

ACS SYMPOSIUM SERIES **750**

# **Infrared Analysis of Peptides and Proteins**

## **Principles and Applications**

**Bal Ram Singh, EDITOR**  
*University of Massachusetts*

fw001



**American Chemical Society**  
American Chemical Society, Washington, DC  
**Library**  
**1155 16th St., N.W.**  
**Washington, D.C. 20036**

In Infrared Analysis of Peptides and Proteins; Singh, B.;  
ACS Symposium Series; American Chemical Society: Washington, DC, 1999.

# Infrared analysis of peptides and proteins



## Library of Congress Cataloging-in-Publication Data

Infrared analysis of peptides and proteins : principles and applications / Bal Ram Singh, editor.

p. cm.—(ACS symposium series , ISSN 0097-6156 ; 750)

Includes bibliographical references and index.

ISBN 0-8412-3636-4

1. Peptides—Analysis. 2. Proteins—Analysis. Infrared spectroscopy.

I. Series. II. Singh, Bal Ram.

QD431.25 .A53 I54 2000  
543 .75046 21—dc21

99-43687  
CIP

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Copyright © 2000 American Chemical Society

Distributed by Oxford University Press

All Rights Reserved. Reprographic copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Act is allowed for internal use only, provided that a per-chapter fee of \$20.00 plus \$0.25 per page is paid to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. Republication or reproduction for sale of pages in this book is permitted only under license from ACS. Direct these and other permissions requests to ACS Copyright Office, Publications Division, 1155 16th Street, N.W., Washington, DC 20036.

The citation of trade names and/or names of manufacturers in this publication is not to be construed as an endorsement or as approval by ACS of the commercial products or services referenced herein; nor should the mere reference herein to any drawing, specification, chemical process, or other data be regarded as a license or as a conveyance of any right or permission to the holder, reader, or any other person or corporation, to manufacture, reproduce, use, or sell any patented invention or copyrighted work that may in any way be related thereto. Registered names, trademarks, etc., used in this publication, even without specific indication thereof, are not to be considered unprotected by law.

PRINTED IN THE UNITED STATES OF AMERICA

**American Chemical Society  
Library**

**1155 16th St., N.W.**

**Washington, D.C. 20036**

Infrared Analysis of Peptides and Proteins; Singh, B.;  
ACS Symposium Series; American Chemical Society: Washington, DC, 1999.

# Advisory Board

## ACS Symposium Series

**Mary E. Castellion**  
ChemEdit Company

**Arthur B. Ellis**  
University of Wisconsin at Madison

**Jeffrey S. Gaffney**  
Argonne National Laboratory

**Gunda I. Georg**  
University of Kansas

**Lawrence P. Klemann**  
Nabisco Foods Group

**Richard N. Loepky**  
University of Missouri

**Cynthia A. Maryanoff**  
R. W. Johnson Pharmaceutical  
Research Institute

**Roger A. Minear**  
University of Illinois  
at Urbana-Champaign

**Omkaram Nalamasu**  
AT&T Bell Laboratories

**Kinam Park**  
Purdue University

**Katherine R. Porter**  
Duke University

**Douglas A. Smith**  
The DAS Group, Inc.

**Martin R. Tant**  
Eastman Chemical Co.

**Michael D. Taylor**  
Parke-Davis Pharmaceutical  
Research

**Leroy B. Townsend**  
University of Michigan

**William C. Walker**  
DuPont Company

## Foreword

THE ACS SYMPOSIUM SERIES was first published in 1974 to provide a mechanism for publishing symposia quickly in book form. The purpose of the series is to publish timely, comprehensive books developed from ACS sponsored symposia based on current scientific research. Occasionally, books are developed from symposia sponsored by other organizations when the topic is of keen interest to the chemistry audience.

Before agreeing to publish a book, the proposed table of contents is reviewed for appropriate and comprehensive coverage and for interest to the audience. Some papers may be excluded in order to better focus the book; others may be added to provide comprehensiveness. When appropriate, overview or introductory chapters are added. Drafts of chapters are peer-reviewed prior to final acceptance or rejection, and manuscripts are prepared in camera-ready format.

As a rule, only original research papers and original review papers are included in the volumes. Verbatim reproductions of previously published papers are not accepted.

ACS BOOKS DEPARTMENT

# Preface

Advancement in the field of structural biology is critical to the advancement of agricultural, biomedical, and pharmaceutical science and technology. In this regard, determination of protein structure and also the correlation between structure and function form the core issues of modern molecular biology. The developments of the past decade in the increased application of IR spectroscopy have been very helpful for the analysis of structure and function relationship of peptides and proteins. However, IR spectroscopy is still not a commonly used method for this purpose, and remains an optional method for experts only. Issues related to protein IR analysis range from instrumentation and sampling methods to theoretical and experimental band assignments and derivations. To address some of these issues, an international symposium on Infrared Analysis of Peptides and Proteins was organized during the 216<sup>th</sup> Annual Meeting of the American Chemical Society (ACS), August 23–27, 1998, at Boston, Massachusetts.

Several spectroscopic techniques are available for studying protein structures, including circular dichroism, UV absorption and fluorescence, Raman, NMR, and IR spectroscopy. However, each of these techniques has its limitations. For example, UV absorption and fluorescence spectroscopic techniques are generally confined to topography of aromatic amino acids. Among the techniques for secondary structure, NMR and Raman spectroscopy need unusually high concentration of proteins, and NMR analysis is generally limited to small proteins, so far. Circular dichroism analysis is limited to clear protein solutions (as opposed to membrane proteins) due to the problem with light scattering. IR spectroscopy is perhaps the most suited technique for protein analysis because of its sensitivity, versatility, and responsiveness to structural changes.

IR spectroscopy is being widely used as a probe of protein molecular structure at the secondary level and has potential for tertiary structural level analysis. This method has been recognized as a tool for studying conformation in polypeptides and proteins since 1950 when Elliot and Ambrose found that polypeptides known to be in the  $\alpha$ -helix and  $\beta$ -sheet forms, displayed absorbance of the C=O stretching and N-H deformation modes characteristically at different frequencies.

In an attempt to find the origin of characteristic peptide vibrations, Miyazawa and co-workers in the 1950s determined the chemical groups that give rise to the characteristic amide vibrations using peptide analogues. The amide I, II, and III vibrational modes were identified to be localized within the peptide CONH group, and became the focus of most attempts to correlate polypeptide and protein secondary structure with IR spectral properties, with the amide I mode being used the most because of its relative

strength. In the 1960s and 1970s, experimental studies by Timasheff, Susi, and others and theoretical normal mode calculations by Shimanouchi, Krimm, and others solidified our understanding of IR spectra–structure correlations of amide bands of polypeptides and proteins. These studies showed that amide vibrational frequencies associated with different conformations in model systems were roughly transferable to proteins, in general, for the  $\alpha$ -helix,  $\beta$ -sheet, and unordered conformations. The information from these studies and theoretical calculations relating spectral characteristics to protein conformations were used to predict the structural composition of globular proteins.

The advent of modern Fourier transform infrared (FTIR) spectrometers has led to significant advances in IR spectroscopy, particularly in protein conformational analysis. The digital nature of FTIR spectroscopic data, combined with high signal-to-noise ratio, also allows the use of “resolution-enhancement” techniques. Derivative spectroscopy and Fourier self-deconvolution (FSD) enhance the spectroscopist’s ability to visualize component peaks in multicomponent, overlapping bands such as those arising from conformation-sensitive modes in protein spectra. These approaches along with curve-fitting process, promoted by Chapman, Mantsch, Susi, and others, have been exploited extensively in the 1980s and 1990s to obtain reasonable estimates of secondary structure contents of peptides and proteins. Current FTIR methods of determination of protein conformation involve analysis of the amide I and amide III bands employing curve-fitting methods for the component peaks disclosed by resolution enhancement techniques. In addition, statistical methods of multifactor regression analysis are now being employed to extract secondary structure contents from amide I, II, and III spectra.

Within the past few years, there has been tremendous development in several areas that are relevant to the effective use of IR spectroscopy to analyze proteins. These include: availability of modern computers, interfaces of FTIR spectrometers to multiple variables (temperature, pressure, and titrators), different sampling techniques; optical fiber probes, chemometric approaches, fast data collection methods for time-resolved analysis, and polarization methods. Additionally, great advances have occurred in the biotechnology of sample preparations such as isotopic labeling of proteins, which allows fine-tuning of protein signals and assignments. While theoretical calculations of IR band positions are still less than adequate for use by themselves, advancement in the field at least has been complementary to experimental results. It is imperative that these advances (1) continue and be developed synergistically, and (2) be communicated to user scientists who are in dire need of structural information about proteins. The symposium and the resulting symposium volume monograph are steps toward these goals.

Although this book by no means represents all the work in the field that has been carried out or is in progress, it has captured the essence of the basic foundations as well as recent advances in the application of IR spectroscopy for the analysis of peptides and proteins. I feel that there is a need for a volume like this to apprise researchers in the area of protein structure and functions of the possibilities that are currently available with IR spectroscopy. This observation was in part supported by the fact that the symposium itself was very well attended. Notably, many attendees were from the industrial research groups.

The volume has chapters ranging from the theoretical basis of the protein IR signals to the application of IR spectroscopy to resolve biological events at nanosecond scale. The chapters are written to include basic introductions of the subject matter, and have extensive examples of applications with graphics and references. The volume could make a reference book for an advanced biological spectroscopy course at graduate level.

## **Acknowledgments**

Organizing a symposium of this nature, and the preparation of a symposium volume, requires coordination of efforts by several individuals. I am very fortunate to have the assistance of very capable people like Henry Blount and Ed Yeung, who initially encouraged me with this project, and provided guidance on the details of the symposium program. J. Michael Ramsey was instrumental in finalizing the program, and along with Michele Buchanan, in arranging for support from the ACS Division of Analytical Chemistry. The ACS Petroleum Fund and the National Cancer Institute of the National Institutes of Health (R13CA79103) provided critically needed funds to support symposium speakers, and the University of Massachusetts at Dartmouth provided the administrative support.

Preparation of the book was coordinated by Anne Wilson and Kelly Dennis of the ACS Books Department. Their patience and assistance in maintaining the timeline for getting the manuscripts processed were invaluable to the completion of this volume. Finally, I am very grateful to the contributors of the volume, who willingly accepted the deadlines and patiently made changes suggested by a dedicated group of reviewers of the chapter manuscripts.

BAL RAM SINGH  
Department of Chemistry and Biochemistry  
and Center for Marine Science and Technology  
University of Massachusetts at Dartmouth  
Dartmouth, MA 02747

## Chapter 1

# Basic Aspects of the Technique and Applications of Infrared Spectroscopy of Peptides and Proteins

Bal Ram Singh

Department of Chemistry and Biochemistry, and Center for Marine Science and Technology, University of Massachusetts at Dartmouth, Dartmouth, MA 02747

Infrared spectroscopy is being increasingly utilized for the analysis of peptides and proteins because it probes the universally available amide (peptide) bonds, which display distinct IR signals for differently folded peptides and proteins. Other spectroscopic techniques useful for studying protein structures in solutions are circular dichroism (CD), ultraviolet absorption and fluorescence spectroscopy, Raman and nuclear magnetic resonance (NMR). Among the techniques for secondary structure, NMR and Raman spectroscopy need unusually high concentration of proteins, and NMR analysis is still limited to small proteins of about 200 amino acid residues. CD analysis is limited to clear protein solution (as opposed to membrane proteins) due to the problem with light scattering. Furthermore, for estimation of secondary structure of protein by CD accurate protein concentration is needed. Infrared spectroscopy is an emerging technique for protein analysis. Amide I, II and III are most commonly used IR spectral regions used for protein structure-function analysis. Recent advances in the development of instrumentation (Fourier transformation, sampling), protein IR data bank (band assignments to different components of secondary structure), and techniques (two-dimensional IR methods, time-resolution, and isotopic labeling) have significantly augmented IR spectroscopy as an analytical tool for peptides and proteins. This chapter provides an overview of the basic technique and some of the applications of IR spectroscopy to examine structure, interaction, and conformational changes in peptides and proteins.



## Techniques of Protein Structure Analysis

Proteins and polypeptides are a set of complex macromolecules, and understanding their structure-function relationship is important but a difficult challenge to the scientific community. While tremendous progress has been made in identifying functional properties of proteins, structural information on these proteins has been very limited. To understand complete structure of a protein, one would have to know its primary amino acid sequence along with its folding patterns at the secondary and tertiary structure level. Methods for the secondary and tertiary structural analysis include circular dichroism (CD), infrared spectroscopy, Raman spectroscopy, nuclear magnetic resonance (NMR) spectroscopy, and x-ray diffraction studies of protein crystals. x-ray diffraction studies of protein crystals are considered the ultimate structural analysis of a protein. However, obtaining protein crystals suitable for x-ray diffraction studies is still a difficult task. In addition, even if a protein crystal structure can be estimated in one condition, it is not practical to prepare protein crystals in several physiologically relevant conditions to analyze structure-function relationship. There are two other aspects of protein structures where x-ray crystallography studies become limited. (i) The crystallization conditions could alter the protein structure. (ii) The information on the protein dynamics is not directly available. Nuclear magnetic resonance spectroscopy is now being used as an alternative method for the protein structure analysis but its application has been still limited to relatively small proteins and polypeptides (approximately 200 amino acid residues) even with multiple isotope labeling (1). In addition, high concentrations of proteins are needed for NMR analysis. Circular dichroism is one of the most commonly used spectroscopic techniques for protein structure analysis but it has its limitation in its utility over a narrow range of concentration, and only optically clear solution can be used for CD analysis. In addition,  $\beta$ -sheet and random coil structures have relatively small CD signals thus introducing chances of error in their estimation.

## IR Spectroscopy of Proteins and Peptides

Fourier transform infrared (FT-IR) spectroscopy has been widely used to study the secondary structure of proteins in recent years (2-5). Proteins or polypeptides have a continuous chain of amino acids connected via amide bonds. The frequencies at which amide bond vibrations occur can be attributed to different secondary structures in which the amide bonds are present. The differences in vibration of the amide bonds are due to differential hydrogen bonding among amino acid residues. For example,  $\alpha$ -helix and  $\beta$ -sheet foldings have ordered hydrogen bonding, although differing in their patterns. The differential pattern in H-bonding, along with geometric orientations, of amide bonds in  $\alpha$ -helix,  $\beta$ -sheet and random coil structures allows the different vibration frequencies associated with individual secondary structural foldings. Amide vibrations involve C=O, C-N and N-H groups of a peptide bond (amide bond), which result in characteristic spectral features of proteins. Three major spectral regions (amide I, amide II and amide III) have been identified based on theoretical and experimental studies (4-14).

The amide I vibration region, (1700-1600  $\text{cm}^{-1}$ ), has been widely used due to its strong signal exhibited by proteins in this region. However, there are several difficulties encountered in this region when analyzing protein spectra. The first difficulty encountered in this region is the OH vibrations caused by liquid water and water vapor that need to be subtracted out of the protein solution spectrum. Another problem is the difficulty of peak assignments. Since there is a great deal of overlapping of peaks representative of the different secondary structures, it is difficult to assign bands to their correct structure. For example, bands at 1650-1655  $\text{cm}^{-1}$  could be assigned either to  $\alpha$ -helix or random coils. Using the amide III region (1220-1320  $\text{cm}^{-1}$ ), many of these problems are resolved. In the amide III region, OH vibrations due to water do not interfere with spectrum in the amide III region as much as in the amide I region. The overlapping of bands arising from different secondary structure of a protein is not significantly encountered in this region. In the amide I region, the frequencies at which the different amide bond vibrations occur are not as localized as they are in amide III. In the amide III region, various spectral bands are more resolved in the original protein spectrum than they are in the amide I spectrum. This fact allows a greater ease in peak definition as well as in peak assignment. The only drawback in using amide III is that the protein signal is significantly weaker than the signal obtained in the amide I region (Fig. 1). The lack in signal strength, however, does allow for a more resolved protein spectrum facilitating curve analysis (see below). The amide II protein band is not as sensitive as the amide I and III bands to variation in secondary structure content. Although amide II band by itself is not generally used for estimation of secondary structure, its inclusion with amide I band increases the accuracy of secondary structure prediction in statistical regression methods (15-17).

**Data Collection.** Because IR spectroscopy is largely not affected by light scattering, protein and peptide samples either in solid or liquid form can be used for data collection. Two types of sample accessories are typically used: windows such as calcium fluoride, and attenuated total reflectance (ATR) accessory, generally made of zinc selenide or germanium. Windows, usually of several micron pathlength, are used for recording transmission spectra. The pathlength of the transmission windows are kept small to avoid high absorbance by the aqueous solvent. In transmission mode, high concentration of proteins (> 1mg/ml) are required to obtain accurate protein spectra. For protein adsorption studies, or for studies with solid powder and thin films, ATR accessory is most commonly used (5,18,19) for sampling. One of the advantages of ATR technique in recording protein spectrum is the avoidance of solvent interference in IR spectra, because it limits the effective sample thickness to a thin layer near the surface of an internal reflection element (crystal) (20).

The ATR accessory is generally mounted in the sample compartment of the FT-IR instrument. There are two configurations of ATR accessory commonly used with IR spectrometers: circle cell and horizontal ATR. In circle cell ATR, zinc selenide or germanium crystals are shaped as a cylinder with a narrow passage in the middle for sample flow. The IR radiation is launched in the crystal such that it creates evanescent radiation at the walls of the narrow passage (see below for evanescent wave principle). The horizontal ATR accessory, on the other hand, is a flat rectangular configuration, where the evanescent wave is generated at the open flat surface of the crystal. A

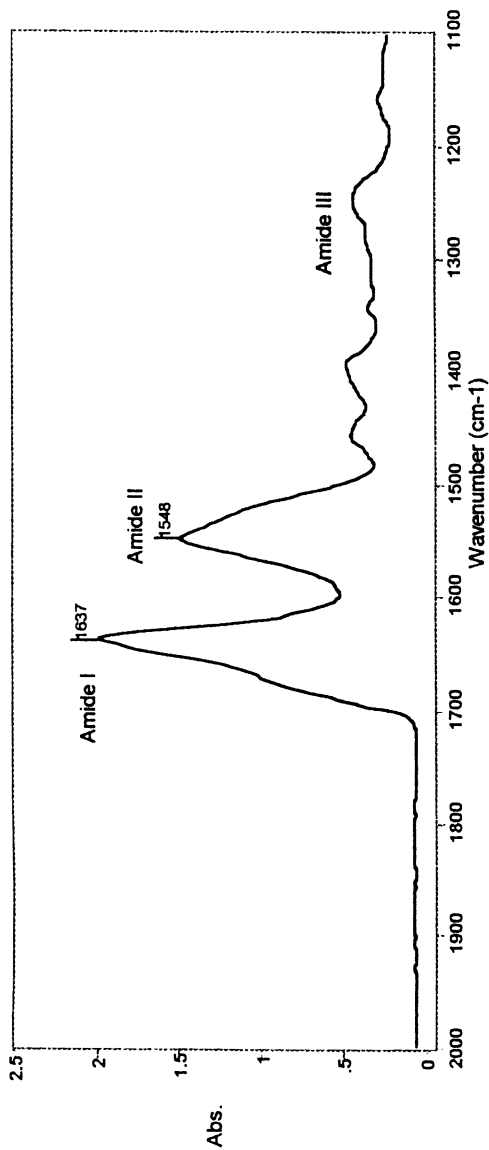


Figure 1. IR spectrum of  $\alpha$ -chymotrypsin. The amide I region (1600-1700 cm<sup>-1</sup>) corresponds to the C=O stretch weakly coupled with C-N stretch and N-H bending. The amide II region (1500-1600 cm<sup>-1</sup>) represents C-N stretch strongly coupled with N-H bending. The amide III region (1200-1350 cm<sup>-1</sup>) is N-H in-plane bending coupled with C-N stretching and also includes C-H and N-H deformation vibrations.

schematic diagram of a horizontal ATR accessory is shown in Figure 2. From practical point of view of cleaning the ATR crystal after each set of experiment, horizontal ATR is more convenient due the accessibility of the surface for cleaning.

At the heart of the ATR accessory is a crystal of infrared transparent material with high reflective index. Typical materials used are zinc selenide, KRS-5 (thallium iodide/thallium bromide), and germanium. The IR beam is directed into the face of the crystal. With angle of light incidence higher than the critical angle, the radiation undergoes the phenomenon of total internal reflection the top surface of the crystal. The critical angle of incidence ( $\theta_c$ ) depends on the refractive index of the crystal material

$$\theta_c = \sin^{-1} n_2/n_1$$

Where,  $n_1$  and  $n_2$  are the refractive indices of dense and rare media, respectively.

The IR radiation is then reflected off both bottom and top surfaces of the crystal many times before exiting. A standing wave of radiation is set up at every reflection point, and is referred to as evanescent wave. The evanescent wave travels beyond the surface of the crystal, so it can interact with matter layered on the surface of the crystal. A protein sample brought into contact with the crystal can interact with the evanescent wave, absorb infrared radiation, and can have its infrared spectrum detected. The evanescent wave is attenuated due to absorbance by the sample, reducing the amount of reflected light, and hence the name attenuated total reflectance.

The depth that the infrared radiation penetrates into the space beyond the surface of the crystal (or into the sample) is known as the depth of penetration ( $d_p$ ), and is analogous to the idea of pathlength in transmission sampling techniques. Depth of penetration is defined as the depth at which the evanescent wave is attenuated to 36.8% (1/e) of its total intensity.  $d_p$  is given by the following equation:

$$d_p = \frac{1}{2\pi W N_c (\sin^2 \theta - N_{sc}^2)^{1/2}}$$

where,  $d_p$  = depth of penetration;  $W$  = the wavenumber;  $N_c$  = the crystal refractive index;  $\theta$  = the angle of incidence;  $N_{sc} = N_{\text{sample}}/N_{\text{crystal}}$

Certain notable aspects of ATR crystal are as follows: The depth of penetration is dependent on the wavenumber. The  $d_p$  decreases as the wavenumber increases. Thus low wavenumber light penetrates farther into the sample than the high wavenumber light. As a result, ATR spectra show peaks that are more intense at low wavenumbers than at high wavenumbers. Second, the depth of penetration decreases as the refractive index of the crystal increases. Thus, a germanium crystal with refractive index of 4.0 has a significantly shallower depth of penetration than a ZnSe crystal with refractive index of 2.5.

For the structural analysis of peptides and proteins in aqueous solution, typically a concentration of 1 mg/ml solution in an appropriate buffer is prepared. While instruments with high sensitivity (such as with MCT detector) and resolution (less than  $1 \text{ cm}^{-1}$  wavenumber) are always better, in most cases an instrument with

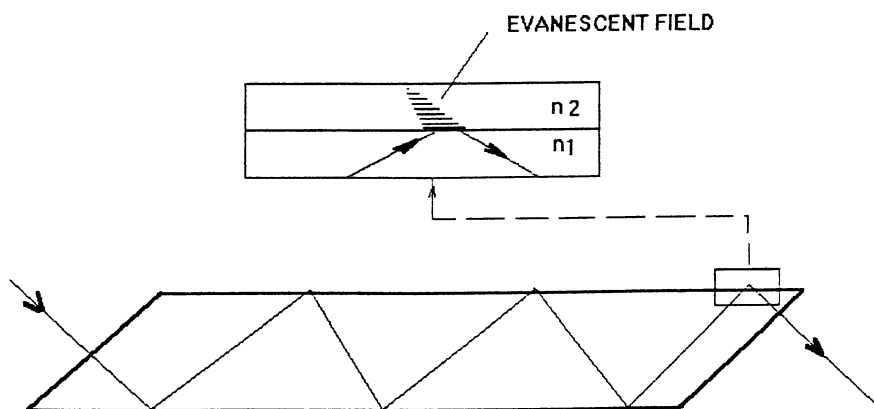


Figure 2. Schematic diagram of light undergoing multiple reflections in an ATR crystal.  $n_1$  is the refractive index of the crystal (dense medium) and  $n_2$  is the refractive index of the sample (protein solution, rare medium).

resolution of up to  $4\text{ cm}^{-1}$  equipped with a DTGS-based detector is adequate. The spectrometer is generally purged with  $\text{CO}_2$ - free dry air, or nitrogen to remove water vapors, which interfere with the protein bands. Typically, a total of 256 to 1024 scans are coadded to obtain the final spectrum. Prior to obtaining the protein spectrum, an open beam background spectrum through the clean ATR crystal is recorded. This step is followed by recording the spectrum of the buffer solution. Subsequently, the spectrum of the protein solution is obtained in the same manner.

The spectra (generally recorded in transmission mode) of the buffer and protein solutions are separately ratioed against the background spectrum, and are then converted into absorbance spectra. To obtain the protein spectrum, the buffer spectrum is subtracted in an iterative manner until a straight baseline in  $\sim 2500\text{-}1700\text{ cm}^{-1}$  spectral region is reached. The iterative subtraction for following disappearance of the water band at  $2200\text{ cm}^{-1}$  in the protein signal silent region of  $1800\text{-}2500\text{ cm}^{-1}$  is an effective approach to derive pure protein spectrum. The difference spectrum thus obtained is then smoothed typically with the use of a 9 point Savitsky-Golay algorithm.

**Spectral processing.** Protein band positions are identified after spectral processing by Fourier self-deconvolution and/or second-order derivatization. We routinely use both deconvolution and second derivatization methods to ensure the fidelity of each method (21) A variety of software packages are available commercially. In case of Fourier self deconvolution, varying half-bandwidths typically in the range of  $15\text{-}30\text{ cm}^{-1}$  and enhancement factors, commonly referred to as K-values, typically in the range of 1.5-3.5 are used to obtain proper spectral band resolution. Second-derivative spectra are obtained typically with a five-data-point window, to confirm the initial identification of the band positions by deconvolution.

The data obtained from the deconvolved and second-derivative spectra are used to determine the number of bands and their positions in order to resolve the protein spectrum into their components. This is accomplished with a curve-fitting process employing computer softwares commercially available from the manufacturers of IR instruments. The softwares resolve the original protein spectrum to individual bands that fit the spectrum. Two main parameters control the fitting process: (1) the individual bands, and (2) the baseline position. Each individual band in turn is controlled by three parameters: (1) the height of the band, (2) the position of the band (wavenumber), and (3) the bandwidth at half-height. The program iterates the curve-fitting process, and each iteration flows (increases or decreases) each parameter (height, bandwidth, position, and baseline) to determine individual parameters in order to achieve the best Gaussian, Lorentzian, or a mixture of Gaussian/Lorentzian-shaped curves that fit the original protein spectrum. A best fit is determined by the root mean square (rms) of differences between the original protein spectrum and the strength of individual bands. In order to estimate the strength of individual band, the band area or intensity is used to calculate the relative contribution of each band(s) to a particular secondary structure of the protein.

Previous studies have utilized curve analysis methods that use Fourier self-deconvolution, second derivative and curve fitting for analysis of the individual bands. The methods of analyzing the protein spectrum via curve analysis are, however, varied. Dong *et al* (14) have used both self-deconvolution followed by the second derivative of the enhanced spectrum for determining the structure assignments and for estimating

the % of each structure present in the protein. Byler and Susi (11) used self-deconvolution of the spectra followed by curve fitting on the deconvolved spectra to estimate secondary structure amounts. The approach employed by us is to use the deconvolved and second derivative spectrum for peak assignment, followed by curve fitting on the original protein spectrum. This method has allowed us to assign bands based on the original spectrum and fit Gaussian curves to the protein spectrum.

Another approach employed more recently for extracting secondary structure component estimation is based on statistical analysis. Partial least square (PLS) method has been used by recording IR spectra of a set of calibration proteins of known secondary structures from x-ray crystallography, and then fitting the data by least square regression to obtain prediction of secondary structure content of unknown proteins (15,22; Chapter 5). A multivariate matrix data analysis incorporating single value decomposition (SVD), PLS, inverse least squares (ILS), and ridge regression with a calibration set of the IR spectra in amide I and II regions for 39 standard proteins has been successfully employed to obtain prediction of  $\alpha$ -helix and  $\beta$ -sheets (16). More recently, Keiderling and coworkers (23) have developed quantitative tools based on factor analysis of spectral band shapes and subsequent restrictive multiple regression analyses of the loadings to predict average fractional secondary structure of proteins optical spectra, including IR data. Their focus has been electronic CD, and vibrational CD. They have found that VCD analysis provides a better reliability than normal FT-IR spectral analysis. Wi et al. (17) have recently used the effect of Fourier self deconvolution factors of the amide I and II protein spectra on the improvement of predictability of factor analysis based restricted multiple regression methods. These statistical methods have achieved significant degree of accuracy, and are welcome developments. Since they do not require spectral processing, and softwares are easily available, these methods have potential to be more commonly used. As is true with band assignment methods, large calibration sets are needed for statistical methods also to provide broad structural variations.

**Sensitivity of IR Spectroscopy of Proteins.** Sensitivity of a spectroscopic technique is a critical issue for protein structure analysis for several reasons. (i) In many cases only small quantities of purified proteins are available. (ii) Peptides and proteins are insoluble at high concentrations. (iii) The structure of proteins changes at high concentrations. Of the spectroscopic techniques available for protein analysis, fluorescence spectroscopy is considered a very sensitive technique, primarily due to the high extinction coefficient of the electronic transitions, and the sensitivity of the excited state to the environment of the chromophore. However, fluorescence provides only selective information of protein tertiary structure, and not on the secondary structural folding. Techniques used for secondary structure estimation, NMR and Raman spectroscopy require very high concentrations (over 20-30 mg/ml). Far UV circular dichroism, a very commonly used technique for protein secondary structure, normally requires a minimum concentration of 0.1-0.2 mg/ml of proteins. In a direct comparative study, Singh and Fuller (24) analyzed immunoglobulin G (IgG) with fluorescence, CD and FT-IR techniques. At a concentration of 2  $\mu$ g/ml, neither the far-UV CD signal nor the Trp fluorescence signal was detectable (Fig. 3). However, IR signal was clearly observed in the amide I region (Fig. 4). We were able to not only

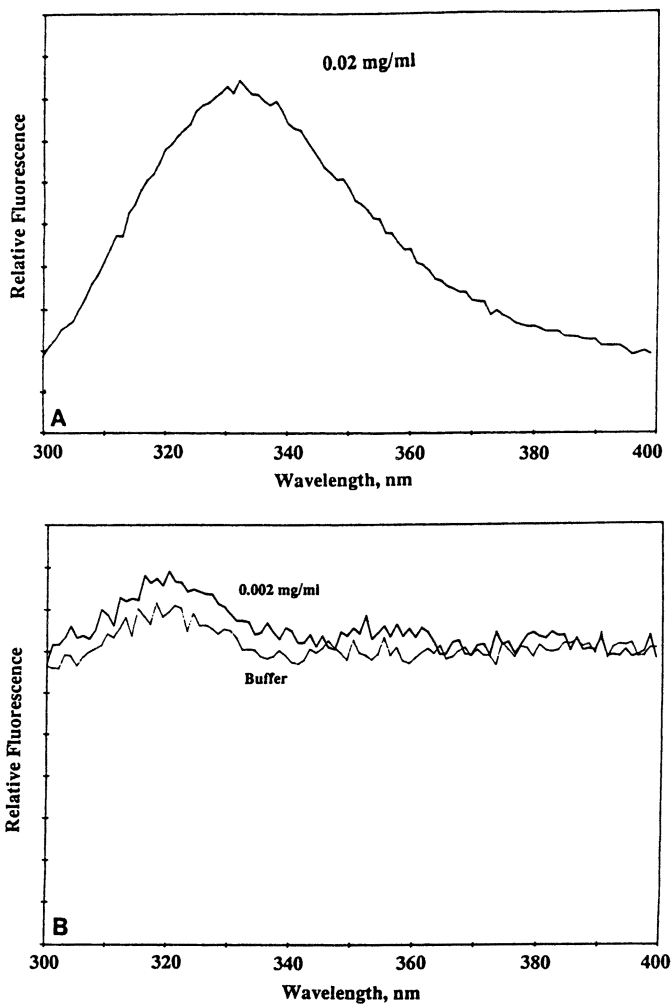
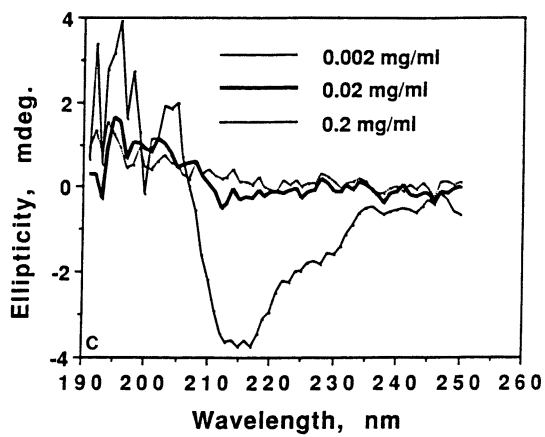


Figure 3. (A) Intrinsic fluorescence spectrum of 0.02 mg/ml IgG solution. (B) Intrinsic fluorescence spectrum of 0.002 mg/ml IgG solution, along with recording of control spectrum of the buffer (10 mM sodium phosphate buffer, pH 6) used to dissolve the protein. (C) Far UV-CD spectra of 0.002, 0.02 and 0.2 mg/ml IgG solutions. (Reproduced with permission from reference 24. Copyright 1991.)



Figure 3. *Continued.*

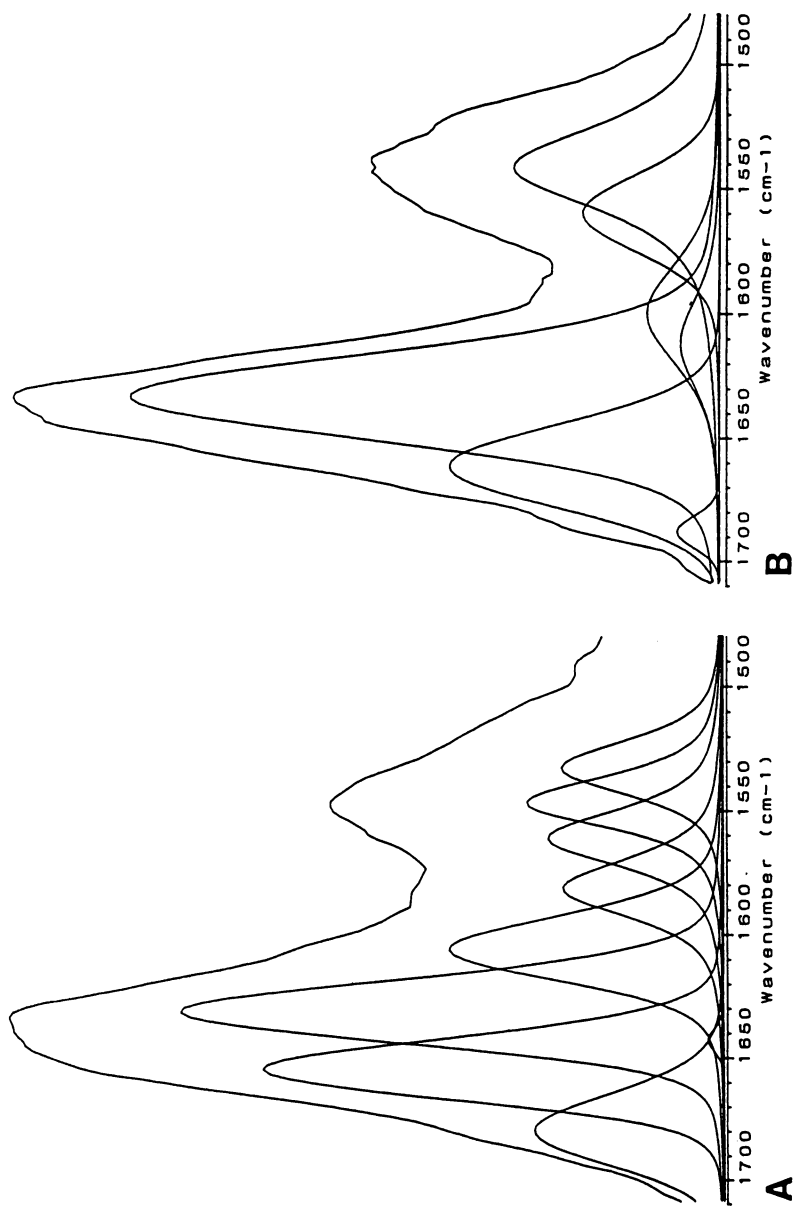
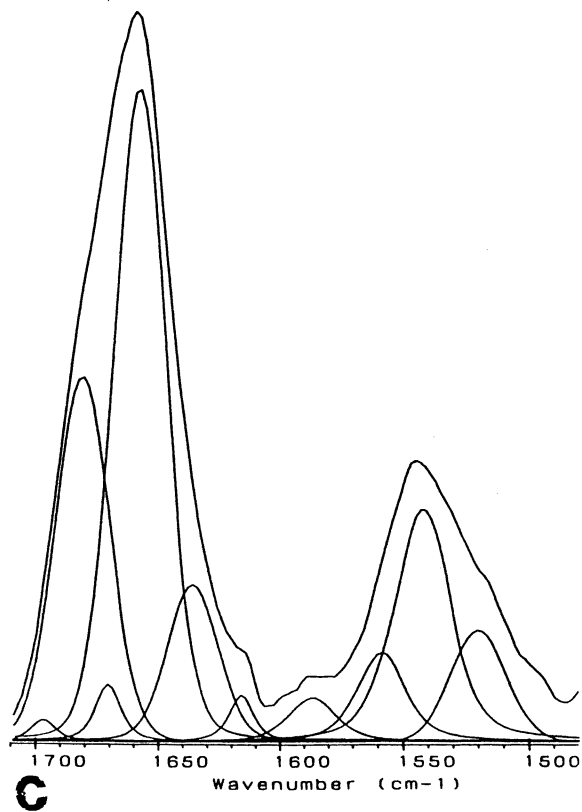


Figure 4. Curve-fitted infrared spectrum of (A) Lysozyme, (B) IgG, and (C) lysozyme, exposed to methanol, obtained after adsorbing the proteins from 0.002 mg/ml solution. (Reproduced with permission from reference 24. Copyright 1991.)

ch001

Figure 4. *Continued.*

carry out a curve-fitting analysis of the spectra, but also were able to observe effect of organic solvent (methanol) on lysozyme (Fig. 4). Methanol has been used in the past to perturb protein structure, and is known to increase  $\alpha$ -helical structure (25,26).

Possible reasons for higher detectability of proteins by IR-ATR spectroscopic approach are two-folds: (i) Amide bonds have higher extinction coefficients for vibrational transitions observed in IR absorption than for the electronic transitions observed in circularly polarized UV light absorption. Far UV-CD signals are used for secondary structure estimation. (ii) In ATR crystals used for recording IR spectra, the light penetrates only a fraction of a micron into the sample, which enhances the effective pathlength for the adsorbed protein. Furthermore, use of Fourier approach allows a large number of scans at a relatively high energy throughput, which results in a high signal to noise ratio.

In sensitive spectroscopic techniques such as protein fluorescence, the signal can be observed only from a limited number of chromophores (viz., aromatic amino acids). Only Trp and Tyr have significant fluorescence yield. For IR spectroscopy, on the other hand, every amide bond is a chromophore. IR spectroscopy can therefore be used for structural analysis of proteins, which have no Trp or Tyr, or have chromophores that are not efficiently fluorescent.

Use of FT-IR/ATR for conformational analysis of low concentration of proteins can be extremely important to biochemists and molecular biologists who have minute amounts of proteins purified or who have produced small amounts of proteins by genetic engineering. In addition, for proteins or protein fragments which are not readily soluble in aqueous solvents, FT-IR/ATR provides a useful means of analyzing their structure. The FT-IR/ATR technique also provides a way to analyze the structure of proteins at higher concentrations, which is not readily possible with other commonly used techniques such as CD and fluorescence. A comparison of protein structures at high concentrations (such as crystals or high concentration aqueous solutions) vs. low concentration aqueous solutions may provide information concerning the relationship between the biological activity and the structure. FT-IR spectroscopic analysis is the only approach currently available, which allows structural analysis of proteins with such an extreme variation in protein concentration.

One major drawback with the use of ATR sampling method is the possibility of structural change in the protein introduced by its adsorption to the ATR crystal. It is notable that the spectrum collected on ATR crystal is mostly due to the adsorbed protein (27), although a significant fraction can be from the soluble protein molecules. In our experience this behavior depends on the protein and the solvent conditions of the protein solution, and on the ATR crystal. We compared the ATR (ZnSe) and transmission spectra of lysozyme, and did not find any significant differences in the spectral features (24). On the other hand, type A botulinum neurotoxin showed considerable change in its  $\alpha$ -helical and  $\beta$ -sheet content (28). The problem is also dependent on the protein concentration used. Adsorption-induced changes are more acute at low protein concentration. In order to solve the problem, one has to compare transmission and ATR spectra to ensure adsorption does not introduce significant alteration in the protein structure. Another remediation to this problem, which allows correction of ATR spectra for the contribution of adsorbed proteins, has been recently suggested by Fink and coworkers (29).

**Surface Protein Adsorption Analysis.** Adsorption behavior of proteins on various surfaces is an important area of research because of its relevance to biotechnology and medicine (30). Important aspects of protein adsorption on surfaces include quantitation of adsorbed protein, kinetics of adsorption, and structural and functional changes in proteins upon adsorption. FT-IR spectroscopy in combination with ATR technique has been utilized to characterize adsorbed proteins (3,5,18,24,30,31). Fu et al. (18) employed FT-IR spectroscopy for the first time to analyze the monolayer to multilayer adsorption transition of proteins. FT-IR spectroscopy particularly has proven to be effective in examining the adsorption kinetics of proteins, because a wide range of protein concentrations could be analyzed on ATR crystals. Detectability of low concentration of proteins allowed calculation of adsorption density of proteins at very low concentrations (18). For example, at 0.5  $\mu\text{g/ml}$  concentration, the adsorption density of lysozyme was estimated as 10 picomoles/ $\text{cm}^2$  (monolayer), whereas at 8  $\text{mg/ml}$  concentration the adsorption density was 479 picomoles/ $\text{cm}^2$  (multilayer). Adsorption isotherms were observed to be sensitive to pH. In case of lysozyme, maximum adsorption was observed at pH near its pI (Fig. 5), which suggested denser packing of the protein molecules at that pH.

#### **Application of protein adsorption analysis by IR spectroscopy.**

Adsorption of proteins to solid surfaces is critically important to biocompatibility of artificial organs (31). An example of adsorption process involved under physiological conditions is that of lipoproteins. Lipoprotein adsorption to surfaces is not only relevant to biocompatibility and artificial biomaterial, but also to certain diseases and to their biological functions (32,33). Adhesion of LDL to arterial walls leads to plaque formation which then could cause atherosclerosis. On the other hand, adsorption of HDL on cell surfaces is involved in one of its functions to clean up the cholesterol from the arterial walls. HDL and LDL binding to arteries and other tissue surfaces is also critical to the cholesterol transportation and metabolism. In addition, their interaction with surfaces of artificial organs is of prime importance in deciding on the suitability of such organs. Therefore, an understanding of macromolecular adhesion process is of critical importance to several disciplines. Most of the studies performed have involved the binding of lipoproteins to receptors located on cell membranes, and have been aimed at a better understanding of the binding characteristics of LDL and HDL to their receptors (34-36). HDL has been found to inhibit atherogenesis by regulating the binding of LDL to cell receptor molecules (37). It has been observed that the means of inhibition is through competitive binding or steric hindrance of LDL by HDL, suggesting a common basis of HDL and LDL binding to the cell surface. One study suggested that the interaction of HDL and free cholesterol on the cell surface is the major factor of HDL-cellular association (38). Therefore, relative adsorption behavior of LDL and HDL are of interest to their physiological functions.

As mentioned above, adsorption of proteins has been investigated in the past, but several questions on various aspects of the adhesion process remain unanswered. Previous studies have only involved isolated proteins without consideration of the presence of other macromolecules in the surrounding (37,38). Role of the surrounding media in the protein adsorption is especially applicable to HDL and LDL, where proteins are always tightly surrounded by lipids. This problem has not been solved in

the past, because of the signal interference of other macromolecules in the study of the protein adsorption. Infra-red spectroscopy has successfully been used to quantitate adsorbed proteins (5,18,37,38) and to estimate the secondary structure of proteins (39-41). In particular, FT-IR in combination with ATR has been utilized to characterize adsorbed proteins (5,18,42).

In a recent study, we investigated the HDL and LDL adsorption behavior in terms of kinetics and isotherms to analyze their cholesterol binding properties using FT-IR with ATR accessory. Adsorption isotherms revealed monolayer to multilayer adsorption transitions of HDL and LDL (19). These results also indicated significant differences in the adsorption behavior of HDL and LDL, which was affected considerably by cholesterol (Fig. 6).

### Protein Secondary Structure by Amide I vs. Amide III

The amide I vibrational region is widely used because of the intense protein signal. However, several problems arise when one is attempting to utilize it for secondary structure analysis. First, water IR band ( $1640\text{ cm}^{-1}$ ) interferes with the protein amide I band at  $1650\text{ cm}^{-1}$ . Second, due to the serious overlapping of the random coil band and the  $\alpha$ -helix band, the analysis needs to be carried out in  $\text{D}_2\text{O}$ . Although the bands are consequently separated, uncertainty in the NH/ND exchange process causes a certain degree of ambiguity (43). While the intensity in the amide III region is relatively small, it does not have interfering OH vibrations from water. It also shows a relatively discrete spectral feature and has been used as a direct qualitative indicator for conformational change in proteins (3,12,41,44,45). These benefits suggest that utilization of the amide III region should be considered as a complementary, if not an alternative, method for protein structural analysis. Developing a clear assignment for protein secondary structure using amide III, therefore, becomes an urgent need. Several model proteins have been analyzed by Fu et al. (39) to build the correct assignment matrix in the amide III spectral region for the structural analysis of proteins.

**Example of Proteins.** Distinct infrared spectra of proteins arise because of the vibrational modes of atoms in the amide bond. The differential vibrational mode of atoms in an amide bond exhibit discrete spectral bands (10). Figure 1 shows a typical protein IR spectrum with amide I, II, and III spectral regions. Spectral features, especially in amide I and amide III regions are distinct for proteins with primarily  $\alpha$ -helical or  $\beta$ -sheets foldings.

**$\beta$ -Sheet proteins.**  $\alpha$ -chymotrypsin and immunoglobulin G are both primarily  $\beta$ -sheet proteins with little  $\alpha$ -helix structure. The amide III spectra of these two proteins reveal important information concerning the secondary structural folding, even without any curve analysis. The original spectrum of each of these proteins exhibits a large peak at  $\sim 1240\text{ cm}^{-1}$  and a much smaller peak at  $\sim 1317\text{ cm}^{-1}$  (Fig. 7). According to band assignments of Singh et al. (21), and Kaiden et al. (12), these peaks would be characteristic of  $\beta$ -sheet and  $\alpha$ -helix structure, respectively. The band intensities give a good representation of the high  $\beta$ -sheet and low  $\alpha$ -helix characteristics. A survey of these two spectra indicates that both  $\alpha$ -chymotrypsin and

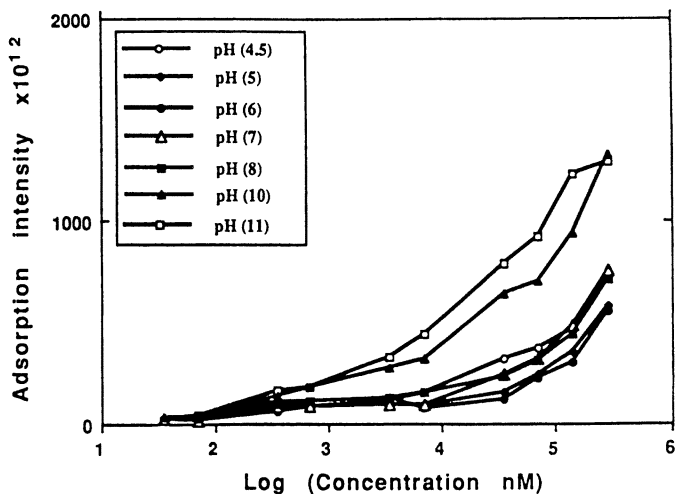


Figure 5. Adsorption isotherm of lysozyme onto a ZnSe ATR crystal, under different pH conditions. A better packing is observed at pH 11, which is the pI of lysozyme. (Reproduced with permission from reference 18. Copyright 1993.)

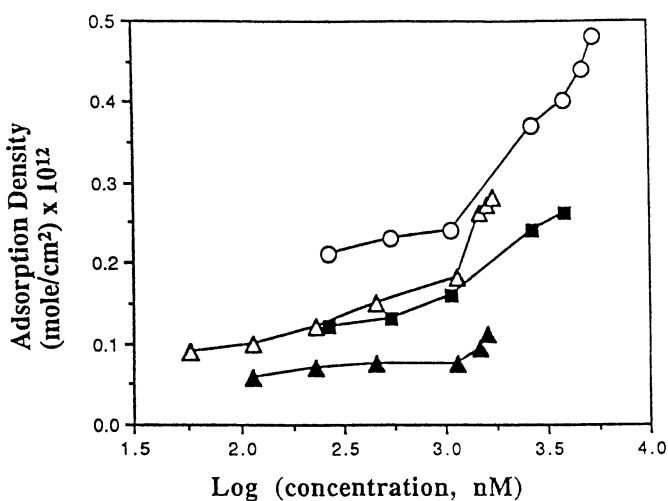


Figure 6. Adsorption isotherm of HDL and LDL onto a ZnSe ATR crystal (open circle and open triangle, respectively), and onto a cholesterol coated ZnSe surface (closed square and closed triangle, respectively). (Reproduced with permission from reference 19. Copyright 1995.)

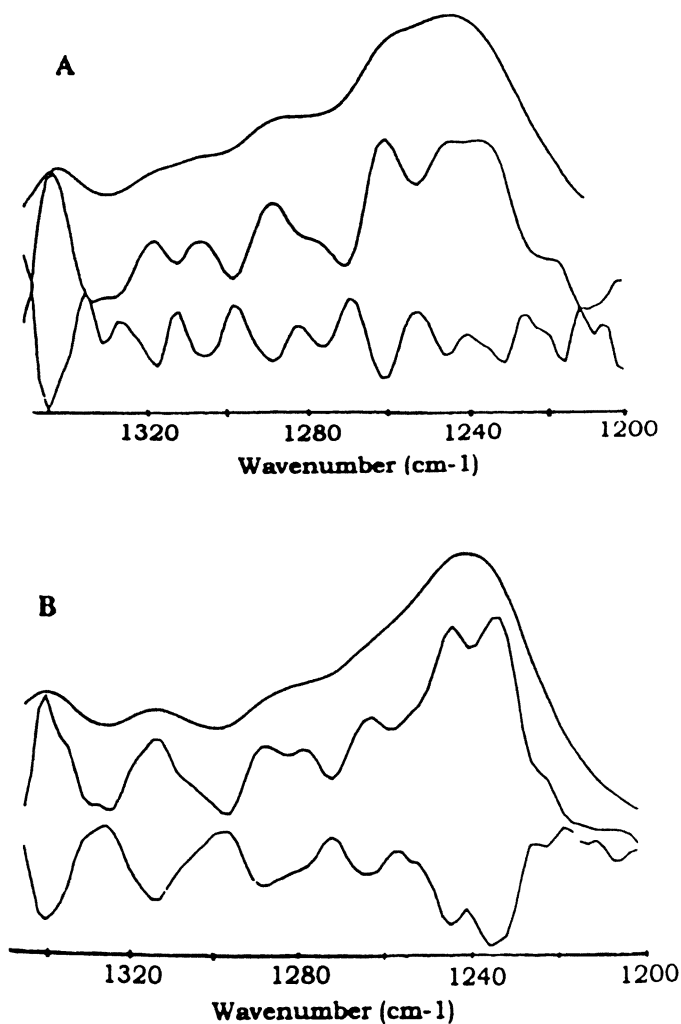


Figure 7. IR spectra of (A)  $\alpha$ -chymotrypsin and (B) IgG in the amide III region. On the basis of original protein spectra (top), a more intense peak around  $1240\text{ cm}^{-1}$  and smaller peaks at  $1317$  and  $1285\text{ cm}^{-1}$  were observed for both proteins. The resolution-enhanced spectra, deconvoluted spectra (middle), and second-derivative spectra (bottom) are also shown.



IgG have a significant amount of  $\beta$ -sheet structures (strong peak at  $\sim 1240\text{ cm}^{-1}$ ; also see below). The broader feature of  $\alpha$ -chymotrypsin at  $1270\text{--}1230\text{ cm}^{-1}$  may be caused by the mixing of an unordered band ( $1260\text{ cm}^{-1}$ ) with the  $\beta$ -sheet band ( $1240\text{ cm}^{-1}$ ). This type of analysis based simply on the protein spectrum is feasible only with amide III spectra. On the contrary, the featureless amide I protein spectrum does not provide such useful structural information unless curve analysis methods are employed (Fig 8; Table 1). Spectral bands obtained from Fourier self-deconvolution and second-derivative analysis of  $\alpha$ -chymotrypsin as well as IgG match very well with those obtained from curve-fitting of their respective original spectra both in amide I and amide III regions (Tables 1 & 2). The secondary structure contents derived on the basis of amide III IR band assignments in the literature (12, 41, 21) are consistent with the results obtained by x-ray crystallographic data analysis (Table 3).

The three-dimensional structure of staphylococcal enterotoxin (SEB), a superantigen, was analyzed by x-ray crystallography (55), but the derived secondary structural content was at a significant variance with that derived from CD analysis (45, 56). Amide I and III IR analysis confirmed the CD results (Table 3), which eventually led to the realization that SEB structure in solution and crystal is different, not only in terms of secondary structure but also in terms of quaternary structure (45).

**$\alpha$ -Helical protein.** The protein spectrum of high  $\alpha$ -helical lysozyme displays a high intensity band at  $\sim 1310\text{ cm}^{-1}$ , indicating a considerable amount of helical structure present in the protein (Fig. 9A). The curve-fitting analysis (Fig. 9B) reveals bands at  $1328$ ,  $1314$ , and  $1300\text{ cm}^{-1}$  which are categorized as being  $\alpha$ -helical with a total relative band strength of 48%. Bands at  $1285$  and  $1269\text{ cm}^{-1}$  are assigned to unordered structure (26%), and the bands at  $1250$  and  $1234\text{ cm}^{-1}$  are assigned to  $\beta$ -sheet structure for 26%. The x-ray data suggests 45%  $\alpha$ -helix, 36% others, and 19%  $\beta$ -sheet structure for lysozyme (Table 3).

**$\alpha$ -Helical- $\beta$ -sheet proteins.** Subtilisin Carlsberg, a protein with a relatively moderate amounts of  $\alpha$ -helix,  $\beta$ -sheet, and unordered structures. The subtilisin spectrum before curve-fitting shows distinct peaks at  $\sim 1310$ ,  $\sim 1265$ , and  $1245\text{ cm}^{-1}$ , which correspond to  $\alpha$ -helix, unordered and  $\beta$ -sheet, respectively. The intensity of the peak at  $\sim 1265\text{ cm}^{-1}$  is higher than that of the other two peaks, which suggests that subtilisin has a higher unordered content in comparison to  $\alpha$ -helix and  $\beta$ -sheet. The spectrum also indicates a similar  $\alpha$ -helix and  $\beta$ -sheet content of subtilisin (Fig. 10A). Curve fitting (Fig. 10B; Table 2) and band assignment result in secondary structure estimates, which closely match with the x-ray data (Table 3).

Secondary structure of botulinum neurotoxin type A (BoNT/A) and tetanus neurotoxin (TeNT) were also examined based on the assignment concluded above. Both neurotoxins are high in  $\beta$ -sheet content with significant amount of  $\alpha$ -helix, as revealed by the curve-fitting process (Table 2). TeNT shows slightly higher  $\beta$ -sheet content when it is compared to the results from CD, which may be due to the ambiguous assignment of peak at  $1253\text{--}1260\text{ cm}^{-1}$ . The peak is likely to be a mixture of  $\beta$ -sheet and unordered structure. Results obtained for BoNT/A is consistent with the one derived from CD (Table 3). Such a fine agreement in structural content derived from amide III spectral analysis with structure derived from other techniques ensures

Table 1: Amide I spectral band positions and their relative band strengths analyzed by Fourier self- deconvolution, second-order derivatization, and curve-fitting procedures.

Proteins	Band Positions <sup>a</sup>			Band percentage <sup>b</sup>	
	FSD	SD	CF	%	Assignment
$\alpha$ -chymotrypsin	1684	1686	1685	5	$\beta$ -sheets <sup>1,5</sup>
	1667	1668	1668	27	turn <sup>1,2,3,4</sup>
	1649	1649	1652	18	helix <sup>5</sup> or helix/coil <sup>2</sup>
	1636	1634	1637	50	$\beta$ -sheet <sup>3,4</sup>
	1616	1616	1616		side chain
IgG	1690	1692	1690	1	$\beta$ -sheet <sup>1,5</sup> or
					turn <sup>3,4</sup>
	1677	1677	1682	14	turn <sup>3</sup>
	1661	1661	1661	28	turn <sup>5</sup>
	1639	1639	1637	57	$\beta$ -sheets <sup>3,5</sup>
	1615	1616	1618		side chain
Subtilisin	1676	1679	1676	19	turn <sup>2,5</sup> or $\beta$ -sheet <sup>4</sup>
	1657	1658	1657	33	helix <sup>1,3,4</sup>
	1647	1645	1640	32	$\beta$ -sheet <sup>3,5</sup>
	1634	1632			
	1623		1629	16	$\beta$ -sheet <sup>1,3,4</sup>
	1614	1618	1615		side chain
Lysozyme	1687	1682	1690	6	turn <sup>2,5</sup> or $\beta$ -sheet <sup>4</sup>
	1673	1673	1674	21	turn <sup>3</sup>
	1655	1655	1654	40	helix <sup>1,3,4,5</sup>
	1638	1638	1638	15	$\beta$ -sheet <sup>3,5</sup>
	1621	1621	1623	18	$\beta$ -sheet <sup>4</sup>
BSA	1681	1684	1676	24	turn <sup>2,5</sup> or $\beta$ -sheet <sup>4</sup>
		1670			
	1653	1655	1654	53	helix <sup>3,5</sup>
	1635	1634	1635	23	$\beta$ -sheet <sup>3,4</sup>
	1617	1617		side chain	

Table 1. *Continued.*

Proteins	Band Positions <sup>a</sup>			Band percentage <sup>b</sup>	Assignment
	FSD	SD	CF	%	
TeNT turn <sup>1,2,4</sup>	1680	1684	1685	9	$\beta$ -sheet <sup>1,2,5</sup> 1669 20
	1653	1653	1653	30	helix <sup>5</sup>
	1638	1638	1368	28	$\beta$ -sheet <sup>5</sup>
	1616	1616	1623	13	side chain
BoNT/A	1687	1682	1687	8	turn <sup>3,4</sup> or $\beta$ -sheet <sup>1,2,5</sup>
	1675		1670	26	turn <sup>3</sup>
	1653	1653	1651	42	coil <sup>3</sup> or helix <sup>5</sup>
	1637	1634	1635	24	$\beta$ -sheet <sup>3,4</sup>
	1615	1615	1619		side chain

<sup>a</sup>FSD = Fourier self-deconvolution, SD = second-order derivatization, CF = curve fitting.

