UV Resonance Raman-Selective Amide Vibrational Enhancement: Quantitative Methodology for Determining Protein Secondary Structure†

Zhenhuan Chi, X. G. Chen,‡ Janet S. W. Holtz, and Sanford A. Asher*

Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

Received May 16, 1997; Revised Manuscript Received October 29, 1997

ABSTRACT: We have directly determined the amide band resonance Raman spectra of the "average" pure α-helix, β-sheet, and unordered secondary structures by exciting within the amide $\pi\rightarrow\pi^*$ transitions at 206.5 nm. The Raman spectra are dominated by the amide bands of the peptide backbone. We have empirically determined the average pure α-helix, β-sheet, and unordered resonance Raman spectra from the amide resonance Raman spectra of 13 proteins with well-known X-ray crystal structures. We demonstrate that we can simultaneously utilize the amide I, II, and III bands and the $C_\alpha-H$ amide bending vibrations of these average secondary structure spectra to directly determine protein secondary structure. The UV Raman method appears to be complementary, and in some cases superior, to the existing methods, such as CD, VCD, and absorption spectroscopy. In addition, the spectra are immune to the light-scattering artifacts that plague CD, VCD, and IR absorption measurements. Thus, it will be possible to examine proteins in micelles and other scattering media.

Any fundamental understanding of protein and peptide structure, dynamics, and function requires methods to measure protein secondary structure. X-ray crystallography and two-dimensional nuclear magnetic resonance (2-D NMR) (*1*−*4*) are the most incisive protein structural probes because they can provide highly accurate three-dimensional structures. However, X-ray diffraction gives a static picture, and NMR gives only a limited amount of dynamical information (*2*, *4*). In addition, these techniques are labor intensive and require protein samples that differ from the normally desired dilute aqueous solution phase.

We report here a new methodology to determine dilute solution protein and peptide secondary structures, using UV resonance Raman spectroscopy (UVRRS) excited with a 206.5-nm CW laser directly into the amide $\pi\rightarrow\pi^*$ transitions of the peptide bonds (*5*−*11*). The resulting spectra are dominated by the amide vibrations, whose frequencies, Raman cross sections, and bandwidths depend sensitively on secondary structure. We have, for the first time, used chemometrics to directly and empirically determine the amide band resonance Raman spectra of the "average" α-helix, β-sheet, and unordered secondary structures (pure secondary structure Raman spectra, PSSRS) for a series of 13 proteins. We simultaneously utilize the amide I, II, and III bands and the $C_\alpha-H$ amide bending vibrations of these PSSRS to directly determine protein secondary structures.

This new methodology is complementary to secondary structural methods, such as CD, VCD, and IR absorption spectroscopy (*12*−*18*). We show here that the ability of this UV Raman methodology to predict secondary structure is, in some cases, superior to that of CD and that the approach has significant advantages compared to CD, VCD, and IR. For example, the Raman spectra are immune from the light-scattering artifacts that plague conventional absorption measurements (*17*). Therefore, UVRRS can be used to examine proteins in micelles and membranes (*19*).

We show in the accompanying paper that this UV Raman methodology can be used to study protein folding and denaturation. In this study of the acid denaturation of myoglobin, we combine the UV Raman spectra, exciting in the peptide backbone region, with spectra that result from excitation directly in the aromatic amino acid absorption bands. The amide bands characterize changes in secondary structure in the protein, such as α-helix melting, while changes in the aromatic amino acid bands suggest which helices undergo the conformational changes (*20*).

EXPERIMENTAL SECTION

Materials and Sample Preparation. Jack bean concanavalin A (Con A), bovine pancreas trypsin (trypsin), bovine serum albumin (BSA), bovine pancreas ribonuclease A (RNase A), horse heart cytochrome *c* (cyt *c*), human hemoglobin (Hb), horse skeletal muscle myoglobin (Mb), bovine pancreas α-chymomyosin (Chy), porcine pancreas elastase (Elas), chicken egg white lysozyme (Lyso), human erythrocyte carbonic anhydrase (Carb), porcine muscle triosephosphate isomerase (Trio), bovine pancreas insulin (Ins), tyrosine (Tyr), phenylalanine (Phe), tryptophan (Trp), histidine (His), arginine (Arg), cysteine (Cys), proline (Pro), methionine (Met), cystine, and the bromide salt of poly(L-lysine) (PLL, MW = 22 000) and the sodium salt of poly(L-glutamic acid) (PGA, MW = 26 500) were purchased from Sigma Chemical Co. (St. Louis, MO). Sodium per-

† We gratefully acknowledge support from NIH Grant R01GM30741-15.

* To whom correspondence should be addressed. Phone: (412) 624-8570. Fax: (412) 624-0588. E-mail: asher+@pitt.edu.

‡ Present address: Section on Metabolic Analysis and Mass Spectrometry, NICHD, National Institute of Health, 10 Center Dr., MSC 1580, Room 6C208, Bethesda, MD 20892.

S0006-2960(97)01160-4 CCC: $15.00
Published on Web 02/11/1998

chlorate was purchased from Aldrich Chemical Co. (Milwaukee, WI). Potassium phosphates (monobasic and dibasic) were purchased from EM Science (Gibbstown, NJ). All compounds were used without further purification.

The proteins were dissolved in 0.010 M aqueous pH 7.0 phosphate buffer. All protein and polypeptide concentrations were 0.50 mg/mL, except for Con A, which was 0.25 mg/mL. Each sample contained 0.15 M sodium perchlorate as an internal intensity standard (*21*).

The β-sheet form of PGA was prepared by heating the PGA solution at pH 4.3 to 95 °C for 5 days. The α-helical and unordered forms of PGA were prepared by maintaining the PGA solution at 25 °C at pH 4.3 and 10.0, respectively. The β-sheet form of PLL as prepared by heating the solution to 52 °C at pH 11.3 for 3 h. The α-helical and unordered forms of PLL were obtained by maintaining the PLL solution at 25 °C at pH 11.0 and 4.0, respectively *(22*, *23*). The "unordered" forms of PLL and PGA may possess some residual local chain order (*24*).

Raman Spectral Measurements. The Raman instrumentation is described in detail elsewhere (*25*). UV excitation at 206.5 nm (2.5 mW) was obtained from a Coherent Inc. CW intracavity frequency-doubled krypton ion laser (*5*). The laser beam was focused into a 1-mm-i.d. fused quartz capillary, through which a 20-mL sample solution was recirculated by using a peristaltic pump. The sampling optics used a 135° backscattering geometry. The total accumulation time was ca. 20 min.

The ca. 1640-cm^{-1} H_2O bending band was subtracted from the protein Raman spectra by using the measured relative intensity of the water to perchlorate bands. The protein Raman spectral baselines were defined to be linear and to span the intensities at 1100 and 1800 cm^{-1}. The protein Raman intensities were normalized to the 932-cm^{-1} perchlorate band intensity.

Spectral Modeling. To a good approximation, the protein amide band resonance Raman spectra can be modeled as resulting from the sum of the individual amide resonance Raman spectral bands contributed by each of the individual amide bonds within the different protein secondary structures. As shown below, this approximation works well for the spectral modeling. This approximation could become problematic in cases where the protein amide Raman bands result from the sum of contributions of Raman bands of both vibrations of individual amide bonds and from the coupled (delocalized) vibrations of a number of adjacent peptide bonds. This would require a more complex model. However, this case is likely to be rare and is unlikely to occur for any but the amide I vibrations, where coupling of adjacent carbonyl stretching is maximized by dipolar interactions between carbonyl stretches.

We modeled the observed 206.5-nm excited protein Raman spectra as a linear combination of the individual spectra of the different secondary structures, weighted by their relative compositions. The Raman bands enhanced in these spectra, excited in resonance with the amide $\pi \rightarrow \pi^*$ transitions, mainly derive from the protein backbone amide vibrations. However, spectral contributions are also observed from the aromatic amino acids ring vibrations (∼1610 cm^{-1}) (*9*, *10*, *26*−*28*) and from the side-chain groups (∼1450 cm^{-1}) (*29*). While we have numerically removed their contributions to the spectra, we do not have sufficient experience in modeling these spectral regions to be sure that any residual contributions will not confound our amide spectral modeling. Thus, we do not include the spectral region around 1610 and 1450 cm^{-1}.

