*Protein Science* (1993), *2*, 411-419. Cambridge University Press. Printed in the USA.


# Frequency analysis of infrared absorption and vibrational circular dichroism of proteins in $D_2O$ solution

PETR PANCOSKA,[1,2] LIJIANG WANG,[1] AND TIMOTHY A. KEIDERLING[1]

[1] Department of Chemistry, University of Illinois at Chicago, Chicago, Illinois 60680
[2] Department of Chemical Physics, Faculty of Mathematics and Physics, Charles University, 121 16 Prague 2, Czech Republic

(RECEIVED September 15, 1992; REVISED MANUSCRIPT RECEIVED November 24, 1992)

**Abstract**

The IR absorption frequencies as derived from second derivatives of the Fourier transform IR spectra of the amide I′ bands of globular proteins in $D_2O$ are compared to those obtained from band fitting of the vibrational circular dichroism (VCD) spectra. The two sets of frequencies are in very good agreement, yielding consistent ranges where amide I′ VCD and IR features occur. Use of VCD to complement the IR allows one to add sign information to the frequency information so that features occurring in the overlapping frequency ranges that might arise from different secondary structures can be better discriminated. From this comparison, it is clear that correlation just of the frequency of a given IR transition to secondary structure can lead to a nonunique solution. Different sign patterns were identified for correlated groups of globular proteins in restricted frequency ranges that have been previously assigned to defined secondary structural elements. Hence, different secondary structural elements must contribute band components to a given frequency range.

**Keywords:** CD of proteins; Fourier self-deconvolution of IR spectra; IR frequency assignment; protein secondary structure

Use of IR frequencies to determine secondary structure in polypeptides and proteins has a long history, stemming from the first physical characterizations of these biomolecules (Parker, 1971). Early on it was recognized that the amide I band gave rise to resolvably different absorption maximum frequencies for the $\alpha$-helix and $\beta$-sheet conformations of homopolypeptides (Susi, 1972). Because these IR maxima at characteristic frequencies are often recognizable in protein spectra, they came into common use for characterization of the dominant secondary structure in proteins. These frequencies become one basis for interpretation of the multiple spectral components that must lie within the typically asymmetric IR band profile of globular proteins, which are composed of a mixture of secondary structural elements. With the advent of Fourier transform (FT) IR spectrometers and the improvements in signal-to-noise ratio (S/N) thereby available, it has become possible to achieve significant resolution enhancement of the amide I (or amide I′ in $D_2O$ solutions) IR band profile so that these components and their associated spectral frequencies could be identified. Subsequently the relative intensities of such bands have been empirically correlated to the distribution of secondary structure components in the respective proteins by several independent researchers (Susi & Byler, 1983; Byler & Susi, 1986; Surewicz & Mantsch, 1988).

FTIR frequency analysis is attractive due to the relative simplicity of the instrumentation required and the current availability of both algorithms and software for the necessary mathematical treatment of the experimental data. This ease of use comes at the cost of necessarily assuming that there is a homomorphic mapping of the mathematically extracted spectral parameters onto a given set of protein secondary structure types. In most cases this means an assumption is made that a unique relationship exists between the secondary structure and IR absorption frequencies *and* that this uniqueness works for transformations in *both* directions. Due to the overall complexity of the protein structure, it is possible and even probable that a given frequency (or range of close lying frequen-

Reprint requests to: Timothy A. Keiderling, Department of Chemistry (M/C 111), University of Illinois at Chicago, Box 4348, Chicago, Illinois 60680.

cies) could correspond to multiple conformations. Because protein structures are far from uniform, one should therefore view the "characteristic frequencies" for given structural types as ranges. Unfortunately these ranges for globular proteins can overlap leading to ambiguity in structural assignment.

We have recently shown that vibrational circular dichroism (VCD) spectra for the amide I band of globular proteins in $D_2O$ solutions can be measured, and that the VCD bandshapes are characteristic of protein secondary structure (Pancoska et al., 1989). This qualitative observation was further strengthened by development of a method for quantitative reduction of the amide I′ VCD data into secondary structure parameters (Pancoska et al., 1991). Bandshape analysis using the principal component method of factor analysis followed by regression analyses of the resulting subspectral coefficients can be used to fit fractional coefficients of protein secondary structure derived from X-ray crystal structures (Kabsch & Sander, 1983) for a training set of "known" proteins. Then the equations and spectral parameters determined can be used to predict structural coefficients for "unknown" proteins (Pancoska et al., 1991).

VCD studies measure transitions to the same excited vibrational states as seen in the more conventional IR studies. Thus the frequency pattern observed in VCD should be directly related to that seen in the FTIR experiments. In practice, the exciton band associated with a particular secondary structure type will cover a range of frequencies (Krimm & Bandekar, 1986). The resulting band contour may have multiple maxima at different frequencies corresponding to the dipolar and rotational strengths of various components of the exciton band. The range for these bands can be quite wide; for example, the anti-parallel $\beta$-sheet band has features $\sim 60\ cm^{-1}$ apart. Because the dipolar and rotational strengths result from different operators, this exciton width can cause the corresponding IR and VCD band maxima to appear at somewhat different positions.

However, VCD has an inherent feature providing a unique advantage in terms of resolution of structurally characteristic frequencies as compared to IR absorption spectroscopy. VCD is strongly dependent on the underlying molecular structure due to its origin as an interference term in the multipolar representation of light absorption. This results in a sign variation that directly relates to structural variations and tends to resolve the underlying bands. Thus VCD can be seen to be a natural tool for resolution enhancement of IR spectra that is not dependent on parameters that can lead to artifacts in the FTIR deconvolution methods commonly used. Furthermore, the VCD sign patterns are physical properties directly dependent on the secondary structure of those residues giving rise to the observed spectral band. They will be constant for a given secondary structure and are expected to be independent of small environmental and nonuniformity effects that can cause absorption frequencies to shift, leading to the ranges alluded to above. We have shown these patterns to be remarkably insensitive to the sequence of the peptide being studied, even when the frequencies of the underlying absorption transitions do change measurably (Keiderling, 1993). If different types of structural sequences give rise to frequency ranges that overlap, their sign patterns may allow one to distinguish which type is the dominant contributor for a given protein.