<sup>b</sup>Band percentage is calculated on the basis of peak height; results from peak area calculation are shown in parentheses. <sup>1</sup>Holloway and Mantsch (46); <sup>2</sup>Olinger et al. (47); <sup>3</sup>Dong et al. (14); <sup>4</sup>Villalain et al. (48); <sup>5</sup>Lee et al. (49).

Table 2: Amide III spectral band positions and their strengths as analyzed by Fourier self-deconvolution, and curve-fitting procedure.

Proteins	Band Positions <sup>a</sup>			Band	Assignment <sup>c</sup>
	FSD	SD	CF	percentage <sup>b</sup> %	
α-chymotrypsin	1317	1317	1309	9	helix
	1306	1306	1294	0	helix
	1287	1288	1283	20	unordered
	1277	1276	1270	0	unordered
	1260	1261	1257	32	unordered
	1245	1245	1246	4	β-sheet
	1235	1233	1237	25	β-sheet
			1224	10	β-sheet
IgG	1317	1317	1317	3	helix
	1290	1286	1286	7	unordered
	1280	1282	1275	4	unordered
	1264	1264	1261	27	unordered
	1244	1244	1248	19	unordered
	1234	1234	1238	28	β-sheet
		1228	12	β-sheet	
Subtilisin	1308	1308	1308	32	helix
			1301	4	unordered
	1284	1285	1282	4	unordered
			1271	5	unordered
	1261	1262	1263	35	unordered
	1242	1242	1241	21	β-sheet
	1231	1230			
Lysozyme	1331		1328	13	helix
	1316	1315	1314	18	helix
	1300	1299	1300	14	helix
	1284	1283	1285	13	unordered
	1269	1269	1269	14	unordered
	1251	1250	1250	17	β-sheet
	1235	1235	1234	11	β-sheet

Table 2. *Continued.*

Proteins	Band Positions <sup>a</sup>			Band	Assignment <sup>c</sup>
	FSD	SD	CF	percentage <sup>b</sup>	
BSA	1318		1319	15	helix
	1307		1308	7	helix
	1292		1293	33	helix
	1280		1280	0	unordered
	1267		1267	18	unordered
	1248		1250	5	$\beta$ -sheet
	1238		1243	17	$\beta$ -sheet
	1225		1225	5	$\beta$ -sheet
TeNT	1320	1322	1321	9	helix
	1311	1310	1309	8	helix
	1293	1294	1294	6	helix
		1286	1283	5	unordered
	1269	1271	1272	6	unordered
	1253	1260	1256	23	$\beta$ -sheet
	1237	1243	1237	39	$\beta$ -sheet
		1223	1226	4	$\beta$ -sheet
BoNT/A	1318	1320	1319	10	helix
	1303	1302	1306	6	helix
	1289	1287	1289	11	unordered
		1271	1269	14	unordered
	1259	1260	1257	14	unordered
	1242	1243	1241	42	$\beta$ -sheet
		1223	1226	3	$\beta$ -sheet

<sup>a</sup>FSD = Fourier self-deconvolution, SD = second-order derivatization, CF = curve fitting.

<sup>b</sup>Band percentage is calculated on the basis of peak height; results from peak area calculation are shown in parentheses.

<sup>c</sup>Unordered structure includes turns and random coils.

Table 3: Comparison of protein secondary structure as determined by FT-IR, CD, and x-ray crystallography<sup>a</sup>.

Proteins	$\alpha$ -helix	$\beta$ -sheet	others	methods
$\alpha$ -chymotrypsin	9	39	52	FT-IR (Amide III) <sup>1</sup>
	8	50	42	X-ray <sup>2</sup>
	9	47	44	FT-IR (Amide I) <sup>3</sup>
	12	51	37	FT-IR (Amide I) <sup>4</sup>
	*18	55	27	FT-IR (Amide I) <sup>1</sup>
IgG	3	59	38	FT-IR (Amide III) <sup>1</sup>
	3	67	30	X-ray <sup>2</sup>
	3	64	33	FT-IR (Amide I) <sup>3</sup>
	9	76	15	FT-IR (Amide I) <sup>4</sup>
	0	58	42	FT-IR (Amide I) <sup>1</sup>
Subtilisin	32	21	47	FT-IR (Amide III) <sup>1</sup>
	28	23	49	X-ray <sup>2,5,6</sup>
	33	48	19	FT-IR (Amide I) <sup>1</sup>
Lysozyme	45	28	27	FT-IR (Amide III) <sup>1</sup>
	45	19	36	X-ray <sup>2</sup>
	35	10	55	CD <sup>7</sup>
	40	33	27	FT-IR (Amide I) <sup>1</sup>
BSA	55	27	18	FT-IR (Amide III) <sup>1</sup>
	55	0	45	CD <sup>8</sup>
	53	23	24	FT-IR (Amide I) <sup>1</sup>
TeNT	23	66	11	FT-IR (Amide III) <sup>1</sup>
	20	50	30	CD <sup>9</sup>
	22	49	29	FT-IR (Amide III) <sup>9</sup>
	*30	50	20	FT-IR (Amide I) <sup>1</sup>
BoNT/A	16	45	39	FT-IR (Amide III) <sup>1</sup>
	28	42	30	CD <sup>10</sup>
	29	45-49	22-26	FT-IR (Amide III) <sup>11</sup>
	*42	32	26	FT-IR (Amide I) <sup>1</sup>
SEB	11	71	18	FT-IR (Amide III) <sup>12</sup>
	20	37	43	X-ray <sup>13</sup>
	10	75	15	CD <sup>12</sup>
	12	57	31	FT-IR (Amide I) <sup>12</sup>

<sup>a</sup>Results are calculated on the basis of peak intensity; results from area calculation are shown in parenthesis. (\*) High helix content might be due to the mixture of helix and random coil structure. <sup>1</sup>Fu et al. (41); <sup>2</sup>Levitt and Greer (50); <sup>3</sup>Dong et al. (14); <sup>4</sup>Byler and Susi (11); <sup>5</sup>Wright et al. (51); <sup>6</sup>Petsko and Tsimoglov (52); <sup>7</sup>Saxena and Wetlaufer (53); <sup>8</sup>Schechter and Blout (54); <sup>9</sup>Singh et al. (21); <sup>10</sup>Singh et al. (26); <sup>11</sup>Singh et al. (28); <sup>12</sup>Singh et al. (45); <sup>13</sup>Swaminathan et al. (55)

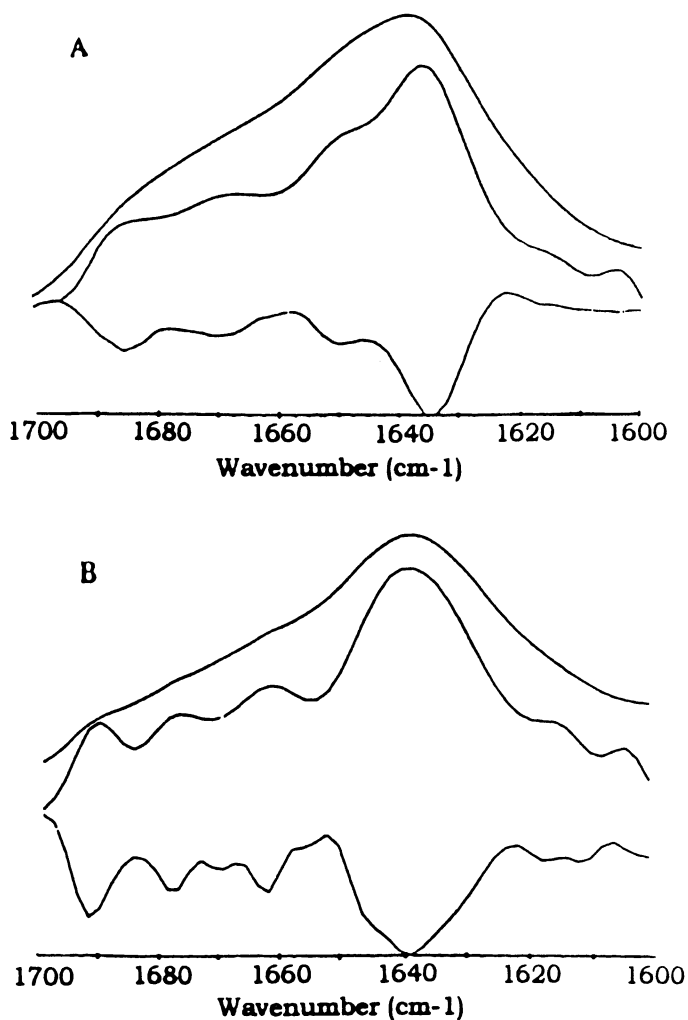


Figure 8. IR spectra of (A)  $\alpha$ -chymotrypsin and (B) IgG in the amide I region. Without the resolution-enhancement (deconvolution, middle, and second derivatization, bottom), a major band only at  $\sim 1640\text{ cm}^{-1}$  appeared in the original spectra of both the proteins.

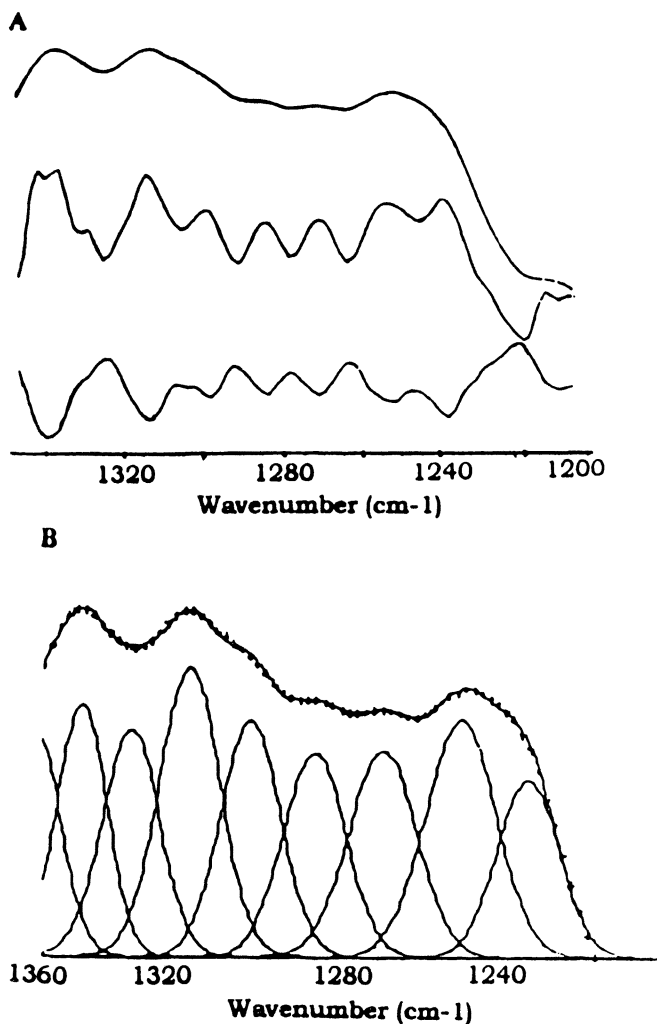


Figure 9. The IR spectra of lysozyme. (A) The original (top), deconvolved (middle), second-derivative (bottom) and (B) the curve-fitted spectrum. A band ( $\sim 1300\text{ cm}^{-1}$ ) with higher intensity appears in  $\alpha$ -helix region.



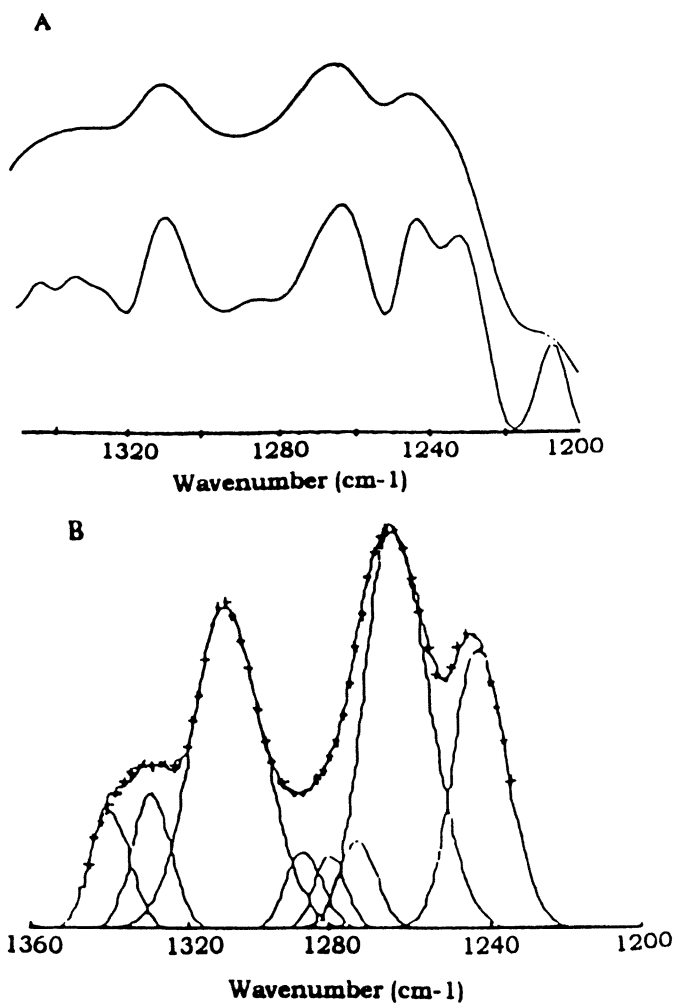


Figure 10. The IR spectrum of subtilisin. (A) The IR spectrum of subtilisin (top) and the deconvolution spectrum (bottom). The higher intensity of band at  $\sim 1260\text{ cm}^{-1}$  was observed on the original spectrum, which suggested the higher unordered structural content in subtilisin. (B) Curved-fitting results of subtilisin (—) on the original spectrum (+++).

the usage of the amide III IR spectra as one of the structure determination method for the following studies. It is notable that amide I analysis of both TeNT and BoNT/A revealed  $\alpha$ -helical content higher than expected from CD results.

**Application of Amide III approach to examine biological systems.** In order to evaluate the method and assignment, the amide III FT-IR analysis has been applied to several biological systems such as calsequestrin, a protein whose secondary structure has been analyzed in the past with the use of several spectroscopic techniques to investigate the structural change induced by calcium binding (57). Calsequestrin is a 40 kDa protein, which is found in the terminal cisternae of the sarcoplasmic reticulum and has been shown to bind up to 50 moles  $\text{Ca}^{2+}$  per mole protein (58,59). Therefore, it is an ideal candidate for the storage of the calcium pumped into the sarcoplasmic reticulum (60). The effect of calcium binding to calsequestrin has been examined by CD,  $^1\text{H}$  NMR, Raman spectroscopy, and other spectroscopic methods (57-62). Although all these studies have suggested that calcium binding results in a conformational change in the protein, leading to an increase in the  $\alpha$ -helical folding, the  $\alpha$ -helical content itself has varied widely (11-40 %) from one report to another. In addition, calsequestrin is known to aggregate upon treatment with  $\text{Ca}^{2+}$ , which makes it a difficult system for spectral recordings. However, normal IR spectra can still be recorded because of the unique sampling mode of ATR spectroscopy (5,24).

In a study the technique of ATR was employed in combination with FT-IR to examine the calcium-induced secondary structural change of protein (41). Amide III FT-IR spectra of calsequestrin before and after treatment with different concentrations of calcium are shown in Fig. 11. While treatment with 0.1 mM  $\text{Ca}^{2+}$  does not alter the IR spectrum significantly, 0.5 mM  $\text{Ca}^{2+}$  introduces considerable change at 1240 and 1310  $\text{cm}^{-1}$  (Fig. 11). It may be noted here that the protein starts aggregating when treated with 0.5 mM  $\text{Ca}^{2+}$ , as indicated by precipitation and turbidity. The broad band at 1310-1320  $\text{cm}^{-1}$ , representing  $\alpha$ -helix structure, increases with the increase in the  $\text{Ca}^{2+}$  concentration. Thus,  $\text{Ca}^{2+}$  binding increases the  $\alpha$ -helical content of calsequestrin. After the curve-fitting analysis of the amide III region, we estimate 36%  $\alpha$ -helix, 31%  $\beta$ -sheet, and 33% random coils and turns in the calcium-free form of calsequestrin and 46%  $\alpha$ -helix, 21%  $\beta$ -sheet, and 33% random coils and turns in calsequestrin treated with 1 mM  $\text{Ca}^{2+}$ .

As mentioned earlier, the secondary structure of calsequestrin, and the effect of the  $\text{Ca}^{2+}$  binding on the secondary structure, has been the subject of several previous studies (57-62). The following two different types of results have been obtained: (1) At low concentrations (less than 1 mg/ml), circular dichroism results have suggested low  $\alpha$ -helical contents (11-13%) in calsequestrin that could be enhanced to 19-35% with the binding of  $\text{Ca}^{2+}$ . (2) At high concentrations (100 mg/ml), Raman spectroscopy results have suggested high  $\alpha$ -helical content (32%) that increased to 41% upon treatment with 30 mM  $\text{Ca}^{2+}$ .

Results from Fu et al. (41) clearly show a structural change in calsequestrin upon treatment with 0.5 mM  $\text{Ca}^{2+}$  (Fig. 11). The observed change in  $\alpha$ -helical content (36 to 46%) with 1 mM  $\text{Ca}^{2+}$  is consistent with  $\alpha$ -helical content determined by Raman spectroscopy (62), even though the protein concentration used was 0.9 mg/ml (low calsequestrin concentration) as compared to 100 mg/ml calsequestrin (high

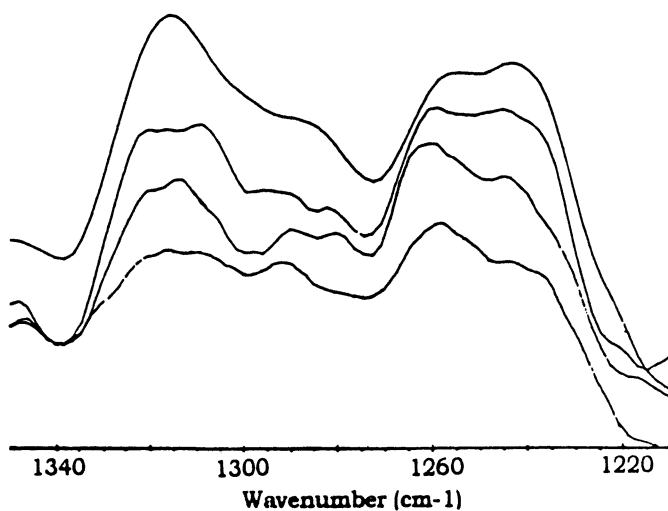


Figure 11. The original amide III IR spectra of calsequestrin with or without treatment with calcium. From bottom to top, calcium-free calsequestrin, and calsequestrin after treatment with 0.1, 0.5, 1 mM  $\text{Ca}^{2+}$ , respectively. Upon binding of calcium, the peak intensity at 1250-1260  $\text{cm}^{-1}$  (the combination of  $\beta$ -sheet and random coil/turn) relatively decreases while the peak at 1320  $\text{cm}^{-1}$  is enhanced significantly, indicating an increase in  $\alpha$ -helical content of the protein.

calsequestrin concentration) used for Raman spectroscopy (62). It is possible that Raman spectroscopy results will be similar to those of FT-IR if Raman spectra could also be determined at low concentrations. However, Williams and Beeler (62) pointed out that determining the Raman spectra of low concentration of calsequestrin was not possible because of the interference from other luminescent groups in the protein. Such difficulty could also obscure the clear structural change we observed with FT-IR that was contrary to the conclusion made by Williams and Beeler (62). In any case, the inconsistent estimate of  $\alpha$ -helical content from CD and Raman spectra does not seem to be due to the concentration difference, as concluded from Raman spectroscopy data (63). It is also notable that FT-IR results suggest that the increase in the  $\alpha$ -helix is at the expense of the  $\beta$ -sheet structure, and not at the expense of random coils, as suggested by Aaron et al. (58).

The changes in the calsequestrin secondary structure do not seem to occur as a result of the binding of only a few cations to the specific groups of the protein. This conclusion is reached because 100  $\mu\text{M}$   $\text{Ca}^{2+}$  ions (a 4.5 molar excess) do not alter the secondary structure of the protein significantly. Structural changes are obvious at 22-molar excess, which are further altered at a 45-fold molar  $\text{Ca}^{2+}$  excess (Fig. 11), suggesting that at higher molar excess  $\text{Ca}^{2+}$  binding could occur, even if the protein is already precipitated. In summary, the amide III region of the IR band not only allows a clear visualization of the structural change in calsequestrin upon binding with calcium, but it also reveals new information about this system not readily possible with other spectroscopic techniques because of their various limitations.

**Critical assessment of amide I and III.** The amide III spectral region has not been recognized as a potential region for secondary structure analysis primarily because of the weak protein signal in this region (5). However, a great advantage of using amide III for structural analysis of protein is the absence of the OH vibration interference from water, which is a major concern in the amide I region. Therefore, use of the amide III protein band eliminates the need for  $\text{D}_2\text{O}$ , which is commonly used to deal with water interference (11). The error introduced from the subtraction of water and water vapor present in the light path is also eliminated (44). Furthermore, ambiguity associated with the assignment of bands to a specific protein secondary structure in the amide I region (26,64) is absent in the amide III region. For instance, bands that are located between 1650 and 1655  $\text{cm}^{-1}$  can be assigned to either a  $\alpha$ -helix or random coils structure in the amide I region. In the amide III region, the frequencies at which the different amide bond vibrations occur are more localized and provide a better-resolved protein spectrum. This factor allows a greater ease in the peak definition. In the amide III region, distinct peaks are visualized before any curve-analysis methods are applied. These characteristic bands afford a simple illustration of secondary structure content of a protein.

Curve-analysis methods have been employed in previous studies that include the use of Fourier self-deconvolution (resolution enhancement), second-derivatization, and curve-fitting for the determination of individual bands that comprise the protein spectrum. The use of these methods for curve analysis are, however, varied; Dong et al. (14) have used the method of calculating band area under the second-derivative spectrum for estimating the percent of each structure present in a protein. Byler and

Susi (11) used self-deconvolution of the spectra followed by curve-fitting of the deconvolved spectra to estimate the secondary structure present. Since second-derivative and deconvolution processes are very sensitive to change in the spectra, it is possible that noise, or the inappropriately subtracted water band, may be amplified during this process (64). Thus, any curve-analysis method that involves curve-fitting on a second-derivative and deconvolved spectrum is not recommended. Our approach to overcome this obstacle employs second-derivative and deconvolution techniques as peak-assigning methods. A curve-fitting process on the original IR spectrum is then applied with these defined peak positions.

In our studies, we have used both band intensity and band area to calculate the contribution of each band of the protein spectrum. Direct estimation from peak height gives parallel results based on the band area, and the information of peak intensity is much easier to obtain. The secondary structural content information of proteins obtained from these studies is in good agreement with the results obtained from other researchers' work (Table 3; 41), especially, with the results from x-ray crystallography, which is considered the most precise method for protein structure analysis.

Comparative analysis of the amide I spectral region of proteins (Table 1) suggests that this region produces a relatively featureless spectrum. In such spectrum, it is difficult to readily extract characteristic information with the originally recorded spectrum, especially if the protein does not have predominantly  $\alpha$ -helical or  $\beta$ -sheet folding. With the aid of the resolution enhancement, curve-fitting can be achieved to a reasonable extent. Difficulty arises when one is attempting to assign bands. For example, the band at  $1676\text{ cm}^{-1}$  would be assigned to turn structure according to Lee et al. (49) and Olinger et al. (47), whereas, Villalain et al. (48) assigned this band to  $\beta$ -sheet. Misassignment of the only band, which contributes considerably to the IR spectra of certain proteins, would introduce significant error (41). In addition, because of its inability in distinguishing  $\alpha$ -helix and random coil structures in aqueous environment, the amide I spectral region loses its accuracy in structural analysis (5,26).

The band assignments made with standard proteins were used as guidelines for analyzing the secondary structure of tetanus and botulinum neurotoxins, casequestrin, and staphylococcal enterotoxin B, which yield nearly identical results when compared to those obtained from CD, Raman and x-ray crystallographic analyses (Table 3). This observation supports the use and accuracy of the amide III spectral region in secondary structural estimation of proteins of unknown x-ray crystal structure.

As with other spectroscopic methods, the IR spectroscopy has not been successfully used to distinguish subtle differences within the helical, sheet, turn and random coil structures, although it is theoretically feasible to observe influence of local environment on the vibrational frequencies of these structures. A comparison of the use of amide I and amide III regions suggest that while amide III region shows greater promise of detailed spectral features, this region has its limitations vis-à-vis amide I region. The limitations of the amide III band region for secondary structure analysis include the following: (1) the lower signal band that requires higher signal-to-noise ratio spectra for accurate quantitative analysis, and (2) there is an interference from the side chains in the amide III region. The signal contribution from the amide vibration is only approximately 30%, and amide III bands are extensively mixed with CH vibrations of amino side chains (65). However, a similar potential energy distribution

for the amide III spectral region in Raman spectroscopy has been proven useful and sensitive in protein conformational analysis (10), which suggests that the signal from the backbone vibration might also be sensitive to protein structure in this region. Some researchers have already observed that the amide III spectral region is particularly useful for distinguishing secondary structures (5,12,66). Although some of the frequencies are mixed mode, the behavior of the bands still clearly indicates that the bands are due to the amide III vibrations and can be used to identify the secondary structures of proteins. In addition, possible interferences from other groups including those of lipids and carbohydrates present in membranes and glycoproteins could be a problem in the amide III region. For obtaining protein spectra in such conditions, one needs to subtract the contribution of these interfering groups. As an example, one way to correct the interference from lipid is to monitor the disappearance of the band at  $1737\text{ cm}^{-1}$  (C=O vibration of the ester group) while subtracting the lipid from protein spectrum.

A major problem in using the amide III region for secondary structure estimation has remained the lack of distinction between random coil and turn frequencies (41,45,67). In a recent study, our research group has attempted to address this issue by employing an approach to obtain totally unordered form of protein with the use of 6 M guanidine-hydrogen chloride (68). Monitoring of structural changes accompanying protein denaturation with guanidine.HCl has allowed identification of the representative bands for random coil structures. The following frequency ranges are considered the best available at present for the correlation of amide III modes with secondary structure (Table 4):  $1330\text{-}1295\text{ cm}^{-1}$ ,  $\alpha$ -helix;  $1295\text{-}1270\text{ cm}^{-1}$ ,  $\beta$ -turn;  $1270\text{-}1250\text{ cm}^{-1}$ , random coil;  $1245\text{-}1220\text{ cm}^{-1}$ ,  $\beta$ -sheet. The boundary region of  $1245\text{-}1250\text{ cm}^{-1}$  may have contributions from both random coil and  $\beta$ -sheet structures; the boundary region of  $1300\text{-}1295\text{ cm}^{-1}$  may have contributions from both  $\alpha$ -helix and  $\beta$ -turn structures.

In summary, applying data processing methods to the amide III spectral region eliminates the concerns of water interference and the ambiguity in identifying the  $\alpha$ -helix and random coil bands, in contrast to the amide I region. In addition, although the amide III mode is  $\sim 5\text{-}10$ -fold weaker in IR band strength than the amide I mode, it is detectable with modern instrumentation. Thus the amide III region of protein IR spectra appears to be a valuable tool in estimating secondary structural contents of individual protein. However, the strong signal of amide I band, and the availability of a large amounts of amide I protein data available in the literature makes this band an indispensable region for protein secondary structure analysis. Therefore, it seems that both amide I and amide III regions can be used to obtain complementary spectral information for estimation of protein secondary structure.

### Modern advanced IR techniques

In recent years, several modern approaches have been employed to use IR spectroscopy for protein analysis. These include two-dimensional IR spectroscopy, polarized IR spectroscopy, isotope-edited IR spectroscopy, and time-resolved IR spectroscopy. In two-dimensional IR spectroscopy, IR spectra are recorded as a function of certain perturbations such as temperature, H-D exchange, etc., and correlation contours are

Table 4. Band assignments for Amide III frequencies.

coils	$\alpha$ -helix	$\beta$ -sheets	$\beta$ -turns	random (other)
Kaiden <i>et al.</i> (12)	1300 - 1250 $\text{cm}^{-1}$	1240 - 1230 $\text{cm}^{-1}$	-	1270 - 1240 $\text{cm}^{-1}$
Singh <i>et al.</i> (21)	1317 - 1280 $\text{cm}^{-1}$	1245 - 1230 $\text{cm}^{-1}$	-	1271 - 1245 $\text{cm}^{-1}$
Fu <i>et al.</i> (41)	1320 - 1295 $\text{cm}^{-1}$	1255 - 1225 $\text{cm}^{-1}$	-	1294 - 1255 $\text{cm}^{-1}$
Cai and Singh (67)	1330-1295 $\text{cm}^{-1}$	1245-1220 $\text{cm}^{-1}$	1295-1270 $\text{cm}^{-1}$	1270-1255 $\text{cm}^{-1}$

plotted (69-72) to obtain synchronous and asynchronous spectra. Cross-correlation of peaks have been very effective in identifying amide I and amide II peaks with different secondary structural features. Polarized IR spectroscopy has been particularly useful in evaluating orientations of  $\alpha$ -helical or  $\beta$ -sheet segments of proteins (73; Chapter 4). Isotope-edited IR spectroscopy is gaining popularity with scientists interested in protein-protein interactions, as labeling of one of the proteins with  $^{13}\text{C}$  or  $^{15}\text{N}$ , allows to distinguish contribution of two proteins in a complex (74; Chapter 9). Time-resolved IR spectroscopy is becoming very effective in examining both fast and slow kinetic changes in proteins in responses to external stimulants. Currently instruments are available for resolving protein spectra at ms,  $\mu\text{s}$  and ns level (75; Chapter 8; 76-79). In addition, various statistical analysis methods are being developed to use IR spectra in amide I, amide II, amide III, and various combination among them, as well as with other spectroscopic techniques such as electronic CD and vibrational CD, for deriving secondary structure content of proteins (16,17,22,23,80,81; Chapter 5)

### Concluding Remarks

With the high-sensitivity, detectability of opaque or turbid solution, and the convenience of sample handling, Fourier transform infrared spectroscopy has become an effective technique for the study of protein secondary structures, and changes in these features under various experimental conditions (82-84). In addition, since vibrational frequency of the atoms in biological molecules are of the order of  $10^{-13}$  s (covering mid-IR region,  $100\text{-}4000\text{ cm}^{-1}$  wavenumbers). The time scale of these vibrations, which are the basis of IR spectroscopy, is much faster than most chemical reactions (ms to  $\mu\text{s}$ ), cellular diffusion processes (ns), and even protein dynamics (ns to ps). Therefore, IR spectroscopy has potential to capture information on atomic dynamics of protein structure and function. The technique needs further development in correlating theoretical prediction with experimental observations with respect to band assignments (85; Chapter 2). There are still many practical difficulties in applying the technique for peptides and proteins (82; Chapter 3). However, use of IR spectroscopy, particularly in sample conditions where other spectroscopic techniques are not possible (such as the case of protein aggregates or membrane samples), IR is emerging as a formidable technique of analysis (74; Chapter 4; 83; Chapter 6). Future efforts are needed to generate large data bank of IR spectra of proteins of known structural composition from x-ray crystallography and NMR spectroscopy. This exercise in combination with 2-dimensional IR spectroscopy will permit assignment of spectral bands to various components of secondary structural folding. Although IR spectroscopy of proteins and peptides is not likely to provide static structural information at par with x-ray crystallography and NMR spectroscopy, it has potential to provide dynamic structural information, which will surpass any other current methods.



## References

1. Wagner, G. *Nature Struct. Biol.* **1997**, 4, 841.
2. Susi, H.; Byler, D. M. *Methods Enzymol.* **1986**, 130, 290.
3. Jakobsen, R. J.; Wasacz, F. M. *Appl. Spectrosc.* **1990**, 44, 1478.
4. Haris, P. I.; Chapman, D. *Trends. Biochem.* **1992**, 17, 328.
5. Singh BR, Fu F-N.; Fuller, M. P. in *Techniques in protein chemistry III* (ed. R. H. Angeletti), Academic Press, Orlando, FL., **1992**, pp 385-398.
6. Elliot, A.; Ambrose, E. J. *Nature* **1950**, 165, 921.
7. Miyazawa, T.; Shimanouchi, T.; Mizushima, S. I. *J. Chem. Phys.* **1956**, 24, 408.
8. Krimm, S. *J. Mol. Biol.* **1962**, 528.
9. Krimm, S.; Abe, Y. *Proc. Natl. Acad. Sci. USA* **1972**, 69, 2788.
10. Krimm, S.; Bandekar, J. *Adv. Prot. Chem.* **1986**, 38, 181.
11. Byler, D. M.; Susi, H. *Biopolymers* **1986**, 25, 469.
12. Kaiden, K.; Matsui T.; Tanaka, S. *Appl. Spectrosc.* **1987**, 41, 180.
13. Surewicz, W. K.; Mantsch, H. H. *Biochim. Biophys. Acta* **1988**, 952, 115.
14. Dong, A.; Huang, P.; Caughy, W. S. *Biochemistry* **1990**, 29, 3303.
15. Dosseau, F.; Pezolet, M. *Biochemistry* **1990**, 29, 8771.
16. Rahmelow, K.; Hubner, W. *Anal. Biochem.* **1996**, 241, 5.
17. Wi, S.; Pancoska, P.; Keiderling, T. A. *Biospectrosc.* **1998**, 4, 93.
18. Fu, F. -N.; Fuller, M. P.; Singh, B. R. J. *Appl. Spectrosc.* **1993**, 47, 98-102.
19. Hu, Y.; Singh, B. R. *Appl. Spectrosc.* **1995**, 49, 1356.
20. Havel, H. A. *Spectroscopic Methods for Determining Protein Structure in Solution*, VCH Publishers, New York, pp. 2-10.
21. Singh, B. R.; Fuller, M. P.; Schiavo, G. *Biophys. Chem.* **1990**, 36, 155.
22. Cai, S.; Singh, B. R. in: *Infrared Analysis of Peptides and Proteins: Functional Molecular Features* (ed. B. R. Singh), American Chemical Society, Washington, D.C. **1999**; Chapter 5.
23. Pancoska, P.; Janota, V.; Keiderling, T. A. *Anal. Biochem.* **1999**, 267, 72.
24. Singh, B. R.; Fuller, M. P. *Appl. Spectrosc.* **1991**, 45, 1017.
25. Wasacz, F. M.; Olinger, J. M.; Jakobsen, R. J. *Biochemistry* **1987**, 26, 1464.
26. Singh, B. R.; Wasacz, F. M.; Strand, S.; Jakobsen, R. J.; DasGupta, B. R. *J. Prot. Chem.* **1990**, 9, 705.
27. Fink, D. J.; Gendreau, R. M. *Anal. Biochem.* **1984**, 139, 140.
28. Singh, B. R.; Fuller, M. P.; DasGupta, B. R. *J. Prot. Chem.* **1991**, 10, 637.
29. Oberg, K. A.; Fink, A. L. *Anal. Biochem.* **1998**, 256, 92.
30. Horbett, T. A.; Brash, J. L. in: *Proteins at Interfaces: Physicochemical and Biochemical Studies* (Eds. J. L. Brash and T. A. Horbett), American Chemical Society, Washington, D.C. **1987**, ACS Symposium Series 343, pp. 1-33.
31. Pitt, W. G. and Cooper, S. L. in: *Proteins at Interfaces: Physicochemical and Biochemical Studies* (Eds. J. L. Brash and T. A. Horbett), American Chemical Society, Washington, D.C. **1987**, ACS Symposium Series 343, pp. 324-338.
32. Brown, M. S.; Goldstein, J. L. *Sci. Amer.* **1984**, 252, 58.
33. Reichl, D. *Atherosclerosis* **1994**, 105, 117.

34. Chappell, D. A.; Fry, G. L.; Waknitz, M. A.; Berns, J. J. *J. Biol. Chem.* **1991**, 266, 19296.
35. Brinton, E. A.; Oram, J. F.; Bierman, E. L. *Biochim. Biophys. Acta* **1987**, 920, 68.
36. Mendel, C. M.; Kunitake, S. T. *J. Lipid Res.* **1988**, 29, 1171.
37. Alexander, J. J.; Miguel, R.; Graham, D. *J. Sur. Res.* **1990**, 49, 248.
38. Tabas, I.; Tall, A. R. *J. Biol. Chem.* **1984**, 259, 13897.
39. Gendreau, R. M.; Leininger, R. I.; Winters, S.; Jakobsen, R. J. in: *Biomaterials: Interfacial Phenomena and Applications* (Eds. S. L. Cooper and N. A. Peppas), American Chemical Society, Washington, D.C. **1987**, pp. 371-394.
40. Casstilo, E. J.; Koenig, J. L.; Anderson, J. M.; Lo, J. *Biomaterials* **1984**, 25, 319.
41. Fu, F.-N.; DeOliveira, D. B.; Trumble, W. R.; Sarkar, H. K.; Singh, B. R. *Appl. Spectrosc.* **1994**, 48, 1432.
42. Gendreau, R. M.; Winters, S.; Leininger, R. I.; Fink, D. J.; Jakobsen, R. J. *Appl. Spectrosc.* **1981**, 35, 353.
43. Englander, S. W.; Downer, N. W.; Teitelbaum, H. *Ann. Rev. Biochem.* **1972**, 41, 903.
44. Anderle, G.; Mendelsohn, P. *Biophys. J.* **1987**, 52, 69.
45. Singh, B. R.; Fu, F. -N.; Ledoux, D. N. *Nature Struct. Biol.* **1994**, 1, 358.
46. Holloway, P. W.; Mantsch, H. H. *Biochemistry* **1989**, 28, 931.
47. Olinger, J. M.; Hill, D. M.; Jakobsen, R. J.; Brody, R. S. *Biochim. Biophys. Acta* **1986**, 869, 89.
48. Villalain, J.; Gomez-Fernandez, J. C.; Jakobsen, M.; Chapman, D. *Biochim. Biophys. Acta* **1989**, 978, 305.
49. Lee, D. C.; Herzky, E.; Chapman, D. *Biochemistry* **1987**, 26, 5775.
50. Levitt, M.; Greer, J. *J. Mol. Biol.* **1977**, 114, 181.
51. Wright, C. S.; Alden, R. A.; Kraut, J. *Nature* **1969**, 221, 235.
52. Petsko, G. A.; Tsernoglou, P. *J. Mol. Biol.* **1976**, 106, 453.
53. Saxena, V. P.; Wetlaufer, D. B. *Proc. Natl. Acad. Sci. USA* **1971**, 68, 962.
54. Schechter, E.; Bout, E. R. *Proc. Natl. Acad. Sci. USA* **1964**, 51, 695.
55. Swaminathan, S.; Furey, W.; Pletcher, J.; Sax, M. *Nature* **1992**, 359, 801.
56. Singh, B. R.; Evenson, M. L.; Bergdoll, M. S. *Biochemistry* **1988**, 27, 8735.
57. He L.; Dunker, A. K.; Wesson, C. R.; Trumble, W. R. *J. Biol. Chem.* **1993**, 268, 24635.
58. Aaron, B. M.; Oikawa, K.; Reithmeier, R. A.; Sykes, B. D. *J. Biol. Chem.* **1984**, 259, 11876.
59. Cozens, B.; Reithmeier, R. A. *J. Biol. Chem.* **1984**, 259, 6248.
60. MacLennan, D. H.; Campbell, K. P.; Reithmerier, R. A. F. in *Calcium and Cell Function* (Ed. W. Y. Cheung), Academic Press, New York, **1983**, pp. 151-193.
61. MacLennan, D. H.; Wong, P. T. S. *Proc. Acad. Sci. USA* **1971**, 68, 1231.
62. Williams, R. W. and Beeler, T. J. *J. Biol. Chem.* **1986**, 261, 12408.
63. Markley, J. L.; Urich, E. L. *Ann. Rev. Biophys. Bioeng.* **1984**, 13, 493.
64. Surewicz, W. K.; Mantsch, H. H.; Chapman, D. *Biochemistry* **1993**, 32, 389.
65. Carey, P. R. in *Biochemical Applications of Raman and Resonance Raman Spectroscopy*, Academic Press, New York, **1982**, p. 71.

66. Jakobsen, R. J.; Wasacz, F. M. in *Proteins at interfaces: physical and biochemical studies* (Eds. J.L. Brash and Horbett TA), American Chemical Society, Washington, D.C. **1987**, pp 361.
67. Griebenow, K.; Klibanov, A. *Proc. Natl. Acad. Sci. USA* **1995**, 92, 10969.
68. Cai, S.; Singh, B. R. *Biophysical Chem.* **1999**, in press.
69. Noda, I. *Bull. Am. Phys. Soc.* **1986**, 31, 520.
70. Nabet, A.; Pezolet, M. *Appl. Spectrosc.* **1997**, 51, 466.
71. Ozaki, Y.; Liu, Y.; Noda, I. *Appl. Spectrosc.* **1997**, 51, 526-535.
72. Sefara, N. L.; Magtoto, N. P.; Richardson, H. H. *Appl. Spectrosc.* **1997**, 51, 536-540.
73. Goormaghtigh, E.; Ruyschaert, J. -M. in: *Infrared Analysis of Peptides and Proteins: Functional Molecular Features* (ed. B. R. Singh), American Chemical Society, Washington, D.C. **1999**; Chapter 4.
74. Li, T.; Arakawa, T.; Horan, T. P.; Chang, B. in: *Infrared Analysis of Peptides and Proteins: Functional Molecular Features* (ed. B. R. Singh), American Chemical Society, Washington, D.C. **1999**; Chapter 9.
75. Gerwert, K. in: *Infrared Analysis of Peptides and Proteins: Functional Molecular Features* (ed. B. R. Singh), American Chemical Society, Washington, D.C. **1999**; Chapter 8.
76. Braiman, M. S.; Ahl, P. L.; Rothschild, K. J. *Proc. Natl. Acad. Sci.* **1987**, 84, 5227.
77. Rath, P.; Degrip, W. J.; Rothschild, K. J. *Biophys. J.* **1998**, 74, 192.
78. Uhmann, W.; Becker, A.; Taran, C.; Siebert, F. *Appl. Spectrosc.* **1991**, 45, 390.
79. Rammelsberg, R.; Hebling, B.; Chorongiewski, H.; Gerwert, K. *Appl. Spectrosc.* **1997**, 51, 558.
80. Baumruk, V.; Pancoska, P.; Keiderling, T. A. *J. Mol. Biol.* **1996**, 259, 774.
81. Baello, B. I.; Pancoska, P.; Keiderling, T. A. *Anal. Biochem.* **1997**, 250, 212.
2. Haris, P. I. in: *Infrared Analysis of Peptides and Proteins: Functional Molecular Features* (ed. B. R. Singh), American Chemical Society, Washington, D.C. **1999**; Chapter 3.
83. Fink, A. L.; Oberg, K.; Seshadri, S.; Khurana, R.; Gillespie, J. R. in: *Infrared Analysis of Peptides and Proteins: Functional Molecular Features* (ed. B. R. Singh), American Chemical Society, Washington, D.C. **1999**; Chapter 6.
84. Byler, D. M.; Lee, D. L.; Randall, C. S. in: *Infrared Analysis of Peptides and Proteins: Functional Molecular Features* (ed. B. R. Singh), American Chemical Society, Washington, D.C. **1999**; Chapter 7.
85. Krimm, S. in: *Infrared Analysis of Peptides and Proteins: Functional Molecular Features* (ed. B. R. Singh), American Chemical Society, Washington, D.C. **1999**; Chapter 2.

## Chapter 2

# Interpreting Infrared Spectra of Peptides and Proteins

Samuel Krimm

**Biophysics Research Division and Department of Physics,  
University of Michigan, Ann Arbor, MI 48109**

Characteristics of peptide group modes are analyzed, with a view to assessing the reliability of conformational assignments based on band deconvolution and frequency-structure correlations. It is concluded that this approach has severe problems, rendering it minimally dependable. A more definitive method has to be based on normal mode analysis of defined structures in comparison with observed spectra. Developments in theoretical and experimental directions are discussed that make such a program feasible.

The power of vibrational spectroscopy in determining structure and intra- and intermolecular forces of molecules is well known. The sensitivity of infrared (ir) bands to chemical composition has been the basis of valuable qualitative and quantitative analytical studies. The further observation that such absorption bands can reflect differences in conformation of a given chemical structure has raised hopes that ir spectroscopy can become an important tool in defining the three-dimensional structures of molecules, for example peptides and proteins. Much effort has been devoted to such studies, with results, however, that are of increasingly questionable significance. It is of importance, therefore, to assess where we stand at present in this endeavor and where we go from here.

The basis of any ir analysis of peptide conformation has to be a detailed understanding of the normal modes of vibration of each three-dimensional structure to be identified and a correlation of these modes with observed spectral bands. [It should, of course, not be forgotten that such bands result from transitions between energy levels of the molecule that are independently determined by these normal mode frequencies, which in turn are determined by the three-dimensional structure and the intramolecular and intermolecular forces.] Although initial results of this kind have been obtained on many standard conformations (1) (e.g., infinite models of the  $\alpha$ -helix and the  $\beta$ -sheet), we still lack such analyses for the many complex conformations found in peptides and proteins (e.g., distorted  $\alpha$ -helices, twisted  $\beta$ -sheets, finite structures, etc.). As a result, it has been tempting to pursue an easier course, viz., to assume that an observed spectral feature can be deconvoluted into a sum of "standard" contributions from a limited number of model structures. Thus, the broad amide I band contour of a protein would be "resolved" into individual components attributable to  $\alpha$ -helix,  $\beta$ -sheet, turns, "unordered", etc. (2), it being assumed that each structure

makes its contribution in a definable narrow frequency range. This is clearly an approximation, and it is beginning to be recognized that this approach has some basic problems associated with it (2-4), not only in the uncertainties involved in curve deconvolution but in the oversimplifications in treating normal modes of composite structures in terms of simple combinations of modes of individual model systems. This is therefore an opportune time to face up to the problems associated with this approach and to inquire into the directions needed to bring ir spectroscopy of peptides and proteins into its deservedly more mature phase.

To gain a better appreciation of the limitations associated with the band deconvolution approach, it is important to understand the nature of the characteristic peptide group modes (1), the so-called amide modes, on which assignments of chain conformation can be and are being made. We can then place in context the need for, and state of the art in, achieving more comprehensive analyses of actual peptide and protein structures.

### Characteristics of Peptide Group Modes

Early ir studies of molecules with amide groups (5), led to the recognition that specific vibrations in these groups could be associated with bands in roughly definable frequency regions. The nature and designation of these modes were given for N-methylacetamide (NMA) (6), and were explicitly described by a normal mode calculation of the in-plane modes on a simplified model (7) (CH<sub>3</sub> taken as a point mass). A recent complete ab initio calculation on the full molecule (8) does not lead to a substantial change in the nature of these modes, and they have since then (6, 7) been the basis for referring to and analyzing structure-sensitive bands in the ir spectra of peptides and proteins.

However, it is important to recognize that a simple model like NMA does not necessarily represent the additional chemical complexity of a peptide chain, nor can NMA serve as a guide to the additional interactions between peptide groups that must be present in a polypeptide chain. It is, therefore, relevant that we understand the nature of such amide modes in the more complex structures, starting with dipeptides (9, 10) and extending to polypeptides (1).

**Amide A.** Amide A is the designation for the band in the region near 3300 cm<sup>-1</sup> that is one component of a Fermi resonance pair (the other being amide B in the region near 3050 cm<sup>-1</sup>) between essentially 100% NH stretch (s), which predominates in amide A, and combinations and/or overtones from the 1500 -1700 cm<sup>-1</sup> region. Analysis of this resonance pair permits recovery of the unperturbed NH s frequency,  $\nu(\text{NH s})$  (11-13).

It has been thought that, since NH s is such a localized mode, it should not depend on conformation and that its frequency would be affected only by external interactions such as hydrogen bonding. Thus, a lower  $\nu(\text{NH s})$  would be associated with a stronger hydrogen bond, and vice versa. In fact, ab initio calculations on the alanine dipeptide (CH<sub>3</sub>-CO-NH-C<sub>α</sub>H(CH<sub>3</sub>)-CO-NH-CH<sub>3</sub>) show (10), Table I, that this is not the case, and  $\nu(\text{NH s})$  of even non-hydrogen-bonded groups is influenced by conformation. [The specific values of the  $\phi(\text{NC}_\alpha)$  and  $\psi(\text{C}_\alpha\text{C})$  conformational angles and the frequencies depend on basis set and scale factors, but calculations with larger basis sets and electron correlation (Mirkin, N.G.; Krimm, S., personal communication) confirm the principle of such a conformation dependence.] The underlying reason is not difficult to discern:  $\nu(\text{NH s})$  is fundamentally determined by the force constant,  $f(\text{NH})$ , which is related to the bond length,  $r(\text{NH})$  (9), and this can be influenced by the differing intramolecular interactions in the different three-dimensional structures,  $f(\text{NH})$  in general being inversely related to  $r(\text{NH})$ .

**Table I. Calculated Stretch Frequencies in (cm<sup>-1</sup>) of NH Groups in Non-Hydrogen-Bonded Conformers of CH<sub>3</sub>-CO-NH-C<sub>α</sub>H(CH<sub>3</sub>)-CO-NH-CH<sub>3</sub><sup>a</sup>**

Group	Conformer			
	$\beta_2$ (-134, 38) <sup>b</sup>	$\alpha_R$ (-92, 6) <sup>b</sup>	$\alpha_L$ (61, 41) <sup>b</sup>	$\alpha'$ (-162, -55) <sup>b</sup>
NH2	3455	3467	3461	3474
NH1	3431	3403	3413	3418

<sup>a</sup> Adapted from reference 10.

<sup>b</sup>  $\phi, \psi$  torsion angles (in degrees).

These results thus show that caution must be exercised in attributing shifts in unperturbed values of  $\nu(\text{NH s})$  entirely to changes in hydrogen-bond strength; some of the variation may be due to conformational changes. [Of course, shifts in amide A itself cannot be directly translated into changes in conformation or hydrogen bonding; this can be done only after recovering  $\nu(\text{NH s})$  from a Fermi resonance analysis.]

**Amide I.** The amide I mode is predominantly CO s, with small contributions to the potential energy distribution (PED) by CN s, C<sub>α</sub>CN deformation (d), and NH in-plane bend (ib) (the latter being mainly responsible for the downshift in  $\nu(\text{I})$  on N-deuteration). Conformation again is seen to affect  $r(\text{CO})$ , and thus  $f(\text{CO})$  and  $\nu(\text{CO s})$ , as found in the alanine dipeptide even in the absence of hydrogen bonding (10), Table II. [Differences also persist at higher level ab initio calculations.]

An additional influence on the frequency arises from the more composite nature of the amide I mode. [Table II only gives PED contributions  $\geq 10$ , but for example at the  $\geq 5$  level the 1694 cm<sup>-1</sup> mode of  $\beta_2$  is CO s2(81), NH ib2(9), C<sub>α</sub>CN d(8), C<sub>α</sub>C s(6).] Because the frequency of a normal mode is also determined by its eigenvector, a small change in this, resulting from structural differences (and also associated with possible changes in off-diagonal force constants) can alter the amide I frequency even though  $f(\text{CO})$  does not change much. An example of this is seen in the comparison of  $\nu(\text{CO s1}) = 1719$  of  $\beta_2$  and  $\nu(\text{CO s2}) = 1703$  of  $\alpha_L$  (Table II):  $f(\text{CO1}) = 11.400$  of  $\beta_2$  is only slightly larger than  $f(\text{CO2}) = 11.385$  of  $\alpha_L$ , but the significant difference in eigenvectors, CO s1(78), NH ib1(10), MCN d(7), MC s(6) for  $\beta_2$  (M = methyl C) vs CO s2(83), C<sub>α</sub>CN d(8), NH ib2(8), C<sub>α</sub>C s(5) for  $\alpha_L$ , results in a much larger than expected frequency difference. It is, of course, through this non-local character of a normal mode that we expect and find a significant conformational dependence of its bands.

Hydrogen bonding also gives rise to shifts in amide I frequencies, but caution is necessary in simply interpreting relative frequencies of observed bands. This is illustrated in Table III for two hydrogen-bonded conformers (CO1...NH2) of the alanine dipeptide, C<sub>7</sub>(eq),  $\phi, \psi = -85, 73$ , and C<sub>7</sub>(ax),  $\phi, \psi = 75, -62$  (9). On the basis of  $r(\text{O...H})$ ,  $f(\text{CO1})$ , and  $f(\text{NH2})$  values, it is clear that the C<sub>7</sub>(ax) conformer forms the stronger hydrogen bond, yet, counterintuitively, its CO s1 frequency is higher. This is undoubtedly a reflection of the difference in eigenvectors (additional components are MCN d(7), MC s(7), CO s2(6) for C<sub>7</sub>(eq) and MCN d(6) for C<sub>7</sub>(ax)), and emphasizes again the kinds of factors that determine frequencies of normal modes. [We note that the total energy of C<sub>7</sub>(ax) is higher than that of C<sub>7</sub>(eq), showing the importance of other interactions in stabilizing structure.]

A similar situation is encountered in the analysis of the amide I modes of  $\alpha$ -helix and  $\beta$ -sheet poly(L-alanine), (L-Ala)<sub>n</sub> (1). The unperturbed frequencies are

**Table II. Calculated Amide I Frequencies (in  $\text{cm}^{-1}$ ) and CO Bond Properties in Non-Hydrogen-Bonded Conformers of  $\text{CH}_3\text{-CO-NH-C}_\alpha\text{H(CH}_3\text{)-CO-NH-CH}_3$  <sup>a</sup>**

Property	Conformer			
	$\beta_2$	$\alpha_R$	$\alpha_L$	$\alpha'$
$\nu(\text{CO s1})$	1719	1726	1725	1710
$r(\text{CO1})^b$	1.220	1.218	1.219	1.222
$f(\text{CO1})^c$	11.400	11.542	11.486	11.300
PED <sup>d</sup>	CO s1(78), NH ib(10)	CO s1(84)	CO s1(79)	CO s1(82), NH ib(10)
$\nu(\text{CO s2})$	1694	1679	1703	1720
$r(\text{CO2})^b$	1.223	1.226	1.220	1.220
$f(\text{CO2})^c$	11.221	11.044	11.385	11.479
PED <sup>d</sup>	CO s2(81)	CO s2(84), NH ib2(10)	CO s2(83)	CO s2(83)

<sup>a</sup> Adapted from reference 10.

<sup>b</sup> Bond length in Å.

<sup>c</sup> Force constant in  $\text{mdyn}/\text{Å}$ .

<sup>d</sup> Potential energy distribution, contributions  $\geq 10$ . s = stretch, ib = in-plane bend.

**Table III. Calculated Amide I Frequencies (in  $\text{cm}^{-1}$ ) and CO Bond Properties of Hydrogen-Bonded Conformers of  $\text{CH}_3\text{-CO-NH-C}_\alpha\text{H(CH}_3\text{)-CO-NH-CH}_3$  <sup>a</sup>**

Conformer	Property <sup>b</sup>				
	$\nu(\text{CO s1})$	$r(\text{O}\cdots\text{H})$	$f(\text{CO1})$	$f(\text{NH2})$	PED
C <sub>7</sub> (eq)	1676	2.07	10.652	6.239	CO s1(70), NH ib1(15)
C <sub>7</sub> (ax)	1682	1.94	10.550	6.134	CO s1(55), CO s2(19), NH b1(15)

<sup>a</sup> Adapted from reference 8.

<sup>b</sup> See Table II.

$\nu(\alpha) = 1662$  and  $\nu(\beta) = 1670$ , yet the hydrogen-bond properties [ $\alpha$ :  $r(\text{O}\cdots\text{H}) = 1.88$ ,  $f(\text{CO}) = 10.029$ ,  $f(\text{NH}) = 5.830$ ;  $\beta$ :  $r(\text{O}\cdots\text{H}) = 1.75$ ,  $f(\text{CO}) = 9.882$ ,  $f(\text{NH}) = 5.674$ ] clearly show that the  $\beta$ -sheet has the stronger hydrogen bond. The reason that we observe a lower  $\beta$ -sheet (1632) than  $\alpha$ -helix (1658) frequency is not due, as has been claimed (2), to the stronger hydrogen bond in the  $\beta$ -sheet, but to the effect of another important interaction, transition dipole coupling (TDC).

Two vibrating (or transition) dipoles of the same frequency give rise to an interaction energy, and therefore a force constant, that depends on their spatial separation and relative orientation (14-16), thus making such an interaction conformation dependent. When a normal mode analysis of  $\beta$ -sheet polyglycine I,  $(\text{Gly})_n\text{I}$ , based only on a valence force field failed to account for the large observed splitting ( $49 \text{ cm}^{-1}$ ) of the amide I bands (17), the TDC interaction was calculated and was found to account for such splittings (17-19), with values of the transition moment consistent with ab initio values (16). This coupling is particularly important for amide

I because of the large transition moment ( i.e., dipole derivative) associated with this mode. The TDC interaction has been shown to explain many features of observed amide I bands (1), such as the larger ( $62\text{ cm}^{-1}$ ) splitting of antiparallel-chain pleated sheet (L-Ala)<sub>n</sub> compared to the smaller splitting of antiparallel-chain rippled sheet (Gly)<sub>n</sub>I and the very small splitting ( $\sim 1\text{ cm}^{-1}$ ) of  $\alpha$ -helical (L-Ala)<sub>n</sub>. Therefore, its effects must be incorporated in any analysis of the amide I band of a peptide or a protein. Whether it is the only important intergroup interaction needed to explain the general features of amide I band contours of proteins (20), i.e., to the exclusion of valence and hydrogen-bond interactions, remains to be seen. However, the presence of TDC emphasizes the interconnectedness of all peptide groups in a compact protein structure, and indicates that the effects of its incorporation certainly cannot be excluded in the analysis of amide I band contours.

In this connection, it has become clear that a TDC interaction treatment based on the previously accepted weak coupling (21) and perturbation (14-16, 21, 22) approximations cannot explain important details of amide I region spectra. It has been found that, in order to account for the observed A-E<sub>1</sub> species frequency separation and for the frequency of the newly assigned E<sub>2</sub> species Raman-active mode of  $\alpha$ -(L-Ala)<sub>n</sub> at  $1665\text{ cm}^{-1}$  (33), it is necessary to invoke a general non-weak-coupling-non-perturbation treatment of vibrations of helical molecules (24). With this generalization, not only are the above observations reproduced but an ab initio-based force field refinement for  $\alpha$ -(L-Ala)<sub>n</sub> can be implemented that reproduces all observed ir and Raman bands above  $200\text{ cm}^{-1}$  to better than  $5\text{ cm}^{-1}$  (25). The lesson to be learned from this is that if we want to extract the full structural content in vibrational spectra of complex molecules, we neglect subtle features of intermolecular interactions at our peril.

