rotational contributions are calculated for the entire ab initio/buffer/EFP system.[14] In practice, the translational terms and the rotational free energy of the protein (but not for water) are unaffected by deuteration. The ratio of partial pressures at 277 K is obtained from the experimentally measured values of Kakiuchi.[25] Measurements were done for 273−373 K in 10 K increments. The 277 K value ($\sqrt{P_{D_2O}/P_{H_2O}} = 0.908$) was obtained by linear interpolation of the 273 and 283 K values [0.903 and 0.915, i.e., 0.903 + 0.4(0.012)].

II.d. Vibrational SCF Calculations. The ab initio vibrational SCF method of Chaban, Jung, and Gerber[15] is used to calculate anharmonic vibrational energies and wave functions for select modes. No coupling between modes is considered, so that for each mode a one-dimensional (1-D) potential energy surface (PES) is generated by 16 equally spaced displacements (in mass-weighted atomic units) along each mode. In our study, only the ab initio atoms were displaced. The energy and wave function of the ground state and first excited state are computed numerically for each mode. The energy difference between the ground and first exited state is the vibrational frequency. Isotopic substitution changes the harmonic modes, so 1-D PESs are recomputed to obtain vibrational energies and wave functions for each isotope.

II.e. Chemical Shift Calculations. The accurate prediction of an experimentally measured NMR chemical shift presents a challenge to theory, due to the very high level of theory necessary for converged results and the effect of the molecular environment (e.g., solvent or protein). Chesnut[26] has proposed a linear scaling technique to address these effects, and Rablen, Pearlman, and Finkbiner[16] have obtained the necessary parameters for proton chemical shifts relative to TMS in nonpolar solvents ($CDCl_3$ and CCl_4):

$$\delta_H(TMS_{CDCl_3}) = 30.60 - 0.957\sigma_H^A \tag{8}$$

Here, σ_H is the isotropic chemical shielding calculated at the B3LYP/6-311++G(d,p)//B3LYP/6-31G+(d) level of theory, which in this study is approximated by B3LYP/6-311++G-(d,p)//B3LYP/6-31G(d), referred to hereafter as method A.

Ideally the chemical shift of the $N^{\delta 1}$-H proton would be calculated at this level using a molecular system that includes a sizable part of the protein. However, this direct approach is computationally unfeasible. Instead, we use the ONIOM-NMR method of Karadov and Morokuma[27] for calculating chemical shifts in large molecules. Here we use a three-layered approach to estimate the chemical shift of the $N^{\delta 1}$-H proton for a system that includes the molecular environment within roughly 7 Å of the proton (Figure 3b) calculated using method A

$$\sigma_H^A[7\text{ Å}] = \sigma_H^A[D] + \sigma_H^{B'}[5\text{ Å}] - \sigma_H^B[D] + \sigma_H^C[7\text{ Å}] - \sigma_H^C[5\text{ Å}] \tag{9}$$

"5 Å" refers to a system that includes the molecular environment within roughly 5 Å of the proton (in Figure 3a), while D refers to the formate−imidazolium dimer. In all cases, the geometry of the atoms in the ab initio region is that optimized at the B3LYP/6-31G(d) level of theory in the presence of the EFP. Methods B and C refer to B3LYP/6-31+G(d,p) and RHF/STO-3G chemical shift calculations, respectively, while B′ refers to a B3LYP/mixed basis set chemical shift calculations with 6-31+G(d,p) for the atoms in the ab initio region and STO-3G elsewhere (Figure 3a). Chesnut and co-workers[28] have previously used mixed basis sets successfully for chemical shift calculations, while the use of a STO-3G layer to model chemical shifts in large systems has been used previously by Mennucci,

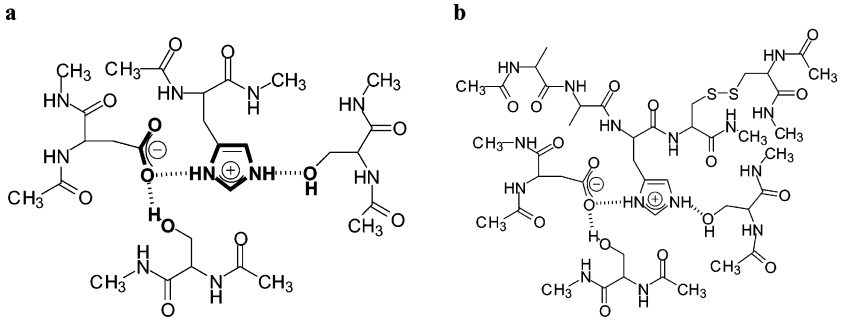

Figure 3. Subsystem of α-chymotrypsin corresponding to the protein environment within roughly (a) 5 Å and (b) 7 Å of the active site, respectively, used for the NMR chemical shift calculations. The bold region in (a) denotes the atoms described by the 6-31G(d) basis set.

