Influence of Electrostatic Environment on the Vibrational Frequencies of Proteins

Tim M. Watson and Jonathan D. Hirst*

School of Chemistry, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom

Received: February 21, 2003; In Final Form: May 22, 2003

Structure−spectra correlations in protein infrared spectroscopy are well established. Features observable in the amide I region of the spectrum (∼1600−1700 cm^{-1}) correspond well with α-helical (∼1645 cm^{-1}) and β-sheet (∼1610 and 1690 cm^{-1}) structure. To provide a better theoretical understanding of how structure and spectra relate, in this work, we have studied how the electrostatic environment of a protein affects the vibrational characteristics of two small amide molecules (*trans-N*-methylacetamide and *N*-acetylglycine-*N*′-methylamide) when they replace residues at structurally diverse locations within the protein. Four representative environments were examined: α-helical and β-sheet residues that are buried or solvent-accessible. We employed a quantum mechanics/molecular mechanics model using the EDF1 density functional with the 6-31+G* basis set as the quantum calculation and CHARMM22 atom-centered charges for the molecular mechanics model. The electrostatic environment generated by the point charges has a significant effect on the vibrational frequencies, lowering their values from gas-phase values into the typical protein range. The local structure of the protein also has a substantial effect. Finally, calculations incorporating a cage of point charge water molecules showed that solvent can have a profound effect near the surface of the protein.

Introduction

Secondary structure motifs are important determinants of the infrared (IR) spectra of peptides and proteins.[1] This relationship has led to the development and widespread use of techniques, based on experimentally observed spectra, to calculate secondary structure content for proteins.[2,3] These correlations arise from the fundamental physics of the system, and understanding how the IR spectrum of the amide group is affected by its environment is critical for the fullest interpretation of IR experiments. To this end, small amides such as formamide and *N*-methylacetamide (NMA) have been extensively studied.

The computationally cheapest calculation considers the isolated molecule and corresponds to a gas-phase experimental study. We have shown previously[4] that density functional theory (DFT) calculations with the empirical density functional EDF1,[5] in combination with the modest size 6-31+G* basis set,[6] yield unscaled harmonic vibrational frequencies in excellent agreement with experiment. For nine amide systems including formamide, NMA, and *N*-acetyl-L-alanine-*N*′-methylamide, we used MP2/6-31+G*, B3LYP/6-31+G*, B3LYP/6-311++G-(2d,2p), and EDF1/6-31+G* calculations to optimize geometries and compute harmonic vibrational frequencies. On the basis of the mean errors, standard deviations in errors, and root-mean-square deviations from experiment for the amide I, II, and III band positions compared to experimental data, the accuracy follows this pattern:

MP2/6-31+G* < B3LYP/6-31+G* < B3LYP/6-311++G(2d,2p) < EDF1/ 6-31+G*

For all systems, EDF1/6-31+G* gave significantly better agreement with experiment than MP2 or B3LYP calculations with similar and more complete basis sets and at lower computational expense. The remaining discrepancies in the calculations may be due to anharmonicity. For example, the difference between the calculated and experimental amide I band in *trans*-NMA (tNMA) corresponded well with experimental estimates[7,8] of the anharmonicity for this mode.

Many experimental studies of amide systems are conducted in condensed phases. The solution phase is challenging to model theoretically, and two strategies are generally adopted.[9,10] The simpler continuum method involves placing the solute in a cavity and modeling the solvent as a bulk dielectric medium. The alternative technique is to add explicit solvent molecules to the gas-phase system. However, this rapidly becomes computationally expensive for realistic solvation shells. Increasingly, a combination of both strategies is employed, whereby a small number of explicit solvent molecules are placed at key points, e.g., at hydrogen-bond donor/acceptor sites, and bulk effects are treated with a continuum model. While solvent undoubtedly affects the vibrational modes of the protein, in this paper we focus on how the protein itself, in particular its electrostatic environment, affects the vibrational structure. We study this by comparison of gas-phase vibrational frequencies with those computed for a small molecule within an electrostatic environment representative of the protein.

The application of hybrid quantum mechanics/molecular mechanics (QM/MM) models has become more widespread over the past few years,[11,12] with implementations in packages such as CHARMM,[13] GAMESS (U.S.[14] and U.K.[15]), and CADPAC.[16] Initially, semiempirical models were employed for the QM region. However, these have limited applicability and accuracy and have been superseded by Hartree−Fock (HF) or DFT treatments of the QM region. Lyne et al.[17] implemented an interface between CHARMM and CADPAC using both HF and DFT and found that QM/MM simulations compared well with full QM calculations for structures, binding energies, and charge distributions. To enable calculations of vibrational frequencies of large systems with coupling between the QM and MM components, Cui and Karplus[18] formulated a fully analytic Hessian for the QM/MM model. However, codes to perform

10.1021/jp0344500 CCC: $25.00
Published on Web 08/06/2003

these calculations are not yet generally available. In this paper, we employ an uncoupled QM/MM model.

The vibrational properties of small amides, on first inspection, do not resemble those of large proteins, for many well-known reasons. The chain length and conformation affect the vibrational frequencies. In dipeptides, the vibrational modes depend strongly on the dihedral angles ϕ and ψ.[19] These are the through-bond effects. Hydrogen bonding to the amide groups is another example of this class of interaction. Calculations on small amides with explicit water molecules show shifts of up to 20 cm^{-1} in the amide I band upon hydrogen bonding of the water,[20] and studies of hydrogen-bonded dimers have shown similar results.[4,21] Through-space effects are also important and are often understood in terms of a transition dipole coupling mechanism; Torii and Tasumi[22] showed that the amide I band profiles in proteins could be described in this manner. Both through-bond and through-space effects are significant in proteins where chains are long and conformations can vary significantly between secondary structure types. Also, the tertiary/quaternary folded structure of the protein may bring strands into close enough proximity to hydrogen-bond, and therefore their vibrational modes may couple. Teasing apart these various factors would aid our understanding of protein IR.