In summary, it is well established, from decades of study, that the secondary structure gives rise to a variety of frequencies even within a well-characterized transition such as amide I′ band in polypeptides and proteins. The important question for quantitative studies is: What are the limits of using such frequency information when one knows that the other aspects of the structure and the environment can cause IR frequency shifts? VCD can be used to help define some of those limits. Furthermore, while it may be no surprise that the VCD frequencies parallel those of FTIR, this has not been demonstrated in detail. Such a demonstration requires some sort of deconvolution of the VCD bandshape, which, due to the two-signed character of VCD measurements, is an even more complex task than for absorption spectra. In this paper we address these two questions. We have decomposed the VCD spectra of a set of proteins into component bands and compared their frequencies to frequencies derived from FTIR spectra of the same proteins, which we have remeasured in a consistent manner for sake of this comparison. Two different band decompositions, in some sense limiting cases, will be used to describe the VCD spectra: an independent fit to a minimal number of bands and a fit based on frequency parameters derived from second derivative analysis of the related FTIR spectra.

## Results

In Table 1 are summarized the peak positions determined from the second derivatives of the FTIR spectra of the proteins studied in order of decreasing $\alpha$-helix content. The values tabulated for the fractional content (FC) of $\alpha$-helix were taken from our analysis of the Kabsch and Sander (1983) data set for known proteins and from our earlier work with VCD (Pancoska et al., 1991) for the unknowns. In Table 2 are given the results of Fits I and II of VCD data in terms of band positions, signs, and widths. While Gaussian band shapes give better fits of VCD spectra than do Lorentzians, it is likely that little physical interpretation can be made of this fact, given the signal-to-noise ratio (S/N) limitations of the VCD experiment. On the other hand, there are good physical reasons to anticipate that the component bandshapes would be of at least mixed character.

The FTIR second derivative calculation gave frequencies that were in good agreement with those derived from

**Table 1.** *Frequencies of amide I' component bands as determined by the second derivative of Fourier transform (FT) IR spectra*

| | $FC_\alpha^a$ (%) | Band 1 | Band 2 | Band 3 | Band 4 | Band 5 | Band 6 | Band 7 | Band 8 | Band 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| Myoglobin | 77 | – | – | 1,678 | – | 1,648 | – | – | 1,629 | 1,612 |
| Hemoglobin | 67 | – | – | 1,674 | 1,660 | 1,651 | – | 1,633 | – | 1,615 |
| Albumin | 60 | – | – | 1,676 | – | 1,651 | 1,640 | – | 1,628 | – |
| Triose phosphate isomerase | 43 | 1,692 | 1,682 | 1,673 | 1,663 | – | 1,647 | 1,637 | 1,629 | 1,615 |
| Lysozyme | 29 | – | 1,681 | 1,672 | 1,665 | 1,652 | 1,640 | – | – | 1,605 |
| Cytochrome *c* | 26 | – | – | 1,673 | – | 1,650 | – | 1,639 | 1,628 | 1,611 |
| Ribonuclease A | 25 | – | 1,687 | 1,676 | 1,661 | – | 1,647 | 1,637 | – | 1,607 |
| Papain | 23 | – | 1,685 | 1,678 | 1,668 | 1,653 | 1,645 | 1,632 | – | 1,611 |
| Lactoferrin | 22 | 1,697 | 1,682 | – | 1,665 | 1,652 | 1,642 | 1,632 | – | 1,603 |
| Ribonuclease S | 18 | – | 1,680 | – | 1,662 | 1,653 | – | 1,632 | – | 1,607 |
| Lactoglobulin | 15 | 1,692 | – | 1,679 | 1,666 | 1,660 | 1,648 | 1,633 | 1,624 | 1,611 |
| Trypsin inhibitor | 14 | – | 1,684 | 1,672 | 1,666 | 1,659 | 1,640 | 1,634 | – | 1,609 |
| Trypsin | 8 | 1,693 | 1,683 | 1,673 | 1,662 | 1,652 | 1,640 | 1,633 | 1,625 | 1,609 |
| Carbonic anhydrase | 7 | – | 1,689 | 1,676 | 1,662 | 1,652 | – | 1,635 | 1,624 | – |
| Chymotrypsin | 6 | – | 1,687 | 1,673 | 1,665 | 1,654 | 1,644 | 1,636 | – | 1,602 |
| Chymotrypsinogen | 5 | – | 1,688 | 1,674 | 1,665 | 1,652 | – | 1,636 | 1,628 | 1,603 |
| Elastase | 5 | – | 1,684 | 1,671 | 1,661 | – | 1,645 | 1,634 | – | 1,614 |
| Casein | 2 | – | – | – | 1,666 | – | 1,640 | – | – | 1,611 |
| Concanavalin | 0 | 1,692 | 1,681 | 1,668 | 1,656 | 1,650 | – | 1,633 | 1,620 | – |
| Average frequencies | | 1,693 ± 2 | 1,684 ± 3 | 1,674 ± 3 | 1,663 ± 3 | 1,653 ± 3 | 1,643 ± 3 | 1,634 ± 2 | 1,626 ± 3 | 1,609 ± 4 |

[a] Fractional content (FC) of α-helix.

**Table 2.** *Frequencies, signs, and bandwidths of component vibrational circular dichroism (VCD) bands in the amide I' frequency region*