TABLE 1: Summary of Experimental and Computed Observables Using a Minimum Energy Geometry (MEG) or Vibrational Average (VA) for Low-pH α-Chymotrypsin and α-Lytic Protease

	α-chymotrypsin			α-lytic protease		
		predicted			predicted	
properties	exptl	VA	MEG	exptl	VA	MEG
r_{NO} (Å)	2.67[20]	2.67	2.67	2.72[21]	2.73	2.73
δ_H (ppm)	18.2[35]	18.3	17.8	17.1[2]	16.8	16.3
ϕ (277 K)	0.64 ± 0.02[35]	0.67	0.85		0.85	0.95
$\Delta\delta_{H-D}$ (ppm)	1.0 ± 0.4[6]	0.28			0.21	
$\Delta\delta_{H-T}$ (ppm)		0.44			0.35	
ν^H_{st}/ν^D_{st}		1.18	1.32		1.25	1.33
$N^{\delta1}$-H (Å)		1.102	1.07	1.113 ± 0.018[2]	1.086	1.059
$N^{\delta1}$-D (Å)		1.087			1.075	
$N^{\delta1}$-T (Å)		1.076			1.068	

Martinez, and Tomasi[29] to reproduce experimental solvent-induced shifts on chemical shielding tensors.

Finally, the aqueous phase value of the chemical shielding of the proton is given by[30]

$$\delta_H = \delta_H(DSS_{aq}) = \delta_H(TMS_{CDCl_3}) + 0.33 \qquad (10)$$

This correction is almost entirely due to solvent effects, since DSS and TMS have very similar chemical shifts in the same solvent.[31]

II.f. Calculation of $\langle\delta_H\rangle$ and $\Delta\langle\delta_{H-X}\rangle$. The change in chemical shift of the $N^{\delta1}$-H proton upon isotopic substitution (X = D, T) is

$$\begin{aligned}\Delta\delta_{H-X} &= \langle\delta_H\rangle - \langle\delta_X\rangle \\ &= \langle\sigma_X\rangle - \langle\sigma_H\rangle\end{aligned} \qquad (11)$$

The latter equation makes the usual assumption that anharmonic effects have a negligible effect on the TMS chemical shielding reference (the anharmonic correction to $\sigma_{H(TMS)}$ was calculated using method B and found to be 0.01 ppm). The expectation value of the chemical shielding is given by[32]

$$\langle\sigma_X\rangle = \sum_{i=1}^{16}(\psi^X_{str}(i))^2\sigma_H(i)\Delta x \qquad (12)$$

where Δx is the displacement-length in mass-weighted Cartesian coordinates determined by the vibrational SCF (VSCF) algorithm. Thus, the chemical shielding value is calculated at each point along the VSCF 1D PES for each isotope. Since a total of 48 chemical shielding calculations are required to calculate $\Delta\delta_{H-D}$ and $\Delta\delta_{H-T}$ for each protein, the chemical shielding tensors are estimated using method B

$$\sigma_H \approx \sigma^B_H[7\text{ Å}] = \sigma^{B'}_H[5\text{ Å}] + \sigma^C_H[7\text{ Å}] - \sigma^C_H[5\text{ Å}] \qquad (13)$$

Using the ONIOM approach, $\langle\sigma_x\rangle$ has two terms

$$\langle\sigma_X\rangle = \langle\sigma^{B'}_X[5\text{ Å}]\rangle + \langle\sigma^C_X[7\text{ Å}] - \sigma^C_X[5\text{ Å}]\rangle \qquad (14)$$

For X = H, the latter term is very similar to the harmonic value (0.14 vs 0.17 ppm). Thus, the latter value is used to calculate $\langle\sigma_H\rangle$, $\langle\sigma_D\rangle$, and $\langle\sigma_T\rangle$. Finally, the expectation value of the chemical shielding of the proton is given by

$$\langle\delta_H\rangle = \delta_H + \sigma^B_H - \langle\sigma^B_H\rangle \qquad (15)$$

where the correction is calculated to be 0.50 ppm for both proteins.