We model the observed protein Raman spectra, in which aromatic side-chain contributions have been removed, as a linear combination of PSSRS, weighted by the secondary structure abundancies:

$$[D] = [R][C] \quad (1)$$

where the ($r \times p$) data matrix, [D], of the r Raman intensities, at the r Raman frequencies, for the p proteins, is determined by the product of the ($r \times c$) matrix, [R], of r normalized Raman spectral intensities at the r frequencies for the different c secondary structure conformations, with the ($c \times p$) protein conformational matrix, [C], containing the fractional composition of the different secondary structural forms, c, for the different proteins, p.

Although we expected the protein spectra to be dominated by the α-helix, β-sheet, and unordered conformations, we used factor analysis (FA) (*30*, *31*) to independently determine the minimum eigenvectors and eigenvalues which could fit the protein spectra within the spectral signal-to-noise ratios. We found that the data could be accurately modeled with only three eigenvectors, with the remaining calculated eigenvectors at the level of the spectral noise.

As indicated in detail in appendices I and II, we calculated the PSSRS using a least-squares deviation minimization method from the measured data matrix [D] and the matrix $[C]^{-1}$ derived from the known protein secondary structure composition matrix [C]. We can then utilize these PSSRS to calculate the secondary structure of unknown proteins.

RESULTS AND DISCUSSION

Protein Spectra. Figure 1 shows the 206.5-nm excited Raman spectra of Con A, trypsin, RNase A, cyt *c*, BSA, Hb, Mb, Chy, Elas, Lyso, Carb, Trio, and Ins. The amide I band occurs at ∼1655 cm^{-1}, while aromatic amino acid ring breathing bands occur in the ∼1610-cm^{-1} region. The amide II band frequency is fairly constant and occurs at ∼1555 cm^{-1}. The amide C_α−H symmetric bending band, which shows a variable intensity for different protein secondary structures, occurs at ∼1390 cm^{-1}. The amide III band, which shows a strong frequency and width dependence on secondary structure, occurs in the ∼1200−1300-cm^{-1} spectral region. The remaining band at ∼1450 cm^{-1}, which appears in the protein spectra with variable intensity, is neither an amide band nor an aromatic amino acid band. Its assignment is, at present, uncertain.

The amide I band consists of amide carbonyl C═O stretching, with smaller contributions of C−N stretching and N−H bending. The amide II and III bands involve significant C−N stretching, N−H bending, and C−C stretching. The C_α−H bending band involves C_α−H symmetric bending, with a variable amount of C−C_α stretching (*22*, *32*−*37*).

The Raman spectra demonstrate that the amide III bands of proteins with low α-helical content occur at the lowest frequencies, e.g., 1250 cm^{-1} for Con A (2% α-helix), while the amide III bands of proteins with high α-helical content occur at the highest frequencies, e.g., 1300 cm^{-1} for Mb

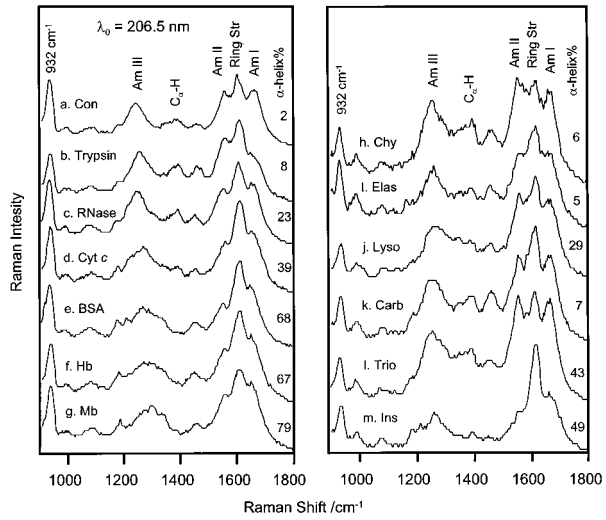

FIGURE 1: UV Raman spectra of aqueous proteins at 206.5-nm excitation. (a) Concanavalin A (Con), (b) trypsin, (c) ribonuclease A (RNase), (d) cytochrome *c* (cyt *c*), (e) albumin (BSA), (f) hemoglobin (Hb), (g) myoglobin (Mb), (h) chymomyosin (Chy), (i) elastase (Elas), (j) lysozyme (Lyso), (k) carbonic anhydrase (Carb), (l) triosephosphate isomerase (Trio), (m) insulin (Ins). The proteins (0.50 mg/mL) were dissolved in 0.10 M aqueous pH 7.0 phosphate buffer solutions containing $NaClO_4$ (0.15 M), except for Con, which was 0.25 mg/mL. The 932-cm^{-1} band derives from the ClO_4^- stretching band. The spectral resolution is 20 cm^{-1}.

(85% α-helix). Both the frequency and bandwidth of the amide III band increase with α-helical content. The intensity of the 1390-cm^{-1} amide $C_\alpha-H$ bending band decreases as the protein α-helical content increases. The band at 932 cm^{-1} results from the internal standard, perchlorate.

Amino Acid Side-Chain Raman Spectra. We quantitatively determined the magnitude of the spectral contributions of amino side chains to the 206.5-nm excited Raman spectra by measuring the Raman spectra of amino acids in water. These aqueous amino acid Raman cross sections will not differ dramatically from those of the side chains within hydrophobic or unique hydrogen-bonding protein environments.

The average band intensities observed for each amino acid band within a protein will be the product of the occurrence frequency of the amino acid and the Raman cross section of the band. Some side-chain Raman bands will show significant environmental frequency dependencies. Thus, the observed Raman band from a protein containing a number of these residues will derive from the sum of these frequency-shifted overlapping bands. This information can be used to examine side-chain conformation and environment. As shown below, however, only a few side chains, such as the Phe, Tyr, and Trp rings, show significant contributions to the 206.5-nm Raman spectra.

Figure 2 shows the 206.5-nm excited Raman spectra of aqueous solutions of Phe, Tyr, Trp, His (pH = 7.0 and 2.0), Arg, Pro, cysteine, Met, and cystine, while Table 1 lists their differential Raman cross sections. Phe, Tyr, and Trp all give rise to intense Raman bands in the 1610-cm^{-1} spectral region, with cross sections of 0.42, 0.63 and 0.76 barn $molecule^{-1}$ sr^{-1}, respectively. The other Raman bands are generally smaller, except for Trp, which shows 1550−1590-cm^{-1} Raman bands with cross sections of ∼0.50 barn $molecule^{-1}$ sr^{-1}.

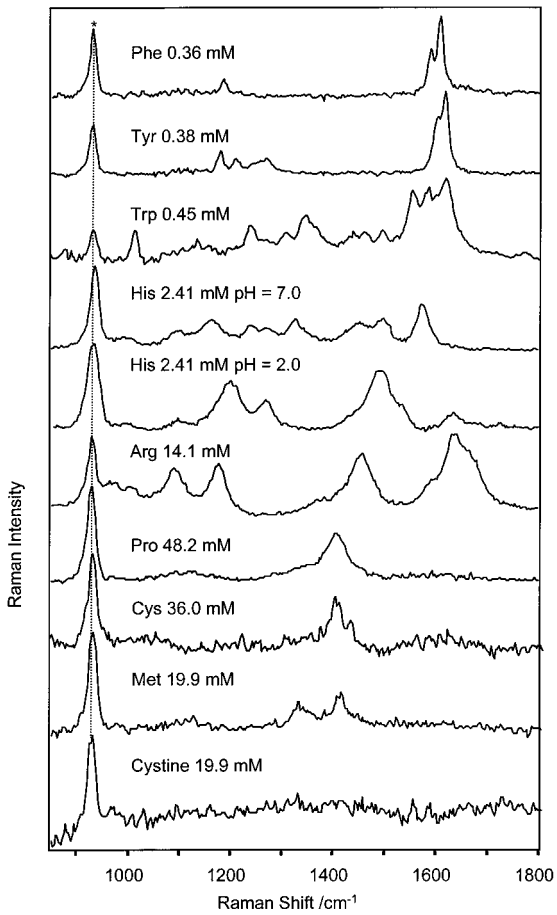

FIGURE 2: 206.5-nm excited Raman spectra of phenylalanine (0.36 mM), tyrosine (0.38 mM), tryptophan (0.45 mM), histidine (2.41 mM, pH = 7.0), histidine (2.41 mM, pH = 2.0), arginine (14.1 mM), proline (48.2 mM), cysteine (36.0 mM), methionine (19.9 mM), and cystine (19.9 mM) in aqueous solution containing 0.15 M $NaClO_4$. The peaks indicated by asterisks are from the ClO_4^- 932-cm^{-1} band.

