Quantitative IR spectrophotometry of peptide compounds in water (H2O) solution. III. Estimation of the protein secondary structure.

ARTICLE *in* BIOPOLYMERS · JANUARY 1990

Impact Factor: 2.39 · DOI: 10.1002/bip.360301311

3 AUTHORS, INCLUDING:

Sergei Y Venyaminov
Mayo Foundation for Medical Education and …

131 PUBLICATIONS **3,512** CITATIONS

SEE PROFILE

Quantitative IR Spectrophotometry of Peptide Compounds in Water (H_2O) Solutions. III. Estimation of the Protein Secondary Structure

N. N. KALNIN, I. A. BAIKALOV, and S. YU. VENYAMINOV

Institute of Protein Research, Academy of Sciences of the USSR, Puschino, Moscow Region, USSR

SYNOPSIS

Infrared spectra of 13 globular proteins have been obtained in the 1800–1480-cm^{-1} region for H_2O solutions. A method for estimating protein secondary structure from the ir spectrum has been developed. The method can also be used for estimating polypeptide and fibrous protein conformation. For the globular and fibrous proteins and polypeptides analyzed, the correlation coefficients between the ir and x-ray estimates of ordered helix, disordered helix, ordered β-structure, disordered β-structure, turns, and remainder were 0.98, 0.80, 0.99, 0.87, 0.90, and 0.92 respectively.

INTRODUCTION

Infrared spectroscopy is one of the earliest experimental methods recognized as potentially useful for estimating the secondary structure of polypeptides and proteins. Up to now, the practical usefulness of the method was, however, severely limited by absorption from H_2O interfering with the protein one. That is why D_2O is a commonly used solvent for ir spectroscopy of biological samples.

A set of powerful methods for protein secondary structure estimation has been developed for D_2O solutions during the last several years.[1-4] All of them require complete deuterium exchange of the labile protons in a protein. This problem involves partial unfolding and refolding of the protein. In any case, one should be sure that the proton–deuterium exchange is really completed and the protein structure is at least close to the native one.

The purpose of the present work was to develop a new method of estimating the protein secondary structure for H_2O solution using high sensitivity of a protein ir spectrum to the polypeptide chain conformation. The method stems from the results of various approaches to that problem, which were developed during the last few years. We tried to create a universal useful technique, which combines the main advantages of the existing ones.

CCC 0006-3525/90/13-141273-08 $04.00
Biopolymers, Vol. 30, 1273–1280 (1990)

EXPERIMENTAL

Proteins

Hen egg white lysozyme from Calbiochem (no. 4403, lot 001838) was dissolved in 0.1 *M* Na-phosphate buffer, pH 7.0.

Bovine pancreas ribonuclease A from Calbiochem (no. 55674, lot 003490) was dissolved in 0.1 *M* Na-phosphate buffer, pH 7.0.

Bovine α-chymotrypsin A from Calbiochem (no. 23083, lot 801657) was dissolved in 0.1 *M* NaCl, 1 m*M* HCl, pH 3.0.

Jack bean concanavalin A from Calbiochem (no. 234567, lot 810319) was dissolved in 0.1 *M* Na-phosphate buffer, 0.2 *M* NaCl, pH 6.7.

Sperm whale met-myoglobin kindly provided by Dr. Yu. V. Griko was dissolved in 0.1 *M* Na-phosphate buffer, pH 6.7.

Staphylococcal nuclease kindly provided by Dr. Yu. V. Griko was dissolved in 0.1 *M* Na-phosphate buffer, pH 7.1.

Horse heart cytochrome C from Serva (no. 18020) was dissolved in 0.1 *M* Na-phosphate buffer, pH 7.1. The ferri-form was obtained by oxidation with potassium ferricyanide.

Bovine insulin from Calbiochem (no. 40769, lot 102118) was dissolved in 1 m*M* HCl, pH 2.9.

Human carbonic anhydrase B kindly provided by Dr. G. V. Semisotnov was dissolved in 0.1 *M* Na-phosphate buffer, pH 7.0.

Papain was isolated from *Carica papaya* fruit latex (Calbiochem, no. 5125, lot 102388) according to the known procedure.[5] The isolated protein was modified to obtain a methyl acetamide derivative. The sample was then dissolved in 0.1 *M* Na-phosphate buffer, pH 6.5.

