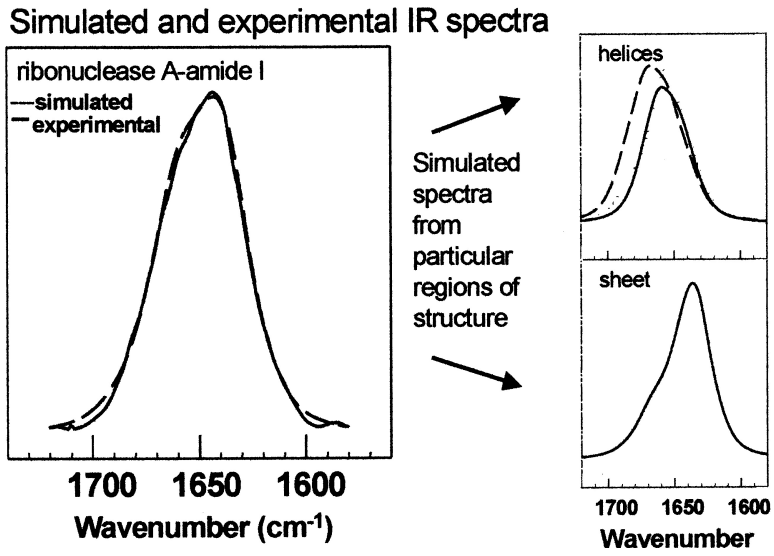
Article

# A Quantitative Reconstruction of the Amide I Contour in the IR Spectra of Globular Proteins: From Structure to Spectrum

Joseph W. Brauner, Carol R. Flach, and Richard Mendelsohn

*J. Am. Chem. Soc.*, **2005**, 127 (1), 100-109• DOI: 10.1021/ja0400685 • Publication Date (Web): 15 December 2004






## More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 10 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 12/15/2004

# A Quantitative Reconstruction of the Amide I Contour in the IR Spectra of Globular Proteins: From Structure to Spectrum

Joseph W. Brauner,* Carol R. Flach, and Richard Mendelsohn*

*Contribution from the Department of Chemistry, Rutgers University, Newark College of Arts and Sciences, 73 Warren Street, Newark New Jersey 07102*

Received February 27, 2004; Revised Manuscript Received September 20, 2004; E-mail: mendelso@andromeda.rutgers.edu

***Abstract:*** The Amide I contours of six globular proteins of varied secondary structure content along with a peptide model for collagen and pulmonary surfactant protein C have been simulated very closely by using a modified GF matrix method. The starting point for the method uses the three-dimensional structure as obtained from the Protein Data Bank. Elements of the interactions between peptide groups (e.g., transition dipole coupling) are very sensitive to tertiary structure, thus the current formalism demonstrates that the Amide I contour may be useful for a more detailed probe of 3-D conformation that goes beyond the traditional use of this band to probe the percentages of particular elements of secondary structure. For example, postulated changes to a known structure can be tested by comparing the new simulated band to the experimental band. A number of refinements to the transition dipole interaction calculation have been made. Most of the important interactions between the C═O oscillators that define the Amide I mode appear to have been identified, including through space transition dipole coupling, through valence bond and through hydrogen bond coupling. The eigenvector matrix produced by the method permits the contribution of each peptide group to the spectrum to be precisely determined. Analysis of the results shows that the often-used structure-frequency correlations are at best approximate and at worst misleading. The subbands from helices, sheets, turns, and loops are much broader and more overlapped than has been commonly assumed. Furthermore, the traditional α-helical marker band may be substantially distorted in short segments. Difference spectra based on isotope editing, a technique thought capable of revealing the spectral contributions of individual peptide groups, are shown to be prone to misinterpretation.

The understanding of protein-mediated biology at the molecular level requires techniques to monitor protein structural changes induced by ligand interactions or other external factors. As an example of the importance of protein structure, many disease states such as Alzheimer's, Kreutzfeldt-Jakob, diabetes, arthritis, etc. are caused by misfolded proteins that cannot respond properly to other proteins or substrates because they lack their normal three-dimensional structure. Among the techniques available for monitoring protein structural changes, Fourier transform infrared spectroscopy (FTIR) provides a rapid, inexpensive, convenient, noninvasive, nondestructive method for monitoring elements of protein structure in a variety of physically relevant states including solutions, crystals, monomolecular films, single cells and tissues. The feature of protein IR spectra most widely used as a structure probe is the Amide I vibration (mostly peptide bond C═O stretch) which falls between 1600 and 1700 $cm^{-1}$.

Spectra-structure correlations for protein IR spectra have evolved for half a century since the pioneering work of Miyazawa and Blout.[1,2] Hundreds of studies (for example see refs 3,4) based on the analysis of the Amide I band have attempted to determine the relative amounts of secondary structures such as helices, sheets, loops, and turns. This is accomplished by reproducing the contour after summing a number of putative subbands, each suggested to arise from a particular type of secondary structure. Resolution-enhancing techniques such as Fourier self-deconvolution or second derivative spectra are often used to delineate the positions of subbands in the contour which is then curve fit with bands (broadened for example, with Gaussian−Lorentzian line shapes) at the indicated positions. Certain frequency ranges are then assumed to correspond to particular secondary structures, e.g., α-helices are generally assumed to absorb between 1650 and 1660 $cm^{-1}$.[5,6] The relative band areas for each of the subbands are often taken to represent the percentages of the particular secondary structures.

As has been pointed out, there are several possible criticisms of this approach.[7] From a spectroscopic point of view, the intrinsic absorptivity (extinction coefficients) of different secondary structures may differ among proteins and within a single protein. In addition, the resolution enhancement technique of

(1) Miyazawa, T.; Shimanouchi, Y.; Mizushima, S. *J. Chem. Phys.* **1958**, *29*, 611−616.
(2) Miyazawa, T.; Blout, E. R. *J. Am. Chem. Soc.* **1961**, *83*, 712−719.
(3) Braiman, M. S.; Rothschild, K. J. *Annu. Rev. Biophys. Chem.* **1998**, *17*, 541−570.
(4) Surewicz, W. K.; Mantsch, H. H. *Biochim. Biophys. Acta* **1988**, *952*, 115−130.
(5) Byler, D. M.; Susi, H. *Biopolymers* **1986**, *25*, 469−487.
(6) Dong, A.; Huang, P.; Caughey, W. S. *Biochemistry* **1990**, *29*, 3303−3308.
(7) Jackson, M.; Mantsch, H. H. *Crit. Rev. Biochem. Mol. Biol.* **1995**, *302*, 95−120.


**10.1021/ja0400685 CCC: $30.25** 

Fourier deconvolution can produce artifacts (such as negative contributions to a contour), especially if the bands involved have different widths, which is often the case. Second derivative spectra are sensitive to trace water vapor and noise. Each level of derivation reduces the signal/noise ratio by a factor of 5. The group who first developed these tools have detailed these sources of error.[5] They also noted the main source of error in this procedure, namely, that in globular proteins extensive regions of perfectly regular secondary structure rarely exist. Only such regions would produce sharp bands at characteristic frequencies. The present work shows that the bands uncovered by these deconvolution procedures simply do not represent all of the oscillators in a given structure class, and, that they actually include substantial contributions from other structure types. These caveats have generally been ignored.

In light of these and related issues, some authors have suggested using statistical approaches such as factor analysis or analysis based on neural networks.[8] Such procedures, while robust, offer no molecular insight into the origin of the different frequency components in the Amide I contour. Sets of calibration spectra are needed and unusual structures are not handled well. In general, limits of error are on the order of ±10% for the estimation of the relative amount of various types of structure when compared to the results of X-ray structure determination.