II.g. Miscellaneous. The GAMESS[33] program was used for all calculations, except the NMR calculations, which were performed using GAUSSIAN98.[34]

III. Results and Discussion

III.a. Optimized Geometries of the QM Regions. The 16 atoms in the ab initio region[12] of both models are optimized at the B3LYP/6-31G(d) level of theory. The optimized geometry of α-chymotrypsin (Cht) reproduces the experimental $N^{\delta1}-O^{\delta1}$-distance (r_{NO}) of 2.67 Å and predicts a relatively short $N^{\delta1}-H$ distance of 1.070 Å (Table 1). The optimized geometry of α-lytic protease (Alp) yields an r_{NO} of 2.73 Å (the experimental value is 2.72 Å) and predicts an $N^{\delta1}-H$ distance of 1.059 Å, shorter than in Cht as expected due to the longer r_{NO}. We note that the same models optimized at RHF/6-31G(d) yield less satisfactory distances: r_{NO} = 2.72 Å ($N^{\delta1}-H$ = 1.030 Å) for Cht and r_{NO} = 2.78 Å ($N^{\delta1}-H$ = 1.026 Å) for Alp. The B3LYP/6-31G(d) minimum energy structures for both proteins are verified as minima by computing vibrational frequencies associated with the ab initio region by the partial Hessian vibrational analysis.[14]

III.b H/D Fractionation Factor. The H/D fractionation factor is calculated using eq 6. Using harmonic frequencies and the

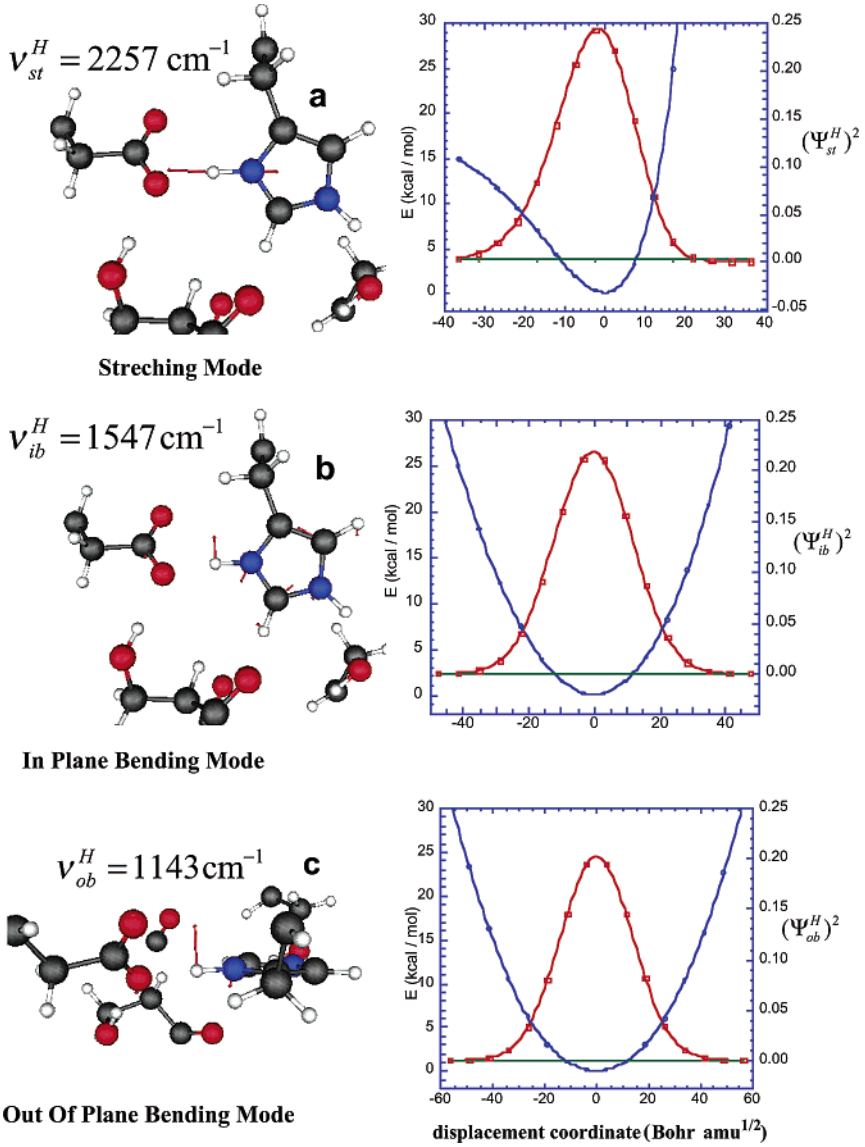

Figure 4. α-Chymotrypsin: plot of the normal mode of the $N^{\delta 1}-H$ (a) stretch, (b) in-plane bend, and (c) out-of-plane bend along with plots of the respective energy surfaces and vibrational probability densities.