Horse liver alcohol dehydrogenase from Reanal (no. 01107, lot 80033692) was dissolved in 0.1 *M* Na-phosphate buffer, pH 7.6.

Baker yeast phosphoglycerate kinase from Sigma (no. P-7634, lot 64F-8130) was dissolved in 0.1 *M* Na-phosphate buffer, pH 6.8.

Porcine pancreas elastase from Sigma (no. E-6883, lot 74F-80951) was dissolved in 0.1 *M* NaCl, 0.01 m*M* HCl, pH 5.0.

All proteins were purified with high performance liquid chromatography using a TSK column G-2000SW. The homogeneity of the samples was tested by SDS electrophoresis in 15% polyacrylamide gel in the presence of β-mercaptoethanol. The solutions were then concentrated up to 20–40 mg/mL with the Centricon-10 concentrator from Amicon. These protein solutions were then used for spectral measurements.

The sample concentrations were determined from uv-visible absorption spectra recorded with a Cary-219 spectrophotometer. A cell 0.1 mm thick was used. Extinction coefficients were obtained by nitrogen analysis of the samples.[6]

All data used for calculation of the ir and uv extinction coefficients are represented in Table I.

IR Spectra

Infrared spectra were recorded in the 1800–1000-cm^{-1} region by the method of solvent absorption compensation. We obtained from five to eight spectra for each protein using various cells, concentrations, and sample preparations. These spectra were averaged. The method of registration of ir spectra in H_2O was discussed in detail earlier.[9,10] The average experimental errors of the spectra at the maxima of amide I and II bands were less than 3 and 1.5%, respectively.

Amino acid side-chain absorption curves, computed according to the known procedure,[9] were subtracted from the experimental spectra to obtain the peptide absorption.

Infrared peptide absorption spectra in the 1800–1480-cm^{-1} region were used for secondary structure computation. We experimentally found the optimal data interval of 4 cm^{-1} (81 points per spectrum): a

Table I Some Characteristics of the Basic Proteins

No.	Protein	Code	The Number of Residues	Mean Residue Weight[a]	Nitrogen Content, %[b]	UV-Visible Spectrum	
						λ_{max} (nm)	$A_{1\%,1mm}$ (λ_{max})
1	Lysozyme	LYZ	129	110.9	18.8	281.0	2.82 ± 0.05[c]
2	Ribonuclease	RSA	124	110.3	17.5	277.5	0.730 ± 0.015
3	Chymotrypsin	CHA	241	104.5	16.6	282.0	2.00 ± 0.04
4	Concanavalin	CNA	237	107.9	16.4	278.5	1.37 ± 0.04
5	Myoglobin	MBN	153	116.4	17.3	409.0	10.0 ± 0.3
6	Nuclease	SNS	149	113.1	17.4	277.0	0.97 ± 0.02
7	Cytochrome	CYT	104	118.8	16.8	409.0	8.8 ± 0.2
8	Insulin	INS	51	109.7	15.5	276.5	1.06 ± 0.03
9	Carbonic anhydrase	CAB	260	110.5	17.2	280.5	1.93 ± 0.04[c]
10	Papain	PAP	212	110.5	17.5	277.5	2.50 ± 0.10
11	Alcohol dehydrogenase	ADH	374	106.4	16.6	280.0	0.455 ± 0.004
12	Phosphoglycerate kinase	PGK	415	107.4	16.8	278.0	0.50 ± 0.02
13	Elastase	EST	240	107.9	17.8	282.0	2.02 ± 0.04

[a] The mean residue weight is determined as the protein molecular weight divided by the number of amino acid residues.

[b] The nitrogen content is determined as $[N] \cdot 14/MW$, where $[N]$ is the number of nitrogen atoms in the protein and MW is the protein molecular weight.

[c] Most of the protein extinction coefficients obtained coincide with the published ones except for lysozyme (the published value is 2.64)[7] and carbonic anhydrase (1.63).[8]

decrease of the interval did not affect the results of computations.