It was recognized many years ago by Miyazawa[9] that a quantitative analysis of the contour may be understood from the fact that the broad Amide I band envelope results from vibrational coupling between localized Amide I motions of the individual peptide groups where each Amide I motion was described by a single oscillator with a common unperturbed frequency. The coupling produces Amide I modes of the collective system with a wide range of frequencies. The vibrational coupling is expressed in computing a force field or an F matrix of self-and interaction force constants for the oscillators. A G matrix based on reciprocal oscillator masses must also be calculated. The eigenvalues and eigenvectors of the resulting GF matrix product are related to the frequencies and intensities of the collective modes of vibration as will be explained.

Three basic approaches have been used to compute the force field for large peptides. The first approach is classical normal coordinate analysis or the Wilson GF matrix method where the oscillators being coupled represent individual or compound bond motions such as stretches, bends, etc. as used by Krimm and co-workers.[10,11] They developed F matrices of self-and interaction force constants for small peptide molecules and then pieced these together to get the F matrices for large molecules. This procedure can only be applied to polypeptides with regular structure; practical limits on the size of the model molecule (at that time) became evident as the size of the F matrix (n by n) grows rapidly for a molecule with n internal coordinates. The approach did not reproduce the broad Amide I contour and transition dipole coupling was added to shift the Amide I frequencies obtained.[10,11]

The second approach is illustrated by the relatively recent work of Lee and Krimm.[12] They used ab initio calculations of *N*-methylacetamide and L-alanyl-L-alanine to improve the force field previously estimated to calculate the Amide I band for α-helical poly (L-alanine). In a similar vein, Keiderling and co-workers have used ab initio calculations on somewhat larger but still small peptides to obtain force fields. These have been pieced together to simulate the Amide I bands of larger regular structures such as α-helices and *β*-sheets.[13,14] Small molecules are used in the ab initio calculations described here because the calculations are extremely time-consuming, sometimes requiring days. The off-diagonal elements in the force field so obtained are not broken down in terms of specific interactions such as hydrogen bonding, etc. An additional level of calculation is still required to simulate long range interactions, namely, transition dipole coupling (TDC). Using a slightly different approach, Dannenberg and colleagues performed ab initio calculations on peptides in a several geometrically optimized secondary structures.[15,16] By systematically choosing the positions of particular amino acids in the peptides, they were able to extract information about H-bond cooperativity and relative H-bond strengths.

The third approach and the one used here is the *modified* Wilson GF matrix method initiated by Miyazawa.[9] Here, the polypeptide is regarded as a collection of interacting oscillators, one oscillator for each peptide group. The first two types of calculation cannot be applied to globular proteins which have very little perfectly regular structure. In the current approach, the F matrix elements are calculated by modeling various interactions using geometry-sensitive formulas suggested by physical considerations. This approach allows for comparisons of the relative magnitudes of the various group interactions. A list of the interactions and the formulas used for their calculation are discussed later.

Variations of the modified Wilson GF matrix method have been reported with different authors emphasizing different coupling mechanisms. Miyazawa,[9] in his pioneering analysis of the Amide I contour of an α-helix, only allowed for interactions through valence bonds between adjacent groups and through H-bonds between the *i*th and the $i + 3$th groups. Torii and Tasumi[17] considered only TDC to model the Amide I bands of globular proteins. These approaches have each had some success (i.e., varying levels of agreement between simulated and experimental spectra) which indicates that simulation of the contour is a feasible goal.

In the current work, all of the generally agreed upon oscillator interactions are included. The numerical parameters in these formulas have been chosen empirically and no detailed a priori knowledge is claimed. However, (as shown below) excellent agreement is obtained between simulated and experimental Amide I contours using a single set of parameters for proteins containing a wide range of secondary structure.

It is unlikely that it will be possible to determine the complete three-dimensional structure of a protein from its Amide I contour. This can be understood by considering a typical globular protein with one hundred peptide bonds. The collection of oscillators will have one hundred vibrational modes in the

(8) Rahmelow, K.; Hubner, W. *Anal. Biochem.* **1996**, *35*, 5−13.
(9) Miyazawa, T. *J. Chem. Phys.* **1960**, *32*, 1647−1652.
(10) Krimm, S.; Abe, Y. *Proc. Natl. Acad. Sci. U.S.A.* **1972**, *69*, 2788−2792.
(11) Krimm, S.; Bandekar, J. *Adv. Protein Chem.* **1986**, *38*, 181−364.
(12) Lee, S.-H.; Krimm, S. *Biopolymers* **1998**, *46*, 283−317.
(13) Kubelka, J.; Keiderling, T. A. *J. Am. Chem. Soc.* **2001**, *123*, 6142−6150.
(14) Kubelka, J.; Keiderling, T. A. *J. Am. Chem. Soc.* **2001**, *123*, 12048−12058.
(15) Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2003**, *125*, 14065−14071.
(16) Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2003**, *125*, 8124−8129.
(17) Torii, H.; Tasumi, M. *J. Chem. Phys.* **1992**, *96*, 3379−3387.

Amide I region, each of small intensity. Due to coupling, individual oscillators lose importance and collective modes within this frequency range are formed by interactions of unknown magnitudes among an unknown number of oscillators. Because of the irregular structure of globular proteins, generally only a few groups participate significantly in a given collective mode of the system. This accounts for the small intensity of each collective mode. No known method of deconvolution can resolve the envelope of these inhomogeneously broadened bands. Even if the subbands could be resolved, it would be difficult a priori to ascribe each one to a particular collection of peptide groups or to deduce their geometrical relationships to other groups. This situation is in contrast to that for a regular, periodic peptide structure where a few intense collective modes result from the participation of many oscillators. By symmetry, their transition dipoles add to produce a large resultant.

The experimental Amide I spectra of the proteins in the current work were acquired in $D_2O$ solution, the usual medium for recording IR spectra of proteins. Although one might expect small but possibly noticeable geometry changes in going from a crystalline state (well-known however to be heavily hydrated) to a solution state, this problem does not seem to have materialized significantly. In addition, the solvent itself may distort the spectrum in ways beyond the scope of these calculations. However, for globular proteins, each subband in the envelope is weak and has only a weak electrical interaction with the solvent. This may account for the ability of a method based only on internal interactions to yield good agreement between the simulated and experimental spectra for a variety of diverse proteins.

The observed agreement between simulated and experimental spectra is taken as confirmation of the general correctness of the calculation. Therefore, any detailed conclusions which can be extracted about the contribution of individual or subsets of peptide groups to the spectrum are considered to be meaningful. The current approach also permits the prediction of the spectral changes resulting from proposed geometric modifications of the molecule due to, for example, ligand binding, or from isotopic substitution in particular peptide groups. The latter is a strategy commonly employed[18−24] for observing structure in particular regions in a protein.

## Materials and Methods

**Materials.** Five of the six globular proteins studied were purchased from Sigma (St. Louis, MO) and were of the highest purity commercially available. Recombinant rat intestinal fatty acid binding protein was obtained and purified as previously described.[25] Purified porcine lung surfactant protein SP−C was generously provided by Prof. Kevin Keough (Memorial University of Newfoundland). The synthetic collagen model peptide, $(PPG)_{10}$, was generously supplied by Prof. Barbara Brodsky (UMDNJ-Robert Wood Johnson Medical School, NJ). Deuterium oxide ($D_2O$) with 99.9% isotopic enrichment was purchased from Cambridge Isotope Laboratories (Andover, MA). All other reagents were of the highest quality available.

