Ab Initio Calculation of Amide Carbonyl Stretch Vibrational Frequencies in Solution with Modified Basis Sets. 1. *N*-Methyl Acetamide

Jan Kubelka and Timothy A. Keiderling*

Department of Chemistry (M/C 111), University of Illinois at Chicago, 845 West Taylor Street, Chicago, Illinois 60607-7061

Received: August 17, 2001; In Final Form: October 12, 2001

Density functional theory DFT(BPW91) level calculations with modified 6-31G(d) basis sets are tested for a small amide, *N*-methyl acetamide (NMA), as an efficient way for calculating amide I and amide II frequencies that are directly comparable to those commonly measured in solution. The calculational results are compared to experimentally measured FTIR spectra in gas and solution phases. The 6-31G(d) basis set at the DFT level yields vibrational frequencies that have the best agreement with the gas-phase experiment, as compared to amide I and II frequencies calculated with the same basis at the HF, CASSCF, MP2, QCISD, and CCD levels. The DFT(BPW91)/6-31G(d) level calculation for the NMA·$3H_2O$ hydrogen-bonded complex with an Onsager or CPCM reaction field yields amide I, II, and III frequencies comparable to the experiment in aqueous solution. The amide I and, to a smaller degree, amide II frequencies are found to be sensitive to the exponent of the d function in the basis set. Use of more diffuse (smaller exponent) d functions in the 6-31G(d) basis set results in a calculated amide I frequency closer to the solution experimental values. Such modified, relatively small basis sets may provide a computationally efficient means of approximating the solvent effects on amide vibrational frequencies.

I. Introduction

Accurate simulations of vibrational infrared (IR) absorption and vibrational circular dichroism (VCD) spectra using ab initio density functional theory (DFT) methods for small molecules are now widespread.[1−10] Calculations with valence double-ζ polarized or higher quality basis sets are capable of predicting vibrational frequencies and IR and VCD intensities in good agreement with experiment.[10−16] For peptides, our DFT level calculations on sizable oligopeptides having up to 102 atoms using a polarized 6-31G(d) basis set have been shown to give useful spectral predictions.[17−22] It has been normal, and in most cases computationally necessary, to assume that these isolated molecule calculations were relevant for predicting spectral properties of real systems, which are typically in solution. However, the experimental vibrational frequencies for oligopeptides, in both aqueous and nonaqueous solutions, differ substantially from those predicted from gas-phase (isolated molecule) calculations. The amide I (mainly C═O stretch) frequencies are systematically predicted to be too high (by at least 100 cm^{-1}),[17,18,21−25] which often correlates with the C═O bond length being too short.[17,26] On the other hand, DFT-predicted amide II (N−H bend and C−N stretch) frequencies are closer to experiment but are typically lower.[17,18,21,22]

Thus, the calculated separation between the amide I (mainly C═O stretch) and amide II frequencies is too large by more than 100 cm^{-1}. This has several negative consequences. First, direct comparison between the experimental amide I frequencies and those predicted from ab initio calculations is difficult. This is unfortunate because the amide I band in terms of frequency and band shape is most sensitive to and is most commonly used for secondary structure determination in peptides and proteins.[27−29] Second, overestimation of the amide I−amide II separation could result in an underestimation of the interaction between these two modes (i.e., mixing of local vibrational motions), causing inaccurate predictions for detailed spectral band shapes in terms of both vibrational frequencies and intensities. Use of a uniform scaling factor, a common method for correcting ab initio force fields, does not solve this problem because, while the amide I mode is too high, the amide II and III modes are typically predicted too low. Because of mixing with CH modes, the errors for amide III prediction are more difficult to delineate, but they appear to be smaller than those for amide II.

Theoretical analysis of the solvent effects is quite problematic. Use of explicit solvent in ab initio calculations is feasible only for small systems. Implicit solvent models, on the other hand, cannot account for hydrogen bonding. The simplest implicit solvent model is the Onsager reaction field[30,31] based on interaction of a solute dipole moment with a continuum dielectric outside a spherical cavity. While this model is robust and computationally efficient, it is in principle suitable only for polar, approximately spherical molecules, but the dipole/sphere approximations may be inappropriate for systems with several polar groups and an extended shape, such as oligopeptides. Implementations of more sophisticated polarized continuum solvent models,[32−35] which adopt more realistic cavity shapes as well as reflect solute charge distributions, usually suffer from convergence problems and, because of the unavailability of analytical second derivatives in available DFT programs, do not offer any computational advantage for frequency calculations on large molecules.

The aqueous solvent problem has been previously addressed for amino acids[36,37] and small amides.[24,26,38] These studies concluded that quantitative agreement between the calculated vibrational frequencies and those measured in aqueous solution

* To whom correspondence should be addressed. Tel: (312) 996-3156. Fax: (312) 996-0431. E-mail: tak@uic.edu.

10.1021/jp013203y CCC: $20.00
Published on Web 11/07/2001

requires both explicit solvent and implicit solvent correction (Onsager reaction field) and may even require a large number of the explicit solvent molecules.[37] Obviously, such complete calculations are a desirable approach to solve this problem, but for larger oligopeptides, such calculations would become prohibitively costly. Thus, it is important to identify efficient approximations for solvent effects, which is the origin of the present work.

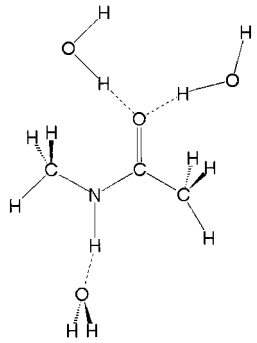
It has been shown for small molecules that DFT-calculated vibrational frequencies are generally superior to those obtained at the Hartree−Fock (HF, SCF)[1,8] level and often also better than those at the MP2[7,9,39] and even QCISD[9] levels. While the computational cost of high-level, correlated ab initio methods is very high and thus their applicability to peptide studies very limited, it is still of interest for this problem to compare the performance of DFT with such methods for amide vibrational frequency calculations.

