Use of Periodic Boundary Conditions to Calculate Accurate 
β-Sheet Frequencies Using Density Functional Theory

Stefan Franzen  
Department of Chemistry  
North Carolina State University  
Raleigh, NC 27695

Abstract  
Calculations of vibrational spectra of peptides that represent the major structural motifs, α-helix, β-sheet and extended conformations were carried out using density functional theory (DFT). Although the calculations agree qualitatively with experiment, the frequencies are not in the correct range due to the lack of inclusion of intermolecular interactions in the calculated model. One solution to this problem for the parallel β-sheet structure is demonstrated using periodic boundary conditions (PBC). A model consisting of four glycines with a pleated parallel β-sheet structure in a box of appropriate dimensions was calculated using DFT methods to obtain accurate frequencies of the amide bands. This model is compared and it is shown that intramolecular hydrogen bonding can be included to quantitatively account for the amide I and amide II spectrum of the β-sheet.

Introduction  
The two most common secondary structure motifs in proteins are the α-helix and the β-sheet. A large number of computational and experimental studies have been carried out on α-helical peptides, 24,5,10,27 but studies of β-sheet peptide models are rarer. 19,20,13,16,23 On the experimental side, β-sheets are often insoluble in water because they tend to form aggregates. The β-sheet structure inherently involves interstrand hydrogen bonding, while an α-helical peptide or domain has intrastrand hydrogen bonds. Consequently, computation models of α-helices are more tractable. Since hydrogen bonding results in frequency shifts, 33,32,27,21 an ab initio calculation of frequencies is likely to be closer to experimental values for an α-helix than for a β-sheet.

The power of vibrational spectroscopy as a tool in biology has not been realized. Often the interpretation of vibrational spectra is a qualitative association of bands and particular structures or chemical groups. Nuclear magnetic resonance has the advantage that nuclear spin is associated with one peak or multiplet. In contrast, the observed bands in vibrational spectra are the result of the coupled motion of all of the nuclei in the sample. For N nuclei there are at most 3N – 6 observed fundamental bands and yet the matrix of internuclear force interactions (the second derivative matrix in the harmonic approximation) has (3N – 6)(3N – 5)/2 unique terms. The mathematical problem of obtaining the force constants from vibrational frequencies is underdetermined. 36 Even isotope labeling is of limited utility in making a complete assignment. Although each isotope provides a constraint, one would need (3N – 5)/2 isotopomers to make a complete assignment. If one adds to this the problem that many bands are not observed, for example there are overtones and combination bands and deviations from the harmonic approximation, the problem seems intractable. The advent of efficient codes for the
computation of the ground state potential energy surface provides a potential solution to the problem. The \textit{ab initio} computation of the force constants provides a method for developing vibrational spectroscopy into a method of general utility. For small isolated molecules, these approaches are nearing the point of complete assignment of experimental bands including deviations from the harmonic approximation. However, for biopolymers there are a number of issues that remain.

In the application of vibrational spectroscopy to biology a second level of difficulty is encountered. Biological molecules maintain their structure by virtue of strong intermolecular interactions, such as hydrogen bonding and the hydrophobic effect. Moreover, proteins and nucleic acids are biopolymers with strong interactions of similar groups in a chain. Carbonyl groups and amino groups at identical frequencies in the monomer interact by excitonic coupling in biopolymers. The interpretation of vibrational spectra in such systems requires new methods in order to extract essential features while keeping the size of the calculation at a reasonable level.

The interpretation of peptide and protein vibrational spectra requires a method that will accurately determine the bonding, as well as intermolecular interactions. Density functional theory (DFT) methods are well suited to the calculation of ground state properties and potential energies. \textsuperscript{10,31,24,16,22,23} For this reason, DFT is an excellent method for calculating vibrational spectra from first principles. While DFT models appear to represent the bonding quite well and permit comparison with experimental trends, the absolute values of the frequencies are high compared to experiment. There are two factors that can account for this discrepancy. Anharmonicity and hydrogen bonding both result in a lower frequency for a particular normal mode than predicted by the harmonic approximation. While hydrogen bonding can be accounted for using bound water, the structures in proteins involve interactions between peptide units. In the present study, we are concerned with the calculation of accurate frequencies when hydrogen bonding between peptide units is taken into account.