**Sample Preparation and FT-IR Spectroscopy.** Lyophilized protein samples were dissolved in a 5 mM $D_2O$ Tris buffer (∼10 mg/mL) at pD 7 containing 100 mM NaCl. The $(PPG)_{10}$ peptide was dissolved (∼5 mg/mL) in $D_2O$ at pD 3. Spectra were collected on a Mattson Instruments Research Series spectrometer equipped with a sample shuttle and MCT detector. Sample solutions (∼30 $\mu$L) were placed between $CaF_2$ windows separated with a 25 $\mu$m spacer. Spectra were obtained at room temperature by co-adding 4 blocks of 256 scans. The spectra were acquired at ∼4 $cm^{-1}$ resolution, apodized with a triangular function and Fourier transformed with one level of zero-filling to yield spectra encoded every ∼2 $cm^{-1}$. Using Grams/32 software (Galactic Industries), an appropriate spectrum of $D_2O$ buffer was subtracted from each sample spectrum. Linear baseline correction was performed except in those cases where side chain intensity in the low frequency region precluded the determination of a precise baseline endpoint. A spectrum of the highly hydrophobic lung surfactant protein SP−C was obtained using infrared reflection−absorption spectroscopy (IRRAS) at the air/water interface. SP−C was dissolved in chloroform (1 mg/mL) and ∼10 $\mu$L were spread on a buffered $D_2O$ subphase. Details of the IRRAS instrument have been described previously.[26] The spectrum was acquired at a surface pressure value of 30 mN/m using a sample shuttle and acquisition parameters as noted above.

**Calculations.** The programs for the calculations made here were written in Mathcad 2001 (Mathsoft, Inc., Cambridge, MA) and carried out on a PC with a speed of 2.4 GHz. A "C" program was written to cull the coordinates of the CON or CONH atoms that make the peptide groups from the PDB listing for a protein. The groups based on proline-N's are noted. The same basic program is customized with the proper data file names and other unique listings for the various proteins studied. The execution of a program for a given protein, that is, the time to obtain a simulated spectrum takes from less than a minute for a protein of 100 peptide groups to several minutes for a protein with 500 groups. The various "data mining" functions take only seconds to run. The program will not be published for external use but the authors will be pleased to assist those interested in the program.

**A Model for Simulation of the Amide I Contour.** Since the geometries of the peptide groups in proteins are all more or less similar we assume that any normal modes of vibration localized in the peptide group such as the Amide I mode have a common frequency in the absence of any vibrational coupling. It is well-known[1] that about 80% of the potential energy of Amide I mode comes from the C═O stretch. Minor contributing internal coordinates are the C−N stretch and N−H in-plane bend. The small contribution of the N−H in-plane bend explains the small (∼0.5%) frequency reduction in Amide I on deuteration. This N−H contribution is absent for proline-N based groups. A much larger frequency reduction in the Amide I mode occurs upon $^{13}C$ substitution in the group.

In the current model, each group is represented as a simple harmonic oscillator having a mass and a force constant. The derivative of the electric dipole moment of the group with respect to the normal coordinate of the mode is proportional to the transition moment of the mode.

The fact that the Amide I absorption can span a frequency range of 1600 $cm^{-1}$ to 1700 $cm^{-1}$ is due to appreciable coupling between the oscillators. The system of coupled oscillators itself has normal modes of vibration whose frequencies encompass the cited range. The number of normal modes for the collective system equals the number of oscillators.

(18) Ludlam, C. F. C.; Sonar, S.; Lee, C.-P.; Coleman, M.; Herzfeld, J.; Rajbhandary, U. L.; Rothschild, K. J. *Biochemistry* **1995**, *34*, 2−6.
(19) Ludlam, C. F.; Arkin, I. T.; Liu, X. M.; Rothman, M. S.; Rath, P.; Aimoto, S.; Smith, S. O.; Engleman, D. M.; Rothschild, K. J. *Biophys. J.* **1996**, *70*, 1728−1736.
(20) Haris, P. I.; Robillard, G. T.; van Dijk, A. A.; Chapman, D. *Biochemistry* **1992**, *31*, 6279−6284.
(21) Halverson, K.; Sucholeiki, I.; Ashburn, T. T.; Lansbury, R. T. *J. Am. Chem. Soc.* **1991**, *113*, 6701−6703.
(22) Sonar, S.; Lee, C.-P.; Coleman, M.; Patel, N.; Liu, X. M.; Marti, T.; Khorana, H. G.; Rajbhandari, U. L.; Rothschild, K. J. *Nat. Struct. Biol.* **1994**, *1*, 512−517.
(23) Tadesse, L.; Nazarbaghi, R.; Walters, L. *J. Am. Chem. Soc.* **1991**, *113*, 7036−7037.
(24) Zhang, M. J.; Fabian, H.; Mantsch, H. H.; Vogel, H. J. *Biochemistry* **1994**, *33*, 10833−10888.
(25) Hsu, K. T.; Storch, J. *J. Biol. Chem.* **1996**, *271*, 13317−13323.
(26) Gericke, A.; Flach, C. R.; Mendelsohn, R. *Biophys. J.* **1997**, *73*, 492−499.

A modified Wilson GF matrix method is used.[27] The G matrix is diagonal. The elements are the reciprocal oscillator masses. The reciprocal mass of proline-N based oscillators has been decreased by a factor of 0.978 compared to nonproline-N based oscillators. For $^{13}C$ substitution the reciprocal mass is decreased by a factor of 0.951. The diagonal F matrix elements are the force constants for the oscillators which are initially all the same. The off-diagonal F matrix elements are the interaction force constants which are calculated as described below. They are expressed as a certain fraction of the diagonal force constants. The off-diagonal elements in a given row are added into the diagonal element as is made clear in Symon's exposition of coupled oscillators.[28] The GF matrix product is then diagonalized. The eigenvalues are related to the frequencies of the normal modes ($4\pi^2c^2\nu^2 = \lambda_0$) of the collection of oscillators while the *ij* th element of the reciprocal eigenvector matrix, $L^{-1}$, gives the contribution of the *j*th oscillator to the *i*th normal mode. This is the proper weighting factor to be used when computing the vector sum of the individual transition dipoles to get the length of the resultant dipole for a given normal mode ($L^{-1}R = Q$).[27] The square of the length of this resultant is proportional to the intensity of a given mode. In assessing the *relative* contribution of any oscillator to a given mode it is the square of the $L^{-1}$ matrix element that must be used. The sum of the squares of the elements in any row equals one.

In performing calculations the mass and force constant for the general oscillator need not be individually specified. Rather, we set $k/m = 4\pi^2c^2\nu^2 = \lambda_0$ where $\nu = 1650$ cm$^{-1}$. The F matrix is divided by *k* while the G matrix is multiplied by *m* so that the eigenvalues of the GF matrix product are multiples of $\lambda_0$. The various interaction constants are defined below:

**1. Through Space Electrical Interactions. Transition Dipole Coupling.** The vibrating peptide group has an electric dipole moment that varies with change in length of the normal coordinate; the derivative formed from these changes is proportional to the transition dipole moment. Physically, the transition dipole acts as an oscillating electric dipole but it can be regarded as quasi-static if its physical size and the range of its action is small compared to the wavelength of its radiation. The dipole−dipole interaction formula based on the point dipole approximation was not used. Rather, a finite size of 2.2 Å, the N to O distance in a peptide group, was ascribed to each. The orientation of the transition dipole in the molecular coordinate system is obtained by rotating a vector off the CO bond 21° away from the CN bond for nonproline-N based groups. For the proline-N based groups the angle was taken as −7°. For cis-proline-N groups the angle used was 2°. The transition dipole strength of a nonproline-N based group was set to 300 cm$^{3/2}$/s while that for a proline-N group was set to 400 cm$^{3/2}$/s. Because of the units used for the transition dipole strength, the TDC interaction terms were divided by $\lambda_0$ to make them dimensionless, as are the other terms in the F matrix. For globular proteins, these interaction terms were also divided by $r^{0.9}$ to model the dielectric constant of the intervening matter.

The details of the transition dipole aspect of the calculation have been described earlier.[29] An improvement in the calculation was made in that the receiving dipole, *j*, 2.2 Å in length and in the field of the sending dipole, *i*, was divided into 10 0.22 Å contiguous dipoles end-to-end such that they add to give the original dipole. The separate interactions for each sub-dipole in the field of dipole *i* were then added. This was done because the electric field of sending dipole *i* varies in space and the variation is reduced over a smaller receiving dipole and thus conforms more closely to the interaction formula $W_{ij} = \mathbf{p}_j \cdot \nabla\Phi_i$ which is based on a point dipole.