| | Fit | $FC_\alpha$ | Band 1 | Band 2 | Band 3 | Band 4 | Band 5 | Band 6 | Band 7 | Band 8 | Band 9 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Myoglobin | I | 77 | – | 1,672 | – | 1,656 | 1,646 | – | 1,622 | – | 1,597 |
| | | | – | −/26 | – | −/15 | +/20 | – | −/15 | – | +/13 |
| | II | | – | 1,670 | – | 1,656 | 1,645 | – | – | – | – |
| | | | – | −/21 | – | −/17 | +/16 | – | – | – | – |
| Hemoglobin | I | 67 | – | – | 1,663 | 1,653 | 1,647 | – | 1,631 | – | 1,605 |
| | | | – | – | +/25 | −/16 | +/20 | – | −/20 | – | +/12 |
| | II | | – | – | – | 1,658 | 1,645 | – | 1,624 | – | – |
| | | | – | – | – | −/30 | +/13 | – | −/17 | – | – |
| Albumin | I | 60 | – | 1,668 | – | 1,654 | 1,644 | – | 1,620 | – | 1,614 |
| | | | – | −/21 | – | −/23 | +/17 | – | −/25 | – | +/16 |
| | II | | – | – | – | 1,660 | 1,644 | – | 1,624 | – | – |
| | | | | – | – | −/25 | +/13 | – | −/9 | – | – |
| Triose phosphate isomerase | I | 43 | – | 1,670 | – | 1,659 | 1,649 | 1,639 | 1,626 | – | 1,612 |
| | | | – | −/26 | – | −/16 | +/16 | +/10 | −/21 | – | −/35 |
| | II | | – | – | – | 1,658 | 1,650 | 1,640 | 1,630 | – | – |
| | | | – | – | – | −/32 | +/11 | +/12 | −/30 | – | – |
| Lysozyme | I | 29 | – | 1,671 | – | 1,656 | 1,651 | 1,633 | – | 1,619 | – |
| | | | – | −/14 | – | −/18 | +/25 | −/17 | – | −/32 | – |
| | II | | – | 1,671 | – | 1,659 | 1,650 | 1,632 | – | 1,610 | – |
| | | | – | −/14 | – | −/11 | +/7 | −/18 | – | −/22 | – |
| Cytochrome *c* | I | 26 | – | – | – | 1,654 | 1,643 | – | 1,623 | – | 1,614 |
| | | | – | – | – | −/25 | +/25 | – | −/29 | – | +/23 |
| | II | | – | – | – | 1,655 | – | – | 1,627 | 1,606 | – |
| | | | – | – | – | −/20 | – | – | −/14 | −/14 | – |
| Ribonuclease A | I | 25 | 1,684 | 1,673 | – | 1,649 | 1,644 | 1,634 | – | 1,616 | 1,599 |
| | | | −/30 | −/25 | – | +/25 | −/25 | −/20 | – | −/20 | −/20 |
| | II | – | 1,690 | 1,671 | – | – | – | 1,635 | – | – | – |
| | | | −/14 | −/19 | – | – | – | −/15 | – | – | – |

(*continued*)

**Table 2.** *Continued*

| | Fit | $FC_\alpha$ | Band 1 | Band 2 | Band 3 | Band 4 | Band 5 | Band 6 | Band 7 | Band 8 | Band 9 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Papain | I | 23 | 1,678 | 1,671 | – | 1,655 | 1,645 | 1,638 | 1,624 | – | – |
| | | | −/19 | −/6 | – | −/22 | +/25 | −/37 | −/19 | – | – |
| | II | | – | 1,672 | – | 1,657 | – | 1,639 | 1,626 | – | – |
| | | | – | −/19 | – | −/10 | – | +/13 | −/20 | – | – |
| Lactoferrin | I | 22 | – | – | – | 1,655 | 1,639 | 1,635 | 1,625 | 1,615 | 1,597 |
| | | | – | – | – | −/25 | +/20 | −/20 | −/14 | −/25 | +/25 |
| | II | | – | 1,674 | – | 1,657 | – | – | 1,623 | – | – |
| | | | – | −/18 | – | −/14 | – | – | −/21 | – | – |
| Ribonuclease S | I | 18 | – | 1,675 | – | 1,649 | 1,643 | – | 1,626 | 1,618 | – |
| | | | – | +/21 | – | +/29 | −/25 | – | −/18 | −/18 | – |
| | II | | – | 1,674 | – | 1,657 | – | – | 1,623 | – | – |
| | | | – | +/12 | – | +/16 | – | – | −/23 | – | – |
| Lactoglobulin | I | 15 | – | 1,661 | – | 1,656 | 1,639 | 1,636 | 1,622 | 1,616 | – |
| | | | – | −/15 | – | −/21 | +/23 | −/23 | −/14 | −/20 | – |
| | II | | – | 1,678 | – | 1,656 | – | – | 1,622 | – | – |
| | | | – | −/18 | – | −/16 | – | – | −/15 | – | – |
| Trypsin inhibitor | I | 14 | 1,683 | 1,670 | 1,662 | 1,652 | – | 1,638 | 1,626 | – | 1,606 |
| | | | +/36 | +/10 | +/18 | +/25 | – | −/33 | −/25 | – | −/30 |
| | II | | – | – | – | – | 1,645 | – | 1,628 | – | – |
| | | | – | – | – | – | +/17 | – | −/27 | – | – |
| Trypsin | I | 8 | 1,677 | 1,669 | – | 1,658 | 1,648 | 1,636 | 1,630 | 1,621 | 1,605 |
| | | | +/20 | +/20 | – | +/20 | +/18 | −/21 | −/15 | −/18 | −/20 |
| | II | | – | 1,679 | – | – | 1,652 | – | 1,628 | – | – |
| | | | – | +/28 | – | – | +/28 | – | −/27 | – | – |
| Carbonic anhydrase | I | 7 | – | – | 1,669 | – | 1,647 | – | 1,626 | – | – |
| | | | – | – | +/22 | – | +/25 | – | −/20 | – | – |
| | II | | – | – | 1,670 | – | 1,645 | – | 1,625 | – | – |
| | | | – | – | +/24 | – | +/20 | – | −/22 | – | – |
| Chymotrypsin | I | 6 | 1,692 | – | 1,663 | – | 1,646 | 1,633 | 1,623 | – | 1,597 |
| | | | +/21 | – | +/7 | – | +/18 | −/18 | −/22 | – | −/20 |
| | II | | 1,692 | – | 1,663 | – | 1,645 | – | 1,628 | – | – |
| | | | +/13 | – | +/9 | – | +/15 | – | −/23 | – | – |
| Chymotrypsinogen | I | 5 | – | 1,676 | – | 1,652 | – | 1,636 | – | 1,619 | – |
| | | | – | +/25 | – | +/25 | – | −/29 | – | −/27 | – |
| | II | | – | 1,676 | – | 1,659 | – | 1,632 | – | 1,616 | – |
| | | | – | +/23 | – | +/17 | – | −/25 | – | −/18 | – |
| Elastase | I | 5 | – | – | 1,661 | – | 1,642 | – | – | 1,614 | 1,600 |
| | | | – | – | −/18 | – | +/18 | – | – | −/19 | −/12 |
| | II | | – | – | 1,662 | – | 1,643 | – | – | 1,613 | 1,598 |
| | | | – | – | +/16 | – | +/19 | – | – | −/18 | −/10 |
| Casein | I | 2 | – | 1,676 | 1,662 | – | – | 1,633 | – | 1,618 | 1,608 |
| | | | – | +/25 | +/18 | – | – | −/18 | – | −/17 | −/25 |
| | II | | – | – | 1,666 | – | – | – | 1,628 | – | 1,605 |
| | | | – | – | +/31 | – | – | – | −/28 | – | −/21 |
| Concanavalin | I | 0 | – | 1,673 | – | – | 1,646 | 1,639 | 1,628 | – | 1,607 |
| | | | – | +/19 | – | – | +/20 | −/15 | −/20 | – | −/15 |
| | II | | – | 1,676 | – | 1,654 | – | 1,636 | 1,624 | – | – |
| | | | – | +/18 | – | +/10 | – | −/12 | −/13 | – | – |
| Averaged frequencies | | | 1,685 ± 6 | 1,672 ± 4 | 1,663 ± 3 | 1,655 ± 3 | 1,645 ± 3 | 1,636 ± 3 | 1,625 ± 3 | 1,615 ± 4 | 1,604 ± 6 |