In this report, we attempt to find an alternative, computationally efficient way to obtain amide I and II vibrational frequencies corresponding to those routinely measured in solution. We examine the effects of basis sets on the amide I and amide II frequencies in a small model amide, *N*-methyl acetamide (NMA). It has been noted before[40−42] that modified 6-31G(d) basis sets, with a more diffuse d polarization function, can mimic much larger basis sets in predicting some molecular properties. The 6-31G($d^{0.3}$) and 6-31G($d^{0.25}$) basis sets (in which the default Gaussian exponent, 0.8, of the d-function in the standard 6-31G(d) or 6-31G* is reduced to 0.3 and to 0.25, respectively) were found to provide better prediction for vibrational frequencies and intensities,[40,41] and DNA base stacking interactions[42,43] than the default 6-31G(d) and even larger basis sets. We show here that the calculated amide I vibrational frequency and the amide I−amide II separation are very sensitive to the exponent of the polarization (d) functions used. Such a modified basis set, with a more diffuse polarization function, thus can be used to obtain amide vibrational frequencies that are in closer correspondence to the solution frequencies. These results are compared to experiment and to results from calculations using larger basis sets, using correlated ab initio methods, and to DFT calculations including implicit and explicit solvent (water).

II. Materials and Methods

Experimental Section. Spectral measurements were carried out on a BioRad Digilab FTS-60A FTIR using an MCT detector. Spectra were collected as an average of 1024 scans at a nominal resolution of 4 cm^{-1}. NMA was purchased from Acros Organics, acetonitrile from Aldrich, and D_2O from Cambridge Isotope Laboratories. For the gas-phase measurement, a homemade, cylindrical glass cell, ∼20 cm long, with fixed CaF_2 windows was used. NMA was placed into a small compartment at the bottom of the cell, out of the beam path. The cell was then evacuated and heated to ∼100 °C, but accurate temperature and pressure could not be determined because of nonuniform heating. The spectrum of the empty gas cell was used as a background for absorbance calculation. Acetonitrile and D_2O solution measurements were performed at concentrations of 14 mg/mL (190 mM) and 10 mg/mL (137 mM), respectively, using a semipermanent IR cell (Specac) with CaF_2 windows and a 100 μm Teflon spacer. The cell path length for determination of extinction coefficient was measured using interference fringes for the assembled empty cell. The H_2O measurement was done at a concentration of 60 mg/mL (820 mM) using the same cell with a 6 μm Mylar spacer. Spectra of just the solvent were collected prior to each sample measurement and used as a background.

CHART 1



Calculations. All calculations were performed using the Gaussian 98 program package[44] on the UIC Chemistry Department Beowulf-type PC cluster (Paralogic Inc., Bethlehem, PA). The calculations with modified basis sets were carried out at the DFT level using a BPW91 functional.[45,46] All modified basis sets used Cartesian d functions, which is the default for the 6-31G(d) (also referred to as 6-31G*) basis set.[35] Larger basis sets with only default parameters (see Results) were tested using those commonly available in Gaussian 98.[35]

In comparison calculations at the MP2[47−50] level, all electrons were used for correlation, while in QCISD[51] and CCD[52] level calculations only the valence electrons were used (frozen core). For the CASSCF[53] calculation, the active space contained six electrons and five orbitals (CASSCF(6,5)): two A″ (π) and one A′ (n) doubly occupied and two unoccupied A″ (π^*) orbitals. Prior to frequency calculation with each method and basis set, the NMA geometry was fully optimized at the same level of theory using the default convergence criteria.

For implicit solvent model (reaction field) calculations, the Onsager[30,31] dipole−sphere and the conductor-like polarized continuum solvent model (CPCM)[34,54] with default parameters for water as a solvent (dielectric constant, $\epsilon = 78.39$) and, for the Onsager reaction field, with the recommended cavity radius of 3.49 Å, obtained from the previous volume calculation for the isolated molecule, were used as implemented in Gaussian 98.[35] An NMA·$3H_2O$ complex was constructed according to coordinates reported previously[55−57] and fully optimized prior to frequency calculation for each basis set (Chart 1). For the NMA·$3H_2O$ calculation, a cavity radius of 4.57 Å was used with the Onsager reaction field. The NMA·$3H_2O$ calculation with the CPCM solvent model was performed using all parameters as default. To estimate the basis set superposition error (BSSE)[58] in the NMA·$3H_2O$ complex, a frequency calculation with a counterpoise correction (CP)[58,59] was carried out at both the BPW91/6-31G(d) and 6-31G($d^{0.3}$) levels. First, the CP-corrected geometry optimization was carried out. Starting from the fully optimized NMA·$3H_2O$ complex geometry, the water molecule atoms were replaced by "ghost" atoms (sets of basis functions with no nuclei or electrons associated with them) the positions of which were fixed while the NMA geometry was reoptimized. The frequency calculation was then run for this NMA complex with "ghost" atoms, the NMA force constants were extracted, and the vibrational frequencies were calculated using a program written in-house.

III. Results

Experimental amide I and amide II spectra for NMA in the gas phase and acetonitrile and aqueous (H_2O and D_2O) solutions are shown in Figure 1. Dipole strengths (**D**) were determined by integration of the amide I′ absorption in D_2O (11.1×10^{-2}