Thus, the explicit expression for the interaction force constant is $F_{i,j} = (\partial p_i/\partial q_i)(\partial p_j/\partial q_j)\ \mathbf{u}_j \cdot [\mathbf{i}_r(-2\cos\theta/r^3 - \mu^2(5\cos^3\theta - 3\cos\theta)/2r^5) + \mathbf{i}_\theta(-\sin\theta/r^3 + \mu^2(-15\cos^2\theta\sin\theta + 3\sin\theta)/8r^5)]$.[29] Here, $(\partial p_i/\partial q_i)$ represents the transition dipole strength of group *i*, $\mathbf{u}_j$ is a unit vector along transition moment *j*, $\mathbf{i}_r$ and $\mathbf{i}_\theta$ are unit vectors along a line in local polar coordinates (origin at the center of dipole *i*) that connects the center if dipole *i* with the center of the sub-dipole of dipole *j* (both dipoles are centered on the carbon atom of the respective peptide group), *r* and $\theta$ are the local polar coordinates of the center of the sub-dipole of dipole *j*, and $\mu$ is the finite size of the transition dipole of a group, namely, 2.2 Å.

**2. Through Valence Bond Interactions.** Each peptide group is separated from its neighboring group by an intervening $C^\alpha$ atom which is covalently bound to the N of the first group and the carbonyl C of the second. The groups themselves are generally planar. The $\psi$, $\phi$ angles at a given $C^\alpha$ determine the relative orientation of the adjacent peptide planes. Accuracy in the simulations requires two components to the bond-electron-mediated interaction. These are a through $\sigma$-bond term which is constant and a $\pi-\pi$ component that is sensitive to the relative orientation of the peptide group planes. We have used the following expression for this interaction:

$$(0.00244 + 0.0151\cos\xi)\,s$$

where $\cos\xi$ is the cosine of the angle between the normal of the first plane and the negative of the normal to the second plane. The normals are found by crossing the CN bond vector into the CO bond vector. The strength factor, s, was set to 0.6.

**3. Electrostatic through Hydrogen Bond Interactions. A. Off-Diagonal Elements.** Hydrogen atoms are not located in X-ray structures but they are determined in NMR structures. When the locations are not supplied, they have been calculated and included in the group structure by assuming a published standard geometry for the peptide group.

The hydrogen bonds between peptide groups are "moderately" strong and the interaction is basically electrostatic, that is, a dipole−dipole interaction. There is no build up of electron density between the O and H atoms.[30] The hydrogen bonded pairs of peptide groups are found by a function that limits the CO···HN distance to 2.6 Å or less and the C−O, H−N angle to 120° or more. The following expression is used for this interaction:

$$\{[3\cos^2(\kappa\text{-}116) - 1]/d^3\}s$$

where $\kappa$ is the C−O, H−N angle and *d* is the CO···HN distance. The strength factor, *s*, was set to −0.011. This formula was suggested from the equation describing the axial component of the field of a point dipole. The offset of 116° was chosen empirically because the unshared electron pairs on the carbonyl oxygen in a simple Lewis model are 120° from the C−O bond and the location of the center of the electron density on the carbonyl oxygen determines the orientation of the interacting dipole on the acceptor moiety.

**B. Diagonal Elements.** The "weak coupling" assumed by Miyazawa implies that the interaction does not modify the oscillators being coupled. However, the hydrogen bond exerts considerable force and it is well-known to modify the vibrational frequencies of both the acceptor and donor groups. If this happens in the peptide group interactions being studied here, then the diagonal force constants will have to be directly modified compared to non-hydrogen bonded groups. This diagonal force constant modification is in addition to those produced by adding in the off-diagonal hydrogen bonding interactions as discussed above (Section 3A). In optimizing the simulations it was not necessary to change the diagonal force constant of the donor group, i.e., the NH supplying group, but a correction to the diagonal element of the acceptor was needed. Therefore, an amount equal to −11 times

(27) Wilson, E. B.; Decius, J. C.; Cross, P. C. *Molecular Vibrations. The Theory of Infrared and Raman Vibrational Spectra*; McGraw-Hill: New York, 1955.
(28) Symon, K. R. *Mechanics*; Addison-Wesley: Reading, 1960; p 188.
(29) Brauner, J. W.; Dugan, C.; Mendelsohn, R. *J. Am. Chem. Soc.* **2000**, *122*, 677−683.
(30) Jeffrey, G. A. *An Introduction to Hydrogen Bonding*; Oxford University Press: New York, 1997; p 20.

the hydrogen bonding interaction constant as calculated above was added to the diagonal force constant for the CO supplying groups. This increases its value by less than one to as much as five percent.

**4. Through Hydrogen Bond $\pi-\pi$ Interactions. A. Off-Diagonal Elements.** In explaining the anomalous intensity of $^{13}C$ substituted $\beta$-sheet forming peptides[29] we took the through hydrogen bond interaction in a sheet to be about twice as strong as that in a helix even though the O···H distances and the N−H···O−C angles are similar. This assumption was necessary in order to get the large Amide I splitting between the in plane parallel and perpendicular sheet modes seen for the peptides being studied. However, we have come to appreciate there is another source of interaction across hydrogen bonds in sheets, namely, a $\pi-\pi$ interaction between the $\pi$ electrons in the peptide groups. This interaction is not important in helices because the peptide planes do not align as well as in a sheet. The effect is relatively weak and it depends on the number of strands in the sheet. For the values of the parameters used here some forty strands are required to achieve the approximately 70 $cm^{-1}$ splitting observed for small model peptides. This is consistent with the results of Kubelka and Keiderling[13] whose ab initio study showed the splitting observed for similar peptides increases with the number of strands in the sheet.

The expression for this interaction is

$$s\,[\cos\eta]^2\,(N-2)$$

where cos $\eta$ is the cosine of the angle between the normal of the first peptide group plane and the negative of the normal to the second plane. The normals are found by crossing the CN bond vector into the CO bond vector. The number of groups laterally bonded in a sheet is $N$. This number was found from the function mentioned above which looks for serial connections between the hydrogen bonded pairs The strength factor, $s$, was set to $-0.0002$.

**B. Diagonal Elements.** As in the electrostatic through hydrogen bond interaction case the diagonal force constant of the acceptor groups were additionally modified by $-10$ times the $\pi-\pi$ interaction for those groups.

## Results

**1. Simulated and Experimental IR Spectra for 6 Globular Proteins and 2 Peptides.** Simulated and experimental spectra of six globular proteins (ribonuclease A (1AFU), rat intestinal fatty acid binding protein (1IFC and 1AEL), alpha chymotrypsin (1EX3), hen egg white lysozyme (5LYM), horse heart myoglobin (1YMB), papain (9PAP), and 2 peptides (a synthetic collagen model, $(PPG)_{10}$ (1K6F) and pulmonary surfactant SP−C (1SPF)) are shown in Figures 1A-H, respectively. The alphanumeric symbol in parentheses provides the PDB file number. These molecules encompass a wide variety of secondary structure distributions from myoglobin which is largely helical, to the fatty acid binding protein which is predominantly $\beta$-sheet. The structural data are from X-ray diffraction or NMR experiments. In the simulations of the IR spectra, line-widths of the subbands are chosen to be between 8 and 12 $cm^{-1}$ for the globular proteins, a single value being used for all the subbands of a particular molecule. In addition, the overall simulated contours are shifted to make the maxima coincide with the experimental spectra. All the IR data are from bulk phase measurements in $D_2O$ solution, with the exception of the IR data for SP−C which are from (inverted) IR reflection−absorption spectroscopy (IRRAS) measurements of a monolayer film of the protein at the air/water interface on a $D_2O$ subphase. These latter data (Figure 1H) exhibit some apparent noise due mostly to residual water vapor. In general, the agreement between experimental and calculated spectra is excellent,