FSD and with band fitting of the undeconvolved FTIR spectral bandshape. Given this, we chose to use the second derivative determined peak position values as a basis for the Fit I data analysis because these are most objectively determined. Use of these frequencies to fit the amide I′ VCD spectra in the Fit I analysis proceeded smoothly for most spectra. If relatively large variations in the deconvolved bandwidths (by a factor of >2) are allowed, the Fit II analysis also posed no difficulties. In particular, the strongly variable VCD bandshape (even though of lower S/N) was generally better fit by both these approaches than were the undeconvolved absorbance spectra. While the FTIR had virtually no noise, these proteins result in smooth, generally featureless amide I′ band pro-

files that exhibit little variation, which, if it were present, could constrain the multiparameter fits.

### *High $FC_\alpha$ proteins*

The protein subgroup with the highest fractional content (FC) of α-helix consists of hemoglobin ($FC_\alpha$ 67%), myoglobin ($FC_\alpha$ 77%) and albumin ($FC_\alpha$ >60%, estimated). The characteristics of this group are: The FTIR peak at 1,651 $cm^{-1}$ corresponds in all cases to a positive-negative VCD couplet at 1,645–1,657 $cm^{-1}$. No significant VCD counterparts were found for the 1,673-$cm^{-1}$ FTIR features. Fits I and II differ in the presence or absence of the second negative VCD band centered at 1,660–1,670 $cm^{-1}$ along with the dominant negative VCD feature at ~1,657 $cm^{-1}$. Fits to the low frequency region (1,660–1,625 $cm^{-1}$) using both procedures are less reliable due to the generally low VCD intensity and S/N ratio in this region. The VCD bands found here are not in all cases supported by the corresponding second derivative FTIR features.

### *Medium $FC_\alpha$ proteins*

The next subgroup of proteins studied with a medium content of α-helix consists of lactoglobulin (estimated $FC_\alpha$ 20–30%), cytochrome *c* ($FC_\alpha$ 26%), papain ($FC_\alpha$ 23%), lactoferrin (estimated $FC_\alpha$ 20–30%), triose phosphate isomerase ($FC_\alpha$ 43%), ribonuclease A (estimated $FC_\alpha$ 16–25%), and lysozyme ($FC_\alpha$ 29%). VCD spectra for these proteins have a bandshape roughly in the form of a "W." This shape gives the most difficulty in fitting because the cancellation of oppositely signed subbands leads to larger uncertainty in the amplitudes and widths obtained. The results of Fit I and II differed mainly in the positive part of the spectrum at 1,640–1,655 $cm^{-1}$. This can be modelled by single positive peak with low intensity and narrow bandwidth (Fit II) or as a superposition of two oppositely signed subbands (Fit I). That there should be more than one band in this region for this group of proteins is, in most cases, supported by the FTIR second derivatives. On the other hand, the large cancellation of VCD components incumbent in such a choice makes it unattractive. (In some cases, e.g., for lactoferrin, the VCD bandshape can be described successfully without any band present in this frequency region.) No obvious VCD counterparts were found for the FTIR second derivative features at about 1,664 $cm^{-1}$ for this group.

### *Low $FC_\alpha$ proteins*

This subgroup of proteins with a small content of α-helix consists of concanavalin A ($FC_\alpha$ 0%), ribonuclease S ($FC_\alpha$ 18%), chymotrypsinogen ($FC_\alpha$ 0–15%, estimated), trypsin ($FC_\alpha$ 8%), casein ($FC_\alpha$ 0–15%, estimated), chymotrypsin ($FC_\alpha$ 6%), trypsin inhibitor ($FC_\alpha$ 14%), elastase ($FC_\alpha$ 5%), and carbonic anhydrase ($FC_\alpha$ 7%). For concanavalin A and ribonuclease S, Fits I and II differ in the positions and source of the 1,650 $cm^{-1}$ feature as either a single position peak or as the residue of the cancellation of a (+, −) couplet.

For trypsin, casein, and carbonic anhydrase, Fit I replaces one or more bands used in Fit II by a pair of narrower bands of the same sign based on the second derivative FTIR results. Chymotrypsinogen, elastase, and carbonic anhydrase resulted in the same band and sign pattern for both types of fit.