Data Analysis

We analyze the ir spectrum of a protein as a sum of 19 ir spectra of proteins, whose secondary structures are known from x-ray crystallography:

$$y(\nu_k) = \sum_i \gamma_i x_i(\nu_k), \quad k = 1, \ldots, 81 \qquad (1)$$

where $y(\nu_k)$ is the spectrum of the protein analyzed, $x_i(\nu_k)$ is the ith reference spectrum, γ_i is the weight for ith reference spectrum, and ν_k is the kth value of wavenumber. If we can determine the γ_i values, then we immediately have the secondary structure of the protein analyzed:

$$f_j = \sum_i \gamma_i F_{ij} \qquad (2)$$

where f_j is the content of the secondary structure j in the protein analyzed and F_{ij} is the content of the secondary structure j in the ith reference protein. The values for the matrix of structures F are derived from x-ray data.

To obtain a stable solution of this ill-posed problem, we used two different approaches: constrained statistical regularization procedure, developed by Provencher et al. for CD spectra[11-13] and orthogonal basis set analysis, suggested by Johnson et al. also for CD.[14,15]

Computations were done on IBM/PC/AT. A user-oriented program package will be available on request (customer's floppy disk 5.25″,2HD or 3.5″,2DD is desirable).

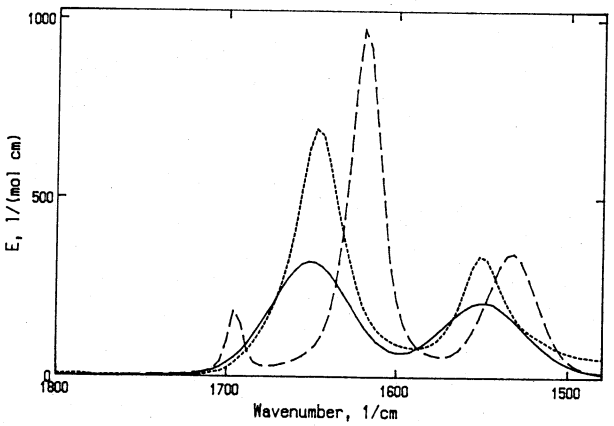

Figure 1. Average polypeptide spectra of 100% α-helix (· · · · ·), 100% β-structure (– –), and 100% random coil (——).

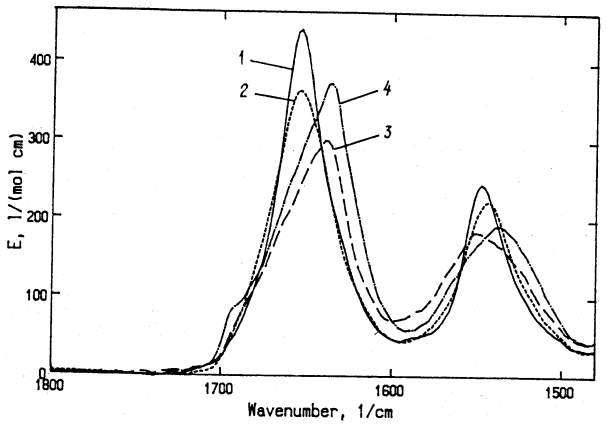

Figure 2. Peptide absorption spectra of myoglobin (1), lysozyme (2), elastase (3), and concanavalin (4).

A statistical test of the method was made according to the commonly used procedure. For 13 globular proteins the secondary structure estimates were done by removing each protein from the reference set and analyzing its structure in terms of the remaining proteins. The results were compared with the x-ray data by computing the Pearson correlation coefficients (PCC) and rms deviations (RMS):

$$\mathrm{PCC} = \frac{N \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{N \sum x_i^2 - (\sum x_i)^2}\sqrt{N \sum y_i^2 - (\sum y_i)^2}} \qquad (3)$$

$$\mathrm{rms} = \sqrt{\frac{\sum (x_i - y_i)^2}{N - 1}} \qquad (4)$$

where x_i and y_i are the compared arrays.

Basis

The basis contains globular proteins with the known spatial structures represented in the x-ray Protein Data Bank (1985).[16] We tried to choose the reference proteins so that the basis would have the maximum variety of structures.