TABLE 2: Ground-State Vibrational Energies and Anharmonic Frequencies for the Stretch, In-plane (IP), and Out-of-Plane (OP) Bend of α-Chymotrypsin (Cht), and α-Lytic Protease (Alp), for Both Proton and Deuterium at the $N^{\delta 1}$-H Position[a]

	ground-state energy		anharmonic frequency		harmonic frequency	
mode	H	D	H	D	H	D
Cht						
stretch	1275.66	993.33	2256.99	1906.44	2658.83	2014.00
IP bend	773.19	530.35	1547.49	1062.40	1544.92	1059.82
OP bend	561.77	399.10	1143.37	823.13	1098.63	768.16
Alp						
stretch	1381.62	1052.41	2522.18	2015.58	2850.35	2135.68
IP bend	772.12	526.09	1544.52	1052.69	1544.63	1051.97
OP bend	550.31	377.84	1122.84	767.37	1073.25	742.35
water						
sym str	1940.64	1415.64	3932.85	2859.48	3827.89	2802.29
asym str	1853.16	1337.93	3640.74	2641.47	3723.48	2684.72
Bend	851.25	623.74	1694.77	1244.46	1705.01	1247.68

[a] The corresponding harmonic frequencies are also given. Similar information is also listed for the three vibrational modes of H_2O and D_2O. All values are in cm^{-1} (0.00286 kcal/mol).

rigid rotor approximations to respectively compute the vibrational and translational plus rotational free energies[14] (eq 7), this method results in $\phi = 0.85$ for Cht (somewhat larger than the experimental value of 0.64[35]) and $\phi = 0.95$ for Alp.

The large experimental value of $\Delta\delta_{H/D}$ for Cht[6] (Table 1) indicates the importance of anharmonic effects in the study of the hydrogen-bonding properties of $H^{\delta 1}$(His57) in the active site of these proteins. While computational expense prohibits an anharmonic treatment of all modes, the work by Westler, Weinhold, and Markley,[11] as well as an early study by Weil and Dixon,[36] suggest that good predictions of ϕ can be obtained by an anharmonic treatment of only the zero-point energy of the N−H stretch. To determine which modes to correct for anharmonicity we first determine the major contributions to ϕ of Cht.

For Cht, $\Delta G_{PH/PD}$ (2.079 kcal/mol) is nearly equal to the change in the zero-point energy ($\Delta ZPE = 2.058$ kcal/mol). ΔZPE has significant contributions from only three frequencies: the stretch ($^1/_2 h\Delta\nu_{st} = 0.922$ kcal/mol), in-plane bend ($^1/_2 h\Delta\nu_{ip} = 0.694$ kcal/mol), and out-of-plane bend ($^1/_2 h\Delta\nu_{op} = 0.472$ kcal/mol) of the NH bond (these modes are displayed in Figure 4). All other contributions are more than 1 order of magnitude smaller. A similar conclusion is reached for Alp. Thus, the change in $\Delta G_{PH/PD}^{vib}$ due to anharmonicity is ap-

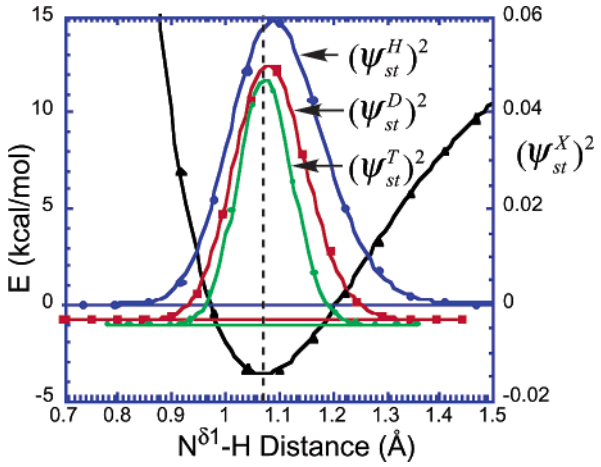

Figure 5. α-Chymotrypsin: plot of the energy surface and the vibrational probability densities for the $N^{\delta 1}$−H/D/T stretch. The horizontal lines denote the ground-state energies of the stretches. The vertical line indicates the minimum energy NH distance of 1.072 Å.