Table 1 displays the average relative Raman intensities ($RI = f\sigma/\sigma_{av}$) expected in 206.5-nm protein Raman spectra of the most intense amino acid side chains. We quantitated the relative contribution of each of these bands to the 206.5-nm Raman spectrum by normalizing its cross section to that of the amide II band, which, as shown in Figure 1, shows the least spectral variation between proteins. *f*, the average occurrence frequency for each amino acid side chain in proteins of known sequence (*38*), is also tabulated in Table 1. σ represents the differential Raman cross sections of the amino acid side-chain bands in aqueous solution, and σ_{av} is the average Raman cross section of the amide II band averaged over the α-helix, β-sheet, and unordered conformations, as calculated from their PSSRS (σ_{av} = 50 mbarn $molecule^{-1}$ sr^{-1} per amide linkage at 206.5 nm).

The average amino acid occurrence frequency is 5.0%, while the aromatic amino acids have lower actual occurrence frequencies of between 1.3% and 3.9%. It is presently straightforward to detect the presence of side-chain Raman bands with RI exceeding 2% (i.e., for a band with an intensity of 2% of that of the amide II band), given our 206.5-nm Raman spectral signal-to-noise ratios for 15-min spectral accumulations. Thus, Tyr, Trp, and Phe will show significant spectral contributions; the RI values of their ∼1610-cm^{-1} ring breathing bands are ∼30%. However, these Tyr and

Table 1: 206.5-nm Amino Acid Differential Raman Cross Sections and Their Predicted RI

Raman band (cm^{-1})	σ^b (mbarn molecule^{-1} sr^{-1})	RI^c (%)
	phenylalanine (3.9^a)	
1183	79	6.2
1586	210	16
1606	420	33
	tyrosine (3.2^a)	
1178	170	11
1209	100	6.4
1601	400	26
1617	630	40
	tryptophan (1.3^a)	
1013	320	8.4
1238	180	4.6
1348	270	7.1
1556	530	14
1584	440	11
1601	170	4.5
1620	760	20
	histidine (2.2^a), pH = 7.0	
1162	19	0.9
1327	17	0.7
1452	18	0.9
1575	33	1.5
	histidine (2.2^a), pH = 2.0	
1194	48	2.1
1265	21	0.9
1486	51	2.2
1633	10	0.5
	arginine (5.7^a)	
1091	9	1.0
1176	12	1.4
1454	13	1.5
1633	16	1.9
1669	11	1.2
	proline (5.1^a)	
1335	0.3	0.03
1410	1.3	0.14
	cysteine (1.7^a)	
1402	2.4	0.09
1431	0.95	0.03
	methionine (2.4^a)	
1410	2.4	0.12
	cystine (<0.85^a)	
	<0.4	<0.01

^{a}The numbers in parentheses are frequency of occurrence of amino acid residue in primary structures of 1021 unrelated proteins of known sequence (*38*). ^{b} σ is the differential Raman cross sections of these amino acids in aqueous solution. ^{c} RI is relative intensity: RI = $f\sigma/\sigma_{av}$, where σ_{av} is the average Raman cross section of the amide II band in these PSSRS spectrum.

Trp bands are best studied with 229-nm excitation, where they show significantly increased cross sections, compared to those of other chromophores in proteins.

Of all of the other amino acid side chains, only protonated His has an RI above 2% for 206.5-nm excitation. Protonated His (pH = 2.0) shows two dominating bands at 1194 and 1486 cm^{-1}, with Raman cross sections of ~50 mbarn molecule^{-1} sr^{-1}. If all the His residues in a protein were protonated, these bands would have RI values slightly above 2%. In contrast, neutral His shows bands at 1575, 1162, 1327, and 1452 cm^{-1} with RI < 2%. His residues should show larger cross sections with excitation at slightly longer wavelength.

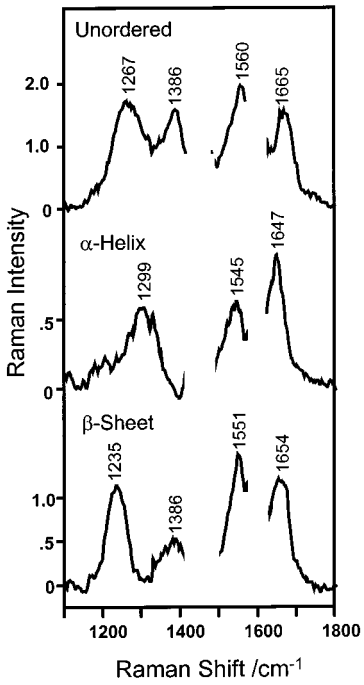

FIGURE 3: Calculated pure secondary structure Raman spectra (PSSRS) obtained from aqueous solution 206.5-nm protein Raman spectra. The regions around 1610 and 1450 cm^{-1} are excluded due to interference from aromatic amino acid ring vibrations and side-chain vibrations.

Table 2: Frequencies and Raman Cross Sections of Amide Bands in Pure Secondary Structure Raman Component Spectra (PSSRS), Excited at 206.5 nm

	unordered			α-helix			β-sheet		
	freq^a	σ^b	width^a	freq^a	σ^b	width^a	freq^a	σ^b	width^a
amide I	1665	56	60	1647	32	53	1654	48	66
amide II	1560	71	91	1545	22	64	1551	57	66
amide III	1267	65	93	1299	24	87	1235	53	62
C_α−H^c	1386	57	66				1386	23	86

^{a} Units, cm^{-1}. ^{b} Units, mbarn molecule^{-1} sr^{-1}. ^{c} Bending band.

Table 1 shows that Arg, Pro, Cys, Met, and cystine have Raman band cross sections below 16 mbarn molecule^{-1} sr^{-1}, with resulting RI < 2%; thus, they will negligibly contribute to the average 206.5-nm protein Raman spectrum.

Quantitative Determination of the Protein Secondary Structure. As discussed in the Experimental Section, we calculated the PSSRS of the α-helix, β-sheet, and unordered conformations from the measured protein UV Raman spectra (Figure 1) and their known protein secondary structures. Figure 3 shows the calculated PSSRS spectra, while Table 2 lists their calculated frequencies, differential Raman cross sections, and bandwidths. As discussed above, the underlying assumption (tested by factor analysis) is that these three secondary structures dominate the 206.5-nm excited amide UV Raman spectra. The PSSRS spectra are thus the calculated "average" Raman spectra of each secondary structure averaged over the 13 proteins studied, and we may expect some variation from that observed for a specific peptide in a particular secondary structure.

PGA and PLL Spectra. We can compare the PSSRS spectra to spectra of PLL and PGA samples which were treated to force them to predominantly assume the unordered,

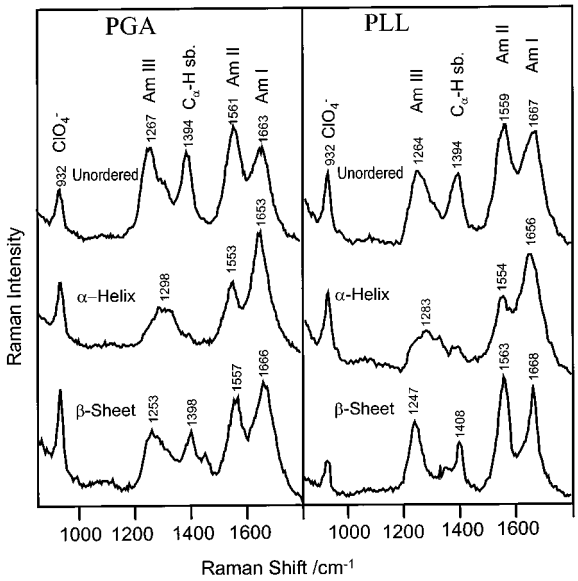

FIGURE 4: UV Raman spectra of 1.9 μM poly(L-glutamic acid) (PGA) and 2.3 μM poly(L-lysine) (PLL) in water at 206.5-nm excitation, containing $NaClO_4$ (0.15 M) as an internal standard. The 932-cm^{-1} band derives from ClO_4^- stretching. PGA unordered form: pH = 10.0, 25 °C. α-Helix form: pH = 4.3, 25 °C. β-Sheet form: pH = 4.3. PLL unordered form: pH = 4.0, 25 °C. α-Helix form: pH = 11.0, 25 °C. β-Sheet form: pH =11.3. See text for details.