The basis consists of 19 compounds: 13 globular proteins, represented in Table I; poly(D,L-alanine)—100% random coil (PLA); poly(L-glutamic acid)—55% helix and 45% coil (PGE); silk fibroin—30% β-structure and 70% random coil (SFB); average polypeptide 100% helix (HEX); 100% β-structure (BET); and 100% random coil (RCL). The spectra and structures for polypeptides were obtained in our earlier work.[10] Some of the reference spectra are represented in Figures 1 and 2.

Matrix of Structures

We used in the present work a method of automatic identification of secondary structure from atomic coordinates of globular proteins developed by Levitt and Greer.[17] α-Helices and β-sheets were assigned by the hydrogen-bond method alone because it gives the best statistical parameters PCC and RMS.

The F_{ij} values used coincide with those published by Levitt and Greer except for insulin. The H-bond method applied to file 1INS of the Protein Data Bank gives 54% of the helix content instead of 61% as reported. The difference appears to be the result of using different x-ray data.

The secondary structure assignment by the method of Cabsch and Sander gave poor results.[18]

Following Williams,[19] we supposed that it is possible to obtain more detailed information about the protein secondary structure than the content of α-helix, β-structure, and turns. The segments assigned to the helix with the H-bond method of Levitt and Greer were divided into ordered and disordered parts. Two residues on each end of helical segments were assigned to the disordered helix. The rest of segments were assigned to the ordered one.

We also divided the β-structure into ordered and disordered. From the total amount of residues assigned to the β-sheet with the H-bond method, we picked up residues having "classical" H bonds with at least one neighbor β-strand (Figure 3). The rest of β-sheets were assigned to disordered ones.

The turns were assigned with the α-angle method according to Levitt and Greer.

An attempt to divide the β-structure into parallel and antiparallel was not successful.

RESULTS AND DISCUSSION

The results of the statistical test of the method of estimating the protein secondary structure from ir spectra are represented in Table II. Statistical parameters are calculated for 13 globular proteins from the reference set. There are also statistical parameters for the most widespread methods of secondary structure estimation.

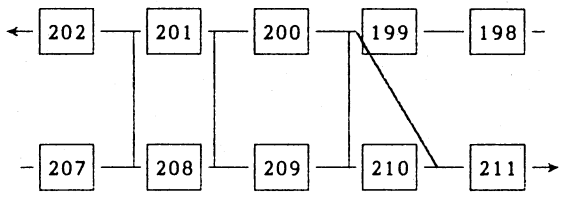

Figure 3. Fragment of a β-sheet of chymotrypsin. Residues 200, 201, 208, and 209 are assigned to the ordered β-structure because they have "classical" hydrogen bonds. All the rest residues are assigned to disordered β-structure.

We calculated the statistics for the methods R1 and R2 from Eqs. (3) and (4) for 14 proteins considered in the original publication.[19] The data for the methods CD1 and CD2 were taken from the papers.[13,14] The parameters for the method IRD were calculated for 11 proteins.[1] The proteins with a high helical content were not considered.

Numerical values of various types of secondary structure computed from ir spectra are represented in Table III.

As mentioned above, the basis consists of 13 globular proteins and 6 polypeptides. Since the reference set contains polypeptide spectra, the method can be used for secondary structure estimation of polypeptides and fibrous proteins. The data obtained are represented in Table IV. The values in the rows F were taken from our previous publication.[10] Table IV lists the statistical parameters calculated for all 19 compounds from the reference set.

The differences of the secondary structure fractions of disordered helix, disordered β-structure, and turns in Table IV should be attributed to disadvantages of method F because we could not estimate these values for polypeptides. We assumed them to be equal to zero due to a high spread and regularity of polypeptide structures.

Computation of the secondary structure with the method of generalized inverse (IR2 in Tables II and III) yields orthogonal spectra as an intermediate result.[14] The number of orthogonal spectra that have an amplitude larger than the experimental error gives the informational content of the basis. In our case it is equal to 11. Approximately the same number of degrees of freedom was obtained for the chosen solution by the method of regularization.[13] The five most significant orthogonal spectra are represented in Figure 4.