We present QM/MM studies with EDF1/6-31+G* calculations for the QM region and static CHARMM22 atom-centered point charges for the MM region. We study the effect of protein electrostatic environment on the vibrational modes of small amides in different environments, to dissect the various effects at work in proteins. The environments we study are α-helix and β-sheet, buried and solvent-accessible. For this purpose we select the 76-residue chromosomal protein ubiquitin, due to its small size, monomeric nature, and the presence of all the desired environments. We calculate unscaled harmonic vibrational frequencies for tNMA and for *N*-acetylglycine-*N*′-methylamide (Ac-Gly-NHMe) in the four chosen environments within an unsolvated ubiquitin molecule and then with ubiquitin solvated by MM water molecules.

Computational Details

Small-molecule geometries were built with the molecular editor of Spartan '02.[23] Hydrogen atoms missing from the ubiquitin Protein Data Bank[24] (PDB) file (code 1UBI[25]) were rebuilt with the CHARMM software package.[13,26] Geometries were optimized with the Q-Chem 2.01 ab initio program,[27] and unscaled harmonic vibrational frequencies were calculated by finite difference methods with Q-Chem 1.2 Parallel.[28] Tighter convergence criteria than the Q-Chem defaults were required for optimization; maximum change in the gradient $= 1.0 \times 10^{-5} E_h a_0^{-1}$, maximum atomic displacement $= 1 \times 10^{-4} a_0$, and maximum energy change $= 1.0 \times 10^{-7} E_h$ were used in all cases. Geometries were fully optimized before frequency calculations were performed at the same level of theory. All reported vibrational frequencies are unscaled, and we present the unmodified results of the vibrational calculation.

In our previous DFT study of small amides,[4] the standard SG-1[29] numerical integration grid proved inadequate for fully converging geometries and calculating frequencies. In this work, by contrast, the SG-1 grid performed well; structures were almost identical and the vibrational frequencies within 0.5 cm^{-1} of the results obtained with the more expensive 70-point Euler-Maclaurin radial grid[30] combined with the two-dimensional Lebedev grid[31] with 302 angular points. This suggests that the gas-phase calculations involve a flatter potential and grid errors were therefore a larger factor. In the present point charge model systems, the potential appears to be such that grid errors are less significant.

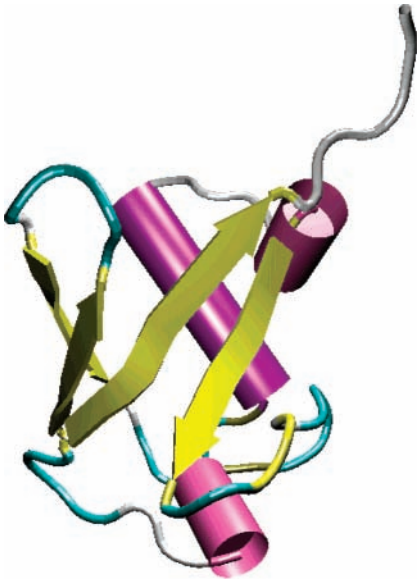

Figure 1. Cartoon representation of the human chromosomal protein ubiquitin, rendered with VMD.[43]

TABLE 1: Chosen Residues for QM/MM Calculation and Their Environments

QM molecule	residue	secondary structure	main-chain rel accessibility (%)
tNMA	^{32}Asp	α-helix	71
tNMA	^{26}Val	α-helix	0
tNMA	^{71}Leu	β-strand	64
tNMA	^{69}Leu	β-strand	0
Ac-Gly-NHMe	^{32}Asp	α-helix	71
Ac-Gly-NHMe	^{26}Val	α-helix	0
Ac-Gly-NHMe	^{71}Leu	β-strand	64
Ac-Gly-NHMe	^{5}Val	β-strand	0

Figure 1 shows a cartoon representation of the human chromosomal protein ubiquitin, depicting secondary structure elements and their relative orientations. In the calculations, we selected two contiguous residues in the protein. We then disconnected these residues from the protein and treated the non-side-chain atoms of this fragment using QM. Selection of the fragment within the protein was based upon two criteria: local secondary structure (α-helix or β-sheet) and solvent accessibility. Secondary structure was determined with DSSP[32] and the solvent accessibilities were calculated with the Naccess[33,34] program. Table 1 shows the environments studied. When the valency of the QM fragment was saturated with hydrogen atoms, the problem of QM atoms overlapping MM charges was overcome by removing the charges from the calculation. Charges that overlapped QM atoms were deleted in a manner that retained the overall charge and minimized local charge polarization. Charges were deleted two neutral units back along the backbone to prevent excessively strong local interactions. Figure 2 depicts an Ac-Gly-NHMe fragment from the α-helical protein backbone, and Figure 3 shows a schematic of the deletion process.