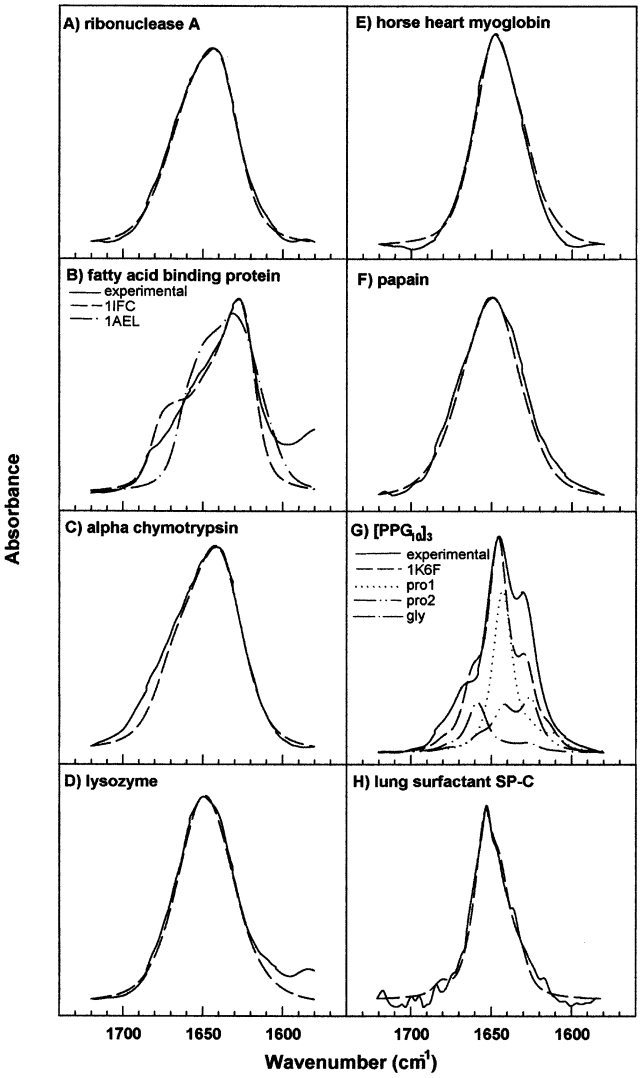



***Figure 1.*** Experimental and simulated Amide I contours for six globular proteins (A−F) and two peptides: (G) a synthetic collagen model peptide, $[PPG_{10}]_3$, and (H) pulmonary surfactant protein SP−C. The solid line is the experimental spectrum and the dashed line is calculated from the protein data bank entries listed in the text. Simulations for two PDB entries are shown for (B) the fatty acid binding protein: 1IFC from X-ray crystallography and 1AEL from NMR experiments. Calculated subbands corresponding to contributions from specific amino acid residues are shown (G) for $[PPG_{10}]_3$.

especially for ribonuclease A. This simulation will be "data mined" to exemplify the main points of this paper. In addition, some points will made about the myoglobin, $(PPG)_{10}$, and the fatty acid binding protein results.

**2. Data Mining. A. Ribonuclease A.** One useful result of the current algorithm is that Amide I features arising from any desired segments of the protein may be isolated. The protein data bank entry (1AFU) provided the atomic coordinates for the band simulation of ribonuclease A. The biological unit is a single chain of 124 residues. The entry listed the coordinates of a dimer so only the coordinates of one chain were used. For chain A, twenty-six residues were listed as being in three helices of about equal length and 24 residues were ascribed to two sheets. No residues for turns or loops were listed.

The current program has functions that report hydrogen bonding patterns which in turn permit one to determine whether a group is part of a helix, sheet, turn, or loop. Groups not

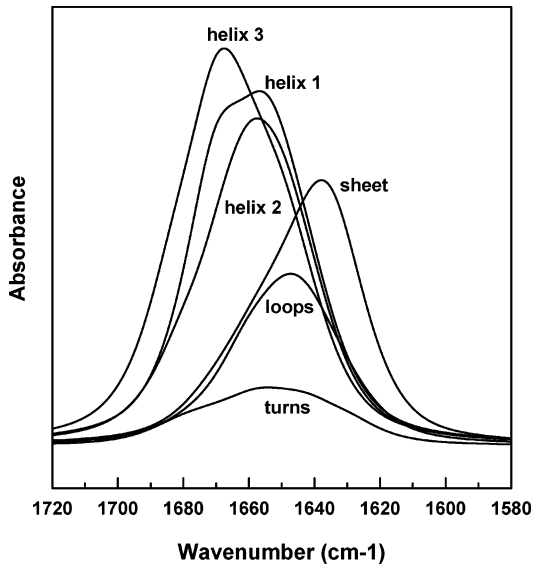



***Figure 2.*** "Data mining" for ribonuclease A. The calculated spectral contributions for oscillators in the indicated secondary structures are noted. Helical intensities have been multiplied by ten. Helix 1 and 2 have the classical α-helical hydrogen bonding pattern and helix 3 is more tightly wound. The sum for several different sheet regions of various length and width (total of 53 peptide groups) is shown including three turns within sheet structures. Various types of turns differing in the number of groups separating the hydrogen bonded residues as described in the text and loop (random coil) regions composed of 40 peptide groups yield broad bands.

hydrogen bonded to other groups according to the criteria mentioned in the through H-bond interaction section were ascribed to loops. Using this function, we found the three helices as given in 1 AFU but the sheet structure found was more extensive than that listed. This may be due to the fact that the cutoff CO···HN distance and angle in the hydrogen bond finding function were 2.6 Å and 120° which exceeds the typical sheet values of 1.88 Å and 155°. Actually, most of the additional interactions were found with values well inside this extended cutoff. These structures and the IR subbands they produce will now be described in some detail. The calculated subbands are plotted in Figure 2 to allow comparison with the analysis below.

The PDB starts the residue numbering at one. In our numbering, residues one and two form peptide group zero. Residue 1 supplies the CO and residue 2 supplies the NH of the peptide group. Helix 1 spans peptide groups 2 to 10 and the simulation yields a somewhat flattened asymmetric band shape. Helix 2 is wound from groups 23 to 30 and its band has a maximum at 1657 $cm^{-1}$ with a slight shoulder at ~1680 $cm^{-1}$. These two helices have the classical α-helical structure in that the carbonyl oxygen of one group is hydrogen bonded to the amine hydrogen of a group three units down the chain. Helix 3 spans groups 49 to 57 and is more tightly wound than the other two in that some of the hydrogen bonds are between groups separated by two units as well as three units. One can actually see the tighter coiling by plotting the groups involved in Mathcad's 3-D plotting facility and looking down the helix axis. This visualization also serves as a check as to exactly which groups belong to a given sub-structure. The CO···HN angle in helix 3 is on the order of 125° as compared to about 155° in the α-helical case. According to the formula just given the hydrogen bonding interaction is calculated to be stronger for the smaller angles. Since the frequency increase for CO supplying groups is proportional to the hydrogen bond interaction, the more tightly wound helix has more oscillators with higher frequencies, hence the band due to helix 3 has a peak maximum at around 1668 $cm^{-1}$. These three subbands all have a fwhh of some 45 $cm^{-1}$ which is much broader than is commonly assumed in curve fitting exercises.

As can be seen from the ribbon and coil representation of this molecule as supplied by the PDB, the sheet structure is complicated. On the basis of the hydrogen bond finding function in the simulation program and a second function that looks for serial connections among the bonded pairs, 53 groups were found in sheet structures. These results were mapped and a clear and consistent picture of the sheet structure emerged. In one part of the molecule, runs of six groups form a three stranded sheet. In another part, runs of three groups form a four stranded sheet. The remaining groups are in two stranded sheets. Together these groups produce a broad contour with a maximum at 1638 $cm^{-1}$ and substantial asymmetry on the high frequency side. This band spans practically the entire Amide I region and it has considerable intensity in the region usually assigned to helices and loops.