Method IR2 enables one to compute ir spectra of 100% secondary structures.[15] The spectra are represented in Figure 5. The secondary structure spectra are derived from the basis consisting of only 13 globular proteins. The spectra of the ordered α- and β-structures (Figure 5a) resemble those derived from polypeptides (Figure 1). The spectra of disordered structures have positive as well as negative regions. This gives the way to shift absorption bands for fitting the protein and polypeptide spectra. According to theoretical considerations the amide I band positions are very sensitive to spreading of the secondary structure segment.[20-22] The linear approximation, which was used to solve the problem, could

Table II Comparison of the Statistical Parameters for Various Methods of the Globular Protein Secondary Structure Estimation

	Types of Secondary Structure[a]								
Method	H_o	H_d	B_o	B_d	T	R	H_t	B_t	Statistics[b]
IR1[c]	0.97	0.77	0.94	0.91	0.80	0.48			PCC
	0.04	0.05	0.03	0.04	0.04	0.08			RMS
					0.80	0.48	0.93	0.97	PCC
					0.04	0.08	0.09	0.04	RMS
						0.74	0.93	0.97	PCC
						0.09	0.09	0.04	RMS
IR2[d]	0.98	0.80	0.93	0.90	0.75	0.48			PCC
	0.03	0.05	0.04	0.04	0.04	0.07			RMS
					0.75	0.48	0.96	0.96	PCC
					0.04	0.07	0.07	0.05	RMS
						0.71	0.96	0.96	PCC
						0.08	0.07	0.04	RMS
R1[e]	0.99	0.76			0.82	0.39	0.97	0.97	PCC
	0.04	0.05			0.03	0.03	0.06	0.04	RMS
R2[f]	0.98	0.83			0.87	0.31	0.99	0.99	PCC
	0.04	0.04			0.02	0.03	0.05	0.03	RMS
CD1[g]					0.31	0.49	0.96	0.94	PCC
					0.10	0.11	0.05	0.06	RMS
CD2[h]					0.25	0.72	0.95	0.65	PCC
					0.08	0.10	0.08	0.10	RMS
IRD[i]						0.87	0.98	0.98	PCC
						0.03	0.02	0.03	RMS

[a] Abbreviations used: H_o, ordered helix; H_d, disordered helix; B_o, ordered β-structure; Bd, disordered β-structure; T, turns; R, remainder; H_t, total helix; B_t, total β structure.
[b] PCC: Pearson correlation coefficient; RMS: root mean square deviation.
[c] From 13 proteins in Table I, using the method of regularization.
[d] From 13 proteins in Table I, using the method of orthogonal spectra.
[e] Estimation from Raman spectra (method R1).[19]
[f] Estimation from Raman spectra (method R2).[19]
[g] Estimation from CD spectra.[13]
[h] Estimation from CD spectra.[14]
[i] Estimation from ir spectra in D_2O.[1]

not shift the absorption bands. Thus the program effectively moves the bands along the wavenumber axis using S-shaped spectra of disordered secondary structures.

To estimate the effect of amino acid residue absorption on the secondary structure computation, we used the protein spectra without eliminating this contribution to compute the secondary structure. The results are represented in Table V. The statistical parameters become worse in comparison with the first row of Table II.

As mentioned above, the matrix of structures was obtained according to the H-bond method of Levitt and Greer.[17] Therefore the structures estimated from ir spectra also correspond to the criteria of the H-bond method. Unfortunately, the majority of the modern methods are based on the secondary structure assignment from the original x-ray publications or some other subjective approaches. As a result, it is difficult to compare the data obtained with different methods.

A set of criteria for x-ray data interpretation should be both internally consistent and oriented toward contribution of the secondary structure to the ir spectrum. The first demand was satisfied by using an automatic algorithm for the secondary structure assignment. To find a spectral-oriented method of the secondary structure assignment, we

Table III Comparison of Secondary Structure Content as Estimated from ir and X-Ray Data