Once coordinates had been built, the geometries of the small molecules were optimized within the static point charge environment. Unlike traditional QM/MM methodology, we did not place any constraints upon the positions or geometry of the molecules. This is significant when vibrational frequency calculations are considered. Analysis and use of frequency data for constrained systems may be undesirable,[35] since nonvan-

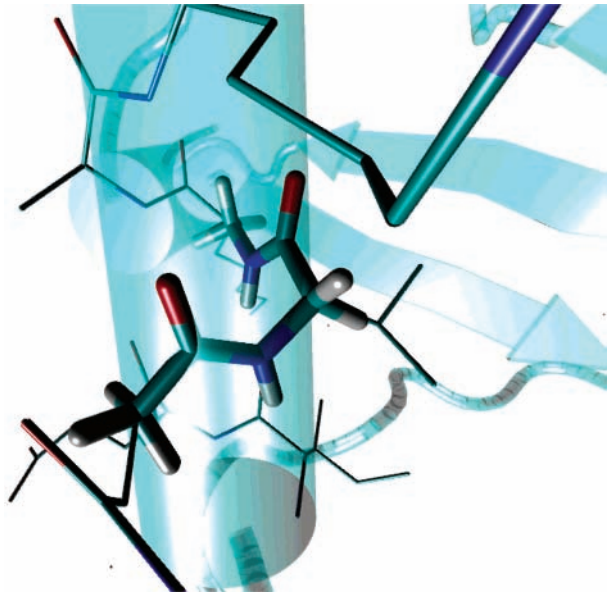

Figure 2. Initial structure of the Ac-Gly-NH-Me molecule (thick tubes) superimposed on the ^{26}Val residue of the backbone (thin tubes), rendered with VMD.[43]

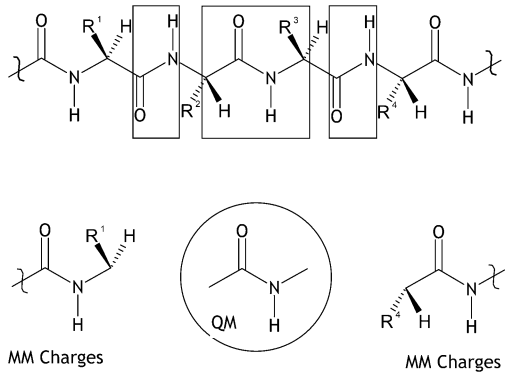

Figure 3. Schematic of the construction of the QM/MM model. Starting from the QM atoms (square solid box), two charge-neutral groups (dashed boxes) are removed back along the backbone and back along the side chain (R^2 and R^3), leaving a QM region (comprising tNMA, as shown, or Ac-Gly-NHMe) in a field of MM point charges.

ishing forces could alter the modes of interest.[36] Although we would not anticipate that constraints on dihedral angles would influence the amide I band significantly, it seemed preferable to avoid the use of constraints. The lack of constraints did not significantly affect the conformation or position of the tNMA molecules in the static point charge environment. The optimization pathway consisted mostly of a torsional relaxation of the methyl groups. The dipeptide systems, however, have much greater conformational flexibility. Hence, we monitored the change in ϕ and ψ dihedral angles of the dipeptides to confirm that they remained in a conformation close to that found in ubiquitin.

In the model described, residues at the surface of the protein are in an environment more akin to the gas phase than at the surface of a solvated protein. Hence, we extended our calculations to include a solvation sphere. The all-atom ubiquitin molecule was solvated in a 62 Å^3 box (containing 7522 TIP3P[37] waters) with cubic periodic boundary conditions. All protein coordinates were fixed to maintain the X-ray crystal structure and to facilitate comparison with the unsolvated model. The system was minimized for 2000 cycles with the CHARMM22

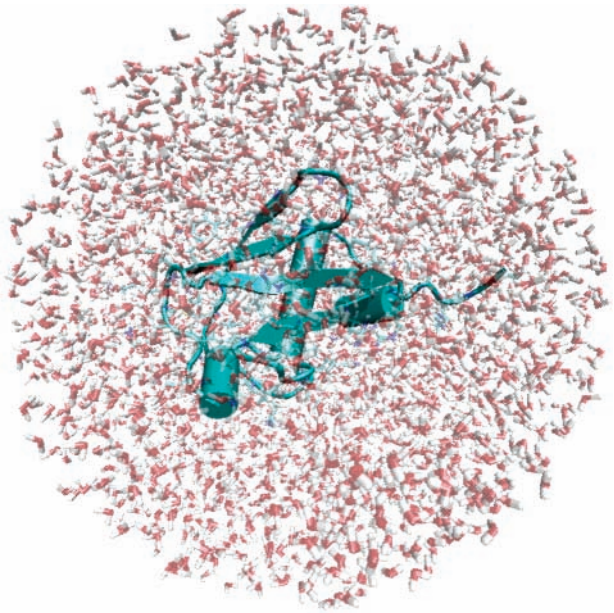

Figure 4. Spherical water solvation shell surrounding ubiquitin.

TABLE 2: Selected EDF1/6-31+G* Optimized Structural Data for tNMA in the Gas Phase and in the Four Unsolvated Protein Environments

	gas phase	^{32}Asp	^{26}Val	^{71}Leu	^{69}Leu
r_{CO} (Å)	1.235	1.259	1.262	1.241	1.239
r_{CN} (Å)	1.374	1.357	1.358	1.368	1.379
r_{NH} (Å)	1.013	1.025	1.024	1.022	1.023
r_{CC} (Å)	1.524	1.518	1.518	1.523	1.519
$r_{NC(Me)}$ (Å)	1.454	1.457	1.457	1.461	1.459
$\angle_{NCO}$ (deg)	122.9	123.3	122.6	122.8	122.3
$\angle_{CCO}$ (deg)	121.6	121.5	122.5	122.1	122.0
$\angle_{CNC}$ (deg)	123.5	123.5	122.7	122.0	121.2
$\angle_{CNH}$ (deg)	118.4	117.3	118.0	120.9	118.7
T1[a] (deg)	0.1	164.5	164.1	−167.1	−6.4
T2[b] (deg)	179.8	−18.1	154.5	−159.4	−164.5

[a] T1 = torsion [C(O)-N−C−H(ip)]. [b] T2 = torsion [H(ip)-C−C=O].

force field and the adopted basis-set Newton−Rhapson algorithm. After minimization, a 10 ps period of dynamics was run with the protein coordinates still fixed. We utilized the constant pressure/temperature method (the temperature was maintained at 298 ± 5 K) with the leapfrog algorithm and the bycube image nonbond list generator. A 2 fs time step and a nonbonded cutoff of 12 Å were used. The cube was then trimmed to form a spherical solvation shell of radius 30 Å (Figure 4) and the resulting water molecules were appended as point charges to the QM/MM calculation.