### *Summary*

Some typical fits for proteins from each of the above groups are shown in Figures 1–3. The bands and their sum are indicated with dashed lines. The deviation from or match to the experimental spectrum is indicated by the very good overlap of the solid (experimental) and dashed (fit sum) bandshapes. The two Fits, I and II, do not exhibit a systematic difference in degree of replication of the experimental data. It is not our intention to give a detailed interpretation of the fitting results summarized above. The preceding paragraphs demonstrate very clearly the ambiguities inherent in such an attempt. Nevertheless, the results do indicate that there is a consistent sign pattern for the corresponding deconvolved VCD component bands in the amide I′ region. This can be demonstrated in a systematic and unambiguous way by comparing the results of the two limiting types of fit employed here and

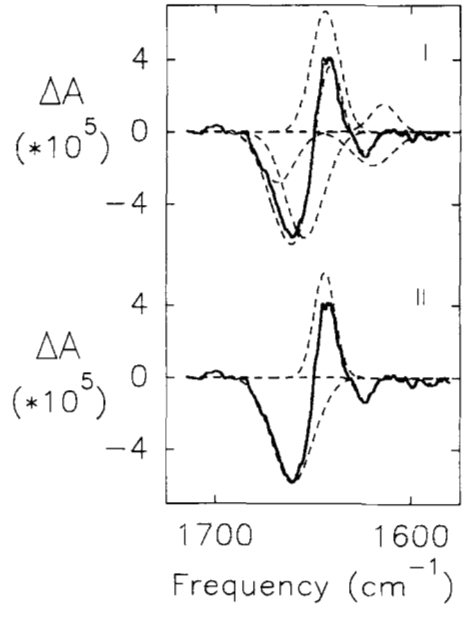


**Fig. 1.** Comparison of Fits I and II to VCD spectrum of albumin (example protein with high content of α-helix). Solid line – experimental VCD spectrum; dashed line – component Gaussian bands. The sum of the components is also indicated with a dashed line overlapping the experimental spectrum.

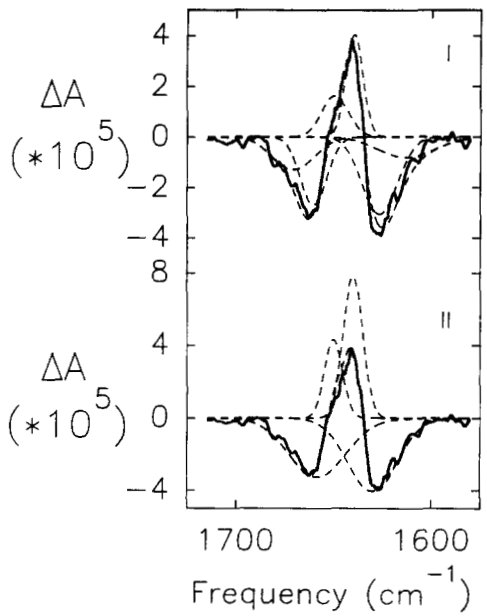


**Fig. 2.** Comparison of Fits I and II to VCD spectrum of triose phosphate isomerase (example protein with medium content of α-helix). As in Figure 1.

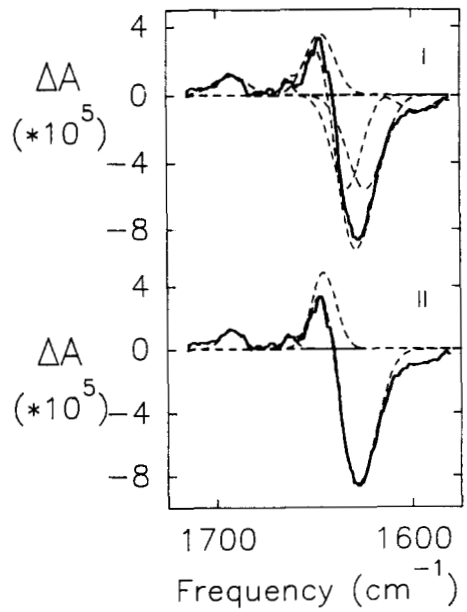


**Fig. 3.** Comparison of Fits I and II to VCD spectrum of chymotrypsin (example protein with low content of α-helix). As in Figure 1.

finding trends within the subgroups of spectrally similar proteins. A general sign/frequency pattern emerges as will be discussed below. The frequencies found in the VCD spectra also coincide (within reasonable error) with the averages of frequency ranges used by Byler and Susi (1986) for characterization of different secondary structures. There is a continuous stretch of positive VCD at 1,645–1,655 $cm^{-1}$, a continuous stretch of negative VCD around 1,625 $cm^{-1}$ and two frequency ranges with the sign discontinuities over this protein set: 1,600–1,620 $cm^{-1}$ and 1,655–1,700 $cm^{-1}$.

A closer look at the band positions, with the added characteristic provided by the VCD sign patterns, allows us to define a series of intervals (typically ±3 $cm^{-1}$ about a characteristic frequency value) that "box in" stretches of frequencies for each protein subgroup. The average frequencies found in our VCD and FTIR data as well as by Byler and Susi (1986) are listed in Table 3. Starting with the highly α-helical protein VCD of amide I′, the average central frequencies appear to be 1,645 $cm^{-1}$ (positive) and 1,657 $cm^{-1}$ (negative). The characteristic negative VCD at 1,657 $cm^{-1}$ is also present in the second cluster of proteins (medium α-helical content), while the positive feature at 1,645 $cm^{-1}$ is replaced by a pair of bands at 1,639 and 1,651 $cm^{-1}$ (averages) indicating more heterogeneity in the structure.

For the low α-helical content proteins, a broad positive VCD to high frequency is typical with components at 1,651 $cm^{-1}$, 1,645 $cm^{-1}$, and 1,660 $cm^{-1}$. This last group of bands might have been considered to be the same as the 1,657 $cm^{-1}$ band seen in high α-helical proteins, if one used FTIR frequencies alone; but, due to the VCD sign change, its source must arise from different structural elements in these low α-helical proteins.

