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, North Carolina 27695

Received: May 5, 2003; In Final Form: August 26, 2003

Calculations of vibrational spectra of peptides that represent the major structural motifs, α-helix, β-sheet, and extended conformations, carried out using density functional theory (DFT) agree only qualitatively with experiment because of the lack of inclusion of intermolecular interactions in the calculated model. One solution to this problem for the parallel β-sheet structure is demonstrated in this study 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 to gas-phase calculation of β-sheet and extended conformations, 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,[1−4] but studies of β-sheet peptide models are rarer.[5−9] 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, whereas an α-helical peptide or domain has intrastrand hydrogen bonds. Because hydrogen bonding results in frequency shifts,[4,10−12] and α-helices have internal N−H···C═O hydrogen bonding the ab initio calculation of frequencies is likely to be closer to experimental values for an α-helix than for a β-sheet, which has uncompensated C═O and N−H groups. The inclusion of water does not solve this problem because water is not the physiologically relevant hydrogen bond partner in the interior of a protein. The hydrogen bonding interactions of β-sheets inside a protein are dominated by interstrand C═O···H−N interactions.

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.[13] 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 ab initio computation of the force constants provides a method for developing vibrational spectroscopy into a method of general utility.[14] 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 to extract essential features while the size of the calculation is kept 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.[3,7,15−21] For this reason, DFT is an excellent method for calculating vibrational spectra from first principles. Although 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 interrelated 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. Though 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

* Corresponding author. Phone: (919)-515-8915. E-mail: Stefan_Franzen@ncsu.edu.

10.1021/jp035215k CCC: $25.00
Published on Web 10/25/2003

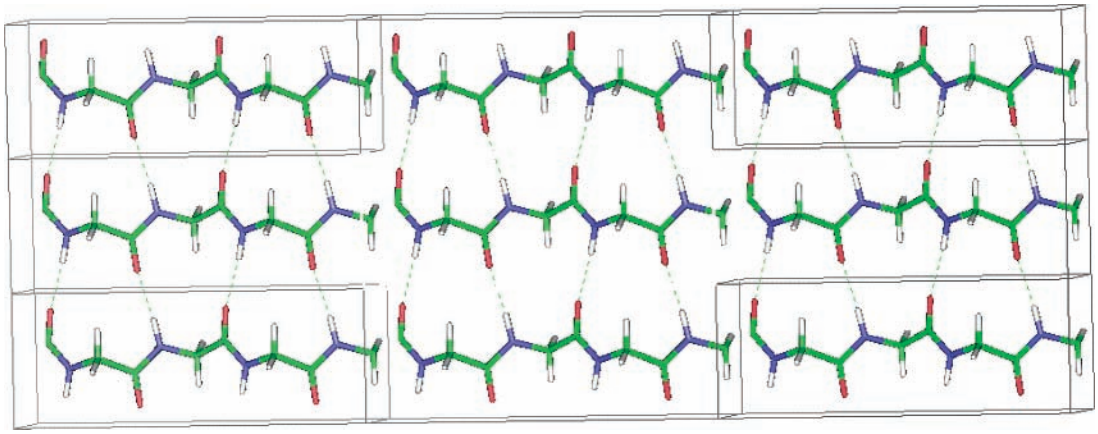

Figure 1. Model β-sheet structure used for vibrational frequency calculation using periodic boundary conditions. The C═O···H—N hydrogen bond distance is 1.94 Å.

frequencies when hydrogen bonding between peptide units is taken into account.

Polyglycine peptide models can be used to determine the vibrational spectrum of amide bands (amide A, I, II, and III) for β-sheets. Because glycine is a helix breaker and alanine is a helix former, polyalanines are well-suited to studies of α-helical structure and vibrational spectra.[1,8] The amide bands are important markers of conformation and show pronounced differences between α-helix and β-sheet structures. 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 have no side chains and are the simplest case for the study of β-sheets. However, the frequencies calculated for such models are always significantly higher than the experimental values because hydrogen bonding to solvent molecules is not included. The calculation is a model of the spectrum of crystalline polyglycine but has relevance for β-sheets in proteins because of the similarity of their spectra.[22]

Methods

DMol3 was used to perform DFT calculations for a number of model peptides to validate the method in comparison with previous work (see Supporting Information). The β-sheet peptides are compared with a model that uses periodic boundary conditions (PBC) for the parallel β-sheet shown in Figure 1. The optimized ground-state geometries were obtained using the generalized gradient approximation (GGA) of Perdew and Wang[23] as implemented in DMol3 (Accelerys Inc.).[24] A numerical basis set was used that corresponds to a double-ζ 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 × 4.0 × 5.0 Å. The C═O···H—N hydrogen bonding distance is 1.94 Å. The three-dimensional PBC model was compared to a gas-phase calculation of a tetraglycine − sheet that was capped by $-NH-CH_3$ and $-C{=}O-CH_3$ on the N- and C-termini, respectively. The geometry of all molecules was optimized with a convergence criterion of $<10^{-6}$ au 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.[25,26]