An important consequence of the (relatively long range) TDC interaction is that the profile of the amide I band will depend on the dimensions of a structure. This is because, while for an infinite regular structure only a few of the infinitely many phase differences between similar vibrations in adjacent groups can lead to optical activity, for a finite regular structure all modes with allowed phase differences are potentially active. An example of this is seen in the case of the  $\alpha$ -helix by comparing the amide I band contour of a 10-residue helix (26) with that of a 26-residue helix (27), Figure 1. [The calculations are based on an  $\alpha$ -helix with a point-mass side chain, a force field optimized for this structure (28), and dipole derivatives obtained from ab initio calculations of hydrogen-bonded NMA (29).] As can be seen in the 10-residue case, all 9 coupled peptide group modes are ir active, with varying intensities and TDC shifts; in the absence of TDC these modes span a frequency range of only  $2\text{ cm}^{-1}$ , compared to  $\sim 30\text{ cm}^{-1}$  with TDC. In the 26-residue case, the 25 coupled modes span similar ranges but with different TDC shifts and relative intensities, thus giving rise to a different band peak position and contour. Departures from regularity, either in variations in  $\phi, \psi$  or in  $f(\text{CO})$  (due to variations in hydrogen-bond strength), which are likely in proteins, lead to significant changes in coupling and thus in band contour, as can be seen in the calculated example of deoxymyoglobin (27). Of course, if the departures are known or reasonably hypothesized, these studies show that it is now possible to calculate an expected amide I band contour that compares favorably with the observed spectrum, as was shown to be the case for the coiled coil (30). Size-dependent band contours are also expected for  $\beta$ -sheet structures (22).

Thus, the position and shape of an amide I band of even an  $\alpha$ -helix are subject to a variety of structural and force constant variables, and the assumption of a "standard" spectrum is clearly an oversimplification. At the very least, any attempt to deconvolute an amide I band contour from a protein into single independent



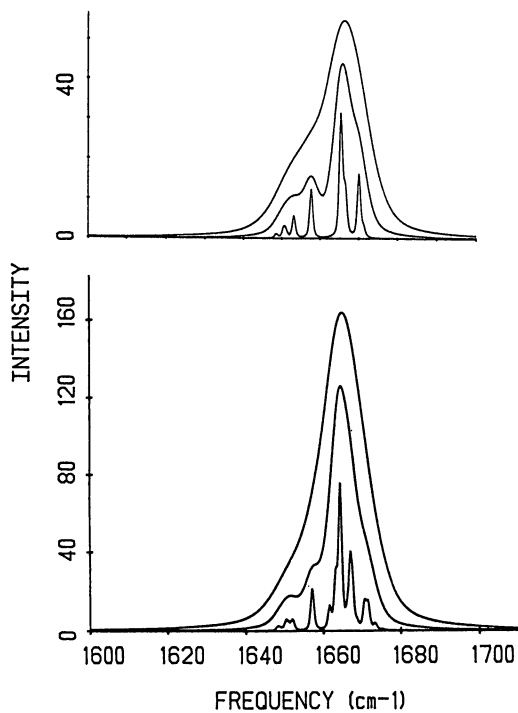


Figure 1. Amide I band profile of  $\alpha$ -helix of 10 residues (top) and 26 residues (bottom). Profiles are given for gaussian-lorentzian components at mode frequencies with FWHM bandwidths of 1, 5, and 10  $\text{cm}^{-1}$ . (Adapted from refs. 26 and 27.).

components with quantitatively assignable conformations is suspect, in view of the broad frequency range associated with contributions from a given structure and thus the extensive overlaps in a given frequency region of contributions from different structures.

**Amide II.** The dependence of frequency on eigenvector (and therefore on structure) that was noted in the amide I mode is even more striking in the case of amide II, a mode dominated by NH *ib* and CN *s* but with significant contributions from C $\alpha$ C *s*, CO *ib*, and NC $\alpha$  *s* (*I*). Although TDC also plays an important role in determining amide II frequencies, the dominant factor seems to be the PED contribution of NH *ib* (*I*). This can be seen from Table IV, which gives amide II frequencies of a number of

**Table IV. Observed Amide II Infrared Bands of Polypeptides <sup>a</sup> (in cm<sup>-1</sup>)**

Structure	$\nu(\text{obs})$	PED <sup>b</sup>				
		NH <i>ib</i>	CN <i>s</i>	C $\alpha$ C <i>s</i>	CO <i>ib</i>	NC $\alpha$ <i>s</i>
$\beta$ -(GluCa) <sub>n</sub>	1560	58	18		8	
(Gly) <sub>n</sub> II	1550	55	21	11	8	6
$\alpha$ -(GluH) <sub>n</sub>	1550	46	33	10	11	6
$\alpha$ -(Ala) <sub>n</sub>	1545	46	33	10	11	6
$\beta$ -(Ala) <sub>n</sub>	1524	41	26	13	14	
(Gly) <sub>n</sub> I	1517	35	28	17	14	

<sup>a</sup> Adapted from reference 1.

<sup>b</sup> Potential energy distribution, contributions  $\geq 5$ , from calculated normal modes.

different polypeptides in different conformations:  $\alpha$ -helix [(Ala)<sub>n</sub>, GluH]<sub>n</sub>,  $\beta$ -sheet [(Ala)<sub>n</sub>, (GluCa)<sub>n</sub>, (Gly)<sub>n</sub>I], and 3<sub>1</sub>-helix [(Gly)<sub>n</sub>II]. If these frequencies are extrapolated to zero PED of NH *ib*, a value near 1460 cm<sup>-1</sup> is obtained, which is in about the middle of the range of values of amide II', the dominant CN *s* mode of the N-deuterated polypeptides.

This spread of frequencies over side-chain composition and main-chain conformation does not make the amide II mode alone a dependable basis of secondary structure determination in polypeptides.

**Amide III.** On the basis of early normal mode studies of NMA (6, 7), it was assumed that the amide III mode was simply the counterpart of amide II, viz., essentially a mode with the opposite phase relation between NH *ib* and CN *s*. However, even the earliest normal mode analyses of  $\beta$ -(Ala)<sub>n</sub> (31) and  $\alpha$ -(Ala)<sub>n</sub> (32) showed that what we call the amide III mode is really a misnomer. In fact, amide III should be considered to be a set of complex modes involving various proportions of NH *ib*, CN *s*, H $\alpha$  bend (b), and other contributions, the specifics depending on side-chain composition and main-chain conformation. These conclusions are supported by more detailed normal mode analyses (*I*), and an example of this is given in Table V. We see that, if the criterion of an amide III mode is sensitivity to N-deuteration, many modes qualify by virtue of their NH *ib* PED component. The PEDs also show that it should not be surprising that these bands shift on C $\alpha$ -deuteration (33) or <sup>13</sup>C labeling (34) as well as on N-deuteration. When such treatment removes an internal coordinate (or changes a mass slightly), the remaining coordinates rearrange into a new set of eigenvectors,

**Table V. Observed Amide III Modes of Some Polypeptide Structures <sup>a</sup>**

Structure	$\nu(\text{obs})^b$	PED <sup>c</sup>					
		NH ib	CN s	H <sub>α</sub> b1	H <sub>α</sub> b2	C <sub>α</sub> C s	Other
β-(Ala) <sub>n</sub>	1402 ir	15			33	11	CH <sub>3</sub> sb(20)
	1399 R	13			32	14	CH <sub>3</sub> sb(17)
	1332 ir, R	25			42	12	CH <sub>3</sub> sb(11)
	1310 ir, R	23	13			14	CO ib(18)
	1243 R	19	13		34		NC <sub>α</sub> s(19)
	1224 ir, R	18	11		28		NC <sub>α</sub> s(24)
β-(GluCa) <sub>n</sub>	1260 ir, R	11	15		23		NC <sub>α</sub> s(17), C <sub>γ</sub> H <sub>2</sub> tw(14)
	1223 ir, R	11			22		C <sub>β</sub> H <sub>2</sub> tw(16), C <sub>γ</sub> H <sub>2</sub> tw(15), NC <sub>α</sub> s(10)
α-(Ala) <sub>n</sub>	1338 R	12		25	20	17	
	1278 R	23		18			NC <sub>α</sub> s(18)
	1270 ir, R	28		11	15		NC <sub>α</sub> s(14)
	1265 ir, R	40			28		
α-(GluH) <sub>n</sub>	1343 ir	11			16		C <sub>γ</sub> H <sub>2</sub> w(20)
	1296 R	15			24		C <sub>γ</sub> H <sub>2</sub> tw(23)
	1283 ir	25			29		C <sub>γ</sub> H <sub>2</sub> tw(11)

<sup>a</sup> Adapted from reference 1.<sup>b</sup> ir: infrared; R: Raman.<sup>c</sup> Potential energy distribution, contributions ≥10, from calculated normal modes.H<sub>α</sub> b1: bending in H<sub>α</sub>C<sub>α</sub>C<sub>β</sub> plane; H<sub>α</sub> b2: bending perpendicular to H<sub>α</sub>C<sub>α</sub>C<sub>β</sub> plane; sb: symmetric bend; w: wag; tw: twist.

whose frequencies can shift up or down with respect to their "parent" bands. In fact, by accurately predicting such isotopic shifts we can definitively validate a proposed structure.

Although these results seem to support many earlier assignments that separate  $\alpha$ -helix (in the 1265 -1295  $\text{cm}^{-1}$  region) from  $\beta$ -sheet (in the 1220 -1260  $\text{cm}^{-1}$  region) bands, the comparable  $\alpha$ -(Ala)<sub>n</sub> (1265) and  $\beta$ -(Glu)<sub>n</sub> (1260) bands should give one pause in firmly committing to this region for conformation "marker" bands. However, it is possible that, since NH ib and H $\alpha$  b couple in these modes, more specific information about the  $\phi$  angle may be obtainable from bands in this region (1).

**Amide V.** Another characteristic mode of the peptide group, designated amide V from the study of NMA (6, 7), involves a coupled NH out-of-plane bend (ob) and CN torsion (t). It has thus far not been very useful in identifying structure because, as a result of contributions in peptides from other coordinates, modes in this category have frequencies that are sensitive to hydrogen-bond strength and side-chain structure as well as main-chain conformation (1).

**Other Peptide Group Modes.** The study of NMA (6, 7) led to the designation of other peptide group modes: amide IV (CO ib), amide VI (CO ob), and amide VII (CN t). Because of the significant mixing in polypeptides of these with other coordinates, such modes are not easily identifiable and directly relatable to conformation (1). Only a normal mode analysis can provide a secure assignment, e.g., showing that an  $\alpha$ -helix band near 740  $\text{cm}^{-1}$ , though sensitive to  $^{13}\text{C}$  labeling, should be assigned to amide VI and not amide IV, as proposed (34).

Where does the above understanding of peptide group modes leave us with respect to achieving reliable spectral band-to-structure assignments? It must be said: not in a very strong position. While some general statements about conformation may be possible in cases of unambiguous spectral features, strong inferences about structural detail—and certainly in a quantitative setting—will never be achieved with certainty by such a "band aid" approach. Vibrational spectroscopy is capable of providing exquisitely detailed information about molecular structure and interactions, and it should be our goal to obtain such results for peptides and proteins. Only in this way can we achieve the maturity inherent in the method and hope to match the sophistication of other structural techniques such as NMR and x-ray diffraction.

## Future Directions

If our aim is to develop the full potentialities of peptide and protein spectroscopy, the path is clear: we must concentrate on the structure-to-spectrum route, i.e., on validating structural hypotheses by reliable normal mode predictions that can be compared with observed spectra, not incidentally extending the evaluation of agreement to the entire spectrum instead of depending on the analysis of a single band. Such a procedure at least provides unambiguous criteria for assessing congruence between proposal and reality.

It may be said that such a protocol is not likely to be productive, given the complexity of the structures we are dealing with and the difficulties in doing such calculations. With respect to the first objection, the situation is far from bleak: structural knowledge and insights often provide good starting hypotheses; molecular mechanics (MM) energy calculations are guides to possible structures; the set of similar conformations that are provided by NMR distance-geometry constraints can be analyzed and perhaps narrowed down; subtle changes in known structures can be

specified more accurately, particularly in environments not amenable to study by other methods; and many other important applications can be envisioned. With respect to calculational difficulties, modern computers and programs eliminate the computational problem (calculations have already been done on small (35) and large (20, 27) proteins), and developments in theoretical and experimental directions are making the proposed approach more feasible.

**Theoretical Directions.** We consider here two directions in theoretical developments that enhance the prospect of achieving accurate predictions of ir band frequencies and intensities for comparison with observed spectra.

**Improved Force Fields.** The key to obtaining accurate normal mode frequencies is a valid spectroscopic force field. In the absence of a full set of ab initio force constants, which can only be obtained for relatively small molecules, there are only two ways to proceed: improved empirical force fields or force constants derived from dependable MM potential energy functions. There has been recent progress in both areas.

The development of good empirical force fields for the polypeptide chain has led to deeper insights into the nature of the normal modes of such molecules (1). However, inherent in the empirical approach is the limitation imposed by the selection of which off-diagonal force constants to include in the optimization. This deficiency can be minimized by using results from ab initio force fields of small representative molecules, and such an ab initio-based force field has recently been optimized for  $\alpha$ -(Ala)<sub>n</sub> (25). Together with improved spectral data (23, 25), this resulted in the previously mentioned  $<5 \text{ cm}^{-1}$  agreement between observed and calculated frequencies. The consequence is not only that we can have greater confidence in the description of the normal modes, but we can use this improved force field to reliably examine other spectral properties, such as modes of finite helices (Lee, S.-H.; Krimm, S., to be published), thereby avoiding the limitations of approximate treatments (20), which, though qualitatively useful, do not give quantitative reproduction, for example of the asymptotic amide I A-E<sub>1</sub>, splitting of a regular  $\alpha$ -helix (36). Of course, such a force field is not directly transferable to other conformations, such as the  $\beta$ -sheet, since there is some conformational dependence of the force constants (1). Unless such dependence is explicitly known, empirical force fields are limited in their general applicability.

A natural way of obtaining conformation-dependent force constants is from the second derivatives of an MM energy function. This is because such functions consist of bonded terms (some of which may contain well-known conformation dependences) and non-bonded terms, and the latter (generally electrostatic and van der Waals) have an explicit dependence on conformation. The problem has been that most current MM functions have not been designed for spectroscopic accuracy, and when used in normal mode calculations give root-mean-square (rms) deviations of the order of  $30 \text{ cm}^{-1}$  or worse. Not only are such results spectroscopically unacceptable, but if one is interested in understanding the structural origins of small frequency shifts, they are useless.

We have been developing spectroscopically viable MM functions, what we call a spectroscopically determined force field (SDFF). This is based on an analytical transformation of a set of spectroscopic force constants, which together with structures and energies we obtain from (scaled) ab initio calculations on conformers of model molecules, into MM energy function parameters (37). With a systematic criterion for the reduction in the number of unimportant off-diagonal force constants, the ab initio

frequencies can be reproduced to any desired accuracy. We have implemented this procedure for linear (38) and branched (39) saturated hydrocarbon chains, with the result that 791 ab initio non-CH s frequencies of 21 linear and branched alkane conformers are reproduced with an rms deviation of  $6\text{ cm}^{-1}$ . An SDFP for the polypeptide chain is now under development (Palmo, K.; Mirkin, N.G.; Krimm, S., to be published), and should open the way to more routine normal mode studies of peptides and proteins.

**Intensity Predictions.** An ir spectrum contains more information about a molecule than is given by band frequencies alone, which are determined by structure and forces, viz., that provided by band intensities, which are also determined by the normal modes but in addition by the dynamic charge properties of the molecule (i.e., dipole derivatives). Thus, the ability to reproduce band intensities not only validates the correctness of a force field but also gives insight into electronic characteristics of a molecule. Knowledge of dipole derivatives associated with displacement coordinates, obtainable from ab initio calculations, make such spectral predictions possible, as was shown in the case of  $(\text{Gly})_n\text{I}$  (40), and provides a simple way of describing the coordinates that contribute to the ir intensity of a mode through a dipole derivative distribution (DDD) (41), which may differ from the PED. Variations of the DDD with conformation make it understandable that amide I absorptivities could vary with structural state (42). Dipole derivatives can also be converted into a model of fixed charges and charge fluxes to calculate intensities (43).

The above approaches are based on the use of a spectroscopic force field in the normal mode calculation. If one is to move to an MM normal mode calculation, as in the SDFP method, the constraints on intensity predictions are increased, since fixed charges must be consistent with values derived to reproduce structures and energies while fluxes must be consistent with intermolecular interactions that can determine intermolecular normal modes and band splittings. Despite these increased demands, we have shown that an extended electrostatic model in the context of an SDFP can satisfactorily account for ir intensities of linear alkanes (44). Figure 2 gives an example of the agreement that this model gives with ab initio results. There is every reason to believe that a similar model will be possible for peptides, raising the hope of producing complete simulations of their ir spectra.

**Experimental Directions.** Since experimental spectra are the necessary starting point for all analyses, it is useful to consider emphases in future directions that will enhance the development of a better understanding of peptide and protein structure. Of many possibilities, we discuss two, isotopic substitution and solvation studies.

**Isotopic Substitution.** In the very earliest ir studies of peptide systems (5), it was already evident that the relatively easy NH to ND substitution was important in identifying modes involving NH motions. The importance of more extensive use of this technique has been emphasized (1), and many N-deuteration studies, and even some  $^{12}\text{C}$  to  $^{13}\text{C}$  substitution experiments (45, 46, 34), have been undertaken. However, the full power of this technique remains to be exploited, and when compared to the broad-ranging specific isotopic substitution probes used in NMR, the ir effort is seen to be in its infancy.

Isotopic substitution at specific sites, combined with normal mode calculations, can be a powerful determinant of conformation in peptide systems. This principle is illustrated by the analysis of the  $\gamma$ -turn (47). The  $\gamma$ -turn is a reverse turn formed by three amino acid residues, with a hydrogen bond between the CO of residue 1 and the

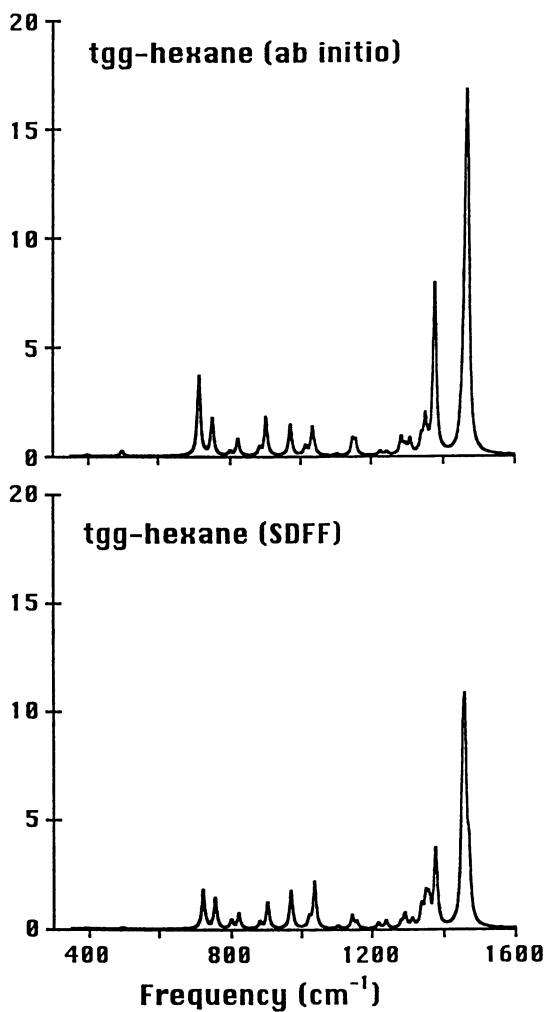


Figure 2. Simulated ab initio (top) and SDFP (bottom) infrared spectra of tgg hexane. Bands are based on lorentzian profiles with FWHM bandwidths of 10 cm<sup>-1</sup>. (Reproduced with permission from ref. 44. Copyright 1998 John Wiley & Sons, Inc.)

NH of residue 3. It can occur in three low-energy conformations: the  $\gamma$ -turn ( $\gamma$ ), a C<sub>7</sub>(ax) structure; the mirror-related  $\gamma$ -turn ( $\gamma_M$ ), a C<sub>7</sub>(eq) structure; and the inverse  $\gamma$ -turn ( $\gamma_I$ ), also a C<sub>7</sub>(eq) structure. Normal mode calculations were done on CH<sub>3</sub>-CO-(L-Ala)<sub>3</sub>-NH-CH<sub>3</sub> models of all three turns as well as ones with <sup>14</sup>C(1O) substitution (<sup>14</sup>C was used for illustrative purposes; <sup>13</sup>C would be the more appropriate experimental substituent). The resulting amide I frequencies are shown in Table VI. It can be seen that in principle it should be possible to distinguish these

**Table VI. Calculated Amide I Frequencies (in cm<sup>-1</sup>) of  $\gamma$ -Turns of a Model of CH<sub>3</sub>-CO-(L-Ala)<sub>3</sub>-NH-CH<sub>3</sub> <sup>a</sup>**

$\gamma^b$		$\gamma_M^b$		$\gamma_I^b$	
$\nu[^{12}\text{C}(1)]$	$\nu[^{14}\text{C}(1)]$	$\nu[^{12}\text{C}(1)]$	$\nu[^{14}\text{C}(1)]$	$\nu[^{12}\text{C}(1)]$	$\nu[^{14}\text{C}(1)]$
1684 (0, 3)	1683 (0, 3)	1675 (0, 3)	1675 (0, 3)	1675 (0)	1675 (0)
1670 (1)		1668 (2)	1666 (2)	1667 (2)	1664 (2)
1655 (2)	1654 (2)	1656 (1)		1660 (1)	
1653 (3, 0)	1653 (3, 0)	1654 (3, 0)	1659 (3, 0)	1649 (3)	1648 (3)
	1588 (1)		1570 (1)		1580 (1)

<sup>a</sup> Adapted from reference 46.

<sup>b</sup>  $\gamma$ :  $\gamma$ -turn;  $\gamma_M$ : mirror-related  $\gamma$ -turn;  $\gamma_I$ : inverse  $\gamma$ -turn. Number in parenthesis designates group(s) with significant CO s contributions to mode.

turns from the spectral results, in part from the order of shifted bands and in part from the magnitudes of the shifts. If ambiguities exist, these can probably be eliminated by further judiciously chosen isotopic substitution patterns. This approach undoubtedly involves more effort, but the reward is confidence instead of speculation in structure determination. [In this connection, it should be noted that since isotopic substitution obviously changes the eigenvector of a normal mode (48), the transition moment direction of a <sup>13</sup>CO amide I mode will be different from that of a <sup>12</sup>CO amide I mode. If they are assumed to be equal, as in some measurements of orientation from dichroic ratios (49), this could lead to significant errors. This problem does not arise in cases of conformation change alone, such as between the  $\alpha_I$ -helix and the  $\alpha_{II}$ -helix (50).]

**Solvation Studies.** Peptides and proteins are usually studied in solvated (generally aqueous) environments. It is therefore important to understand the effects of such environments on the vibrational properties of these molecules. It is well known that dielectric properties of the surrounding medium can cause large frequency shifts in a molecule. Such changes can now be computed (by reaction field calculations), and we need to develop a better understanding of these effects in assigning bands and understanding solvent perturbations. Two examples illustrate the potential for developing such understanding.

In analyzing intermolecular coupling effects in liquid NMA (51), it became important to identify the amide I mode of the monomeric species, which might be assignable to a component at 1674 cm<sup>-1</sup>. A similar band was observed at 1675 cm<sup>-1</sup> in the simpler NMA-CD<sub>3</sub>CN solution system, where intermolecular hydrogen bonding is not a major contributor. A reaction field calculation predicted that, for the dielectric constant of CD<sub>3</sub>CN, the gas-phase NMA frequency of 1722 would shift to 1665-1691 cm<sup>-1</sup> (depending on the cavity size taken for the NMA molecule). This provided convincing evidence for the monomer assignment in liquid NMA.



An analysis of the composite amide I band of NMA in water suggested that two resolved components arose from coupling of amide I to the water bending modes of associated hydrogen-bonded water molecules (52), the coupling giving rise to an observed splitting of  $20\text{ cm}^{-1}$ . Although an ab initio calculation of an isolated model system predicted a splitting of only  $6\text{ cm}^{-1}$ , addition of a reaction field resulted in an increase to  $23\text{ cm}^{-1}$  (48).

These and other (53) results show that solvation effects can be predicted, in the context of normal mode calculations, and that we can aim for a more complete understanding of environmental perturbations to vibrational spectra of peptides and proteins.

### Concluding Remarks

The discussion of the nature of the amide modes in a polypeptide chain makes it clear that we cannot accept a simple band frequency-to-structure correlation at face value. The characteristics of the normal modes and the ways in which they are influenced by details of the structure of the molecule and the interactions it experiences imply that such an approach can only provide a very blurred view of the structural landscape. Added to this general basic problem is the specific fact that the "unordered" component of protein structures is poorly defined, and the attempts to model it by the spectra of polyelectrolyte polypeptides are misguided. It is now well established (54, 55) that, as proposed earlier (56, 57), local order of the "extended helix" (also referred to (54) as a polyproline II) type is present in such systems, thus incorrectly biasing protocols based on such "standard" spectra. We should, therefore, seek more critical approaches.

We can sharpen our insight by employing more powerful methods of analysis, the most important of which is normal mode calculation in conjunction with a broad range of experimental studies. This approach will at least permit strong inferences to be made about peptide and protein structure from ir spectra.

The means to accomplish the computational part of such a program, primarily in having a reliable conformation-dependent polypeptide force field, are soon to be available. This should stimulate more extensive experimental studies, particularly in the examination of a wider variety of isotopic derivatives. With this kind of approach, we could indeed claim that the ir spectroscopy of peptides and proteins has come of age.

### Acknowledgments

My research in this area has been made possible by my collaborators and by financial support from the National Science Foundation, most recently through grants MCB 9601006 and DMR 9627786.

### Literature Cited

1. Krimm, S.; Bandekar, J. *Adv. Protein Chem.* **1986**, *38*, 181-364.
2. Jackson, M.; Mantsch, H.H. *Crit. Rev. Biochem. Mol. Biol.* **1995**, *30*, 95-120.
3. Surewicz, W.K.; Mantsch, H.H.; Chapman, D. *Biochemistry* **1993**, *32*, 389-394.
4. Goormaghtigh, E.; Cabiaux, V.; Ruyschaert, J.-M. *Subcell. Biochem.* **1994**, *23*, 405-450.
5. Sutherland, G.B.B.M. *Adv. Protein Chem.* **1952**, *7*, 291-318.
6. Miyazawa, T.; Shimanouchi, T.; Mizushima, S. *J. Chem. Phys.* **1956**, *24*, 408-418.

7. Miyazawa, T.; Shimanouchi, T.; Mizushima, S. *J. Chem. Phys.* **1958**, *29*, 611-616.
8. Mirkin, N.G.; Krimm, S. *J. Mol. Struct.* **1991**, *242*, 143-160.
9. Cheam, T.C.; Krimm, S. *J. Mol. Struct. (Theochem)* **1989**, *188*, 15-43.
10. Cheam, T.C.; Krimm, S. *J. Mol. Struct. (Theochem)* **1990**, *206*, 173-203.
11. Miyazawa, T. *J. Mol. Spectrosc.* **1960**, *4*, 168-172.
12. Krimm, S.; Dwivedi, A.M. *J. Raman Spectrosc.* **1982**, *12*, 133-137.
13. Lee, S.-H.; Mirkin, N.G.; Krimm, S. *Biopolymers* **1999**, in press.
14. Krimm, S.; Abe, Y. *Proc. Natl. Acad. Sci. U.S.A.* **1972**, *69*, 2788-2792.
15. Moore, W.H.; Krimm, S. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 4933-4935.
16. Cheam, T.C.; Krimm, S. *Chem. Phys. Lett.* **1984**, *107*, 613-616.
17. Abe, Y.; Krimm, S. *Biopolymers* **1972**, *11*, 1817-1839.
18. Moore, W.H.; Krimm, S. *Biopolymers* **1976**, *15*, 2439-2464.
19. Dwivedi, A.M.; Krimm, S. *Macromolecules* **1982**, *15*, 177-185.
20. Torii, H.; Tasumi, M. *J. Chem. Phys.* **1992**, *96*, 3379-3387.
21. Higgs, P.W. *Proc. Roy. Soc. London* **1953**, *A220*, 472-485.
22. Chirgadze, Yu. N.; Nevskaya, N.A. *Biopolymers* **1976**, *15*, 607-625.
23. Lee, S.-H.; Krimm, S. *J. Raman Spectrosc.* **1998**, *29*, 73-80.
24. Lee, S.-H.; Krimm, S. *Chem. Phys.* **1998**, *230*, 277-295.
25. Lee, S.-H.; Krimm, S. *Biopolymers* **1998**, *46*, 283-317.
26. Reisdorf, Jr., W.C., Ph.D. Thesis, University of Michigan, 1994.
27. Krimm, S.; Reisdorf, Jr., W.C. *Faraday Discuss.* **1994**, *99*, 181-197.
28. Dwivedi, A.M.; Krimm, S. *J. Phys. Chem.* **1984**, *88*, 620-627.
29. Cheam, T.C. *J. Mol. Struct. (Theochem)* **1992**, *257*, 57-73.
30. Reisdorf, Jr., W.C.; Krimm, S. *Biochemistry* **1996**, *35*, 1383-1386.
31. Moore, W.H.; Krimm, S. *Biopolymers* **1976**, *15*, 2465-2483.
32. Rabolt, J.F.; Moore, W.H.; Krimm, S. *Macromolecules* **1977**, *10*, 1065-1074.
33. Oboodi, M.R.; Alva, C.; Diem, M. *J. Phys. Chem.* **1984**, *88*, 501-505.
34. Overman, S.A.; Thomas, Jr., G.J. *Biochemistry* **1998**, *37*, 5654-5665.
35. Tasumi, M.; Takeuchi, H.; Ataka, S.; Dwivedi, A.M.; Krimm, S. *Biopolymers* **1982**, *21*, 711-714.
36. Torii, H.; Tasumi, M. *J. Chem. Phys.* **1992**, *97*, 92-98.
37. Palmo, K.; Pietila, L.-O.; Krimm, S. *J. Comput. Chem.* **1991**, *12*, 385-390.
38. Palmo, K.; Mirkin, N.G.; Pietila, L.-O.; Krimm, S. *Macromolecules* **1993**, *26*, 6831-6840.
39. Palmo, K.; Mirkin, N.G.; Krimm, S. *J. Phys. Chem. A* **1998**, *102*, 6448-6456.
40. Cheam, T.C.; Krimm, S. *J. Chem. Phys.* **1985**, *82*, 1631-1641.
41. Qian, W.; Krimm, S. *J. Phys. Chem.* **1993**, *97*, 11578-11579.
42. Jackson, M.; Haris, P.I.; Chapman, D. *Biochim. Biophys. Acta* **1989**, *998*, 75-79.
43. Torii, H.; Tasumi, M. *J. Mol. Struct.* **1993**, *300*, 171-179.
44. Palmo, K.; Krimm, S. *J. Comput. Chem.* **1998**, *19*, 754-768.
45. Deber, C.M. *Macromolecules* **1974**, *7*, 45-51.
46. Tadesse, L.; Nazarbaghi, R.; Walters, L. *J. Am. Chem. Soc.* **1991**, *113*, 7036-7037.
47. Bandekar, J.; Krimm, S. *Int. J. Pept. Protein Res.* **1985**, *26*, 407-415.
48. Mirkin, N.G.; Krimm, S. *J. Mol. Struct.* **1996**, *377*, 219-234.
49. Arkin, I.T.; MacKenzie, K.R.; Brünger, A.T. *J. Am. Chem. Soc.* **1997**, *119*, 8973-8980.
50. Reisdorf, Jr., W.C.; Krimm, S. *Biophys. J.* **1995**, *69*, 271-273.
51. Schweitzer-Stenner, R.; Sieler, G.; Mirkin, N.G.; Krimm, S. *J. Phys. Chem. A* **1998**, *102*, 118-127.

52. Chen, X.G.; Schweitzer-Stenner, R.; Krimm, S.; Mirkin, N.G.; Asher, S.A. *J. Am. Chem. Soc.* **1994**, *116*, 11141-11142.
53. Torii, H.; Tatsumi, T.; Tasumi, M. *J. Raman Spectrosc.* **1998**, *29*, 537-546.
54. Woody, R.W. *Adv. Biophys. Chem.* **1992**, *2*, 37-79.
55. Torii, H.; Tasumi, M. *J. Raman Spectrosc.* **1998**, *29*, 81-86.
56. Tiffany, M.L.; Krimm, S. *Biopolymers* **1968**, *6*, 1379-1382.
57. Krimm, S.; Mark, J.E. *Proc. Natl. Acad. Sci. U.S.A.* **1968**, *60*, 1122-1129.

## Chapter 3

# Fourier Transform Infrared Spectroscopic Studies of Peptides: Potentials and Pitfalls

Parvez I. Haris

Department of Biological Sciences, De Montfort University, Hawthorn Building, The Gateway, Leicester LE1 9BH, United Kingdom. (e-mail: pharis@dmu.ac.uk)

This article presents a discussion of the advantages and drawbacks of FTIR spectroscopy for structural characterisation of peptides. FTIR spectra of small peptides can be highly complex and extremely difficult to interpret. Amide bands associated with different secondary structural elements in peptides can show large deviations from those observed with large water-soluble proteins. This is partly due to the fact small peptides are often highly flexible and are solvated to a greater extent compared to globular proteins. Peptide aggregation can also result in the appearance of additional bands in the amide I region that can be erroneously attributed to peptide structure. These are amongst a host of factors that needs to be taken into consideration when interpreting infrared spectra of peptides.

Infrared spectroscopy is one of the earliest techniques to be used for protein conformational analysis (1). From the frequency, intensity and width of the amide bands in the infrared spectrum of a protein, it is possible to obtain highly detailed information on its structure. A survey of the current literature clearly indicates the popularity of Fourier transform infrared (FTIR) spectroscopy as a tool for structural characterisation of peptides and proteins (for reviews see ref. 2-12). This popularity can be attributed to a number of factors, including the ability to use this technique for recording spectra of biomolecules in diverse environments. It enables conformational analysis of a peptide to be carried out in aqueous solution, crystals, organic solvents, detergents, micelles as well as in phospholipid membranes. As such FTIR spectroscopy is one of the few techniques that can be used to monitor the role of the surrounding environment on protein conformation. This is important in light of increasing evidence

ch003

suggesting that the environment can have a significant role in determining the secondary structure formed by an amino acid sequence (13,14). A large proportion of FTIR studies on proteins, reported in the literature, involves the use of this technique for characterising structures in membrane systems. This can be partly explained by the fact that both X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy cannot be readily applied to study the conformation of peptides and proteins in phospholipid membrane systems. Although FTIR spectroscopy has many advantages for peptide structure analysis, it has its limitations. Indeed, there is no single technique that can provide information on all aspects of protein structure. Therefore, whenever possible it is always preferable to use a battery of techniques for structural characterisation of a peptide so that a more comprehensive picture of its molecular structure is derived. Often in the study of peptide structure, infrared spectroscopy is used in conjunction with other biophysical techniques such as NMR and electron spin resonance (ESR) spectroscopy. These techniques complement and supplement each other. For example, infrared spectroscopy provides a fast snap shot ( $10^{-12}$  s) of the peptide structure and complements the time scale of the NMR ( $10^{-5}$  s) and ESR ( $10^{-8}$ ) techniques.

With the advent of automated peptide synthesisers, there has been a tremendous growth in the use of synthetic peptides for diverse applications. Many studies utilise synthetic peptides for understanding protein folding and peptide-membrane interactions. Conformational analysis on peptides are mainly carried out using NMR, circular dichroism (CD) and FTIR spectroscopy. This Chapter presents a discussion on the use of FTIR spectroscopy for structural analysis of peptides in diverse environments. Examples are given of how the amide I band frequency can change depending on the type of environment encountered by a particular peptide secondary structure. It particularly focuses on some problem areas that needs to be taken into consideration when using FTIR spectroscopy for conformational analysis of peptides.

## Measurement and Analysis of Peptide FTIR spectra

**Water and Water Vapour - Problem of Overlap with Amide Bands.** Water is an essential component of all biological systems. It is therefore important to characterise the structure of biomolecules in an aqueous media. For a long time it was not possible to record infrared spectra of proteins in aqueous media due to the strong absorption of H<sub>2</sub>O over much of the mid-infrared region. Therefore, most of the early infrared studies were restricted to measurements in the solid state and in <sup>2</sup>H<sub>2</sub>O. Thanks to the advent of computerised FTIR spectrometers, it is now possible to digitally subtract the overlapping H<sub>2</sub>O absorption from the spectrum of the protein solution. A relatively high protein concentration (~10mg/ml) and short path length cells (6 μm) are needed for recording spectra of proteins in H<sub>2</sub>O. Especially in situations where the protein concentration is low, accurate H<sub>2</sub>O subtraction can be a problem. Over-subtraction or under-subtraction of the H<sub>2</sub>O spectrum can lead to the appearance of negative and positive bands in the final protein spectrum. This can lead to misleading conclusions being made regarding the secondary structure of a peptide or protein. In order to avoid such artefacts it is essential to ensure that the spectrum of H<sub>2</sub>O solvent is recorded under identical conditions as the protein solution spectrum (15,16,17). It is particularly important to ensure that the cell pathlength, temperature, pH, buffer concentration, number of scans

and spectral resolution are kept identical for the sample and solvent spectrum. Whenever possible, it is always a good practice to record the spectrum of a protein in  $^2\text{H}_2\text{O}$  for comparison with the spectrum in  $\text{H}_2\text{O}$ . Unlike  $\text{H}_2\text{O}$ ,  $^2\text{H}_2\text{O}$  does not absorb strongly in the amide I region and is therefore relatively free from problems encountered with solvent subtraction. In general, the amide I components observed for a peptide in  $^2\text{H}_2\text{O}$  are similar to those seen in  $\text{H}_2\text{O}$ , with the exception that bands are shifted to lower frequency in the former due to H- $^2\text{H}$  exchange. Besides  $^2\text{H}_2\text{O}$ , there are other solvents, including many organic solvents, that do not have strong absorption bands in the amide I region (1600-1700) and these can be readily used for FTIR analysis of peptides. Such solvents also offer the possibility of monitoring the effect of peptide concentration on the FTIR spectra of peptides, since it is possible to use much longer path length cells. In  $^2\text{H}_2\text{O}$  path lengths of up to 100 $\mu\text{m}$  can be employed and hence much lower peptide concentration ( $\sim 0.5\text{mg/ml}$ ) is required to obtain high quality FTIR spectra. This is particularly useful for studies on peptides and proteins that tend to aggregate and form inter-molecular  $\beta$ -sheet structures at high concentrations (see later).

Water present, in its gaseous form, in the sample compartment of an FTIR spectrometer can also be a problem in FTIR analysis of peptides. Acquisition of high signal-to-noise ratio spectra of peptides and proteins necessitates the elimination of water vapour from the sample compartment of the spectrometer. This is because the narrow water vapour bands overlap with the conformation sensitive amide I band. Water vapour contributions can be reduced by purging the instrument with dry air or nitrogen, and by employing a sample shuttle which allows the background to be signal-averaged concurrently with the sample. Also a pre-recorded water vapour absorption spectrum can be subtracted from the protein absorption spectrum to reduce the vapour bands still further. However, subtraction of water vapour bands can be complicated and over-subtraction or under-subtraction can introduce artefacts.

**Recording FTIR spectra: Transmission Vs ATR.** One of the important advantages of FTIR spectroscopy is that spectra of biological samples can be recorded in virtually all types of environment. Measurements can be performed in both solid, solution and even in the gaseous state. New approaches are constantly being developed for recording FTIR spectra of peptides and proteins under different conditions. For example, coupling of optical microscopes to an FTIR spectrometer enables the study of very small and minute quantities of biological samples in solution and in the solid state. Spectra of small crystals of peptides and proteins can be analysed using this method. Spectral analysis using the transmission method is by far the most popular approach for recording solution spectra of peptides and proteins. Attenuated total reflection (ATR) method is another approach for recording infrared spectra of biomolecules (18-20). The use of the ATR method for recording FTIR spectra of water soluble peptides and proteins is a relatively recent phenomenon. Most of the early studies using this approach have involved the investigation of orientation of components of thin films, such as lipids and peptides in biomembranes. While such studies have been particularly useful, Jackson and Mantsch (21) recently published a study highlighting the drawbacks of this approach for characterising water soluble proteins. According to this study, there are two major difficulties with this approach. Firstly, the extent of hydration of proteins in solution and in thin films may vary significantly, resulting in large spectral differences as observed by

these authors for concanavalin A. Secondly, proteins may undergo changes in their secondary structure, once deposited onto infrared transmitting substrates and dried to form thin films. For example, Jackson and Mantsch (21) reported large changes in the spectrum of myoglobin when it was dried onto the ATR crystal.

**Techniques for qualitative and quantitative analysis of FTIR spectra.** It is the availability of a range of data processing techniques that make possible the extraction of detailed information from infrared spectra of peptides and proteins. The infrared spectrum of a peptide or protein is extremely complicated due to overlap of bands arising from different functional groups present within the system. The narrow frequency dispersion, together with the large width of protein bands, results in highly overlapped spectra that are extremely difficult to interpret. Different approaches have been developed that can help overcome these problems. These include methods such as derivative, deconvolution and difference spectroscopy (2-4, 10-12). It is through the application of these techniques that subtle changes in protein structure can be identified.

Often referred to as "resolution" enhancements procedures, both second-derivative and deconvolution do not increase the spectral resolution. They essentially involve narrowing the widths of infrared bands, allowing increased separation of the overlapping components present within the broad band envelope. This is useful for the identifying different structures present in a protein and also for detecting conformational changes. They provide a means for monitoring small changes in the frequency and intensity of protein amide bands (2-4, 10-12) which in turn can be related to structural changes in a peptide. Application of the second-derivative and deconvolution procedures often results in the magnification of noise and enhancement of bands arising from water vapour in the spectra of peptides. Consequently, such procedures should only be applied to peptide spectra that are of high signal-to-noise ratio, and are essentially free from absorption of water vapour bands.

A more powerful approach for detecting subtle changes in infrared spectra involves the subtraction of a protein absorbance spectrum in state A from that of the protein in state B. Such procedures simplify the spectrum, since absorption from those groups within the protein that remain unchanged in their structure is subtracted. The resultant difference spectrum only shows peaks that are associated with those groups involved in the conformational change. The method works best when spectra of a protein can be recorded in two different states with minimal disturbance. The ideal condition for difference spectroscopy is to have spectra recorded, for a protein in two states, with the same sample and the same cell. This ensures that there are no changes in the pathlength and sample concentration. Consequently, much of this types of spectroscopic studies have been most effectively carried out on proteins that contain a chromophore molecule that can be activated by an external trigger such as light. Light-activated membrane proteins such as the photosynthetic reaction centres, bacteriorhodopsin and rhodopsin have been extensively studied using FTIR difference spectroscopy (5-9). Recently, the development of photochemical caged compounds (6), which can be triggered by light to release a ligand, have extended the application of difference spectroscopy to include other proteins such as the  $\text{Ca}^{2+}$ -ATPase. Electrochemical cells (6) are also available which can be used to investigate redox induced changes in protein structure. Assignment of bands in the difference spectra is not trivial. With large proteins, assignments are made with the aid of site-directed mutagenesis and chemical

modifications such as isotopic labelling of specific residues. With small peptides the assignments of bands should be simpler and the introduction of  $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^2\text{H}$  labelled amino acids would further simplify this task.

Several methods have been developed for the quantitative analysis of protein infrared spectra. As yet there are no methods available that are suitable for characterisation of peptides and membrane proteins. The first method to be introduced is based on curve-fitting analysis of deconvolved amide I band (22). This procedure provided a very good estimate of protein secondary structure and has been adopted by others with some modifications (3,19). There are, however, a number of problems associated with the curve-fitting procedure. These include the need to assign all the bands within the amide I region. Another assumption in this method is that the molar absorptivities of the bands associated with different secondary structural elements are assumed to be identical. Studies on poly-L-lysine have shown that the molar absorptivities of different secondary structures can be significantly different (23,24) There is also a significant element of subjectivity associated with curve-fitting, such as the initial choice of input parameters. In order to overcome some of these problems, other methods have also been developed (25-30). One of these methods is based on pattern recognition. This approach avoids the need to deconvolve the amide I band and the assignment of the bands to different structures. It is similar to the method used in CD and Raman spectroscopy (31). It involves the use of a calibration matrix of the infrared spectra of proteins of known X-ray structure. Mathematical tools such as factor analysis (25) and partial least squares analysis (26) are used in the analysis of the infrared spectra. In general, these methods show a good correlation between the infrared estimates and those calculated from X-ray data and compares well with values obtained using other spectroscopic techniques such as CD spectroscopy. There are a number of problems still associated with these methods. These include normalisation procedures which assumes equal absorptivity of amide bands irrespective of secondary structure variation. Furthermore, the technique encounters difficulties in cases where the spectral properties of the unknown protein lie outside the properties of the spectra within the calibration set. In these situations an incorrect estimation of the secondary structure may be possible. Both the curve-fitting and pattern recognition methods suffer from the problem of overlap of amino acid side chain absorbance in the amide I region (29,32). It has been shown that subtraction of amino acid side absorption from the amide I and amide II region of protein infrared spectra can result in a slight in accuracy of the secondary structure prediction (32). Another drawback of the current FTIR secondary structure prediction methods, is that they are based on infrared spectra of soluble proteins of known X-ray structures. This is the case with the pattern recognition method. Consequently, such methods are not readily applicable for estimating the secondary structure of membrane proteins. This situation is unlikely to change in the immediate future due to the scarcity of membrane proteins with known X-ray crystal structures.

**Impurities in Peptide FTIR Spectra - Detection and Overlap with the Amide Bands.** Infrared spectrum of a sample will contain absorbance bands from all molecules that have infrared-active functional groups. Hence, it can suffer from the problem of spectral overlap due to absorbance arising from impurities. If the band arising from the impurity is known, than FTIR spectroscopy can be used for detection of that impurity in a protein sample. The absorbance of trifluoroacetic acid (TFA) in peptide samples is a



good example of this. TFA is commonly used for the cleavage of peptides from the resin, after synthesis, and also in HPLC purification procedures. This molecule can remain strongly associated with the peptide. This has not been considered to be a problem in functional studies, or in structure characterisation using other spectroscopic techniques such as NMR spectroscopy. However, in the infrared spectrum the strong absorption band associated with COO<sup>-</sup> stretching vibration of the TFA counterion is observed. This overlaps with the amide I band and show a maximum at 1673cm<sup>-1</sup>. In several studies this band has been erroneously attributed to peptide secondary structure. The presence of TFA may have structural and functional implications for peptides that needs to be taken into consideration. As such FTIR spectroscopy can be an extremely valuable tool for detecting presence of TFA impurity in synthetic peptides. In our own studies we have used FTIR spectroscopy to detect absorbance from a host of different molecules, that overlap with protein bands, such as those arising from buffers, salts, detergents and carbohydrates. Although these molecules may not have any effect on the function of a protein, they can certainly give rise to bands in an infrared spectrum causing erroneous spectral interpretations to be made. Therefore, it is critical to identify the chemical composition of a protein sample before finalising spectral interpretations.

### Information on Peptide Secondary Structure from FTIR Spectra

The vast majority of studies related to protein FTIR spectroscopy involves the analysis of protein secondary structure using the amide bands. However, as can be seen from Table I, the problem of band assignment is a serious problem in such studies.

**Relationship between amide bands and protein secondary structure.** As discussed by Krimm in his Chapter (Chapter 2), normal-mode calculations and experimental studies on peptides and proteins have shown a good correlation between peptide structure and amide band frequency. There are a number of amide mode vibrations which are potentially useful for protein conformational studies. By far the most widely used of these bands is the amide I band (2-5,10-12), closely followed by the amide II. The amide III (33,34) and amide A (35,36) bands have also been successfully used. The usefulness of the Amide III band for protein secondary structural analysis has been most effectively demonstrated by Singh and co-workers (*ref. 34 and also see Chapters 1 and 5 for further details*).

The amide I vibrational mode arises predominantly from C=O stretching vibrations of the amide group [2-12]. The absorption for this band occurs in the region 1600-1700cm<sup>-1</sup>. Hydrogen bonding and the coupling between transition dipoles are amongst the most important factors governing conformational sensitivity of the amide bands (see Chapter 2 by Krimm). Correlation between the frequency of the amide I band and peptide secondary structure has been best demonstrated with homopolypeptides which are known to adopt a particular secondary structure under specific experimental conditions. This has been most effectively demonstrated with poly-L-lysine. Alteration in the pH of this polypeptide induces interconversion between various secondary structural classes (23,24,37). The amide I band position for the  $\alpha$ -helical conformation is located at 1638cm<sup>-1</sup>. The  $\beta$ -sheet band frequency is located at 1610cm<sup>-1</sup> and a weaker band associated with high-frequency vibration of anti-parallel  $\beta$ -sheet structure is seen at

1680 $\text{cm}^{-1}$ . The second-derivative analysis of this absorbance spectrum does not reveal the presence of any other significant components. Unlike homopolypeptides, FTIR spectra of most peptides and proteins consist of a mixture of different secondary structures. Consequently, the spectra of such peptides and proteins can have extensive overlap of bands arising from the different secondary structures. There is an additional factor which distinguishes homopolypeptides from other peptides and proteins. This concerns the position of the amide I band components in homopolypeptides, associated with different secondary structures, which can be significantly different from those observed in other peptides and proteins. Thus for example, the  $\alpha$ -helical band frequency is observed at 1638 $\text{cm}^{-1}$  for poly-L-lysine, whereas in most proteins and peptides the  $\alpha$ -helical absorption occurs in the range 1650-1658 $\text{cm}^{-1}$  (see below). This can be attributed to differences in hydrogen-bonding strengths. It appears that the hydrogen-bonding is stronger for the secondary structural elements in homopolypeptides compared to naturally occurring peptides and proteins which usually consist of a mixture of different amino acids.

**Absorption of  $\alpha$ -Helical Structures.** The overwhelming majority of proteins of known structure show infrared absorption in the range 1650 $\text{cm}^{-1}$ -1658 $\text{cm}^{-1}$ , in  $\text{H}_2\text{O}$  and  $^2\text{H}_2\text{O}$  (2-4, 10-12), for the  $\alpha$ -helical conformation. In  $\text{H}_2\text{O}$  there can be extensive overlap of absorption of the  $\alpha$ -helical and random coil structures. Normally, these two types of structures can be distinguished by carrying out experiments in  $^2\text{H}_2\text{O}$ . In this media rapid hydrogen-deuterium exchange of the peptide N-H groups in random coil structure results in a large shift of its absorbance towards lower frequency (1644 $\text{cm}^{-1}$ ) (2-4, 10-12).

The sensitivity of the amide I vibration to different factors can result in significant deviations in the absorption frequency of  $\alpha$ -helical structures. Indeed,  $\alpha$ -helical structure has been reported to absorb between 1664-1630 $\text{cm}^{-1}$  (see Table I), overlapping with regions that are thought to arise from  $\beta$ -sheet and turn structures. For example, as discussed above, the amide I maximum for the  $\alpha$ -helical structure in the homopolypeptide poly-L-lysine occurs at 1638 $\text{cm}^{-1}$ , significantly different from that observed with most proteins and peptides (1650-1658 $\text{cm}^{-1}$ ). Furthermore, with highly solvent exposed helices (38,39) in  $^2\text{H}_2\text{O}$ , the amide I band can shift to a low frequency of 1644 $\text{cm}^{-1}$ . This has been illustrated with calmodulin which displays amide I maximum at 1644 $\text{cm}^{-1}$  in  $^2\text{H}_2\text{O}$ . Its amide I band maximum shifts to higher frequency in the presence of sucrose (see Figure 1). This is due to a reduction in the solvation of calmodulin by  $^2\text{H}_2\text{O}$  molecules. This clearly demonstrates how solvation can have an important role in determining the position of the amide I band frequency of a protein. Recently, amide I band position as low as 1632 $\text{cm}^{-1}$  has been attributed to solvated  $\alpha$ -helices (40). This was reported for a small peptide. It is possible for the band frequency for an  $\alpha$ -helical structure in a short, solvent exposed, peptide to be significantly different from the absorption of helical structures which may be buried within a solvent inaccessible region of a highly folded globular protein. In contrast, there are situations where interaction with water is reduced, as is the case with proteins embedded in the hydrophobic core of phospholipid membranes. Under this condition it is expected that the amide I band frequency will shift to higher frequency. Indeed, this is what is observed from a study of a large range of membrane proteins (41; see Table II). For these proteins  $\alpha$ -helical absorption occurs in the range 1656-1658 $\text{cm}^{-1}$ , compared to water-soluble proteins

**Table I: Conflicting amide I assignments for the different types of protein secondary structures**

<b><u>Standard assignments</u></b>	<b><u>Alternative assignments</u></b>
<b><u>1. <math>\alpha</math>-helix</u></b>	
1630-1645 $\text{cm}^{-1}$	B, R, C, T,3
1646-1659 $\text{cm}^{-1}$	B, R, L, C, T
1660-1665 $\text{cm}^{-1}$	T, B, R, C,3
<b><u>2. <math>\beta</math>-sheet</u></b>	
1610-1624 $\text{cm}^{-1}$	A, S, T
1625-1645 $\text{cm}^{-1}$	T, 3, H, L, R, C
1670-1695 $\text{cm}^{-1}$	T, R,3
<b><u>3. <math>\beta</math>-turn</u></b>	
1610-1645 $\text{cm}^{-1}$	A, B, R, S, C, T, H,L,3
1660-1695 $\text{cm}^{-1}$	B, 3, H, L
1670-1695 $\text{cm}^{-1}$	T, R,3, A
<b><u>4. Random Coil</u></b>	
1640-1660 $\text{cm}^{-1}$	H, T, C, 3, B, L
1661-1675 $\text{cm}^{-1}$	H, T, 3, B

B,  $\beta$ -sheet; C, coiled coil; H,  $\alpha$ -helix; L, loops, 3,  $3_{10}$ -helix; T,  $\beta$ -turn; R, random coil; A, aggregated structures (inter-molecular  $\beta$ -sheet); L, loops; S, side chains. The assignments are based on a survey of the current literature on protein infrared spectroscopy.

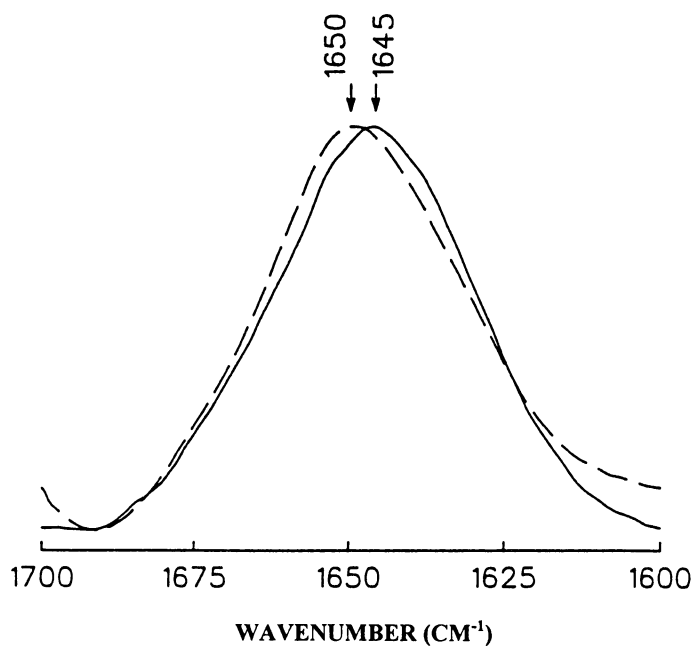


Figure 1. FTIR absorbance spectra in the amide I region of calmodulin in the presence (broken line) and absence (continuous line) of 60% deuterated glycerol. The spectra were recorded for the protein in  $^2\text{H}_2\text{O}$ .

**Table II: Amide I maxima for some membrane proteins studied using FTIR spectroscopy**

<i>Membrane Proteins and their source</i>	<i>Amide I Maxima in Wavenumbers (cm<sup>-1</sup>)</i>	<i>Predominant Conformation</i>
<b>ANIMAL</b>		
Cytochrome C Oxidase	1656	$\alpha$ -helix
Ca <sup>++</sup> -ATPase	1656	$\alpha$ -helix
H <sup>+</sup> /K <sup>+</sup> -ATPase	1656	$\alpha$ -helix
Na <sup>+</sup> /K <sup>+</sup> ATPase	1656	$\alpha$ -helix
Glucose Transporter	1656	$\alpha$ -helix
Rhodopsin	1656	$\alpha$ -helix
<b>PLANT</b>		
Photosystem II reaction centre	1656	$\alpha$ -helix
<b>BACTERIA</b>		
Bacteriorhodopsin from <i>Halobacterium salinarium</i>	1662	$\alpha$ -helix
Reaction Centre from <i>Rhodospseudomonas sphaeroidis</i>	1656	$\alpha$ -helix
Porin from <i>Rhodobacter capsulatus</i>	1629	$\beta$ -sheet
<b>VIRUS</b>		
M13 and Pfl viral coat proteins (in phospholipid membranes)	1658	$\alpha$ -helix

which normally display absorbance in the region  $1650\text{-}1655\text{cm}^{-1}$ . Even in  $^2\text{H}_2\text{O}$ , the  $\alpha$ -helical band position for membrane proteins shows a much reduced shift towards lower frequency compared to water-soluble proteins. This is likely to be due to a reduction in the interaction of  $\text{H}_2\text{O}$  molecules with peptide chains that are embedded in the hydrophobic core of a phospholipid membrane.

Non-standard  $\alpha$ -helical structures can have absorption frequencies well below or higher than the frequency associated with standard  $\alpha$ -helices. For example, it has been demonstrated recently that coiled-coil proteins display their amide I maxima in the region of  $1630\text{cm}^{-1}$  (42). This is in the spectral region normally associated with  $\beta$ -sheet structure. Coiled-coil structure consists of two parallel  $\alpha$ -helices wrapped around one another in a left-handed supercoil. In contrast the  $\alpha_{\text{II}}$  helical structure has been suggested to absorb at higher frequencies (43). The amide I maxima for the  $\alpha_{\text{II}}$  helices occur at a higher frequency due to the fact such structures have longer hydrogen-bond length compared to  $\alpha_{\text{I}}$  helices. The high frequency amide I maximum for bacteriorhodopsin was suggested to be due to the presence of  $\alpha_{\text{II}}$  helices (43) (see below). Reisdorf and Krimm suggested that the  $\alpha_{\text{I}}$  and  $\alpha_{\text{II}}$  helices can be distinguished by infrared linear dichroism studies (42). The frequency associated with the amide I band for an  $\alpha$ -helix can also depend on the length of the  $\alpha$ -helix. Thus one has to take into consideration different aspects of the helical structure in order to reach a reliable conclusion regarding the assignment of an amide I band.

From the discussion presented above, it is clear that the assignment of a band in the region  $1650\text{-}1658$  to  $\alpha$ -helical structure is not straightforward. There are a host of factors that can cause the frequency to deviate significantly from this range. Often the major problem is ascertaining which factor is responsible for the shift. For example, is it due to a change in secondary structure or due to a mere change in the environment (such as an increase in solvation). These are issues that will need to be addressed if unambiguous spectral assignments are to be made. The large differences that one observes in the position of amide I bands for helical conformation in several peptides and proteins are highlighted by the examples presented below.