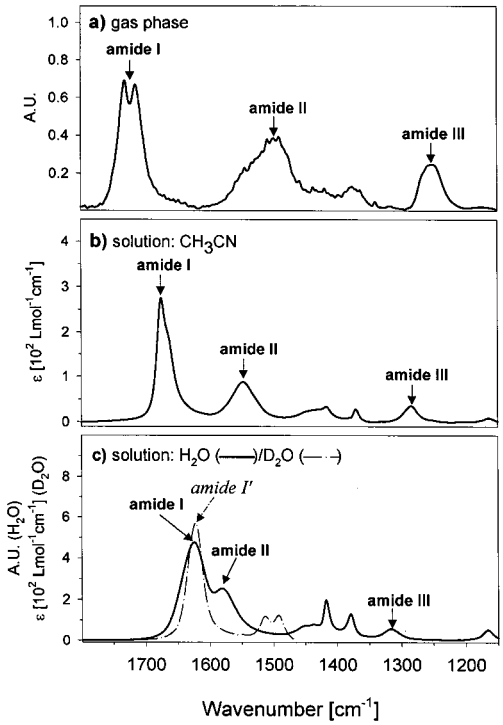


Figure 1. Experimental FTIR spectra in the amide I, II, and III region for *N*-methyl acetamide in (a) the gas phase, (b) acetonitrile solution, and (**c**) aqueous H_2O (solid line) and D_2O (dash−dot) solutions (c). The dipole strengths ($\mathbf{D} = 9.180 \times 10^{-3} \int \epsilon(\nu)\, d\nu/\nu_{max}$) are as follows: for acetonitrile, **D**(amide I) $= 5.2 \times 10^{-2}$ D^2, **D**(amide II) $= 2.6 \times 10^{-2}$ D^2; for D_2O, **D**(amide I′) $= 11.1 \times 10^{-2}$ D^2. Concentrations were 190 mM (acetonitrile) and 137 mM (D_2O).

TABLE 1: Experimental Amide Mode Frequencies for *N*-Methyl Acetamide

phase	amide I (cm^{-1})	amide I′ (cm^{-1})	amide II (cm^{-1})	amide III (cm^{-1})
gas[a]	1731, 1714		1499	1255
gas[b]	1731, 1713		1497	1257
gas[c]	1728		1500	1259
gas[c] (N-deuterated)		1717		
n-hexane[d]	1697			
acetonitrile[a]	1674[f]		1546[g]	1285
acetonitrile[d]	1674			
dichloromethane[d]	1673			
dimethyl sulfoxide[d]	1667			
methanol[d]	1660, 1637			
water (H_2O)[a]	1625		1582	1317
water (H_2O)[d]	1628			
water (H_2O)[e]	1646, 1626		1565, 1585	1313
water (D_2O)[a]		1623[h]		
water (D_2O)[e]		1626		

[a] This study. [b] Reference 60. [c] Reference 61. [d] Reference 62. [e] Reference 63. [f] $\mathbf{D} = 5.2 \times 10^{-2}$ D^2. [g] $\mathbf{D} = 2.6 \times 10^{-2}$ D^2. [h] $\mathbf{D} = 11.1 \times 10^{-2}$ D^2.

D^2) and amide I and II absorption in acetonitrile solution (5.2×10^{-2} and 2.6×10^{-2} D^2, respectively). These frequency results (Table 1) are in agreement with previously published gas-phase[60,61] and solution[61−63] spectra, summarized in Table 1. The gas-phase amide I is a B-type doublet[60] with its center frequency at ∼1723 cm^{-1}, and the amide II has its center frequency at ∼1500 cm^{-1}. Note that the amide II frequency is difficult to determine accurately because of interfering CH_3 deformation modes. In solution, the bands become more distinct, with the amide I frequencies shifting significantly lower and amide II frequencies higher as evident in Figure 1. In addition, the amide I dipole strength in aqueous solution is also significantly greater than that in nonaqueous (Figure 1). Clearly, all solvents have a significant impact on the amide I and II frequencies, with the H_2O effect being the most dramatic and offering unique interactions.

Amide I, II, and III vibrational frequencies, dipole strengths, and C═O bond lengths for NMA calculated using the BPW91 density functional with several common basis sets are listed in Table 2. The unpolarized 6-31G basis set yields a relatively long C═O bond length and amide I and II vibrational frequencies for the isolated NMA molecule that are actually closer to the experimental solution values than to the gas-phase ones (Figure 1, Table 1), as is the amide I−amide II separation. Addition of the polarization (d) function causes a contraction of the C═O bond and a significant shift of the calculated amide I to higher frequency with a much smaller shift of the amide II in the opposite (lower frequency) direction. Thus, the polarized basis yields amide I and II frequencies closer to the gas-phase values. With further increase in the basis set size, both the amide I and amide II shift steadily lower. The amide I−amide II separation is ∼210 cm^{-1} for the 6-31G(d), 6-31G(d,p) and cc-pVDZ basis sets, reduces to ∼180 cm^{-1} when diffuse functions are added (6-31+G(d), aug-cc-pVDZ), and stays approximately the same as the complete basis set limit (6-311+G(3df,p)) is approached (Table 2). Calculated amide III frequencies for the 6-31G(d) basis set slightly underestimate the gas-phase value and essentially follow the same trend as the amide II (shifting to lower frequencies with increasing basis set size) but to a somewhat smaller extent than the amide II. Dipole strengths for both amide I and II show a generally increasing trend with addition of polarization and diffuse functions to the basis set. Note that the amide I and II dipole strengths of the BPW91/6-31G(d) calculation (Table 2) are in a very close agreement with the acetonitrile solution experimental values (Figure 1). The unpolarized 6-31G basis set yields a substantially weaker amide I.

Table 3 lists the results for ab initio calculations with different levels of electron correlation, all using the same polarized 6-31G(d) basis set. In all cases, both the amide I and amide II frequencies are severely overestimated as compared to the gas-phase frequencies, especially at the HF and CASSCF levels with the amide I−amide II separation ≥ 220 cm^{-1}. Including dynamical correlation reduces the amide I frequency by ∼100 cm^{-1}, and the amide I−amide II separation is reduced to ∼180 cm^{-1}. No significant improvement is observed in QCISD and CCD level calculations over the MP2 level results, for which the frequencies are still quite high but, for this set of methods, are those closest to the gas-phase experimental values (Table 3).