Results and Discussion

tNMA. Table 2 shows selected optimized structural data for the four point charge models, alongside the gas-phase EDF1/6-31+G* calculation reported previously. The most significant deviations in the structures compared to gas-phase calculations occur in the methyl group torsion angles and the C=O bond length. The MM charges generate a nonzero electric field that may cause translation or rotation of the QM molecules. Table 3 shows the center of mass displacements and rotations upon optimization of the QM molecule. The center of mass for the molecules is essentially stationary during the optimization. The tNMA molecules rotate within the MM field by between 3.8° and 24.4°. Considering the simplicity of the model, we feel that these changes are acceptable, although it would be difficult to

TABLE 3: Observed Center of Mass Displacements and NCO Plane Rotation for the QM Molecules during Geometry Optimization

	monomer	^{32}Asp	^{26}Val	^{71}Leu	^{69}Leu
center of mass displacement (Å)	unsolvated	0.0040	0.0116	0.0072	0.0048
center of mass displacement (Å)	solvated	0.0088	0.0130	0.0067	0.0039
ω^a (deg)	unsolvated	14.3	18.5	13.5	3.8
ω (deg)	solvated	16.3	24.4	12.6	10.3

	dimer	^{32}Asp	^{26}Val	^{71}Leu	^{5}Val
center of mass displacement (Å)	unsolvated	0.0090	0.0056	0.0048	0.0057
center of mass displacement (Å)	solvated	0.0005	0.0055	0.0067	0.0066

[a] ω is defined as the angle between the normals to the NCO planes of the starting and optimized geometries.

TABLE 4: Calculated Amide I−III Bands for tNMA in the Gas Phase and in the Four Unsolvated Protein Environments from EDF1/6-31+G*

		amide I	amide II	amide III
N_2 matrix isolated data[42]	E (cm^{-1})	1707	1511	1266
gas-phase calculation[4]	E (cm^{-1})	1717	1541	1252
gas-phase calculation[4]	I (km mol^{-1})	280	121	107
^{32}Asp	E (cm^{-1})	1646	1592	1314
^{32}Asp	I (km mol^{-1})	317	177	59
^{26}Val	E (cm^{-1})	1636	1580	1306
^{26}Val	I (km mol^{-1})	281	184	56
^{71}Leu	E (cm^{-1})	1700	1558	1280
^{71}Leu	I (km mol^{-1})	276	179	86
^{69}Leu	E (cm^{-1})	1707	1558	1265
^{69}Leu	I (km mol^{-1})	261	150	118

quantify their effects upon the vibrational frequencies. Table 4 shows the calculated amide I, II, and III bands for the four model systems and for tNMA in the gas phase. Experimentally, the matrix-isolated tNMA amide I, II, and III frequencies have been determined to be 1707, 1511, and 1266 cm^{-1}, respectively. There are changes in the energies and intensities of all three bands. In accord with previous theoretical studies of solvated amide systems,[38] we observe a decrease in the amide I frequency, as might be expected from the lengthened C═O bond.

Generally, α-helices have an amide I band at ∼1650 cm^{-1}; β-sheets have a strong 1620 cm^{-1} band and possibly a weak 1690 cm^{-1} band.[22] Our results show for the monomer an amide I band at ∼1645 cm^{-1} in an α-helical environment and ∼1690 cm^{-1} for β-sheet environment. Coupling of vibrational modes between monomer units would need to be included to model the amide I bands of different secondary structures. The coupling could be introduced by using our monomer amide I frequencies as the diagonal force constants in a transition dipole coupling calculation.[22] Its neglect here most likely accounts for the difference between the experimentally observed band in β-sheets at 1620 cm^{-1} and the calculated monomer value of 1690 cm^{-1}. The dimer calculations that follow begin to address the issue of coupling. Nevertheless, it is noteworthy that pure electrostatics significantly affect the vibrational frequencies (with a shift of between 10 and 81 cm^{-1} from gas-phase calculations) and, for α-helices, appear to be a dominating contribution to the decrease in the amide I band frequency between gas-phase and protein environments. Both α-helical and β-sheet residues exhibit small differences between solvent-accessible and buried residues of 10 and 7 cm^{-1}, respectively, for amide I. The amide II band shows similar differences (12 and 0 cm^{-1}, respectively).

Since both the amide I and II bands have considerable intensity, the two bands would be resolved experimentally. In the gas phase the difference in the calculated frequencies of the amide I and II bands is 176 cm^{-1} at the EDF1/6-31+G* level. The α-helical residues exhibit a separation of the amide I and II bands of 56 and 54 cm^{-1} for the buried and solvent-accessible residues, respectively. The β-sheet residues exhibit an amide I−II separation of 149 and 142 cm^{-1} for the buried and solvent-accessible residues, respectively. These data suggest that the amide I−II band separation might be a useful indicator of secondary structure, although again there is the caveat that coupling between amide groups will modify the separation between amide I and II bands.