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 because 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

The calculated frequencies of polyglycine models presented in the Supporting Information agree well with calculations on gas-phase peptide models studied previously.[3,7,9,15−18] Moreover, the polyglycine amide frequencies are in reasonable qualtitative agreement with amide band spectra, as shown by the identity, relative intensity, and general location of the amide bands A, I, II, and III (see Supporting Information). The spectra were generated from a list of frequencies and intensities using a Gaussian broadening function

$$I(n) = \sum_{k}^{N} \frac{A_k}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(\nu - \nu_0)^2}{2\sigma^2} \right\} \qquad (1)$$

for each of the N vibrational modes calculated. The intensity of each band is A_k in km/mol. In agreement with previous work on unsolvated peptide models, the calculated frequencies are higher than the experimental values (see Supporting Information). For example, amide I is calculated between 1671 and 1701 cm^{-1} for the model systems compared to experimental values in peptides and proteins range from 1610 to 1680 cm^{-1}.[5,9,27−30] In each of the peptide models there are dangling C═O groups that have no hydrogen bonding partner. Addition of H_2O molecules will lower the frequency, as discussed in a model calculation of *N*-methylacetamide in the Supporting Information and other studies.[7,22,31,32] However, hydrogen bonding cannot

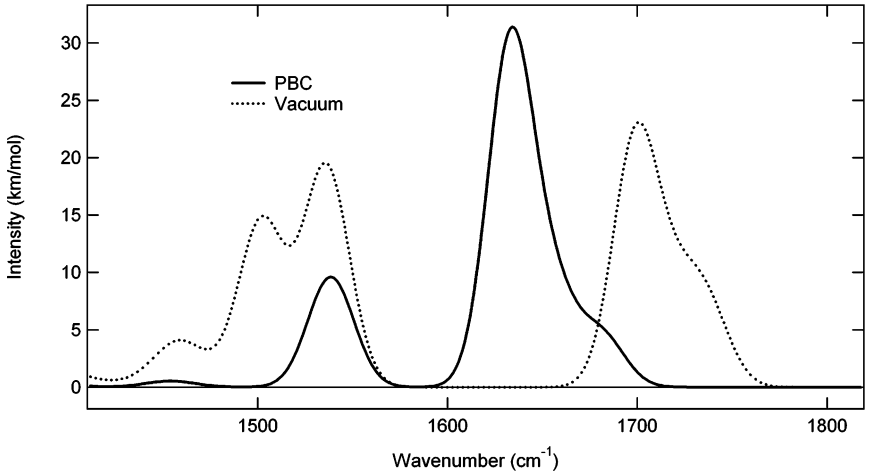

Figure 2. Calculated infrared spectra of tetraglycine in the region of amide I and amide II. The calculated spectra correspond to a tetraglycine in a vacuum (···) and a tetraglycine calculated using periodic boundary conditions to create an infinite chain (—). The spectra were calculated using a Gaussian 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.

TABLE 1: Comparison of the Frequency and Intensity of Characteristic Polypeptide Modes in PBC and Non-PBC DFT Calculations[a]

PBC β-sheet		β-sheet		extended	
amide II (cm^{-1})	intensity (km/mol)	amide II (cm^{-1})	intensity (km/mol)	amide II (cm^{-1})	intensity (km/mol)
1537	147.5	1502	434.2	1485	643.6
1539	374.5	1533	148.3	1487	128.3
1545	15.0	1537	415.0	1500	308.8
1546	7.08	1541	26.2	1543	118.8
		Amide I			
1633	1686.0	1699	640.4	1691	551.0
1656	431.4	1709	19.1	1701	9.37
1661	2.77	1721	175.8	1712	178.7
1681	248.8	1737	192.8	1720	192.4
C−H		C−H		C−H	
3030	0.112	2947	29.2	2953	24.0
3033	0.418	2993	50.3	2991	26.4
3036	1.30	2995	18.5	3011	9.7
3042	0.258	3001	10.7	3030	14.3
		Amide B			
3101	0.529	3083	6.4	3020	12.5
3102	2.05	3090	11.7	3077	9.96
3106	0.294	3098	0.8	3113	0.42
3127	0.0497	3114	4.1	3128	5.54
		Amide A			
3325	86.2	3508	59.0	3431	174.8
3347	67.1	3525	63.6	3452	152.0
3375	66.4	3559	61.7	3468	109.7
3390	52.0	3596	28.4	3573	29.4

[a] Individual vibrational bands and infrared intensities are reported from the DMol3 output file.

account for interresidue 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. To address this problem, a comparison was made of the tetraglycine β-sheet model in a vacuum and in a box of appropriate dimensions with periodic boundary conditions (PBC).