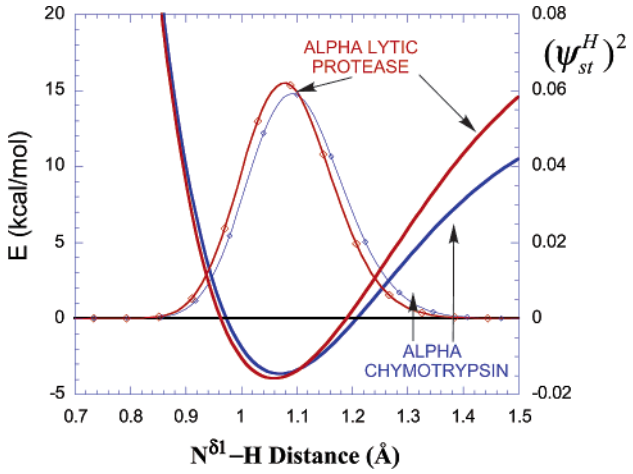

Figure 6. α-Chymotrypsin (blue) and α-lytic protease (red). Plot of the energy surface and the vibrational probability densities for their respective $N^{\delta 1}$-H stretching modes.

proximated by the change in the ground-state vibrational energies ($E^{X}_{st/ib/ob}$) of these three NH and ND modes. These energies are calculated using the ab initio vibrational SCF (VSCF) method of Chaban, Jung, and Gerber[15] for the protonated and deuterated case and used to correct G_{PX} accordingly (X = H or D)

$$G'_{PX} = G_{PX} + E^{X}_{st} + E^{X}_{ib} + E^{X}_{ob} - \frac{1}{2}h(\nu^{X}_{st} + \nu^{X}_{ib} + \nu^{X}_{ob}) \quad (16)$$

The "anharmonic" changes in $\Delta G^{vib}_{PH/PD}$ are −0.127 and −0.061 kcal/mol for Cht and Alp, respectively. The change in $\Delta G^{vib}_{H_2O/D_2O}$ is approximated similarly by considering all three modes (0.008 kcal/mol). Using these corrected G's, the H/D fractionation factor for Cht is 0.67, which is in excellent agreement with the experimental value of 0.64 ± 0.02 (Table 1). The anharmonic effect is almost entirely due to the stretch: if only the stretch is corrected for anharmonic effects, $\phi = 0.68$. The stretch, in-plane, and out-of-plane bend contributes 41%, 35%, and 24% to $\Delta G'_{PH/PD}$ respectively.

For Alp, the corrected G's yield $\phi = 0.85$ (Table 1). While there is no experimental ϕ available for this protein, the larger value is consistent with a larger r_{NO}.[11] It is interesting to note that the corresponding anharmonic correction to the out-of-plane bend energy is greater in magnitude and opposite in sign compared to Cht (−0.020 kcal/mol vs 0.007 kcal/mol). If only the stretch is corrected, $\phi = 0.82$. The difference in $\Delta G^{vib}_{PH/PD}$ for the two proteins is 0.063 kcal/mol while the change in the corresponding anharmonic ΔZPE of the stretch is 0.134 kcal/mol. Thus, the decrease in ϕ on going from Cht to Alp can be qualitatively understood, but not quantitatively predicted, by considering only the stretch. This general point has been made earlier by Edison, Weinhold, and Markley.[12] However, including the change in ΔZPE of the in-plane and out-of plane bends (ΔΔZPE = 0.171 kcal/mol) does not lead to a better approximation of the change in $\Delta G^{vib}_{PH/PD}$ on going from Cht to Alp. The remaining significant contributions to $\Delta\Delta G^{vib}_{PH/PD}$ are small and spread over more than 10 modes, indicating that a quantitative analysis of different ϕ values may be difficult.

III.c. Proton Chemical Shifts (δ_H). Using the minimum energy geometries, the harmonic δ_H's are estimated to be 17.8 and 16.3 ppm for Cht and Alp respectively (Table 1). As shown in Figure 5, the vibrational wave function for the NH stretch (ψ^{H}_{st}) of Cht is clearly skewed toward longer distances. Thus, the expectation value calculated using ψ^{H}_{st} is $<\delta_H>$ = 18.3 ppm (eq 15) and is in good agreement with the experimental value[35] of 18.2 ppm. Similarly, for Alp, $<\delta_H>$ = 16.8 ppm which is also close to the experimental value[2] of 17.1 ppm, considering that eq 8 has a standard deviation of 0.2 ppm.[16]

III.d. Isotope Effects ($\Delta\delta_{H-D}$ and $\Delta\delta_{H-T}$). *(i) α-Chymotrypsin.* The vibrational wave function of the ND stretch is computed and used to predict a $\Delta\delta_{H-D}$ of 0.28 ppm (eq 11), which is smaller than the experimental value[6] of 1.0 ± 0.4 ppm, and a $\Delta\delta_{H-T}$ of 0.44 ppm. However, quadrupole broadening limits the accuracy with which δ_D can be determined,[8] and recently measured[3] $\Delta\delta_{H-T}$ values for related systems support a relatively low δ_{H-D} as discussed next.