Ac-Gly-NHMe. Unlike the tNMA models, the Ac-Gly-NHMe systems studied exhibited some structural changes upon optimization. Table 5 shows how the dihedral angles change from those found in the native protein structure upon optimization of the dipeptide. The mean unsigned deviation is ∼6°, an acceptable shift considering the uncertainty in the experimental structure and the relative simplicity of the model. As for the monomers, Table 3 shows that the center of mass displacements upon optimization are negligible. Rotation data for Ac-Gly-NHMe are difficult to interpret, due to the changes in internal geometry of the dipeptide, but the rotations are similar to those for the monomers. Detailed geometry information is provided as Supporting Information.

Table 6 shows the calculated amide I−III bands of the dipeptide in the four protein environments studied. Unlike tNMA, comparison of our models with gas-phase data is not trivial. The dependence of the vibrational modes upon the ϕ and ψ dihedral angles is well-known and would require us to compare the model dipeptides with gas-phase models at the same ϕ and ψ. As discussed before, constrained frequency calculations are not necessarily meaningful, and no unconstrained minima on the Ac-Gly-NHMe potential energy surface correspond to our model geometries.

The most significant difference between tNMA and Ac-Gly-NHMe is the addition of the second amide unit that will couple to the first. This leads to a second peak in the amide I, II, and III (and a third peak in the amide III) bands. As seen in Table 6, the splitting of the peaks within a band depends on their environment, ranging from 9 to 50 cm^{-1} for the amide I band. These differences are encouraging for the differentiation of secondary structure from spectra. A caveat is that solvent-accessible regions in this gas-phase model are effectively in a vacuum environment and we should consider the solvated calculations for this kind of information.

tNMA/Solvated Protein. Figure 5 shows a schematic of the environment of the tNMA placed at residue 71, and Table 7 shows optimized structural data for the four protein environments. There is little change in the geometries of the tNMA molecules between the unsolvated and solvated environments. Upon solvation, all of the vibrational modes were altered (Table 8). Previous studies of small amides hydrogen-bonded to water have shown an increase in C═O bond length of ∼0.01 Å with a corresponding reduction in the amide I frequency of ∼10−15 cm^{-1}.[20] This is what we observe for our calculations. For the two solvent-inaccessible residues, ^{26}Val and ^{69}Leu, there are small changes in the C═O bond length and the amide I mode (0 and 10 cm^{-1}, respectively) compared to the unsolvated protein (Table 4). For the two solvent-accessible residues, the changes in the amide I frequency are larger, 6 cm^{-1} for ^{32}Asp and 21 cm^{-1} for ^{71}Leu, with a corresponding increase in C═O bond length of ∼0.01 Å. A similar pattern is seen for the amide

TABLE 5: Dihedral Angles[a] (ϕ, ψ) of the Protein Residues Replaced by Ac-Gly-NHMe

residue	^{32}Asp	^{26}Val	^{71}Leu	^{5}Val
experimental protein crystal structure	(−59, −39)	(−63, −43)	(−67, 153)	(−116, 107)
EDF1/6-31+G* (unsolvated protein charges)	(−66, −29)	(−63, −38)	(−72, 151)	(−92, 113)
EDF1/6-31+G* (solvated protein charges)	(−50, −48)	(−68, −52)	(−65, 130)	(−95, 113)

[a] All dihedral angles are given in degrees.

TABLE 6: Calculated Vibrational Data for Ac-Gly-NHMe in the Four Unsolvated Protein Environments at the EDF1/6-31+G* Level

		amide I		amide II		amide III		
^{32}Asp	E (cm^{-1})	1710	1660	1551	1541	1294	1273	1238
^{32}Asp	I (km mol^{-1})	300	233	119	185	24	59	84
^{26}Val	E (cm^{-1})	1671	1662	1572	1528	1288	1258	1256
^{26}Val	I (km mol^{-1})	341	166	136	166	39	150	14
^{71}Leu	E (cm^{-1})	1714	1693	1563	1531	1286	1276	1223
^{71}Leu	I (km mol^{-1})	185	338	167	137	59	60	43
^{5}Val	E (cm^{-1})	1675	1656	1577	1575	1299	1277	1225
^{5}Val	I (km mol^{-1})	246	258	136	177	32	143	38

TABLE 7: Calculated Optimized Selected Structural Data for tNMA in the Gas Phase and in the Four Solvated Protein Environments from EDF1/6-31+G*

	gas phase	^{32}Asp	^{26}Val	^{71}Leu	^{69}Leu
r_{CO} (Å)	1.235	1.259	1.262	1.247	1.243
r_{CN} (Å)	1.374	1.366	1.354	1.366	1.370
r_{NH} (Å)	1.013	1.028	1.022	1.022	1.023
r_{RC} (Å)	1.524	1.510	1.519	1.521	1.520
$r_{NR'}$ (Å)	1.454	1.449	1.459	1.461	1.463
$\angle_{NCO}$ (deg)	122.9	120.7	122.4	121.7	123.0
$\angle_{RCO}$ (deg)	121.6	123.7	122.7	123.5	121.2
$\angle_{CNR'}$ (deg)	123.5	120.2	123.2	120.0	122.0
$\angle_{CNH}$ (deg)	118.4	118.6	118.1	117.7	118.4
T1[a] (deg)	0.1	−177.9	174.4	−165.2	−9.55
T2[b] (deg)	179.8	145.9	24.3	−160.7	175.2

[a] T1 = torsion [C−N−C′−H(ip)]. [b] T2 = torsion [H(ip)-R−C═O].