Alamethicin and melittin are two naturally occurring peptides that form pores in membranes. Alamethicin (44,45) consists of 19 amino acid residues. Crystals of this peptide have been obtained from acetonitrile/methanol solution and its high resolution X-ray structure determined (46). The structure is predominantly  $\alpha$ -helical with a bend in the helix generated by the proline at position 14. The X-ray structure also reveals the presence some short segments of  $3_{10}$ -helical regions. Melittin is a 26 amino acid residue long peptide and its X-ray diffraction structure also shows a predominantly  $\alpha$ -helical conformation (47). Both melittin and alamethicin have been extensively studied using FTIR spectroscopy (44,48-50). These studies reveal some interesting differences and similarities between the two peptides. Most significantly, the amide I maximum for alamethicin occurs near  $1660\text{-}1665\text{cm}^{-1}$  as compared to  $1650\text{-}1655\text{cm}^{-1}$  for melittin. This suggests that melittin has a more normal helical structure as compared to alamethicin. Haris and Chapman (50) suggested that the high band frequency in alamethicin is probably due to an overlap of absorbance arising from a combination of  $\alpha$ -helix and  $3_{10}$ -helix structures (50). Furthermore, although alamethicin does not contain any  $\beta$ -sheet structure, it displays a band near  $1640\text{cm}^{-1}$ . This band was assigned to  $3_{10}$ -helical structure (50). Others have also assigned bands in this region, in the spectra of peptides

and proteins, to  $3_{10}$ -helical structure (51,52). Detailed FTIR analysis has been carried out on peptides which are known to exist in  $3_{10}$ -helical structures (53). Results of these studies indicate that in deuterated chloroform, peptides known to contain  $3_{10}$ -helical structure show their main amide I band at  $1666\text{-}1662\text{cm}^{-1}$ . This is similar to that observed for alamethicin (50). Second-derivative and deconvolution analysis of these model peptides also reveal weak components near  $1646\text{cm}^{-1}$  and  $1679\text{cm}^{-1}$ . These bands were tentatively attributed to type III  $\beta$ -turns (53). Results of the studies on these model peptides as well as with alamethicin indicate that FTIR spectroscopy is a sensitive tool for distinguishing between  $3_{10}$  and standard  $\alpha$ -helical structures.

Bacteriorhodopsin is a retinal containing light-driven proton pump found in the purple membrane of *Halobacterium salinarium*. Light activation results in the transfer of protons across the biomembrane from the cytoplasmic to the extracellular side via the Schiff base. The structural model of bacteriorhodopsin based on the electron diffraction and X-ray diffraction data has been used in conjunction with FTIR spectroscopy to obtain a detailed understanding of the molecular changes associated with the functional cycle of this protein (5,7). Figure 2 compares the absorbance spectrum of bacteriorhodopsin and porin. The latter displays its amide I maximum near  $1629\text{cm}^{-1}$  whilst the former shows its maximum at  $1662\text{cm}^{-1}$ . Currently, these two proteins have the most unusual amide I band maxima for any membrane proteins studied using FTIR spectroscopy. Bacteriorhodopsin has the highest amide I maxima and porin has the lowest. The low frequency of porin can be explained as being due to the presence of  $\beta$ -sheet structure in accordance with its X-ray diffraction data (54). Whilst, the band maximum for bacteriorhodopsin has been a matter of controversy for many years as discussed below.

Figure 3 compares the second-derivative spectrum of bacteriorhodopsin and myoglobin recorded in an aqueous media. According to X-ray crystallographic data both proteins are predominantly  $\alpha$ -helical with a distinct lack of  $\beta$ -sheet structure. The most notable difference in the spectrum of myoglobin and bacteriorhodopsin is in the position of their amide I maximum. Myoglobin shows an amide I maximum at  $1655\text{cm}^{-1}$  consistent with what is normally observed for  $\alpha$ -helical structures. However, it is difficult to explain the unusually high amide I maximum of  $1662\text{cm}^{-1}$  for bacteriorhodopsin. Different explanations have been put forward to explain this high frequency, including the possibility that this may be due to the presence of some distorted or  $\alpha_{II}$ -helices (5,43) or a mixture of short  $3_{10}$ -helical and normal  $\alpha$ -helices (50). The high-resolution structure of bacteriorhodopsin does not show the presence of  $\alpha_{II}$ -helices, and  $3_{10}$  helices were only present in the ends of some of the helices. Like alamethicin, it is possible that the high frequency band in bacteriorhodopsin is likely to be due to the presence of a mixture of  $3_{10}$  and normal helical structures (50).

As can be seen from Figure 3 both myoglobin and bacteriorhodopsin display another band near  $1630\text{cm}^{-1}$ , raising the possibility that these proteins contain some  $\beta$ -sheet structure. However, the crystal structures of these proteins do not support such a possibility. It has been suggested that bands in this region in the spectrum of myoglobin can be assigned to short extended segments that do not form  $\beta$ -sheet structures. More recently, model calculations by Torii and Tasumi, 1992 (55) for the protein myoglobin indicate that bands below  $1640\text{cm}^{-1}$  may also arise from vibrational motions of  $\alpha$ -helical structures. Similar explanations can also apply for bacteriorhodopsin. However, the

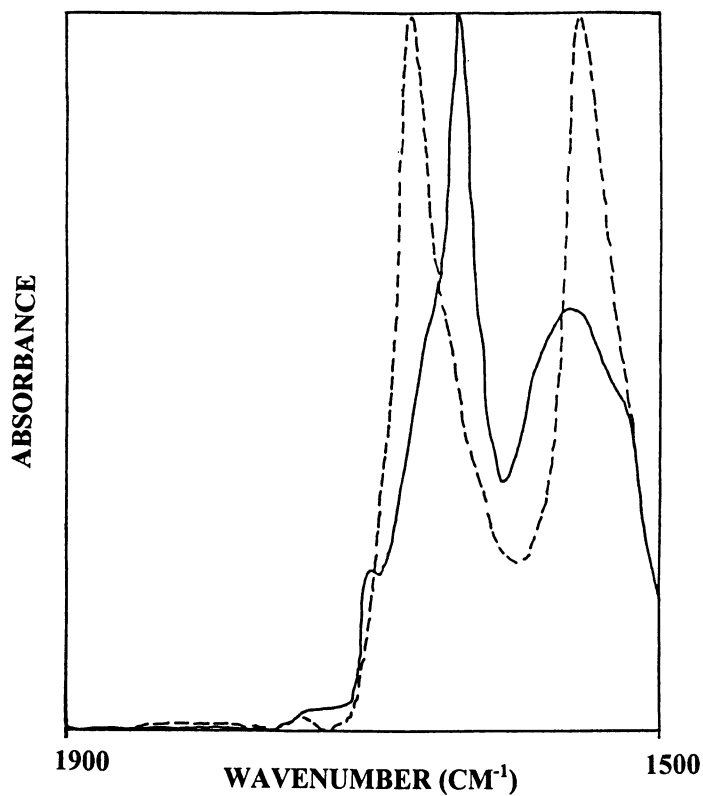


Figure 2. Comparison of the FTIR absorbance spectra of the membrane proteins bacteriorhodopsin and porin. Bacteriorhodopsin displays an unusually high amide I maximum near  $1662\text{cm}^{-1}$ . Spectra were recorded for samples in  $\text{H}_2\text{O}$  buffer.



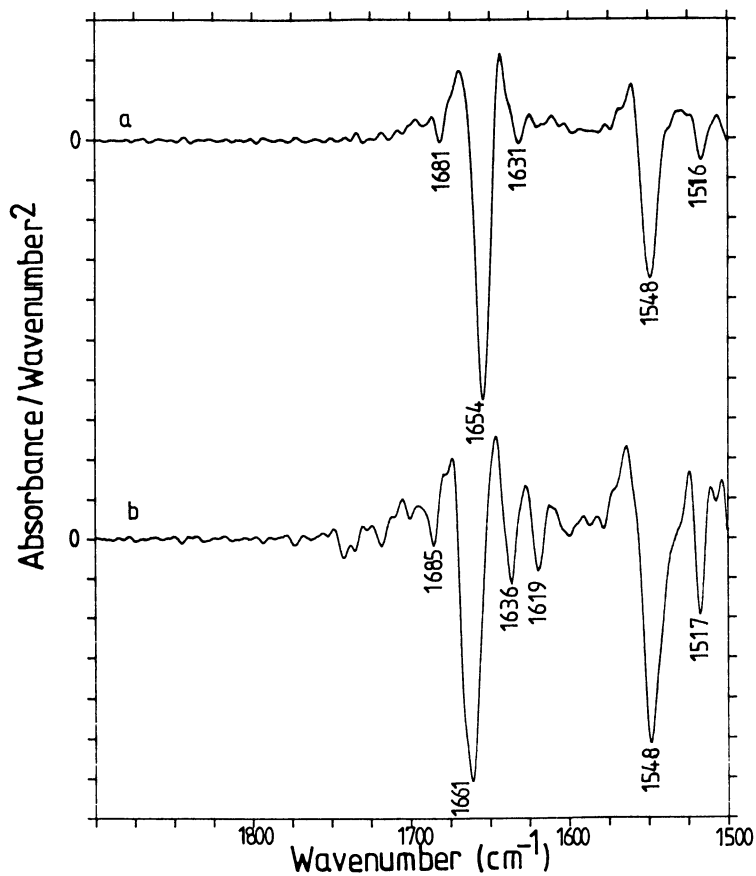


Figure 3. Comparison of the FTIR second-derivative spectra of myoglobin (top spectrum) and bacteriorhodopsin (lower spectrum). Spectra were recorded in H<sub>2</sub>O buffer.

similarity in the spectrum of bacteriorhodopsin and alamethicin led Haris and Chapman (50) to suggest that the  $1640\text{cm}^{-1}$  band in the spectra of these systems may arise from 3-10 helical structures. Others have suggested that the  $1640\text{cm}^{-1}$  band in bacteriorhodopsin may have strong contributions from C=C ethylenic stretch and C=N Schiff base modes of the retinal chromophore (56). However, bands near  $1640\text{cm}^{-1}$  has also been observed with other predominantly  $\alpha$ -helical proteins, that do not contain either  $\beta$ -sheet or a retinal chromophore, such as haemoglobin, myoglobin and alamethicin.

Like bacteriorhodopsin, rhodopsin is a retinal containing integral membrane protein. It belongs to a family of G protein-coupled receptors with seven transmembrane helices. The secondary structure and hydrogen-deuterium exchange characteristics of this protein have been well characterised (56,57). Polarised FTIR spectroscopy reveals a perpendicular orientation for the helices with respect to the plane of the membrane (56). FTIR spectroscopy has been used to compare (57) the structure bacteriorhodopsin with rhodopsin (see Figure 4). It can be seen that there are marked similarities and differences in the spectra of the two membrane proteins. Both proteins show an amide I maximum indicative of a predominantly  $\alpha$ -helical structure. Unlike bacteriorhodopsin the amide I maximum of rhodopsin is centred at  $1657\text{cm}^{-1}$ . In contrast to bacteriorhodopsin, rhodopsin appears to have a more "standard" helical structure like most of the other membrane proteins (see Table II).

**Absorption of  $\beta$ -Sheet Structures.** In general, peptides adopting  $\beta$ -sheet structure display Amide I bands in the spectral range  $1620\text{cm}^{-1}$  and  $1640\text{cm}^{-1}$  (2-4,10-12). Variation in hydrogen-bonding strength as well as differences in transition dipole coupling in different  $\beta$ -strands can strongly influence the position of this band (see Chapter 2 by Krimm). The assignment of bands in the region  $1620\text{-}1640\text{cm}^{-1}$  to  $\beta$ -sheet structure is consistent with both theoretical calculations and experimental studies on a large number of peptides and proteins. However, bands in this region can sometimes be observed for proteins which are known to contain very little or no  $\beta$ -sheet structure as discussed already for alamethicin, bacteriorhodopsin and myoglobin. Band assignment in this region is further complicated by the possibility of  $\beta$ -turn structure also absorbing in this region (58,59). Anti-parallel  $\beta$ -sheet structure can be identified by the presence of another band near  $1670\text{-}1695\text{cm}^{-1}$ . This component is normally weak and its precise assignment is often made difficult by the overlap of absorption from  $\beta$ -turn and unordered structure. A yet further complication arises from the fact that aggregated peptides adopting inter-molecular  $\beta$ -sheet structures can also absorb in this region (see later).

Most membrane associated peptides and proteins, that we and others have studied, adopt a predominantly  $\alpha$ -helical conformation (see Table II). There are some notable exceptions such as the porins (54) and tachyplesin I. The former is example of a large membrane protein and the latter is a small peptide. Tachyplesin I is a cyclic antimicrobial peptide consisting of 17 amino acid residues (60). This peptide can be isolated from the acid extracts of hemocytes of a horseshoe crab (*Tachyplesus tridentatus*). 2D NMR studies have shown that the peptide forms a rigid anti-parallel  $\beta$ -sheet structure, because of the two S-S linkages. Permeabilization of bacterial cell membranes is believed to be a major factor responsible for its bactericidal action. The amphiphilicity of the peptide is thought to play an important role in the binding of the

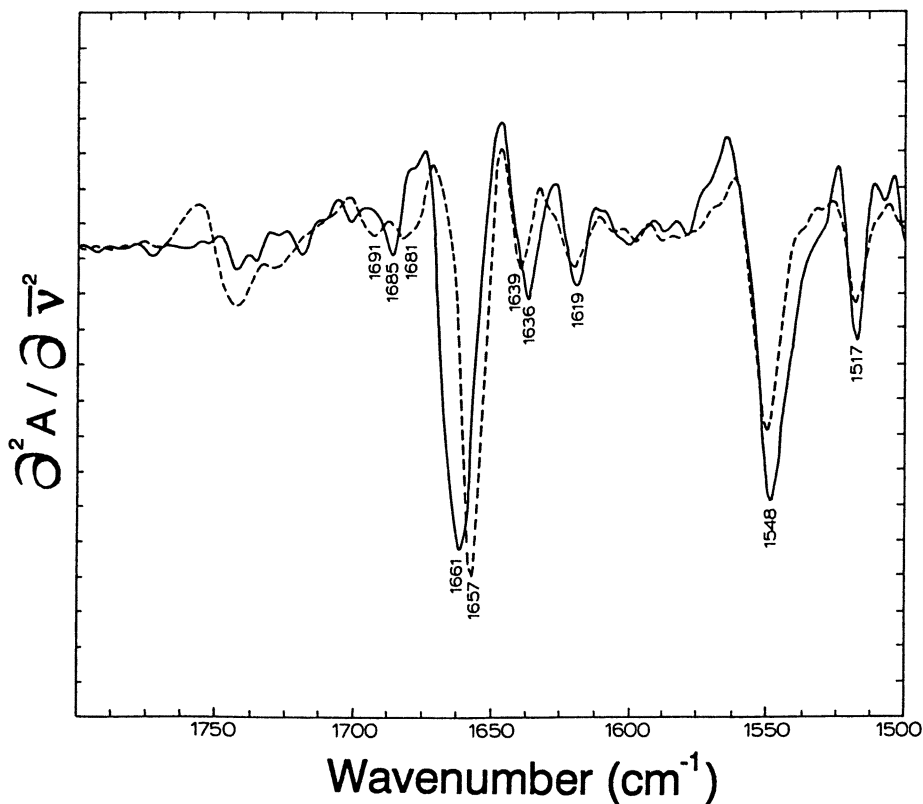


Figure 4. Comparison of the second-derivative FTIR spectra of bacteriorhodopsin (continuous line) and rhodopsin (broken line). The spectra were recorded for membranes suspended in H<sub>2</sub>O.

peptide to the bacterial cell membrane. ATR-FTIR spectroscopy has been employed to investigate the molecular aspects of the peptide-lipid interactions of tachyplesin (61). Strong amide I bands in the range 1620-1630 $\text{cm}^{-1}$  and the weak bands around 1685 $\text{cm}^{-1}$  were assigned to anti-parallel  $\beta$ -sheet structures. For a linear version of the peptide the main amide I band was centred at 1620 $\text{cm}^{-1}$  whereas for the cyclic peptide it was observed near 1630 $\text{cm}^{-1}$ . The lower frequency for the linear peptide was suggested to indicate a stronger interchain hydrogen bonding within its  $\beta$ -sheet structure. The polarised ATR-FTIR experiments revealed that the  $\beta$ -sheet structure is oriented parallel to the membrane surface (61). Such an arrangement was suggested to be highly likely due to the amphiphilic character of the peptides.

**Absorption of  $\beta$ -Turn and Random Coil Structures.**  $\beta$ -turns have been identified to have important role in the structure and function of many proteins. Table I shows the region where  $\beta$ -turn structures absorb and how this is effected by overlap from other structures. FTIR spectroscopy can be used to identify bands associated with this structure (58,59). The results of experimental studies are often in good agreement with the force-field calculations for such structures. Bands in the region 1620-1640 $\text{cm}^{-1}$  and 1650-1695 $\text{cm}^{-1}$  have been attributed to  $\beta$ -turn structures by many authors (2-4,10-12,58,59). Both of these regions also happen to be the area where  $\beta$ -sheet structures also absorb. For example, Hollosi and co-workers investigated eight cyclic penta- and hexapeptides with hydrogen-bonded  $\beta$ - and  $\gamma$ -turn structures using FTIR spectroscopy (58,59). Measurements were carried out in trifluoroethanol and dimethylsulphoxide solvents. Bands located at 1650 $\text{cm}^{-1}$ , 1642 $\text{cm}^{-1}$ , 1629  $\text{cm}^{-1}$  and 1625-15 $\text{cm}^{-1}$  were assigned to turn structures. It is noteworthy, that each of these bands occur at regions that can be associated with other secondary structure. For example, the 1650 $\text{cm}^{-1}$  and 1629-1625 $\text{cm}^{-1}$  bands can be assigned to helical and  $\beta$ -sheet structures, respectively. This clearly shows the difficulty associated with assigning turn structures in the spectra of large proteins which often contain a very small proportion of turn structures.

Attempts are being made to distinguish between  $\beta$ -sheet and  $\beta$ -turn structures. For example, Arrondo and co-workers (62) used FTIR spectroscopy to characterise the conformation of a 16 amino acid peptide. According to NMR results the peptide adopts a  $\beta$ -hairpin structure. Arrondo and co-workers found two features that distinguish the infrared spectrum of this  $\beta$ -hairpin peptide as compared to the spectrum of an anti-parallel  $\beta$ -sheet peptide. Firstly, the low frequency amide I component is observed at 1620 $\text{cm}^{-1}$  instead of  $\sim$ 1632 $\text{cm}^{-1}$  as seen with anti-parallel  $\beta$ -sheet structures. The high frequency band that shifts, due to hydrogen-deuterium exchange, from 1690 to 1675 $\text{cm}^{-1}$  in  $\beta$ -sheet proteins do not occur for the  $\beta$ -hairpin peptide (62). These studies are amongst a host of studies that are aimed towards distinguishing  $\beta$ -sheet and  $\beta$ -turn structures. Unfortunately, this distinction is unlikely to be simple due to the narrow frequency dispersion of the amide I band.

Table I shows the regions where random coil structures absorb in the infrared spectra of proteins. Generally, it is suggested that random coil structures absorb in the region 1650-1655 $\text{cm}^{-1}$  in  $\text{H}_2\text{O}$  but shift to 1643 $\text{cm}^{-1}$  in  $^2\text{H}_2\text{O}$  (2-4,10-12). However, there are reports of random coil structure absorbing at frequencies of 1660-1670 $\text{cm}^{-1}$  (2-4,10-12). It is highly likely that the frequency of the random coil structure will

vary depending on the particular arrangement of the peptide chains and its surrounding environment. For example, if the structure is highly solvated it may absorb at low frequency as compared to when it is buried within the protein core shielded away from solvent.

### The Problem of Peptide Aggregation

Several studies have shown the appearance of two characteristic bands that accompany temperature-induced protein denaturation. This is for example shown in Figure 5 for the human transferrin receptor protein. The deconvolved spectrum of this protein shows the main amide I components near  $1655\text{cm}^{-1}$  and  $1640\text{cm}^{-1}$ . The amide II band is centred at  $1549\text{cm}^{-1}$ . As can be seen from Figure 5a the gradual increase in temperature results in the appearance of the characteristic bands at  $1617\text{cm}^{-1}$  and  $1683\text{cm}^{-1}$ . Figure 5b shows a plot of the changes in the intensities of the amide I and amide II components as a function of temperature. A decrease in the intensity of the bands at  $1640\text{cm}^{-1}$ ,  $1655\text{cm}^{-1}$  and  $1549\text{cm}^{-1}$  is associated with an increase in temperature. Simultaneously, we see an increase in the intensity of the bands at  $1617\text{cm}^{-1}$ , and  $1683\text{cm}^{-1}$ . From the plots presented in Figure 5b, it is possible to both estimate the general denaturation temperature of the protein and also to obtain the precise unfolding temperature of the different secondary structural classes. With few exceptions, the denaturation of most proteins co-incides with the appearance of two bands, one near  $1610\text{-}1630\text{cm}^{-1}$  and the other at  $1680\text{-}1690\text{cm}^{-1}$ . These bands differ from the other bands, in the infrared spectra of proteins, by their extremely narrow band width. As such they are extremely useful as markers for monitoring protein denaturation. The major problem with these bands is that they occur in a spectral region that is normally associated with the absorption of  $\beta$ -sheet structure. The bands at  $1610\text{-}1630\text{cm}^{-1}$  and  $1680\text{-}1690\text{cm}^{-1}$  observed in the spectra of thermally denatured proteins are also associated with  $\beta$ -sheet but are a result of aggregation of polypeptide chains after protein unfolding. These bands have been explained as being due to polypeptide chains hydrogen-bonded in an inter-molecular  $\beta$ -sheet arrangement. The simultaneous appearance of the band near  $1680\text{-}1690\text{cm}^{-1}$  has been used to suggest that the  $\beta$ -sheets are of an anti-parallel nature. Unfortunately, these aggregation bands are also observed in the spectra of some synthetic peptides. This has led to numerous studies where conflicting results have been presented regarding the conformation. This has been particularly so with extremely hydrophobic peptides. It is difficult to ascertain if a band appearing in the  $1615\text{-}1630\text{cm}^{-1}$  region arises from aggregated structure or is due to normal intra-molecular  $\beta$ -sheet structure. In a number of studies we have encountered this problem. This is discussed below, firstly with the M13 coat protein.

**FTIR Spectra of Viral Coat Proteins.** The conformation of the M13 coat protein can change depending on the method used for its purification and reconstitution into phospholipid membranes (63). The coat protein of the filamentous bacteriophage M13 is a peptide composed of 45-50 amino acids. During infection of bacterial cells, the major coat protein is stripped off from the DNA and inserted into the cytoplasmic membrane of the host. Newly synthesised coat proteins are inserted into the bacterial cytoplasmic membrane prior to their assembly into viral particles. Several spectroscopic

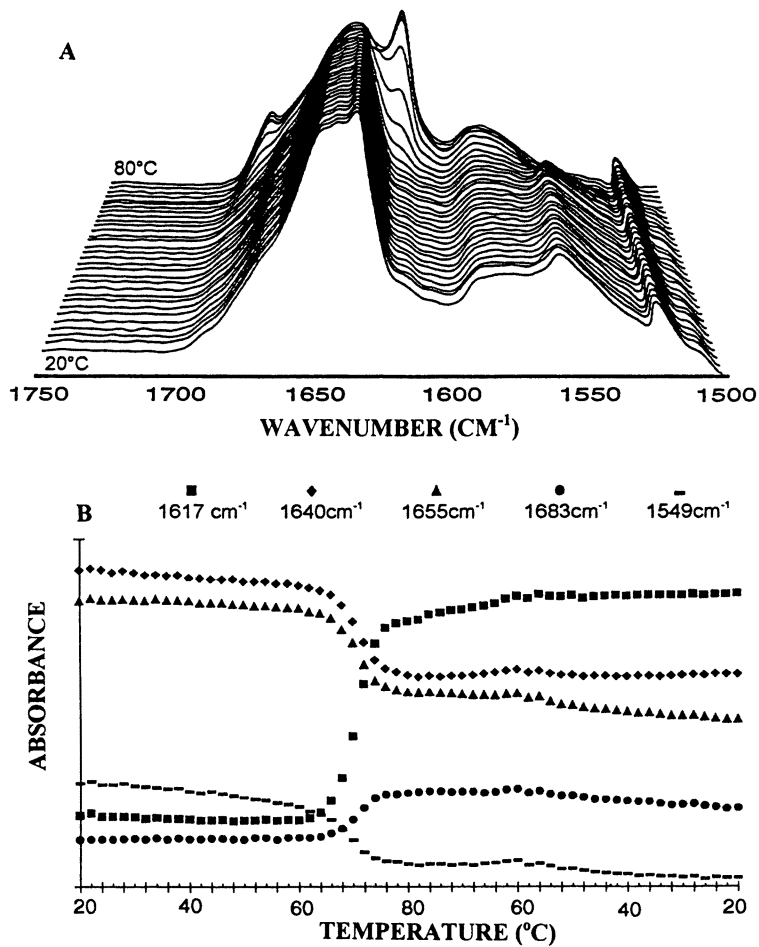


Figure 5. (A) Temperature induced changes in the deconvolved FTIR spectrum of human transferrin receptor in <sup>2</sup>H<sub>2</sub>O buffer at pH 7.4. (B) A plot of the changes in the intensity of selected amide I bands, in the FTIR spectrum presented in (A), as a function of temperature. Thermal denaturation of proteins is associated with the appearance of two new bands near 1617cm<sup>-1</sup> and 1683cm<sup>-1</sup>.

techniques have been applied to investigate the conformational changes associated with the insertion of the coat protein into a lipid bilayer. High resolution NMR spectroscopy was applied to determine the structure of the M13 coat protein solubilized in sodium dodecyl sulphate micelles. This study revealed that micellar-bound M13 coat protein consists of two  $\alpha$ -helical domains linked by a short region of uncertain conformation (64). Studies on the secondary structure of the M13 coat protein in phospholipid bilayers have been carried out using FTIR, CD and Raman spectroscopy (63). From the latter study, we found that when the M13 coat protein is extracted using phenol, it adopts a  $\beta$ -sheet conformation in phospholipid membranes. However, if the extraction procedure is carried out using cholate, the protein adopts a predominantly helical structure. Figure 6 shows the absorbance spectra of the  $\alpha$ -helical and  $\beta$ -sheet forms of the M13 coat protein recorded for samples reconstituted in dioleoylphosphatidylglycerol. The amide I maximum for the  $\beta$ -sheet and  $\alpha$ -helical forms of the protein are centred at  $1628\text{-}1630\text{cm}^{-1}$  and  $1654\text{-}1657\text{cm}^{-1}$ , respectively. The anti-parallel nature of the  $\beta$ -sheet structure is evident from the presence of a band near  $1694\text{cm}^{-1}$ . We concluded that the  $\beta$ -sheet form of the protein represents irreversibly aggregated M13 coat proteins. As such it is an artificial state that has no relevance to the biological activity. The helical form of the protein is assumed to be the biologically active form. This study was important in demonstrating how sample preparation can have a dramatic effect on the conformation adopted by a peptide. Such problems are more commonly encountered with synthetic peptides. Unlike, native proteins, synthetic peptides are exposed, during synthesis and purification, to a whole series of organic solvents and scavenger molecules which can impart significant influence on the conformation of the peptide. The effect of some of these molecules may remain even after their removal. TFA, as already mentioned, often remains strongly bound to peptides and are difficult to remove. Hence conformational characterisation, and even functional characterisation, of synthetic peptides needs to be carried out with great caution. We have been using synthetic peptides to study the conformation of a range of potassium channels. These studies have shown how peptide aggregation can be a serious problem in the interpretation of bands in the spectra of peptides. Here we present a summary of our studies.

**FTIR Analysis of Ion-Channel Peptides.** Voltage-gated potassium channels are large transmembrane proteins that conduct potassium ions across biomembranes and are responsible for electrical excitability (*for a review see ref. 65*). They are thought to be constructed from four similar domains with each of the domains consisting of six membrane spanning segments, S1 to S6. The relatively small size of voltage-gated  $\text{K}^+$ -channels makes them particularly attractive for structure-function analysis (65). Recent studies employing site-directed mutagenesis in conjunction with electrophysiological measurements have identified the functional role of several regions of this protein (*for a review see ref. 65*). For example, a linker region (designated as SS1-SS2 or H5) connecting the S5 and S6 segment is thought to constitute the pore-forming sequence. The structures of voltage-gated potassium channels are not known as yet. This is principally due to the lack of sufficient quantities of these proteins from natural sources. An attractive approach towards the structural characterisation of ion-channels is to simplify these large complex proteins by conceptually dividing them into various

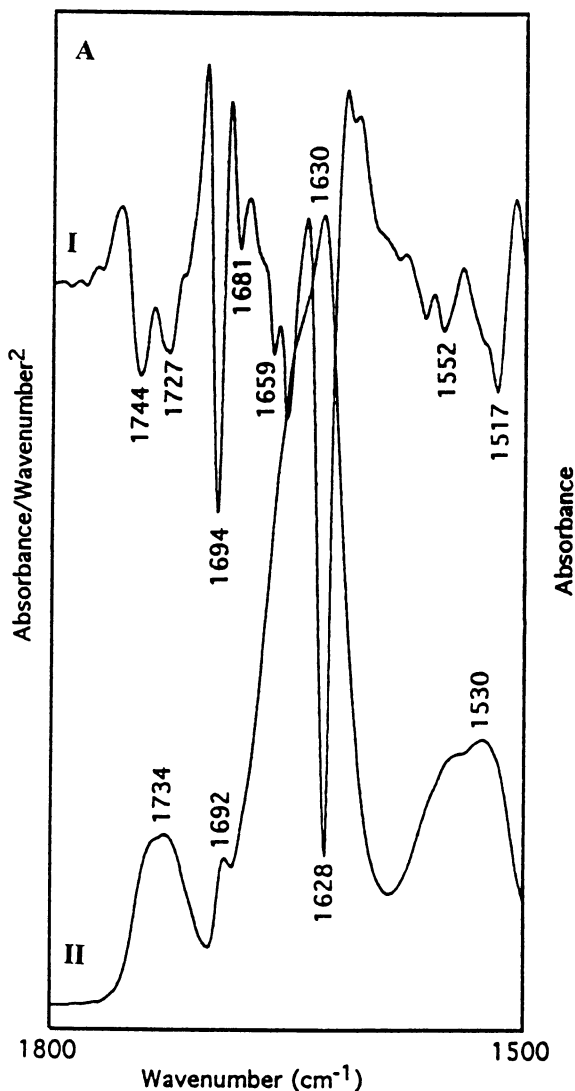
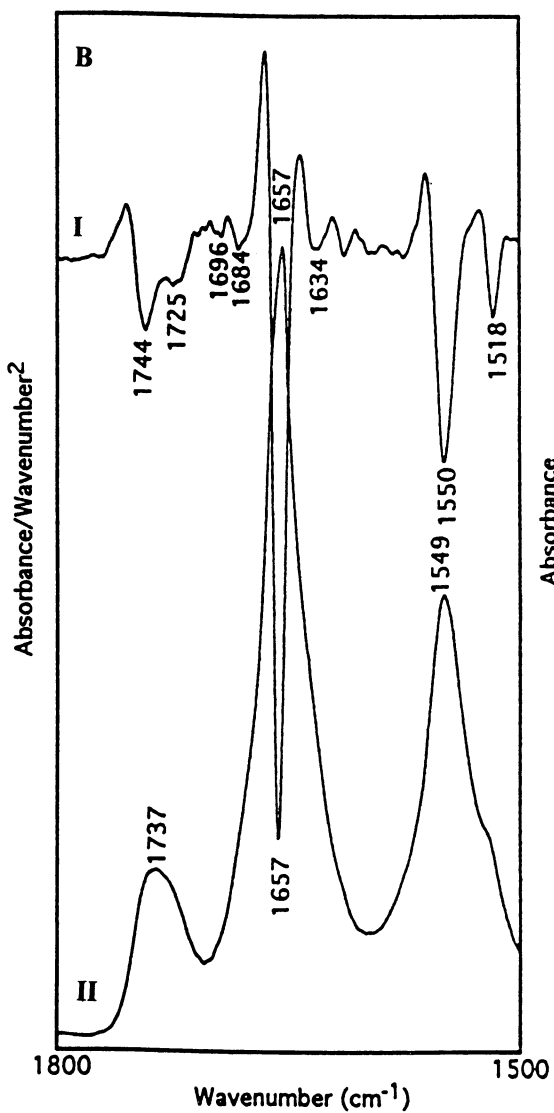


Figure 6. Absorbance (I) and second-derivative (II) FTIR spectra of the M13 coat protein reconstituted into phospholipid membranes composed of dioleoylphosphatidylglycerol. (A) M13 coat protein in the aggregated  $\beta$ -sheet form. (B) M13 Coat protein in the disaggregated  $\alpha$ -helical form. Both spectra were recorded for samples reconstituted into phospholipid membranes and suspended in H<sub>2</sub>O buffer.



Figure 6. *Continued.*

domains/motifs, identified from the primary structure. These domains can then be chemically synthesised, making possible the analysis of their structure and function. The use of peptide domains as models to study specific ion channels and pore-forming proteins is becoming increasingly popular (44,45,66,67). This approach is employed in the study of a range of potassium channels including voltage-gated potassium channels.

The sequence corresponding to the ion-selective pore region (H5) of the potassium channel from *Drosophila* was chemically synthesised and spectroscopic measurements carried out to investigate its structure (68). An extended form of the H5 peptide that includes residues on its C- and N- termini was also synthesised. This was done to investigate if the presence of additional residues influences the conformation of the H5 domain. Furthermore, a tetrameric version of H5 was synthesised (68) as it is suggested that in the native protein, the pore is lined by four H5 chains, one from each subunit. Conformational analysis on the synthetic peptides were carried out in different environments such as in lysophosphatidylcholine micelles and in phospholipid bilayers composed of dimyristoylphosphatidyl choline and dimyristoylphosphatidyl glycerol. FTIR and CD spectroscopy were used for the conformational analysis (68). Ion conductivity properties of the peptides were assessed using the planar bilayer technique. This revealed that when the H5 peptide is incorporated into planar lipid bilayers it induces ion channel activity (68). Figure 7 shows the absorbance spectrum of the H5 peptide in phospholipid membranes. From this spectrum it can be seen that the peptide appears to consist of a mixture of  $\alpha$ -helical ( $1650\text{cm}^{-1}$ ) and  $\beta$ -sheet structure ( $1623\text{cm}^{-1}$ ). However, it rather surprising that a small 19 amino acid peptide consists of roughly equal proportions of helical and  $\beta$ -sheet structures. Therefore, further experiments were carried out to ascertain if the  $1623\text{cm}^{-1}$  was really due to  $\beta$ -sheet structures or was due to some other structures, such as inter-molecular  $\beta$ -sheet structure arising from peptide aggregates. From our past experience with peptides, and denatured proteins, we knew that aggregated inter-molecular  $\beta$ -sheet structures can give rise to absorbance in the region  $1615\text{-}1630\text{cm}^{-1}$ . After carrying out an extensive series of experiments under different conditions we found the band at  $1623\text{cm}^{-1}$  in the spectrum of the H5 was indeed a reflection of peptide aggregation. We had to develop procedures for disaggregating the peptides prior to their reconstitution into phospholipid membranes. Under these conditions, the strong absorbance in the region of  $1623\text{cm}^{-1}$  is absent as can be seen from Figure 8. The problem of peptide aggregation was also encountered with all the transmembrane domains of the *Shaker* channel with the exception of the S4 peptide (see later). This can be explained by the fact that the S4 peptide is water soluble, whereas the other peptides were insoluble in an aqueous media. Besides providing valuable information regarding the secondary structure of domain in potassium channels, this study also highlighted the necessity for carrying out diverse experiments under different conditions before committing to specific interpretations of FTIR spectra.

Similar problems to those discussed above were encountered in the conformational analysis of synthetic peptides corresponding to membrane spanning domains of two other potassium channels namely the ROMK (69) and minK (70). Both FTIR and CD spectroscopy were used to investigate the secondary structures of synthetic peptides corresponding to the P and M1 and M2 domains of ROMK1 in aqueous solution, organic solvent and in phospholipid membranes. Like the *Shaker* channel peptides, the ROMK peptides also show aggregated  $\beta$ -sheet structure under

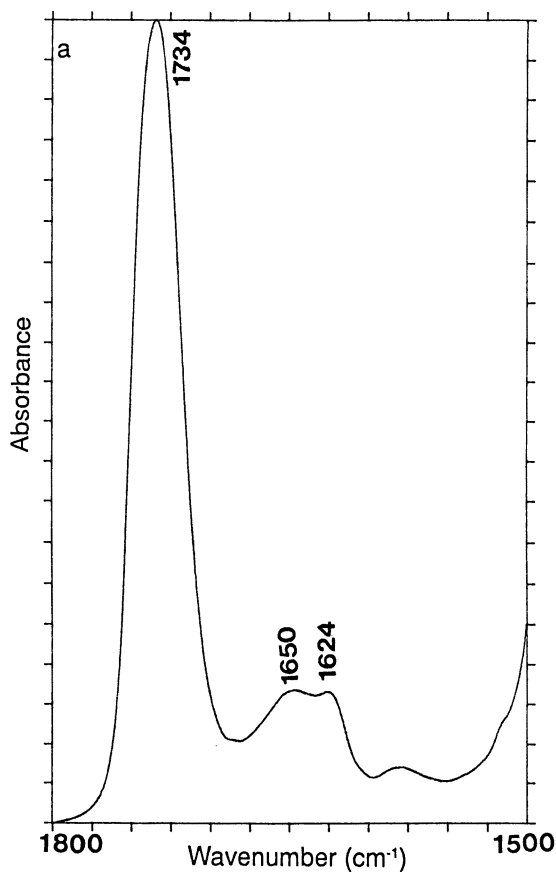


Figure 7. FTIR absorbance spectrum of the H5 peptide, corresponding to the ion-selective pore of voltage-gated potassium channel. The peptide, present in an aggregated state, was reconstituted in phospholipid membranes composed of dimyristoyl phosphatidyl choline. The spectrum was recorded for samples in <sup>2</sup>H<sub>2</sub>O buffer.

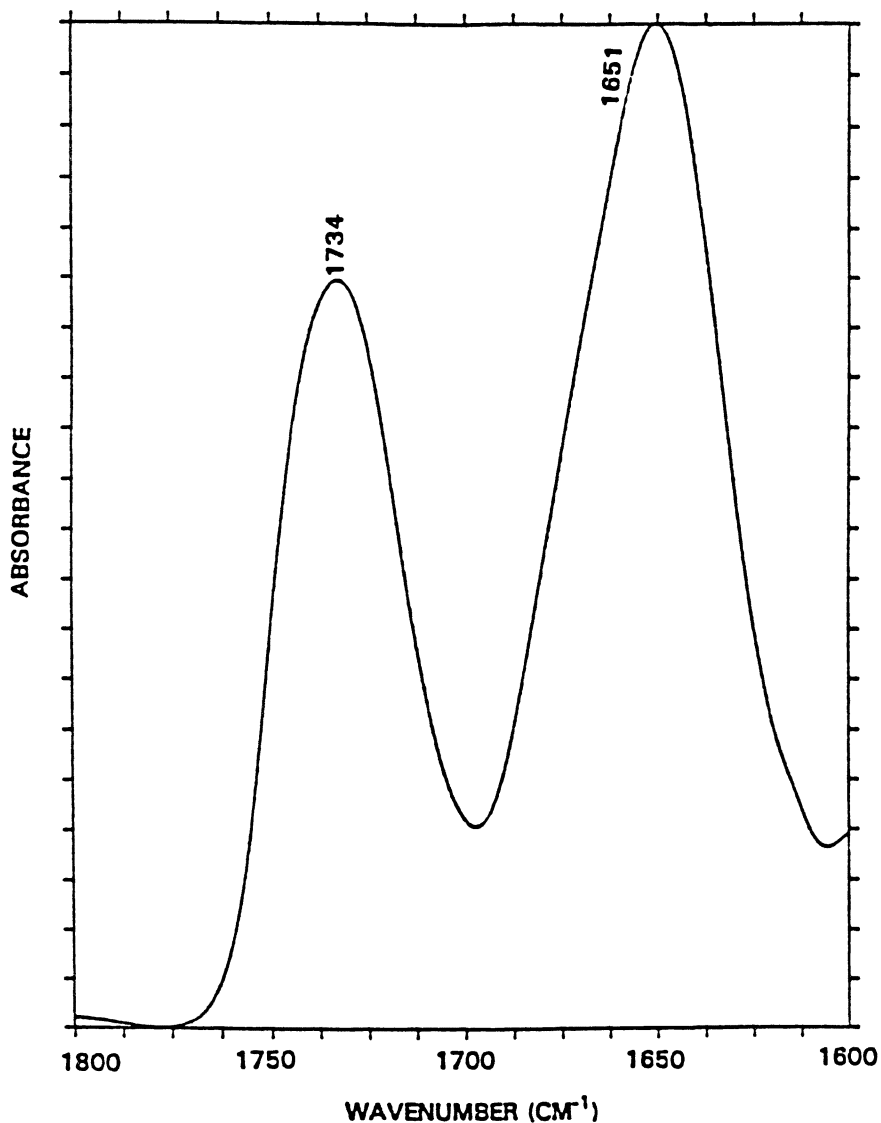


Figure 8. FTIR absorbance spectrum of the H5 peptide, corresponding to the ion-selective pore of voltage-gated potassium channel. The peptide was disaggregated prior to its reconstitution in phospholipid membranes composed of dimyristoyl phosphatidyl choline. The spectrum was recorded for samples in <sup>2</sup>H<sub>2</sub>O buffer.

certain conditions but are predominantly helical once disaggregated (69). FTIR spectral changes associated with the FTIR spectra of the M2 peptide is presented in Figure 9. In  $^2\text{H}_2\text{O}$  it displays an amide I maximum at  $1627\text{cm}^{-1}$ . There is also an additional component at  $1694\text{cm}^{-1}$ . Presence of these two bands suggests that the peptide is likely to be arranged in an anti-parallel  $\beta$ -sheet structure. However, the M2 peptide was insoluble in  $^2\text{H}_2\text{O}$  and the spectrum presented in Figure 9 was recorded for an aqueous suspension of the peptide. This raised the possibility that the  $1627\text{cm}^{-1}$  and  $1694\text{cm}^{-1}$  bands are due to peptide aggregation and reflects inter-molecular  $\beta$ -sheet structure. This was indeed the case since the disaggregated peptide adopts a helical structure in trifluoroethanol (see Figure 9). In this media the amide I maximum is centred at  $1654\text{cm}^{-1}$ . In the disaggregated state, the peptide also adopts a helical conformation in phospholipid membranes (69). The problem associated with peptide aggregation, and subsequent problems with spectroscopic interpretation, were found to be extremely complicated as was highlighted by our studies with the M1 peptide (69). For this peptide it was only possible to effectively eliminate the interference from peptide aggregates by removing them using a sucrose gradient procedure.

It is interesting to note that the aggregation bands we observe in the spectra of hydrophobic peptides, corresponding to domains of membrane proteins, are also seen in the infrared spectra of amyloid and prion protein aggregates. It appears that the structure of the peptide aggregates is virtually identical for denatured proteins, amyloid/prion fibrils as well as for hydrophobic membrane peptides. However, the position of the band can vary ranging from  $1610$ - $1630\text{cm}^{-1}$ . Under certain situations FTIR spectroscopy can be a very useful tool for detecting protein aggregates in diverse samples such as in amyloid and prion fibrils.

### Side Chains and Chromophores - Structure and Overlap with Amide Bands

If the amino acid side chains and chromophore groups of a protein contain infrared-active functional groups then these will be manifested as bands in an infrared spectrum. For example, the absorption bands associated with the retinal chromophore of bacteriorhodopsin and rhodopsin has been extensively studied using FTIR spectroscopy. Often the absorption of the chromophore bands is weak and they can also overlap with other bands such as the amide bands. In order to overcome such problems the retinal group has been isotopically labelled so that its absorption bands can be distinguished from other groups within the protein. As discussed earlier, a band near  $1640\text{cm}^{-1}$  in the spectrum of bacteriorhodopsin has been attributed to retinal absorption rather than to the protein backbone. Difference spectroscopy has been particularly useful for monitoring structural changes in chromophore groups. For example, Kandori et al. (71) have used time-resolved FTIR spectroscopy to investigate the last step in the photocycle of bacteriorhodopsin (BR) (the red shifted O state). They used mutants of bacteriorhodopsin for these studies. Figure 10 shows the FTIR difference spectra of O minus BR spectra of E204Q and E204D. Difference spectra of both of these mutants display negative bands at  $1695$ ,  $1672$ , and  $1642\text{cm}^{-1}$ . Positive bands are seen at  $1684\text{cm}^{-1}$ ,  $1663\text{cm}^{-1}$ ,  $1655\text{cm}^{-1}$ ,  $1628\text{cm}^{-1}$ , and  $1618\text{cm}^{-1}$  (Figure 10A) with slight variations in intensities (Figure 10B). The negative band at  $1642\text{cm}^{-1}$  and the positive band at  $1628\text{cm}^{-1}$  were assigned to the C=N stretches of the protonated *Schiff* base of BR and the O intermediate, respectively.

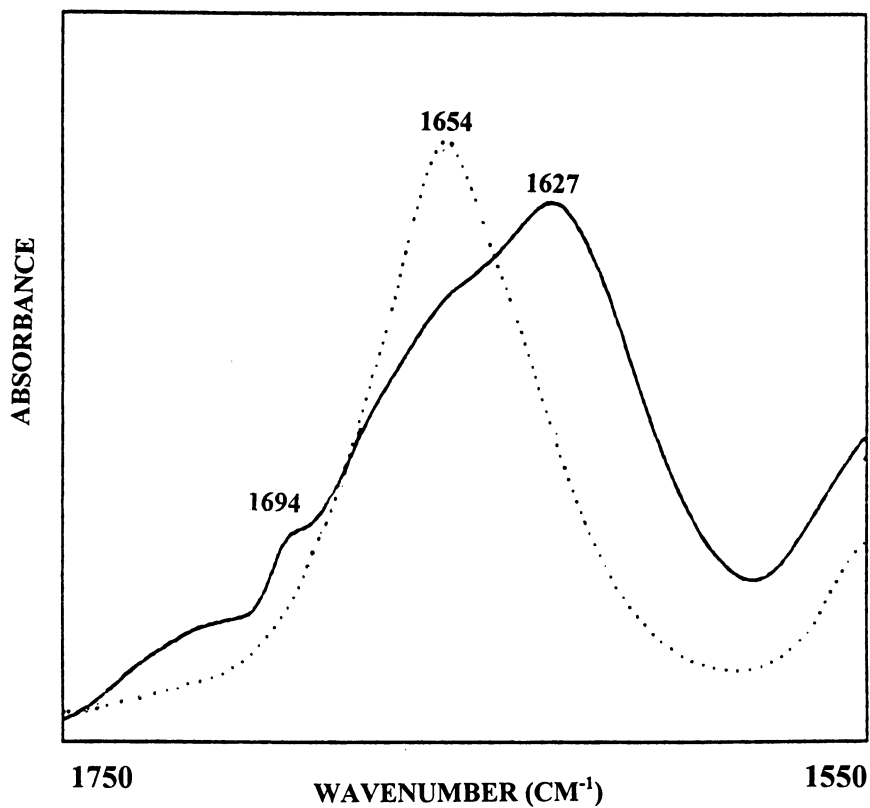


Figure 9. FTIR absorbance spectra for the ROMK M2 peptide recorded as a suspension in <sup>2</sup>H<sub>2</sub>O (continuous line) and as a solution in 95% TFE/5% <sup>2</sup>H<sub>2</sub>O (dotted line).

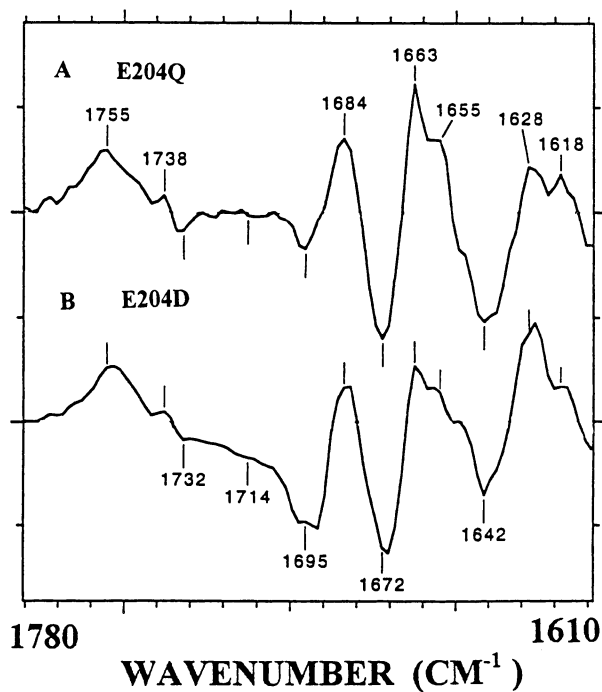


Figure 10. FTIR difference spectra of bacteriorhodopsin (BR). O minus BR difference spectra for E204Q and E204D mutants of bacteriorhodopsin. Adapted from ref. 71 with the permission of the authors.

It is noteworthy that of these bands occur in a region normally assigned to protein secondary structure. Other bands seen in Figure 10 were suggested to be due to the carbonyl vibration of the peptide groups.

Kandori and Maeda (72) have used FTIR difference spectroscopy to identify spectral changes associated with the chromophore and protein structure for rhodopsin in the region  $4000\text{-}1800\text{cm}^{-1}$ . In this region absorbance bands are observed from the O-H, N-H, and S-H stretching vibrational modes. For example, the isomerization of the retinal chromophore results in the shift of a band at  $3643\text{cm}^{-1}$  (*11-cis* form rhodopsin) to  $3487\text{cm}^{-1}$  (the *all-trans* form bathorhodopsin) or to  $3481\text{cm}^{-1}$  for the (9-*cis* form isorhodopsin). This band was suggested to arise from the N-H bond from an indole of the tryptophan residue (72). It is present in a hydrophobic environment around the beta-ionone ring and/or polyene chain of the retinal, and undergoes a change in its geometric alignment depending on the isomeric state (72). It is the only distinct band present in the high frequency spectral region between rhodopsin and isorhodopsin. This result was taken to suggest that specific interaction between the N-H group and the retinal chromophore contributes to the more efficient isomerisation in rhodopsin and isorhodopsin. The stretching vibrations of the water O-H, cysteine S-H, and the amide N-H of the peptide backbone shifts towards lower frequency upon formation of bathorhodopsin. This indicates that hydrogen bonding around the chromophore becomes stronger in bathorhodopsin. This suggests that at least part of the energy absorbed in the chromophore is already transferred to the protein in bathorhodopsin by strengthened hydrogen bonding (72).

Infrared spectroscopy has been successfully utilised to gain information on the structure and ionisation of the amino acid side chains (5-9). Most of the amino acid side chain absorptions are very weak in intensity and they can often strongly overlap with the amide bands. This can not only cause problems in the accurate assignment of the amide I band, but can also prevent information being derived regarding the amino side chains. However, through the use of site-directed mutagenesis, isotopic labelling and difference spectroscopy, it has been possible to information on the structure of different amino acid residues in a protein. This has been most successfully applied with bacteriorhodopsin and rhodopsin. Potential problems with overlap from amino acid side chains can be overcome if one takes into consideration the amino sequence of the protein when carrying out assignments in the amide I region. In order to reduce possible errors, some authors have recorded spectra of amino acids, which are subsequently subtracted from the spectrum of the protein. While such methods are useful, they can introduce possible errors. For example, the structure of the amino acid side chains when they are part of a protein may have different structures as compared to the situation when they are in the form of free amino acids or dipeptides. Amino acid side chains in a protein participate in diverse forms of interactions such as hydrogen-bonding and salt-bridge interactions.

Figure 10, shows a number of peaks that has been assigned to amino acid side chains of BR. Take the case of the weak band at  $1714\text{cm}^{-1}$  seen in the spectrum of E204D and which is absent in E204Q. This band has been assigned to the deprotonation of Asp-204 as the same negative band that has been observed in the M minus BR spectrum of E204D with a peak at  $1714\text{cm}^{-1}$ . The presence of the negative band at  $1714\text{cm}^{-1}$  for Asp-204 and the positive peak at  $1755\text{cm}^{-1}$  for Asp-85 in the O minus BR spectrum of E204D suggests that the proton transfer from Asp-85 to Asp-204 occurs in the final O-to-BR process in the photocycle. The negative band at  $1732\text{cm}^{-1}$  in E204Q



and E204D was attributed to the carbonyl stretching vibration of the Asp-115 because the carbonyl stretch is at  $1743\text{cm}^{-1}$  in the N minus BR spectrum. Since the frequencies of Asp-115 and Asp-96 are similar to those observed previously for the wild type (73), Kandori et al. (71) suggested that surroundings of these residues are not altered in the E204 mutants. It can be seen from Figure 10 that Asp-115 band frequency shifts towards higher frequency (from  $1732$  to  $1738\text{cm}^{-1}$ ) upon O formation. This reflects weakening of the hydrogen bonding of the carbonyl group of Asp-115 that has been previously observed for the M and N intermediates of the wild type (73). These types of studies show how infrared spectroscopy can be used to probe changes at the level of individual bonds in a large and complex protein such as bacteriorhodopsin.

### Hydrogen-Deuterium Exchange Analysis Using FTIR Spectroscopy

Along with NMR and Neutron Diffraction techniques, FTIR spectroscopy is one of the few biophysical techniques that can be used to monitor hydrogen-deuterium exchange in proteins. Hydrogen-deuterium exchange rates can be a powerful probe of protein structure providing information on solvent accessibility and hydrogen bond stability of amide protons. FTIR spectra of proteins in  $^2\text{H}_2\text{O}$  can be readily used for the estimation of slowly- and rapidly-exchanging amide protons (2-5,10-12). The amide A (35) and amide II bands (2-5,10-12,15) have been used for monitoring hydrogen-deuterium exchange. However, the amide II band ( $1550\text{cm}^{-1}$ ) is most widely used. Exposure of a peptide to  $^2\text{H}_2\text{O}$  results in a shift of this band by about  $100\text{cm}^{-1}$  to  $\sim 1450\text{cm}^{-1}$ . The amide II band is dominated by N-H bending vibration and isotopic substitution of the peptide bond N-H to N- $^2\text{H}$  results in this large shift. The amide I band, associated primarily with C=O stretching of the peptide group, is effected to a much lesser extent. The shift of this band varies depending on different factors such as the secondary structure of the protein. Normally, the amide I band in  $^2\text{H}_2\text{O}$  shifts towards lower frequency by  $5\text{-}15\text{cm}^{-1}$ . Even this small shift in frequency can be useful for identifying different secondary structures and the nature of their surrounding environment (15,56,57). Detailed changes in hydrogen-deuterium exchange can be monitored by following the shift in the position of the amide I components in the second-derivative and deconvolved spectra. Following the changes in the intensity of the amide II band as a function of time can reveal quantitative information on the number of amide protons that are exchanged (15,56,57). The particular advantage of FTIR spectroscopy for hydrogen-deuterium exchange analysis is that there are no limitations due to the size or the physical state of the protein, measurements can be performed with small sample quantities. It is one of the few techniques that can be readily used to investigate hydrogen-deuterium exchange of membrane-bound proteins (2-5,10-12). With time-resolved FTIR spectrometers it is possible to follow exchange processes on a millisecond time scale.

**Detection of Conformational Changes using Hydrogen-deuterium Exchange.** Measurement of hydrogen-deuterium exchange is often a more powerful probe for detection of conformational changes in proteins. We have found this to be the case in the study of a range of metalloproteins (33,38,74). These results indicate that the metal free proteins exchange to a greater extent than the corresponding metal-bound proteins. This difference can be attributed to changes in flexibility/mobility in the proteins since

little or no change in secondary structure content is detected. We have also detected conformational differences between the native and reactive-centre cleaved forms of serine proteinase inhibitors such as  $\alpha_1$ -antitrypsin (75). Differences were observed in the spectra indicative of greater hydrogen bonding occurring with the reactive-centre cleaved form of the protein. The hydrogen-deuterium exchange within the polypeptide backbone is also found to be less for the cleaved form. This difference in the hydrogen-deuterium exchange was further supported by NMR spectroscopy (75). The ability of FTIR spectroscopy to detect subtle changes in protein conformation makes it an extremely useful tool for comparing the structures of wild type proteins and their site-directed mutant forms.

**Hydrogen-Deuterium Exchange of Membrane Proteins.** FTIR spectroscopy is one of the few techniques that can be used for investigating hydrogen-deuterium exchange of membrane proteins. It has been used to study a large number of proteins including bacteriorhodopsin, rhodopsin, the ATPases, glucose transporter etc (2-4,10-12,56,57,76). The extent of hydrogen-deuterium exchange in membrane proteins can sometimes depend on the proportion of the protein buried in the membrane. Bacteriorhodopsin, for example, has a greater number of unexchanged amide protons than rhodopsin, i.e. 71% and 51%, respectively (57). Bacteriorhodopsin has approximately 80-90% of its total protein mass embedded in the membrane, the remainder being small N- and C-termini and linker regions which are most likely in random coil conformation. In contrast, rhodopsin is thought to have only 50% of its protein embedded in the membrane (57). Consequently, the higher hydrogen-deuterium exchange observed for rhodopsin is likely to be due to it containing a higher proportion of its protein mass exposed to the aqueous media outside the membrane. Hydrogen-deuterium exchange thus offers an approach for identifying the proportion of a protein embedded in the membrane. However, such conclusion needs to be carried out with caution, since it is always possible for polypeptide segments outside the membrane to exchange at a significantly reduced rate compared to transmembrane segments. For example, this can be the case for strongly hydrogen-bonded segments, buried in a globular domain, outside the membrane.

**Hydrogen-Deuterium Exchange and Assignment of Amide I Bands.** It is important to have an understanding of how hydrogen-deuterium exchange can influence the position and intensity of amide bands before interpreting FTIR spectra of proteins recorded in  $^2\text{H}_2\text{O}$ . This has been clearly demonstrated in a study performed with ribonuclease A and S (15). The spectra of these two proteins were virtually indistinguishable in  $\text{H}_2\text{O}$ . This suggested there are no differences in the secondary structure of these two proteins which have identical sequence and size and only differ by the presence of a peptide bond cleavage in ribonuclease S. However, in  $^2\text{H}_2\text{O}$  there were significant differences in the intensity and frequency of their amide I components. This was explained as not arising from differences in their secondary structure but due to differences in hydrogen-deuterium exchange which results in shifts in the frequency of the different amide I components (15). It was shown that at equivalent extent of hydrogen-deuterium exchange there was no spectral difference in the amide I region of the two proteins. This study was important in illustrating the necessity for taking into

consideration the differences in hydrogen-deuterium exchange in proteins before reaching conclusions regarding differences in their secondary structures. However, differences in the rates and extent of hydrogen-deuterium exchange can reveal differences in the structure of proteins. So for example, in the case of ribonuclease A and S, it was demonstrated that the proteins have identical secondary structures but have different tertiary structures as revealed by their differences in hydrogen-deuterium exchange. Ribonuclease S undergoes complete hydrogen-deuterium exchange after exposure to  $^2\text{H}_2\text{O}$  for few days. In contrast, it is necessary to heat ribonuclease A to  $62^\circ\text{C}$  for few minutes in order to complete its hydrogen-deuterium exchange (15). Figure 11 displays the FTIR spectra of ribonuclease A after partial and complete hydrogen-deuterium exchange. From the comparison of the two spectra it is clear that variations in the frequency and intensity of the amide components occur as a result of differences in hydrogen-deuterium exchange.