Results of DFT vibrational frequency and intensity calculations for NMA using BPW91 and modified exponent basis sets are listed in Table 4. The 6-31G(d$^{0.3}$) and 6-31G(d$^{0.25}$) basis sets were constructed from the 6-31G(d) set by changing the exponent from the default value of 0.8 to 0.3 and 0.25, respectively. (The 6-31G(d) basis set corresponds to 6-31G(d$^{0.8}$) in this notation.) More diffuse d functions primarily affect the amide I frequencies, which shift down more significantly than the amide II frequencies, resulting in amide I−amide II differences of 147 and 132 cm^{-1} for d-function exponents 0.3 and 0.25, respectively. Amide III frequencies, on the other hand, stay essentially unaffected. The dipole strengths for the amide I decrease somewhat for modified basis sets with respect to 6-31G(d) but are the same for both 6-31G(d$^{0.3}$) and 6-31G(d$^{0.25}$). Amide II dipole strengths do not change significantly from those for the 6-31G(d) basis set.

TABLE 2: Calculated Amide Frequencies, Dipole Strengths, and C═O Bond Lengths for NMA with DFT(BPW91) and Several Common Basis Sets

	amide I		amide II				
basis set	frequency (cm^{-1})	**D** (10^{-2} D^2)	frequency (cm^{-1})	**D** (10^{-2} D^2)	amide I−amide II separation (cm^{-1})	amide III frequency (cm^{-1})	C═O bond lenght (Å)
6-31G	1657	3.8	1542	2.1	115	1267	1.262
6-31G(d)	1738	4.9	1533	2.4	205	1240	1.235
6-31G(d,p)	1735	4.9	1522	2.6	213	1231	1.235
6-31+G(d,p)	1702	6.5	1516	3.3	186	1231	1.238
6-31++G(d,p)	1701	6.5	1516	3.3	185	1232	1.238
6-311+G(3df,2p)	1696	6.2	1508	2.9	188	1225	1.228
cc-pVDZ	1724	5.1	1506	3.2	218	1218	1.234
aug-cc-pVDZ	1684	6.1	1510	3.6	174	1223	1.237

TABLE 3: Amide I and II Frequencies for NMA at Different Levels of Theory with a 6-31G(d) Basis Set

level	amide I (cm^{-1})	amide II (cm^{-1})	C═O bond length (Å)
HF/6-31G(d)	1956	1733	1.201
CASSCF(6,5)[a]/6-31G(d)	1950	1724	1.206
MP2/6-31G(d)	1806	1612	1.232
QCISD/6-31G(d)	1805	1614	1.229
CCD/6-31G(d)	1848	1626	1.224

[a] The active space includes six electrons and five orbitals (two A″ and one A′ occupied, and two A″ virtual orbitals).

From comparison of the last two rows in Table 4, it can be seen that if the lower d function exponent is applied to just the oxygen atom, only a small frequency shift is seen. Thus, the modified basis effect on the vibrational frequencies is not solely due to the oxygen lone pair electrons. On the other hand, employing a diffuse d function on just the C and O atoms shifts the amide I frequency even slightly lower than the case when it is employed on all second row (C, N, O) atoms. The amide II frequency remains roughly halfway between the 6-31G(d) and 6-31G($d^{0.3}$) (all second row atoms) results.

One might have guessed that lowering the exponent on the d function would depopulate the d orbital, in comparison to the 6-31G(d) result, thus causing the C═O bond electron density and the resulting frequencies (and intensities) to become closer to those obtained with the 6-31G basis. However, population analysis shows that the opposite is true: the d-orbital population is higher for the 6-31G($d^{0.3}$) than for 6-31G(d) and even higher for the 6-31G($d^{0.25}$) basis set.

Results of vibrational frequency calculations for NMA plus an implicit (reaction field) solvent with parameters for water, NMA plus explicit solvent (3 hydrogen-bonded water molecules), and a combination of both are shown in Table 5. Both types of reaction field lead to a small amide I frequency lowering and little amide II frequency change, yielding an amide I−amide II separation of ∼170 cm^{-1}. The CPCM solvent model a causes more significant shift of the amide I than does the Onsager model (the predicted C═O bond length is also higher) but also more shift of the amide II and the amide III frequencies.

Because of vibrational mixing between the amide I and H−O−H bending of (protonated) water in the NMA·$3H_2O$ complex, four amide I components are calculated, all having some admixture of local amide I coordinate. Amide I′ (N-deuterated) frequencies with deuterated water (D_2O), the bending modes of which are shifted to ∼1220−1250 cm^{-1} and do not mix appreciably with the amide I, are also listed in Table 5. Since the amide I in the NMA−water complex is split by mixing with HOH bending modes into four components, it is impossible to get the amide I−amide II separation exactly. However, because the amide I shift on N-deuteration is typically small (∼4 cm^{-1}, Table 5, rows 2 and 3), the separation can be estimated as the amide I′−amide II difference. This is consistent with experimental results (Table 1) in which the amide I′ is only ∼3 cm^{-1} below the amide I.

The amide I′−amide II separation obtained for NMA·$3H_2O$ with the 6-31G(d) calculation is slightly higher, by ∼5 and 24 cm^{-1}, respectively, than the separation for isolated NMA computed with modified basis sets 6-31G($d^{0.3}$) and 6-31G($d^{0.25}$). The C═O bond length, however, is the same as that obtained with the 6-31G($d^{0.25}$) isolated NMA calculation. The amide frequencies obtained for NMA + $3H_2O$ with the Onsager or CPCM solvent models and those for the complex with just the modified 6-31G($d^{0.3}$) basis set are much closer to the experimental frequencies in aqueous solution (Table 1). The amide I−amide II separations in these calculations and the amide III frequencies, with exception of CPCM, almost exactly match the experimental ones.

Addition of explicit water causes a significant increase in amide I intensities, as do the implicit solvent models, which is consistent with stronger observed dipole strengths in aqueous solution with respect to those in nonaqueous solution (Figure 1). The impact of solvent on the amide II intensities is smaller (except in the case of CPCM); thus, the amide I intensity increases relative to the amide II intensity (Table 5).