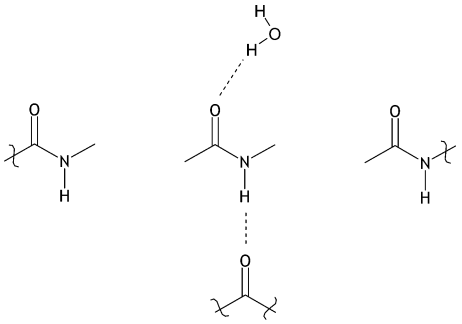

Figure 5. Local solvated residue environment for tNMA at ^{71}Leu. The central tNMA molecule is the QM molecule; all other atoms are MM point charges.

II and III bands; the frequencies of the solvent-accessible residues are more sensitive to the solvation shell than the buried residues.

Ac-Gly-NHMe/Solvated Protein. As shown in Table 5, there is a change in dihedral angles associated with the solvation and subsequent optimization. As before, we believe that these changes are within acceptable ranges. Interpretation of the Ac-Gly-NHMe data is more complicated due to the slight changes in dihedral angles of the dipeptides. However, we do see broadly the same changes in band positions (Table 9). The two buried residues exhibit shifts of 2−11 cm^{-1} in the amide I frequencies, and the solvent-accessible residues show a 13−49 cm^{-1} decrease. These latter shifts are quite considerable and are not wholly uniform. For ^{32}Asp we observe a 50 cm^{-1} splitting of

TABLE 8: Calculated Amide I−III Bands for tNMA Molecule in the Gas Phase and in the Four Solvated Protein Environments from EDF1/6-31+G*

		amide I	amide II	amide III
N_2 matrix isolated data[42]	E (cm^{-1})	1707	1511	1266
gas phase	E (cm^{-1})	1717	1541	1252
gas phase	I (km mol^{-1})	280	121	107
^{32}Asp	E (cm^{-1})	1640	1574	1306
^{32}Asp	I (km mol^{-1})	231	228	91
^{26}Val	E (cm^{-1})	1636	1582	1310
^{26}Val	I (km mol^{-1})	287	187	55
^{71}Leu	E (cm^{-1})	1679	1565	1285
^{71}Leu	I (km mol^{-1})	284	174	91
^{69}Leu	E (cm^{-1})	1697	1573	1280
^{69}Leu	I (km mol^{-1})	296	171	101

TABLE 9: Calculated Vibrational Data for Ac-Gly-NHMe in the Four Solvated Protein Environments at the EDF1/6-31+G* Level

		amide I		amide II		amide III		
^{32}Asp	E (cm^{-1})	1679	1647	1572	1564	1304	1292	1245
^{32}Asp	I (km mol^{-1})	366	163	174	149	60	121	9
^{26}Val	E (cm^{-1})	1682	1665	1568	1538	1299	1272	1258
^{26}Val	I (km mol^{-1})	433	65	148	193	29	139	21
^{71}Leu	E (cm^{-1})	1665	1660	1578	1556	1296	1249	1155
^{71}Leu	I (km mol^{-1})	313	173	153	167	84	67	18
^{5}Val	E (cm^{-1})	1673	1659	1577	1576	1280	1228	1159
^{5}Val	I (km mol^{-1})	227	329	142	159	130	26	8

the amide I band in the unsolvated protein but a splitting of 32 cm^{-1} in the solvated model. The other solvent-accessible residue, ^{71}Leu, also shows a decrease in the splitting of the amide I band, from 21 cm^{-1} in the unsolvated protein to 5 cm^{-1} in the solvated protein. The relative intensities of the two transitions in the amide I band become more pronounced in the solvated protein model. In the case of ^{71}Leu, the relative intensities actually reverse and the higher energy transition is calculated to be more intense in the solvated protein model. This is most likely attributable to the difference in dihedral angles.

Carbonyl Bond Lengths. The amide I vibration consists mainly of the C═O stretch with smaller contributions from the N−H in-plane bend and the C−N stretch. Intuitively we expect a high degree of correlation between the C═O bond length and the amide I stretch frequency. Previous studies have reported this effect. QM studies[20] of the association of tNMA with up to three water molecules showed a striking correlation between C═O bond length and amide I stretch frequency. A similar study[21] of hydrogen-bonded tNMA dimers and trimers exhibit the same effect. For our tNMA monomer calculations, the amide I stretch is, by definition, localized in one peptide group. Since the C═O stretch coordinate dominates the normal mode and on the basis of prior work, a strong correlation was expected, and this is what we observe. Figure 6 includes all calculated bond lengths and frequencies in this study. For the dipeptides, the amide I band is delocalized across the two amide groups. Hence a stretch is assigned to the particular peptide group with the greatest contribution to the normal mode. As for the monomers, we observe a linear relationship, although there is greater scatter in the data. This scatter may be representative of the degree of coupling between the two peptide groups. While

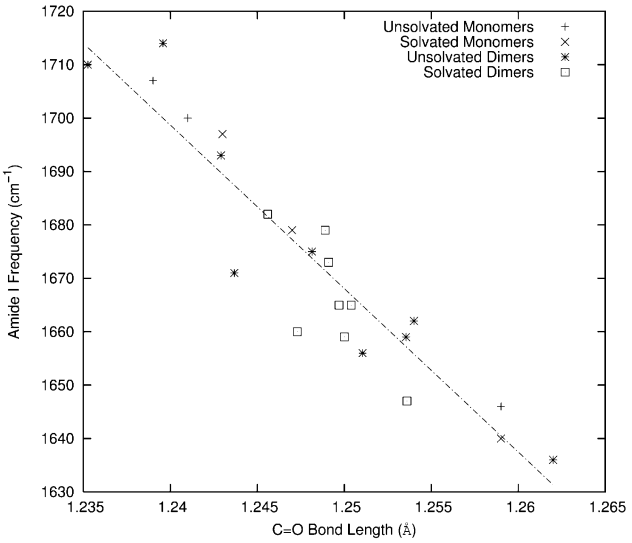

Figure 6. Variation of C═O bond length with amide I stretch frequency for tNMA and Ac-Gly-NHMe.

the absolute bond lengths differ from previous studies (most likely due to the differing levels of theory), the trend is essentially identical.