Figure 2 shows a comparison of the calculated infrared spectra for a β-sheet using DFT in a vacuum and with PBC. Examination of the frequencies in Table 1 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. 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.[7−9,22,33−43]

The calculated spectra for β-sheets and extended structures the gas phase differ primarily in the amide II region. Extended structures have a large intensity amide II band that is near 1485 cm^{-1} (see Supporting Information). The calculated amide II frequency in β-sheet tetraglycine shown in Figure 2 occurs at 1502 cm^{-1} and is significantly lower in intensity than for the

extended conformation. The calculated amide I bands for the extended and β-sheet tetraglycine structures in the gas phase are quite similar (Table 1).

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.[27,33,34,36−40,43−48] 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 transition dipole coupling between the oscillators in the molecule, e.g., the carbonyl stretching oscillations responsible for amide I. Although the theory for coupling has been available for some time,[49] 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.[3,9,15−19,25,26,32] 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.[16] 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.[32] 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.

As a test of frequency calculation for short peptides in the gas phase, geometry optimization and frequency calculation were carried out for α-helix, β-sheet, β-turn or hairpin, and extended conformations using DFT. The model calculations agree with previous ab initio calculations,[3,9,12,32] but the quantitative agreement with experiment 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 the dielectric constant $\epsilon = 1$ and the dielectric effect of chain packing are not included. Perhaps more serious, the hydrogen bonding of the models is not correct. For example, in a nonaglycine α-helix, $(Gly)_9$, the terminal glycines have incorrect (non-α-helical) hydrogen bond patterns to the next two residues in the calculation of nonaglycine. Out of nine glycine residues in the model, only three are truly α-helical in their hydrogen bond pattern. Despite these deficiencies, the *relative* frequencies and intensities of various amide bands show reasonable agreement with experimental trends; e.g., the frequencies of β-sheets are lower than those of α-helices. It is logical to assume that the origin of the systematically higher calculated frequencies relative to experimental data arises from the lack of intermolecular interactions in the computational models. Interactions with water have been discussed in other studies.[7,31,32] 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 antiparallel β-sheets to model vibrational spectra. 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. Although these calculations are 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 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 hybrogen bonding 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.

Acknowledgment. I acknowledge support through NSF grant MCB-9874895 and a grant and technical support from the North Carolina Supercomputer Center (NCSC).

Supporting Information Available: Model calculations and spectra of *N*-methylacetamide and polyglycine model peptides are presented. Eigenvectors showing the normal modes for amide I, II, III, and A are given using *N*-methylacetamide as the model. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Kubelka, J.; Silva, R. A. G. D.; Keiderling, T. A. *J. Am. Chem. Soc.* **2002**, *124*, 5325−5332.
(2) Decatur, S. M. *Biopolymers* **2000**, *54*, 180−185.
(3) Elstner, M.; Jalkanen, K. J.; Knapp-Mohammady, M.; Frauenheim, T.; Suhai, S. *Chem. Phys.* **2000**, *256*, 15−27.
(4) Manas, E. S.; Getahun, Z.; Wright, W. W.; DeGrado, W. F.; Vanderkooi, J. M. *J. Am. Chem. Soc.* **2000**, *122*, 9883−9890.
(5) Krimm, S.; Abe, Y. *Proc. Natl. Acad. Sci.* **1972**, *69*, 2788−2792.
(6) Krimm, S.; Bandekar, J. *Biopolymers* **1980**, *19*, 1−29.
(7) Gresh, N.; Guo, H.; Salahub, D. R.; Roques, B. P.; Kafafi, S. A. *J. Am. Chem. Soc.* **1999**, *121*, 7885−7894.
(8) Hilario, J.; Kubelka, J.; Syud, F. A.; Gellman, S. H.; Keiderling, T. A. *Biopolymers* **2002**, *67*, 233−236.
(9) Kubelka, J.; Keiderling, T. A. *J. Am. Chem. Soc.* **2001**, *123*, 6142−6150.
(10) Thijs, R.; Zeeger-Huyskens, T. *Spectrochim. Acta* **1984**, *40A*, 307−313.
(11) Susi, H. *Methods Enzymol.* **1986**, *26*, 381−391.