### **Influence of Environment on Protein Infrared Spectra**

During the functional cycle of a peptide it may undergo diverse changes in its environment. Some peptides are present in aqueous solution and yet can interact with phospholipid membranes or membrane-bound receptors. The peptide is likely to encounter a different environment in each of these cases. Unfortunately, there is a lack of techniques that can be used to monitor changes in peptide structure as a function of environment. FTIR spectroscopy is one of the few exceptions as it can be used for conformational analysis in  $\text{H}_2\text{O}$ , crystals, thin films, organic solvents, detergent micelles as well as in a lipid bilayer matrix. FTIR spectroscopy therefore provides a means of assessing the effect of different environment on the secondary structure and side-chain structure of proteins. It can be used for identifying the most physiologically relevant conformation of a peptide.

How the surrounding environment influences the conformation of a peptide is illustrated here by discussing the results obtained for the S4 peptide, the Pfl coat protein and proteins in the crystalline state.

**Environment Induced Coil-Helix Transition.** The S4 transmembrane domain in voltage-gated ion-channels has been suggested to have a role in sensing the transmembrane electric potential (65). This segment contains a positively charged amino acid residue (lysine or arginine) at every third position. Mutagenesis studies have shown that these basic residues in the S4 region play a central role in the voltage-dependence of activation of potassium channels. Although these studies have provided valuable information on the possible functional role of the S4 sequence, little is known about its structure. We have used FTIR spectroscopy for conformational analysis of a synthetic peptide corresponding to the S4 sequence of the *Shaker Drosophila* potassium channel (77). Spectra of the 20 amino acid residue peptide were obtained in the organic solvent, aqueous solution as well as in phospholipid membranes. The peptide in  $^2\text{H}_2\text{O}$  solution shows its amide I band at  $1643\text{cm}^{-1}$ . This band frequency can be attributed to random coil structure (77). This is further supported by our CD study which shows a spectrum typical of unordered conformation. In trifluoroethanol and in phospholipid membranes the amide I maximum shifts towards higher frequency ( $1650\text{-}1654\text{cm}^{-1}$ ). The shift of the amide I maximum from  $1643\text{cm}^{-1}$  to  $1654\text{cm}^{-1}$  can be associated with a conformational

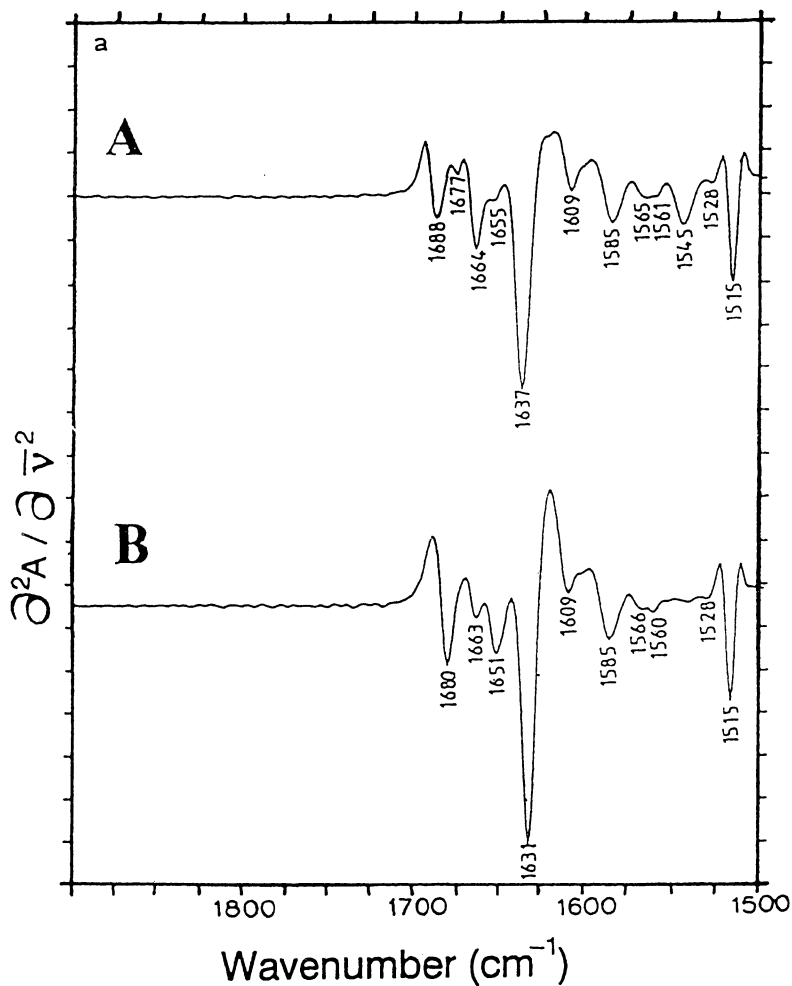


Figure 11. FTIR second-derivative spectra of ribonuclease A in  $^2\text{H}_2\text{O}$  buffer after (I) partial exchange and (II) complete exchange.

change from random coil ( $1643\text{cm}^{-1}$ ) to  $\alpha$ -helical ( $1654\text{cm}^{-1}$ ) structure. There are other channel peptides such as the ROMK M2 peptide (see Figure 9) which goes a transition from aggregated  $\beta$ -sheet structure to a helical conformation upon transfer from an aqueous environment to a hydrophobic media. The predominantly helical structure observed, by FTIR spectroscopy for the S4 peptide in membrane systems, is in good agreement with a previous 2D NMR study on a synthetic S4 peptide corresponding to the rat brain sodium channel (78). A model for the structure of S4 has been proposed in which as much 90% of the S4 segment is suggested to moves outside the lipid bilayer in the open state of the channel (79). Such a drastic change in environment is likely to lead to a significant change in the structure of the S4 peptide. Our FTIR spectroscopic results clearly show an environment dependent conformational switching for the S4 peptide. This information can be helpful for rationalising the structural and functional role of the S4 segment in voltage-gated ion channels.

**Environment Dependent Amide I frequency for a Viral Coat Protein.** We have already discussed earlier how the conformation of the M13 coat protein change depending on the method used for its reconstitution into phospholipid membranes. Here we discuss about the case of the Pfl coat protein and how its structure change under different conditions. FTIR spectroscopy (80,81) has been used to study the structure of the Pfl coat protein when present in the phage, in detergent micelles, and in a phospholipid membrane system. Figure 11 presents FTIR spectra of the Pfl coat protein in the phage and membrane systems. Suspensions of the Pfl phage in  $\text{H}_2\text{O}$  (Figure 12) and  $^2\text{H}_2\text{O}$  show an amide I band at  $1652\text{cm}^{-1}$ , indicative of a high content of  $\alpha$ -helical structure. When the Pfl coat protein is present in sodium dodecyl sulphate micelles and in lipid membrane (Figure 12) systems the amide I band is located at  $1657\text{-}1658\text{cm}^{-1}$ . This also indicates the presence of  $\alpha$ -helical structure but the amide I band in these media shifts to a higher frequency as compared to the protein in the phage. This shift is indicative of the presence of a weaker hydrogen-bonding  $\alpha$ -helix for the protein in detergent and membrane systems. This is probably due to the reduction in the hydrogen-bonding of water molecules to Pfl helix. We discussed earlier how the amide I maximum for the helical structure in calmodulin occurs at an unusually low frequency due to hydrogen-bonding of its polypeptide segments with water molecules. The frequency observed for the Pfl protein in phospholipid membranes is similar to what we observe with other membrane proteins (see Table II). This result clearly shows how the amide I band frequency for a protein can change upon its insertion into the low dielectric environment of a phospholipid membrane.

**Spectra of Proteins in the Crystalline and Solution States.** The most detailed structure of a protein can be elucidated using X-ray crystallography. However, this requires the protein to be in a crystalline state. Consequently, there has been ongoing debates regarding the validity of such structures considering that proteins in a physiological environment are in an non-crystalline state. In order to address this question, NMR spectroscopy has been used to obtain the solution structure of proteins whose structure has also been determined in the crystalline state by crystallography. The general conclusion from such studies is that the crystal structure does not significantly differ from the solution structure. Minor differences have, however, been shown to

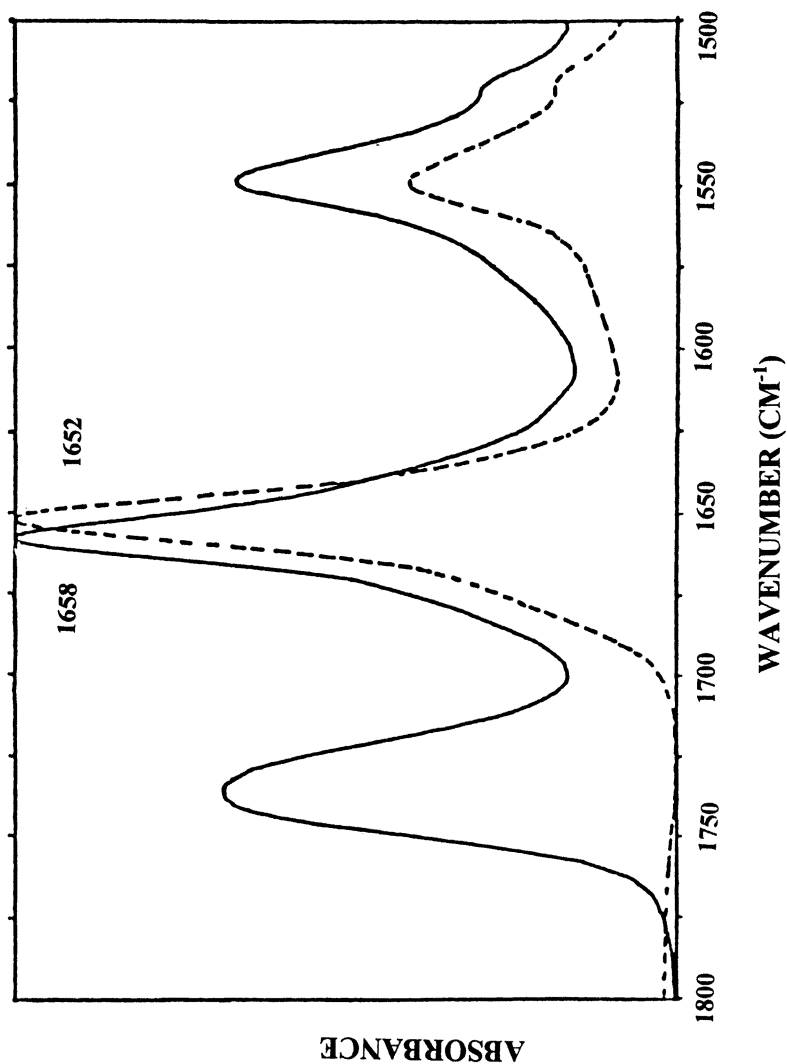


Figure 12. FTIR absorbance spectrum of the Pfl viral coat protein in the Phage form (broken line) and after reconstitution into phospholipid membranes (continuous line). The spectra were recorded for samples in H<sub>2</sub>O buffer.

originate from the differences in environments which occur in the crystal and in solution state (82). Only a limited number of proteins have been compared in this way. This number is unlikely to increase at a significant rate due to the complexity of protein structural determination using NMR spectroscopy. Indeed, some proteins are far too large to be studied by NMR. FTIR spectroscopy is an alternative method that can be used for comparing the structure of a protein in the solution and in the crystalline state. Whilst it cannot provide the complete 3D structures of proteins, it can be useful for monitoring similarities and differences in protein secondary structure in a rapid manner. Using a FTIR spectrometer connected to a microscope, infrared spectra of six proteins have been recorded both in aqueous solution and in the crystalline state (83). The structures of these proteins have been previously determined using X-ray crystallography. The crystals used were those prepared for X-ray diffraction analysis. The FTIR analysis revealed minor variation between the spectra of the proteins in the solution and crystal states. The minor differences are primarily confined in the 1670-1680 $\text{cm}^{-1}$  region.

The studies discussed above clearly show how FTIR spectroscopy can be used for comparing the structures of peptides and proteins in diverse environments including aqueous solution, phospholipid membranes and in the crystalline state. It also demonstrates how the amide I band position can be extremely sensitive to the surrounding environment.

### Complexity of Membrane Protein FTIR Spectra

Membrane proteins are often large and complex, and can have their polypeptide chains exposed to different environments. For example, there are membrane embedded segments in direct contact with phospholipids and extra-membranous segments that are exposed to the aqueous media as well as segments in the protein interior that may be in contact with cofactors or are in contact with extrinsic or intrinsic proteins. FTIR spectroscopy is not restricted by the large size or the membrane-bound nature of membrane proteins. However, difficulties are encountered when studying membrane proteins that consist of a large proportion of extramembranous domains. In such situations it is difficult to distinguish between infrared bands arising from the transmembrane region as compared to those arising from the extramembranous region. The proportion of extra-membranous protein mass can vary significantly from protein to protein. Thus for example, with bacteriorhodopsin only 20-30% of the protein is exposed to the aqueous media whereas in the case of prostoglandin H it is as much as 90%. In order to overcome such problems in the interpretation of infrared spectra different approaches are being used. One such method involves proteolytic digestion of extra-membranous regions. This approach has been used with several proteins which are discussed below.

**Comparison of Membrane Protein Amide I bands.** The X-ray structures of membrane proteins that have been solved using X-ray crystallography show either an all- $\alpha$ -helical or all- $\beta$ -sheet transmembrane structure. Table II gives examples of the different types of membrane proteins that have been studied by FTIR spectroscopy. It can be seen that the vast majority of these membrane proteins, from a wide variety of sources, show a

remarkable similarity in their main amide I maximum frequency consistent with an  $\alpha$ -helical conformation (41). The amide I band is very sensitive to hydrogen-bonding and to the local environment. Therefore, the similarity in the position of the amide I maximum suggests a strong homology in the nature of the helical structure and the surrounding environment in different membrane proteins. As already mentioned previously, the notable exceptions to this are bacteriorhodopsin and porin. FTIR spectra of some membrane proteins give an impression of a transmembrane structure that consists of both  $\alpha$ -helical and  $\beta$ -sheet structures. At least it is not possible to exclude the possibility that both of these structures may be embedded in the membrane. In order to overcome this problem and distinguish between the extramembranous and membranous regions of proteins it has become quite popular to use proteolytic enzymes to remove the extramembranous domains of membrane proteins. This is done on the basis that the spectrum of the resulting fragment should only contain contribution from structural elements that are transmembrane. This approach emerged after the controversy regarding the origin of the  $1640\text{cm}^{-1}$  band in bacteriorhodopsin. It was suggested that this band may arise from extramembranous regions of this protein. This led Chapman and co-workers (84) to proteolytically remove extramembranous regions of bacteriorhodopsin and characterise the structure of the membranous fragment. The result of their FTIR study showed that the  $1640\text{cm}^{-1}$  band was still present in the proteolytically digested fragment. On the basis of these results it was suggested that  $\beta$ -strand structure must be associated with the membrane-associated structure and is not exposed to proteolytic digestion (84).

**Proteolysis as a Tool for Simplifying Membrane Protein Spectra.** Recently it has been suggested that a third motif comprising of a mixture of  $\alpha$ - and  $\beta$ -structure may also be present in some membrane proteins. This arose after an electron microscopic study on acetylcholine receptor (nAChR) which suggested the possible existence of some transmembrane  $\beta$ -sheet structure (85). FTIR spectroscopy has provided conflicting results regarding the secondary structure of this protein. Gorne-Tschelnokow et al. (86) used FTIR spectroscopy to study nAChR rich membrane preparations which were treated with proteinase K to remove the extramembranous part of the receptor. The remainder of the protein accounts approximately for the four hydrophobic transmembrane stretches. In this way, the FTIR spectrum was simplified so that absorbance in the amide I can be specifically attributed to the transmembrane domain of the protein. It was found that the portion of the receptor not accessible to the proteinase contained about equal amounts of  $\alpha$ -helical and  $\beta$ -structure (86). Results of polarised infrared spectroscopy indicates the  $\alpha$ -helices to span the membrane with an uniaxial orientation. However, no such orientation was observed for the  $\beta$ -structure (86). It was suggested that the lack of uni-directional orientation corresponds to the 'diffuse'  $\beta$ -structures deduced from electron-microscopical investigations (85). In contrast, there are other studies that suggests an  $\alpha$ -helical structure for the transmembrane domain (87). Marked changes in the amide II band of nAChR was detected upon its exposure to  $^2\text{H}_2\text{O}$  (87). Approximately 30% of the amide protons were exchanged with seconds, 50% after 30 min and, 60% after 12 h and 75% after prolonged exposure of the nAChR for several days. On the basis of these results it was suggested that 30% of the amide protons that exchange within seconds are highly exposed to solvent and are likely to be involved in turn or random coil structures. The 25% of the hard-to-exchange amide protons are



inaccessible to solvent and are likely to be located within the transmembrane domains of nAChR. This study suggests that this protein has a predominantly  $\alpha$ -helical structure. They did not find any evidence for the existence of a large number of exchange-resistant  $\beta$ -strands (87).

### Polarised infrared spectroscopy

Information regarding molecular orientations can be determined from polarised infrared spectra of oriented samples (see Chapter 4 by Ruyschaert). This approach has been used to determine the orientation of the different secondary structures within the lipid bilayer matrix of biomembrane systems (10,18-20). Many membrane proteins including bacteriorhodopsin, rhodopsin, cytochrome C oxidase and photosynthetic reaction centres have been studied in this way. In most cases, the helices are found to be oriented perpendicular to the plane of the biomembrane. Polarised infrared difference spectroscopy has also been used to measure the average orientation of individual groups such as the chromophore during the photocycle of bacteriorhodopsin (5,7). Like all other aspects of peptide FTIR spectroscopy, there are potential pitfalls that one needs to take into consideration. The results of polarised spectroscopy can be highly influenced by the method used for reconstitution of the peptides. In many cases the biomembrane system may be composed of different populations - one with peptides adhered to the biomembrane surface, and other in which the peptide is fully inserted. Under such conditions the polarised infrared data can give a very misleading picture.

How sample preparation can significantly influence the results of polarised infrared spectroscopy of biomembranes have been demonstrated by Frey and Tamm (88). They used polarised ATR-IR spectroscopy to determine the orientation of melittin bound to two different types of model phospholipid membranes. They initially carried out measurements on melittin which was bound to fully hydrated single supported planar phospholipid bilayers (SPBs) on germanium substrates. These results were subsequently compared with measurements on melittin in multibilayers (MBLs) prepared by solvent evaporation. Results obtained revealed that in the case of hydrated SPBs, the  $\alpha$ -helical structure of melittin was preferentially oriented parallel to the plane of the of the supported membrane. However, in dry MBLs, the  $\alpha$ -helical structure was oriented parallel to the phospholipid hydrocarbon acyl chains. It was concluded, that the orientation of melittin in phospholipid membranes is strongly dependent on the degree of hydration of model membranes (88). This clearly demonstrates how important the surrounding environment, such as hydration, can be in modulating peptide interaction with membranes.

### Future prospects

There are several areas of peptide FTIR spectroscopy that is likely to be extremely popular in the future. Time-resolved FTIR spectroscopy is certainly one such area. The Chapter by Gerwert and co-workers (Chapter 8) discusses this approach for protein conformational studies. Other workers are also using this approach for monitoring protein folding. For example, Callender and co-workers (89) are using time-resolved FTIR spectroscopy to investigate the folding of proteins like ribonuclease and

apomyoglobin. The major problem hindering the interpretation of the results from such studies is the assignments of the amide I components to specific structural features. Another area that has seen an increase in popularity is isotope-edited FTIR spectroscopy (see Chapter 9 by Li et al.). This is already making possible significant progress in the assignment of bands in the infrared spectra of proteins. Using solid-phase peptide synthesis it is possible to isotopically label specific amino acid residues in proteins and characterise their conformation using FTIR spectroscopy. Furthermore, uniformly labelling of proteins has opened the possibility of studying protein-protein (90) and receptor-ligand interactions (ref. 91 and see Chapter 9 by Li et al. ).

## Conclusions

The discussion presented here clearly show that FTIR spectroscopy offers far more advantages than disadvantages for conformational characterisation of peptides. The major problem with the technique concerns the accurate assignment of the different amide I components. A shift in the position of an amide I component may not necessarily represent a change in secondary structure. For example, a helical structure highly solvated by water molecules can display a lower amide I band frequency compared with a helix buried inside the hydrophobic core of a phospholipid membrane. The precise position of the band may vary depending on the length of the  $\alpha$ -helix. The extinction coefficient of the different secondary structure elements may also vary depending on the nature of the surrounding environment. Further studies are necessary to understand how these and other factors govern the position, intensity and width of bands in the spectra of peptides. Assignment of infrared bands is likely to improve in the future thanks to the development of methods such as isotope-edited FTIR spectroscopy. By combining such procedures with time-resolved FTIR spectroscopy, it should be possible to investigate protein folding and stability in far greater detail than it is possible currently. By far the most attractive aspect of FTIR spectroscopy, as compared to other techniques, is the ease with which it can be used for conformational analysis of peptides in a wide range of environment, especially in phospholipid membranes. Although there are many obstacles that needs to be overcome, the future of biomolecular FTIR spectroscopy is a bright one.

## References

- 1 Elliot, A. and Ambrose, E.J. *Nature*, **1950**, *165*, 921-922
- 2 Susi, H. and Byler, D.M. *Methods. Enzymol.* **1986**, *150*, 290-311.
- 3 Surewicz, W.K. and Mantsch, H.H. *Biochem. Biophys. Acta.* **1988**, *952*, 115-130.
- 4 Haris, P.I. and Chapman, D. *Trends. Biochem. Sci.* **1992**, *17*, 328-333.
- 5 Braiman, M.S. and Rothschild, K.J. *Ann. Rev. Biophys. Biophys. Chem.*, **1988**, *17*, 541-570.
- 6 Mantele, W. *Trends in Biochem. Sci.*, **1993**, *18*, 197-202.
- 7 Maeda, A., Kandori, H., Yamazaki, Y., Nishimura, S., Hatanaka, M., Chon, Y.S., Sasaki, J., Needleman, R., Lanyi, J.K. *J. of Biochemistry*, **1997**, *121*, 399-406.
- 8 Siebert, F. *Mikrochimica Acta.*, **1997**, *S14*, 43-50
- 9 Gerwert, K. *Curr. Opin. in Struct. Biol.*, **1993**, *3*, 769-773.

- 10 Tamm, L.K. and Tatulian, S.A. *Qtr. Rev. Biophys.*, **1997**, *30*, 365-429.
- 11 Goormaghtigh, E. and Ruyschaert, J.M. In *Molecular Description of Biological Components by Computer Aided Conformational Analysis* (Brasseur, R., Ed.) CRC Press, Boca Raton, FL. pp 1998, 285-329.
- 12 Arrondo, J.L.R., and Goni, F.M. In *Protein-Lipid Interactions* (Watts, A., Ed.) Elsevier, Amsterdam, 1993, pp 321-349.
- 13 Jackson, M. and Mantsch, H.H. *Biochim. Biophys. Acta.* **1992**, *1118*, 139-143.
- 14 Waterhous, D.V., and Johnson, W.C. *Biochemistry*, **1994**, *33*, 2121-2128.
- 15 Haris, P.I., Lee, D.C. and Chapman, D. *Biochem. Biophys. Acta.*, **1986**, *874*, 255-265.
- 16 Dosseau, F., Therrien, M., and Pezolet, M. *Appl. Spectrosc.* **1989**, *43*, 538-542.
- 17 Powell, J.R., Wasacz, F.M., and Jakobsen, R.J. *Appl. Spectrosc.* **1986**, *40*, 339-344.
- 18 Fringeli, U.P. and Gunthard, H.H. In *Membrane Spectroscopy* (Grell, E., Ed.) Springer-Verlag, Berlin, 1981, Vol. 31, pp 270-332.
- 19 Goormaghtigh, E., Brasseur, R., Huart, P. and Ruyschaert, J.M. *Biochemistry* **1987**, *26*, 1789-1794.
- 20 Harrick, N.J. *Internal Reflection Spectroscopy*, Interscience, Wiley, New York, 1967.
- 21 Jackson, M., and Mantsch, H.H. *Appl. Spectros.* **1992**, *46*, 699-701.
- 22 Byler, D.M. and Susi, H. *Biopolymers* **1986**, *25*, 469-487.
- 23 Jackson, M., Haris, P.I., Chapman, D. *Biochim. Biophys. Acta.* **1989**, *998*, 75-79.
- 24 Venyaminov, S.Y. and Kalnin, N.N. *Biopolymers*, **1991**, *30*, 1243-1257.
- 25 Lee, D.C., Haris, P.I., Chapman, D. and Mitchell, R.C. *Biochemistry*, **1990**, *29*, 9185-9193.
- 26 Dosseau, F. and Pezolet, M. *Biochemistry*, **1990**, *29*, 8771-8779.
- 27 Pribic, R., van Stokkum, I.H.M., Chapman, D., Haris, P.I. and Bloemendal, M. *Anal. Biochem.*, **1993**, *214*, 366-378.
- 28 Sarver, R.W., and Krueger, W.C. *Anal. Biochem.*, **1991**, *194*, 89-100.
- 29 Rahmelow, K, Hubner, W *Anal. Biochem.*, **1996**, *241*, 5-13
- 30 Baumruk, V, Pancoska, P, Keiderling, T. A *J. Mol. Biol.*, **1996**, *259*, 774-791.
- 31 Williams, R.W. *J. Mol. Biol.* **1983**, *166*, 581-603.
- 32 Rahmelow, K, Hubner, W., Ackermann, T. *Anal. Biochem.*, **1998**, *257*, 1-11
- 33 Alvarez, J., Haris, P.I., Lee, D.C. and Chapman, D. *Biochim. Biophys. Acta.*, **1987**, *916*, 5-12.
- 34 Fu, F.N., Deoliveira, D.B, Trubmble, W.R, Sarkar, H.K, Singh, B.R. *Appl. Spectrosc.*, **1994**, *48*, 1432-1441
- 35 Mitchell, R.C., Haris, P.I., Fallowfield, C., Keeling, D.J. and Chapman, D. *Biochim. Biophys. Acta.*, **1989**, *941*, 31-38.
- 36 Dieudonne, D., Gericke, A., Flach, C.R., Jiang, X., Farid, R.S. and Mendelsohn, R. *J. Am. Chem. Soc.*, **1998**, *120*, 792-799.
- 37 Susi, H., Timasheff, S.N. and Stevens, L. *J. Biol. Chem.*, **1967**, *242*, 5460-5466.
- 38 Jackson, M., Haris, P.I. and Chapman, D. *Biochemistry*, **1991**, *30*, 9681-9686.
- 39 Trewhella, J., Liddle, W.K, Heidorn, D.B, Strynadka, N. *Biochemistry*, **1989**, *28*, 1294-1301.

- 40 Gilmanshin, R, Williams, S, Callender, R.H, Woodruff, W.H and Dyer, R.B. *Proc. Natl. Acad. Sci. U.S.A.*, **1997**, *94*, 3709-3713.
- 41 Haris, P.I. and Chapman, D. *Biochem. Soc. Trans.*, **1989**, *17*, 161-162.
- 42 Reisdorf, W.C. and Krimm, S. *Biochemistry*, **1996**, *35*, 1383-1386.
- 43 Krimm, S. and Dwivedi, A.M. *Science*, **1982**, *216*, 407-408.
- 44 Sansom, M.S.P. The biophysics of peptide models of ion channels. *Prog. Biophys. Molec. Biol.* **1991**, *55*, 139-235.
- 45 Woolley, G.A. and Wallace, B.A. *J. Membrane Biol.* **1992**, *129*, 109-136.
- 46 Fox, R., and Richards, F.M. *Nature*, **1982**, *300*, 325-330.
- 47 Terwilliger, T.C. and Eisenberg, D. *J. Biol. Chem.* **1982**, *257*, 6010-6015.
- 48 Vogel, H. *Biochemistry*, **1987**, *26*, 4562-4572.
- 49 Brauner, J.W., Mendelsohn, R. and Prendergast, F.G. *Biochemistry*, **1987**, *26*, 8151-8158.
- 50 Haris, P.I. and Chapman, D. *Biochim. Biophys. Acta.* **1988**, *943*, 375-380.
- 51 Miick, S.M., Martinez, G.V., Fiori, W.R., Todd, A.P., Millhauser, G.L. *Nature*, **1992**, *359*, 653-655.
- 52 Prestrelski, S.J., Byler, D.M. and Thompson, M.P. *Int. J. Peptide Protein Res.* **1991**, *37*, 508-512.
- 53 Kennedy, D.F., Crisma, M., Toniolo, C. and Chapman, D. *Biochemistry*, **1991**, *30*, 6541-6548.
- 54 Landau, E.M. in *Biomembrane Structures* (Haris, P.I. and Chapman, D. Eds.), IOS Press, Amsterdam, 1998, pp. 1-19.
- 55 Torii, H. and Tasumi, M. *J. Chem. Phys.* **1992**, *96*, 3379-3387.
- 56 Earnest, T.N., Herzfeld, J. and Rothschild, K.J. *Biophys. J.*, **1990**, *58*, 1539-1546.
- 57 Haris, P.I., Coke, M. and Chapman, D. *Biochim. Biophys. Acta.*, **1989**, *995*, 160-167.
- 58 Mantsch, H.H., Perczel, A., Hollosi, M. and Fasman, G.D. *Biopolymers*, **1993**, *33*, 201-207.
- 59 Vass, E, Kurz, M, Konat, R.K, Hollosi, M. (1998) *Spectrochim. Acta. Pt. A-Mol. & Biomol. Spectrosc.* **1998**, *54*, 773-786
- 60 Nakamura, T., Furunaka, H., Miyata, T., Tokunaga, F., Muta, T., Iwanaga, S., Niwa, M., Takao, T., and Shimonishi, Y. *J. Biol. Chem.* **1988**, *263*, 16709-16712.
- 61 Matsuzaki, K. In *Biomembrane Structures* (Haris, P.I. and Chapman, D. Eds.), IOS Press, Amsterdam, 1998, pp. 205-227.
- 62 Arrondo, J.L.R., Blanco, F.J, Serrano, L., Goni, F.M., *FEBS Lett.* **1996**, *384*, 35-37.
- 63 Sanders, J.C., Haris, P.I., Chapman, D., Otto, C. and Hemminga, M.A. *Biochemistry*, **1993**, *32*, 12446-12454.
- 64 Henry, G.D. and Sykes, B.D. *Biochemistry*, **1992**, *31*, 5284-5297.
- 65 Pongs, O. *Physiol. Rev.* **1992**, *72*, S69-S88.
- 66 Haris, P.I. and Chapman, D. In *Biomembrane Structures* (Haris, P.I. and Chapman, D. Eds.), IOS Press, Amsterdam, 1998, pp. 134-168.
- 67 Montal, M. *Ann. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 31-57.
- 68 Haris, P.I., Ramesh, B., Sansom, M.S.P., Ker, I.D., Srail, K.S. and Chapman, D. *Protein Engineering.* **1994**, *7*, 255-262.

- 69 Brazier, S.P., Ramesh, B., Haris, P.I., Lee, D.C. and Srail, S.K.S. *Biochem. J.*, **1998**, *335*, 375-380.
- 70 Mercer, E.A.J., Brazier, S. Abbott, G.W., Ramesh, B., Haris, P.I and Srail, S.K.S. *Biochem. J.*, **1997**, *325*, 475-479.
- 71 Kandori, H, Yamazaki, Y, Hatanaka, M, Needleman, R, Brown, L.S, Richter, H.T, Lanyi, J.K, Maeda, A. *Biochemistry*, **1997**, *36*, 5134-5141
- 72 Kandori, H. and Maeda, A. *Biochemistry*, **1995**, *34*, 14220-14229.
- 73 Hessling, B. Souvignier, G. and Gerwert, K. *Biophys. J.*, **1993**, *65*, 1929-1941.
- 74 Hadden, J.M., Bloemendal, M., Haris, P.I., Srail, S.K.S, Chapman, D. *Biochim. Biophys. Acta.*, **1994**, *1205*, 59-67.
- 75 Haris, P.I., Chapman, D., Harrison, R.A., Smith, K.F. and Perkins, S.J. *Biochemistry*, **1990**, *29*, 1377-1380.
- 76 Alvarez, J., Lee, D.C., Baldwin, S.A. and Chapman, D. *J. Biol. Chem.*, **1987**, *262*, 3502-3509.
- 77 Haris, P.I., Ramesh, B., Brazier, S. and Chapman, D. *FEBS Lett.*, **1994**, *349*, 371-374.
- 78 Mulvey, D., King, G.F., Cooke, R.M., Doak, D.G., Harvey, T.S. and Campbell, I.D. *FEBS Lett.* **1989**, *257*, 113-117.
- 79 Durrell, S.R. and Guy, H.R. *Biophys. J.*, **1992**, *62*, 238-250.
- 80 Azpiazu, I., Gomez-Fernandez, J.C. and Chapman, D. *Biochemistry*, **1993**, *32*, 10720-10726.
- 81 Wolkers, W.F., Haris, P.I., Pistorius, A.M.A., Chapman, D. and Hemminga, M.A. *Biochemistry*, **1995**, *34*, 7825-7833.
- 82 Billeter, M. *Qtr. Rev. Biophys.*, **1992**, *25*, 325-377.
- 83 Hadden, J.M., Chapman, D. and Lee, D.C. *Biochim. Biophys. Acta.*, **1995**, *148*, 115-122.
- 84 Lee, D.C, Herzyk, E., and Chapman, D. *Biochemistry*, **1987**, *26*, 5775-5783.
- 85 Unwin, N. *J. Mol. Biol.*, **1993**, *229*, 1101-1124.
- 86 Gorne-Tschelnokow, U, Strecker, A., Kaudik, C., Naumann, D., Hucho, F. *EMBO J.*, **1994**, *13*, 338-341.
- 87 Baenziger, J.E. & Methot, N. *J. Biol. Chem.* **1995**, *270*, 29129-29137.
- 88 Frey, S. and Tamm, L.K. (1991) *Biophys. J.* **1991**, *60*, 922-930.
- 89 Callender, R.H, Dyer, R.B, Gilmanshin, R. and Woodruff, W.H. *Ann. Rev. Phys. Chem.*, **1998**, *49*, 173-202..
- 90 Haris, P.I., Robillard, G.T., Van Dijk, A.A. and Chapman, D. *Biochemistry*, **1992**, *31*, 6279-6284.
- 91 Li, T.S., Horan, T., Osslund, T., Stearns, G., Arakawa, T. *Biochemistry*, **1997**, *36*, 8849-8857.

# Attenuated Total Reflection Infrared Spectroscopy: Orientation and Tertiary Structural Changes of Proteins or Peptides Inserted into a Lipid Bilayer

E. Goormaghtigh and J.-M. Ruyschaert

LCPMI, Free University of Brussels, B-1050 Brussels, Belgium

We describe here some of the potential uses of ATR-IR spectroscopy. Our first purpose is to demonstrate that orientation of proteins and peptides in a lipid environment can be determined by polarized attenuated total reflection (ATR) spectroscopy on thin hydrated films deposited on attenuated total reflection crystals. Second, shifts to lower wavenumbers upon exposure of a protein to an  $^2\text{H}_2\text{O}$  environment are shown to give access to information on the stability of the individual secondary structure types of a protein. The effect of ligand binding on the hydrogen-deuterium exchange is investigated in order to determine how ligand binding influences the dynamic properties of the protein.

Specific functions essential to cell life are carried out by membrane proteins inserted into a lipid bilayer. Therefore determination of the structure of membrane proteins has become a major problem. To date, the structure of about 8 membrane proteins (photosynthetic reaction centre, porin, bacteriorhodopsin, cytochrome c oxidase, light-harvesting complexes,  $\alpha$ -hemolysin and prostaglandin synthase, mitochondrial ATPase) have been solved with a reasonably high resolution.

High quality crystals of membrane proteins are extremely difficult to obtain. The method is inherently static and provides no dynamical information. In most cases the three-dimensional crystals require the replacement of the lipid bilayers by detergents. Recently, Rosenbush et al. provided evidence that crystals can be formed in a lipid cubic phase (1)

Nuclear magnetic resonance spectroscopy remains limited to small soluble proteins because the line broadening effect with large structures and non isotropic motion of protein within the lipid matrix. Membrane vesicles containing both lipids and proteins or peptides are therefore still difficult to examine. Solid state NMR offers promising

opportunities but is limited to small proteins. Circular dichroism is extremely sensitive to light scattering and spectra have to be corrected for flattening effects. Raman spectroscopy offers valuable information on amino acids and side chains, that are complementary to IR data but the low signal to noise ratio is a serious limitation.

A technique that has become popular for membrane spectroscopists is Attenuated Total Reflection (ATR) Fourier transform infrared (FTIR) spectroscopy. The method is technically simple and is extremely sensitive. It requires microgram quantities of sample. Spectra are recorded in a matter of minutes. The light scattering problems are non-existent and spectra of large membrane fragments can be recorded. One of the main advantages of the method is that it allows a simultaneous study of the structure of lipids and proteins in biological membranes. It is worth to mention that IR spectroscopy has a very short time scale ( $10^{-13}$  sec) as compared to trans-gauche isomerization ( $5 \cdot 10^{-9}$  sec), rotational diffusion ( $10^{-8}$  to  $10^{-10}$  sec) and lipid lateral diffusion constant ( $10^{-8}$  cm<sup>2</sup>/s). Therefore, IR spectroscopy provides an instantaneous description of a molecule whereas average values are usually obtained from ESR and NMR measurements.

Several excellent reviews, emphasizing different aspects of IR spectroscopy are available (2-8) and in some of them (9-13) ATR-IR spectroscopy was treated in more details.

Our major goal in this review is to focus on two aspects of ATR-FTIR spectroscopy. The first deals with polarized ATR-FTIR as a tool for determining the orientation of proteins and lipids in biological membranes. The second focuses on the use of the hydrogen/deuterium exchange of protein amide hydrogens to gain information on the stability and/or solvent accessibility of different secondary structures in various experimental conditions.

## RESULTS

### Polarized ATR-FTIR for orientation determination in proteins and lipids - General principles

The method is based on the fact that the infrared light absorption is maximum if the dipole transition moment is parallel to the electric field component of the incident light.

In an ordered membrane system deposited on the IRE (Internal Reflection Element) surface, all the molecules, and therefore the transition dipole moments within the membrane, have the same orientation with respect to a symmetric axis, which is normal to the IRE plane. By measuring the spectral intensity for different polarization orientation with a polarizer, it is possible to obtain the orientation of the transition dipoles. In fact, all the orientation information is contained in the dichroic ratio  $R^{\text{ATR}}$  which is the ratio of the integrated absorbance of a band measured with a parallel polarization of the incident light  $A_{//}$  to the absorbance measured with a perpendicular polarization of the incident light  $A_{\perp}$  :

$$R^{\text{ATR}} = \frac{A_{//}}{A_{\perp}} \quad (1)$$

For the geometry described in Figure 1, the relation between  $R^{ATR}$  and the dipole orientation can be computed by averaging for all orientations around the IRE normal as well as for all tilts with respect to this normal. The former averaging is trivial since we suppose here a uniaxial symmetry for the distribution around this normal. The latter averaging must take into account the distribution of the orientations with respect to the normal described by a density distribution function  $C(\beta)$  (see below). In fact, distribution of the orientations around the mean value and other source of disordering must be taken into account in real systems. The dichroic ratio  $R^{ATR}$  is related to an orientational order parameter  $S$  by

$$R^{ATR} = \frac{Ex^2}{Ey^2} + \frac{E^2 z}{E^2 y} \left( 1 + \frac{3S}{1-S} \right) \quad (2)$$

Where  $Ex^2$ ,  $Ey^2$  and  $Ez^2$  are the time averaged square electric field amplitude of the evanescent wave in the film at the IRE/film. Interestingly, the evanescent wave exhibits electric field components in all directions. Therefore, in contrast to the transmission situation, no tilt of the oriented membrane is necessary to gain full information on dipole orientation.  $S$  is defined as the projection of the density distribution function  $C(\beta)$  (around the mean value of the orientation) on the second Legendre polynomial. For uniaxial symmetry, it is computed as:

$$S = \int_0^\pi \frac{3 \cos^2 \Gamma - 1}{2} C(\Gamma) \sin(\Gamma) d\Gamma \quad (3)$$

where  $\beta$  refers to the considered distribution. It appears here that if the angular distribution function is developed in Legendre polynomials, only the second term of this development governs the dichroism in ATR-FTIR. The reader is referred to Goormaghtigh et al.(10) for derivation of these equations for the specific case of ATR.

For the clarity of the subsequent discussion, we consider below the fact that the orientation of an  $\alpha$ -helix long axis with respect to the IRE surface normal must be determined. The measured order parameter  $S$ , denoted below  $S_{\text{experimental}}$ , obtained from ATR through relation (1) can be generally expressed as a set of nested, uniaxial distributions (Fig. 2)

$$S_{\text{experimental}} = S_{\text{membrane}} \cdot S_{\text{helix}} \cdot S_{\text{dipole}} \quad (4)$$

Where  $S_{\text{membrane}}$  describes the distribution function of the lipid membrane patches (smallest planar membrane unit) with respect to the internal reflection element.  $S_{\text{helix}}$  describes the orientation of the helices within the membrane plane and  $S_{\text{dipole}}$  describes the dipole orientation of either amide I or amide II with respect to helix axis. A schematic representation of these nested distributions appears in Figure 2.  $S_{\text{helix}}$  can be



further split by describing it as the convolution of an order function  $S_{\text{helix order}}$  by a delta function describing the mean angle of the helix with respect to a membrane perpendicular  $S_{\text{helix angle}}$

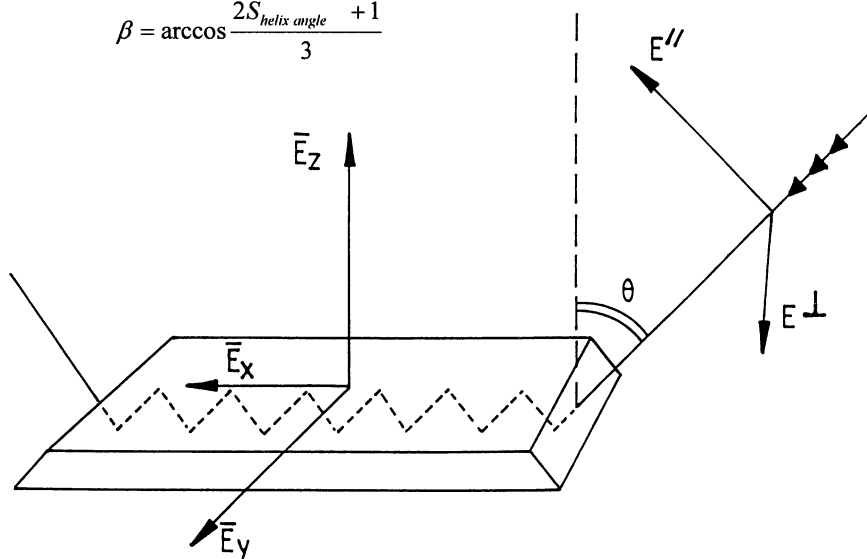
$$S_{\text{helix}} = S_{\text{helix angle}} \cdot S_{\text{helix order}} \quad (5)$$

Because the transition dipole has a unique distribution of the angle  $\alpha$ , it can generally be described by a delta function.

$$S_{\text{dipole}} = \frac{3 \cos 2\alpha - 1}{2} \quad (6)$$

$S_{\text{helix angle}}$  can now be evaluated from  $S_{\text{experimental}}$ . Assuming  $S_{\text{helix order}} = 1$  places clear limits on  $S_{\text{helix angle}}$  as will be discussed below. In this hypothesis, the tilt  $\beta$  of the helix axis with respect to the membrane normal can be determined provided that  $S_{\text{dipole}}$  and  $S_{\text{membrane}}$  are known. It is usually considered that  $S_{\text{membrane}} = 1$  and  $S_{\text{dipole}}$  is a characteristic of the secondary structure.

$$\beta = \arccos \frac{2S_{\text{helix angle}} + 1}{3}$$



### Internal reflection element for ATR

Figure 1. Schematic representation of the internal reflection element (IRE) and of the light pathway. The cartesian components of the electric field are shown along the X, Y and Z axis. Two plane of polarization of the incident light are indicated by  $E_{//}$  (polarization parallel to the incident plane) and  $E_{\perp}$  (polarization perpendicular to the incident plane). The incident beam makes an angle  $\theta$  with respect to a normal to the IRE surface. The edges of the IRE are beveled so that the incident beam penetrates the IRE through a surface that is perpendicular to its propagation.

## Potentialities for orientation determination of membrane components - Alignment of the Apolipoporphin-III $\alpha$ -helices in complexes with dimyristoylphosphatidylcholine

Apolipoporphin-III from *Manduca sexta* can exist in two states : as a globular, lipid-free helix bundle or a lipid surface associated apolipoprotein. Previous papers had investigated the structures of this protein in the lipid-free state and associated to lipids (14). Association of apolipoporphin III with dimyristoylphosphatidylcholine leads to the formation of lipid discs with an average diameter and thickness of  $18.5 \pm 2$  and  $4.8 \pm 0.8$  nm, respectively. Each disc contains six molecules of apolipoporphin-III. Geometrical calculations together with x-ray crystallographic data obtained for an homologous apolipoporphin-III (15) have allowed the presentation of a model of lipid-protein interaction in which the  $\alpha$ -helices of the apolipoporphin III orient perpendicular to the phospholipid chains and surround the lipid discs (Fig.3). This model is nevertheless quite hypothetical regarding the helix orientation. We want to illustrate here how polarized attenuated total reflection Fourier transform infrared spectroscopy has provided the first experimental evidence of such an unique perpendicular orientation. From the spectrum recorded with  $90^\circ$  polarized light a dichroic spectrum obtained by subtracting the spectrum recorded with  $0^\circ$  polarized light shows a negative deviation located in the helix region of the amide I band, indicating an orientation of the  $\alpha$ -helices parallel to the germanium plate. This orientation of the helices is confirmed by the positive deviation in the amide II band at  $1549 \text{ cm}^{-1}$  corresponding to the  $\alpha$ -helix amide II band. Indeed the amide I and amide II polarizations are perpendicular (2) (Fig.4).

It should be noted that for an isotropic distribution of the dipoles

$$R_{iso}^{ATR} = \frac{E_x^2 + E_z^2}{E_y^2}$$

and due to the particular magnitudes of  $E_x, E_y, E_z$ , it is interesting to note that

$$R_{iso}^{ATR} \neq 1$$

While the isotropic dichroic ratio measured for the C=O lipid dipole was 1.51, it was 1.21 for the amide I component. The calculated helix axis orientation makes a maximum tilt with the germanium surface of  $20^\circ$ . The orientation of the lipid acyl chains was also assessed. In DMPC, the hydrocarbon chain in the  $\alpha$ -position in the gel state is all-trans from the ester groups to the methyl groups. This conformation allows a resonance to occur between the ester groups and the  $\text{CH}_2$  groups of the chain, giving rise to the so-called  $\gamma_w(\text{CH}_2)$  progression between  $1180$  and  $1350 \text{ cm}^{-1}$  (peaks at  $1201, 1230, 1257, 1280, 1304$  and  $1328 \text{ cm}^{-1}$ ). The peak at  $1200 \text{ cm}^{-1}$  was chosen to characterize the lipid acyl chain orientation. The strong  $90^\circ$  polarization of this absorption peak indicates that the all-trans hydrocarbon chains of DMPC are oriented normal to the germanium plate. The measured dichroic ratio is 3.4-4.0. Accordingly, the maximum tilt between the acyl chains and the normal to the Germanium surface is equal to  $20^\circ$ . These data provide strong evidence that acyl chains and  $\alpha$ -helices are almost perpendicular to each other in the apolipoporphin III -DMPC complex (17).

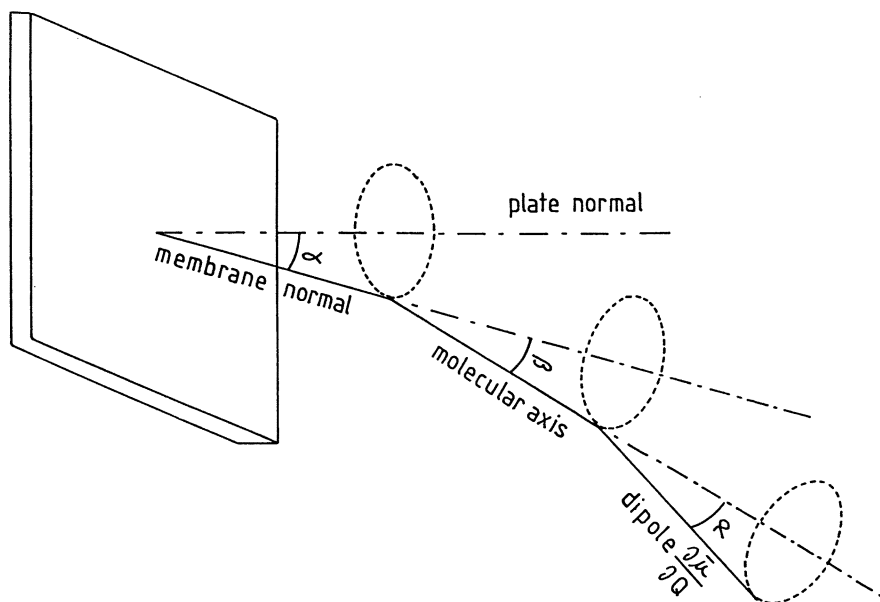


Figure 2 : Set of nested axially symmetric distributions. The membrane normal is distributed about the plate normal (angle  $\gamma$ ) ; the molecular axis is about the membrane normal (angle  $\beta$ ) ; and the transition dipole moment  $\delta\mu/\delta Q$  is about the molecular axis (angle  $\alpha$ ).

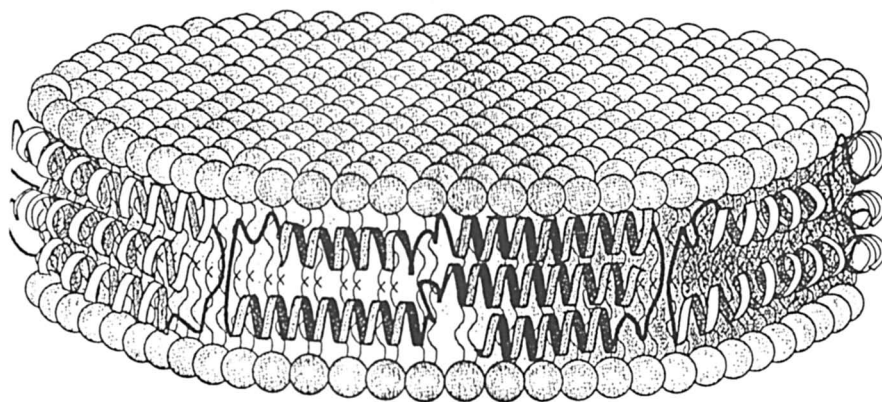


Figure 3 : Hypothetical representation of apoLp-III:DMPC complex

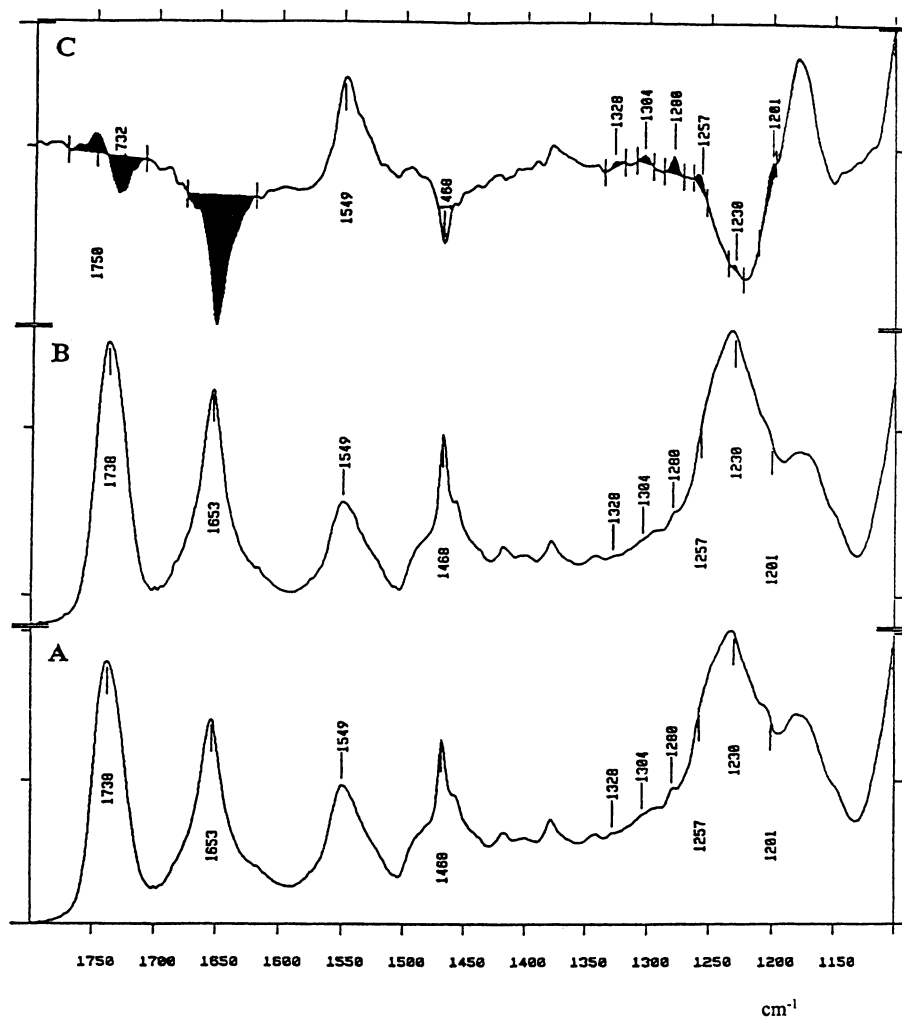


Figure 4 : IR spectra of the apoLp-III-DMPC complex. Spectrum A was obtained with  $90^\circ$  polarized light and spectrum B with  $0^\circ$  polarized light, trace C represents the difference between spectrum A and spectrum B. The optical density amplitudes of spectra A-C are 0.239, 0.141 and 0.30 respectively.

When presented to a lipid surface, the protein probably undergoes a major conformational change, opening about hinge regions located in the loops between helices 2 and 3 and helices 4 and 5. These structural modifications would lead to exposure of the hydrophobic interior of the protein which becomes available for lipid interaction. The model schematised (Fig. 3) is in agreement with an IR analysis of the secondary structure of the apolipoprotein-III revealing that its  $\alpha$ -helical structure is retained as the protein alternates between lipid-free and lipid associated states (data not shown)

**LDL receptor active conformation of apolipoprotein E. Helix organization in N-terminal domain - phospholipid disc particles.** Apolipoprotein (Apo E) is an important plasma protein that functions in triacylglycerol and cholesterol transport. Apo E serves as ligand for at least 3 lipoprotein receptors, the low density lipoprotein (LDL) receptor, the LDL receptor-related protein (LRP), and the very low density lipoprotein (VLDL) receptor. The N-terminal domain of Apo E is capable of transforming multilamellar vesicles of DMPC into disc complexes. Interestingly, in the absence of lipid, the receptor binding region does not interact with the LDL receptor of fibroblasts (18). Thus, it is postulated that lipid binding induces a conformational change in this domain which confers receptor recognition properties. Using linear dichroism, we provided experimental evidence that Apo E 3-N terminal domain  $\alpha$  helices orient perpendicular to the acyl chains of the lipids in discoidal complexes (19). From the x-ray crystal structure of Apo E3 N-terminal domain (20), one can identify a helix which connects helix 1 and helix 2 in the bundle and which is oriented perpendicular to the four long helices. These helical segments in the bundle can be separated into two groups with different orientations. Since analysis of the Amide I band revealed that the percentages of helical secondary structure were quite similar in both lipid free and DMPC bound Apo E3, there was a unique possibility to detect orientations of these two kinds of helices when associated to a lipid matrix.

$$R_a = \frac{A_a''}{A_a^\perp} \quad \text{and} \quad R_b = \frac{A_b''}{A_b^\perp}$$

$R_a$  and  $R_b$  : dichroic ratios of 2 helix population

$$R_{\text{exp}} = \frac{A_a'' + A_b''}{A_a^\perp + A_b^\perp}$$

Resolving these equations :

$x$  : fraction of population a

$1-x$  : fraction of population b

$$R_a = \frac{xR(R_b + 2) + 2(1-x)(R - R_b)}{x(R_b + 2) - (1-x)(R - R_b)}$$

$$R_b = \frac{(1-x)R(R_a + 2) + 2x(R - R_a)}{(1-x)(R_a + 2) - x(R - R_a)}$$

$$x = \frac{(R - R_b)(R_a + 2)}{(R_b + 2)(R_a - R) - (R_a + 2)(R_b - R)}$$

Using the equations described above, we found that the experimental dichroic ratio (1.25) could be decomposed into 2 components with different orientations. The first component representing 92% of the helix content has a 1.11 dichroic ratio while the second component, representing the rest of the helix content, has a dichroic ratio of 4.5. It seems reasonable that the 1.11 ratio representing 92% of the helices could be attributed to the pair of long helices (representing 110 amino acids of a total of 118 in helix conformation) which orient perpendicular to a vector normal to both the face of the disc and the germanium plate (70° minimum tilt). The 4.5 dichroic ratio is attributed to the short 8 amino acids long helix suggesting this helix orients parallel to the normal vector with a maximum tilt of 10-15°.

Experimental data provided evidence that each disc employed in the IR experiments, contained 3 apo E3. In our model, when the bundle opens with helices 1 and 2 on one side and helices 3 and 4 on the other, we estimated that the elongated structure is about 116 Å long. Assuming a helix diameter of 15-17 Å; 2 helices oriented side by side are large enough to effectively cover the DMPC bilayer thickness. Thus, three elongated, open molecules cover 348 Å, in agreement with a disc circumference of 345 Å, as derived from gradient gel electrophoresis. This illustrates how combining dichroism IR and X-ray crystallography measurements allow to describe the molecular organisation (Fig.5) of a lipid-protein complex that none of the 2 methods would have independently.

#### Amide hydrogen/deuterium exchange kinetics

**Tertiary structural changes - Ligand binding:** Detection of the changes in intensity in the amide II (1600-1500 cm<sup>-1</sup>) or amide II' (1500-1400 cm<sup>-1</sup>) regions upon exposure of the protein to <sup>2</sup>H<sub>2</sub>O, provides information on the tertiary structure stability of a

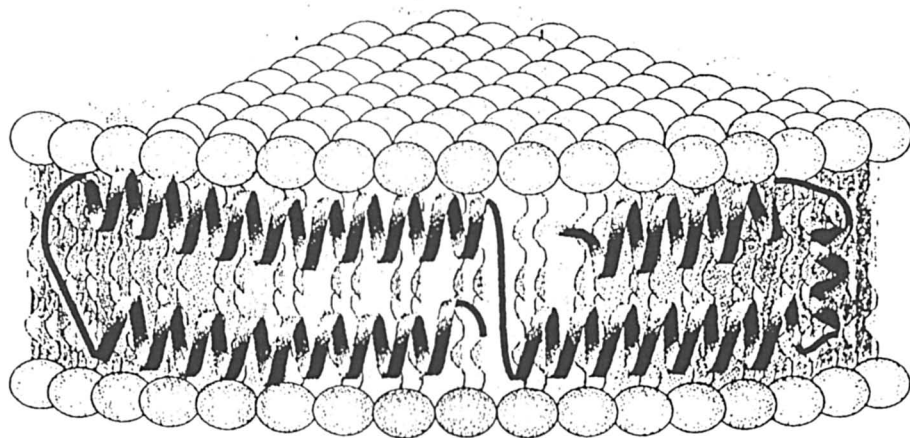


Figure 5 : Model representation of apoE3 (1-183) helix orientation and alignment in DMPC disc complexes.

protein at the molecular level. One particularity of IR spectroscopy is that it allows to monitor the exchange rate of the amide protons only, yielding data which are directly proportional to the number of amino acid residues in the protein.