On the low-frequency side of the positive VCD area are negative VCD peaks centered around 1,634 $cm^{-1}$ and 1,625 $cm^{-1}$. The high frequency region has additional negative VCD in frequency intervals around 1,671 and

**Table 3.** *Comparison of averaged frequencies of groups of bands obtained from the analysis of FTIR and VCD spectra of proteins in amide I′ region*

| | Band 1 | Band 2 | Band 3 | Band 4 | Band 5 | Band 6 | Band 7 | Band 8 | Band 9 |
|---|---|---|---|---|---|---|---|---|---|
| FTIR[a] | 1,693 ± 2 | 1,684 ± 3 | 1,674 ± 3 | 1,663 ± 3 | 1,653 ± 3 | 1,643 ± 3 | 1,634 ± 2 | 1,626 ± 3 | 1,609 ± 4 |
| FTIR[b] | 1,693 ± 1 | 1,685 ± 2 | 1,675 ± 3 | 1,663 ± 2 | 1,653 ± 2 | 1,645 ± 2 | 1,631 ± 3 | 1,624 ± 3 | – |
| VCD[a] | – | 1,685 ± 6 | 1,672 ± 4 | 1,663 ± 3 | 1,655 ± 3 | 1,645 ± 3 | 1,636 ± 3 | 1,625 ± 3 | 1,609 ± 5 |

[a] This work.
[b] Byler and Susi (1986).

1,680 $cm^{-1}$ which should be considered to be only tentatively identified due to their low S/N.

## Discussion

The comparisons in Table 3 show that frequencies derived from the FTIR second derivative analysis are very similar to those derived from the VCD band fits. Although expected, this is the first such demonstration of alignment of FTIR and VCD bands.

One can try to utilize the variety of samples studied within the globular protein set and the corresponding variability of the VCD spectra to go one step further from the above statement. These results can be utilized to investigate the possibilities of improving the FTIR analyses by utilization of a selection of more focused reference sets composed of groups of structurally similar proteins. For such restricted analyses there will be a better possibility of finding the homomorphism sought in the frequency–structure relationship.

To visualize our findings, the protein spectra are represented in a two-dimensional plane in Figure 4. In this representation, one dimension is given by the fraction of $\alpha$-helical conformation in the secondary structure in a given protein,[1] and the second dimension is given by the frequencies of the resolved components within the amide I′ band as derived from FTIR or VCD. This plane is then divided up into areas of positive and negative VCD sign as derived from the deconvolved (Fit I and II) VCD spectra for various positions on the plane.

It is clear from our results that the assumption that there is a homomorphous mapping of protein structure onto frequencies is an oversimplification. Figure 4 demonstrates that assigning a specific frequency to a given structure, which is the underpinning of much of modern FTIR and Raman analyses of secondary structure, can lead to a serious misinterpretation of that structure. This is made obvious in two ways, both of which arise from intrinsic properties of VCD, sign, and intensity, which arise from the structural type itself. First, the VCD signs that correlate to the IR absorption bands at 1,600–1,620 $cm^{-1}$ and 1,655–1,700 $cm^{-1}$ change as the character of the proteins changes. The bands in these regions for proteins with high and low $FC_\alpha$ thus cannot arise from the same type of structural elements even though the frequencies appear from FTIR measurement to be the same.

Furthermore, even at frequencies where the VCD sign is constant, it is clear that two or more structures contribute. For example the "continuity" of the 1,645-$cm^{-1}$ positive band region is interrupted within the second group

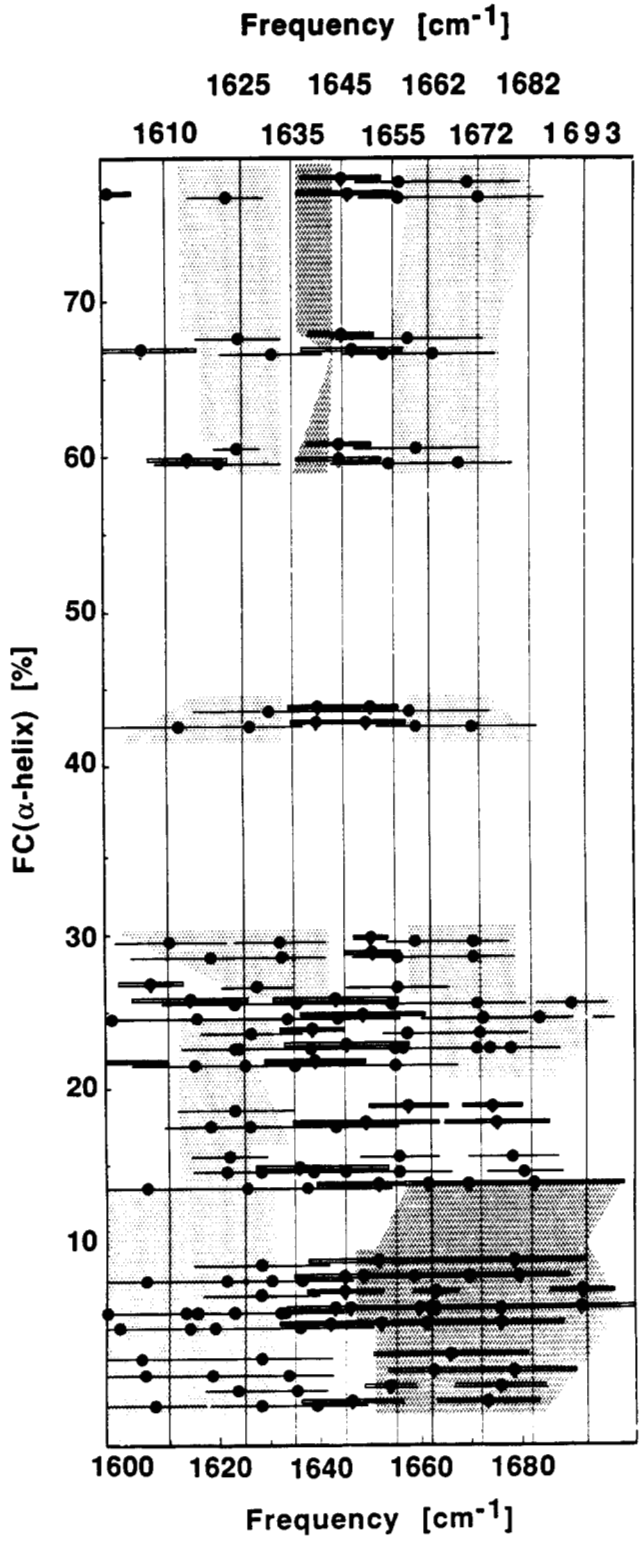


**Fig. 4.** Amide I′ frequency intervals of constant VCD sign for various globular proteins. Lower $x$-axis—frequency scale; upper $x$-axis and vertical lines—averaged positions of conformationally sensitive component bands (see Table 3). The horizontal lines represent frequency intervals corresponding to the bandwidths of positive resolved VCD bands and the horizontal rectangles are frequency intervals corresponding to the bandwidths of negative resolved VCD bands. Circles are placed at the position of component band maxima from Fits I and II. Shaded areas are "VCD-homogeneous" areas (light—positive; dark—negative).