The last two lines of Table 5 show the 6-31G(d) and 6-31G($d^{0.3}$) amide I and amide II vibrational frequencies with the counterpoise (CP) correction for basis set superposition error (BSSE) due to the basis functions of water molecules. Both amide I and amide II frequencies in the CP-corrected calculation with a 6-31G(d) basis set shift slightly lower (by ∼15 cm^{-1}) as compared to the vacuum 6-31G(d) calculation, and the C═O bond is ∼0.002 Å longer (Tables 2 and 4). An even smaller effect of BSSE is observed for the 6-31G($d^{0.3}$) basis set. The BSSE is thus not very significant, especially for the amide I−amide II separation.

IV. Discussion

There are discrepancies in measured vibrational frequencies of NMA, especially for the gas phase,[60,61] with some IR data being rather dated and some solution studies focused on only the amide I (amide I′) without providing amide II frequencies.[62] For comparison with the calculated frequencies, we therefore remeasured NMA IR absorption spectra both in gas phase and in solution using a modern FTIR spectrometer. From Figure 1, a distinct (>100 cm^{-1}) contraction in the amide I−amide II splitting with increase in solvent polarity is shown to be a dominant characteristic. The magnitude of these changes is large enough that reasonable calculations should reflect them. That and the strong solvent sensitivity of the amide I frequency[26,62] are two features that we sought to model in our calculations.

Amide I and II vibrational frequencies calculated at the BPW91/6-31G(d) level without scaling are in reasonable

TABLE 4: Calculated Amide I, II, and III Frequencies for NMA Using DFT(BPW91) and 6-31G(d) Basis Sets Including Modified Exponents on Polarization (d) Functions

	amide I		amide II				
basis set	frequency (cm^{-1})	**D** (10^{-2} D^2)	frequency (cm^{-1})	**D** (10^{-2} D^2)	amide I−amide II separation (cm^{-1})	amide III frequency (cm^{-1})	C=O bond length (Å)
6-31G(d) = 6-31G($d^{0.8}$)	1738	4.9	1533	2.4	205	1240	1.235
6-31G($d^{0.3}$)	1666	4.2	1519	2.4	147	1236	1.254
6-31G($d^{0.25}$)	1653	4.2	1521	2.4	132	1243	1.255
6-31G(d) (C, N, H), 6-31G($d^{0.3}$) (O)	1712	4.6	1532	2.3	180	1239	1.240
6-31G(d) (N, H), 6-31G($d^{0.3}$) (C, O)	1663	4.2	1525	2.6	138	1232	1.257

TABLE 5: Calculated Amide I, Amide I′ (N-deuterated), and Amide II Frequencies for Isolated NMA and NMA + $3H_2O$ (D_2O) with the Onsager and CPCM Continuum Solvent Models at the DFT(BPW91) Level

molecule/ basis set/ reaction field	amide (+ HOH)[a] frequency (cm^{-1})	amide I′		amide II				
		frequency (cm^{-1})	**D** (10^{-2} D^2)	frequency (cm^{-1})	**D** (10^{-2} D^2)	amide I′−amide II separation (cm^{-1})	amide III frequency (cm^{-1})	C=O bond length (Å)
NMA/6-31G(d)/ Onsager	1703	1698	9.1	1536	3.7	162	1236	1.239
NMA/6-31G(d)/ CPCM	1690	1686	8.7	1523	7.6	163	1272	1.246
NMA·$3H_2O$/ 6-31G(d)	1728 (14) 1713 (36) 1678 (93) 1672 (73)	1676	9.8	1523	3.3	153	1316	1.255
NMA·$3H_2O$/ 6-31G(d)/ Onsager	1763 (25) 1702 (4) 1681 (4) 1640 (99)	1642	21.3	1582	4.8	60	1318	1.274
NMA·$3H_2O$/ 6-31G(d)/ CPCM	1693 (23) 1688 (26) 1636 (98) 1616 (46)	1636	15.7	1580	6.5	56	1332	1.264
NMA·$3H_2O$/ 6-31G($d^{0.3}$)	1688 (18) 1667 (25) 1648 (7) 1616 (99)	1617	9.2	1556	3.2	61	1308	1.274
NMA CP[b]/ 6-31G(d)	1727	1723	5.6	1520	3.4	203	1251	1.237
NMA CP[b]/ 6-31G($d^{0.3}$)	1661	1657	5.0	1509	3.4	148	1251	1.255

[a] For NMA·$3H_2O$ complexes, in parentheses is given the relative contribution of the NMA motion to the potential energy of the normal mode. [b] Isolated NMA with counterpoise correction for basis set superposition error due to the three H_2O molecules.

agreement with gas-phase experimental data (Figure 1) for NMA. The HF and higher correlated level calculations (with the same, 6-31G(d), basis set) drastically overestimate both amide I and amide II frequencies. The improved predictions using DFT are within the established patterns of vibrational frequency calculations for other molecules.[1,3,7−9,39] The amide I−amide II separation in the HF and correlated ab initio calculations is, however, comparable to that obtained with DFT (BPW91) with the same basis set. It is well-known that basis set convergence of energies and molecular properties for correlated methods is much slower than that for the HF or DFT level calculations.[64,65] However, even use of the larger, correlation-consistent cc-pVDZ basis set for MP2 level calculations (data not shown) did not yield significantly different results from the MP2 calculation with a 6-31G(d) basis set (Table 3). Thus, it is not at all clear whether more accurate vibrational frequencies would be obtained with correlated ab initio methods using very large basis sets. Moreover, these methods even with relatively small basis sets are computationally prohibitive even for smallest peptides, and we could not test larger basis sets for NMA. DFT is thus clearly superior to any other method in terms of optimizing both accuracy and computational cost.