We demonstrate here that monitoring exchange rate provides dynamical information about tertiary structural changes resulting from interactions between a protein and its specific ligands.

P-glycoprotein is a plasma membrane protein member of the ABC transporter (ATP-binding cassette) superfamily. P-glycoprotein consists of 2 homologous domains, each including a hydrophobic domain and a hydrophilic region containing the ATP binding domain. P-glycoprotein is thought to mediate multidrug resistance by acting as an ATP-dependent pump with a broad specificity. It is able to transport drugs against a concentration gradient to the exterior of the cell (21). P-glycoprotein was recently purified from Chinese hamster ovary CHRB30 cells overexpressing the protein by a combination of anion exchange and immunoaffinity chromatography (22). Our group has reconstituted the P-glycoprotein in asolectin vesicles, and has characterized the reconstituted system with respect to protein orientation, protein structure, ATPase activity, and drug transport (23)

The transport function of the reconstituted P-glycoprotein was assayed by measuring ATP-dependent accumulation of [ $^3\text{H}$ ] daunomycin in the proteoliposomes. Free drug was eliminated by gel filtration on Sephadex G50. In the absence of MgATP, the uptake measured likely represents passive diffusion of the molecule into the vesicles and partitioning into the membrane bilayer, as well as binding to the P-glycoprotein. The difference between daunomycin accumulation in the presence and in the absence of MgATP was therefore considered to represent specific ATP-dependent transport. After 10 min., proteoliposomes treated with MgATP accumulated twice as much daunomycin as proteoliposomes without MgATP.

In order to investigate whether modifications of the secondary structure occurred upon binding of the substrates, the IR spectra of P-glycoprotein recorded after addition of MgATP, MgADP or MgATP-verapamil were subtracted from the spectrum of P-glycoprotein with no substrate added (data not shown). Subtractions were carried out after rescaling all the spectra to the same amide I area. Only minor deviations from the base line were observed in the difference spectra, indicating similar secondary structure contents (data not shown). A quantitative evaluation of the secondary structure was obtained by a Fourier deconvolution and a curve fitting analysis of the amide I region. No significant difference in the protein secondary structure can be detected upon binding of the different molecules to P-glycoprotein (data not shown).

Kinetics of  $^2\text{H}/\text{H}$  exchange of the reconstituted P-glycoprotein was determined in the absence and presence of verapamil-MgATP, MgATP and MgADP in order to check whether, upon binding of these molecules, the protein undergoes tertiary structure modifications. The infrared spectroscopy methodology allowed us to measure the kinetics of deuteration by recording a series of spectra as a function of the time of exposure to  $\text{D}_2\text{O}$ . Ten spectra of the protein were recorded before starting the exchange to verify the reproductibility of the measurements.  $\text{D}_2\text{O}$  saturated  $\text{N}_2$  was

then flushed in the sample chamber and infrared spectra were automatically recorded at increasing time intervals as the exchange proceeded.

Upon deuteration, the amide II band associated with the  $\delta(\text{N-H})$  shifts from the 1500-1580  $\text{cm}^{-1}$  region to the 1460  $\text{cm}^{-1}$  region associated with the  $\delta(\text{N-D})$ . The rate of exchange was estimated from the ratio of the amide II area to the corresponding lipidic  $\nu(\text{C=O})$  area (band at 1748  $\text{cm}^{-1}$ ) to take into account variations of the ATR spectra intensity due to the swelling of the sample film upon hydration. The decrease of the amide II area computed between 0 and 100% , is reported as a function of the deuteration time in Fig.6.

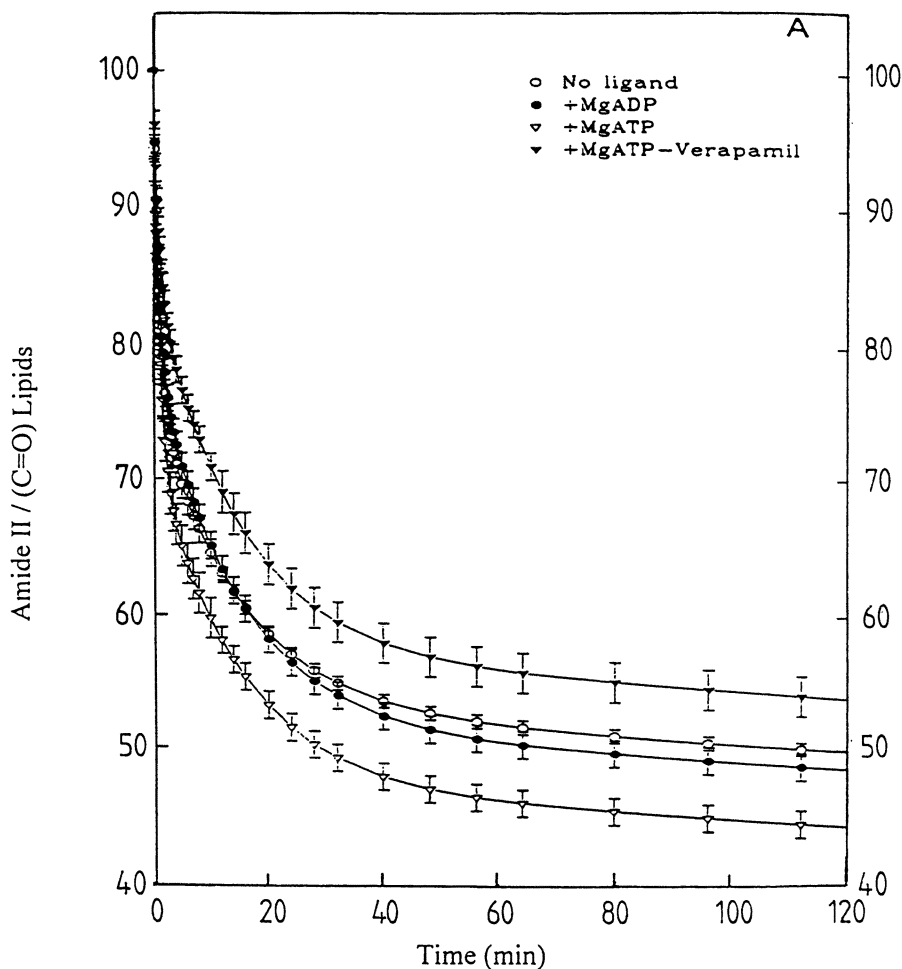


Figure 6 : Evolution of the proportion of exchanged amide bond computed between 0 and 100%, as a function of the deuteration time. (○) no substrate added, (●) MgADP (P-glycoprotein / MgADP molar ratio = 1 :56), (▽) MgATP (P-glycoprotein / MgATP molar ratio = 1 :56), (▼) MgATP-verapamil (P-glycoprotein / verapamil molar ratio = 1 :1).



As shown in Fig. 6, about 50% of P-glycoprotein is poorly accessible to the aqueous medium in the absence of ligand. This percentage is not significantly affected upon binding of MgADP (P-glycoprotein/mgADP molar ratio = 1 :56). Addition of MgATP to P-glycoprotein (P-glycoprotein/MgATP molar ratio = 1:56) clearly results in an increase of the protein exchange rate. This increase is observed as soon as the exchange begins (as shown in Fig. 6), and after 2 h of deuteration, the fraction of exchanged amino acids is increased from 50%(P-glycoprotein free of ligands) to 56% (P-glycoprotein with MgATP) (Fig. 6). On the other hand, P-glycoprotein shows a significant reduction of its exchange rate upon addition of MgATP-verapamil (P-glycoprotein-verapamil molar ratio =1:1). The proportion of exchanged amino acids is reduced from 50% (no ligands) and 56% (MgATP) to 46% (MgATP-verapamil) (Fig.6). Verapamil alone has no effect on the  $^2\text{H}/\text{H}$  exchange kinetics of P-glycoprotein (data is not shown).

Hydrogen atoms with different exchange rate are implicated in the exchange process. Since the  $^2\text{H}/\text{H}$  exchange is a first order reaction, the exchange curve is expected to display a multiexponential decay corresponding to the different groups of amide protons characterized by a common period  $T_i$ .

$$H(t) = \sum_i a_i \exp(-t/T_i)$$

Where  $a_i$  is the proportion of each amide group with identical  $T_i$  values.

The large number of protons makes it impossible to obtain the individual rate constants. One approach to this problem is to choose arbitrarily a number of exponentials to fit the exchange curves. Three exponentials characterized by their period  $T_i$  ( $i = 1-3$ ) and by their proportion  $a_i$  of amide groups were chosen. A nonlinear fitting of all the experimental curves without constraints on the period  $T_i$  and on the  $a_i$  yields three period  $T_1$ ,  $T_2$  and  $T_3$ . Since these periods  $T_1$ ,  $T_2$  and  $T_3$  were similar for all the kinetic curves analyzed, a second fitting was performed for each curve, setting the  $T_i$  to their average value ( $T_1 = 1, T_2 = 15$ , and  $T_3 = 2000\text{min}$ ) in order to compare the proportion  $a_i$  of each amide group for the different experimental curves. The result is shown in Table I.

Addition of MgATP resulted in an increased accessibility to the solvent of a population of slowly exchanging amino acids (76 out of 676 amino acids). These slowly exchanging amide protons ( $T_3 = 2000$  min) are essentially changed into fast ( $T_1 = 1$  min) and intermediate ( $T_2 = 15$  min) exchanging species. Conversely, addition of MgATP-verapamil resulted in protection of a population of fast exchanging amino acids (102 out of 306), demonstrating a verapamil-induced major conformational change in the enzyme. Minor changes in amino acid accessibility were observed upon addition of MgADP, characterized by a decrease of about 25 amino acid residues from both the slow and fast exchanging population to the intermediate one (Table I), resulting in the crossing over of the exchange curves.

**Assignment of the exchanging protons to a secondary structure (24,25)** The amide I, primarily arising from C=O stretching (70-85% of the potential energy) together with an out-of-phase CN stretching, a CCN deformation, and a small NH in-plane bending contribution (2), generally shifts by 5-10  $\text{cm}^{-1}$  to lower wavenumber upon exposure of

Table I

Proportion ( $a_1, a_2, a_3$ ) of the three exponential components characterized by a half-decay of  $T_1 = 1, T_2 = 15$ , and  $T_3 = 2000$  min

The proportions are yielded from the analysis of the exchange curves recorded for P-glycoprotein in the presence of different ligands.

Substrate	$a_1$	$a_2$	$a_3$
		%	
No ligand	24(306) <sup>o</sup>	23(295)	53(676)
MgATP	28(357)	25(320)	47(600)
MgATP-verapamil	16(204)	27(345)	57(727)
MgADP	22(281)	27(345)	51(650)

<sup>o</sup> Numbers in parentheses are the number of amino acids involved.

a protein to an  $^2\text{H}_2\text{O}$  environment (5). Detection of these shifts in different regions of amide I should give access to information on the stability of the individual secondary structure types of a protein. In this work, we will describe the exchange of the amide-protons of BPTI as the sum of four different classes of exchanging components. The choice made to define these classes are based on the objective to relate them to the different secondary structure types present in a protein. For this reason, BPTI was chosen in this study as a test case for this approach, since its structure (26,27,28) and amide-proton exchange (29,30,31,32) have been described in detail by X-ray and NMR techniques.

Figure 7A shows the amide I region after deconvolution of four ATR-FTIR spectra of a film of BPTI, at distinct time intervals after initiating amide-proton exchange by exposure of the film to  $^2\text{H}_2\text{O}$ -saturated  $\text{N}_2$  gas. A clear shift of the maximum of the amide I band from 1648 to 1643  $\text{cm}^{-1}$  upon exposure to  $^2\text{H}_2\text{O}$  can be observed.

**Class I and Class II Exchange Kinetics.** The wavenumbers of maximal absorption of  $\alpha$ -helical and random coiled structures are generally close (1658 – 1652  $\text{cm}^{-1}$ ) (5), prohibiting a clear discrimination of the contributions of these two structure types. Any intensity change monitored in this spectral region is a result of the exchange of both the nonstructured and the helical parts. However, when these intensities shift 5-10  $\text{cm}^{-1}$  to lower wavenumbers upon exchange, both of these contributions will exhibit an isobestic point in their absorption spectra.

Either successive infrared difference spectra or deconvolution spectra could be used to search for the isobestic point. The advantage of using deconvoluted spectra over difference spectra is that kinetics curve can easily be obtained as shown in Figure 8. For simplicity we discriminate a fast exchanging component (defined as class I) and a component containing all other exchanging contributions monitored in the 1660-1640  $\text{cm}^{-1}$  region (class II). Monitoring the intensity at the isobestic point of one of these two classes will reveal information on the exchange of the other class. The isobestic

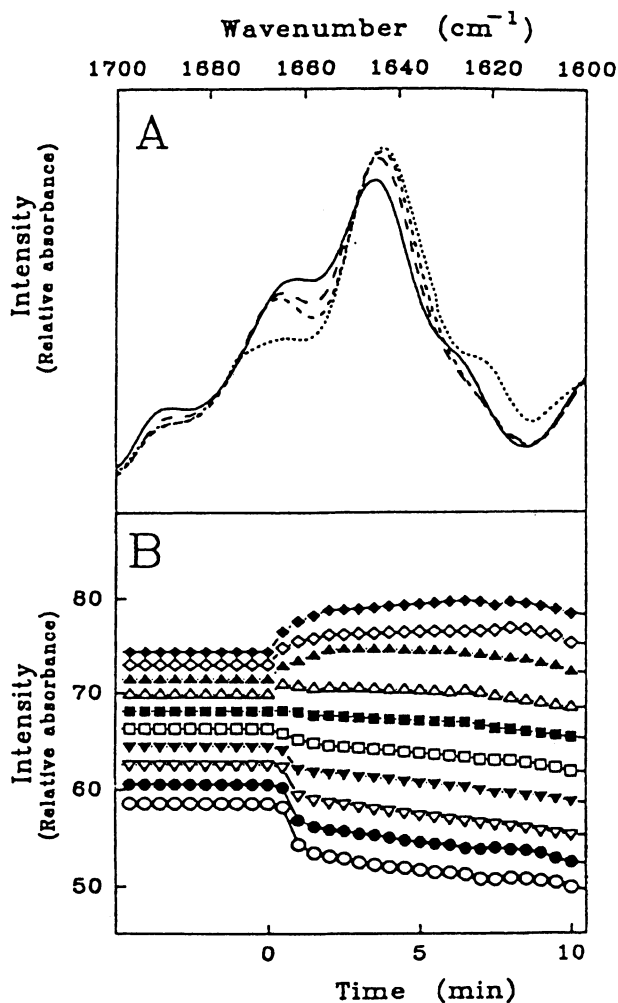


Figure 7: (A) IR spectra recorded of different time periods of exposure of the film to  $2\text{H}_2\text{O}$  saturated  $\text{N}_2$  gas : — 0 min., — — :2.5 min, ---- :11.5 min, ..... : 1140 min, after deconvolution using a Lorentzian line shape with a line width of  $40\text{ cm}^{-1}$ , followed by a Gaussian apodization with a line width of  $20\text{ cm}^{-1}$ . (B) The intensity monitored at various positions in the amide I region of the IR spectra of a film of 200 mg of BPTI, corrected for all contributions related to the replacement of  $\text{H}_2\text{O}$  by  $^2\text{H}_2\text{O}$  as described in the text, during the prescans and the first 10 min. of exposure of a film to  $^2\text{H}_2\text{O}$  ( $\blacklozenge$ ) 1649.5. ( $\diamond$ ) 1650. ( $\blacktriangle$ ) 1650.5. ( $\triangle$ ) 1651. ( $\blacksquare$ ) 1651.5. ( $\square$ ) 1652. ( $\blacktriangledown$ ) 1652.5. ( $\triangledown$ ) 1653. ( $\bullet$ ) 1653.5 and ( $\circ$ ) 1654  $\text{cm}^{-1}$ .

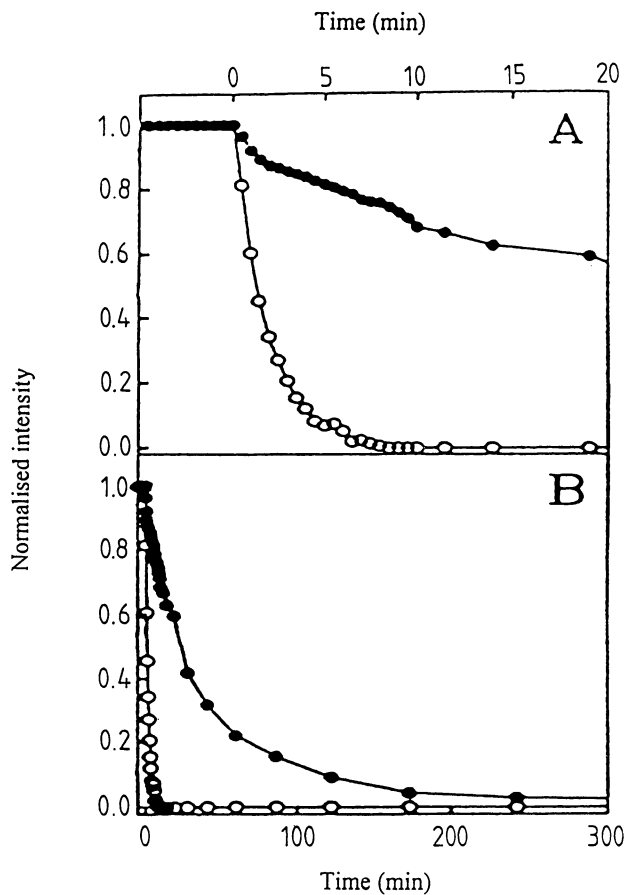


Figure 8 : Kinetics of exchange of the random coiled (○) and helical (●) structures of BPTI, detected at 1640 and 1652  $\text{cm}^{-1}$  respectively, in the deconvoluted spectra displayed for the first 20 min (A) and shown up to 5 h of exposure of the film to  $^2\text{H}_2\text{O}$  (B). In order to compare class I and class II, we rescaled both kinetics between 1 and 0 with the value of 1 corresponding to the undeuterated sample ( $t=0$ ).

point of the class I is searched by analyzing the first 10 min of the exchange kinetics, as shown in Figure 7B, by evaluation of the intensity of the deconvoluted spectra. Whereas an initial strong decrease in intensity can be observed at  $1654\text{ cm}^{-1}$  in the first minute of exchange, an increase of intensity is found at  $1649.5\text{ cm}^{-1}$ , indicating that the isobestic point of the class I exchange must be intermediate to these wavenumbers. Since no initial intensity change is observed at  $1652\text{ cm}^{-1}$  during the first moments of the exposure of the film to  $^2\text{H}_2\text{O}$ , we define this wavenumber as the isobestic point of the class I exchange. Any intensity change observed at this wavenumber does not belong to the fast exchanging component and is therefore attributed to class II exchange only.

Figure 8 shows the kinetics of the class II exchange, monitored at  $1652\text{ cm}^{-1}$  in the deconvoluted spectra, upon exposure of the film to  $^2\text{H}_2\text{O}$  (closed circles). The data have been scaled between 0 and 1 for amide deuteration by making use of data provided by a spectrum of BPTI where all labile protons are replaced by deuterons. Apparently, the class II amide-proton exchange of BPTI is completed in approximately 4 h (Figure 8).

In the next step the exchange kinetics for the class I will be determined below (at  $1640\text{ cm}^{-1}$ ) and above ( $1656\text{ cm}^{-1}$ ) its expected isobestic point by subtraction of the class II contribution. By monitoring the intensity at  $1640\text{ cm}^{-1}$  in the deconvoluted spectra during exchange, one detects the increase of intensity of both the random and helical structures (class I + II), since this wavenumber is well below both their expected isobestic points. Knowing the exchange kinetics of class II (as obtained at  $1652\text{ cm}^{-1}$ ), this contribution can be eliminated from the observed increase of intensity at  $1640\text{ cm}^{-1}$  by subtraction. The amplitude of the class II contribution at this wavenumber is obtained by subtraction of kinetics of this component from the observed intensity at  $1640\text{ cm}^{-1}$ , by making the slope of the exchange curve zero between the 14 and 19 h time points. This procedure is based on the assumption that class I exchange does not contain any contribution with a slow exchange behavior. The result of this subtraction is shown in Figure 8 (open circles), after rescaling between 0 and 100% using the data provided by a spectrum of BPTI where all labile protons are replaced by deuterons. Obviously, the rate of class I exchange is much faster than that found for the class II contribution of BPTI, and the exchange is completed in approximately 10 min.

**Class III and class IV Exchange Kinetics** While class I and II were defined to account for the nonstructured and  $\alpha$ -helical parts of the protein, the class III contribution is supposed to reflect predominantly the exchange of the  $\beta$ -stranded structures. The increase of intensity during the course of exchange, for example, at  $1620\text{ cm}^{-1}$  in the IR spectra (Figure 9) displays the shift to lower wavenumber of pleated  $\beta$ -strands, which, in an  $\text{H}_2\text{O}$  environment, absorb around  $1635\text{-}1630\text{ cm}^{-1}$  (5). At  $1620\text{ cm}^{-1}$  the other structure types are expected to have a negligible contribution. A slow exchange is observed for the class III exchange in BPTI (Figure 9), reaching approximately a 60% exchange after 19 h, based on the intensity at  $1620\text{ cm}^{-1}$  for a sample where all labile protons have been exchanged for deuterons. When the class III exchange at other wavenumbers in the  $1630\text{-}1610\text{ cm}^{-1}$  region is monitored, similar exchange curves are found (not shown), with the difference that, at higher

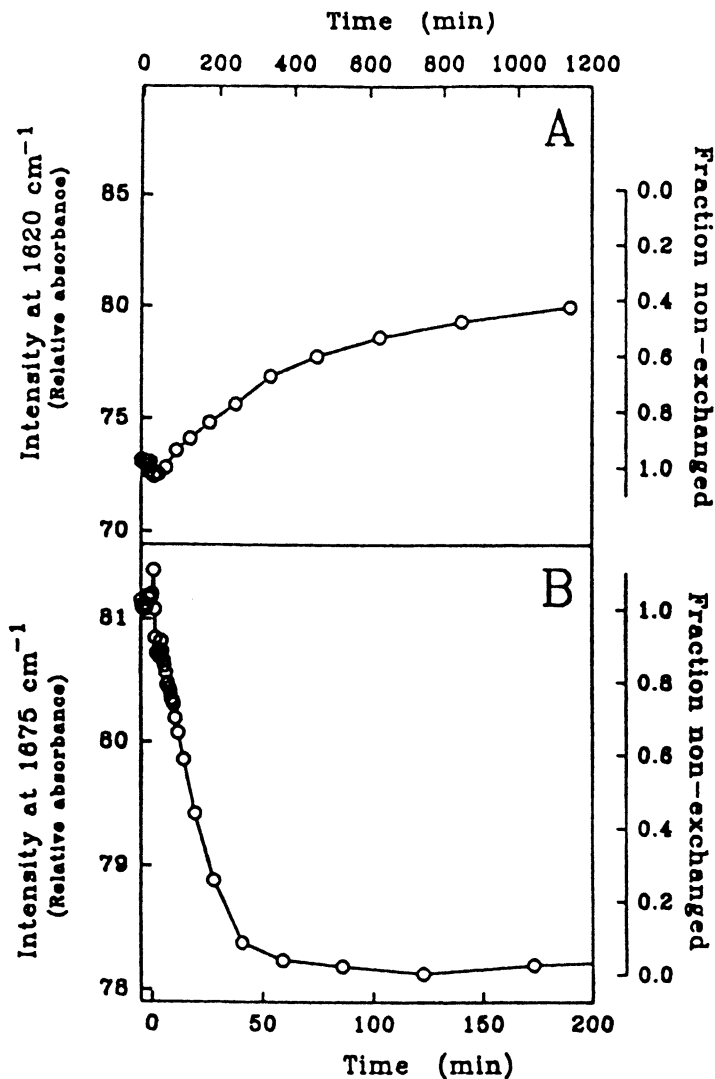


Figure 9 : (A) The IR intensity monitored at 1620 cm<sup>-1</sup> reflecting the exchange of  $\beta$ -strands and (B) at 1675 cm<sup>-1</sup>, representing the exchange of  $\beta$ -turns, is plotted as a function of the time of exposure of a film of 200 mg of BPTI. The right axis represents the scale after normalization of the data using the spectrum of a film of BPTI in which all labile protons have been exchanged

wavenumbers, the interference of class I contributions becomes stronger, as reflected by more pronounced initial intensity changes in the first minutes of the time course.

$\beta$ -turns generally absorb in the 1682-1662  $\text{cm}^{-1}$  region (5). By monitoring then the IR intensity (Figure 9) at 1675  $\text{cm}^{-1}$ , one could detect the shift of this secondary structure type to lower wavenumber upon amide-proton exchange. However,  $\beta$ -strands are also known to have a component of relatively low intensity in the 1689-1682  $\text{cm}^{-1}$  region (5), which results in an increase of intensity at 1675  $\text{cm}^{-1}$  upon exchange. The kinetics of this interfering contribution has been described by the class III exchange, and consequently its contribution to the observed intensity at 1675  $\text{cm}^{-1}$  can be eliminated by subtraction, resulting in what is defined as class IV exchange. For this subtraction the assumption is made that at a later stage of the kinetic experiment the slope of the curve is determined by class III amide-proton exchange only. The class IV exchange is shown in Figure 9, where it can be observed that the class IV contribution to BPTI exchanges within approximately 70 min. The difference in the absolute magnitude of the intensity change recorded for each class (before deuteration minus after complete amide-proton exchange) might be correlated to the secondary structure content of BPTI. From the fit shown in Figure 10 it could be deduced that of the total change of intensity at 1656  $\text{cm}^{-1}$  is 20.4 units of relative absorbance, the contributions of class I and class II being 9.5 and 10.9 units, respectively. From Figure 9 it can be evaluated that the total intensity changes upon deuteration at 1620 and 1675  $\text{cm}^{-1}$  are 13.1 and 3.0 units for class III and class IV, respectively. If these classes represent the individual secondary structures present in BPTI, then these relative contributions should reflect their percentages in the protein. Yet, if a correlation between the intensity changes upon total deuteration for each class and a secondary structure content is to be established, the intensity changes have to be weighted by the molar absorptivities known for each secondary structure. The relative molar absorptivities were shown to 1 :0.69 :0.33 :0.38 respectively for  $\beta$ -strands : helices :  $\beta$ -turns : random coils (33). In agreement with usual assignments for the different secondary structure, classes I-IV are expected to correspond to disordered, helical,  $\beta$ -strand, and  $\beta$ -turns structures, respectively. The assumption that the molar absorptivities of the secondary structures are not significantly influenced by deuteration has to be made (34,35,36). After weighting the intensity changes characteristic of every class by the relative molar absorptivities of the corresponding secondary structure, the contribution of the four classes to BPTI are computed to be 21, 20, 14, and 45%, respectively. Remarkably, these numbers correspond closely to the percentages of the reported individual secondary structures of BPTI 19.0, 22.4, 10.3, and 48.3%, respectively (26,27), suggesting a clear correlation between the different classes and the secondary structure they are intended to represent

**Comparison of the Exchange Kinetics at the Secondary and entire Molecule Level.** Summation of the normalized exchange kinetics of the four individual secondary structures, as shown in Figure 8 for class I and II, and in Figure 9 for class III and IV, multiplied by their proportion in the protein, should then make up the exchange kinetics of the entire protein. In Figure 10 both the normalized exchange obtained by monitoring the amide II/I region (open circles ; experiment not shown) and

that obtained by summation of the individual classes (dashed line) are shown. At a longer time scale (Figure 10B) there is a close agreement between these two data sets. However, within the first 10 min of the kinetics some discrepancy appears (Figure 10A), which might be explained by the fact that in the amide II :I ratio no corrections for H<sub>2</sub>O or side chain contributions have been made, the intensities of which vary only within those first minutes and lead to a slight overestimation of the amide II/I ratio. The general agreement between the two data sets displayed in figure 10 indicates that the exchange of all amide-protons of the protein could be detected and described by four classes of amide-protons. The fact that the secondary structure content correlated to the individual classes is in close agreement with the reported structure of BPTI stresses the potentialities of the present approach to record the exchange characteristics of the different secondary structure types.

### Conclusion

In conclusion, ATR-FTIR spectroscopy of biological membrane components is a very flexible method that allows the study of molecular conformation, molecular orientation within the membrane, and protein folding stability and solvent accessibility in a variety of environments. The method is very sensitive when difference spectroscopy can be applied, i.e. change at the level of one amino acid because measurable for large protein and in the presence of phospholipids. The possibility to modulate the environment of the sample opens the way to the study of induced conformational changes and to the understanding of protein dynamic or enzyme catalytic cycle. With the advent of the fast scan techniques or step scan FTIR, the molecular details of complex processes such as enzyme catalytic cycle is becoming feasible. The description of the of the amino acid protonation/deprotonation, minor secondary structure change and protein segment reorientation in a catalytic cycle will shed a new light on essential processes dynamic in nature that could not be described before by high resolution but static methods such as X-ray diffraction.

### References

- 1) Landau, E.M. and Rosenbusch, J.P., Proc. Natl. Acad. Sci. U.S.A. (1996) 93, 14532-
- 2) Krimm, S. and Bandekar, J., Adv. Prot. Chem. (1986) 38, 181-364.
- 3) Braiman, M.S. and Rothschild, Ann. Rev. Biophys. Biophys. Chem (1988) 17,541-570.
- 4) Arrondo, J.L.R., Muga, A., Castresana, J and Goni, F.M., Prog. Biophys. Mol. Biol. (1993) 59, 23-56.
- 5) Goormaghtigh, E., Cabiliaux, V. and Ruyschaert, J.-M., Subcellular Biochemistry (1994) 23, 329-450.
- 6) Haris, P.I. and Chapman, D., Biopolymers (Peptide Science) (1995) 37, 251-263
- 7) Jackson, M. and Mantsch, H.H., Crit. Rev. Biochem. Mol. Biol. (1995) 30, 95-120.
- 8) Tamm, L.K. and Tatuliam, S.A., Quartely Rev. Biophys. (1997) 30, 365-429.
- 9) Fringeli, U.P. and Günthard, H.H., Mol. Biol. Biochem. Biophys. (1981) 31, 270-332.
- 10) Goormaghtigh, E. and Ruyschaert, J.-M. (1990), Polarized attenuated total reflection infrared spectroscopy as a tool to investigate the conformation and orientation of membrane components. In : Molecular Description of Biological



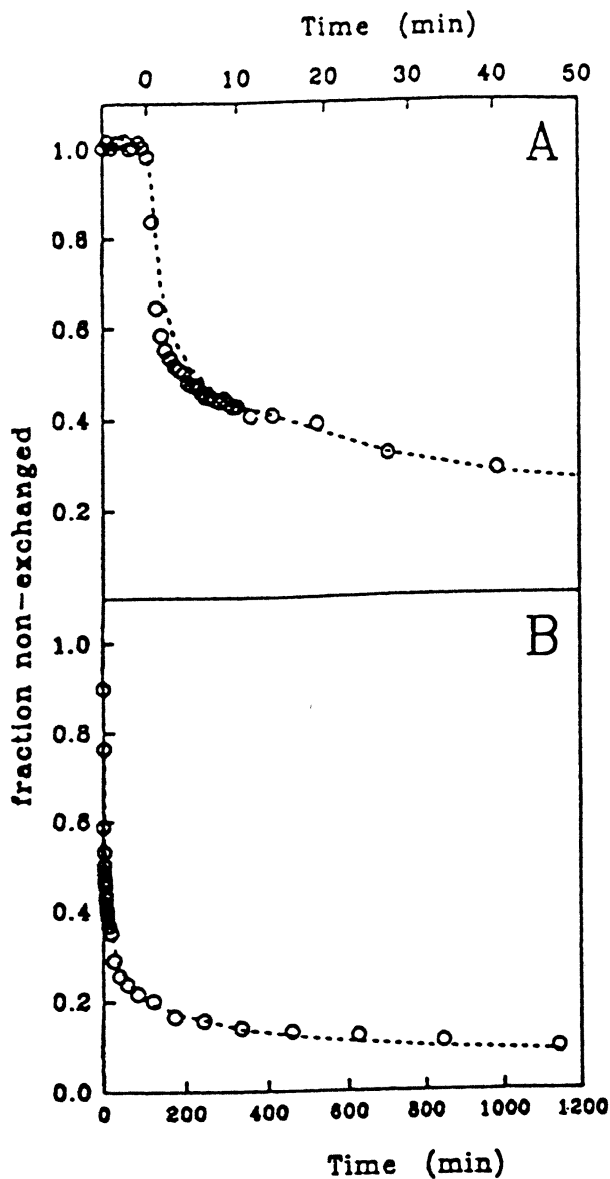


Figure 10 : Normalized amide II' / amide I plotted as a function of time of exposure of a film of 200 mg of BPTI to  $^2\text{H}_2\text{O}$  (open circle) (A) during the prescans and the first 50 min and (B) up to 19 h. The dashed line represents the summation of the individual secondary structure contributions to the amide-proton exchange, based on the secondary structure of the protein and their exchange kinetics.

- Membranes by Computer-Aided Conformational Analysis (ed. R. Brasseur), pp. 285-329, Boca Raton, FL, CRC Press.
- 11) Fringeli, U.P. (1993), In situ infrared attenuated total reflection membrane spectroscopy. In *Internal Reflection Spectroscopy, Theory and Application* (ed. F.M. Mirabella, Jr.), pp 255-324, Marcel Dekker, Inc. : New York.
  - 12) Dluhy, R.A., Stephens, S.A., Widayati, S. and Williams, A.D., *Spectrochim. Acta* (1995) Part A, 51,1413-1447.
  - 13) Axelsen, P.H. and Citra, M.J., *Prog. Biophys. Molec. Biol.* (1996) 66, 227-253.
  - 14) Ryan R.O., Oikawa, K. and Kay, C., *J. Biol. Chem.* (1993) 268, 1525-1530.
  - 15) Breiter, D.R., Kanost, M.R., Benning, M.M, Wesenberg, G, Law, J.H., Wells, M.A., Rayment, I. and Holden, H.M., *Biochemistry* (1991) 30,603-608
  - 16) Bechinger, B., Ruyschaert, J.-M. and Goormaghtigh, E. (in press).
  - 17) Raussens, V., Narayanaswami, V., Goormaghtigh, E., Ryan, R. and Ruyschaert, J.-M., *J. Biol. Chem.* (1995) 270, 12542-12547.
  - 18) Innerarity, T.L., Pitas, R.E. and Mahley, R.W., *J. Biol. Chem.* (1979) 254, 4184-4190.
  - 19) Raussens, V., Fisher, C.A., Goormaghtigh, E., Ryan, R.O and Ruyschaert J.-M, *J. Biol. Chem.* (in press)
  - 20) Wilson, C., Wardell, M.R. Weisgraber, K.H, Chuen-Sang, C.W and Lindgren, F.T., *J. Biol. Chem.* (1988) 263,6249-6258.
  - 21) Shapiro, A. and Ling, V., *J. Biol. Chem* (1995) 270, 16167-16175.
  - 22) Shapiro, A. and Ling, V., *J. Biol. Chem.* (1994) 269, 3745-3754.
  - 23) Sonveaux, N., Shapiro, A, Goormaghtigh, E., Ling, V. and Ruyschaert, J.-M., *J. Biol. Chem.* (1996) 271, 24617-24624.
  - 24) de Jongh, H.H., Goormaghtigh, E. and Ruyschaert, J.-M., *Biochemistry* (1997) 36, 13593-13602.
  - 25) de Jongh, H.H., Goormaghtigh, E. and Ruyschaert, J.-M., *Biochemistry* (1997) 36, 13603-13610.
  - 26) Deisenhofer, J. and Steigemann, W., *Acta Crystallogr* (1975) B31, 238-250.
  - 27) Berndt, K.D., Guntert, P. Orbons, L.P. and Wüthrich, K., *J. Mol. Biol.* (1992) 227, 757-775.
  - 28) Schiffer, C.A., Huber, R., Wüthrich, K. and van Gusteren, W.F., *J. Mol. Biol.* (1994) 241, 588-599.
  - 29) Wagner, G. and Wüthrich, K., *J. Mol. Biol.* (1982) 160, 343-361.
  - 30) Rashin, A.A., *J. Mol. Biol.* (1987) 198, 339-349.
  - 31) Gallagher, W., Tao, F. and Woodward, C.K., *Biochemistry* (1992) 31, 4673-4680.
  - 32) Kim, K.-S., Fuchs, J.A. and Woodward, C.K., *Biochemistry* (1993) 32, 9600-9608.
  - 33) de Jongh, H.H., Goormaghtigh, E. and Ruyschaert, J.-M., *Anal. Biochem.* (1996) 242, 95-103.
  - 34) Chirgadze, Y.N., Shestopalov, B.V. and Venyaminov, S.Y., *Biopolymers* (1973) 12, 1337-1351.
  - 35) Chirgadze, Y.N. and Brazhnikov, E.V., *Biopolymers* (1974) 13, 1701-1712.
  - 36) Venyaminov, S.Y. and Kalnin, N.N., *Biopolymers* (1991) 30, 1259-1271.

# Determination of the Secondary Structure of Proteins from Amide I and Amide III Infrared Bands Using Partial Least-Square Method

S. Cai and Bal Ram Singh

Department of Chemistry and Biochemistry,  
University of Massachusetts at Dartmouth, Dartmouth, MA 02747

Fourier transform infrared spectroscopy is now becoming an important method to study protein secondary structure. The amide I region of protein infrared spectrum is the widely used region, and the amide III region has been comparatively neglected due to their low signal. Since there is no water interference in amide III region, and more importantly, the different secondary structures of proteins have more significant differences in their amide III spectra, it is quite promising to use amide III region to determine protein secondary structure. In our current study, partial least-squares (PLS) method was used to predict protein secondary structures from protein IR spectra. The IR spectra of H<sub>2</sub>O solutions of 9 different proteins of known crystal structure have been recorded, and amide I, amide III, and amide I combined with amide III region of these proteins were used to set up the calibration set for PLS algorithm. Our results correlate quite well with the data from X-ray studies, and the prediction from amide III region is better than that from amide I or combined amide I and amide III regions.

Elucidation of the structure of proteins and of their structure-function relationships is one of the main targets in modern biochemistry and bio-analytical chemistry. Several methods are currently used for the analysis of protein structures. Of these methods, x-ray crystallography provides the most detailed information about protein conformation and structure if high quality single crystals of protein can be prepared. However, it is usually difficult to get high quality single crystals, especially for membrane proteins. Nuclear-magnetic resonance (NMR) spectroscopy is another important tool to analyze proteins, which can provide high resolution structural information of proteins in solution. Unfortunately, NMR is a low sensitivity technique, and therefore at present, it is only

restricted to small proteins of less than 20 kDa (1). The practical limitations encountered in high-resolution structural techniques for proteins (e.g., x-ray crystallography and NMR methods) have stimulated progress in the development and improvement of "low-resolution" spectroscopic methods, which provide global insight into the overall secondary structure of proteins without being able to establish the precise three-dimensional location of individual structural elements. One of such techniques is Fourier-transform infrared (FT-IR) spectroscopy, which is becoming an increasingly useful tool for the structural analysis of peptides and proteins.

The most important advantage of FT-IR spectroscopy is that it can readily follow the conformational change of a protein as a function of change in its environment, say from aqueous to organic solvent (2), or from detergent to lipid bilayer (3). This advantage renders FT-IR spectroscopy especially useful for probing membrane-associated proteins which are difficult to be probed by other spectroscopic techniques.

Methods currently being used to extract information on protein secondary structure from infrared spectra are based on empirical correlations between the frequencies of certain vibrational modes and types of secondary structure of polypeptide chains such as  $\alpha$ -helix,  $\beta$ -sheet,  $\beta$ -turn, and random coil. The mode most often used and by far best characterized in this respect is the so-called amide I mode, which represents primarily the C=O stretching vibrations of amide groups and gives rise to infrared band(s) in the region between 1600 and 1700  $\text{cm}^{-1}$ . Due to strong absorption of water between 1640-1650  $\text{cm}^{-1}$ , most structure determinations by amide I mode are performed in  $\text{D}_2\text{O}$  solutions. However, uncertainty in the NH/ND exchange process may cause a certain degree of ambiguity (4). Also, serious overlapping of the random coil and the  $\alpha$ -helix bands in amide I region makes it difficult to accurately predict  $\alpha$ -helix contents in proteins. Attempts have also been made to exploit other vibrational modes, particularly amide II and amide III bands. Unfortunately, even though the intensity of amide II region is relatively strong, it is not very sensitive to secondary structure changes of proteins. Furthermore, the amide II bands are strongly overlapped by bands originating from amino acid side chain vibrations (5). On the other hand, amide III bands, which are predominantly due to the in-phase combination of N-H in-plane-bending and C-N stretching vibrations, are highly sensitive to the secondary structure folding (6,7). In addition, there is no  $\text{H}_2\text{O}$  interference in this region. Therefore, even though the signal of amide III bands are  $\sim 5$ -10-fold weaker than that of amide I bands, amide III region is still very promising to estimate protein secondary structure content.

In recent years, several researchers have used amide III region to study protein structures (8-14). Since both amide I and amide III regions are broad band, it is not possible to resolve individual bands corresponding to different secondary structure elements from the original spectra. Most common method used to solve this problem is to employ resolution enhancement or band narrowing methods (14). However, there is a certain degree of subjectivity associated methods based on the band narrowing method such as initial choice of input parameters, and assignment of secondary structures in boundary regions. Chemometric methods have been developed to resolve the individual species in a multi-components system (15). Since this type of method is based on full-spectra analysis, it involves the use of a calibration matrix of the IR spectra of proteins of known x-ray structure, and therefore avoids the need to deconvolve the spectra and assign

the bands. Thus, such method can avoid most of the subjectivity associated in the band assignment process. PLS method is one of these methods, which is commonly used in analytical chemistry (16). Previously, the PLS method has been applied to amide I and amide II regions of protein IR spectra to predict protein secondary structure (17). In this article, we describe application of the PLS method in amide III region of protein IR spectra to predict the protein secondary structures. We used 9 different proteins, whose structures have already been known by x-ray crystallography, to set up the calibration matrix, and employed the amide I, amide III, and a combination of amide I and amide III as the spectral region for estimating protein secondary structure.

### Proteins and infrared spectroscopic measurement

The following proteins were purchased from Sigma Chemical Co. (St. Louis, MO), and used without further purification:  $\alpha$ -chymotrypsin, concanavalin A, cytochrome c hemoglobin, immunoglobulin G, lysozyme, myoglobin, ribonuclease A, and trypsin. These proteins were chosen because their secondary structures have already been known from x ray crystallography. All protein solutions were prepared in 20 mM sodium phosphate buffer, pH 7.2. The concentration of these proteins used in our experiments was 1 mg/ml.

A Nicolet Model 8210 FT-IR spectrometer, equipped with a zinc selenide attenuated total reflectance (ATR) accessory and DTGS detector, was used for spectra recordings at room temperature. The spectrometer was purged with CO<sub>2</sub>-free dry air for at least 24 hours before recording spectra. For each spectrum, a 256-scan interferogram was collected at a resolution of 4 cm<sup>-1</sup>. In every case, the single beam spectrum of the buffer and the protein solutions were divided by the background single beam spectrum and then converted to the absorbance spectra. To obtain the protein spectra, the buffer spectra were subtracted. The following criteria were used to judge the water subtraction: a flat baseline between 2000 to 1700 cm<sup>-1</sup> and no negative lobe between this range (18, 19). The spectra were then multiplied by 100. All spectra were smoothed with a nine-point Savitsky-Golay function to remove the possible noise before further data analysis.

### Partial least-square method (PLS)

The basis of all quantitative analysis using spectroscopy is Beer's law. The multicomponent analyses are based on the additivity of Beer' law, i.e., the absorbance at a specific wavenumber (wavelength) is the sum of the absorbance of all sample components which absorb at that wavenumber (wavelength). If the absorbance at different wavenumbers are collected, the Beer's law can be written in the following equation form:

$$\mathbf{A} = \sum (\mathbf{KLC} + \mathbf{E}_a) \quad (1)$$

where  $\mathbf{A}$  = The vector of absorbances

$\mathbf{K}$  = The matrix absorptivities

$\mathbf{L}$  = The vector of pathlengths

$\mathbf{C}$  = The vector of concentration

$\mathbf{E}_a$  = The matrix of the spectral error

The classic least-square method (CLS) belongs to this kind of method and is most widely used. For CLS method, it is assumed that protein spectra are linear combinations of  $\ell$  pure-structure spectra, i.e.,  $\alpha$ -helix,  $\beta$ -sheet,  $\beta$ -turn and random coil. Suppose that there are  $m$  calibration proteins measured at  $n$  wavenumbers, then:

$$\mathbf{A} = \mathbf{C}\mathbf{K} + \mathbf{E}_a \quad (2)$$

where  $\mathbf{A}$  is an  $m \times n$  matrix of the spectra of the  $m$  calibration proteins,  $\mathbf{C}$  is an  $m \times \ell$  matrix of the concentration,  $\mathbf{K}$  is an  $\ell \times n$  matrix in which rows are the pure-structure spectra.

The  $\mathbf{K}$  matrix can be obtained from the calibration set as follows:

$$\mathbf{K} = \mathbf{A}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-1} \quad (3)$$

where  $\mathbf{C}^T$  is the transpose of matrix  $\mathbf{C}$ , and superscript  $-1$  stands for the inverse of a matrix.

Once  $\mathbf{K}$  matrix is known, the unknown concentration can be calculated as follows:

$$\mathbf{c} = (\mathbf{K}^T\mathbf{K})^{-1}\mathbf{K}^T\mathbf{A} \quad (4)$$

where  $\mathbf{A}$  is the spectrum of the protein to be analyzed.

The CLS method is used widely because the mathematical steps are straightforward, and many standards and wavenumbers can be used in the calibration to obtain an averaging effect (therefore, the CLS method actually is a full-spectrum method). A major disadvantage of the CLS method is that all interfering chemical components in the spectral region of interest need to be known and included in the calibration. This is not an easy task for complex spectra such as protein spectra. In addition, the two matrix inversions required by this method are a major source of errors.

The partial least-squares method (PLS) is a factor analysis method, which has many of the full-spectrum advantages of the CLS method (16). In the PLS method, the calibration spectra can be represented as follows:

$$\mathbf{A} = \mathbf{T}\mathbf{B} + \mathbf{E}_a \quad (5)$$

where  $\mathbf{B}$  is a  $h \times n$  matrix with the rows of  $\mathbf{B}$  being the new PLS basis set of  $h$  full-spectrum vectors, called loading vectors or loading spectra.  $\mathbf{T}$  is an  $m \times h$  matrix of intensities (or scores) in the new coordinate system of the  $h$  PLS loading vectors for the calibration spectra, and  $\mathbf{E}_a$  is an  $m \times n$  matrix of spectral residuals not fitted by the PLS model.

The analogy between the PLS model and the CLS model is quite clear since both equations involve the decomposition of  $\mathbf{A}$  into the product of two smaller matrices. However, rather than the basis vectors being the pure-component spectra in the CLS, they are the loading vectors generated by the PLS algorithm. The intensities in the new coordinate system are no longer the concentrations as they were in CLS, but they are linearly related to the concentrations. The new basis set of full-spectrum loading vectors is composed of linear combinations of the original calibration spectra. The amount (i.e., intensities) of each of the loading vectors which are required to reconstruct each calibration

spectrum are the scores. In general, only a small number of the full-spectrum basis vector are required to represent the calibration spectra (**A**), therefore, the PLS algorithm reduces the number of intensities (*n*) of each spectrum in the spectra matrix **A** to a small number of intensities (*h*) in the new coordinate system of the loading vectors. This data compression step also reduces the noise because noise is distributed throughout all loading vectors while the true spectral variation is generally concentrated in the early loading vectors.

The **c** vector, of size *m*, containing the concentration, can be related to the spectral intensities (**T**) in the new coordinate system by solving the following set of equations:

$$\mathbf{c} = \mathbf{T}\mathbf{v} + \mathbf{e}_c \quad (6)$$

where **v** is the *h*\**1* vector of coefficients relating the scores (intensities) to the conformation fractions and **T** is the matrix of scores (intensities) from the PLS spectral decomposition, **e<sub>c</sub>** is the vector of concentration errors. During calibration, the least-squares solution of eq. 6 is

$$\mathbf{v} = [\mathbf{T}^T\mathbf{T}]^{-1}\mathbf{T}^T\mathbf{c} \quad (7)$$

During prediction, the unknown concentration is obtained by solving the following equation:

$$\mathbf{c} = \mathbf{t}\mathbf{v} \quad (8)$$

where **t** is the vector of size *h* of the intensities of the PLS loading vector in the new coordinate system for the spectrum of the unknown sample.

The advantage of the PLS method is that it is insensitive to the presence of impurities, and makes full use of all spectral data. In addition, this method eliminates the problem of calculating the two matrix inversions in CLS method. Since columns of **T** are orthogonal in the PLS algorithm, the least-squares solution to **v** involves trivial inversion of the diagonal (**T<sup>T</sup>T**) matrix.

For this study, we used the PLSplus program created by Galactic Industries Corporation (Salem, NH) to calibrate and predict the secondary structure of proteins.

The spectral regions we used are amide I (1700-1600 cm<sup>-1</sup>), amide III (1350-1200 cm<sup>-1</sup>), and the combination of amide I and amide III. Before calibration and prediction, the spectra were normalized to total intensity of one in each region. This step helped to remove the protein concentration effect. In order to predict the secondary structure of unknown proteins, a series of calibration spectra whose secondary structures are already known must be set up. After loading these spectra into calibration file, and inputting the contents of each structure, this calibration file is ready to be used to predict the protein with unknown structure. To predict the secondary structure of a protein, the protein is eliminated from the calibration file, and its structure is predicted by using other 8 proteins as the calibration set.

### Prediction of secondary structure contents

Amide I and amide III regions of the infrared spectra of proteins are the two regions which are very sensitive to their secondary structure. For example, Figure 1 shows the spectra of

aqueous solutions of myoglobin and immunoglobulin G, which are two proteins with completely different conformations. Myoglobin is a high  $\alpha$ -helix protein (85% helix) with no  $\beta$ -sheet, and IgG is a high  $\beta$ -sheet protein (67%  $\beta$ -sheet) with 3%  $\alpha$ -helix. These significant differences in shape and frequency render the amide I and amide III region particularly useful for predicting the conformation of proteins in aqueous solutions from their infrared spectra.

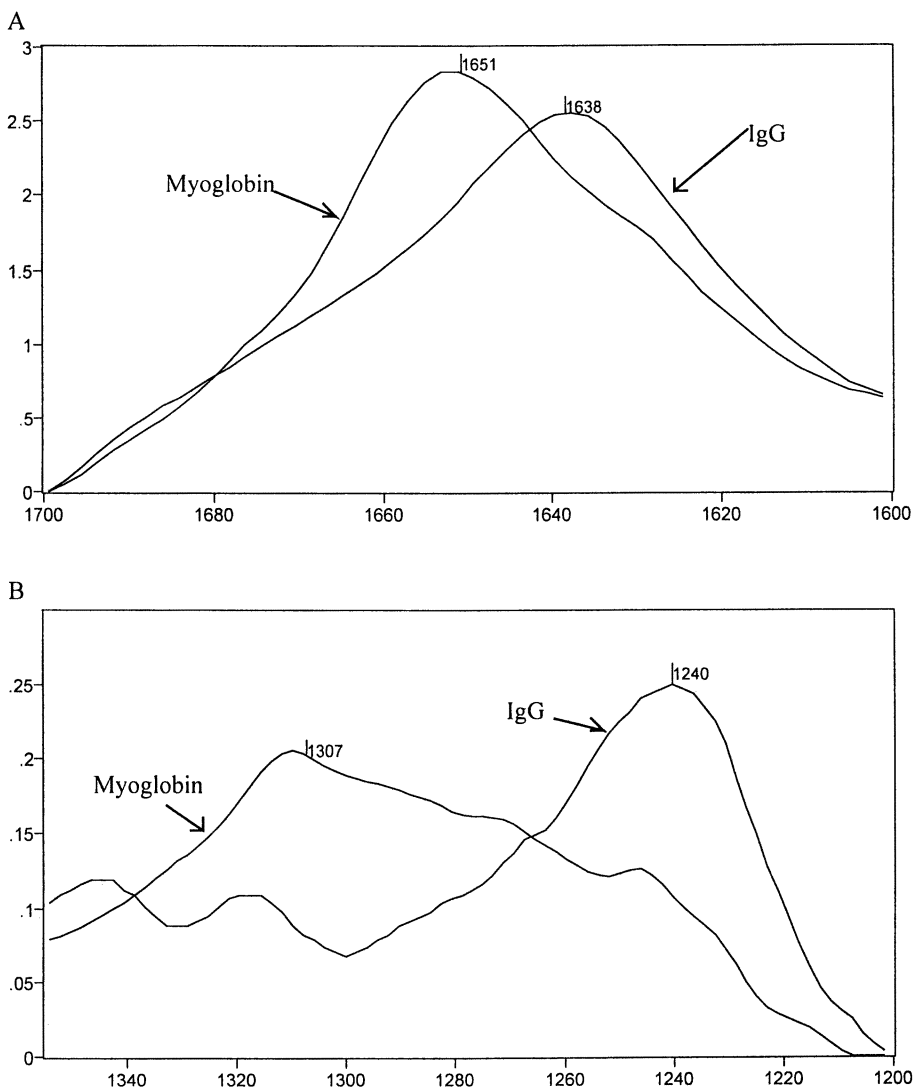


Figure 1. Comparison of FT-IR spectra of  $\alpha$ -helix protein (myoglobin) and  $\beta$ -sheet protein (IgG). A, amide I region and B amide III region.



The PLS algorithm was applied to amide I and amide III spectra to predict the protein secondary structure. The results are listed in Table I, and are compared with the data from the x-ray studies (20).

In order to assess correlation between secondary structure estimation based on amide III IR spectra and secondary structure estimated from x-ray crystallography, we also

**Table I Comparison of Protein Secondary Structures Determined by PLS and X-ray Crystallography**

Proteins	$\alpha$ -Helix	$\beta$ -Sheet	$\beta$ -Turn	Random Coil	Methods
$\alpha$ -Chymotrypsin	8	50	27	15	X-ray (20)
	21	41	23	15	PLS Amide III
	0	63	25	12	PLS Amide I
	2	68	20	10	PLS Amide I & III
Concanavalin A	3	60	22	15	X-ray (20)
	9	56	21	14	PLS Amide III
	3	72	15	14	PLS Amide I
	2	64	20	14	PLS Amide I & III
Cytochrome C	48	10	17	25	X ray (20)
	56	13	15	15	PLS Amide III
	62	14	16	8	PLS Amide I
	61	9	17	13	PLS Amide I & III
Hemoglobin	87	0	7	6	X-ray (20)
	68	8	12	12	PLS Amide III
	69	6	12	13	PLS Amide I
	68	7	11	14	PLS Amide I & III
IgG	3	67	18	12	X-ray (20)
	-8	67	26	15	PLS Amide III
	21	49	21	9	PLS Amide I
	12	54	24	10	PLS Amide I & III
Lysozyme	45	19	23	13	X-ray (20)
	41	23	21	14	PLS Amide III
	64	13	15	18	PLS Amide I
	56	14	15	15	PLS Amide I & III
Myoglobin	85	0	8	7	X-ray (20)
	83	-8	10	15	PLS Amide III
	91	-1	2	8	PLS Amide I
	83	-8	11	14	PLS Amide I & III
Ribonuclease A	23	46	21	10	X-ray (20)
	16	43	25	15	PLS Amide I
	28	35	24	13	PLS Amide I & III
Trypsin	9	56	24	11	X-ray (20)
	-5	67	25	13	PLS Amide III
	8	52	25	15	PLS Amide I
	4	55	23	18	PLS Amide I & III

performed the statistical tests. The protein secondary structures from PLS method were compared with the data from x-ray crystallography by computing the Pearson correlation coefficients (PCC) and standard deviations(S) (21).

$$PCC = (\sum x_i y_i - \sum x_i \sum y_i / N) / \{ [N \sum x_i^2 - (\sum x_i)^2]^{1/2} [N \sum y_i^2 - (\sum y_i)^2]^{1/2} \}$$

$$S = [(\sum (x_i - y_i)^2 / (N-1))]^{1/2}$$

where  $x_i$  is the result from the PLS method, and  $y_i$  is the result from x-ray studies. The statistical test results are listed in the Table II. For a perfect correlation, PCC of 1.0 and S of 0 are expected.

**Table II. Correlation analysis of protein secondary structures between PLS method and x-ray data**

Statistical Parameter	$\alpha$ -Helix	$\beta$ -Sheet	$\beta$ -turn	Random Coil	Spectral Region
PCC	0.943	0.968	0.821	0.412	Amide III
S	0.11	0.068	0.040	0.051	Amide III
PCC	0.930	0.935	0.779	0.240	Amide I
S	.013	0.096	0.051	0.073	Amide I
PCC	0.954	0.939	0.718	0.206	Amide I & III
S	0.10	0.098	0.048	0.066	Amide I & III

As can be derived from Table II, the secondary structure predicted from amide III spectra were better correlated with x-ray crystallographic structures than those obtained from either amide I or amide I plus amide III spectra. For  $\alpha$ -helix, the Pearson correlation coefficient (PCC) between x-ray data and amide III IR spectra was 0.943, and PCC was reduced to 0.930 when using amide I IR spectra. When using amide I plus amide III, the PCC increased to 0.954. For  $\beta$ -sheet, the PCC for amide III estimates was the highest observed with 0.968, while it reduced to 0.935 and 0.939 when using amide I spectra and amide I plus amide III spectra respectively. For  $\beta$ -turn and random coil, the PCCs for amide III spectral estimates indicated the best correlations. The correlation analysis between PLS result and x-ray data also showed that the  $\alpha$ -helix and  $\beta$ -sheet were highly correlated with the x-ray data (PCC above 0.900).  $\beta$ -turn and random coil, especially random coil (PCC was only 0.412), on the other hand, were not well correlated. While  $\beta$ -turn structures are significantly correlated with x-ray crystallographic structures (PCC > 0.602, the critical value of PCC with 95% confidence for 9 samples), the random coil estimation from FT-IR spectra was not well correlated with that from crystallographic structure (PCC < 0.602) (see Figure 2 and Table 2). Low correlation may be due to the narrow range of  $\beta$ -turn and random coil in the calibration proteins. The unclear definition of random coil structures in x-ray crystallographic studies (22, 23) was also contributed to the poor correlation of random coil structures.