The eight pairs of groups found in turns give rise to an IR absorption band that spans the entire Amide I region. The wide frequency span results from the variety of turn types present, i.e., the group number separations of the pairs are: 2,2,4,5,2,2,2, and 1. Three of the turn pairs are also counted as part of the sheet structure and they connect beta strands, three are at the ends of helices, and two form wide turns.

The 40 groups that are not hydrogen bonded to other groups are ascribed to loops or what are sometimes called random coils. These structures are of course, neither random nor coils. For the IR spectrum of ribonuclease A, these give rise to a band with a broad maximum between about 1635 $cm^{-1}$ to 1660 $cm^{-1}$.

The intensity per group ratios in arbitrary units for helix 1, helix 2, helix 3, sheet, turns, and loops were found to be 2.20, 2.17, 2.43, 2.57, 2.50, and 2.37, respectively. These results indicate that the intrinsic absorbances of different structural elements (even of the same secondary structure type) are not the same. A range of about 15% is noted.

**B. Myoglobin.** Horse heart myoglobin also exhibits excellent agreement between the simulated and experimental spectra (Figure 1E). This structure is comprised of eight helical segments connected by turns. The group numbers in each of the eight helices are: 2...16, 19...34, 35...40, 50...55, 57...77, 81...94, 100...117, and 123...148. The simulated bands for each helix are shown in Figure 3. They all produced bands of the usual shape except for helix 3 which displays a triple humped band with the low-frequency peak at 1610 $cm^{-1}$ and a high-frequency peak at 1660 $cm^{-1}$. The peak frequencies of the normal looking bands fell between 1644 and 1653 $cm^{-1}$, consistent with the usual spectra-structure correlations. Deviations from the usual assignment may occur in shorter helices.

**C. $PPG_{10}$.** $PPG_{10}$ is a synthetic polypeptide consisting of 10 triplets each of which consists of two proline and one glycine residue in the indicated sequence. It serves as a model for the triple helix of collagen. The PDB entry used for the atomic coordinates of $PPG_{10}$ was 1K6F. This entry lists the coordinates for $[(PPG)_{10}]_3$. Actually, the coordinates of a dimer were listed, but only the coordinates of the first triple helix were used. The structure is a tightly wound rod much different than the globular proteins used in the other simulations. One change was required in the simulation parameters used to get an adequate fit. The

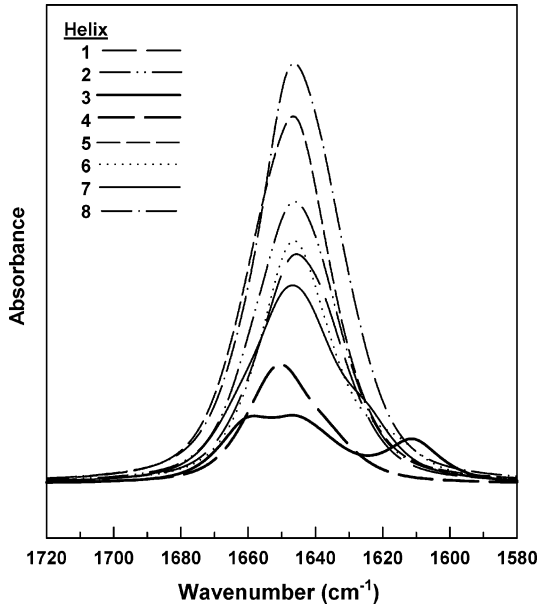



***Figure 3.*** "Data mining" for horse heart myoglobin. The calculated spectral contributions for oscillators in the eight helical regions are plotted. The two short helices (3 and 4) have the lowest intensity Amide I contours and are highlighted by thicker lines. The triple humped contour of helix 3 may be a manifestation of "end effects" in short helical segments.

dielectric constant function in the TDC interaction term was changed to $r^{0.4}$ instead of the $r^{0.9}$ used in the globular protein simulations. No water is inside the rod and the structure is no doubt more rigid compared to the globular proteins.

The simulated band reproduces the main features of the experimental band which was obtained in $D_2O$ solution at 16 °C but the match is far from perfect (Figure 1G). This may be due to a bulk solvent distortion of the experimental spectrum. A few of the subbands have considerable intensity because of the larger amount of regular structure compared to the globular proteins considered here and their larger transition dipoles will experience a stronger interaction with the solvent. Furthermore, the thin, elongated, rodlike structure means more total bulk solvent exposure for the transition dipoles of all the various modes.

There is only one secondary structure, the triple helix, but the Amide I contour is broken down into contributions that come from three classes of peptide groups, i.e., those whose N atom comes from a Pro1, Pro2, or Gly residue where the numbering corresponds to the PPG sequence. This will permit comparison with the analysis of the Amide I band of $(Gly\text{-}Pro\text{-}Pro)_8$ made by Lazarev et al.[31]

Lazarev et al. assigned each of the three distinct lobes of the Amide I contour to the three different kinds of peptide groups. We find this to be only approximately true in that all three peptide types contribute to some extent to all three lobes (Figure 1G). Also, we find a reversal of the 1628 and 1642 $cm^{-1}$ assignments. Their first imine carbonyl $C_1O_1$ is part of our Pro1. Lazarev assigns the 1628 $cm^{-1}$ band to this group while our results show Pro1 groups contribute strongly to the 1642 $cm^{-1}$ band. According to Lazarev these carbonyls are free or hydrogen bonded to the solvent. In this case an important source of intergroup coupling is missing and a more localized spectral pattern for Pro1 is expected like the one seen here.

**D. Fatty Acid Binding Protein.** Two structures have been used to simulate the band for fatty acid binding protein as shown in Figure 1B. These are 1IFC (an X-ray crystal structure) and 1AEL (an NMR solution structure). The use of a second structure was prompted by the divergence of the experimental and simulated band based on the X-ray structure in the high wavenumber region. Visualization of the two structures shows them to be similar but the hydrogen bond finding function in the simulation program shows them to be quite different. It finds 44 hydrogen bonded oscillator pairs in the 1AEL structure and 89 in the 1IFC structure. The X-ray structure appears to be the tighter of the two. We note that the experimental band measured in $D_2O$ solution falls between the two simulations.

## Discussion

As previously mentioned, the starting point for the present Amide I simulation method is a structure, either actual or proposed. The reverse process of spectrum to structure is not considered here for reasons already outlined. The main benefit of the present calculation is to obtain an understanding of how the important geometry sensitive peptide group interactions can be modeled and combined to produce the observed Amide I contour. With a reliable model in hand the effect of proposed structure changes or isotopic substitution can be calculated. The overall band shape is very sensitive to localized conformational changes and isotopic substitution. For example, Naumann and co-workers[32] found a substantial change in the Amide I band of ribonuclease T1 on introducing four $^{13}C$-proline isotope labels, the details of which will be discussed in a future publication. This sensitivity exists because changes made in one part of the F matrix affect all the eigenvalues and eigenvectors. It is impossible to tweak one interaction parameter to produce a change in just one portion of the band. To fit the diverse group of proteins considered here, the range of values for the set of parameters was found to be highly constrained. The importance of each of the interaction terms can be seen in Figure 4 where the simulated contours for ribonuclease A obtained after turning off one interaction at a time are shown.

In the present discussion, a history of the current formalism and the effects of isotopic substitution will first be reviewed, followed by descriptions of several potential applications.

**1. Comparison of the Current Approach with Prior Methods.** Attempts at quantitative analysis of the IR peptide bond vibrations began with Elliot and Ambrose[33] who noted a correlation between the structure of certain geometrically uniform polypeptides and the peak position of their Amide I band. No explicit consideration of the intergroup forces involved was attempted. In 1960, Miyazawa[9] developed a perturbation treatment for the Amide I and Amide II vibrations in ordered structures to understand the origin of frequency shifts from those of an isolated peptide group. Computational limitations determined the complexity of the model that could be utilized. Miyazawa noted that interactions mediated by both valence and hydrogen bonds should cause a first-order frequency shift. No detailed recipe was given for calculating the off-diagonal $G^{-1}$ and F matrix elements which are the kinetic and potential energy interaction terms, only that they varied in a regular way with

(31) Lazarev, Y. A.; Grishkovsky, B. A.; Khromova, T. B. *Biopolymers* **1985**, *24*, 1449−1478.