The sensitivity of amide frequencies to the solvent polarity and hydrogen bonding was previously studied theoretically,[23,24,26] and both explicit solvent and reaction field were found to be important for the description of the solvent. Specifically for NMA, only amide I frequency calculations were reported, and moreover, scaled HF frequency calculations were used.[26] In our unscaled DFT(BPW91)/6-31G(d) treatment, inclusion of explicit H_2O solvent for the first hydration shell along with either reaction field model yields both amide I and II frequencies very close to aqueous solution experimental values and with a separation of ~58 cm^{-1} (experimentally ~43 cm^{-1}). By contrast, for NMA with only the explicit solvent and no reaction field or just the reaction field and no explicit solvent, the amide II is much lower and the amide I higher making the separation in both cases more than 150 cm^{-1}.

The progression of change in the amide I with change in the solvent model was suggested by Suhai and co-workers in their previous work on *N*-acetyl-L-alanine-*N*′-methylamide including explicit H_2O's and an Onsager reaction field.[23,24] However, even though they correctly predicted the amide I−amide II separation, both the amide I and amide II vibrational frequencies were significantly overestimated, most likely because of their use of the B3LYP functional.[24] Similar studies by the same group for L-alanine,[36,37] which is not a peptide and therefore not directly related to our work, concluded that both explicit and implicit solvent are necessary for accurate predictions of vibrational

spectra in aqueous solutions. We note that, until our work, only the Onsager reaction field had been used to implicitly model the solvent.[23,24,26,37,38] Although the difference between the Onsager and CPCM solvent models is minor for NMA, the Onsager reaction field is likely to become a much worse approximation for larger extended oligopeptides. This is not only due to their nonspherical shape but, more importantly, due to their complicated electronic density distribution that is poorly described by just a single dipole moment.

Modification of the basis set, strictly speaking, does not improve the predicted amide frequencies because the 6-31G(d) results are a better match to the gas-phase frequencies than are those of the 6-31G($d^{0.3}$). Rather, more diffuse polarization functions seem to provide a correction for the solvent effects to a certain degree, primarily for the amide I mode. Incorporation of diffuse d functions in the basis set elongates the C═O bond, the length of which becomes comparable to that of the explicitly solvated NMA (without a reaction field correction), thus reducing its force constant and lowering its vibrational frequency. The resulting amide I and II frequencies with the DFT(BPW91)/6-31G($d^{0.3}$) or 6-31G($d^{0.25}$) roughly approach the experimental values found in non-hydrogen-bonding solvents such as acetonitrile (Table 1), with both predicted values being somewhat low but their separation being relatively close to the experimental value. However, the amide I and II frequencies for the isolated NMA calculation, even with this modified basis set, are still far from those observed in aqueous solution as well as those calculated with explicit water molecules and a reaction field.

Explicit solvent coupled with this modified, 6-31G($d^{0.3}$), basis set, while underestimating both the amide I and II frequencies slightly, does yield an amide I−amide II separation very close to that found in the calculations using both explicit solvent and reaction field for solvent description. Correlated with this is the longer C═O bond obtained in both methods. Increased population of the more diffuse d orbitals decreases the electron density between the carbonyl C and O atoms, which reduces the bond strength, resulting in longer C═O bond length and lowering the C═O stretching frequency. The hydrogen bonding to the solvent and the polarization of the dielectric (reaction field) have similar effects. However, the modified basis set avoids the computational problems associated with reaction field models, especially those with more realistic solvent cavities, and difficulties with solute charge distributions that go beyond the spherical cavity and single dipole approximation in the Onsager reaction field. Allowing more diffuse d electron density thus may implicitly incorporate the effect of electrostatic interaction with the solvent to some degree. As basis set convergence studies have shown,[66,67] diffuse, high-angular-momentum functions are important for accurate geometry and vibrational frequency calculations, particularly for polar and anionic molecules in solution. When used before, the 6-31G($d^{0.3}$) basis set for vibrational spectral calculations[40,41] and 6-31G($d^{0.25}$) for description of DNA base stacking interactions[42,43] in fact were applied to study of solvated (condensed-phase) species, even though in the former case[41] experimental data were obtained in low-polarity solvents.

It has been shown that diffuse (reduced exponent), high-angular-momentum (polarization) basis functions are generally required for accurate calculation of the electric properties of molecules because the external field stimulates such polarization of the electronic density in the outermost shells.[64,68,69] Use of the highest-angular-momentum function with exponents reduced to about 1/3 of the energy optimized value has been recommended as a minimum for quantitative electric property calculations.[64] It is thus interesting to note the trends in our predicted absorption intensities. Experimentally, the amide I dipole strength increases in aqueous solvent compared to that in a nonaqueous one (Table 1). This is reflected in our calculations, in which addition of explicit solvent as well as reaction field dramatically increases the amide I intensity, resulting in significant overestimation of aqueous amide I dipolar strength in combined explicit and implicit solvent calculations (Table 5). Modified basis sets, on the other hand, consistently yield somewhat weaker dipolar strengths than corresponding calculations with the standard 6-31G(d) basis set. Inclusion of the solvent in the calculations and modifying the basis set with more diffuse d functions thus yield opposite trends in calculated absorption intensities. Nevertheless, the 6-31G($d^{0.3}$) calculation provided reasonably close agreement with NMA intensities in acetonitrile (and perhaps gas phase) for an isolated NMA calculation and with the aqueous experiment for NMA·$3H_2O$ complex calculations, canceling the overestimation.