## Advantages and Disadvantages of the approach

The PLS method does not need to deconvolve spectra or perform a curve-fitting. As we mentioned in the introduction part, there is a certain degree of subjectivity associated with the methods based on the band narrowing methods. However, since the PLS method is based on the pattern recognition, the pretreatment of the data can be kept to the minimum. Therefore, it is possible to set up a database from the known protein structure, which can be used to predict the unknown structures of proteins. This kind of method has already been successfully used in protein structure analysis by CD spectroscopy (24, 25) and vibrational spectroscopy (17, 26, 27). Since prediction of protein secondary structure both by CD and FT-IR spectroscopy is based on the Beer's law, the chemometric method can be used in both techniques. Moreover, since CD is at its best when it is used for determining  $\alpha$ -helix content, and is less accurate for  $\beta$ -sheet (28), whereas FT-IR spectroscopy is good for both  $\alpha$ -helix and  $\beta$ -sheet prediction, there is good reason to believe that CD and IR spectra should complement one another.

Our results indicate that the amide III region gives a better prediction than either the amide I region or amide I plus amide III region. This is because different secondary structures reflect a larger spectral difference in the amide III region than in the amide I region (see Figure 1). According to equation 1, if two spectra are very similar, the equation will become mortal, and the error of the prediction will become larger. For some special cases, if the two spectra are the same, no prediction can be obtained. Therefore, the more difference the spectra of the species in the system have, the less possibility to get the mortal equation, and therefore, the more accurate result will be obtained. Since amide I region of protein IR spectra has a higher risk to get co-linear spectra for different secondary structures, amide III region is a better region for use in the PLS algorithm.

The interference from the side chain of peptide is a concern in the use of amide III region. However, the results from our previous work (14, 31) indicated that side chain effect is not a major problem for the use of amide III region to predict protein secondary structure. Moreover, since the PLS method can tolerate some unknown species in the system, the side chain effect is not as important as it is in curve fitting method.

ATR sampling technique also raises some concern, since binding of proteins on the ATR crystal may change the protein structure. However, several researchers (12, 13, 32, 33) suggest that binding of protein on the crystal does not change protein structure significantly. The result from our current work also supports this point.

Two assumptions were made for the development of the PLS method: first, the secondary structure of proteins in solution is considered to be the same as that determined by x-ray diffraction in the crystalline form; second, the normalization procedures assumes equal absorptivity of amide bands irrespective of secondary structure variation.

The most definitive comparisons of solution versus crystal structure are provided by multidimensional NMR studies, which can yield solution structures for small proteins (<20 kDa), comparable to those from X-ray diffraction. One review (1) concluded that, in most cases, the solution and crystal structures are very similar, allowing for some differences in the local conformation and dynamics of surface residues. Thus, the practice of using crystal structures to calibrate spectroscopic methods for determining solution structures is generally satisfactory, but anomalous cases may be encountered (9).

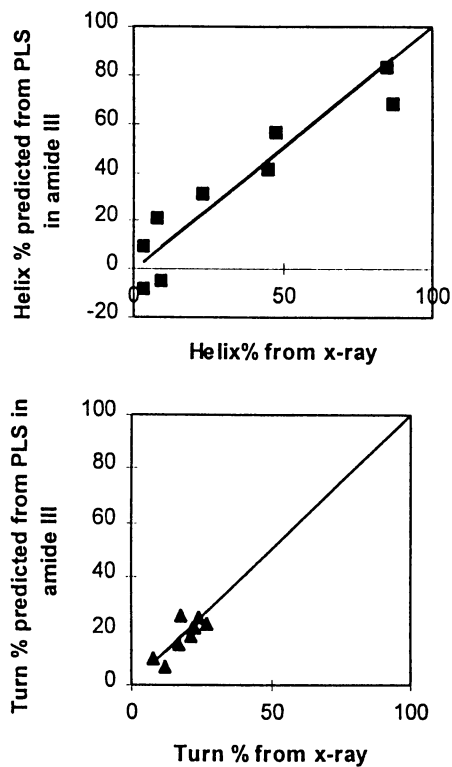
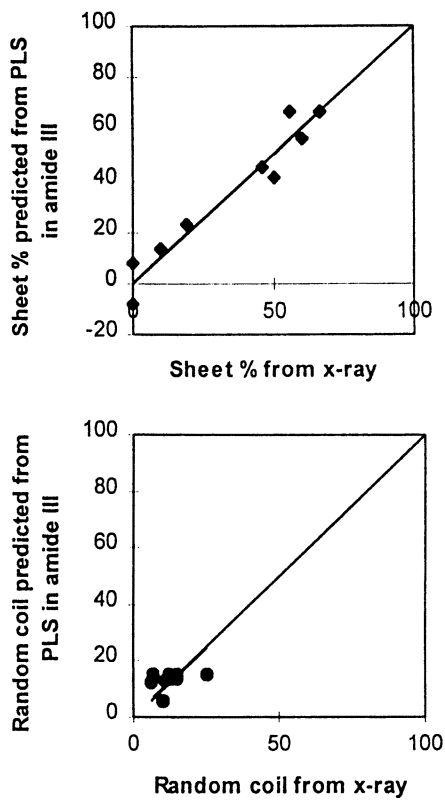


Figure 2. Correlation of secondary structure determined by PLS method from amide III with values from x-ray crystallography determined (20).

Figure 2. *Continued.*

By normalizing the spectra in the amide I and amide III region so that the sum of the absorbances in each region is equal to unity, the assumption was made that the integrated area in each region (amide I or amide III) is nearly constant for different proteins, and therefore, assumed that there are equal molar absorptivities of amide bands regardless of the type of secondary structure (as in the case of band-narrowing analysis). The correlation study done by Susi and Byler (29) suggested that this may be a reasonable assumption. However, studies on poly-lysine have shown that the molar absorptivities of different secondary structures are significantly different (30).

Another problem associated with the PLS method is the design of the calibration set. A statistically significant number of samples is critical in both evaluating the analysis and obtaining good calibration data. Our calibration set does not seem to be adequate. If we have more calibration proteins, we believe the estimation and correlation results will improve. In addition, the technique encounters difficulties in those cases where the spectral properties of the unknown protein lies outside the properties of the spectra within the calibration set. This can be seen clearly from the prediction of random coil content in cytochrome c. The random coil content of cytochrome c from x-ray studies is 25%. When the cytochrome c is removed from the calibration set, the range of random coil contents in the calibration set is between 6% to 15%. Such an observation explains why the prediction of random coil content in cytochrome c is not very good. When the cytochrome c data was discarded and the statistical parameters was calculated again for random coil estimation from the amide III region, the PCC was 0.537 and standard deviation was 3.6%, which was much better than the results obtained when cytochrome c was included in the calibration set (0.412 and 5.1%, respectively for estimates with amide III spectral analysis).

An additional common problem for both the band-narrowing method and the pattern recognition method (such as PLS) is the overlap of amino acid side chains in the amide I and amide III region. However, since the PLS method can predict the number of conformations that is smaller than the total number of conformations in the proteins, the interference due to the noise in the spectra, such as the side chain bands, is less important for the structure prediction in PLS than in curve-fitting method.

## Concluding Remarks

Based on all of our experimental results taken together, we believe that the PLS method is a very promising method for predicting the protein secondary structure, especially using amide III region. It is possible to set up a commercial bank of reference spectra of proteins with known three-dimensional structures and use this approach as a routine method to predict the protein secondary structure as was used in CD spectroscopy.

**Acknowledgments.** This study was in part supported by a grant (NS33740) from National Institutes of Health-National Institute of Neurological Disorders and Strokes.

## References

1. Wagner, G. *Nature Structural Biology* **1997**, *4*, 841.
2. Griebenow, K.; Klibanov, A. M. *J. Am. Chem. Soc.* **1996**, *118*, 11695.

3. Zhang, Y. P.; Lewis, R. N. A. H.; Henry, G. D.; Sykes, B. D.; Hodges, R. S.; McElhaney, R. N. *Biochemistry*, **1995**, *34*, 2348.
4. Englander, S. W.; Downer, N. W.; Teitelbaum, H. *Ann. Rev. Biochem.* **1972**, *41*, 903.
5. Chirgadze, Y. N.; Fedorov, O. V.; Trushina, N. P. *Biopolymers* **1975**, *14*, 679.
6. Anderle, G.; Mendelsohn, R. *Biophys. J.* **1987**, *52*, 69.
7. Kaiden, K.; Matsui, T.; Tanaka, S. *Appl. Spectrosc.* **1987**, *41*, 861.
8. Singh, B. R.; Fuller, M. P.; Schiavo, G. *Biohys. Chem.* **1990**, *46*, 155.
9. Singh, B. R.; Fu, F.-N.; Ledoux, D. N. *Nature Structural Biology* **1994**, *1*, 358.
10. Griebenow, K.; Klibanov, A. M. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 10966.
11. Costantino, H. R.; Griebenow, K.; Mishra, P.; Langer, R.; Klibanov, A. M. *Biochim. Biophys. Acta*, **1995**, *1253*, 69.
12. Bramanti, E.; Benedetti, E. *Biopolymers* **1996**, *38*, 639.
13. Bramanti, E.; Benedetti, E.; Sagripanti, A.; Papineschi, F.; Benedetti, E. *Biopolymers* **1997**, *41*, 545.
14. Fu, F.-N.; DeOliveira, D. B.; Trumble, W.; Sarkar, H. K.; Singh, B. R. *Appl. Spectrosc.* **1994**, *48*, 1432.
15. Lober, A.; Wangen, L. E.; Kowalski, B. R. *J. Chemom.* **1987**, *1*, 19.
16. Haaland, D. M.; Thomas, E. V. *Anal. Chem.*, **1988**, *60*, 1193.
17. Dosseau, F.; Pezolet, M. *Biochemistry*, **1990**, *29*, 8771.
18. Haris, P. I.; Lee, D. C.; Chapman, D. *Biochim. Biophys. Acta* **1986** *874*, 255.
19. Mitchell, R. C.; Haris, P. I.; Fallowfield, C.; Keeling, D. J.; Chapman, D. *Biochim. Biophys. Acta* **1988** *941*, 31.
20. Levitt, M.; Greer, J. *J. Mol. Biol.* **1977**, *114*, 181
21. Kalnin, N. N.; Baikalov, I. A.; Venyaminov, S. YU *Biopolymers*, **1990**, *30*, 1273.
22. Fetrow, J. S.; Zehfus, M. S.; Rose, G. D. *Biothenology*, **1988**, *6*, 167.
23. Lesczynski, J. F.; Rose, R. D. *Science*, **1986**, *234*, 849.
24. Yang, J. T.; Wu, C. S.; Martinez, H. M. *Methods Enzymol.*, **1986**, *130*, 208.
25. Manavalan, P.; Johnson, W. C., Jr. *Anal. Biochem.*, **1987**, *167*, 76.
26. Lee, D. C.; Haris, P. I.; Chapman, D.; Mitchell, R. C. *Biochemistry*, **1990**, *29*, 9185.
27. Compton, L. A.; Johnson, W. C. *Anal. Biochem.*, **1986**, *155*, 155.
28. Woody, R. *Methods Enzymol.*, **1996**, *246*, 34.
29. Susi, H.; Byler, D. M. *Methods Enzymol.*, **1986**, *130*, 290.
30. Jackson, M.; Haris, P. I.; Chapman, D. *J. Mol. Struct.*, **1989**, *214*, 329.
31. Cai, S., Singh, B. R., *Biophys. Chem.*, **1999**, (in press).
32. Jakobsen, R. J.; Wasacz, F. M. *Appl. Spectrosc.* **1990**, *44*, 1478.
33. Bocheva, M.; Vogel, H. *Biophys. J.*, **1997**, *73*, 1056.

## Chapter 6

# Determination of Secondary Structure in Protein Aggregates Using Attenuated Total Reflectance FTIR

A. L. Fink, S. Seshadri, R. Khurana, and K. A. Oberg

Department of Chemistry and Biochemistry, University of California,  
Santa Cruz, CA 95064

Protein aggregation is a major problem in many fields. Currently the mechanism of formation and the structure of protein aggregates are poorly understood. The insoluble nature of protein aggregates limits the number of techniques which can be used to ascertain the conformation of the aggregated protein. Attenuated total reflectance (ATR) FTIR is one method which can readily provide information about the secondary structure content of both soluble and insoluble proteins. ATR-FTIR was used to examine the structure of inclusion bodies, folding aggregates, amorphous precipitates and amyloid fibrils. A common feature of the aggregated proteins is the presence of additional  $\beta$ -structure compared to the native conformation.

Protein aggregates are a major problem, not only for many protein chemists performing basic research, but the aggregation of proteins can present significant technical and economic problems in the biotechnology and pharmaceutical industries, and lead to lethal and debilitating situations when present in the body in the form of a protein deposition disease.

The intrinsic insoluble nature of protein deposits (as in amorphous aggregates, inclusion bodies, amyloid fibrils, etc.) places severe restrictions on the availability of methods for ascertaining the structure of the material. FTIR spectroscopy, especially in the attenuated total reflectance (ATR) mode, is well suited for determining structural features of proteins. Proteins in the form of solutions, thin films (hydrated or dry, from solutions or precipitates), solids (including lyophilized or spray-dried powders), or suspensions of precipitates (e. g. inclusion bodies, amyloid fibrils), can be used for ATR-FTIR analysis.

After a brief discussion of some of the key features of protein aggregation, this chapter focuses on the use of ATR-FTIR in the determination of the structure of

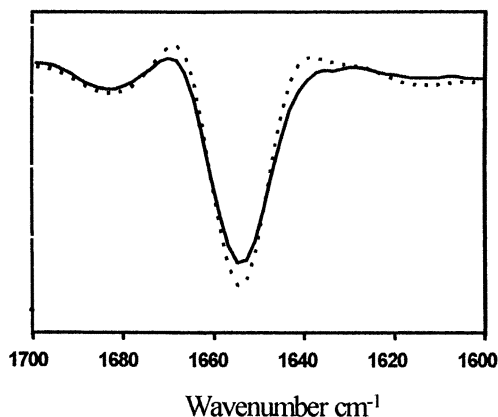


protein aggregates, and in investigations to unravel the underlying molecular mechanisms of protein aggregation.

### Protein Aggregation

Protein aggregation is now recognized as a major problem in many fields and has become the focus of increasing research efforts (1-5). Substantial evidence is accumulating to support the hypothesis that the key precursors of protein aggregation are partially-folded intermediates, which may arise either during the folding of newly synthesized proteins, as with inclusion bodies, for example, or from the native state, as appears likely for at least some extracellular amyloid deposits (1,5-14). Circumstances that lead to the population of partially-folded intermediates, especially if their concentration is high, are thus likely to lead to aggregation; these include mutations, or environmental conditions, which produce differential destabilization of the native state relative to the partially-folded intermediate. Furthermore, the characteristics and properties of the intermediates may be significantly different from those of the native (and unfolded) conformation (15).

Protein aggregation is conveniently classified into ordered and disordered deposits. Amyloid fibrils (both *in vivo* and *in vitro*) are examples of ordered aggregates, whereas inclusion bodies are examples of *in vivo* disordered aggregates. Corresponding disordered *in vitro* aggregates are folding aggregates, formed during the refolding of denaturant-unfolded protein at high protein concentrations, or under weakly native conditions at high protein concentration. Native, folded proteins may aggregate under certain conditions, most notably salting out and isoelectric precipitation. Such precipitates of native protein are readily distinguished from “pathological” aggregates by their solubility in buffer under native-like conditions. In contrast, “pathological” aggregates dissolve/dissociate only in the presence of high concentration of denaturant or detergent. Using ATR-FTIR, we have shown that the native conformation is retained in “salting out” precipitates (Figure 1).



**Figure 1.** Second derivative spectrum of the amide I region of interleukin-2: the solid line is for the native protein the dotted line represents the spectrum for ammonium-sulfate-precipitated interleukin-2.

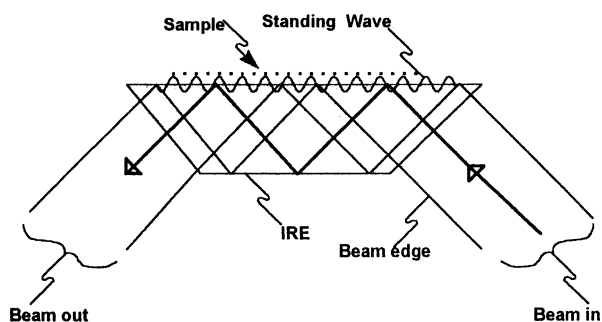
Inclusion bodies and other aggregates formed during protein folding have been assumed to arise from hydrophobic interactions of the unfolded or denatured states, whereas amyloid fibrils and other extracellular aggregates have been assumed to arise from native-like conformations in a process analogous to the polymerization of hemoglobin S (3). However, if aggregation arises from specific partially-folded intermediates, then one would expect that aggregation will be favored by factors and conditions that favor population of these intermediates. Indeed, this is usually found to be the case. Furthermore, the characteristics and properties of the intermediates may be significantly different from those of the native and unfolded conformations.

There are many unanswered questions relating to protein aggregation, including details of the mechanism of aggregation and the underlying kinetics, the structure of the aggregates, the factors which discriminate between ordered aggregates (amyloid) and disordered aggregates (inclusion bodies, folding aggregates, amorphous deposits), the nature of the specific intermolecular interactions, and how aggregation may be prevented.

**Problems Due To Protein Aggregation.** Several dozen protein deposition diseases are known (16). The most familiar include the amyloid diseases such as Alzheimer's disease, and prion diseases such as bovine spongiform encephalopathy (BSE - Mad Cow disease), and Creutzfeldt-Jacob disease (CJD) in humans). In both amyloid and prion diseases the aggregated protein is usually in the form of ordered fibrils. Amyloid fibril formation has been observed to arise from both peptides and proteins. Several protein deposition diseases involve non-ordered protein deposits, such as light-chain deposition disease and cataracts. A number of neurological diseases, e. g. Parkinson's disease, involve deposited proteins in the form of "inclusions". Classical inclusion body formation is very common when proteins are overexpressed, especially in bacteria; these inclusion bodies are usually highly homogeneous. Protein aggregation is also a problem in the storage or delivery of protein drugs, and in the lyophilization and rehydration of pharmaceutical proteins.

### Attenuated Total Reflectance (ATR) FTIR

In ATR-FTIR, the sample is placed in contact with the surface of material having a high refractive index, known as the internal reflection element (IRE). The IRE is usually made of germanium or zinc selenide. When infrared radiation penetrates the IRE at an angle beyond the critical angle of incidence, total internal reflection of this incident radiation produces an evanescent (standing) wave at the boundary between the IRE and the sample (Figure 2). This reflected beam penetrates beyond the crystal, and can be absorbed by materials in contact with its surface. The depth of penetration varies, depending on the material of the IRE and the wavelength. Since the IR beam is not transmitted through the sample, the spectra are unaffected by turbidity. Several different ATR cell designs are available; we favor out-of-compartment, horizontal trapezoidal-shaped IREs, which may be used in either a trough or flow-cell configuration. ATR may be used with samples in solution (17), as a thin-film (starting with either solution or suspension), as a suspension or in the dry solid state. Interactions between the protein and the IRE surface (which may perturb the protein conformation) are minimal with suspensions of aggregated proteins.



**Figure 2.** Schematic outline of a trapezoidal internal reflection element.

ATR has been used to analyze spectra of soluble and insoluble proteins and peptides (including inclusion bodies and fibrils) and for solid (powdered) proteins/peptides (13,18,19). Details of the ATR method have been published (13,17,20,21).

The simplest way in which to study samples of aggregated proteins is as a thin film deposited from a suspension. There are few or no interfering bands from either solvents or dispersion media and the limited sample preparation avoids conditions which could affect the conformation of the material. Typically 50 to 100  $\mu\text{l}$  of 0.5 to 1 mg/ml solution or suspension (i.e. 10-200  $\mu\text{g}$  of protein) is placed on the IRE and the sample is dried down using nitrogen gas or dry air or a vacuum. The thin films have been shown to maintain protein molecules in a hydrated state and to conserve their 3D structure (17,22,23). The limited penetration depth in thin film ATR is a substantial advantage in that it minimizes the contribution of any liquid water present. ATR samples containing materials of high refractive index may result in significant shifts in the observed band positions, making comparisons between samples difficult.

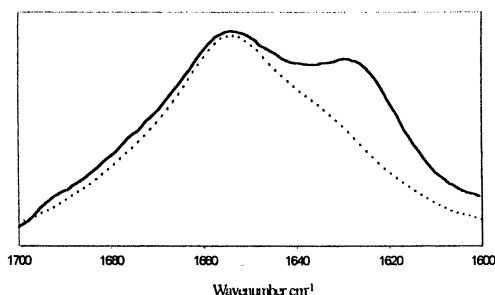
Our analysis is usually focused on the amide I region (1600-1700  $\text{cm}^{-1}$  region, corresponding to C=O stretch), which gives information about the secondary structure (17,24), although we also have used the amide II and III regions (13,17). Quantitative analysis of secondary structure is determined either from assignment of deconvoluted bands (following second derivative and Fourier self-deconvolution) or from partial least-squares multivariate analysis (17). The fine structure in the amide I band is most readily seen in the form of Fourier self-deconvoluted or second derivative spectra, or curve-fit component bands.

**Data Analysis.** Two general methods have been used to analyze the IR spectra of proteins to ascertain their secondary structure: deconvolution (resolution enhancement) and assignment of the component bands, or factor analysis, using pattern recognition methods. The fundamental difficulty encountered in the analysis of the amide I spectra arises from the fact that the widths of these component bands are usually greater than the separation between the maxima of adjacent peaks.

The preferred method to deconvolute the amide I region is as follows: first both FSD and second derivative spectra are run to determine the positions of the components. A robust analysis can only be assured if both methods give identical peak locations. Curve fitting can be done using the band positions derived from the FSD and second derivative spectra to fit the raw spectrum to a combination of Gaussian-Lorentzian peaks. The area under each peak is then used to compute the percentage of the individual component contributing to the amide I region; it is assumed that the extinction coefficients are identical for each type of secondary structure, and consequently that component peak area is directly proportional to the fraction of secondary structure accounting for that component.

### Partially-folded Intermediates As Aggregation Precursors

Most partially-folded intermediates have a strong propensity to aggregate (25). In studies on partially-folded intermediates we have observed that their aggregation is usually accompanied by an increase in  $\beta$ -structure. For example, the helical apomyoglobin forms three distinct ensembles of partially-folded intermediates at low pH, A<sub>1</sub>, A<sub>2</sub>, and A<sub>3</sub> (15). The least structured of these intermediates shows half as much  $\alpha$ -helix as the native state, and aggregates at protein concentrations above  $\sim 0.25$  mg/ml as determined by dynamic light and small-angle X-ray scattering. Thin-film ATR-FTIR spectra of the soluble aggregated intermediate show increasing amounts of  $\beta$ -structure as the protein concentration is increased (Figure 3). At higher protein concentrations the intermediate precipitates; the precipitated material also shows a large  $\beta$ -sheet component in the ATR-FTIR spectrum. The data clearly demonstrate that as the concentration of partially-folded intermediates increases the intermediates associate, initially to form soluble aggregates, and at higher protein concentrations to form insoluble precipitates, in which the aggregated material is substantially enriched in  $\beta$ -sheet or extended chain conformation.



**Figure 3.** Increasing protein concentration leads to aggregation of an apomyoglobin partially-folded intermediate: the dotted line is the amide I spectrum of the monomeric intermediate, the solid line corresponds to the soluble aggregate; note the substantial increase in  $\beta$ -structure shown by the shoulder at  $1627\text{ cm}^{-1}$ .

## Aggregation Can Induce Secondary Structure Into Partially-Folded Intermediates

Recently, in studies on partially-folded intermediates of staphylococcal nuclease (SNase) we have discovered a rather interesting phenomenon, namely that aggregation of a partially-folded intermediate may induce substantial additional secondary structure in the protein (26-27). Staphylococcal nuclease, like many proteins (25,28), forms partially-folded intermediates at low pH and low ionic strength (29). Under conditions of pH < 3.5, 100 mM sodium sulfate or 0.5 M KCl, and low protein concentrations (<0.2 mg/ml) SNase is monomeric, and approximately 50% folded, based on the amount of secondary structure and the radius of gyration from small-angle X-ray scattering (SAXS) studies. Increasing the protein concentration leads to formation of dimers, with a significant increase in secondary structure (corresponding to about 70 % of that of the native state) (27). Further increases in protein concentration induces a second transition leading to formation of larger (multimeric) soluble aggregates with more secondary structure. Above 4 mg/ml protein, the far-UV CD spectrum is almost identical to that of the native protein, although there are some differences when analyzed by ATR-FTIR (30). However, this apparently native-like conformation has a near-UV CD spectrum close to that of the unfolded molecule. Thus the increase in concentration of the partially-folded intermediate induces the appearance of new internal structure, as manifested by both increased secondary structure and the globular core. The stable dimers, in particular, are consistent with an association mechanism in which domain swapping (31) is involved.

### ATR-FTIR Investigations Of Inclusion Bodies And Folding Aggregates

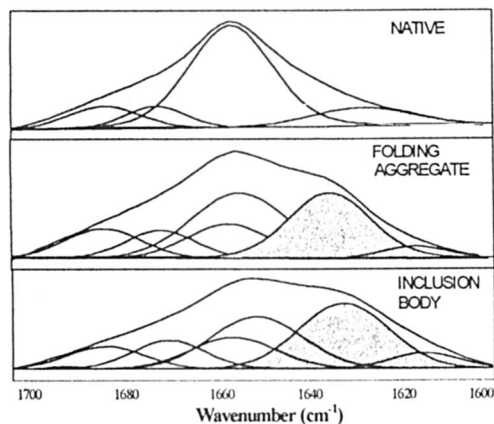
Most proteins, when refolded from a denatured state by dilution of the denaturant, will aggregate and form insoluble precipitates if the protein is above some threshold concentration. This concentration varies from nM to mM depending on the particular protein. These aggregates, folding or refolding aggregates, appear to be the *in vitro* equivalent of inclusion bodies formed *in vivo*. Similar aggregates also occur for many proteins when they are brought under destabilizing conditions, e. g. to a temperature close to that at the beginning of the thermal denaturation transition, or a concentration of denaturant close to the beginning of the unfolding transition (32).

We have used ATR-FTIR to examine the structure of proteins in various aggregated forms, including inclusion bodies and folding aggregates (1,13,17). The samples included representative all- $\beta$ , all- $\alpha$  and  $\alpha$ + $\beta$  proteins. Typical ATR spectral data, for the helical protein interleukin-2, are shown in Figure 4 and Table I.

The overall results are summarized in Table II. Several common features were observed: both the inclusion bodies and the folding aggregates exhibited substantial secondary structure, typically between 50 and 70% of that of the native conformation. We interpret this to mean that the inclusion bodies arose from the association of partially-folded intermediates containing substantial native-like structure. The structure of a given protein in inclusion bodies, refolding aggregates, and thermal aggregates (formed by heating a concentrated protein solution to a temperature just below the beginning of the thermal denaturation transition) is the same.

**Table I.** Secondary structure content of interleukin-2 determined by ATR-FTIR in different conformations. The peak position and percentage of total area are given.

Secondary Structure	Native (FT-IR)	Native (X-ray)	Inclusion Bodies	Refolding Aggregates
$\alpha$	1655 59%	55%	1655 16%	1655 15%
$\beta$	1638 9%	11%	1630 34%	1633 28%
	1626 12%			
Turns	1681 10.5%		1681 9.8%	1681 12%
	1670 8.5%		1668 12%	1669 10%
	1693 1%		1693 0.5%	1693 1%
Disordered			1650 27%	1652 33%



**Figure 4** The amide I region of IL-2 in the native, inclusion body and folding aggregate states, showing the curve-fit components. The bands corresponding to  $\beta$ -structure are shaded.

Thus we conclude that a given protein will have one partially-folded intermediate that is particularly prone to aggregate, and consequently most aggregates of that protein will arise from that intermediate and thus have similar structures.

Further, in all cases, even for all- $\beta$  proteins, significant new  $\beta$ -structure, compared to that in the native conformation, was observed: typically this amounted to about 20-25% of the total secondary structure. We take this to indicate that the nature of the intermolecular interactions leading to the aggregation involve  $\beta$ -sheet-like interactions. The exact nature of the intermolecular interactions is unknown, and could be different in different aggregates. However, what is clear is that aggregation

**Table II.** Changes in secondary structure of inclusion bodies relative to the corresponding native state.

<i>Protein</i>	<i>% Native secondary structure</i>	<i>% increase in non-native <math>\beta</math></i>	<i>% increase in disordered structure</i>
Interleukin-2	57	22	27
Interferon- $\gamma$	65	25	19
Haptoglobin	73	22	6
Apomyoglobin	35	26	26
Interleukin-1 $\beta$	72	26	15

usually leads to an increase in the amount of secondary structure with IR components in the 1623-1637  $\text{cm}^{-1}$  region, which corresponds to the region where  $\beta$ -sheet structure is observed. Interestingly the amount of secondary structure in the inclusion body varies from one protein to another, as does the amount of disordered structure.

Three possible mechanisms for inclusion body formation have been proposed: aggregation of native protein of limited solubility; aggregation of the unfolded state; and aggregation of partially-folded intermediate states. The results reported here indicate that substantial native-like structure is present in inclusion bodies and folding aggregates, and strongly implies that inclusion bodies arise from the association of partially-folded intermediates formed relatively late in the folding process. This is in contrast to the prevailing view that inclusion body formation is an aggregation process mediated by non-specific interactions in the unfolded state (5).

### ATR-FTIR Investigations Of Amyloid Fibrils

A number of human diseases involve the deposition of protein aggregates; a subset of these, known as amyloidoses, are lethal diseases involving the extracellular deposition of amyloid fibrils and plaques. Amyloid fibrils are customarily defined by three features: characteristic birefringent staining by Congo Red, fibrous morphology by electron microscopy and a distinctive X-ray fiber diffraction pattern. Amyloid fibrils typically are 7-12 nm in diameter, and can be dissociated by high concentrations of denaturant. The molecular mechanisms leading to amyloid formation are not well understood. It has long been known that one of the structural characteristics of amyloid fibrils is the presence of  $\beta$ -sheet structure. A recently proposed model by Blake and Serpell (33) for the structure of amyloid fibrils from patients with familial amyloidotic polyneuropathy (FAP), which are derived from transthyretin (TTR) variants, nicely accounts for the fiber diffraction patterns in aligned fibrils. Depending on the protein or peptide, four or five protofilaments intertwine to form a fibril with the characteristic diameter of  $\sim 10$  nm. The proposed structure of the fibril indicates that the TTR building blocks are different from the native tetramer conformation (33).

The ATR-FTIR spectra of amyloid fibrils are not significantly different from those of disordered aggregates, in that both show increased intensity in low frequency components corresponding to  $\beta$ -structure. Interestingly, it has been suggested (34) that

the  $\beta$ -sheet in some amyloid fibrils may be in the form of a  $\beta$ -helix, the structure found in pectate lyase, acyltransferase and the P22 tailspike protein. The structures of several amyloidogenic proteins and peptides can be plausibly modeled by  $\beta$ -helices to form fibrils with the observed structural properties. The IR bands of  $\beta$ -helices are found in the same regions of the amide I band as other forms of  $\beta$ -sheet ( $1623\text{--}1636\text{ cm}^{-1}$ ) (R. Khurana and A. L. Fink, unpublished observations).

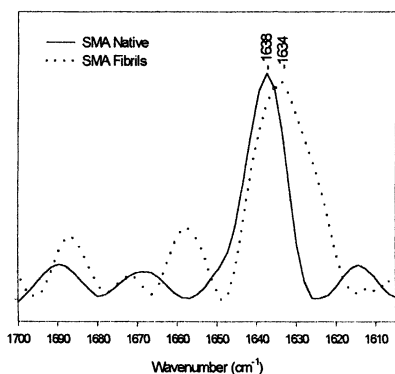
AL amyloidosis is the most common form of systemic amyloidosis, and is one of four common immunoglobulin light-chain-related deposition diseases in which the amyloid fibrils are derived from the light-chain variable domain of immunoglobulins ( $V_L$ ). The immunoglobulin light chain consists of two domains, the variable ( $V_L$ ) (N-terminal) and the constant ( $C_L$ )(C-terminal), each with a MW of  $\approx 12,000$ . Each domain has the classical immunoglobulin fold, a  $\beta$ -sandwich consisting of two  $\beta$ -sheets connected by a buried disulfide bond. The amyloidogenic  $V_L$  domain, SMA, has been cloned and expressed in *E. coli* (35). The properties of AL amyloidosis fibrils are similar to those of other amyloidoses, and it is believed that many of the factors involved in amyloid formation are common to all the amyloid diseases (36).

Characterization of the light chain domain SMA by circular dichroism, tryptophan fluorescence, ANS binding and ATR-FTIR reveals the population of a partially-folded intermediate in the pH 4-6 range, whose structure is relatively native-like. Solutions of SMA readily formed aggregates under conditions in which the partially-folded intermediate was populated. Under typical conditions it takes about one week for fibrils to form, although the lag period can be eliminated by seeding with existing crystals. Interestingly, under most aggregating conditions amorphous aggregates formed initially and then slowly converted (directly or indirectly) into fibrils.

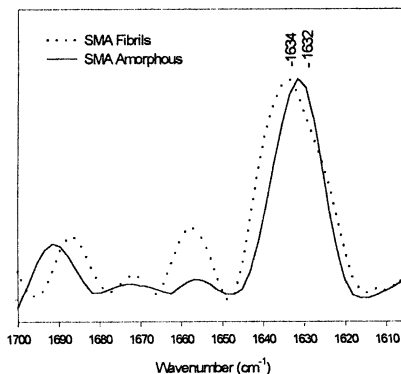
ATR-FTIR was used to characterize the various states of SMA. Figure 5 shows a comparison of the amide I region of SMA in the native and fibrillar states: both spectra are dominated by a  $\beta$ -structure band around  $1630\text{--}1640\text{ cm}^{-1}$ . Noteworthy points are: a shift to lower frequency in the beta components in the fibrils, consistent with our observations on other aggregated proteins and fibrils; the  $1634\text{ cm}^{-1}$  band of the fibrils is broad, suggesting an underlying distribution of closely related conformations; the  $1658\text{ cm}^{-1}$  band in the fibrils corresponds to either disordered or loop structure, which is absent in the native state. The fibrils have about 8 % more  $\beta$ -structure than the native state.

Comparison of the ATR-FTIR spectra of amorphous aggregates of SMA with that of fibrils (Figure 6) shows them to be quite similar, with the exception of a small downfield shift in the  $\beta$  band for the amorphous aggregates, and less of the disordered/loop component. The amorphous aggregate was obtained by incubating SMA at pH 5.6, 150 mM NaCl,  $37^\circ\text{C}$  at 2 mg/ml for one day. The partially-folded intermediate, at pH 5.6, has a spectrum quite distinct from that of the native and aggregated conformations (Figure 7), exhibiting two broad beta components, one of which, at  $1634\text{ cm}^{-1}$ , corresponds to a component in the fibrils.





**Figure 5.** FTIR amide I region, second-derivative spectra (inverted) of native SMA (solid line) and fibrils at pH 5.6 (dotted line).

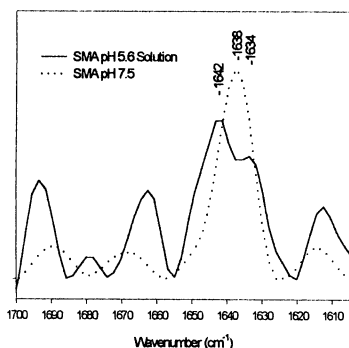


**Figure 6.** Second-derivative FTIR spectra (inverted) of amorphous SMA (solid line) and fibrils (dotted), both at pH 5.6, 37°C.

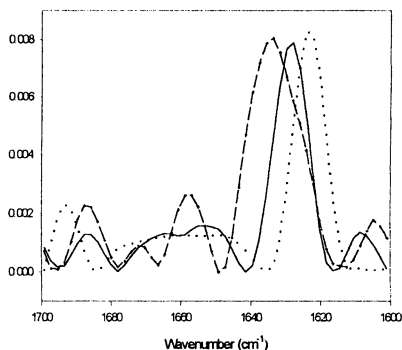
**Do all amyloid fibrils have similar  $\beta$ -structure?** We have also used ATR-FTIR to examine the conformation of fibrils from Alzheimer's A $\beta$ 40 peptides and a small peptide (KLEG (37)). Interestingly, a comparison of the FTIR spectra of fibrils from A $\beta$ 40, SMA and KLEG illustrate the diversity of detailed structure in different fibrils (Figure 8).

### Models For The Mechanism Of Protein Aggregation

Based on our investigations of aggregation in inclusion bodies, folding aggregates and amyloid fibrils, our working model for aggregation and fibril formation is as follows: conditions leading to increased population of the key partially-folded intermediate lead to intermolecular interactions in which one part of one molecule forms a  $\beta$ -strand with part of another molecule. Whether these are existing  $\beta$ -strands, disordered polypeptide chain or some other type of secondary structure is currently unknown. It is likely that the new 2-stranded  $\beta$ -sheet formed by this intermolecular interaction is antiparallel, based on our FTIR analysis, but the interpretation is not unambiguous. The driving force for the intermolecular interaction (aggregation) is undoubtedly predominantly hydrophobic in nature, but may also involve key electrostatic interactions. This is especially likely in view of the significant pH-dependence of most protein aggregation reactions. It is also most likely that in many cases aggregation is specific in the sense that the intermolecular interactions are quite particular and not random i. e. only certain regions of the molecule interact with each other. We have proposed a model to account for such specificity (1). It is also possible that aggregation may often involve "domain" swapping (31); a particularly intriguing notion is that a limited number of strands (e. g. a single hairpin) in one  $\beta$ -sheet may switch places with those in adjacent molecules, thus forming a tightly linked aggregate.



**Figure 7.** Second derivative FTIR (inverted) spectrum of SMA under native conditions, pH 7.5, and as the partially-folded intermediate, pH 5.6.



**Figure 8.** Second-derivative FTIR spectra (inverted) of fibrils from the light chain domain SMA (dashed line), A $\beta$ 1-40 (solid) and a short peptide KLEG (37) (dotted).

It is very interesting to note our finding that the structural changes on formation of amorphous aggregates such as inclusion bodies and folding aggregates involve a significant increase in  $\beta$ -structure. Thus both amyloid fibrils and amorphous aggregates share the common structural feature of substantially increased  $\beta$ -structure, especially in the form of lower frequency amide I infrared components compared to those in the native conformation (a decrease in amide I frequency is usually taken to indicate stronger H-bonding, and thus more ideal alignment of the  $\beta$ -strands). This raises the possibility that there might be local order in the apparently amorphous aggregates of inclusion bodies and folding aggregates. For example, it is possible that the amorphous aggregates actually involve local  $\beta$ -sheet formation between the associating molecules, but the ordered structure does not extend beyond the immediate vicinity.

### Concluding remarks

ATR-FTIR is one of the most readily applied techniques for ascertaining structural information about aggregated proteins. Its advantages include the small amounts of material needed, and the relative ease of data acquisition and analysis. The technique is quite sensitive to small changes in structure, and comparisons can be made between similar samples in solution, suspension or solid forms. As with other modes of FTIR analysis of protein structure, the assignment of secondary structure is not always unambiguous. ATR-FTIR has contributed significant new information about the structure of inclusion bodies, folding aggregates, amorphous precipitates and amyloid fibrils. A common feature of all these aggregated proteins is the presence of additional  $\beta$ -structure compared to the native conformation, as well as substantial native-like secondary structure, supporting the hypothesis that the inclusion bodies arose from the association of partially-folded intermediates. We anticipate that the

advantages of ATR-FTIR in studying aggregated proteins will lead to increasing numbers of investigations using this technique.

### Acknowledgments

The research reported herein was supported by grants from the National Science Foundation to ALF. We thank the following for supplying material used in these studies: T. Arakawa, D. Dowling, F. Stevens, and D. Teplow.

### Literature cited

1. Fink, A. L. *Folding & Design* **1998**, *3*, R9-15.
2. DeYoung, L. R.; Fink, A. L.; Dill, K. A. *Acc. Chem. Res.* **1993**, *26*, 614-620.
3. Wetzel, R. *Trends Biotechnol* **1994**, *12*, 193-8.
4. Jaenicke, R. *Phil. Trans. Roy. Soc. Lond. Series B: Biological Sciences* **1995**, *348*, 97-105.
5. Wetzel, R. *Cell* **1996**, *86*, 699-702.
6. London, J.; Skrzynia, C.; Goldberg, M. E. *Eur. J. Biochem.* **1974**, *47*, 409-15.
7. Mitraki, A.; King, J. *Bio/Technology* **1989**, *7*, 690-697.
8. King, J.; Haase-Pettingell, C.; Robinson, A. S.; Speed, M.; Mitraki, A. *FASEB J* **1996**, *10*, 57-66.
9. Speed, M. A.; Wang, D. I.; King, J. *Protein Sci* **1995**, *4*, 900-8.
10. Wetzel, R. *Protein Engineering : A Practical Approach*; Rees, A. R.; Sternberg, M. J. E., Eds.; IRL Press at Oxford University Press: Oxford. New York, 1992; pp 191-219.
11. Chan, W.; Helms, L. R.; Brooks, I.; Lee, G.; Ngola, S.; McNulty, D.; Maleeff, B.; Hensley, P.; Wetzel, R. *Folding & Design* **1996**, *1*, 77-89.
12. Kelly, J. W. *Structure* **1997**, *5*, 595-600.
13. Oberg, K.; Chrnyk, B. A.; Wetzel, R.; Fink, A. L. *Biochemistry* **1994**, *33*, 2628-2634.
14. Kim, D.; Yu, M. H. *Biochem. Biophys. Res. Comm.* **1996**, *226*, 378-84.
15. Fink, A. L.; Oberg, K. A.; Seshadri, S. *Folding & Design* **1997**, *3*, 19-25.
16. Thomas, P. J.; Qu, B. H.; Pedersen, P. L. *Trends Biochem Sci* **1995**, *20*, 456-9.
17. Oberg, K. A.; Fink, A. L. *Anal. Biochem.* **1998**, *256*, 92-106.
18. Bauer H. H. ; Muller, M.; Goette, J.; Merkle H. P. ; Fringeli U. P. *Biochemistry* **1994**, *33*, 12276-82.
19. Dong, A. C.; Prestrelski, S. J.; Allison, S. D.; Carpenter, J. F. *J. Pharm. Sci.* **1995**, *84*, 415-424.
20. Seshadri, S.; Khurana, R.; Fink, A. L. *Methods Enzymol.* **1998**, *in press*.
21. Oberg, K. A.; Fink, A. L. in *Techniques in Protein Chemistry*; W. Crabb, Ed.; Academic Press, Inc. 1995; Vol. VI, pp 475-484.
22. Goormaghtigh, E.; Ruyschaert, J.-M. *Molecular Description of Biological Membranes by Computer Aided Conformational Analysis*, Vol. 1 ed.; Ed. Brasseur, R., Editor; CRC Press: Boca Raton, Ann Arbor, Boston, 1990; pp 285-329.
23. Safar, J.; Roller, P. P.; Ruben, G. C.; Gajdusek, D. C.; Gibbs, C. J. *Biopolymers* **1993**, *33*, 1461-1476.
24. Byler, D. M.; and Susi, H. *Biopolymers* **1986**, *25*, 469-487. 12336.
25. Fink, A. L. *Ann. Rev. Biophys. Biomol. Struct.*, **1995**, *24*, 495-522.
26. Uversky, V. N.; Karnoup, A. S.; Segel, D. J.; Seshadri, S.; Doniach, S.; Fink, A. L. *J. Mol. Biol.* **1998**, *278*, 879-894.

27. Uversky, V. N.; Karnoup, A. S.; Segel, D. J.; Doniach, S.; Fink, A. L. *Proc. Nat. Acad. Sci. US* **1998**, *95*, 5480-5483.
28. Fink, A. L.; Calciano, L. J.; Goto, Y.; Kurotsu, T.; Palleros, D. R. *Biochemistry* **1994**, *33*, 12504-12511.
29. Fink, A. L.; Calciano, L. J.; Goto, Y.; Nishimura, M.; Swedberg, S. A. *Protein Science* **1993**, *2*, 1155-1160.
30. Uversky, V. N.; Karnoup, A. S.; Khurana, R.; Segel, D. J.; Doniach, S.; Fink, A. L. *Prot. Science* **1999**, *8*, 161-173.
31. Bennett, M. J.; Schlunegger, M. P.; Eisenberg, D. *Protein Sci* **1995**, *4*, 2455-68.
32. Raman, B.; Ramakrishna, T.; Rao, C. M. *J. Biol. Chem.* **1996**, *271*, 17067-17072.
33. Blake, C. C.; Serpell, L. C.; Sunde, M.; Sandgren, O.; Lundgren, E. *Ciba Found Symp* **1996**, 199:6-15; discussion 15-21, 40-6.
34. Lazo, N. D.; Downing, D. T. *Biochemistry* **1998**, *37*, 1731-1735.
35. Stevens, P. W.; Raffin, R.; Hanson, D. K.; Deng, Y. L.; Berrios-Hammond, M.; Westholm, F. A.; Murphy, C.; Eulitz, M.; Wetzell, R.; Solomon, A.; et, a. l. *Protein Sci* **1995**, *4*, 421-32.
36. Lansbury, P. T. Jr *Biochemistry* **1992**, *31*, 6865-70.
37. Lazo, N. D.; Downing, D. T. *Biochem. Biophys. Res. Comm.* **1997**, *235*, 675-679.

# Thermal Denaturation of Elastase in the Presence and Absence of Guanidinium Chloride: An IR Spectroscopic and DSC Investigation

D. M. Byler<sup>1,2</sup>, D. L. Lee<sup>2</sup>, and C. S. Randall<sup>3</sup>

<sup>1</sup>Department of Chemistry, St. Joseph's University, Philadelphia, PA 19131-1395

<sup>2</sup>SmithKline Beecham Pharmaceuticals, King of Prussia, PA 19406-0939

<sup>3</sup>Sanofi Research, Malvern, PA 19355

The response of the enzyme elastase to thermal stress was examined by second-derivative infrared (IR) spectroscopy and differential scanning microcalorimetry (DSC). In particular, we have measured the mean temperature ( $T_m$ ) for the unfolding of the peptide backbone in aqueous solution and after the addition of guanidinium chloride (GdmCl). At weakly basic pH, solutions of the native enzyme devoid of chemical denaturant are found to unfold with a  $T_m$  of  $\sim 59$  °C. Addition of guanidinium ion drives this parameter to lower values: 51 °C at 1M, 44 °C at 2M, and  $\sim 39$  °C at 3M [GdmCl]. Yet, even with 3M denaturant, the IR results indicate that a large fraction of the protein maintains its original conformation to within a few degrees of  $T_m$ . This suggests that GdmCl destabilizes protein secondary structure by altering the energy level for the native state relative to that of the denatured state without seriously disrupting the *intramolecular* hydrogen bonds linking the peptide carbonyls to nearby amide N-H groups for most of the protein molecules.

In an ongoing effort to elucidate fully the nature of protein structure and the process by which the peptide skeleton unfolds and refolds, biochemists have examined the effect of a range of factors known to perturb these macromolecular attributes, including variations in temperature, pressure, pH, ionic strength, concentration, solvent, and added chemical denaturants. These studies have employed such analytical techniques as ultracentrifugation, chromatography, light scattering, spectroscopy, and calorimetry. In much of the early work, available instruments had rather low sensitivity and could provide but limited information concerning the effect of the environment on the secondary structure of the macromolecule. Due to recent advances in technology, spectroscopic techniques such as nuclear magnetic resonance

(NMR), UV- and vibrational (infrared) circular dichroism (CD and VCD), Raman, and infrared (IR) are all now routinely used for probing protein conformation. For example, after the introduction in the mid-1980's of Fourier self-deconvolution and second derivative spectroscopy to "enhance" band resolution, IR found renewed favor as a tool for such studies (1, 2, 3).

Proteins, like polypeptides in general, consist of chains of amino acid residues joined end-to-end by secondary amide bonds. Different amino acids sport different side-chain groups. To accommodate their diverse shape, size, and polarity and attain a favorable free energy state, the protein folds into a compact geometry. Thus, the protein's unique sequence of amino acid residues (primary structure) prescribes the three dimensional curve traced by the atoms of the peptide chain. For most proteins, different segments of the backbone have their atoms arrayed in one of a few of simple, recurring architectural motifs or secondary structures. The most common of these exhibit periodically repeating geometries ( $\alpha$ -helices and  $\beta$ -strands); others are regular, but aperiodic (turns, bends, and loops). Those segments that cannot be so cataloged are said to have irregular or unordered structures. All these conformations that comprise the protein's framework are stabilized in part by a network of hydrogen bonds. Many of the latter link peptide carbonyls with nearby amide  $-NH$  groups.

As a result of the characteristic oscillations of the atoms of the amide group, the vibrational spectra of nearly all proteins display a set of nine diagnostic bands. In the infrared, the one most commonly investigated and also the most intense is the so-called Amide I band (1, 2). This appears between  $1700-1620\text{ cm}^{-1}$ . The major contributor to this absorption is the stretching vibration of the peptide carbonyls ( $C=O$ ). The observed wavenumber frequency at which a given subset of these carbonyls absorbs depends both upon their proximity and relative orientation in space and on the strengths of any hydrogen bonds that they may form with other polar groups. These structural parameters depend on the conformation of the framework atoms in the vicinity of the carbonyl. Thus, each distinct type of secondary structure will absorb in a different part of the amide I region.

Unfortunately, when carbonyls or other polar groups hydrogen bond, their IR bands are significantly broader than the analogous band of a  $C=O$  in a non-polar environment. Because the amide I components for the different types of conformation often lie closer in frequency than their bandwidths at half height, these absorptions commonly overlap one another badly. Hence, for most proteins, the observed amide I band is an unresolved composite of several different peaks and so has a complex, asymmetric shape. Although increased instrument resolution cannot reduce the natural width of a band, differentiation or Fourier self-deconvolution of the original protein band envelope will considerably narrow the underlying components. Often, this will be sufficient to visualize the individual peaks associated with specific classes of protein secondary structure. Thus, in most cases, the distinctive signature in the "resolution-enhanced" amide I band will provide qualitative identification of the major components of the protein's conformation. Furthermore, the relative integrated intensity of individual amide I band components is roughly proportional to the relative number of carbonyls (and hence peptide residues) that populate the corresponding subclass of conformational structure within the protein's architecture. Because of this, the judicious application iterative curve fitting will often yield a quantitative estimate of the composition of a protein's secondary structure (1, 2, 3). Just as importantly,

any observed variation in the amide I portion of the IR spectrum serves to signal that the secondary structure of a protein has undergone a corresponding change.

Urea or guanidinium ion is frequently used to unfold and denature proteins. Yet, the structure of this unfolded state is still poorly understood. Unfortunately, both substances absorb IR radiation in or near the amide I region of the spectrum even more strongly than does the protein itself. Thus, it has long been assumed that IR spectroscopy could offer little insight into the effect of these chemical perturbants on the structure of these macromolecules. Within the last five years, however, a few reports have shown, that with the careful choice of sampling conditions and the use of high-sensitivity interferometers, such studies can indeed bear fruit (4, 5, 6, 7, 8, 29).

Elastase is of potential interest to the pharmaceutical industry from two perspectives. First, this protein is a member of a class of proteolytic enzymes that have been implicated in the induction of several disease states, including emphysema, rheumatoid arthritis, and pancreatitis (9, 10). In addition, elastase itself may have therapeutic applications in the treatment of atherosclerosis, diabetic neuropathy, hypercholesterolemia, etc. (11, 12). Consequently, increased understanding of the factors and conditions that affect the stability of elastase should prove beneficial to the design of optimal formulations for this and other proteins under consideration as therapeutic agents.

The current study focuses on the use of DSC and IR spectroscopy to examine the effect of the chemical denaturant, guanidinium chloride [**GdmCl**], on the thermal stability of this structurally well characterized enzyme.

## Materials and Methods

Porcine pancreatic elastase [EC #3.4.21.11] (MW = 25.9 kD) was obtained from Worthington Biochemical Corp., Freehold, NJ (Cat. #6365, code ESFF, 95% protein, chromatographically prepared, lyophilized). All protein lyophiles were from the same lot, stored at *ca.* -20 °C. Guanidinium chloride (guanidine hydrochloride) [**GdmCl**] (Ultrapure) came from ICN Biomedicals, Inc., Aurora, OH. H<sub>2</sub>O was deionized in house. Tris(hydroxymethyl)aminomethane (**Tris**) was from Sigma Chemical Co., St. Louis, MO. D<sub>2</sub>O (99.9% D) was from Aldrich Chemical Co., Milwaukee, WI.

**Differential Scanning Microcalorimetry (DSC).** DSC studies were carried out with an MC-2 microcalorimeter (Microcal, Northampton, MA) equipped with an external nanovolt preamplifier. Samples were dissolved in 10mM Tris buffer (pH=8) prepared with deionized H<sub>2</sub>O; the same buffer was used in the reference cell. These protein solutions had weight/volume (w/V) concentrations between 1.5 and 3.0 mg/mL (0.15% to 0.30% w/V). Solutions were degassed; the cells were pressurized under nitrogen to minimize bubbling and evaporation. Scan rates were 1 °C/min from 7 to 95 °C; data were collected every fifteen seconds (13). After cooling the sample, a second DSC trial was run to judge if denaturation of the protein were reversible.

**Infrared Spectroscopy (IR).** All protein solutions for IR spectroscopic analysis were prepared by reconstituting a sample of lyophilized enzyme in neat D<sub>2</sub>O, without buffer or other excipients, under a dry N<sub>2</sub> atmosphere. Most of this protein's exchangeable hydrogens are readily replaced by deuteriums at ambient temperatures within less than one-half hour from the time of mixing; this is confirmed by the disappearance of the amide II band (due to amide NH groups) near 1550 cm<sup>-1</sup> and its

replacement by a new band near  $1450\text{ cm}^{-1}$  (due to ND moieties) (*I*, *14*). Final protein concentrations were estimated from the ratio of the measured dry mass of the lyophile and the volume of  $\text{D}_2\text{O}$  added by means of a calibrated micropipette (typically  $150\text{--}200\text{ }\mu\text{L}$ ). The values for solutions with  $1\text{--}3\text{ M GdmCl}$  ranged between  $36$  and  $45\text{ mg/mL}$  ( $3.6\%$  and  $4.5\%$  w/V). For the solution without GdmCl, the protein concentration was  $20\text{ mg/mL}$  ( $2.0\%$  w/V). Spectra were also collected of  $1.0\text{ M GdmCl}$  in  $\text{D}_2\text{O}$  solution without protein at  $4\text{--}5^\circ\text{C}$  temperature intervals from  $25$  to  $88^\circ\text{C}$ . This allowed the very intense band due to this molecule near  $1600\text{ cm}^{-1}$  to be effectively eliminated from the protein spectra by spectral subtraction.

The protein concentration in all of these solutions, even those at  $3\text{M GdmCl}$ , was sufficiently large to buffer the solution at  $\text{pD} = 7.3$ . These  $\text{pD}$ 's represent uncorrected values measured with a semi-micro combination pH electrode and a digital pH meter calibrated with  $\text{pH } 4.00$  and  $\text{pH } 7.00$  buffers (Baxter Diagnostics, Inc., Deerfield, IL) at ambient temperature ( $\sim 23 \pm 1^\circ\text{C}$ ).

The demountable, heatable IR cell with  $\text{CaF}_2$  windows was from Graseby Specac (Smyrna, GA). To avoid any possible problem in intensity measurements that might arise because of non-linear detector response, cell path lengths and total sample concentrations were kept sufficiently low to ensure that no band for any species in the region of interest ( $1800\text{--}1300\text{ cm}^{-1}$ ) had a peak maximum  $> 1.2$  absorbance units. For solutions of elastase without GdmCl, a nominal  $50\text{ }\mu\text{m}$  lead spacer was used. To keep the intensity of the GdmCl band near  $1600\text{ cm}^{-1}$  within reasonable bounds, samples with this denaturant necessitated the use of thinner, tin spacers:  $12\text{ }\mu\text{m}$  thick with  $1\text{M GdmCl}$ , and only  $6\text{ }\mu\text{m}$  with  $2\text{M}$  or  $3\text{M GdmCl}$ .

To collect spectra of solutions heated above room temperature, the IR cell was positioned in a water-jacketed cell holder. The latter was connected to a circulating water bath with a heater thermostatted to better than  $\pm 0.5^\circ\text{C}$ . Bath temperatures were estimated to the nearest  $0.3^\circ\text{C}$  with a calibrated Hg thermometer. These values were converted to calculated cell temperatures by a linear calibration based on a separate set of coincident temperature measurements of the bath with the Hg thermometer and of the cell with a calibrated, digital electronic thermometer connected to a Cu/Constantan (type K) thermocouple in direct contact with the cell body. After collecting data at a given temperature, the thermostat was manually set to the next higher temperature and the solution was allowed about  $10\text{ min}$  to attain thermal equilibrium. Then the process was repeated. After the data indicated that the protein was fully denatured, at least one or two additional spectra were collected at temperatures up to  $15^\circ\text{C}$  higher to ensure no further change in the amide I region of the second derivative IR spectrum had occurred.

**Data Collection Parameters.** IR spectra were obtained with a Nicolet 510 Fourier-transform infrared (FTIR) spectrometer equipped with an air-cooled Globar source, a Ge/KBr beamsplitter, a room-temperature DTGS detector, an 18-bit analog-to-digital converter, and a Nicolet 620 workstation with NICOS and DX software. All spectra were collected at  $\text{Gain}=2$  and at a nominal instrument resolution of  $4\text{ cm}^{-1}$ . The mirror velocity was  $\sim 1.4\text{ scans/s}$  ( $\text{VEL} = 30$ ). To ensure a sufficient signal-to-noise ratio, each spectrum was derived from  $256$  or  $512$  single-sided interferograms, co-added. After one level of zero-filling, each interferogram consisted of  $10240$  data points; all were single precision, except for  $512$  surrounding the centerburst. The latter were



double precision. Thus, data points were spaced  $0.9644\text{ cm}^{-1}$  apart. After apodization by the Happ-Genzel function, the signal-averaged interferogram was Fourier transformed to give a single-beam spectrum. Finally, the single beam spectrum for each sample is ratioed against a single beam spectrum of the empty sample compartment to yield the usual infrared absorbance spectrum. The spectrometer and its sample chamber were continuously purged with dry nitrogen gas. Residual traces of water vapor lines were eliminated from the spectra as previously described (15, 16) Before initiating any additional data handling, the IR data in Nicolet DX format was converted with NIconvert software to PC format. This enabled us to use Nicolet's OMNIC software to calculate derivative spectra, measure peak parameters, etc.

**Normalization of IR Intensities.** To facilitate the comparison among spectra collected from samples with different protein concentrations and those measured at different cell pathlengths, all ordinate values were normalized. The Y-coordinate for every data point in each spectrum was multiplied by a factor calculated to make the integrated area of the middle lobe of the second derivative peak of the tyrosine ring band at  $1517\text{ cm}^{-1}$  (1, 17, 18) the same for all spectra. This sharp peak was chosen as the intensity benchmark because it is rather isolated from other absorptions and easily discernible. Also, its frequency and band area are little affected by most changes in the local environment of the molecule. Deuteration of the tyrosine's hydroxyl group causes just a small downward shift of  $\sim 1\text{ cm}^{-1}$  (18). Thus, for a protein, the integrated area of this marker band is proportional to the number of tyrosines residues (18). Furthermore, this value is invariant for a fixed amount of a specific protein, even if its conformation changes or if the peptide chain is cut apart by proteolysis.

After smoothing with a 9-point window, normalized peak intensities (measured relative to  $y=0$ ) for selected second derivative bands were plotted versus temperature to estimate the mid-range temperature of protein denaturation,  $T_m$ .