In this study polyglycine peptide models were used to determine the vibrational spectrum of amide bands (amide A, I, II, and III). Amide A consists mainly of amide N-H stretching. Amide I consists of a major contribution to the potential energy distribution from C=O stretching and a minor contribution from in-plane C-N-H bending. Amides II and III involve in-plane deformation of the C-N-H coupled to other in-plane bending motions. Polyglycine (Gly)$_n$ models were chosen first because they have no side chains and are the simplest case. Geometry optimization and frequency calculation were carried out for \(\alpha\)-helix, \(\beta\)-sheet, \(\beta\)-turn and extended conformations using DFT. The first class of models consists of a gas phase calculation of a small peptide that has a defined secondary structure. The quantitative agreement with spectra in these models is poor for two reasons: 1. dielectric effects and 2. end-of-chain effects. The model calculations are gas phase calculations, hence \(\varepsilon = 1\). Perhaps more serious, the hydrogen bonding of the models is not correct. For example, the end glycines have incorrect (non \(\alpha\)-helical) hydrogen bond patterns to the next two residues in the calculation of nonaglycine, (Gly)$_9$. Out of nine glycine residues in the model, only three are truly \(\alpha\)-helical in their hydrogen bond pattern. In spite of these deficiencies, the relative frequencies and intensities of various amide bands show reasonable agreement with experimental data. It is logical to assume that the origin of the deviation from experimental data is the lack of intermolecular interactions. Interactions with water have been discussed in other studies.
In this study we show that the interchain interactions in β-sheet peptides can be quantitatively calculated for a relatively small model using periodic boundary conditions. This approach validates similar studies that use multiple parallel and anti-parallel β-sheets to model vibrational spectra.

**Methods**

DMol3 was used to perform DFT calculations for a number of model peptides shown in Figure 1. The β-sheet peptides are compared with a model that uses periodic boundary conditions (PBC) for the parallel β-sheet shown in Figure 2. The optimized ground state geometries were obtained using the generalized gradient approximation (GGA) of Perdew and Wang \(^{30}\) as implemented in DMol3 (Accelrys Inc.). \(^{6,7}\) A numerical basis set was used that corresponds to a double-\(\zeta\) basis set. The model consists of a chain of four glycine amino acids in a pleated β-sheet structure, i.e. the calculation consisted of 28 atoms per unit cell. The dimensions of the rectangular box used for the periodic boundary conditions were 13.35 x 4.0 x 5.0 Å. The C=O…H-N hydrogen bonding distance is 1.94 Å. The geometry of all molecules was optimized with a convergence criterion of <10\(^{-6}\) a.u. change in the energy and a change of less than 0.003 Bohrs per iteration. The geometry optimization of the gas phase β-sheet polyglycine structures were carried out with fixed positions for the amide nitrogen atoms. If this is not done the optimized structure is an extended structure. Vibrational frequencies were calculated using finite difference methods as described previously. \(^{12,11}\)

There is no method at present for directly determining the infrared intensities for vibrational normal modes determined using PBC methods. Infrared intensities are not available since neither the dipole nor the dipole derivatives can be calculated using PBC methods. Given the importance of intensities for comparison with experimental spectra, a method was implemented that uses the Mulliken charge to estimate the change in the dipole moment. Using the normal mode projections obtained from calculation of the hessian matrix, this method permits calculation of the dipole derivative required for estimation of the infrared intensities. The eigenvectors obtained from the normal mode calculation were used to displace the geometry of the molecule. The difference dipole moment (\(\partial\mu/\partial Q\)) required for calculation of the infrared transition moment was calculated from the dipole moment in each geometry.