Conclusion

We have used an uncoupled QM/MM model to examine the effect of protein electrostatics on the vibrational spectra of tNMA and Ac-Gly-NHMe. Our protocol (shown in Figure 3) avoids the link atom problem and allows direct comparison with isolated small molecules, which would otherwise not be possible. It has been used to assess the potential contribution of electrostatics, which has turned out to be sizable. The magnitude of the contribution may even be larger than our calculations suggest, as a relatively conservative protocol was employed in which the closest charged MM atoms were excluded. Nevertheless, the model is not perfect, and effects other than electrostatics are also important. The model neglects any steric effects as well as any changes due to the anchoring of the peptide units to the rest of the chain. However, inclusion of these would require the systems to be treated with a fully coupled QM/MM calculation. It is not consistent to generate geometries with these effects included unless the potential has been correctly modified and, as noted before, codes to perform these types of calculations are not yet widely available.

In previous work,[4] we have shown how the local ϕ and ψ dihedral angles affect the vibrational modes for each of the gas-phase minima on the Ac-Gly-NHMe potential energy surface and noted that the effects could be substantial. In this study, we have observed both local and nonlocal effects. We have seen how discrimination between α-helical and β-sheet type environments could be possible due to the differential effect of the local protein environment on the amide I, II, and III bands. We have also observed changes due to hydrogen bonding at the protein surface when solvent-accessible residues interacted with explicit water molecules. Bulk effects have also been observed. Comparison of vibrational frequencies for the buried residues in the unsolvated and solvated models shows a small change in frequencies attributable to long-range electrostatic interactions with the bulk solvent.

The interplay between various factors seems not to be simple. For example, the calculated amide I frequencies for tNMA in the unsolvated protein (Table 4) and solvated protein (Table 8) indicate a discernible difference between the α-helical and β-sheet residues. Calculations of Ac-Gly-NHMe in the unsolvated protein suggest that solvent exposure is the more important determinant of the amide I frequencies. However, the calculations on the Ac-Gly-NHMe in the solvated protein show rather smaller differences between the amide I bands in α-helices versus β-sheets and between buried and exposed residues. The last of these calculations is arguably the most realistic model. It suggests that the electrostatic environments of the solvent and of the protein have similar effects on the amide I frequencies, and little difference emerges between the solvent-accessible and buried residues. There is some indication that the α-helical residues compared to β-strand residues have slightly higher amide I frequencies, that the splitting within the amide I band is slightly greater, and that the intensity of the band is more localized in one of the vibrations.

Recent developments in 2D-IR experiments have shown promise in extracting detailed structural information from molecular vibrations.[39] The technique allows the study of couplings between oscillators. The link between coupling and structure is usually considered within a transition dipole coupling formalism that explicitly includes a distance and orientation dependence. However, the simple transition dipole coupling mechanism has been found to be too crude for accurate prediction and interpretation of 2D-IR when used to describe nearest neighbors. Limitations of the transition dipole coupling model have been discussed previously.[40] The coupling parameter has been shown to be best determined by performing ab initio calculations. These calculations are usually performed in the gas phase or with simple solvent models. Although 2D-IR is currently being used to study small molecules, it is hoped that it will be possible to study complete proteins in a way similar to 2D-NMR techniques. It is unlikely that the gas-phase coupling data will provide a good description within the protein environment, and our calculations could be used to generate more suitable parameters.

We have used molecular dynamics simulations with fixed protein coordinates to generate a single structure to represent the solvated protein. It would be interesting to explore the effect of dynamics by generating an ensemble of structures. Currently, computational demands would only permit low-level calculations. Despite some impressive calculations,[41] fully ab initio calculations of the IR spectra of proteins are likely to be computationally infeasible for some time yet, and more approximate calculations, based on empirical spectroscopic force fields,[1] offer a tractable alternative. Using data from these QM/MM studies, we aim to improve the accuracy of such methods.

Acknowledgment. We thank the EPSRC for a studentship to support T.M.W. We thank Dr. Andrew Gilbert for useful discussions and technical assistance with Q-Chem.

Supporting Information Available: Tables of optimized (un)solvated monomer and dimer geometries and frequency data. This information is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Krimm, S.; Bandekar, J. *Adv. Protein Chem.* **1986**, *38*, 181.
(2) Simonetti, M.; Di Bello, C. *Biopolymers* **2001**, *62*, 95.
(3) Wi, S.; Pancoska, P.; Keiderling, T. A. *Biospectroscopy* **1998**, *4*, 93.
(4) Watson, T. M.; Hirst, J. D. *J. Phys. Chem. A* **2002**, *106*, 7858.
(5) Adamson, R. D.; Gill, P. M. W.; Pople, J. A. *Chem. Phys. Lett.* **1998**, *284*, 6.
(6) Hehre, W. J.; Radom, L.; Schleyer, P. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; John Wiley & Sons: New York, 1986.