Table 3: Frequencies and Raman Cross Section of Amide Bands in Pure Secondary Structural Raman Spectra (PSSRS), Excited at 206.5 nm

	unordered			α-helix			β-sheet		
	freq[a]	σ[b]	width[a]	freq[a]	σ[b]	width[a]	freq[a]	σ[b]	width[a]
amide I	1665	56	60	1647	32	53	1654	48	66
amide II	1560	71	91	1545	22	64	1551	57	66
amide III	1267	65	93	1299	24	87	1235	53	62
C_α−H[c]	1386	57	66				1386	23	86

[a] Units, cm^{-1}. [b] Units, mbarn $molecule^{-1}$ sr^{-1}. [c] Bending band.

α-helix, and β-sheet forms, shown in Figure 4. As observed for the protein spectra, these polypeptide Raman spectra are dominated by the intense amide I (1650 cm^{-1}), amide II (1554 cm^{-1}), and amide III (between 1250 and 1290 cm^{-1}) bands and the C_α−H bending band (1398 cm^{-1}). Table 3 lists the band frequencies, bandwidths, and Raman cross sections of the PGA and PLL amide bands.

The calculated unordered PSSRS spectrum is essentially identical to the unordered PGA and PLL spectra; the largest intensity occurs for the amide II band, with similar smaller intensities for the amide I (1665 cm^{-1}) and III (1265 cm^{-1}) bands and the 1394-cm^{-1} C_α−H bending bands. The unordered PSSRS amide III and the C_α−H bending bands are broader than those that in PGA and PLL, presumably because of the distribution of geometries which exist in this ill-defined secondary structure conformation. The amide I band seems to be less sensitive to this distribution, since it is somewhat narrower than those of PLL and PGA. The C_α−H bending band occurs ~8 cm^{-1} lower in frequency in the PSSRS compared to that in PGA and PLL. The frequency alterations presumably result from the side-chain dependence of the Raman frequencies.

The α-helix PSSRS spectrum is similar to the PGA and PLL α-helix Raman spectra. These spectra are dominated by the amide I band (1655 cm^{-1}), while the amide II bands (1554 cm^{-1}) are relatively weak. No bands appear in the ca. 1400-cm^{-1} C_α−H bending band region. The amide III band at ~1290 cm^{-1} is very broad and appears to be composed of overlapping subbands (*39*, *40*). The amide III band of the α-helical PSSRS occurs at a frequency similar to that of PGA, which is, however, ca. 15 cm^{-1} lower in frequency than that of PLL. The PSSRS amide I and II bands appear to be downshifted by ~10 cm^{-1} compared to those of PGA and PLL.

The β-sheet PSSRS spectrum is similar to that of β-sheet PLL, with similar relative intensities for the amide I (1667 cm^{-1}), II (1560 cm^{-1}), and III (1250 cm^{-1}) bands and the C_α−H bending band (1400 cm^{-1}). However, the amide I and C_α−H bending PSSRS bands are broader than those of PLL. This presumably results from the smaller distribution of β-sheet geometries in the PLL oligopeptide.

In contrast, the β-sheet PGA Raman spectrum appears to be dramatically different from that of either PLL or the PSSRS. These spectral differences probably result from incomplete conversion of the PGA into the β-sheet conformation (vide infra).

Krimm and Bandekar's normal-mode calculations (*37*) suggest that the conformational dependence of the amide band frequencies results from the dependence of the amide bond force constants on the polypeptide geometry and the differences in hydrogen bonding of the different secondary structure forms. These force constant alterations change the normal-mode compositions, which results in amide band frequency changes. For example, the α-helix geometry difference and its weaker hydrogen bonding, compared to that in the β-sheet structure, shifts the calculated α-helix amide I frequency from 1655 to 1674 cm^{-1} for the β-sheet (*37*, *39*, *40*).

Figure 3 and Table 2 indicate a significant amide band Raman cross section conformational dependence. For example, the α-helix amide II and III band Raman cross sections are the lowest among these three secondary structures due to the decreased oscillator strength of the ~190-nm amide $\pi \rightarrow \pi^*$ transition in the α-helix structure, since the resonance Raman cross section is proportional to the square of the oscillator strength of the resonant transition.

We can utilize the PSSRS to calculate the secondary structural composition of the PGA and PLL samples. This calculation determines the best fit of the linear sum of PSSRS spectra, and the composition is given by the relative contribution of each of the PSSRS spectra to the calculated spectrum. Figure 5 compares the calculated and measured PGA and PLL Raman spectra and shows the residuals between them. The figures also lists the calculated percentage compositions.

The PGA and PLL unordered samples are calculated to be 90% and 78% unordered, respectively, with the remainder β-sheet form. The residual features in the amide II and III and C_α−H bending regions most likely derive from a shift in the center frequency of these bands between the PSSRS and the PGA and PLL oligopeptide spectra. This probably signals the fact that the oligopeptide band unordered secondary structure differs somewhat from that of the average protein unordered conformation.

For the α-helix forms of PGA and PLL, we calculate 88% and 72% α-helix conformations, respectively. We observe

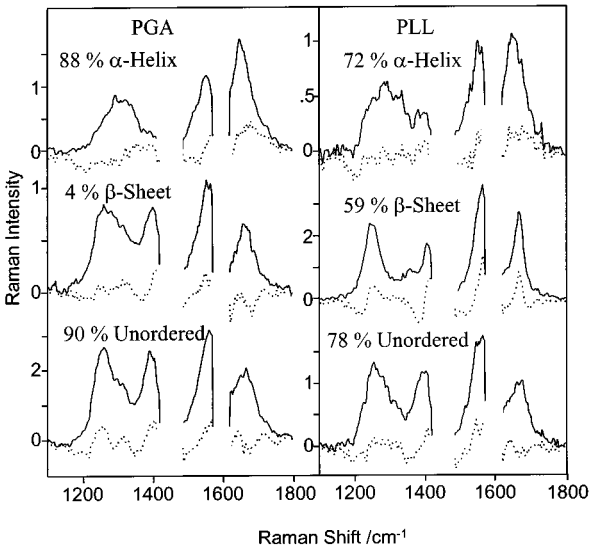

FIGURE 5: Measured PGA and PLL Raman spectra and residuals calculated from the best fit using the PSSRS. The calculated percent relative abundances of the dominant secondary structural forms are listed.

broad troughs in the residuals on the low-frequency side of the amide III band as well as a broad peak for the amide I band. These residuals probably reflect differences in structure between an "average protein" α-helix structure and that which occurs in these oligopeptides.

The best fits for the PGA and PLL β-sheet samples give nonphysical negative abundance values. If the composition value is constrained to be positive, we calculate 4% and 59% β-sheet structures, respectively. For PGA, we were already suspicious that β-sheet structure did not dominate.

Calculation of Protein Secondary Structures. We utilized the PSSRS to calculate the fractional secondary structure abundances of the proteins studied and also examined the precision of fitting of the resonance Raman spectra. We subtracted bands from Trp, Tyr, and Phe from the observed Raman spectra on the basis of the known composition. We also subtracted the Raman spectrum of water on the basis of the relative intensities of the water band compared to that of the internal standard perchlorate intensity. Due to lack of confidence in reliable subtraction of the aromatic amino acid bands at ~1610 cm^{-1}, we neglected the spectral region around 1610 cm^{-1}. In addition, we neglected the spectral region around the 1450-cm^{-1} band. We used least-squares criteria to determine the best secondary structure composition for each protein. Figure 6 shows the measured protein spectra used for spectral modeling and the residuals that result from subtracting the calculated from the measured spectra.

Except for cyt *c*, the residuals are everywhere less than 10% of the measured Raman band intensities. In addition, no recognizable peaks or troughs are evident in the residuals. Table 4 compares this calculated protein secondary structure composition to that determined by X-ray crystallography. The difference in fractional abundances between X-ray crystallography and UV Raman results is listed as Δ_{R-X}. The correlation coefficient (r) is also listed in Table 4. The correlation coefficients are $r = 0.96$ for α-helix, $r = 0.92$ for β-sheet, $r = 0.90$ for unordered, and $r = 0.93$ overall. Thus, the fractional abundances calculated from the PSSRS UV Raman spectra are close to those determined by X-ray crystallography (*15*, *41*−*44*).