**Results**

Calculation of the polyglycine models gives a reasonable qualitative agreement with amide band spectra. The identity, relative intensity and general location of the amide bands A, I, II and III are given in Figure 3. The spectra were generated from a list of frequencies and intensities using a Gaussian broadening function (Eqn. 1),

\[
I(\nu) = \sum_k^N \frac{A_k}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(\nu - \nu_0)^2}{2\sigma^2} \right\}
\]

for each of the N vibrational modes calculated. The intensity of each band is A\(_k\) in km/mol. The frequencies listed in Table 1 are higher than the experimental values. For example, amide I is calculated between 1671 – 1701 cm\(^{-1}\) for the model systems. The carbonyl group experiences a large frequency decrease for hydrogen bonding. Experimental values in peptides and proteins range from 1610 – 1680 cm\(^{-1}\).
In each of the models there are dangling C=O groups that have no hydrogen bond partner. Addition of H$_2$O molecules will lower the frequency as discussed in a model calculation of N-methyl acetamide in the Supporting Information and other studies. However, hydrogen bonding cannot account for inter-residue interactions inside a protein. In proteins the dominant interaction is C=O…H-N. The fact that this interaction is present in each of the models to a different extent results in a failure of these models to accurately mimic the experimental frequency ordering. In order to address this problem a comparison was made of the tetraglycine β-sheet model in vacuum (Table 1) and in a box of appropriate dimensions with periodic boundary conditions (PBC).

Figure 4 shows a comparison of the calculated infrared spectra for a β-sheet using DFT in vacuum and with PBC. Examination of the frequencies in Table 2 reveals that Amide A, B, I and II are in the correct regions and are within 5% of the correct value for the calculation that uses PBC. The spectrum derived from the DFT calculation using PBC is shown as the solid line in Figure 2. There is a large difference due to the contribution of C=O…H-N hydrogen bonding in the PBC model (Figure 2). Not only are the frequencies of the amide I and II bands in the PBC calculation in agreement with experimental results, but the shape of the amide I band agrees with experimental observations on a wide range of β-sheet peptides. The largest intensity band is calculated at 1633 cm$^{-1}$ and there is a pronounced shoulder at 1681 cm$^{-1}$. This intensity pattern is observed in β-sheet peptides, β-sheet proteins and as a spectral component in proteins that have β-sheet secondary structures. The calculated amide I bands for the extended and β-sheet tetraglycine structures are quite similar (Table 2).

Discussion

Infrared spectra of peptides and proteins consist mainly of the amide band transitions. The most prominent transition is the intense amide I vibration, which is mainly a C=O stretching vibration. The maximum frequency and shape of the amide I band is characteristic for both α-helices and β-sheets. As a result decomposition of the amide I line shape from a protein infrared spectrum can be used to obtain an estimate of the secondary structure of an unknown protein. However, a general understanding of the line shape in terms of the oscillators in a protein macromolecule requires inclusion of the vibrational frequency of each individual oscillator and the coupling between all of the oscillators in the molecule. Although the theory for coupling has been available for some time, there is currently no method for ab initio calculation of the vibrational spectrum of a protein.

Density functional theory (DFT) methods have shown great promise for the calculation of vibrational spectra. DFT is an excellent method for the determination of the ground state properties of a molecule including its potential energy surface. The accuracy of normal mode calculations using DFT methods is sufficiently high that the limitations of the harmonic approximation are often a major source of
disagreement between theory and experiment. Given the success of the method for small molecules, it is logical to extend the strategy to include molecular interactions. For instance, solvation studies can be carried out in explicit solvent and this has already begun in studies that include explicit water. The amide I frequency is a case in point. The potential energy distribution (PED) of the amide I vibration contains a significant contribution from the C=O stretch. The frequency of the C=O stretch is significantly lowered by hydrogen bonding to either an amide N-H group or to water. Explicit water has been used to account for these interactions. Such calculations are much improved in accuracy. This type of approach is excellent for the determination of the vibrational spectra of small peptides in solution. However, the interactions of C=O groups in proteins principally involve N-H protons on neighboring polypeptide chains. Until the present, time these interactions have not been easily modeled using ab initio computational methods. The use of PBC provides an avenue to accurately estimating the vibrational frequency from particular structures. The cost of these calculations is quite low compared to vibrational frequency calculations on a macromolecule.