[1] By choosing $FC_\alpha$ as the prime protein characteristic, we are able to group together proteins with similar secondary structure composition. This follows from our recent demonstration of a statistically significant relationship between the content of $\alpha$-helix and the fractions of other secondary structural types ($\beta$-sheet, bend, turn, "other") for these proteins (Pancoska & Keiderling, 1992).

of proteins that have moderate $\alpha$-helix content. Thus the positive 1,645-$cm^{-1}$ band might come from the helix for high $FC_\alpha$ proteins, but for the low $FC_\alpha$ proteins it must come from a mix of helix and other structures. As a further aid in discriminating between these two sources, it can be noted that the "canonical" $\alpha$-helical frequency of 1,651 $cm^{-1}$ (as found in most FTIR data) corresponds to two oppositely signed VCD bands that lie above and below it in frequency.

Our data clearly illustrate the virtue of using VCD spectroscopy to establish structural aspects of a protein as compared to conventional, frequency-based spectroscopic experiments. The sign variation seen at high and low frequencies establish that different structures give rise to characteristic IR frequencies that fall in comingled ranges. This has serious ramifications for the now growing use of FTIR and Raman data based primarily on frequency assignment to interpret structure. VCD signs (and magnitudes) then can help sort out the ambiguity.

Good examples of this frequency ambiguity problem have been reported. Most recently, FTIR analyses of two epidermal growth hormones, epidermal growth factor and basic fibroblast growth factor, indicated a substantial peak near 1,650 $cm^{-1}$, which was subsequently considered to be an indication of $\alpha$-helical character or ill-defined loop structures (Prestrelski et al., 1991, 1992). All other spectroscopic techniques implied that these proteins had little or no helical content. Subsequent VCD spectral analysis agreed with the results of the other techniques and demonstrated the difficulties of the FTIR analysis (Dukor et al., 1992). Another example is the coil to $\beta$-sheet transformation that characterizes the low pH denaturation of phosvitin, a highly polar glycophosphoprotein (Yasui et al., 1990). Presumably due to its atypical structure, phosvitin has dominant conformations with frequencies out of the range of those in typical proteins, but the VCD sign pattern allowed straightforward assignment of the structure. Finally, to note a much older problem in the polypeptide literature, the secondary structure of poly-L-tyrosine on change of solvent from dimethyl sulfoxide (DMSO) to DMSO:$D_2O$ transforms from coil to right-handed $\alpha$-helix as can be demonstrated using VCD (Yasui & Keiderling, 1986). Early studies attributed the frequency shift seen in this transition to the possible formation of a left-handed $\alpha$-helix, a situation eliminated by the VCD bandshape.

## Materials and methods

### *Sample preparation*

Protein samples were purchased from Sigma and were not further purified. The list of proteins used for this study is contained in Table 1 and is the same in terms of sources as that used in our earlier study (Pancoska et al., 1991). The following procedure has been adopted for sample preparation in FTIR and VCD measurements in $D_2O$ solution. About 7 mg of a dry protein was dissolved in 0.8–1.0 mL $D_2O$ in a centrifuge vial. The solution was left overnight at room temperature and then dried in a vacuum evaporator (Savant Instruments) for typically 6–8 h. The procedure was repeated twice to maximize the H-D exchange. Protein samples were then stored at subzero temperatures in vials held above molecular sieves in a closed container. For the FTIR measurements, a 5-mg/mL concentration in $D_2O$ was used. Samples were filtered prior measurement using 0.2-$\mu$m (Gelman Science) or 0.5-$\mu$m (Schleicher & Schuell) filters. The methods for obtaining VCD spectra were described in detail previously (Pancoska et al., 1989, 1991). That data set was used here directly as processed for the previous study.

### *FTIR measurements and data analysis*

Protein samples were held in a Press-Lok cell (Spectra-Tech, Inc.) composed of $CaF_2$ windows separated by a 25-$\mu$m teflon spacer. Spectra were recorded on a Digilab FTS-60, at a nominal resolution of 2 $cm^{-1}$, by averaging 2,048 scans. The $D_2O$ baseline was scanned under identical conditions. Correction for water vapor absorption was accomplished by subtraction of the $H_2O$ spectrum while monitoring the baseline for flatness over the range of 1,850–1,720 $cm^{-1}$.

Fourier transform self-deconvolution (FSD) computations were carried out for these spectra using both Digilab and SpectraCalc (Galactic, Inc.) routines. Bessel function apodization, a bandwidth of 12 $cm^{-1}$ and a resolution enhancement factor of 2 were found be reasonable parameters for the Kauppinen et al. (1981a,b) FSD analysis of these proteins in $D_2O$. They gave reasonable self-deconvolved spectra with no spurious (artificial) bands at the baseline. Second derivative spectra were calculated with the SpectraCalc routines using a 15-point window. Peak positions found by the second derivative routine and those from FSD of the same spectra were in complete agreement. Further analysis of VCD curves utilized the peak positions derived from the FTIR second derivatives.

Deconvolution of the amide I′ VCD spectra was done using two limiting approaches. Fit I was based on the assumption that there is a VCD counterpart for any band which can be resolved in the second derivative FTIR absorption spectrum of a protein. This series of calculations involved successively constrained fits of Gaussian-shaped components to the overall VCD profile. The first pass used the number of bands and their central frequencies as derived from the second derivatives of the FTIR spectra but the widths and intensities were allowed to vary in optimizing the fit to the observed VCD spectrum. In the second pass, obvious problems (unreasonable bandwidths, extremely low intensities, or exceptionally poor

representation of part of the bandshape) were addressed by fixing additional parameters, or, if necessary, by adding or eliminating bands from the fit. Efforts to use Lorentzian bandshapes for this process gave poor fits particularly due to the increased contributions in the wings arising from a Lorentzian function. Finally, in the last pass, the intensity and width parameters were held fixed, and the frequencies of the VCD component bands were allowed to vary. The center positions of the component bands obtained in this process are denoted above as the Fit I frequencies.