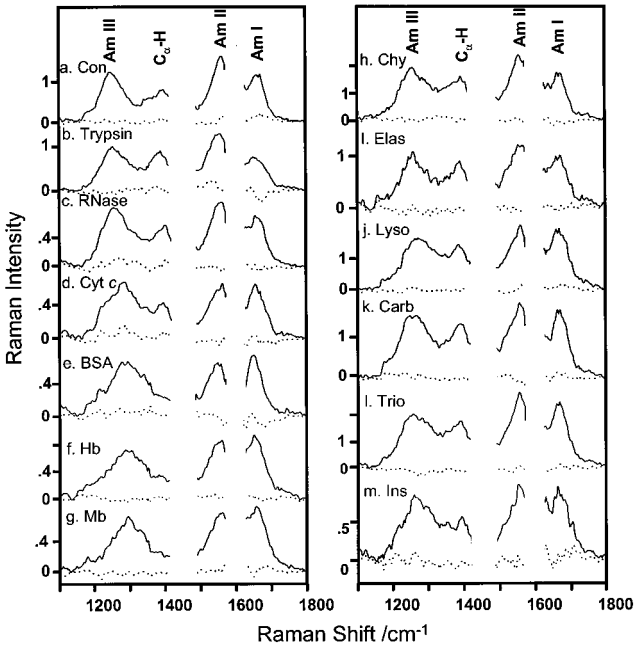

FIGURE 6: 206.5-nm excited measured protein Raman spectra in pH 7.0 water used for modeling and their residuals.

The model we are using is very robust for determining secondary structure composition. For example, Figure 7 shows the impact on the calculated residuals if Mb composition were varied by changing the relative α-helix and β-sheet compositions compared to that calculated for the native protein. As the α-helix composition decreases by 15% and the β-sheet composition increases by 15%, χ^2 over the spectral region 1100−1800 cm^{-1} increases by 2-fold. The initial χ^2 measures the deviation of the PSSRS fit from the measured spectrum and includes a lack of complete modeling (such as specific β-turn spectra, for example) and the contribution of random spectral noise. The increased χ^2 derives directly from a lack of fitting and the appearance of well-defined features, such as the troughs around 1240, 1400, and 1560 cm^{-1} and the peaks around 1310 cm^{-1}. The amide I band region shows the least change in residuals because of the similarity of amide I band spectra for the α-helix and β-sheet forms.

The fact that the fits are excellent for all of the proteins in Figure 6, and that the residuals are small and do not show distinct features, indicates that our model encompasses the major secondary structural elements of these proteins and that 206.5-nm excitation selectively enhances only amide bands, except in the spectral regions which were numerically excised.

The largest residuals occur (Figure 6) for the small, "topologically challenged" protein, cyt *c* (*45*), which has no β-sheet structure but has numerous distorted conformations such as β-turns. Cyt *c* shows the largest deviations in the calculated secondary structure abundancies (Table 4). The two small residual peaks at 1240 and 1300 cm^{-1} probably derive from β-turn structures, for which we have not yet determined the PSSRS; β-turn amide III bands are known to occur at 1240 cm^{-1} and/or at higher frequencies of 1265−1290 cm^{-1} (*46*−*48*). The inability to incorporate cyt *c* β-turn structures would cause the calculation to incorrectly assign β-sheet and unordered conformations. We will experimentally characterize the β-turn Raman spectra soon.

Table 4: Fractional Abundances of Protein Secondary Structure Determined by X-ray Crystallography and UV Raman Spectra

	α-helix (%)			β-sheet (%)			unordered (%)		
	X^a	R^b	$\Delta_{R-X}{}^c$	X^a	R^b	$\Delta_{R-X}{}^c$	X^a	R^b	$\Delta_{R-X}{}^c$
concanavalin	2	13	+11	65	63	−2	33	24	−9
trypsin	8	4	−4	32	44	+12	60	52	−8
RNase A	23	20	−3	40	45	+5	37	35	−2
cytochrome	39	51	+12	0	13	+13	61	36	−25
BSA	68	77	+9	18	7	−11	14	16	−2
hemoglobin	67	75	+8	0	6	+6	33	19	−14
myoglobin	79	81	+2	0	1	+1	21	18	−3
chymotrypsin	6	6	0	33	38	+5	61	57	−4
elastase	5	10	+5	34	27	−7	61	62	+1
lysozyme	29	36	+7	8	2	−6	63	62	−1
carbonic anhy	7	3	−4	27	28	+1	66	69	+3
triosephosphate	43	26	−17	17	18	+1	40	55	+15
insulin	51	45	−6	24	16	−8	25	39	+14
PGA α-helix		88			2			10	
PGA β-sheet		4			4			92	
PGA random		0			10			90	
PLL α-helix		72			9			20	
PLL β-sheet		0			59			41	
PLL random		0			22			78	
r^d		0.96			0.92		0.90	0.93 (overall)	
σ^e		0.084			0.075		0.082	0.082 (overall)	

[a] Fractional abundances of protein secondary structure determined by X-ray crystallography. [b] Fractional abundances of protein secondary structure determined by UV Raman spectra. [c] Difference of fraction abundances between UV Raman spectra and X-ray results. [d] Correlation coefficient between the two fraction abundances determined by UV Raman and X-ray results. [e] Standard deviation between the two fraction abundances determined by UV Raman and X-ray results.

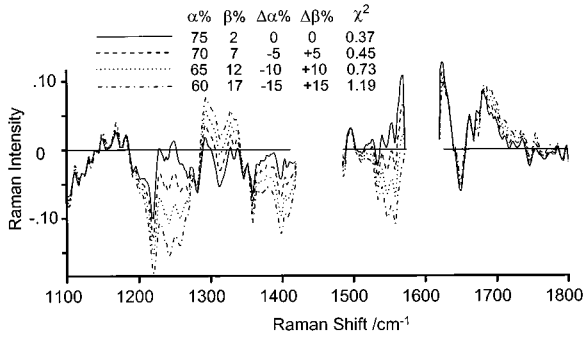

FIGURE 7: Calculated myoglobin residuals at different calculated relative abundances of α-helix and β-sheet structures. See text for details.

Comparison of IR, CD, and UV Raman Secondary Structure Determinations. The UV Raman secondary structure determination methodology described here is complementary to CD, VCD, and IR absorption spectroscopy (*12*−*18*). The information content should be similar to that of CD since the total number of spectral features is approximately the same. The spectral signal-to-noise appears comparable. However, the mechanisms of the two spectroscopies differ dramatically, and the sensitivity of each technique to different secondary structure motifs will differ.

For example, CD measures the chirality associated with the coupling of the electronic transitions of the amide backbone. The largest spectral signature derives from α-helix conformations, with smaller contributions from the unordered and β-sheet conformations (*12*−*17*). The α-helix CD spectrum derives from the coupling of the transition moments of adjacent amide linkages and is helical length dependent. Thus, we monitor a collective phenomenon which is not directly additive over the number of amide bonds involved in the α-helix. The CD spectrum of a short α-helical segment is different from that of a long helix. In contrast, the resonance Raman spectrum involves more local phenomena, where the band frequencies result from localized vibrations which couple to the local electronic transitions within amide segments. The contribution of unordered and β-sheet conformations to the 206.5-nm UV Raman spectra are similar within a factor of about 3; the α-helix contribution is smallest due to the hypochromism of the electronic absorption band.

It should be appreciated, however, that Raman spectra can also be subject to collective phenomena involving adjacent amides; for example, excitonic interactions that alter molar absorptivities impact the Raman cross sections (*22*). However, the Raman band frequencies and band shapes are unaffected, unless the vibrational modes become coupled in phonon-like states, which is unlikely in normal proteins.

The locality and the linearity of the resonance Raman spectroscopy are similar to those for the VCD and IR vibrational spectroscopies. A major advantage of the UV Raman measurements is that we have higher information content compared to VCD and IR because we have a larger number of observed spectral bands.

The UV Raman spectra show similar vibrational resolution, as do the VCD and IR methods. Thus, spectral interference is less common. This becomes a significant problem in CD, where the overlapping CD spectra of the aromatic amino acids can confound amide CD conformational studies. The Raman spectra do not show interference from water, which makes VCD and IR measurements of aqueous protein samples difficult and often compels the use of D_2O as a solvent. The UV Raman spectra are typically measured for relatively dilute protein solutions (∼0.5 mg/mL), but the concentrations could be easily decreased by an order of magnitude for precious samples.