Because the calculated amide I and II vibrational frequencies, as we have shown, are strongly affected by the diffuse basis functions and also by the hydrogen bonding to the solvent, the question arises whether the explicit solvent effect is not, in fact, largely caused by the basis set superposition error (BSSE). This issue, to our best knowledge, has not been addressed before. It seems (Table 5) that BSSE has only a minor effect on the amide I and II frequencies, shifting both slightly down and thus leaving the amide I−amide II separation nearly unaffected. It is interesting to note that the BSSE seems to be slightly smaller for the modified 6-31G($d^{0.3}$) basis set than for the conventional 6-31G(d) one. Because BSSE is a consequence of the basis set incompleteness, this might indicate that the modified basis set more completely describes the electron density. However, such a conclusion is premature because the CP correction provides only a crude estimate of BSSE.[70,71]

Finally, we note that the amide I and II frequencies obtained with the unpolarized 6-31G basis set (Table 2) are relatively close to the aqueous experimental values and the C═O bond length calculated at this level is nearly the same as that calculated for NMA + $3H_2O$ with the CPCM solvent model (Table 5). Thus, one might wish to propose that the unpolarized basis may provide a simple, computationally efficient approximation for the solvent effects. However, this is most probably an artifact due to inadequate description of the amide group without polarization functions. The d functions are important even as they become more diffuse because, as stated above, the population of more diffuse d orbitals becomes higher. Moreover, unpolarized basis sets are likely to give poor predictions of electric properties, such as spectral intensities, because the presence of the field requires additional polarization of the electron density.[64] This is supported by the significantly underestimated amide I dipolar strength (Table 2).

V. Conclusion

We have shown for a small model amide, *N*-methyl acetamide, that DFT (BPW91) calculations with a relatively small basis set give good prediction of amide I and II frequencies for the gas-phase molecule as well as in solution, if the solvent is properly accounted for. Other ab initio methods, such as HF, MP2, QCISD, and CCD, yield much worse frequency predictions, at least with comparable basis sets. Calculated amide I vibrational frequencies, which correlate to the predicted C═O bond lengths, are very sensitive to the exponent on the polarization (d) function. Modified 6-31G(d) basis sets with

more diffuse d function exponents approximate the effect of the solvent in predicting the C═O bond lengths and the amide I and II vibrational frequencies. Use of DFT(BPW91) with either a 6-31G($d^{0.3}$) or 6-31G($d^{0.25}$) basis set thus seems to provide a computationally efficient way for calculating amide I and amide II vibrational frequencies that better correspond to the commonly measured solution-phase experimental values. Improved representation of the amide I−amide II splitting in solution better accounts for the interaction of these two local modes and may provide more reliable simulations of spectral details.

Acknowledgment. This work was funded in part by a grant from the Petroleum Research Fund administered by the American Chemical Society. J.K. was supported as a UIC Dean's Scholar. We thank Prof. Cynthia J. Jameson for computer time on the PC cluster and Devin Sears for system and software administration and maintenance.

References and Notes

(1) Johnson, B. G.; Gill, B. M. W.; Pople, J. A. *J. Chem. Phys.* **1992**, *98*, 5612−5626.

(2) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623−11627.

(3) Devlin, F. J.; Finley, J. W.; Stephens, P. J.; Frisch, M. J. *J. Phys. Chem.* **1995**, *99*, 16883−16902.

(4) Finley, J. W.; Stephens, P. J. *THEOCHEM* **1995**, *357*, 225−235.

(5) Stephens, P. J.; Ashvar, C. S.; Devlin, F. J.; Cheeseman, J. R.; Frisch, M. J. *Mol. Phys.* **1996**, *89*, 579−594.

(6) Cheeseman, J. R.; Frisch, M. J.; Devlin, F. J.; Stephens, P. J. *Chem. Phys. Lett.* **1996**, *252*, 211−220.

(7) Wong, M. W. *Chem. Phys. Lett.* **1996**, *256*, 391−399.

(8) El-Azhary, A. A.; Suter, H. U. *J. Phys. Chem.* **1996**, *100*, 15056−15063.

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

(10) Halls, M. D.; Schlegel, H. B. *J. Chem. Phys.* **1998**, *109*, 10587−10593.

(11) Devlin, F. J.; Stephens, P. J.; Cheeseman, J. R.; Frisch, M. J. *J. Phys. Chem.* **1997**, *101*, 9912−9924.

(12) Devlin, F. J.; Stephens, P. J.; Cheeseman, J. R.; Frisch, M. J. *J. Phys. Chem.* **1997**, *101*, 6322−6333.

(13) Devlin, F. J.; Stephens, P. J. *J. Phys. Chem. A* **1999**, *103*, 527−538.

(14) Ashvar, C. S.; Devlin, F. J.; Stephens, P. J. *J. Am. Chem. Soc.* **1999**, *121*, 2836−2849.

(15) Aamouche, A.; Devlin, F. J.; Stephens, P. J. *J. Am. Chem. Soc.* **2000**, *122*, 7358−7367.

(16) Aamouche, A.; Devlin, F. J.; Stephens, P. J. *J. Am. Chem. Soc.* **2000**, *122*, 2346−2354.

(17) Kubelka, J.; Silva, R. A. G. D.; Bour, P.; Decatur, S. M.; Keiderling, T. A. In *The Physical Chemistry of Chirality*; Hicks, J. M., Ed.; ACS Symposium Series; Oxford University Press: New York, 2002; pp 50−64.

(18) Kubelka, J.; Keiderling, T. A. *J. Am. Chem. Soc.* **2001**, *123*, 6142−6150.

(19) Kubelka, J.; Keiderling, T. A. *J. Am. Chem. Soc.*, in press.

(20) Bour, P.; Kubelka, J.; Keiderling, T. A., to be submitted for publication.

(21) Bour, P.; Kubelka, J.; Keiderling, T. A. *Biopolymers* **2000**, *53*, 380−395.

(22) Silva, R. A. G. D.; Kubelka, J.; Decatur, S. M.; Bour, P.; Keiderling, T. A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8318−8323.

(23) Jalkanen, K. J.; Suhai, S. *Chem. Phys.* **1996**, *208*, 81−116.

(24) Han, W.-G.; Jalkanen, K. J.; Elstner, M.; Suhai, S. *J. Phys. Chem. B* **1998**, *102*, 2587−2602.

(25) Bohr, H. G.; Jalkanen, K. J.; Elstner, M.; Frimand, K.; Suhai, S. *Chem. Phys.* **1999**, *246*, 13−36.

(26) Torii, H.; Tatsumi, T.; Tasumi, M. *J. Raman Spectrosc.* **1998**, *29*, 537−546.