The PBC method can be generalized to the calculation of spectral features for other elements of secondary structure, such as α-helices, β-turns and extended structures. While these calculations are perhaps more closely related to a crystal than a protein, they represent one step closer to direct determination of the factors that govern the infrared line shape of the amide I band. Ultimately, such an approach is needed in order to employ infrared spectroscopy as a quantitative method for the determination of secondary structure. The approach can be generalized to the study of type-I and type-II β-turns, α-helices and extended conformations. These studies will provide a valuable resource in the use of Fourier-transform infrared spectroscopy as a method for the secondary structure determination of proteins.

**Conclusion**

The calculation of vibrational frequencies for the vibrations of the polyamide backbone of peptides and proteins requires inclusion of solvent and hydrogen bonding patterns of the known structural motifs; α-helices, β-sheets and β-turns and extended conformations. Given the current size limitations on calculations using density functional theory (DFT) it is desirable to find methods to incorporate interchain hydrogen bonding in order to study how vibrations are affected. In the present study, the use of periodic boundary conditions has been shown to successfully model the spectrum of the parallel β-sheet using DFT. This approach suggests a general method that can be applied to other small elements of secondary structure to permit a detailed understanding of the origin of the spectral signatures.

**Acknowledgements.** SF acknowledges support through NSF grant MCB-9874895 and a grant and technical support from the North Carolina Supercomputer Center (NCSC).

**Supporting Information.** Model calculations of N-methyl acetamide and spectra are presented. Eigenvectors showing the normal modes for amide I, II, III and A are given using N-methyl acetamide as the model.
References
Table 1. Calculated frequencies for infrared transitions of the amide vibrations of model peptides. All residues in the calculations are glycine. The value reported is the center of a band calculated using a Gaussian function (see text) to broaden the spectra with $\sigma = 12 \text{ cm}^{-1}$ as shown in Figure 3.

<table>
<thead>
<tr>
<th>Secondary Structure</th>
<th>Number of glycines</th>
<th>Amide A (cm$^{-1}$)</th>
<th>Amide I (cm$^{-1}$)</th>
<th>Amide II (cm$^{-1}$)</th>
<th>Amide III (cm$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$-helix</td>
<td>6</td>
<td>3536/3413</td>
<td>1701</td>
<td>1516</td>
<td>1220</td>
</tr>
<tr>
<td>$\alpha$-helix</td>
<td>9</td>
<td>3490</td>
<td>1691</td>
<td>1515</td>
<td>1239</td>
</tr>
<tr>
<td>$\beta$-sheet</td>
<td>6</td>
<td>3419/3239</td>
<td>1651</td>
<td>1501</td>
<td>1233</td>
</tr>
<tr>
<td>$\beta$-sheet and turn</td>
<td>9</td>
<td>3426/3343/3282</td>
<td>1668</td>
<td>1502</td>
<td>1248</td>
</tr>
<tr>
<td>extended</td>
<td>2</td>
<td>3467</td>
<td>1675</td>
<td>1465</td>
<td>1228</td>
</tr>
<tr>
<td>extended</td>
<td>3</td>
<td>3464</td>
<td>1679</td>
<td>1472</td>
<td>1224</td>
</tr>
<tr>
<td>extended</td>
<td>4</td>
<td>3466</td>
<td>1687</td>
<td>1478</td>
<td>1245</td>
</tr>
<tr>
<td>extended</td>
<td>5</td>
<td>3454</td>
<td>1688</td>
<td>1475</td>
<td>1249</td>
</tr>
<tr>
<td>extended</td>
<td>6</td>
<td>3435</td>
<td>1671</td>
<td>1466</td>
<td>1234</td>
</tr>
<tr>
<td>extended</td>
<td>9</td>
<td>3435</td>
<td>1671</td>
<td>1463</td>
<td>1214</td>
</tr>
</tbody>
</table>

Table 2. Comparison of the frequency and intensity of characteristic polypeptide modes in PBC and non-PBC DFT calculations. Individual vibrational bands and infrared intensities are reported from the DMol3 output file.