Our ability to quantitate secondary structure compares very well to that of CD (*49*−*61*), as shown in Table 5, where we have essentially identical correlation coefficients for α-helical

Table 5: Comparison of UV Raman Method and Various CD Methods for Estimating Protein Secondary Structure: Correlation Coefficient (r) between the UV Raman, CD, and X-ray Results

method	α-helix	β-sheet	unordered
UV Raman	0.96	0.92	0.90
PG[a]	0.96	0.94	0.49
HJ[b]	0.95	0.66	0.72
VS[c]	0.97	0.76	0.86
SK[d]	0.96	0.75	0.60
LL[e]	0.96	0.85	0.79
CCA[f]	0.93	0.71	0.73
SC[g]	0.96	0.84	0.74
BPN[h]	0.94	0.63	0.88
SOM[i]	0.85	0.62	0.66
GF[j]	0.92	0.72	
Chen et al.[k]	0.97	0.58	
Chang et al.[l]	0.92	0.83	
Brahms[m]	0.99	0.98	
Bolotina et al.[n]	0.83	0.34	

[a] PG, Provencher and Glockner (*56*). [b] HJ, Hennessey, and Johnson (*58*). [c] VS, Manavalan and Johnson (*55*). [d] SK, Shubin et al. (*54*). [e] LL, van Stokkum et al. (*53*). [f] CCA, Perzel et al. (*52*). [g] SC, Sreermana and Woody (*49*). [h] BPN, Bohm et al. (*51*). [i] SOM Andrade et al. (*50*). [j] GF, Greenfield and Fasman (*61*). [k] Chen et al. (*60*). [l] Chang et al. (*43*) [m] Brahms and Brahms (*59*). [n] Bolotina et al. (*57*).

structure preditions and superior correlation coefficients for β-sheet and unordered conformations.

There are numerous additional advantages for the use of UV Raman to quantitatively determine protein secondary structure. The inherent vibrational resolution permits us to potentially resolve conformation at particular amide linkages within a protein. For example, selective amide ^{13}C, ^{15}N, ^{18}O, and N−D isotopic substitution would shift the amide bands to permit the determination of their frequencies and intensities of the amide linkage of interest. In addition, the technique can be used for membrane proteins and proteins in micelles where the light scattering confounds CD measurements (*19*). The UV Raman spectroscopy can also be easily used in flow systems to study kinetics in the submillisecond time domain. Furthermore, nanosecond laser excitation permits very fast kinetic spectral measurements.

CONCLUSIONS

We have directly determined the average amide band resonance Raman spectra of the α-helix, β-sheet, and unordered secondary structures and have demonstrated a new methodology to determine protein secondary structure. This method simultaneously utilizes the amide I, II, and III bands and the C_α−H amide bending vibration. The sensitivity should be superior to standard methods such as electronic CD, since the different secondary structural spectra have richer spectral details than do the CD spectra. Furthermore, the spectra are immune from the light-scattering artifacts that plague CD measurements (*17*). It is possible to examine proteins in micelles and other scattering media (*19*).

The spectral richness of UVRR is comparable to that of VCD (*14*, *15*), but the Raman spectra can be measured at much lower concentrations with much higher S/N ratios. In addition, there is no interference by the water solvent which limits and complicates IR and VCD spectroscopy.

A unique feature of the UV Raman spectra is that it is possible to selectively examine conformation at individual peptide bonds by isotopically substituting one amide bond by ^{13}C, ^{18}O, and/or ^{15}N. The isotopically substituted peptide linkage bands shift away from those of the other peptide linkages.

ACKNOWLEDGMENT

We thank Drs. Pushing Li, Petr Pancoska, and Richard W. Bormett and Profs. Xingfu Chen and Paul W. Rasmussen for helpful discussions.

APPENDIX I: CALCULATION OF PSSRS SPECTRA BY LEAST-SQUARES METHOD

Suppose $\mathbf{R}_\alpha$, $\mathbf{R}_\beta$, and $\mathbf{R}_u$ are the three predicted component spectra for the α-helix, β-sheet, and unordered conformations, respectively, $\mathbf{D}_1$, $\mathbf{D}_2$, ..., $\mathbf{D}_{13}$ are 13 protein spectra, and C_{kn} is the fractional composition for the kth component in the nth protein. In the calculation reported here, the region from 1100 to 1800 cm^{-1} is utilized, with a data interval of 2 cm^{-1}. Thus each spectrum has 351 points. Matrix $[X]^T$ is the transpose of matrix [X]. Each spectrum has i data points at different frequencies. A bold letter represents a vector. The predicted protein spectra (D'_{ij}) at the ith wavenumber points will be

$$D'_{ij} = R_{i\alpha}C_{\alpha j} + R_{i\beta}C_{\beta j} + R_{iu}C_{uj} \qquad j = 1, 2, ..., 13 \quad (1)$$

where i is the ith wavenumber. The difference between the measured protein spectrum and its predicted spectrum at the ith wavenumber is given as ΔD_{ij}:

$$\Delta D_{ij} = D_{ij} - D'_{ij} = D_{ij} - (R_{i\alpha}C_{\alpha j} + R_{i\beta}C_{\beta j} + R_{iu}C_{uj}) \quad (2)$$

To find the best $\mathbf{R}_\alpha$, $\mathbf{R}_\beta$, and $\mathbf{R}_u$ ($R_{i\alpha}$, $R_{i\beta}$, R_{iu}, $i = 1, 2, ...$), the deviation between the measured spectra and the calculated spectra is minimized by setting the sum of the derivatives of the squares of the difference, given by eq 2, throughout 13 proteins and all spectral regions (all i data points are included) equal to zero.

The derivative of the square of the difference with respect to $R_{i\alpha}$, $R_{i\beta}$, and R_{iu}, is

$$\frac{\partial}{\partial R_{ik}}\sum_i\sum_j (D_{ij} - R_{i\alpha}C_{\alpha j} - R_{i\beta}C_{\beta j} - R_{iu}C_{uj})^2 = 0$$

$$k = \alpha, \beta, u; \quad j = 1, 2, ..., 13; \quad i = 1, 2, ..., 351 \quad (3)$$

For the derivative with respect to $R_{i\alpha}$, we will see

$$\sum_j (D_{ij} - R_{i\alpha}C_{\alpha j} - R_{i\beta}C_{\beta j} - R_{iu}C_{uj})C_{\alpha j} = 0 \quad (4)$$

$$R_{i\alpha}\sum_j C_{\alpha j}^2 + R_{i\beta}\sum_j C_{\alpha j}C_{\beta j} + R_{iu}\sum_j C_{\alpha j}C_{uj} = \sum_j C_{\alpha j}D_{ij} \quad (5)$$

Then, repeating the calculation for the derivative with respect

to $R_{i\beta}$, R_{iR}, we will see

$$R_{i\alpha}\sum_j C_{\alpha j}C_{\beta j} + R_{i\beta}\sum_j C_{\beta j}{}^2 + R_{iu}\sum_j C_{\beta j}C_{uj} = \sum_j C_{\beta j}D_{ij} \quad (6)$$

$$R_{i\alpha}\sum_j C_{\alpha j}C_{Rj} + R_{i\beta}\sum_j C_{\beta j}C_{Rj} + R_{iu}\sum_j C_{uj}{}^2 = \sum_j C_{uj}D_{ij} \quad (7)$$

Let

$$[\mathrm{R}_i] = (R_{i\alpha}, R_{i\beta}, R_{iu}) \qquad i = 1, 2, ..., 351 \quad (8)$$

$$[\mathrm{B}] = (\sum C_{\alpha j}D_{ij}, \sum C_{\beta j}D_{ij}, \sum C_{uj}D_{ij}) \qquad \sum \text{ over } j \quad (9)$$

$$[\mathrm{A}] = \begin{bmatrix} \sum_j C_{\alpha j}{}^2 & \sum_j C_{\alpha j}C_{\beta j} & \sum_j C_{\alpha j}C_{uj} \\ \sum_j C_{\alpha j}C_{\beta j} & \sum_j C_{\beta j}{}^2 & \sum_j C_{\beta j}C_{uj} \\ \sum_j C_{\alpha j}C_{uj} & \sum_j C_{\beta j}C_{uj} & \sum_j C_{uj}{}^2 \end{bmatrix} \quad (10)$$