(12) Silva, R. A. G. D.; Kubelka, J.; Bour, P.; Decatur, S. M.; Keiderling, T. A. *Proc. Natl. Acad. Sci.* **2000**, *97*, 8318−8323.
(13) Wilson, E. B., Jr.; Decius, J. C.; Cross, P. C. *Molecular Vibrations*; Dover Publications: New York, 1955.
(14) Qian, W.; Bandekar, J.; Krimm, S. *Biopolymers* **1991**, *31*, 193−210.
(15) Jishi, R. A.; Flores, R. M.; Valderrama, M.; Lou, L.; Bragin, J. *J. Phys. Chem. A* **1998**, *102*, 9858−9862.
(16) Schweitzer-Stenner, R.; Eker, F.; Huang, Q.; Griebenow, K.; Mroz, P. A.; Kozlowski, P. M. *J. Phys. Chem. B* **2002**, *106*, 4294−4304.
(17) Jalkanen, K. J.; Suhai, S. *Chem. Phys.* **1996**, *208*, 81−116.
(18) Jacob, R.; Fischer, G. *J. Mol. Struct.* **2002**, *613*, 175−188.
(19) Cao, X. L.; Fischer, G. *J. Phys. Chem. A* **2002**, *106*, 41−50.
(20) Perczel, A.; Farkas, O.; Csaszar, A. G.; Csizmadia, I. G. *Can. J. Chem.* **1997**, *75*, 1120−1130.
(21) Wang, B.; Hinton, J. F.; Pulay, P. *J. Comput. Chem.* **2002**, *23*, 492−497.
(22) Cheam, T. C.; Krimm, S. *J. Chem. Phys.* **1985**, *82*, 1631−1641.
(23) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1992**, *46*, 6671−6687.
(24) Delley, B. *J. Chem. Phys.* **2000**, *113*, 7756−7764.
(25) Franzen, S. *J. Am. Chem. Soc.* **2001**, *123*, 12578−12589.
(26) Franzen, S. *J. Am. Chem. Soc.* **2002**, *124*, 13271−13281.
(27) Douseeau, F.; Pezolet, M. *Biochemistry* **1990**, *29*, 8771−8779.
(28) Hering, J. A.; Innocent, P. R.; Haris, P. I. *Proteomics* **2002**, *2*, 839−849.
(29) Jakobsen, R. J.; Wasacz, F. M.; Brasch, J. W.; Smith, K. B. *Biopolymers* **1986**, *25*, 639−654.
(30) Jongh, H.; Goormaghtigh, E.; Ruysschaert, J. *Anal. Biochem.* **1996**, *242*, 95−103.
(31) Kearley, G. J.; Johnson, M. R.; Plazanet, M.; Suard, E. *J. Chem. Phys.* **2001**, *115*, 2614−2620.
(32) Kubelka, J.; Keiderling, T. A. *J. Phys. Chem. A* **2001**, *105*, 10922−10928.
(33) Vedantham, G.; Sparks, H. G.; Sane, S. U.; Tzannis, S.; Przybycien, T. M. *Anal. Biochem.* **2000**, *285*, 33−49.
(34) Zhang, C. T.; Zhang, R. *J. Biomol. Struct. Dynam.* **2000**, *17*, 829−842.
(35) Pribic, R. *Anal. Biochem.* **1994**, *33*, 26−34.
(36) Wi, S.; Pancoska, P.; Keiderling, T. A. *Biospectroscopy* **1998**, *4*, 93−106.
(37) Kumosinski, T.; Unruh, J. *Global-Secondary-Structure Analysis of Proteins in Solution*; Kumsinski, T., Liebmann, M., Eds.; ACS Symposium Series 576; American Chemical Society: Washington, D.C., 1994; pp 71−98.
(38) Levitt, M. *J. Mol. Biol.* **1977**, *114*, 181−239.
(39) Baumruk, V.; Pancoska, P.; Keiderling, T. *J. Mol. Biol.* **1996**, *259*, 774−791.
(40) Baello, B.; Pancoska, P.; Keiderling, T. *Anal. Biochem.* **2000**, *280*, 46−57.
(41) Smith, B. M.; Oswald, L.; Franzen, S. *Anal. Chem.* **2002**, *74*, 3386−3391.
(42) Powell, J.; Wasacz, F.; Jakobsen, R. *Appl. Spectrosc.* **1986**, *40*, 339−344.
(43) Pelton, J. T.; McLean, L. R. *Analy. Biochem.* **2000**, *277*, 167−176.
(44) Byler, D.; Susi, H. *Biopolymers* **1986**, *25*, 469−487.
(45) Pancoska, P.; Bitto, E.; Janota, V.; Urbanova, M.; Gupta, V.; Keiderling, T. *Protein Sci.* **1995**, *4*, 1384−1401.
(46) Baello, B.; Pancoska, P.; Keiderling, T. *Anal. Biochem.* **1997**, *250*, 212−221.
(47) Susi, H.; Byler, D.; Purcell, J. *J. Biochem. Biophys. Methods* **1985**, *11*, 235−240.
(48) Dong, A.; Huang, P.; Caughey, W. S. *Biochemistry* **1990**, *29*, 3303−3308.
(49) Krimm, S. *J. Mol. Biol.* **1962**, *4*, 528−540.