Recently, $\Delta\delta_{H-T}$ values of 0.63−0.68 ppm have been measured for three Cht−inhibitor complexes.[3] Since δ_T and δ_H can be measured with similar accuracy, these $\Delta\delta_{H-T}$ values have significantly smaller errors than the $\Delta\delta_{H-D}$. Compared to Cht, the complexes have larger δ_H's (18.61−18.95 ppm) and lower ϕ's (0.32−0.43). Therefore, the $N^{\delta 1}$−H−$O^{\delta 1}$ HB is generally thought to be stronger in the complexes compared to low-pH Cht. Consequently, a $\Delta\delta_{H-T}$ value of <0.63 ppm is expected for Cht, in agreement with our predicted value of 0.44 ppm. Clearly, $\Delta\delta_{H-T}$ provides an upper bound for $\Delta\delta_{H-D}$. Simple mass considerations[32] predict $\Delta\delta_{H-T}/\Delta\delta_{H-D}$ = 1.44 (in good agreement with our predicted ratio of 1.57), so that $\Delta\delta_{H-D}$ for Cht should be smaller than 0.44 ppm = 0.63/1.44. This is consistent with our predicted value of 0.28 ppm. Thus, the experimental value for $\Delta\delta_{H-D}$ of 1.0 ± 0.4 ppm is likely too large.

As expected,[32] the isotope effects derive from a decrease in the asymmetry of the vibrational wave function of the stretch N−X as the mass of the X is increased (Figure 5).

(ii) α-Lytic Protease. Experimental values for $\Delta\delta_{H-D}$ and $\Delta\delta_{H-T}$ are not available for Alp; hence, our calculations present the first predicted isotope effects on δ_H for this protein (Table 1). The vibrational wave function of the ND stretch and NT are computed as before and used to calculate $<\delta_D>$ and $<\delta_T>$ resulting in $\Delta\delta_{H-D}$ = 0.21 ppm and $\Delta\delta_{H-T}$ = 0.35. These lower values are consistent with a weaker bond HB (as suggested experimentally by the longer r_{NO} and lower δ_H measured for Alp, relative to Cht) and are due to a more symmetric stretch PES for Alp compared to Cht (Figure 6).

III.e. The $N^{\delta 1}$−H Distance and $\nu^{H}_{st}/\nu^{D}_{st}$. The expectation values of the $N^{\delta 1}$−H distance and $N^{\delta 1}$−H−$O^{\delta 1}$ angle of Cht are predicted to be 1.102 Å and 170°, respectively. For Alp,

the corresponding values are slightly smaller, 1.086Å and 168.3°, as expected for a weaker HB. The predicted $N^{\delta 1}-H$ distance for Alp is shorter than the experimental estimate[2] of 1.113 ± 0.018 Å from solid-state NMR measurements. However, we note that a previous calibration study of this technique overestimated the corresponding N−H bond length in L-histidine by 0.02 Å compared to neutron diffraction.[37] For both proteins, the $N^{\delta 1}-D$ and $N^{\delta 1}-T$ distances are predicted to be shorter than the $N^{\delta 1}-H$ distance (Table 1), consistent with their lower chemical shifts.

Hibbert and Emsley[5] suggested a fourth spectroscopic characteristic of strong HB (in addition to low-field δ_H, larger $\Delta\delta_{H-D}$, and $\phi < 1$): $1 < \nu_{st}^H/\nu_{st}^D < 1.4$, where ν_{st}^X is the vibrational frequency of the stretching mode in the N−X bond (X = H, D). To our knowledge, this value has not been measured for any serine protease. Here, we offer the first prediction of this ratio for serine proteases. The first excited vibrational states of the NH and ND stretch of Cht yield calculated frequencies of 2257 and 1906 cm^{-1}, respectively. The resulting ratio of 1.18 is halfway between the limiting values of 1.4 and 1.0, consistent with a strong hydrogen bond.[5] The same ratio for Alp is calculated to be 1.25, and as expected, it is slightly larger than for Cht.