<table>
<thead>
<tr>
<th>PBC $\beta$-sheet</th>
<th>Amide A (cm$^{-1}$)</th>
<th>Intensity (km/mol)</th>
<th>Amide A (cm$^{-1}$)</th>
<th>Intensity (km/mol)</th>
<th>1537</th>
<th>147.5</th>
<th>1502</th>
<th>434.2</th>
<th>1485</th>
<th>643.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$-sheet</td>
<td>Amide II (cm$^{-1}$)</td>
<td>Intensity (km/mol)</td>
<td>Amide II (cm$^{-1}$)</td>
<td>Intensity (km/mol)</td>
<td>1539</td>
<td>374.5</td>
<td>1533</td>
<td>148.3</td>
<td>1487</td>
<td>128.3</td>
</tr>
<tr>
<td></td>
<td>15.0</td>
<td>415.0</td>
<td>1500</td>
<td>308.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1546</td>
<td>7.08</td>
<td>26.2</td>
<td>1543</td>
<td>118.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amide I</td>
<td>Amide I</td>
<td>Amide I</td>
<td>1533</td>
<td>1691</td>
<td>1653</td>
<td>1686.0</td>
<td>1699</td>
<td>640.4</td>
<td>1691</td>
<td>551.0</td>
</tr>
<tr>
<td>1656</td>
<td>431.4</td>
<td>1709</td>
<td>1701</td>
<td>9.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1661</td>
<td>2.77</td>
<td>1721</td>
<td>1712</td>
<td>178.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1681</td>
<td>248.8</td>
<td>1737</td>
<td>1720</td>
<td>1720</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-H</td>
<td>C-H</td>
<td>C-H</td>
<td>3003</td>
<td>2947</td>
<td>29.2</td>
<td>2953</td>
<td>24.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3033</td>
<td>2993</td>
<td>2991</td>
<td>26.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3035</td>
<td>1.30</td>
<td>2995</td>
<td>3011</td>
<td>9.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3042</td>
<td>0.258</td>
<td>3001</td>
<td>3030</td>
<td>14.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amide B</td>
<td>Amide B</td>
<td>Amide B</td>
<td>3012</td>
<td>3083</td>
<td>6.4</td>
<td>3020</td>
<td>12.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3101</td>
<td>0.529</td>
<td>3083</td>
<td>3020</td>
<td>12.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3102</td>
<td>2.05</td>
<td>3090</td>
<td>3077</td>
<td>9.96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3106</td>
<td>0.294</td>
<td>3098</td>
<td>3113</td>
<td>0.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3127</td>
<td>0.0497</td>
<td>3114</td>
<td>3128</td>
<td>5.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amide A</td>
<td>Amide A</td>
<td>Amide A</td>
<td>3325</td>
<td>3508</td>
<td>59.0</td>
<td>3431</td>
<td>174.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3347</td>
<td>67.1</td>
<td>3525</td>
<td>3452</td>
<td>152.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3375</td>
<td>66.4</td>
<td>3559</td>
<td>3468</td>
<td>109.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3390</td>
<td>52.0</td>
<td>3596</td>
<td>3573</td>
<td>29.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1. Nonaglycine models used in the calculations described in the text.

Figure 2. Model $\beta$-sheet structure used for vibrational frequency calculation using periodic boundary conditions. The C=O-H-N hydrogen bond distance is 1.94 Å.
Figure 3. Calculated infrared spectra of nonaglycine models of an α-helix, a β-sheet and turn and an extended conformation shown in Figure 1. The spectral regions of amides I, II and III are shown. The spectra were calculated using a Gaussian model to give width to the infrared frequencies ($\sigma = 12 \text{ cm}^{-1}$) and intensities obtained from a DFT calculation the of vibrational force constants followed by diagonalization of hessian matrix in mass-weighted Cartesian coordinates.
Figure 4. Calculated infrared spectra of tetraglycine in the region of amide I and amide II. The calculated spectra correspond to a tetraglycine in vacuum (****) and a tetraglycine calculated using periodic boundary conditions to create an infinite chain (——). The spectra were calculated using a Gaussian model to give width to the infrared frequencies ($\sigma = 12$ cm$^{-1}$) and intensities obtained from a DFT calculation of the vibrational force constants followed by diagonalization of hessian matrix in mass-weighted Cartesian coordinates.