(27) Surewicz, W.; Mantsch, H. H.; Chapman, D. *Biochemistry* **1993**, *32*, 389−394.

(28) Haris, P. I.; Chapman, D. *Biopolymers* **1995**, *37*, 251−263.

(29) Haris, P. I. In *Infrared Analysis of Peptides and Proteins: Principles and Applications*; Ram Singh, B., Ed.; ACS Symposium Series 750; American Chemical Society: Washington, DC, 2000; pp 54−95.

(30) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486−1493.

(31) Wong, M. W.; Frisch, M. J.; Wiberg, K. B. *J. Am. Chem. Soc.* **1991**, *113*, 4776−4782.

(32) Cossi, M.; Barone, V.; Cammi, R.; Tomasi, J. *Chem. Phys. Lett.* **1996**, *255*, 327−335.

(33) Barone, V.; Cossi, M.; Tomasi, J. *J. Comput. Chem.* **1998**, *19*, 404−417.

(34) Barone, V.; Cossi, M. *J. Phys. Chem.* **1998**, *102*, 1995−2001.

(35) Frisch, A.; Frisch, M. J. *Gaussian 98 User's Reference*; Gaussian, Inc.: Pittsbirgh, PA, 1999.

(36) Tajkhorshid, E.; Jalkanen, K. J.; Suhai, S. *J. Phys. Chem. B* **1998**, *102*, 5899−5913.

(37) Frimand, K.; Bohr, H.; Jalkanen, K. J.; Suhai, S. *Chem. Phys.* **2000**, *255*, 165−194.

(38) Knapp-Mohammady, M.; Jalkanen, K. J.; Nardi, F.; Wade, R. C.; Suhai, S. *Chem. Phys.* **1999**, *240*, 63−77.

(39) Kupka, T.; Gerothanassis, I. P.; Demetropoulos, I. N. *THEOCHEM* **2000**, *531*, 143−157.

(40) Yang, D.; Rauk, A. J. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1996; Vol. 7, pp 261−301.

(41) Yang, D.; Rauk, A. J. *J. Chem. Phys.* **1992**, *97*, 6517.

(42) Hobza, P.; Sponer, J. *Chem. Rev.* **1999**, *99*, 3247−3276.

(43) Florian, J.; Warhel, A. *J. Phys. Chem. B* **1997**, *101*, 5583−5595.

(44) 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*; Gaussian, Inc.: Pittsburgh, PA, 1998.

(45) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(46) Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13244.

(47) Moller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

(48) Krishnan, R.; Pople, J. A. *Int. J. Quantum Chem.* **1978**, *14*, 91.

(49) Krishnan, R.; Frisch, M. J.; Pople, J. *J. Chem. Phys.* **1980**, *72*, 4244.

(50) Frisch, M. J.; Krishnan, R.; Pople, J. *Chem. Phys. Lett.* **1980**, *75*, 66.

(51) Pople, J. A.; Head-Gordon, M.; Raghavachari, K. *J. Chem. Phys.* **1987**, *87*, 5968.

(52) Pople, J. A.; Krishnan, R.; Schlegel, H. B.; Binkley, J. S. *Int. J. Quantum Chem.* **1978**, *14*, 545.

(53) Roos, B. *Adv. Chem. Phys.* **1987**, *69*, 399.

(54) Klamt, A.; Schuurmann, G. *J. Chem. Soc., Perkin Trans.* **1993**, *2*, 799−805.

(55) Guo, H.; Karplus, M. *J. Phys. Chem.* **1994**, *98*, 7104−7105.

(56) Han, W.-G.; Suhai, S. *J. Phys. Chem.* **1996**, *100*, 3942−3949.

(57) Kallies, B.; Mitzner, R. *J. Chem. Soc., Perkin Trans.* **1996**, *2*, 1403−1408.

(58) van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Lenthe, J. H. *Chem. Rev.* **1994**, *94*, 1873.

(59) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 533.

(60) Jones, R. L. *J. Mol. Spectrosc.* **1963**, *11*, 411−421.

(61) Mayne, L. C.; Hudson, B. *J. Phys. Chem.* **1991**, *95*, 2962−2967.

(62) Eaton, G.; Symons, M. R. C.; Rastogi, P. P. *J. Chem. Soc., Faraday Trans. 1* **1989**, *85*, 3257−3271.

(63) Chen, X. G.; Schweitzer-Stenner, R.; Asher, S. A.; Mirkin, N. G.; Krimm, S. *J. Phys. Chem.* **1995**, *99*, 3074−3083.

(64) Helgaker, T.; Taylor, P. R. In *Modern Electronic Structure Theory*; Yarkony, D. R., Ed.; World Scientific: Singapore, 1995; Vol. 2, pp 725−856.

(65) Dunning, T. H.; Peterson, K. A.; Woon, D. E. In *Encyclopedia of Computational Chemistry*; von Rague Schleyer, P., Ed.; Wiley: Chichester, U.K., 1998; Vol. 1, pp 88−115.

(66) Martin, J. M. L.; Taylor, P. R. *Chem. Phys. Lett.* **1994**, *225*, 473−479.

(67) Martin, J. M. L. In *Encyclopedia of Computational Chemistry*; von Rague Schleyer, P., Ed.; Wiley: Chichester, U.K., 1998; Vol. 1, pp 115−128.

(68) Spackman, M. *J. Phys. Chem.* **1989**, *93*, 7594−7603.

(69) Guan, J.; Duffy, P.; Carter, J. T.; Chong, D. P.; Casida, K. C.; Casida, M. E.; Wrinn, M. *J. Chem. Phys.* **1993**, *98*, 4753−4765.

(70) Schwenke, D. W.; Truhlar, D. G. *J. Chem. Phys.* **1985**, *82*, 2418.

(71) Frisch, M. J.; Del Bene, J. E.; Binkley, J. S.; Schaefer, H. F., III *J. Chem. Phys.* **1986**, *84*, 2279.