IV. Conclusions and Future Directions

This paper presents the first predictive model of strong hydrogen bonding in enzymes. The theoretical model is applied to the $N^{\delta 1}-H-O^{\delta 1}$ hydrogen bond (HB) between His57 and Asp102 in the active sites of low-pH α-chymotrypsin (Cht) and α-lytic protease (Alp). The model is constructed using the effective fragment potential (EFP) method: a QM/MM method that treats the active site by ab initio quantum mechanics and the protein environment by a classical potential. The ab initio and EFP regions are separated by a buffer region of frozen density. We note that the only empirical parameters in our model are the atomic coordinates of the buffer and EFP region and, for the fractionation factors, the partial pressures of H_2O and D_2O. Thus, the experimental data serve to benchmark, rather than calibrate, our model. New information derived from the model can thus be viewed with confidence.

The optimized geometry of the active site Cht reproduces the experimental $N^{\delta}-O^{\delta 1}$ distance (2.67 Å). The structure is used to predict values for the H−D fractionation factor (ϕ), the proton chemical shift (δ_H) of the $(N^{\delta 1})-H$, and changes in the chemical shift upon isotopic substitution ($\Delta\delta_{H-D}$ and $\Delta\delta_{H-T}$). Both ϕ (0.67) and δ_H (18.3 ppm) closely reproduce the experimental data (0.64 and 18.2 ppm respectively). As expected, the calculated tritium isotope effect ($\Delta\delta_{H-T} = 0.44$) is smaller than the experimental $\Delta\delta_{H-T}$ values for Cht−inhibitor complexes (0.63−0.68) where the HB stronger is stronger. It is argued that the calculated $\Delta\delta_{H-D}$ of 0.28 ppm for Cht is a more accurate estimate than the experimental value (1.0 ± 0.4).

Computed parameters for the optimized geometry of Alp are also in good agreement with experimental data. The calculated δ_H (16.8 ppm) and $N^{\delta 1}-O^{\delta 1}$ distance (2.73 Å) compare well to experimental values (17.1 ppm and 2.72 Å, respectively). The calculated $N^{\delta 1}-H$ distance of 1.086 Å for Alp is somewhat shorter than the experimental estimate[2] of 1.113 ± 0.018 Å from solid-state NMR measurements. Predictions are made for other properties for which experimental values are not available. Alp is predicted to have a ϕ of 0.85, a $\Delta\delta_{H-D}$ of 0.21 ppm, and a $\Delta\delta_{H-T}$ of 0.35 (all values consistent with a weaker HB than in Cht as expected). ν_{st}^H/ν_{st}^D, the ratio of vibrational stretch frequencies, is predicted to be 1.18 for Cht and 1.25 for Alp, consistent with their relative HB strength.

The use of the vibrational SCF algorithm[15] to account for anharmonic contributions of the N−H stretching mode is shown to be essential to the accurate computation of these HB properties and provides a powerful theoretical tool to gain a physical understanding of these strong hydrogen bonds.

Similar models of low-pH chymotypsinogen and of the chymotrypsin-N−AcF−CF_3 complex are currently being developed to gain a thorough understanding of the relative strengths of hydrogen bonds in the active sites of serine proteases. The good results obtained in the study of these two proteins indicate that the model presented here is transferable to other systems and can be used to elucidate variations in the strength of the $N^{\delta 1}-H-O^{\delta 1}$ hydrogen bond during the catalytic cycle of serine proteases.

Acknowledgment. This work was supported in part by a Research Innovation Award from the Research Corporation and the NSF (MCB-0209941). The calculations were performed on IBM RS/6000 workstations obtained through a CRIF grant from the NSF (CHE-9974502). Amnon Kohen and Daniel Quinn are thanked for helpful suggestions.