(7) Hamm, P.; Lim, M. H.; Hochstrasser, R. M. *J. Phys. Chem. B* **1998**, *102*, 6123.
(8) Peterson, K. A.; Rella, C. W.; Engholm, J. R.; Schwettman, H. A. *J. Phys. Chem. B* **1999**, *103*, 557.
(9) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027.
(10) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161.
(11) Gao, J. L.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467.
(12) Gogonea, V.; Suarez, D.; Vaart, A. v. d.; Merz, K. M., Jr. *Curr. Opin. Struct. Biol.* **2001**, *11*, 217.
(13) MacKerell, A. D.; Brooks, B. R.; Brooks, C. L., III.; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. CHARMM: The Energy Function and Its Parametrization with an Overview of the Program. In *The Encyclopedia of Computational Chemistry*; John Wiley & Sons: Chichester, U.K., 1998; Vol. 1, p 271.
(14) 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.
(15) Guest, M. F.; Lenthe, J. H.; Kendrick, J.; Schoffel, K.; Sherwood, P.; Amos, R. D.; Buenker, R. J.; Dam, H. J. J.; Dupois, M.; Handy, N. C.; Hillier, I. H.; Knowles, P. J.; Bonacic-Koutecky, V.; von Niessen, W.; Harrison, R. J.; Rendel, A. P.; Saunders, V. R.; Stone, A. J.; Tozer, D. J.; de Vries, A. H. GAMESS-UK, 1980.
(16) Amos, R. D.; Alberts, I. L.; Andrews, J. S.; Colwell, S. M.; Handy, N. C.; Jayatilaka, D.; Knowles, P. J.; Kobayahi, R.; Laidig, K. E.; Laming, G.; Lee, A. M.; Maslen, P. E.; Murray, C. W.; Rice, J. E.; Simandiras, E. D.; Stone, A. J.; Su, M. D.; Tozer, D. J. *CADPAC: The Cambridge Analytic Derivatives Package*, Issue 6 ed.; Cambridge University: Cambridge, U.K., 1995.
(17) Lyne, P. D.; Hodoscek, M.; Karplus, M. *J. Phys. Chem. A* **1999**, *103*, 3462.
(18) Cui, Q.; Karplus, M. *J. Chem. Phys.* **2000**, *112*, 1133.
(19) Mirkin, N. G.; Krimm, S. *J. Phys. Chem. A* **2002**, *106*, 3391.
(20) Torii, H.; Tatsumi, T.; Tasumi, M. *J. Raman Spectrosc.* **1998**, *29*, 537.
(21) Torii, H.; Tatsumi, T.; Kanazawa, T.; Tasumi, M. *J. Phys. Chem. B* **1998**, *102*, 309.
(22) Torii, H.; Tasumi, M. *J. Chem. Phys.* **1992**, *96*, 3379.
(23) *Spartan* '02; Wavefunction Inc.: Irvine, CA, 2001.
(24) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235.
(25) Ramage, R.; Green, J.; Muir, T. W.; Ogunjobi, O. M.; Love, S.; Shaw, K. *Biochem. J.* **1994**, *299*, 151.
(26) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(27) Kong, J.; White, C. A.; Krylov, A. I.; Sherrill, D.; Adamson, R. D.; Furlani, T. R.; Lee, M. S.; Lee, A. M.; Gwaltney, S. R.; Adams, T. R.; Ochsenfeld, C.; Gilbert, A. T. B.; Kedziora, G. S.; Rassolov, V. A.; Maurice, D. R.; Nair, N.; Shao, Y. H.; Besley, N. A.; Maslen, P. E.; Dombroski, J. P.; Daschel, H.; Zhang, W. M.; Korambath, P. P.; Baker, J.; Byrd, E. F. C.; Van Voorhis, T.; Oumi, M.; Hirata, S.; Hsu, C. P.; Ishikawa, N.; Florian, J.; Warshel, A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M.; Pople, J. A. *J. Comput. Chem.* **2000**, *21*, 1532.
(28) White, C. A.; Kong, J.; Maurice, D. R.; Adams, T. R.; Baker, J.; M. Challacombe; Schwegler, E.; Dombroski, J. P.; Ochsenfeld, C.; Oumi, M.; Furlani, T. R.; Florian, J.; Adamson, R. D.; Nair, N.; Lee, A. M.; Ishikawa, N.; Graham, R. L.; Warshel, A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon., M. *Q-Chem*, version 1.2; Pittsburgh, PA, 1998.
(29) Gill, P. M. W.; Johnson, B. G.; Pople, J. A. *Chem. Phys. Lett.* **1993**, *209*, 506.
(30) Murray, C. W.; Handy, N. C.; Laming, G. J. *Mol. Phys.* **1993**, *78*, 997.
(31) Lebedev, V. I. *Zh. Vychisl. Mater. Mater. Fiz.* **1975**, *15*, 48.
(32) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577.
(33) Hubbard, S. J. *Naccess*, Version 2.1.1; Manchester, U.K., 1996.
(34) Richards, F. M.; Lee, B. *J. Mol. Biol.* **1971**, *55*, 379.
(35) Schweitzer-Stenner, R.; Eker, F.; Q., H.; Griebenow, K.; Mroz, P. A.; Kozlowski, P. M. *J. Phys. Chem. B* **2002**, *106*, 4294.
(36) Fogarasi, G.; Zhou, X.; Taylor, P. W.; Pulay, P. *J. Am. Chem. Soc.* **1992**, *114*, 8191.
(37) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(38) Han, W. G.; Suhai, S. *J. Phys. Chem.* **1996**, *100*, 3942.
(39) Woutersen, S.; Hamm, P. *J. Phys. Condens. Mater.* **2002**, *14*, R1035.
(40) Torii, H.; Tasumi, M. *J. Raman Spectrosc.* **1998**, *29*, 81.
(41) Bour, P.; Kubelka, J.; Keiderling, T. A. *Biopolymers* **2002**, *65*, 45.
(42) Ataka, S.; Takeuchi, H.; Tasumi, M. *J. Mol. Struct.* **1984**, *113*, 147.
(43) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33.