(32) Moritz, R.; Fabian, H.; Hahn, U.; Diem, M.; Naumann, D. *J. Am. Chem. Soc.* **2002**, *124*, 6259−6264.

(33) Elliott, A.; Ambrose, E. J. *Nature* **1950**, *165*, 921.

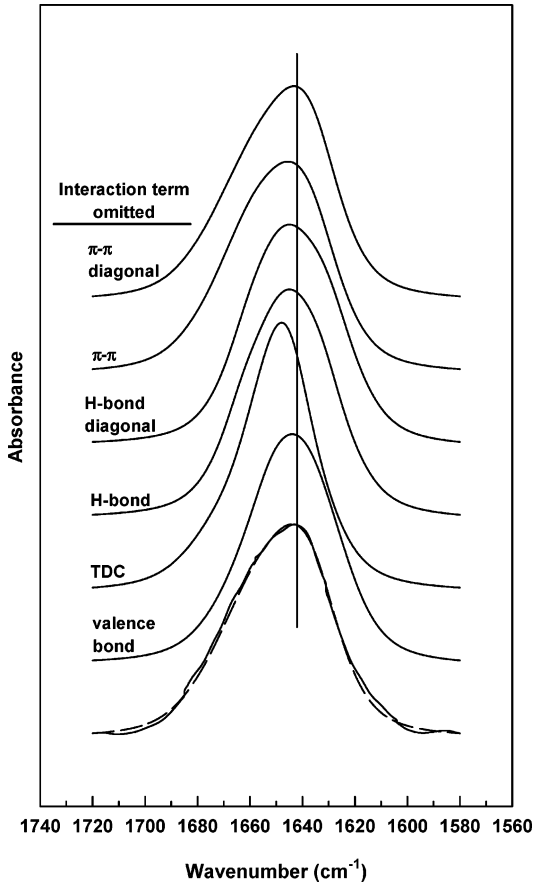



***Figure 4.*** Significance of each interaction term in determining the overall Amide I band contour for ribonuclease A. The bottom, solid line is the experimental spectrum overlaid with the simulated contour (dashed line). Other simulated contours are shown as one interaction term was omitted from the calculation at a time. The vertical line guides the eye to highlight the significance of each term.

the separation of the groups. The F and $G^{-1}$ matrixes had a simple structure amenable to straightforward analytical solutions. When this model is applied to any infinite regular chain, the coupling of the oscillators does not produce a multiplicity of frequencies but merely a shift in a single band for a single structure.

This work was extended[2,9] to empirically evaluate the inter- and intrachain interaction constants from the spectra of polypeptides of known structure. The regular structures dealt with in these two papers include the α-helix, the extended configurations of the parallel and antiparallel pleated sheets, and interestingly enough the "random coil". To justify the name given to this last case, the argument was made that the interactions average to zero. Thus, the infinite helix has two IR active modes closely spaced, the antiparallel pleated sheet has three, two of which are widely separated, while the random coil has only the frequency of an unperturbed peptide group.

As IR technology improved and the acquisition of Amide I spectral data from globular proteins in solution became a routine biochemical experiment, the spectra-structure correlations for the infinite regular structures outlined by Miyazawa and Blout became the accepted basis for analysis of the Amide I contour in the irregular structure of globular proteins. It is now obvious that globular proteins generally do not have very long segments of regular structures and in fact even the short regions of regular structures have a range of bond lengths and angles. An additional factor that tends to broaden the spectral distribution of a given secondary structure, that was not considered directly by Miyazawa, was the occurrence of through-space electrical coupling of transition dipoles. For these reasons, the generally accepted structure-frequency correlations as applied for analysis of the spectra of globular proteins break down. Over the years, the shortcomings of this approach have become apparent. To better understand interactions between peptide groups and how these affect the Amide I contour, more detailed attempts to understand the important physical forces involved have been undertaken as described in the Introduction.

In the current simulations of the Amide I contour, we have not prejudged which forces are most relevant, but have taken into account the various interactions used by prior workers, namely through valence bond and hydrogen bond coupling, and TDC. A number of refinements have been made in the calculation of these forces. These are: a finite size for the transition dipoles, an explicit distance dependent expression for the dielectric constant due to intervening matter in the through space electrical interaction, a geometry dependence of the through bond interactions, a frequency shift in the acceptor hydrogen bonded groups, and a $\pi-\pi$ interaction across extended H-bonds. Bound water and bulk solvent effects have not been considered in the model. The agreement between simulated and experimental spectra (Figure 1) speaks to the utility of the model used.

Our results for the positions and shapes of the subbands for helices, sheets, turns, and loops show some correspondence with the usual structure-frequency correlations assumed for globular proteins. However, several basic differences are apparent. First, the frequency distribution of appreciable intensity due to each secondary structure is much broader than is often assumed. For example, Holloway and Mantsch,[34] in fitting the polar domain of Cytochrome $b_5$ in $D_2O$ using Fourier self-deconvolution as a guide, found six bands between 1700 and 1625 $cm^{-1}$. The fwhh of these features were all between 10 and 15 $cm^{-1}$. The bands at 1625, 1675, and 1683 $cm^{-1}$ were designated as sheet, the band at 1650 $cm^{-1}$ was ascribed to α-helices, the band at 1662 $cm^{-1}$ was said to be due to turns while that at 1639 $cm^{-1}$ was assumed to be from a $3_{10}$ helix. No band was ascribed to loops. This approach is typical of a large number of such analyses in the literature. These narrow bandwidths and unique assignments are in marked contrast to the results reported herein for Ribonuclease A.

**2. Isotope Labeling.** A $^{12}C$ carbonyl group treated as a diatomic harmonic oscillator at 1650 $cm^{-1}$ will shift to 1614 $cm^{-1}$ on $^{13}C$ substitution. It is of interest to see the effect predicted by the present formalism when $^{13}C$ substitution is made in particular peptide groups of globular proteins. Comparisons of calculated and experimental difference spectra should provide a test of both the correctness of the current algorithm, and of the possible utility of isotopic substitution for elucidating the active region of a protein undergoing a conformational change. Also, it has been suggested that isotope editing could be used to resolve the vibrational frequencies of individual peptide bonds in diverse environments. Therefore, we simulated the effects of $^{13}C$ labeling a peptide group in helix 3 (oscillator 52), and in a sheet region (oscillator 80) in Ribonuclease A.

(34) Holloway, P.; Mantsch, H. H. *Biochemistry* **1989**, *28*, 931−935.