Thus, combining eqs 5−7 with eqs 8−10, we obtain

$$[\mathrm{R}_i][\mathrm{A}] = [\mathrm{B}] \quad (11)$$

$$[\mathbf{C}_j]^{\mathrm{T}} = (C_{\alpha j}, C_{\beta j}, C_{uj}) \qquad j = 1, 2, ..., 13 \quad (12)$$

$\mathbf{C}_j$ is a vector. Combining the 13 protein component vectors into a composition matrix [C],

$$[\mathrm{C}] = [\mathbf{C}_1, \mathbf{C}_2, ..., \mathbf{C}_{13}] \quad (13)$$

To compare eqs 10, 12, and 13,

$$[\mathrm{A}] = [\mathrm{C}][\mathrm{C}]^{\mathrm{T}} \quad (14)$$

Similarly, we let

$$[\mathrm{D}_i] = (D_{i1}, D_{i2}, ..., D_{ij}) \quad (15)$$

and combine eqs 9, 12, 13 and 14:

$$[\mathrm{B}] = [\mathrm{D}_i][\mathrm{C}]^{\mathrm{T}} \quad (16)$$

Thus, upon examining eqs 13 and 15,

$$[\mathrm{R}_i][\mathrm{C}_j][\mathrm{C}_j]^{\mathrm{T}} = [\mathrm{D}_i][\mathrm{C}]^{\mathrm{T}} \quad (17)$$

Multiplying by $\{[\mathrm{C}_j][\mathrm{C}_j]^{\mathrm{T}}\}^{-1}$ on both sides,

$$[\mathrm{R}_i] = [\mathrm{B}][\mathrm{A}]^{-1} = [\mathrm{D}_i][\mathrm{C}]^{\mathrm{T}}\{[\mathrm{C}][\mathrm{C}]^{\mathrm{T}}\}^{-1} \quad (18)$$

$[\mathrm{C}]^{\mathrm{T}}\{[\mathrm{C}]\ [\mathrm{C}\ ]^{\mathrm{T}}\}^{-1}$ is the pseudo-inverse of the component matrix [C].

For the component spectra matrix [R] [351 × 3],

$$[\mathrm{R}]^{\mathrm{T}} = [R_1, R_2, ..., R_i] \qquad i = 1, 2, ..., 351 \quad (19)$$

Similarly the data matrix [D] [351 × 13] can be obtained:

$$[\mathrm{D}]^{\mathrm{T}} = [D_1, D_2, ..., D_i] \qquad i = 1, 2, ..., 351 \quad (20)$$

Furthermore, upon examining eqs 18−20, we can see

$$[\mathrm{R}] = [\mathrm{D}][\mathrm{C}]^{\mathrm{T}}\{[\mathrm{C}][\mathrm{C}\ ]^{\mathrm{T}}\}^{-1} \quad (21)$$

Thus PSSRS, matrix [R] can be obtained from the matrix transformation given by eq 21.

APPENDIX II: CALCULATION OF PROTEIN FRACTIONAL COMPOSITION OF SECONDARY STRUCTURE USING LEAST-SQUARES METHOD

Suppose $\mathbf{R}_\alpha$, $\mathbf{R}_\beta$, and $\mathbf{R}_u$ are the PSSRS Raman spectra of the α-helix, β-sheet, and unordered forms, respectively. $\mathbf{D}_1$ is the measured first protein Raman spectrum. $C_{\alpha 1}$, $C_{\beta 1}$, and C_{u1} give the α-helix, β-sheet, and unordered compositions of protein 1. Then $\mathbf{R}_\alpha$, $\mathbf{R}_\beta$, $\mathbf{R}_u$, and $\mathbf{D}_1$ are expressed as below:

$$\mathbf{R}_\alpha{}^{\mathrm{T}} = (R_{1\alpha}, R_{2\alpha}, ..., R_{i\alpha}) \quad (1)$$

$$\mathbf{R}_\beta{}^{\mathrm{T}} = (R_{1\beta}, R_{2\beta}, ..., R_{i\beta}) \quad (2)$$

$$\mathbf{R}_u{}^{\mathrm{T}} = (R_{1u}, R_{2u}, ..., R_{iu}) \quad (3)$$

The calculated spectrum of the protein 1 is shown as

$$\mathbf{D}_1{}^{\mathrm{T}} = (D_{11}, D_{21}, ..., D_{i1}) \quad (4)$$

$$\mathbf{D}_1{}^* = \mathbf{R}_\alpha C_{\alpha 1} + \mathbf{R}_\beta C_{\beta 1} + \mathbf{R}_u C_{u1} \quad (5)$$

$$\Delta\mathbf{D}_1 = \mathbf{D}_1{}^* - \mathbf{D}_1 = \mathbf{D}_1{}^* - \mathbf{R}_\alpha \mathrm{C}_{\alpha 1} - \mathbf{R}_\beta \mathrm{C}_{\beta 1} - \mathbf{R}_u C_{u1} \quad (6)$$

To calculate the best-fit $C_{\alpha 1}$, $C_{\beta 1}$, and C_{u1}, the deviation between the measured spectrum $\mathbf{D}_1$ and the calculated spectrum $\mathbf{D}_1{}^*$ is minimized by setting the sum of derivatives of $(\Delta D_1)^2$, given by eq 6, equal to zero.

$$\frac{\partial}{\partial C_{k1}}\sum_i (D_{i1} - R_{i\alpha}C_{\alpha 1} - R_{i\beta}C_{\beta 1} - R_{iu}C_{u1})^2 = 0, \qquad k = \alpha, \beta, \mathrm{u} \quad (7)$$

$$\sum (D_1{}^* - R_{i\alpha}C_{\alpha 1} - R_{i\beta}C_{\beta 1} - R_{iu}C_{u1})R_{i\alpha} = 0 \quad (8)$$

$$\sum D_{i1}{}^* R_{i\alpha} = C_{\alpha 1}\sum R_{i\alpha}{}^2 - C_{\beta 1}\sum R_{i\alpha}\cdot R_{i\beta} - C_{u1}\sum R_{i\alpha}R_{iu} \quad (9)$$

Repeating the calculation of derivative with respect to $\mathrm{C}_{\beta 1}$ and C_{u1}, we obtain

$$\sum D_{i1}{}^* R_{i\beta} = C_{\beta 1}\sum R_{i\alpha}R_{i\beta} - C_{\beta 1}\sum R_{i\beta}{}^2 - C_{u1}\sum R_{i\beta}R_{iu} \quad (10)$$

$$\sum D_{i1}{}^* R_{iu} = C_{\alpha 1}\sum R_{i\alpha}R_{iu} - C_{\beta 1}\sum R_{i\beta}R_{iu} - C_{u1}\sum R_{iu}{}^2 \quad (11)$$

Matrixes [A], [B], and [C_1] are as follows:

$$[A] = \begin{bmatrix} \sum_i R_{i\alpha}{}^2 & \sum_i R_{i\alpha}C_{i\beta} & \sum_i R_{i\alpha}R_{iu} \\ \sum_i R_{i\alpha}R_{i\beta} & \sum_i R_{i\beta}{}^2 & \sum_i R_{i\beta}R_{iu} \\ \sum_i R_{i\alpha}R_{iu} & \sum_i R_{i\beta}C_{iu} & \sum_i R_{iu}{}^2 \end{bmatrix} \quad (12)$$

Combining eqs 9−11 and using matrix notation, we obtain

$$[C_1]^T = [C_{\alpha 1}, C_{\beta 1}, C_{u1}] \quad (13)$$

$$[B]^T = [\sum D_{i\alpha}R_{i\alpha}, \sum D_{i\alpha}R_{i\beta}, \sum D_{i\alpha}R_{iu}] \quad (14)$$

$$[B] = [A][C] \quad (15)$$

Actually, matrix [R] is the PSSRS matrix.

$$\text{matrix } [R] \equiv [\mathbf{R}_\alpha, \mathbf{R}_\beta, \mathbf{R}_u] \quad (16)$$

According to eqs 11 and 16,

$$[A] = [R]^T[R] \quad (17)$$

Similarly,

$$[B] = [R]^T[D_1] \quad (18)$$

and

$$[R]^T[D_1] = [R]^T[R][C]$$

Premultiplying $\{[R]^T[R]\}^{-1}$ on both sides,

$$[C_1] = \{[R]^T[R]\}^{-1}[R]^T[D_1] \quad (19)$$

$\{[R]^T[R]\}^{-1}[R]^T$ is the pseudo-inverse matrix of PSSRS matrix [R].