Protein[b]	Method[c]	Structure Types[a]					
		H_o	H_d	B_o	B_d	T	R
LYZ	IR1	0.33	0.19	0.06	0.08	0.18	0.15
	IR2	0.30	0.21	0.06	0.06	0.16	0.21
	X-ray	0.27	0.19	0.07	0.10	0.22	0.15
RSA	IR1	0.11	0.08	0.16	0.21	0.21	0.23
	IR2	0.11	0.07	0.16	0.22	0.21	0.22
	X-ray	0.13	0.10	0.19	0.18	0.21	0.19
CHA	IR1	0.06	0.08	0.21	0.28	0.27	0.10
	IR2	0.05	0.03	0.19	0.26	0.27	0.20
	X-ray	0.04	0.03	0.23	0.30	0.24	0.16
CNA	IR1	0.01	0.05	0.27	0.31	0.28	0.08
	IR2	0.01	0.08	0.26	0.28	0.25	0.12
	X-ray	0.00	0.02	0.35	0.30	0.22	0.11
MBN	IR1	0.60	0.27	0.00	0.05	0.09	0.00
	IR2	0.60	0.27	0.00	0.02	0.11	0.00
	X-ray	0.63	0.21	0.00	0.00	0.07	0.09
SNS	IR1	0.20	0.04	0.15	0.14	0.17	0.31
	IR2	0.17	0.07	0.13	0.16	0.18	0.30
	X-ray	0.17	0.08	0.13	0.19	0.22	0.21
CYT	IR1	0.24	0.13	0.09	0.11	0.21	0.22
	IR2	0.26	0.12	0.05	0.10	0.22	0.25
	X-ray	0.28	0.19	0.06	0.04	0.17	0.26
INS	IR1	0.22	0.17	0.08	0.09	0.18	0.26
	IR2	0.24	0.16	0.07	0.09	0.18	0.25
	X-ray	0.30	0.24	0.05	0.10	0.12	0.19
CAB	IR1	0.04	0.06	0.20	0.23	0.25	0.22
	IR2	0.04	0.07	0.19	0.20	0.24	0.25
	X-ray	0.04	0.04	0.19	0.23	0.25	0.25
PAP	IR1	0.14	0.12	0.12	0.19	0.27	0.17
	IR2	0.14	0.13	0.15	0.16	0.22	0.19
	X-ray	0.16	0.09	0.13	0.15	0.25	0.22
ADH	IR1	0.16	0.05	0.19	0.17	0.15	0.28
	IR2	0.17	0.05	0.19	0.17	0.14	0.29
	X-ray	0.15	0.10	0.13	0.23	0.19	0.20
PGK	IR1	0.27	0.19	0.12	0.15	0.20	0.07
	IR2	0.26	0.14	0.13	0.17	0.22	0.09
	X-ray	0.20	0.10	0.12	0.11	0.20	0.27
EST	IR1	0.01	0.01	0.21	0.27	0.27	0.23
	IR2	0.02	0.01	0.21	0.28	0.28	0.20
	X-ray	0.05	0.05	0.20	0.24	0.28	0.18

[a] Abbreviations are as in Table II.
[b] Abbreviations for proteins are as in Table I.
[c] Abbreviations for methods are as in Table II.

used various combinations of the algorithms suggested by Levitt and Greer. Finally, the best statistical parameters PCC and RMS were obtained with the H-bond method alone.

The average accuracy of secondary structure assignment made on similar proteins determined by the same group of workers, whether as dimers in the same crystal or as independent crystal structures, is 3.5% for α-helices and 6.2% for β-strands.[17] These values can be considered as the limits for the precision of secondary structure estimation resulting from the accuracy of the x-ray data.

Up to now we have not found any protein that should be excluded from consideration because of poor correspondence between estimated and x-ray secondary structures despite the large variety of the

Table IV Estimation of Polypeptide and Fibrous Protein Secondary Structure

Sample[c]	Method[d]	Structure Types[a] H_o	H_d	B_o	B_d	T	R	Statistics[b]
PGE	IR1	0.63	0.06	0.00	0.00	0.00	0.31	
	F	0.55	0.00	0.00	0.00	0.00	0.45	
PLA	IR1	0.00	0.03	0.00	0.03	0.00	0.94	
	F	0.00	0.00	0.00	0.00	0.00	1.00	
SFB	IR1	0.00	0.03	0.33	0.19	0.14	0.32	
	F	0.00	0.00	0.30	0.00	0.00	0.70	
HEX	IR1	0.84	0.00	0.00	0.07	0.09	0.00	
	F	1.00	0.00	0.00	0.00	0.00	0.00	
RCL	IR1	0.01	0.00	0.04	0.04	0.05	0.86	
	F	0.00	0.00	0.00	0.00	0.00	1.00	
BET	IR1	0.00	0.09	0.91	0.00	0.00	0.00	
	F	0.00	0.00	1.00	0.00	0.00	0.00	
		0.98	0.80	0.99	0.87	0.90	0.92	PCC[e]
		0.05	0.05	0.04	0.06	0.05	0.12	RMS