By contrast, in the second band fit of the VCD spectra, no link to the absorption spectra was initially assumed. The number of Gaussian component bands and their starting positions were determined empirically as the minimum number of features necessary to describe the experimental amide I′ VCD spectrum. No constraints were imposed on the band parameters, and the spectra were fit to the same level of accuracy ($\chi^2$) as with Fit I. The band centers resulting from this approach are denoted above as the Fit II frequencies.

## Conclusion

This paper seeks to investigate the reliability of frequency analysis of IR absorption data for structural interpretation of globular proteins. It establishes that VCD features can be deconvolved and shown to lie at the same frequencies seen in FTIR experiments within acceptable limits. From this and the sign variation in VCD it is thus possible to demonstrate convincingly that assignment of a given FTIR frequency to a given structural element is not unique even for the most dominant features in the spectrum such as the 1,650–1,655 $cm^{-1}$ range normally assigned to $\alpha$-helices. However, although VCD offers biophysical chemists added features for interpretation of FTIR spectra and to further characterize proteins, we feel it is premature to use frequency analysis of VCD alone for quantitative structural prediction. At present we strongly advocate the use of bandshape analysis of VCD spectra to add a more sensitive dimension to the frequency variables in current use for study for structural analyses of proteins. Finally, in general, it must be stated that use of analysis of data derived from multiple spectroscopic techniques is necessary to achieve a reliable estimation of protein structure. Only by looking at the molecule through its various intersubunit interaction mechanisms can we resolve the ambiguities inherent in all global, structurally sensitive techniques such as CD, VCD, Raman, and FTIR spectroscopies.

## Acknowledgments

This research was funded in part by a grant from the National Institutes of Health (GM 30147 to T.A.K.) and by an International Cooperation Grant from the National Science Foundation (INT91-07588 to P.P. and T.A.K.). We further thank the IBM Corporation for establishment of a computational and electronic communication facility in Prague (IBM–Czechoslovak Academic Initiative), which was instrumental in this cooperation, Charles University for facilitating the cooperation, and the NSF, NIH, and University of Illinois at Chicago for funding the purchase of the FTIR equipment used.

## References

Byler, M. & Susi, H. (1986). Examination of the secondary structure of proteins by deconvolved FTIR spectra. *Biopolymers 25*, 469–487.

Dukor, R.K., Pancoska, P., Keiderling, T.A., Prestrelski, S.J., & Arakawa, T. (1992). Vibrational circular dichroism studies of epidermal growth factor and basic fibroblast growth factor. *Arch. Biochem. Biophys. 298*, 678–681.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers 22*, 2577–2637.

Kauppinen, J.K., Moffat, D.J., Cameron, D.J., & Mantsch, H.H. (1981a). Noise in Fourier self-deconvolution. *Appl. Optics 20*, 1866–1879.

Kauppinen, J.K., Moffat, D.J., Mantsch, H.H., & Cameron, D.J. (1981b). Fourier self-deconvolution: A method for resolving intrinsically overlapped bands. *Appl. Spectrosc. 35*, 271–277.

Keiderling, T.A. (1993). Vibrational circular dichroism of proteins, polysaccharides and nucleic acids. In *Physical Chemistry of Food Processes*, Vol. 2: *Advanced Techniques, Structures, and Applications* (Bainau, I.C., Pessen, H., & Kumosinski, T.F., Eds.), pp. 307–337. Van Nostrand Reinhold, New York.

Krimm, S. & Bandekar, J. (1986). Vibrational spectroscopy and conformation of peptides, polypeptides and proteins. *Adv. Protein Chem. 38*, 181–364.

Pancoska, P., Blazek, M., & Keiderling, T.A. (1992). Relationships between secondary structure fractions for globular proteins. Neural network analyses of crystallographic data sets. *Biochemistry 31*, 10250–10257.

Pancoska, P., Yasui, S.C., & Keiderling, T.A. (1989). Enhanced sensitivity to conformation in various proteins. Vibrational circular dichroism results. *Biochemistry 28*, 5917–5923.

Pancoska, P., Yasui, S.C., & Keiderling, T.A. (1991). Statistical analyses of the vibrational circular dichroism of selected proteins and relationship to secondary structures. *Biochemistry 30*, 5089–5103.

Parker, F.S. (1971). *Applications of Infrared Spectroscopy in Biochemistry, Biology, and Medicine*. Plenum, New York.

Prestrelski, S.J., Arakawa, T., Kenney, W.C., & Byler, D.M. (1991). The secondary structure of two recombinant human growth factors, platelet derived growth factor and basic fibroblast growth factor, as determined by Fourier transform infrared spectroscopy. *Arch. Biochem. Biophys. 285*, 111–115.

Prestrelski, S.J., Arakawa, T., Wu, C.-S.C., O'Neal, K.D., Westcott, K.W., & Nahri, L.O. (1992). Solution structure and dynamics of epidermal growth factor and transforming factor $\alpha$*. *J. Biol. Chem. 267*, 319–322.

Surewicz, W. & Mantsch, H.H. (1988). New insight into protein secondary structure from resolution-enhanced infrared spectra. *Biochim. Biophys. Acta 952*, 115–130.

Susi, H. (1972). Infrared spectroscopy—Conformation. *Methods Enzymol. 26*, 445–472.

Susi, H. & Byler, D.M. (1983). Protein structure by Fourier transform infrared spectroscopy: Second derivative spectroscopy. *Biochem. Biophys. Res. Commun. 115*, 391–397.

Yasui, S.C. & Keiderling, T.A. (1986). Vibrational circular dichroism of polypeptides VI. Polytyrosine $\alpha$-helical and random coil results. *Biopolymers 25*, 5–15.

Yasui, S.C., Pancoska, P., Dukor, R.K., Keiderling, T.A., Renugopalakrishnan, V., Glimcher, M.J., & Clark, R.C. (1990). Conformational transitions in phosvitin with pH variation. Vibrational circular dichroism study. *J. Biol. Chem. 265*, 3780–3788.