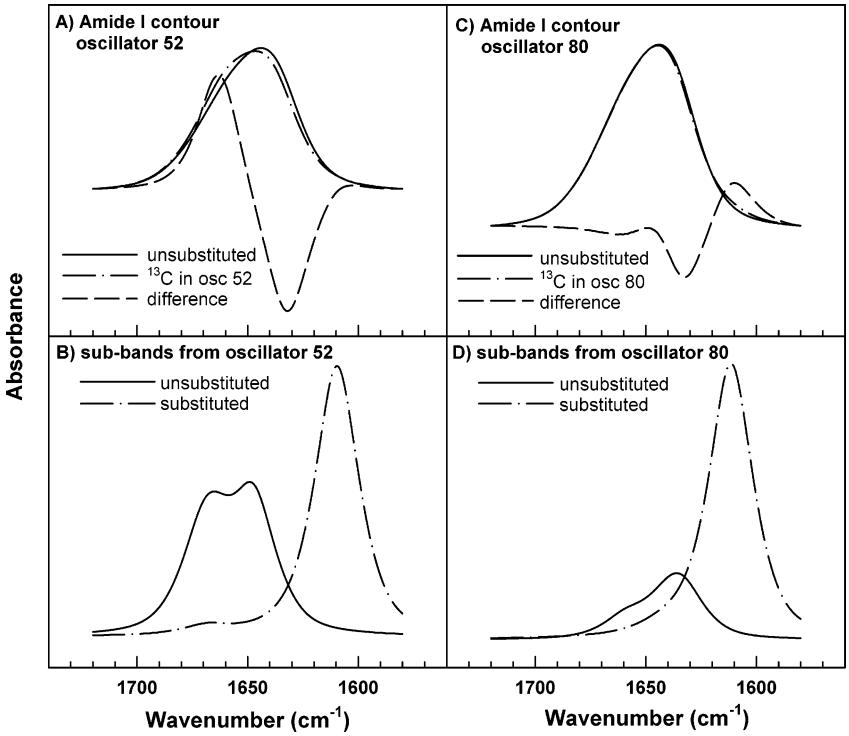



***Figure 5.*** Calculated effect of $^{13}C$ substitution in two different secondary structure regions of ribonuclease A. Oscillator 52 resides in helix 3 and oscillator 80 in a sheet region. (A) and (C) show the Amide I contours for the molecule containing unsubstituted and substituted oscillators 52 and 80, respectively, along with each enlarged (by a factor of 10) difference spectrum. (B) and (D) depict the simulated subbands from both $^{12}C$ and $^{13}C$ isotopomers for oscillators 52 and 80, respectively.

The results of these simulations are shown in Figure 5A−D. Panels A and C show the Amide I contours for the molecule containing unsubstituted and substituted oscillators 52 and 80, respectively and the enlarged (by a factor of 10) difference spectrum. Panels B and D show the spectral contributions of oscillators 52 and 80, respectively, before and after substitution. The results for oscillator 80 agree with our expectations. In the difference spectrum intensity disappears from a higher frequency region and appears at a lower frequency and the band shapes in panels C and D agree. The difference spectrum for oscillator 52 is disconcerting. Intensity disappears from a lower frequency region and appears at a higher frequency. Some explanation is needed. First, we note that panels B and D convey the same information. Before substitution an oscillator participates in many modes mostly to a small extent and to a substantial extent in only one or two modes, all at frequencies above that of the substituted oscillator. This pattern is revealed by examining the inverse L matrix. On broadening these frequencies and summing we get the contour shown. After substitution each group oscillates in a practically pure mode around 1610 $cm^{-1}$. The isolation of a single oscillator by increasing its mass is due to the antagonistic effects of kinetic and potential energy coupling as noted by Miyazawa.[9]

Obviously, the difference spectrum does not always reveal the original and new spectral contributions of a substituted group. The undulations in the difference spectrum are due to the changes in the coupling patterns between the two isotopomers. If the substitution is made in a region undergoing conformational change, then the difference spectrum between the two conformers should be distinguishable from those where the substitution is made in a static part of the molecule. Such a procedure has been used by Rothschild and co-workers in defining the main structural modification responsible for the bR to N transition in bacteriorhodopsin.[18]

**3. Other Applications of the Method.** The atomic coordinates in the Protein Data Bank permit the dihedral $\phi-\psi$ angles in a protein backbone to be calculated. A function can be written to change these to any value desired and so generate a new structure which in turn gives rise to a new simulated Amide I band. Thus, for example, the spectra obtained in the thermal denaturation of $PPG_{10}$, the collagen model peptide, are being simulated in our laboratory by changing the dihedral angles to unwind the triple helix. This is an effort to understand the thermal denaturation of collagen, a protein folding study in reverse.

In a similar vein, any starting structure can be changed in this way to simulate the Amide I contours of stable intermediates formed in protein folding studies. For example, in the bacteriorhodopsin paper cited above, Rothschild and co-workers[18] surmise that the Tyr 185-Pro 186 region may serve as a hinge around which changes in the F-helix orientation occur. Such a change is easily modeled and the band generated from the new structure can be compared to the experimental band to provide a higher level of confidence in the original hypothesis. In addition, this approach may provide information for certain enzyme−substrate interactions or in structural investigations of homologous protein families where only particular protein structures have been solved.

The current formalism can be applied to distinguish between proposed regular secondary structures for smaller peptides as well. There is an active debate in the literature as to whether the 39 to 43 residue strands in $\beta$-amyloid fibrils are organized in a parallel or antiparallel manner. Different experiments have

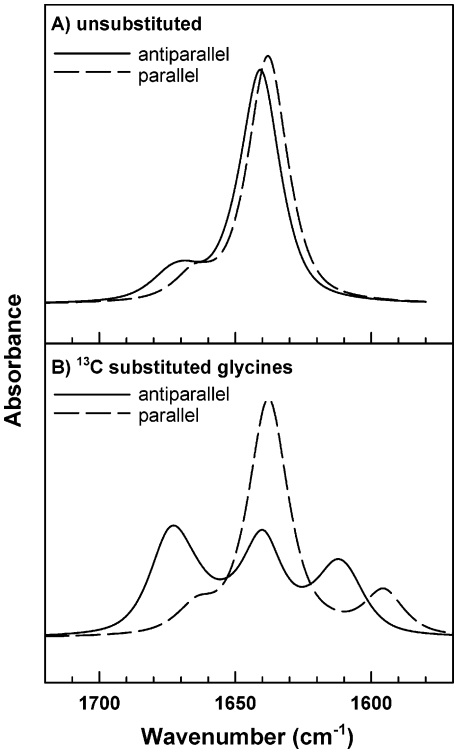



***Figure 6.*** Distinguishing between parallel and antiparallel $\beta$-sheet structure in a $\beta$-amyloid peptide using $^{13}$C substitution: (A) simulated Amide I contours for the unsubstituted peptide in idealized four stranded parallel and antiparallel $\beta$-sheet structures and (B) simulated Amide I contours for the same with six $^{13}$C labeled Gly residues.

produced conflicting results. Tycko and co-workers[35] synthesized a 40 residue $\beta$-amyloid peptide. They used multiple quantum solid-state NMR to show this peptide packs in a parallel fashion. Our simulations of four stranded forty residue idealized $\beta$-sheets yield almost identical Amide I contours for the two packings. However, if the forty residue peptide were synthesized with $^{13}$C labeled glycine (six residues), simulations of the substituted peptide in idealized antiparallel and parallel $\beta$-sheet structures predict an easily discernible difference in the Amide I contour as shown in Figure 6.

**General Comments and Future Possibilities.** As noted above, the excellent agreement between calculated and observed spectra suggests that the general nature of the interactions between oscillators is reasonably well determined. We do not claim complete understanding since some of the parameters utilized are empirical. In compensation, our algorithm is transparent, straightforward and easy to use. A general conclusion from the current study is that the widely used correlations between the Amide I contour and the secondary structure content of a globular protein are at best approximate. Interactions between oscillators are generally not uniform within a given secondary structure and they extend beyond that structure. Thus, the coupling effects that we (and others as cited above) have noted are complicated, widespread, and cannot be ignored in the analysis of the Amide I contour. These effects lead to a wide frequency range for modes of a given secondary structure, different extinction coefficients for different secondary structures and for the same type of secondary structure, and "end effects" in which frequency and intensity patterns for a given secondary structure are altered by the number of residues in the structure. All of these create difficulties with the use of IR for quantitative secondary structure determination in globular proteins.

The very factors which limit the use of IR for quantitative secondary structure determination, may eventually be turned to advantage for investigating tertiary structure changes in localized regions of a molecule in those cases where a new structure can be reasonably proposed. The various factors that have been shown to be important in the prediction of the contour are exquisitely sensitive to altered three-dimensional structure.

**Acknowledgment.** This work was supported by Grant GM 29864 from the US Public Health Service to R.M. We thank Ashleigh Allen and Michael DiPrima for their assistance with the FTIR measurements.

JA0400685

(35) Antzutkin, O. N.; Balbach, J. J.; Leapman, R. D.; Rizzo, N. W.; Reed, J.; Tycko, R. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13045−13050.