[a,b] Abbreviations are as in Table II.
[c] Abbreviations are as in the experimental section.
[d] IR1: method of regularization; F: from data obtained in the second paper of this series.[10]
[e] PCC and RMS are calculated for all 19 compounds in the reference set.

proteins analyzed. This fact seems to be important for reliability of the method suggested.

Each of the methods represented in Table II has some advantages and disadvantages. We believe the main advantage of our method is the possibility of studying the protein structure in the H_2O solution with varying pH, ionic strength, temperature, etc. The sample concentration should be in the range of 1–3%. The total amount of protein required for each spectrum is about 0.1 mg. The use of the Fourier transform ir spectrometer gives the possibility of decreasing the sample concentration to 0.1% and the amount of protein to 0.01 mg.

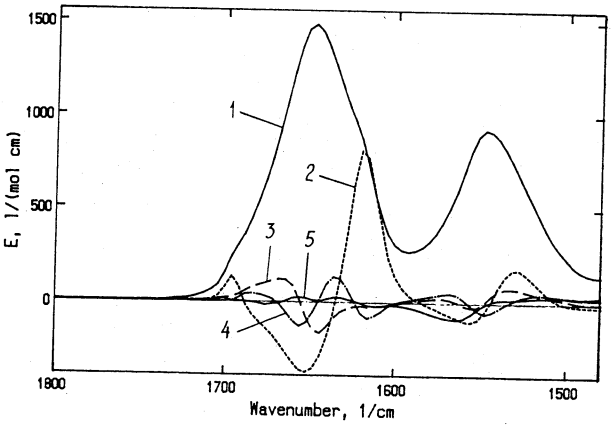

Figure 4. The five most significant orthogonal spectra, which are numbered from 1 to 5 in order of decreasing singular values.

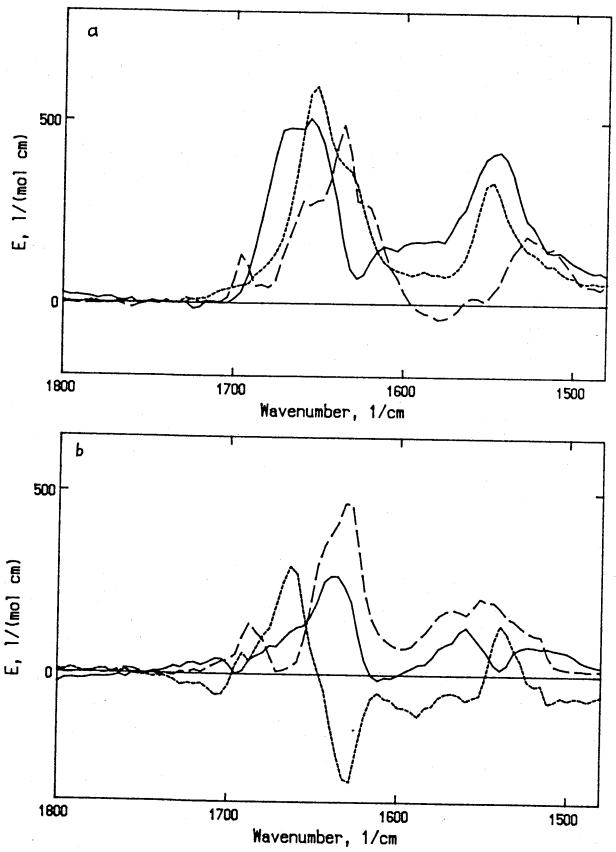

Figure 5. Secondary structure spectra: (a) ordered helix (· · · · ·), ordered β-structure (– –), and turns (——); (b) disordered helix (· · · · ·), disordered β-structure (– –), and remainder (——). The spectra were computed from 13 globular proteins in the reference set.