Thus, with matrix notation,

$$\text{if} \quad [R][C_1] = [D_1],$$

$$\text{then} \quad [C_1] = \{[R]^T[R]\}^{-1}[R]^T[D_1] \quad (19)$$

Therefore, the best-fit $C_{\alpha 1}$, $C_{\beta 1}$, and C_{u1} is calculated directly from the matrix transformation.

SUPPORTING INFORMATION AVAILABLE

Data matrix for protein secondary structure Raman spectra (3 pages). Ordering information is given on any current masthead page.

REFERENCES

1. Kyte, J. (1995) in *Structure in Protein Chemistry*, pp 117−146, Garland Publishing Inc., New York and London.
2. Whthrich, K. (1995) *Acta Crystallogr. 51*, 249.
3. Creighton, T. E. (1993) in *Proteins: Structures and Molecular Properties*, 2nd ed., pp 202−217, W. H. Freeman Co., New York.
4. Billeter, M. (1992) *Q. Rev. Biophys. 25*, 325.
5. Holtz, J. S. W., Bormett, R. W., Chi, Z., Cho, N., Chen, X. G., Pajcini, V., Asher, S. A., Arrogoni, M., Owen, P., and Spinelli, L. (1996) *Appl. Spectrosc. 50*, 1459.
6. Asher, S. A. (1993) *Anal. Chem. 65*, 59A.
7. Asher, S. A. (1993) *Anal. Chem. 65*, 201A.
8. Asher, S. A. (1988) *Annu. Rev. Phys. Chem. 39*, 537.
9. Ludwig, M., and Asher, S. A. (1988) *J. Am. Chem. Soc. 110*, 1005.
10. Asher, S. A. Ludwig, M., and Johnson, C. R. (1986) *J. Am. Chem. Soc. 108*, 3186.
11. Dudik, J. M., Johnson, C. R., and Asher, S. A. (1985) *J. Phys. Chem.* 89, 3805.
12. Woody, R. W., and Dunker, A. K. (1996) in *Circular Dichroism and the Conformational Analysis of Biomolecules* (Fasman, G. D., Ed.), pp 109−158, Plenum Press, New York and London.
13. Johnson, W. C., Jr. (1996) in *Circular Dichroism and the Conformational Analysis of Biomolecules* (Fasman, G. D., Ed.), pp 635−652, Plenum Press, New York and London.
14. Pancoska, P., Bitto, E., Janota, V., and Keiderling, T. A. (1995) *Faraday Discuss. 99*, 287.
15. Keiderling, T. A. (1993) in *Biomolecular Spectroscopy, Part B* (Cark, R. J. H., Hester, R. E., Eds., pp 267−312, John Wiley & Sons, New York.
16. Johnson, W. C., Jr. (1990) *Proteins: Struc., Func., Genet. 7*, 205.
17. Manning, M. C. (1989) *J. Pharm. Biomed. Anal. 7*, 1103.
18. Bussian, B., and Sander, C. (1989) *Biochemistry 28*, 4271.
19. Cho, N., and Asher, S. A. (1996) *Biospectroscopy 2*, 71.
20. Chi, Z., and Asher, S. A. (1998) *Biochemistry 37*, 2865−2872.
21. Dudik, J. M., Johnson, C. R., and Asher, S. A. (1985) *J. Chem. Phys. 82*, 1732.
22. Song, S., and Asher, S. A. (1989) *J. Am. Chem. Soc. 111*, 4295.
23. Itoh, K., Foxman, B. M., and Fasman, G. D. (1976) *Biopolymer 15*, 419.
24. Woody, R. W. (1992) *Adv. Biophys. Chem. 2*, 37.
25. Asher, S. A., Johnson, C. R., and Murtaugh, J. (1983) *Rev. Sci. Instrum. 54*, 1657.
26. Sweeney, J. A., and Asher, S. A. (1990) *J. Phys. Chem. 94*, 4784.
27. Larkin, P. J., Gustafson, W. G., and Asher, S. A. (1991) *J. Chem. Phys. 94*, 5324.
28. Johnson, C. R., Ludwig, M., and Asher, S. A. (1986) *J. Am. Chem. Soc. 108*, 905.
29. Frushour, B. G., and Koenig, J. L. (1975) in *Advances in Infrared and Raman Spectroscopy* (Clark, R. J. H., and Hester, R. E., Eds.), pp 35−142, John Wiley & Sons, New York.
30. Brown, S. D. (1995) *Appl. Spectrosc. 49*, 14A.
31. Malinowski, E. R. (1991) in *Factor Analysis in Chemistry*, pp 32−82, John Wiley & Sons, New York.
32. Mirkin, N. G., and Krimm, S. (1996) *J. Mol. Struct. 377*, 219.
33. Chen, X. G., Asher, S. A., Scheitzer-Stenner, R., Mirkin, N. G., and Krimm, S. (1995) *J. Am. Chem. Soc. 117*, 2884.
34. Chen, X. G., Scheitzer-Stenner, R., Asher, S. A., Mirkin, N. G., and Krimm, S. (1995) *J. Phys. Chem. 99*, 3074.
35. Wang, Y., Purrello, R., Georgiou, S., and Spiro, T. G. (1991) *J. Am. Chem. Soc. 113*, 6359.
36. Wang, Y., Purrello, R., Jordan, T., and Spiro, T. G. (1991) *J. Am. Chem. Soc. 113*, 6368.
37. Krimm, S., and Bandekar, J. (1986) *Adv. Protein Chem. 38*, 181.
38. McCaldon, P., and Argos, P. (1988) *Proteins* 4, 99.
39. Dwivedi, A. M., and Krimm, S. (1982) *Macromolecules 15*, 177.
40. Dwivedi, A. M., and Krimm, S. (1982) *Macromolecules 15*, 186.
41. Yang, J. T., Wu, C. C., and Martinez, H. M. (1986) *Methods Enzymol. 130*, 208.
42. Kabash, W., and Sander, C. (1983) *Biopolymer 22*, 2577.
43. Chang, T. C., Wu, C. C., and Yang, J. T. (1978) *Anal. Biochem. 91*, 13.
44. Levitt, M., and Greer, J. (1977) *J. Mol. Biol. 114*, 181.
45. Takano, T., Trus, B. L., Mandel, N., Mandel, G., Kallai, O. B., Swanson, R., and Dickerson, R. E. (1977) *J. Biol. Chem. 252*, 776.
46. Fox, J., Tu, A. T., Hruby, V. J., and Mosberg, H. I. (1981) *Arch. Biochem. Biophys. 211*, 628.

47. Hseu, T. H., and Chang, H. (1980) *Biochim. Biophys. Acta 624*, 340.
48. Wyssbrod, H. R., and Diem, M. (1992) *Biopolymers 32*, 1237.
49. Sreermana N., and Woody, R. W. (1993) *Anal. Biochem. 209*, 32.
50. Andrade, M. A., Chacon, P., Merelo, J. J., and Moran, F. (1993) *Protein Eng. 6*, 383.
51. Bohm, G., Muhr, R., and Jaenicke, R. (1992) *Protein Eng. 5*, 191.
52. Perzel, A., Hollosi, M., Tusnady, G., and Fasman, G. D. (1991) *Protein Eng. 4*, 669.
53. van Stokkum, I. H. M., Spoelder, H. J. W., Bloemendal, M., van Grondelle, R., and Goren, F. C. A. (1990) *Anal. Biochem. 191*, 110.
54. Shubin, V. V., Khazin, M. L., and Efimovskaya, T. B. (1990) *Mol. Biol. 24*, 165.
55. Manavalan, P., and Johnson, W. C., Jr. (1987) *Anal. Biochem. 167*, 76.
56. Provencher, S. W., and Glockner, J. (1981) *Biochemistry 20*, 33.
57. Bolotina I. A., Chekhov, V. O., Lugauskas, V. Y., and Ptitsyn, O. B. (1981) *Mol. Biol. 15*, 130.
58. Hennessey, J. P., and Johnson, W. C., Jr. (1981) *Biochemistry 20*, 1085.
59. Brahams, S., and Brahams, J. (1980) *J. Mol. Biol. 138*, 149.
60. Chen Y.-H., Yang, J. T., and Martinez, H. M. (1972) *Biochemistry 11*, 4120.
61. Greenfield, N. J., and Fasman G. D. (1969) *Biochemistry 8*, 4108.

BI971160Z