References and Notes

(1) Harris, T. K.; Mildvan, A. S. *Proteins* **1999**, *35*, 275−282.
(2) Bachovchin, W. W. *Magn. Reson. Chem.* **2001**, *39*, S199−S213.
(3) Westler, W. M.; Frey, P. A.; Lin, J.; Wemmer, D. E.; Morimoto, H.; Williams, P. G.; Markley, J. L. *J. Am. Chem. Soc.* **2002**, *124*, 4196−4197.
(4) Robillard, G.; Shulman, R. G. *J. Mol. Biol* **1972**, *71*, 507−511.
(5) Hibbert, F.; Emsley, J. *Adv. Phys. Org. Chem.* **1990**, *26*, 255−379.
(6) Frey, P. A.; Whitt, S. A.; Tobin, J. B. *Science* **1994**, *264*, 1927−1930.
(7) Markley, J. L.; Westler, W. M. *Biochemistry* **1996**, *35*, 11092−11097.
(8) Cassidy, C. S.; Lin, J.; Frey, P. A. *Biochem. Biophys. Res. Commun.* **2000**, *273*, 789−792.
(9) Barich, D. H.; Nicholas, J. B.; Haw, J. F. *J. Phys. Chem. A* **2001**, *105*, 4708−4715.
(10) Garcia-Viloca, M.; Gelabert, R.; Gonzalez-Lafont, A.; Moreno, M.; Lluch, J. M. *J. Phys. Chem. A* **1997**, *101*, 8727−8733.
(11) Westler, W. M.; Weinhold, F.; Markley, J. L. *J. Am. Chem. Soc.* **2002**, *124*, 14373−14381.
(12) Edison, A. S.; Weinhold, F.; Markley, J. L. *J. Am. Chem. Soc.* **1995**, *117*, 9619−9624.
(13) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J. Phys. Chem. A* **2001**, *105*, 293−307.
(14) Li, H.; Jensen, J. H. *Theo. Chem. Acc.* **2002**, *107*, 211−219.
(15) Chaban, G. M.; Jung, J. O.; Gerber, R. B. *J. Chem. Phys.* **1999**, *111*, 1823−1829.
(16) Rablen, P. R.; Pearlman, S. A.; Finkbiner, J. *J. Phys. Chem. A* **1999**, *103*, 7357−7363.
(17) Molina, P. A.; Sikorski, R. S.; Jensen, J. H. *Theo. Chem. Acc.* **2003**, *109*, 100−107.
(18) Minikis, R. M.; Kairys, V.; Jensen, J. H. *J. Phys. Chem. A* **2001**, *105*, 3829−3837.
(19) Kairys, V.; Jensen, J. H. *J. Phys. Chem. A* **2000**, *104*, 6656−6665.
(20) Blevins, R. A.; Tulinsky, A. *J. Biol. Chem.* **1985**, *260*, 4264−4275.
(21) Fujinaga, M.; Delbaere, L. T.; G. D., B.; James, M. N. *J. Mol. Biol.* **1985**, *183*.
(22) Bellido, M. N.; Rullmann, J. A. C. *J. Comput. Chem.* **1989**, *10*, 479−487.
(23) Molina, P. A.; Li, H.; Jensen, J. H. *J. Comput. Chem.* **2003**, accepted.
(24) Graul, S. T.; Brickhouse, M. D.; Squires, R. R. *J. Am. Chem. Soc.* **1990**, *112*, 631−639.
(25) Kakiuchi, M. *Geochim. Cosmochim. Acta* **2000**, *64*, 1485−1492.
(26) Chesnut, D. B. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; 1996; Vol. 8, pp 245−297.
(27) Karadakov, P. B.; Morokuma, K. *Chem. Phys. Lett.* **2000**, *317*, 589−596.
(28) Chesnut, D. B.; Rusiloski, B. E.; Moore, K. D.; Egolf, D. A. *J. Comput. Chem.* **1993**, *14*, 1364−1375.
(29) Mennucci, B.; Martinez, J. M.; Tomasi, J. *J. Phys. Chem. A* **2001**, *105*, 7287−7296.

(30) Porubcan, M. A.; Neves, D. E.; Rausch, S. K.; Markley, J. L. *Biochemistry* **1978**, *17*, 4640−4647.

(31) Harris, R. K.; Becker, E. D.; Cabral de Menezes, S. M.; Goodfellow, R.; Granger, P. *Pure Appl. Chem.* **2002**, *73*, 1795−1818.

(32) Altman, L. J.; Laungani, D.; Gunnarsson, G.; Wennerstrom, H.; Forsen, S. *J. Am. Chem. Soc.* **1978**, *100*, 8264−8266.

(33) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347−1363.

(34) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.6; Gaussian, Inc.: Pittsburgh, PA, 1998.

(35) Bao, D. H.; Huskey, W. P.; Kettner, C. A.; Jordan, F. *J. Am. Chem. Soc.* **1999**, *121*, 4684−4689.

(36) Weil, D. A.; Dixon, D. A. *J. Am. Chem. Soc.* **1985**, *107*, 6859−6865.

(37) Zhao, X.; Sudmeier, J. L.; Bachovchin, W. W.; Levitt, M. H. *J. Am. Chem. Soc.* **2001**, *123*, 11097−11098.