Table V Estimation of the Globular Protein Secondary Structure from IR Spectra Without Eliminating Amino Acid Side-Chain Absorption

	Structure Types[a]						
Method	H_o	H_d	B_o	B_d	T	R	Statistics[b]
IR1[c]	0.92	0.68	0.85	0.90	0.53	0.32	PCC
–	0.06	0.06	0.05	0.04	0.06	0.14	RMS

[a,b,c] Abbreviations are as in Table II.

As mentioned above, H_2O as a solvent is much more preferable than D_2O for studying the protein structure. It is known that H-D exchange leads to an increase of the protein stability.[23] In the process of association–dissociation, D_2O shifts the steady concentrations of enzymes and substrates to association.[24,25] Thus D_2O somewhat changes the protein properties in comparison with the native ones. From this point of view the suggested method is more favorable than the method for D_2O solutions[1] and suspensions.[19]

This method can be used for routine estimations automatically. In principle, to estimate the protein secondary structure one need only correctly obtain the ir spectrum of a protein. All other procedures including elimination of the amino acid side-chain absorption are made with the computer programs.

REFERENCES

1. Byler, D. M. & Susi, H. (1986) *Biopolymers* **25,** 469–487.
2. Chirgadze, Yu. N., Shestopalov, B. V. & Venyaminov, S. Yu. (1973) *Biopolymers* **12,** 1337–1351.
3. Chirgadze, Yu. N. & Brazhnikov, E. V. (1974) *Biopolymers* **13,** 1701–1712.
4. Eckert, K., Grosse, R. M., Malur, J. & Repke, K. R. H. (1977) *Biopolymers* **16,** 2549–2563.
5. Kimmel, J. R. & Smith, E. L. (1954) *J. Biol. Chem.* **207,** 515–531.
6. Jaenicke, L. (1974) *Anal. Biochem.* **61,** 623–627.
7. Span, J., Lenarcic, S. & Lapanje, S. (1974) *Biochim. Biophys. Acta* **359,** 311–319.
8. Finacoshi, S. & Deutsch, H. F. (1969) *J. Biol. Chem.* **244,** 3438–3451.
9. Venyaminov, S. Yu. & Kalnin, N. N. (1990) *Biopolymers,* **30,** 1243–1257.
10. Venyaminov, S. Yu. & Kalnin, N. N. (1990) *Biopolymers,* **30,** 1259–1271.
11. Provencher, S. W. (1982) *Comp. Phys. Commun.* **27,** 213–227.
12. Provencher, S. W. (1982) *Comp. Phys. Commun.* **27,** 229–242.
13. Provencher, S. W. & Glockner J. (1981) *Biochemistry* **20,** 33–37.
14. Hennessey, J. P. & Johnson, W. C. (1981) *Biochemistry* **20,** 1085–1094.
15. Compton, L. A. & Johnson, W. C. (1986) *Anal. Biochem.* **155,** 155–167.
16. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rogers, J. R., Kennard, O., Shimamouchi, T., Tasumi, M. (1977) *Eur. J. Biochem.* **80,** 319–324.
17. Levitt, M. & Greer, J. (1977) *J. Mol. Biol.* **144,** 181–239.
18. Kabsch, W. & Sander, C. (1983) *Biopolymers* **22,** 2577–2637.
19. Williams, R. W. (1983) *J. Mol. Biol.* **166,** 581–603.
20. Chirgadze, Yu. N. & Nevskaya, N. A. (1976) *Biopolymers* **15,** 607–625.
21. Bandekar, J. & Krimm, S. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 774–777.
22. Krimm, S. & Abe, Ya. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 2788–2792.
23. Harrington, W. F. & Hippel, P. H. (1961) *Arch. Biochem. Biophys.* **92,** 100–112.
24. Barhurst, P. A., Nichol, L. W. & Sawyer, W. H. (1972) *J. Biol. Chem.* **247,** 3198–3208.
25. Uratani, Y. (1974) *J. Biochem.* **75,** 1143–1152.

Received June 21, 1989
Accepted June 6, 1990