## Results

Both differential scanning microcalorimetry (DSC) and second-derivative infrared spectroscopy (IR) indicate that the conformational structure of porcine elastase is reasonably stable in neutral or weakly basic aqueous solutions even at temperatures 10-15 °C above typical mammalian body temperatures. Indeed, exposure to such conditions for an hour or more induces rather little change in the observed IR signal.

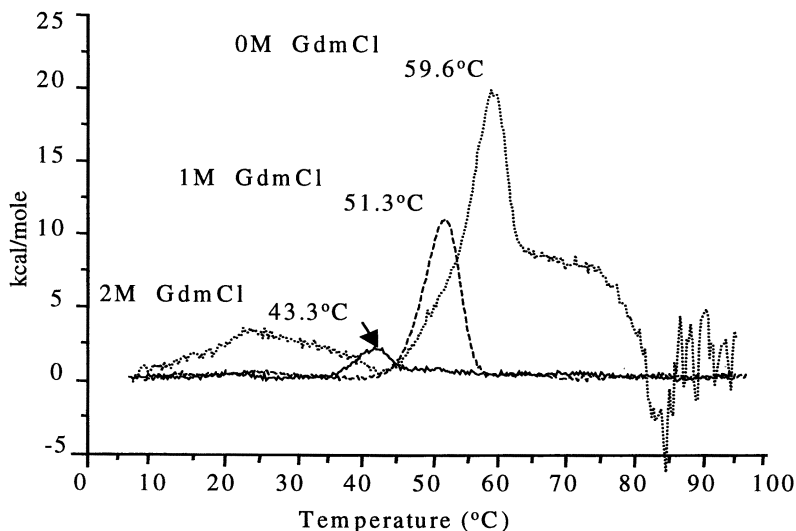
**DSC Measurements.** DSC allows direct measurement of heat transfer associated with protein unfolding and other changes in protein structure brought about by perturbations in temperature (19, 20). Adiabatic DSC measures the rate of heat flow in  $J/\text{min}$  as a function of temperature in or out of a given volume of sample. Heat flow divided by the scan rate in  $^{\circ}\text{C}/\text{min}$  gives the heat capacity,  $C_p$ , in  $J/^{\circ}\text{C}$ . Normalization for protein molar concentration yields the molar heat capacity in  $\text{kJ}/\text{mol}\cdot^{\circ}\text{C}$ . The position of the peak maximum of the thermogram in  $^{\circ}\text{C}$  corresponds to  $T_m$ , the midpoint of the characteristic temperature range over which the aqueous protein denatures (21).

The deviation of the heating curve for the protein solution from that of the baseline thermogram for pure water or of aqueous buffer in the same cell gives the change in heat capacity,  $\Delta C_p$ , of the protein sample at any given temperature. The

integrated area of the baseline-corrected peak on the DSC plot is a measure of the total energy transferred as the native protein unfolds. The observed value for aqueous elastase in the absence of chemical denaturants is +1500 kJ/mol.

DSC data for elastase in 10 mM aqueous Tris buffer (pH 8) with no denaturant (0 M GdmCl) reveal a single, broad, roughly symmetric, endothermic event centered at 59.6 °C (Figure 1), flanked by a broad shoulder stretching some 20 °C higher. The observed full-width at half-height (FWHH) of the DSC peak gives a sense of how broad the range of temperatures around  $T_m$  is over which thermal denaturation occurs. In the absence of chemical denaturants, the observed FWHH for the unfolding of the protein in aqueous solution was 7.3 °C. When the sample had cooled to the starting temperature, another DSC scan was made to learn if thermal denaturation of elastase was reversible. No change in  $C_p$  from baseline was observed for this new trial.

As is well known, addition of guanidinium chloride or other denaturant to a protein solution destabilizes the balance of molecular forces requisite to maintaining the encoded native conformation (secondary structure) of the macromolecule (21, 22). With each increase in concentration of guanidinium ion, the temperature of unfolding ( $T_m$ ) decreases (Figure 1 & Table I). In contrast to the measurements on solutions



**Figure 1.** DSC scans for aqueous elastase with 0.0 M [.....], 1.0 M [----], and 2.0 M [—] GdmCl.  $T_m$  values are noted for each curve.

with no denaturant, the observed DSC peak of the chemically perturbed solutions is more symmetric and has smaller FWHH values. In addition, these data show that increasing the denaturant concentration reduces the measured integrated area under DSC curve and hence the calculated energy change for the transition. The total energy transferred as the protein unfolds in 1.0 M GdmCl is 290 kJ/mol; with 2.0 M denaturant, it is only 110 kJ/mol. Indeed,  $C_p$  observed for a solution of protein in 3M GdmCl is indistinguishable from that of the reference buffer itself throughout the

**Table I: Observed  $T_m$  values for elastase exposed to different concentrations of guanidinium chloride [GdmCl]**

[GdmCl] (M)	DSC pH 8.0 in H <sub>2</sub> O <sup>a</sup>		IR pD 7.3 in D <sub>2</sub> O <sup>b</sup>
	$T_m$ (°C)	FWHH <sup>c</sup> (°C)	(°C)
0.0	59.6	7.3	59
1.0	51.3	6.0	51
2.0	43.3	6.0	44
3.0	(No signal detected)		39

<sup>a</sup> With 10 mM Tris buffer<sup>b</sup> No buffer or other excipients present<sup>c</sup> Full-width at half-height (FWHH)

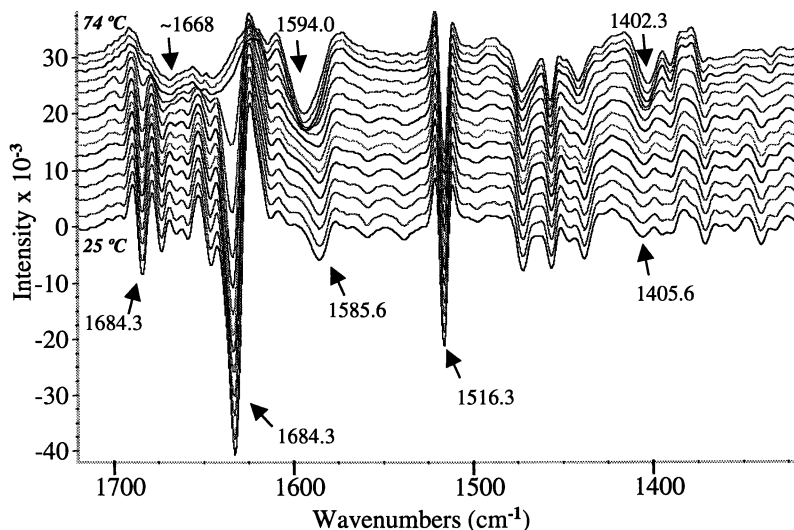
range of temperatures scanned. Thus, within the limits of background noise reported by the instrument, measured  $\Delta C_p$  has fallen to zero throughout the DSC scan and no detectable energy has been transferred to or from the protein as it unfolds.

**IR Spectroscopy.** These same thermodynamic events and structural changes in the elastase molecule were also probed by means of second-derivative IR spectroscopy. Between 1700-1620  $\text{cm}^{-1}$ , the amide I absorption for native elastase in D<sub>2</sub>O at pD = 7.3 is an asymmetric, composite band. Its maximum is at 1636.0  $\text{cm}^{-1}$  trailing off to a very weak shoulder near 1685  $\text{cm}^{-1}$ . These values are typical of globular proteins with the predominant fraction of the peptide residues folded into anti-parallel  $\beta$ -sheets (1,2,3,15,23,24). Taking the second derivative of this spectrum narrows the overlapping amide I components so that six intrinsic bands are observed: 1632.8 (s), 1646.3 (w), 1658.6 (w), 1665.3 (w), 1673.4 (w), and 1684.2 (w)  $\text{cm}^{-1}$  (16) (Figure 2).

As the protein solution is heated and its temperature approaches to within a few degrees of  $T_m$ , the value at which unfolding begins, these six second-derivative bands simultaneously start to diminish in intensity (Figure 2). While the temperature continues to rise, the original peaks are replaced by two broader features (near 1645 and 1668  $\text{cm}^{-1}$ ) much like those that characterize the second derivative spectra of other heat-denatured proteins (6, 13). (The original, undifferentiated amide I absorbance band for this heat-denatured sample crests near 1654  $\text{cm}^{-1}$ .)

Besides these distinct changes in the amide I region, two other bands at 1586 and 1406  $\text{cm}^{-1}$  also undergo a temperature-induced transformation in shape and intensity. Additionally, the former also displays a distinct shift to higher frequency ( $\sim 1590 \text{ cm}^{-1}$ ) at temperatures greater than  $T_m$ . As heat continues to be applied to the solution, the remainder of the spectrum seems scarcely affected by comparison. Other shifts in frequency or intensity do occur throughout the spectrum, but most are nearly imperceptible at the normal scale on which these data are displayed.

**Changes in IR Band Parameters vs. Temperature.** To more easily view and analyze whatever changes may appear in the IR spectra of elastase as the protein solution is heated, plots of selected spectroscopic parameters as a function of temperature were prepared for the bands at 1684, 1633, 1585, 1516, and 1406  $\text{cm}^{-1}$ .



**Figure 2.** Second derivative IR spectra of elastase [0M GdmCl] collected at 25.5, 29.5, 34.5, 38.1, 41.7, 44.5, 47.2, 49.9, 52.6, 55.3, 58.0, 60.7, 63.5, 67.1, and 74.3°C. Note: second derivative peaks have negative intensities.

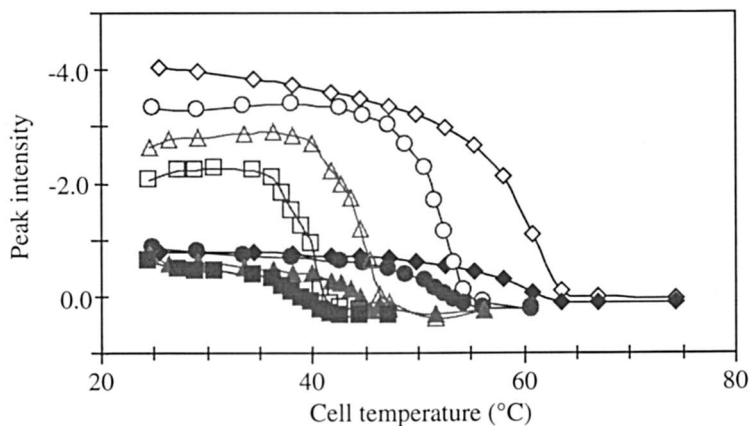
For example, to scrutinize the variation in second derivative intensity for a given band as the sample was heated, the normalized ordinate value at the center of this peak from the spectrum at each temperature in a given series of measurements was plotted versus cell (sample) temperature. The resulting curves are nearly sigmoidal in shape (Figures 3-5). For solutions with just the native protein, the inflection point of these curves lies close to 59 °C (Table I). Other analytical techniques yield similar plots of protein unfolding parameters (21, 22).

For elastase solutions containing 1.0, 2.0, or 3.0 M [GdmCl], peak frequencies measured at room temperature are quite similar to those reported above for elastase free of denaturant. With the possible exception of the 1633  $\text{cm}^{-1}$  band, none of the observed values shifted more than  $\sim 1 \text{ cm}^{-1}$ . Also, for any single spectrum, the intensities of the six amide I bands relative to one another are largely unaffected by the level of guanidinium ion. However, with each increment in GdmCl concentration, the observed normalized intensity of every amide I band diminishes (Figure 3).

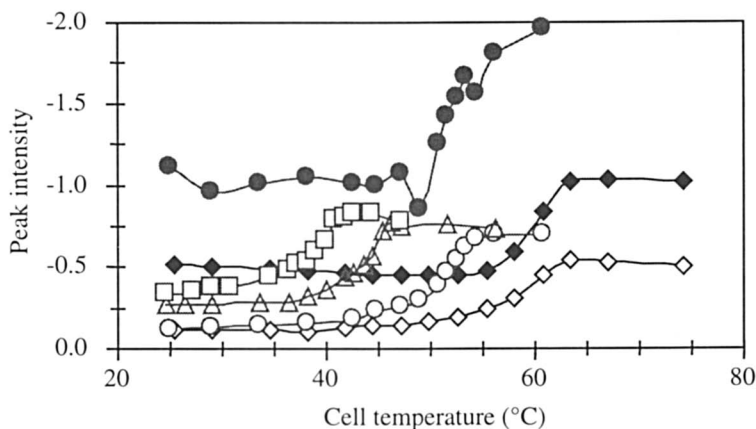
Plots of band parameter values versus temperature for spectra of samples at each of the three [GdmCl] studied (Figures 3-5) give sigmoidal curves analogous to those for the same parameter taken from spectra of protein solutions free of denaturant. But, as [GdmCl] increases, the inflection point of the graph now shifts to lower temperatures. At each concentration, the temperature at maximum slope measured by IR closely matches the protein unfolding temperature ( $T_m$ ) determined by DSC (Table I).

## Discussion

Some small, globular proteins, such as ribonuclease A, when exposed to elevated temperatures or high concentrations of chemical denaturants, can unfold essentially reversibly (6). For such species, removal of the perturbing factor allows the peptide



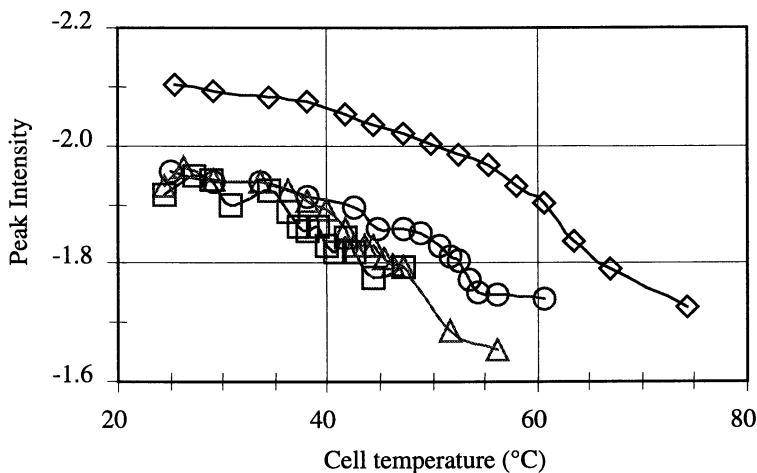
**Figure 3.** Plots of 1633 (open) and 1684  $\text{cm}^{-1}$  (filled) band intensity vs. temperature for elastase solutions with 0M [◇], 1M [○], 2M [△], and 3M [□] GdmCl.



**Figure 4.** Plots of 1406 (open) and 1585  $\text{cm}^{-1}$  (filled) band intensity vs. temperature for elastase solutions with 0M [◇], 1M [○], 2M [△], and 3M [□] GdmCl.

chain spontaneously to reestablish its native conformation. The IR spectrum of the unfolded protein differs significantly from that of the native state, but upon refolding again becomes virtually indistinguishable from that of the original (6). A repetition of the DSC scan of a solution of such a protein would give a second thermogram with essentially the same  $T_m$  and integrated area as the original.

By contrast, our results for elastase from both DSC and IR show that the changes wrought in its secondary structure by the application of high temperatures are clearly



**Figure 5.** Plots of  $1516\text{ cm}^{-1}$  band intensity (open) vs. temperature for elastase solutions with 0M [◇], 1M [O], 2M [Δ], and 3M [◻] GdmCl.

obviously irreversible even in the absence of any GdmCl. Specifically, the second derivative IR spectrum of aqueous elastase at temperatures above  $T_m$  and that of the same solution after it is cooled and held at ambient overnight are indistinguishable. The amide I region of both spectra exhibits only the same two broad bands at  $1667$  and  $1645\text{ cm}^{-1}$  that characterize the denatured protein. This is in clear contrast to the six distinct amide I bands found in the spectrum of the native protein. Previous spectroscopic studies (25) corroborate our observations that, once heat-denatured, no appreciable fraction of an elastase sample returns to its native conformation.

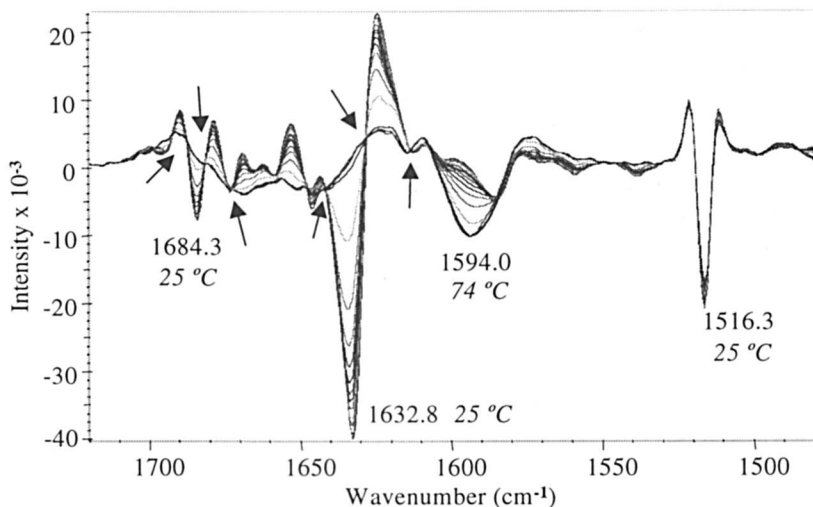
That thermal denaturation of elastase is irreversible is also consistent with our DSC data. If the heat treated protein sample is subjected to a second DSC scan (*not shown*), measured  $\Delta C_p$  remains zero throughout the run. Absence of detectable heat transfer to or from the protein during this new trial implies that again cooling the protein to room temperature is insufficient to make it refold into any conformation with a thermodynamic state distinguishable from that of the unfolded protein.

Therefore, strictly speaking, the unfolding process for elastase under such conditions cannot be represented by a simple two-state transition in which the protein in the native state (N) remains in equilibrium with the fully unfolded protein (U) (22). Instead, the final, irreversible form of the denatured elastase (U) must differ in an undetermined way from some, presumably reversibly unfolded, intermediate state (I). Yet, on the basis of the present results, neither DSC nor IR can distinguish this putative intermediate from the final unfolded product (U). Apparently whatever the differences between them, the transformation of I to U induces no perceptible change in the recorded signal from either physical probe. In the case of DSC, it seems likely either that  $\Delta C_p$  for the process  $I \rightarrow U$  is too small to describe or that perhaps any thermodynamic change is impossible to resolve from that associated with the initial unfolding of the native enzyme (N). The apparent lack of change in the amide I region of the IR spectra during the transition from  $I \rightarrow U$  suggests further that for both

species the spatial interrelationships of the backbone carbonyls and the network of hydrogen bonds to the carbonyl oxygens remain largely unaffected by this molecular transformation. Heating induces similar irreversible changes in the conformation of samples with different [GdmCl] (*spectra not shown*). To reiterate, a two-state model is too simple to represent fully the thermal denaturation of elastase at pH 7. After the DSC experiments, aggregation products visible to the naked eye are apparent in the sample cell. Yet, no obvious visual evidence of aggregation is seen following IR measurements, despite the much larger protein concentrations employed. Possibly the higher final temperatures attained in the former case (95 °C) versus a maximum of just 75 °C for the latter may account for this difference.

For many proteins, thermal denaturation invariably leads irreversibly to the formation of insoluble aggregates, whether their native conformations are dominated by helices, pleated sheets, or some admixture. Generally, the IR spectra of such aggregates exhibit a pair of amide I absorptions near 1620 and 1685  $\text{cm}^{-1}$  with an intensity ratio of about 2:1. The frequency of the first lies outside the 1625 to 1640  $\text{cm}^{-1}$  range where the  $\beta$ -sheet carbonyls of native proteins absorb (2, 3, 5, 13, 26, 27). This pair of bands represents the absorptions of peptide carbonyls in a novel class of  $\beta$ -structures. These are characterized by *intermolecular* hydrogen bonds linking the N-H's of the peptide chain of one molecule to the carbonyls of another. The absence of bands near 1620 and 1685  $\text{cm}^{-1}$  in the IR spectra of heat-denatured elastase suggests that up to  $\sim 75$  °C no such *intermolecular* structures form as this protein unfolds (Figure 2). This is true whether or not the protein was exposed to GdmCl.

The above caveats warn against the temptation to assume arbitrarily that a simple two-state equilibrium adequately represents the overall pathway by which heat leads elastase to unfold. Nevertheless, the IR data as a function of sample temperature does provide some evidence of the contrary case. Plotting an overlay of the second derivative spectra recorded at different temperatures (Figure 6) reveals several



**Figure 6.** Second derivative IR spectra of elastase [0M GdmCl] collected at 15 temperatures from 25 to 74 °C. The arrows point to some isosbestic points.

frequencies at which all the spectra (except perhaps those at the highest temperatures) simultaneously intersect at a common point. Such crossover junctures for a series of spectra are termed isosbestic points (28). As the temperature increases, the peak intensities of the six amide I bands associated with the native protein (Figure 2) progressively decrease. Concurrently, two new, broad peaks associated with the unfolded polypeptide appear and gradually intensify. These effects are especially obvious for the spectra measured over the narrow range of temperatures surrounding  $T_m$ . Generally, true isosbestic points occur only when a chemical process involves a simple equilibrium between two distinct molecular structures (28). Furthermore, when the peak intensities of individual amide I bands are plotted as a function of temperature, the plot profile for all but the values at high temperatures resembles the typical sigmoidal curves observed for various parameters that track the unfolding of many kinds of proteins (21, 22). At temperatures beyond  $T_m$ , the empirical contour deviates from that of the ideal curve. Presumably, in the early stages, the process of unfolding may follow a reversible pathway. Later, the denaturation becomes less cooperative and more irreversible in character. These effects at higher temperatures occur whether or not GdmCl is present in the protein solution.

As shown in Figure 3, if no denaturant is present, the strong  $1633\text{ cm}^{-1}$  amide I band exhibits a small, but steady decrease in intensity even at temperatures close to ambient. This is not due to any change in protein structure, but results from small isotopic shifts in band frequency attributable to continued, slow H-D substitution of some hard-to-exchange peptide hydrogens (1-3, 13). The absence of this effect for the solutions with GdmCl implies that under these conditions H-D exchange is sufficiently rapid that it is nearly complete before the first spectrum at elevated temperature is collected. Perhaps then, one of the effects GdmCl has on the protein, even at room temperature, is to open its tertiary structure, without causing an appreciable, concomitant loss of secondary structure. If this is so, perhaps the protein may then be said to be in the so-called "molten globule" state (21).

The reasons for the increase in the intensity of the two bands associated with side-chain carboxylates at  $1586$  and  $1406\text{ cm}^{-1}$  at temperatures above  $T_m$  are unclear (Figures 2, 4 and 6). Unpublished data show just the opposite effect for a monoclonal immunoglobulin (D. M. Byler and D. L. Lee, 1997). No such obvious changes are observed for other side-chain absorptions of either protein (Figure 2, 5, and 6).

Clearly, future investigations of this enzyme should be directed toward a search for conditions under which unfolding may be kept fully reversible. These studies might include examination of the kinetics of this process at temperatures just a few degrees below  $T_m$  (13). Further scrutiny is also warranted of the factors that lead some proteins to form denatured aggregates with *intermolecular* hydrogen bonding versus those, like elastase, that may denature without doing so. In addition, detailed measurements of H-D exchange rates in the presence and absence of denaturant may offer other clues to the structure of the protein in such disparate environments.

## Conclusion

Not surprisingly, the IR data show that the most obvious change in the structure of elastase as temperature increases is the loss of most of the anti-parallel  $\beta$ -sheets present in its native form. Such secondary structures comprise the largest fraction of this protein's original conformation. Whatever the GdmCl concentration, the spectra



of D<sub>2</sub>O solutions of the heat-modified protein indicate that exposure to temperatures  $\geq T_m$  ultimately unravels its natural structure: the six amide I bands associated with the normal conformation (Figure 2) are lost and two, new, broad features attributed to unordered, aperiodic structures, possibly including a modified form of turns, are now present. The latter bands near 1645 and 1668 cm<sup>-1</sup> are commonly all that is observed in the amide I region of the second derivative IR spectra for a variety of other denatured proteins (1, 8, 17, 27).

The results of this investigation suggest further that at ambient temperatures, even in the presence of up to 3 M guanidinium ion, most of the elastase molecules in aqueous solution maintain their native conformations. By contrast, increasing the concentration of denaturant dramatically lowers the median temperature of denaturation ( $T_m$ ) for this macromolecule. As shown by DSC for elastase dissolved in 3M GdmCl (aq), the endothermic transfer of heat energy from the native protein as it unfolds under these conditions approaches zero. Clearly, the high polarity and hydrogen bonding properties associated with this ion drastically alters the nature of the many *intramolecular* forces and interactions stabilizing the native conformation of the protein under these conditions. Yet, however this is mediated, at room temperature even 3M GdmCl is insufficient to cause all of the protein molecules to unfold within the span of an hour or more.

**Acknowledgments.** The senior author (D.M.B.) dedicates this paper to the memory of his friend and colleague, Dr. Heino Susi (1925-1987). In 1983, they wrote the first report on the use of second derivative IR spectroscopy to “resolve” the amide I band structure of proteins (17). D.M.B. also acknowledges SmithKline Beecham Pharmaceuticals for providing the laboratory facilities and supplies needed to do this research. He especially thanks Dr. John M. Baldoni, Dr. M. Marcia Federici, Dr. Lee Gardella, and Dr. Theodore D. Sokoloski of SB for their support of his research.

## Literature Cited

1. Susi, H.; Byler, D. M. *Methods Enzymol.* **1986**, *130*, 290-311.
2. Surewicz, W. K.; Mantsch, H. H. *Biochim. Biophys. Acta* **1988**, *952*, 115-130.
3. Surewicz, W. K.; Mantsch, H. H. Chapman, D. *Biochemistry* **1993**, *32*, 389-394.
4. Bowler, B. E.; Dong, A.; Caughey, W. S. *Biochemistry* **1994**, *33*, 2402-2408.
5. Fabian, H.; Mantsch, H. H. *Biochemistry* **1995**, *34*, 13651-13655.
6. Reinstadler, D.; Fabian, H.; Backmann, J.; Naumann, D. *Biochemistry* **1996**, *35*, 15822-15830.
7. Li, X.-F.; Zhou, J.-M. *Biospectroscopy* **1997**, *3*, 121-129.
8. From, N. B.; Bowler, B. E. *Biochemistry* **1998**, *37*, 1623-1631.
9. Bieth, J. in *Frontiers of Matrix Biology*; Karger: Basel, Switzerland, 1978, Vol. 6, pp. 1-82.
10. Eriksson, S. *Eur. Respir. J.* **1991**, *4*, 1041-1043.
11. Ooyama, T.; Sakanioto, H. *Ciba Foundation Symp.* **1995**, *192*, 307-320.
12. Nagai, T. *Diabetes Res. Clin. Pract.* **1994**, *24*, 161-165.
13. Byler, D. M.; Lee, D. L.; Sokoloski, T.D.; Dal Monte, P.R.; Baldoni, J.M. *Fourier Transform Spectroscopy: Proc. 11th Internatl. Conf.*, **1998**, CP430, 332-335.
14. Zuber, G.; Prestrelski, S. J.; Benedek, K. *Anal. Biochem.* **1992**, *207*, 150-156.

15. Prestrelski, S. J.; Byler, D. M.; Liebman, M. N. *Proteins* **1992**, *14*, 440-450.
16. Byler, D. M.; Wilson, R. M.; Randall, C. S.; Sokoloski, T. D. *Pharm. Res.* **1995**, *12*, 446-450.
17. Susi, H.; Byler, D. M. *Biochem. Biophys. Res. Commun.* **1983**, *115*, 391-397.
18. Bendit, E. G. *Biopolymers* **1967**, *5*, 525-533.
19. Freire, E. *Meth. Enzymol.* **1995**, *259*, 144-168.
20. Relkin, P. *Thermochim. Acta.* **1994**, *246*, 371-386.
21. Creighton, T. E. *Proteins: Structures and Molecular Properties*, 2<sup>nd</sup> ed.; W. H. Freeman and Co.: New York, NY, 1983; pp. 287-296.
22. Eftink, M. R.; Ionescu, R. *Biophys. Chem.* **1997**, *64*, 175-197.
23. Byler, D. M.; Susi, H. *J. Industr. Microbiol.* **1988**, *3*, 73-88.
24. Byler, D. M.; Susi, H. *Biopolymers* **1986**, *25*, 469-487.
25. Favre-Bonvin, G.; Bostancioglu, K.; Wallach, J. M. *Biochem. Int.* **1986**, *13*, 983-989
26. Clark, A. H.; Saunderson, D. H. P.; Suggett, A. *Int. J. Peptide Res.*, **1981**, *17*, 353-364.
27. Byler, D. M.; Purcell, J. M. *SPIE Fourier Transform Spectroscopy*, **1989**, *1145*, 415-417.
28. Harris, D. C. *Quantitative Chemical Analysis*, 4th ed.; W. H. Freeman and Co.: New York, NY, 1995; pp. 527-528.
29. Kendrick, B. S.; Cleland, J. L.; Lam, X.; Nguyen, T.; Randolph, T. W.; Manning, M. C.; Carpenter, J. F. *J. Pharm. Sci.* **1998**, *87*, 1069-1076.

## Chapter 8

# Molecular Reaction Mechanisms of Proteins Monitored by Time-Resolved FTIR and FT-Raman Spectroscopy

Klaus Gerwert

Ruhr-Universität Bochum, Lehrstuhl für Biophysik, D-44780, Bochum, Germany  
(fax: +49 234 7094238; e-mail: gerwert@bph.ruhr-uni-bochum.de)

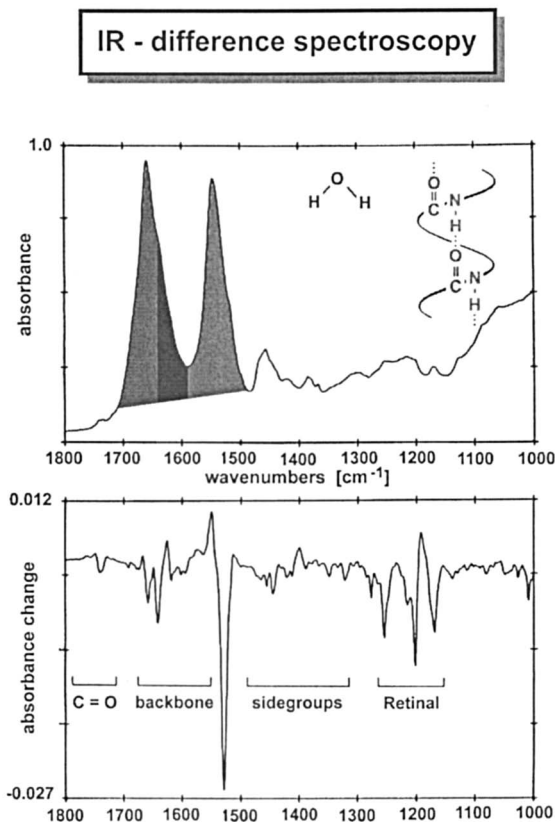
### INTRODUCTION

Time-resolved FT-IR difference spectroscopy has been established as a new biophysical tool for the investigation of molecular reaction mechanisms of proteins at the atomic level. Although infrared spectroscopy is a classical method of analytical and structural investigations<sup>1</sup>, it can also provide valuable insight into the mechanisms of chemical reactions. Especially timeresolved Fourier-Transform-Infrared-difference-spectroscopy has proved itself as a powerful method for studies of molecular reaction mechanisms of large proteins up to 120000 Dalton and time resolution up to nanoseconds<sup>2</sup>.

The infrared spectrum of a protein is dominated by its peptide backbone amide I (C=O) and amide II (C-N, NH) vibrations (Fig. 1). In addition to the strong amide I and amide II bands water also contributes largely to the background absorption (1650 cm<sup>-1</sup>).

The major problem in measuring reactions consists in selecting the small absorption bands of the molecular groups which undergo reactions from the large background absorption of water and of the quiescent entire protein. The absorbance bands are selected by the performance of difference spectra between spectra of the protein in its ground state and in an activated state. Such measurements require highly sensitive instrumentation. The FTIR method has two advantages over conventional dispersive IR spectroscopy: the multiplex and the Jaquinot advantage. This allows reliable detection of very small absorption changes with a time resolution of a few nanoseconds.

A further important step is the assignment of the bands to individual groups of the protein. This is performed by using isotopically labelling, for example isotopically labelled retinal (5) or isotopically labelled quinone (9) or by using site directed mutated proteins for example asp96asn (26) or Glu204Gln (25) bacteriorhodopsin

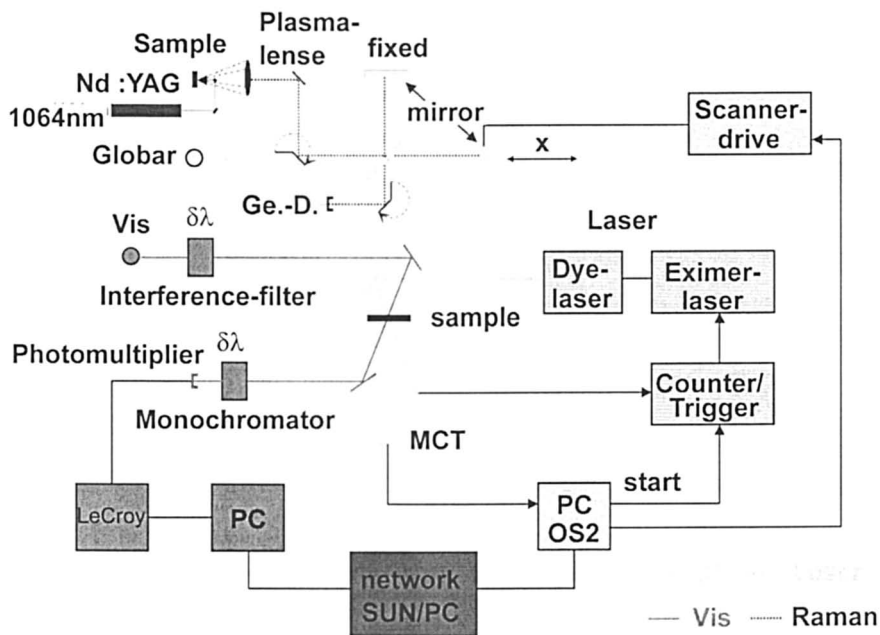


**Fig. 1:** An IR absorbance spectrum of a hydrated protein (bacteriorhodopsin) and a typical difference spectrum is shown.

mutants. The combination of FTIR with molecularbiology methods allows an unequivocal assignment.

In protein chemistry the macromolecule studied most intensely thus far by time-resolved FT-IR difference spectroscopy is the light-driven proton pump bacteriorhodopsin. Many research groups worldwide have contributed to the understanding of the mechanism of this membrane protein<sup>3</sup>. For reviews on FT-IR studies on bacteriorhodopsin, see references<sup>4,5</sup> and<sup>6</sup>. Time-resolved FT-IR spectroscopy has furthermore provided a detailed picture of the light-induced electron transfer mechanism of bacterial photosynthetic reaction centers. For reviews and recent FT-IR work, see references<sup>7,8</sup> and<sup>9</sup>. Beside photobiological proteins, time-resolved FT-IR spectroscopy can also be applied to proteins without an intrinsic chromophore by using photolabile trigger compounds.<sup>10,11</sup> Using "caged" GTP the molecular GTPase mechanism of the oncogenic protein H-ras p21, a molecule that plays a central role in the growth of cancer cells,<sup>12</sup> can be monitored.<sup>13</sup>

Another example is "caged"  $\text{Ca}^{2+}$ , by which the molecular mechanism of the  $\text{Ca}^{2+}$ -ATPase is investigated.<sup>14,15</sup> Furthermore "caged" electrons are used to study redox reactions in cytochrome-oxidase.<sup>16,17</sup>



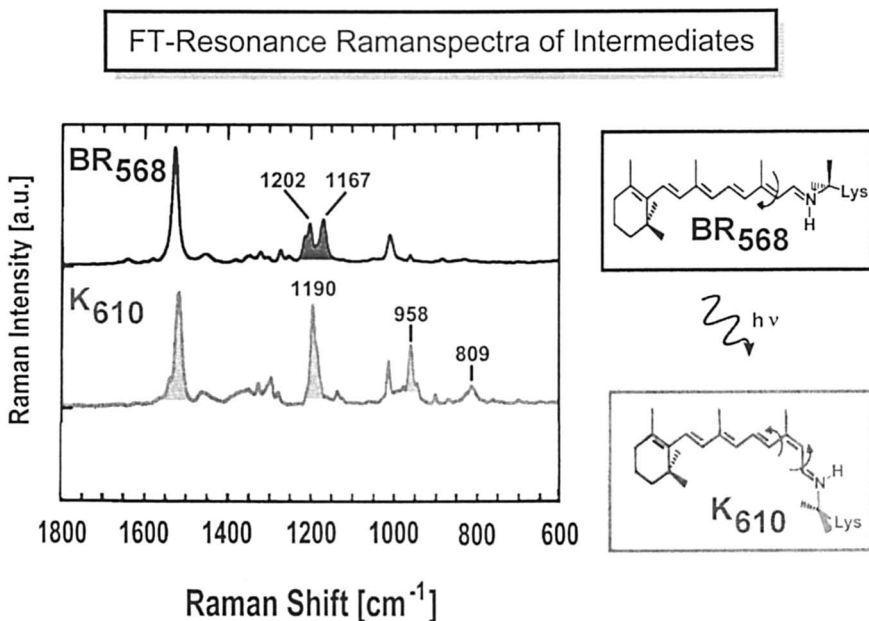
**Fig. 2.** The experimental setup, consisting of an FT-IR apparatus with globar, beamsplitter, mirrors, controller, detector and preamplifier connected to a 200 kHz and a 200 MHz transient recorder; a photolysis setup with light source, interference filters, monochromator, photomultiplier and transient recorder; and an excimer pumped dye laser system to activate the sample. The VIS and IR data are transferred to a workstation (SUN)/PC network for kinetic analysis.

A typical experimental setup is shown in Fig. 2. This setup allows time resolved FTIR, VIS absorbance change and FT Raman measurements. Various time-resolved FT-IR techniques are developed: the rapid-scan technique, which gives millisecond time resolution<sup>18-20</sup>, the stroboscope technique, which gives microsecond time resolution<sup>21,22</sup>, and the step-scan technique, which gives nanoseconds time resolution<sup>23-25</sup>. Here bacteriorhodopsin is presented as an example for the application of time-resolved FTIR spectroscopy to study molecular reaction mechanisms of proteins.

### Bacteriorhodopsin

After light-excitation of bacteriorhodopsin's light-adapted ground-state BR<sub>570</sub> a photocycle starts with the intermediates J<sub>610</sub>, K<sub>590</sub>, L<sub>550</sub>, M<sub>410</sub>, N<sub>530</sub> and O<sub>640</sub> distinguished by the different absorption maxima given by the indices. During the rise of the M-intermediate a proton is released to the extracellular side, and during the M-decay a proton is taken up from the cytoplasmic side. This creates a chemico-osmotic proton gradient across the membrane.

#### The light-induced BR to K transition



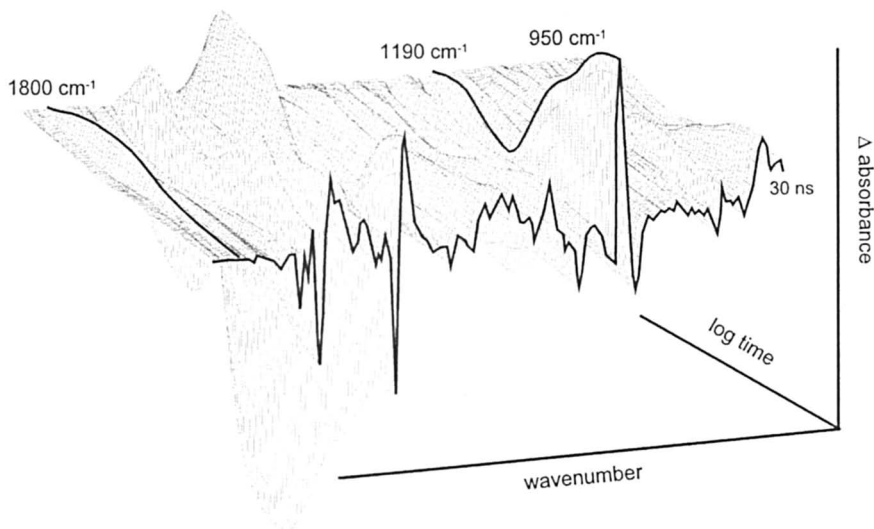
**Fig 3** In a) FT-Resonance Raman spectrum of the BR ground state ( $1202\text{ cm}^{-1}$   $C_{14}-C_{15}$  stretch;  $1167\text{ cm}^{-1}$   $C_{10}-C_{11}$  stretch) and in b) of the K ( $1190\text{ cm}^{-1}$ : ( $C_{14}-C_{15} + C_{10}-C_{11}$  stretch) intermediate stabilized at 77K is shown.

In Fig.3 a BR and K FT-Resonance-Raman spectrum is shown. Because of the Resonance transition only chromophore bands are observed. The K intermediate is stabilized at 77 K. The band pattern in the 'fingerprint' region between  $1300\text{ cm}^{-1}$  and  $1100\text{ cm}^{-1}$  is typical of an all-trans ( $1202\text{ cm}^{-1}$ ,  $1167\text{ cm}^{-1}$ ) configuration in BR and a 13-cis ( $1190\text{ cm}^{-1}$ ) in K. The bands at  $958\text{ cm}^{-1}$  and  $809\text{ cm}^{-1}$  are caused by retinal Hoop (Hydrogen out of plane) vibrations. They are indicative of a strong distortion of the retinal. The BR and K Raman spectra indicate that isomerization of

the chromophore induce tension in the protein that drives the subsequent thermal reactions.

***The thermal reactions:  $K \rightarrow L \rightarrow M \rightarrow N \rightarrow O \rightarrow BR$***

The absorbance changes accompanying this reaction pathway can be monitored under physiological conditions by the step scan FTIR technique with a time resolution of 30 ns.<sup>25</sup>

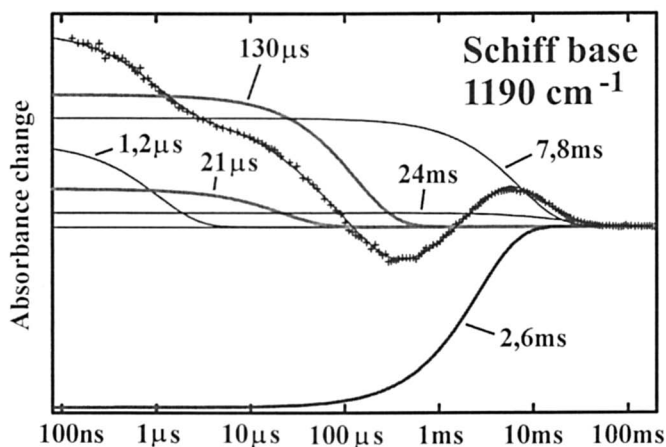


***Fig. 4. A three dimensional representation of the IR absorbance changes between 1800  $\text{cm}^{-1}$  and 1000  $\text{cm}^{-1}$  with 30 ns time-resolution and 3  $\text{cm}^{-1}$  spectral resolution accompanying bacteriorhodopsin's photocycle as revealed by a global-fit analysis. The time axis has a logarithmic scale in order to show the complete bacteriorhodopsin photocycle from 30 ns up to 10 ms in one representation.***

In fig. 4 the IR absorbance changes between 1800  $\text{cm}^{-1}$  and 1000  $\text{cm}^{-1}$  during the photocycle are shown in a three-dimensional representation. Beyond a background absorbance of up to one absorbance unit, changes on the order of  $10^{-3}$  to  $10^{-4}$  are monitored with 3  $\text{cm}^{-1}$  spectral resolution and 30 ns time-resolution.

In Fig. 5 a typical kinetic, the absorbance change at 1190  $\text{cm}^{-1}$ , is shown.

The appearance of the C-C stretching vibration band at 1190  $\text{cm}^{-1}$  indicates the all-trans to 13-cis isomerization of retinal. It takes place within 450 fs (compare also the Raman-bands in Fig 3a and b) and is not time resolved here. Its disappearance at



*Fig. 5: The absorbance change at  $1190\text{ cm}^{-1}$  from fig. 4 is shown in dependence of time. The results of a multi-exponential analysis which yields the rate constants and the respective amplitudes are also presented.*

about  $200\ \mu\text{s}$  indicates the deprotonation of the Schiff base. This loss of charge of the Schiffbase greatly reduces the IR-absorbance of the chromophore. (The Schiff base connects the retinal to Lys216 of the protein (fig. 6).) The deprotonation kinetics of the Schiff base agree nicely with the protonation kinetics of the counterion Asp85, which can be followed at  $1761\text{ cm}^{-1}$ <sup>20</sup> (fig. 4). Recently, for the protonrelease pathway to the protein surface a protonated hydrogen bonded network in a Grothuß like proton transfer mechanism is identified (fig6)<sup>25</sup>. The reappearance of the band at  $1190\text{ cm}^{-1}$  in the millisecond time domain indicates the reprotonation of the Schiff base. It is reprotonated by Asp96<sup>20,27</sup>. The final disappearance at  $1190\text{ cm}^{-1}$  shows the chromophore's relaxation to the all-trans BR-ground state configuration.

Based on the FTIR experiments a detailed model of the lightdriven proton pump bacteriorhodopsin is elucidated and presented in fig. 6.

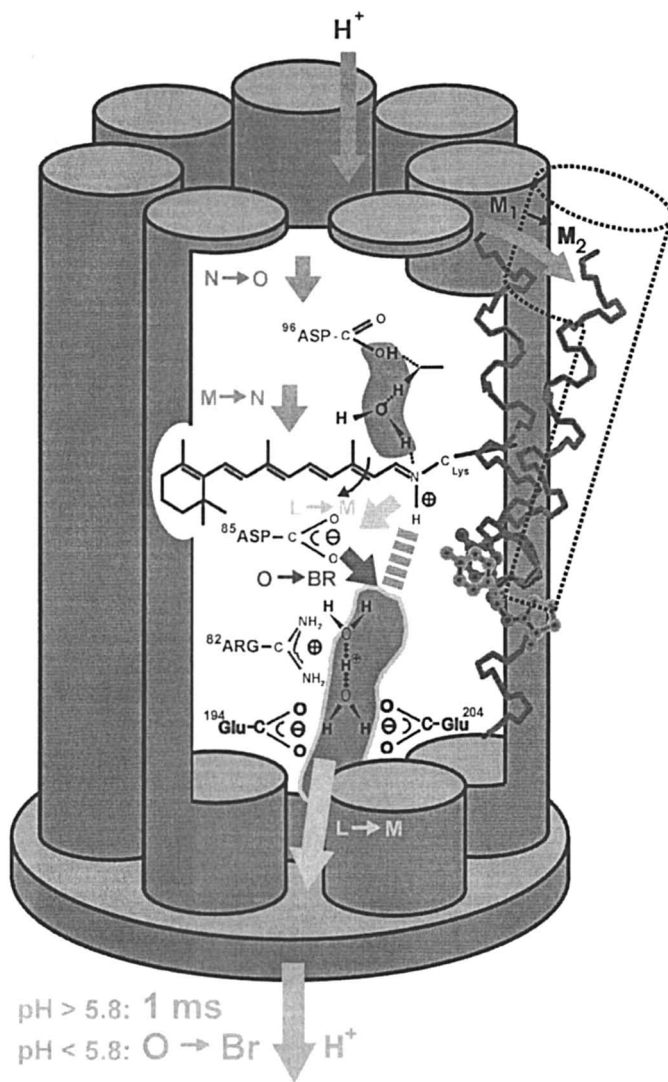
The presented results should demonstrate that time resolved FTIR difference spectroscopy in combination with molecular biological methods allows a detailed analysis of the molecular reaction mechanism of proteins. In future work the use of caged compounds will provide a powerful tool also for non-photobiological proteins.

#### **Acknowledgement:**

I would like to thank my Coworkers Dr. Benedikt Heßling, Dr. Ronald Brudler, and Robin Rammelsberg for contributions to this work.

The financial support of the DFG in the SFB394, Teilprojekt B1 and C2, and SFB 480, TP C3, is gratefully acknowledged





**Fig. 6.** The current model of the proton pump mechanism of bacteriorhodopsin, (for references, see text): After the light-induced all-trans to 13-cis retinal isomerization in the BR to K transition, the Schiff base proton is transferred to Asp85 in the L to M transition. Simultaneously a proton is released from a protonated H-bonded network spanned by internal water molecules to the extracellular site. This network is controlled by Arg 82, Glu204 and Glu194. Asp85 reprotonates the network in the O to BR reaction. The Schiff base is oriented in the  $M_1$  to  $M_2$  transition from the proton release site to the proton uptake site by a small backbone movement of helix F and determines thereby the vectoriality of proton pump. A larger backbone movement is observed in the M to N transition as compared to the  $M_1$  to  $M_2$  transition. Asp96 reprotonates the Schiff base in the M to N transition. Asp96 itself is reprotonated from the cytoplasmic site in the N to O transition.

*References*

1. N.B. Colthup, L.H. Daly and S.E. Wiberley, in *Introduction to Infrared and Raman Spectroscopy*, 3rd edn. Academic Press, Boston (1990)
2. K. Gerwert, *Current Opinion Structural Biology* **3**, 769 (1993)
3. N. Grigorieff, T.A. Ceska, K.H. Downing, J.M. Baldwin and R. Henderson, *J.Mol.Biol.* **259**, 393 (1996)
4. M.S. Braiman and K. Rothschild, *Ann.Rev.Biophys.Chem.* **17**, 541 (1988)
5. K. Gerwert, *Biochimica et Biophysica Acta* **1101**, 147 (1992)
6. F. Siebert in *Biomolecular Spectroscopy*, part A (R.J.H. Clark, R.E. Hester eds.), **1**, Wiley, Chichester, U.K. (1993)
7. B. Robert, E. Navedryk, M. Lutz in *Time Resolved Spectroscopy* (R.J.H. Clark, R.E. Hester, eds.) Wiley, New York (1989)
8. W. Mäntele, in *The Photosynthetic Reaction Center* Vol. II (J. Deisenhofer, J. Norris, eds.) Academic Press, New York, 239 (1993)
9. R. Brudler, H.J.M. de Groot, W.B.S. van Liemt, W.F. Steggerda, R. Esmeijer, P. Gast, A.J. Hoff, J. Lugtenburg, K. Gerwert, *EMBO J.* **13**, 5523 (1994)
10. J.A. McGray and D. Trentham, *Annu.Rev.Biophys.Biophys.Chem.* **18**, 239 (1989)
11. Cepus, V., Ulbrich, C., Allin, Chr., Troullier, A., Gerwert, K. *Methods in Enzymology* eds Mariott, **291**, 223 (1998)
12. A. Wittinghofer, E.F. Pai, *Trends Biochem. Sci.* **16**, 382 (1991)
13. Cepus, V., Goody, R.S., Scheidig, A.J., Gerwert, K. *Biochemistry* **37**, 10263 (1998)
14. A. Barth, W. Kreutz and W. Mäntele, *Biochim.Biophys.Acta* **1057**, 115 (1991)
15. A. Troullier, K. Gerwert, Y. Dupont, *Biophys. J.* **71**, 2970 (1996)
16. M. Lübben and K. Gerwert, *FEBS Letters* **397**, 303 (1996)
17. M. Lübben, A. Prutsch, B. Mamat and K. Gerwert, *Biochemistry* **38**, 2048 (1999)
18. M.S. Braiman, P.L. Ahl and K.J. Rothschild, *Proc. Natl. Acad. Sci. USA* **84**, 5221 (1987)
19. K. Gerwert, *Ber. Bunsenges. Phys. Chem.* **92**, 978 (1988)
20. Gerwert, G. Souvignier and B. Hess, *Proc. Natl. Acad. Sci. USA*, **87**, 9774 (1990)
21. M.S. Braiman, O. Bousche and K.J. Rothschild, *Proc. Natl. Acad. Sci. USA* **88**, 2388 (1991)
22. G. Souvignier, K. Gerwert, *Biophys. J.* **63**, 1393 (1992)
23. W. Uhmann, A. Becker, C. Taran, F. Siebert, *Appl. Spectrosc.* **45**, 390 (1991)
24. Palmer, C.J., Manning, J.L., Noda, A.E. Dorwrey, and C. Marcott, *Appl. Spectrosc.*, **45**, 12 (1991)
25. Rammelsberg, B. Heßling, H. Chorngiewski and K. Gerwert, *Applied Spectroscopy* **51**, 558 (1997)
26. Gerwert, K., Hess, B., Soppa, J. and Oesterhelt, D., *Proc. Natl. Acad. Sci. USA* **86**, 4943 (1989)
27. le Coutre, J. Tittor, D. Oesterhelt and K. Gerwert, *Proc. Natl. Acad. Sci. USA* **92**, 4962 (1995)

## Chapter 9

# Investigation of Cytokine–Receptor Interactions by Isotope-Edited FTIR Spectroscopy

Tiansheng Li, Tsutomu Arakawa, Thomas P. Horan, and Byeong Chang

Amgen Inc., M/S 8–1–C, One Amgen Boulevard, Thousand Oaks, CA 91320

In this chapter, we attempt to provide an overview for the recent progress in the application of isotope edited FTIR spectroscopy to the investigation of protein-protein interactions. To illustrate the usefulness of isotope edited FTIR spectroscopy, we shall discuss our recent work which demonstrated the application of this technique to the structural investigation of granulocyte colony stimulating factor (G-CSF) and the extracellular domain (ECD) of its receptor as well as their complex. To determine unambiguously the protein conformations of G-CSF and the receptor in the complex,  $^{13}\text{C}/^{15}\text{N}$  uniformly labeled G-CSF was prepared for FTIR to resolve the amide I' band of G-CSF from that of the receptor in the FTIR spectrum of the complex. By comparing the FTIR spectra of the isotope-labeled G-CSF and the receptor with that of the complex, we have provided spectral evidence that the AB loop region involving the unique  $3_{10}$  helix segment of G-CSF likely undergoes a conformational change to a regular  $\alpha$ -helix upon binding to the receptor. The IR data also indicate a possible minor increase in  $\alpha$ -helical conformation for the receptor in the complex. Furthermore, FTIR spectra of G-CSF, the receptor, and their complex demonstrate clearly that protein conformations of both G-CSF and the receptor are dramatically stabilized by complex formation. In summary, this work demonstrates that useful structural information of protein-protein complexes such as a cytokine/receptor complex for which there is very limited structural information available can be obtained by employing isotope-edited FTIR spectroscopy.

FT-IR spectroscopy has been proven to be an effective tool to study secondary structures of proteins in various physical states (1-8). However, extensive band overlaps in the amide I/I' region in FTIR spectra of protein-protein or protein-peptide complex curtail the usefulness of conventional FTIR spectroscopy in the

structural analysis of protein complexes. Haris and coworkers were first to demonstrate that by  $^{13}\text{C}$  uniform labeling of amide C=O groups the amide I' band of protein can be shifted to  $\sim 40\text{--}45\text{ cm}^{-1}$  lower in frequency (9), and thus making it possible to determine secondary structures of individual subunit in a protein/protein or protein/peptide complex. This method have recently been applied to several protein/protein and protein/peptide systems (10,11). Since  $^{13}\text{C}=\text{O}$  groups shift amide I' bands to approximately  $40\text{ cm}^{-1}$  lower in frequency relative to those of  $^{12}\text{C}=\text{O}$  amide groups (9, 12-14), it is possible to resolve the amide I' bands of two protein subunits in a protein/protein complex where only one of the constituents is  $^{13}\text{C}$  uniformly labeled. Until recently, there has been few reports on the application of FTIR spectroscopy to the investigation of cytokine/receptor complex because of spectral overlap in amide I/I' region. With the exception of a few ligand/receptor complexes whose crystal structures have been determined, there is limited structural information available with respect to the conformational changes in cytokine/receptor complexes. It is widely accepted that the first step in signal transduction pathway is receptor dimerization induced by the binding of cognate ligand. Conformational changes may also occur in the receptor that can modulate signal transductions (15) and in the ligand that can bring about cooperative ligand/receptor interactions. Recently, we have applied isotope-edited FTIR spectroscopy for the first time to the structural study of granulocyte colony stimulating factor (G-CSF)/receptor interactions and demonstrated its usefulness in determining unambiguously the protein secondary structures of both ligand and receptor in the signal transduction complex (16). In the previous study, we investigated the structural changes in G-CSF and its receptor upon complex formation, and showed that both G-CSF and the receptor likely undergo conformational changes in the complex (16).

## Experimental Methods

The extracellular domain (ECD) of G-CSF receptor has been expressed in CHO cells and purified as described previously (17). Uniform  $^{13}\text{C}/^{15}\text{N}$  isotope labeling of G-CSF was achieved by following the same procedure as described previously (16). Both  $^{13}\text{C}/^{15}\text{N}$  uniformly labeled G-CSF and the labeled G-CSF/receptor complex were purified as described previously (16). The purified G-CSF, the receptor and the G-CSF/receptor complex were dissolved in 10 mM sodium phosphate, pH 7.0 and buffer exchanged into  $\text{D}_2\text{O}$  solutions for FTIR data collection. FTIR spectra of protein solutions were collected on Mattson Research Series 1000 (Mattson Inc.) spectrometer. A thermal coupled temperature controller (Boulder Nonlinear Inc.) was employed to maintain specified sample temperatures during FTIR data collection. Spectral analysis including deconvolution and second derivative calculation were achieved by using SpectraCalc and Grams/32 software (Galactic Industries Corp.).

## Isotope-Edited FTIR Spectroscopy

**Analysis of Protein Secondary Structures** The isotopic shifts of amide I/I' bands in FTIR spectra of  $^{13}\text{C}$  uniformly labeled proteins in relation to those of the unlabeled proteins have been determined in a number of studies in recent years (9,

10, 13, 14, 16). Table I summarizes the frequencies of the amide I' bands corresponding to different type of secondary structures for  $^{13}\text{C}$  uniformly labeled proteins in  $\text{D}_2\text{O}$  solutions. In general,  $^{13}\text{C}$  uniform labeling of protein amide C-N group leads to a 30-40  $\text{cm}^{-1}$  red-shift in frequency for amide I/T' bands (9, 10, 13, 14, 16).

As shown in Fig. 1, the second derivative FTIR spectrum in the amide I' region of  $^{13}\text{C}/^{15}\text{N}$  uniformly labeled G-CSF (spectrum B) is compared to that of the unlabeled protein (spectrum A). The second derivative FTIR spectra in Fig. 1 display essentially identical spectral features except frequency shifts. The second derivative FTIR spectrum of the labeled G-CSF (Fig. 1, spectrum B) exhibits nearly baseline IR features between 1640-1700  $\text{cm}^{-1}$ , and only very weak IR intensities between 1620-1640  $\text{cm}^{-1}$  due to  $\beta$ -turns of the labeled protein. Additionally, the FTIR spectrum of the labeled G-CSF (Fig. 1, spectrum B) shows minimal IR intensity and is close to baseline near 1653  $\text{cm}^{-1}$  where the unlabeled protein (Fig. 1, spectrum A) gives rise to peak IR intensity at the same frequency. It is therefore concluded that isotope labeling of amide C-N groups by  $^{13}\text{C}$  and  $^{15}\text{N}$  atoms are essentially complete in  $^{13}\text{C}/^{15}\text{N}$  uniformly labeled G-CSF. In fact, mass spectroscopy indicates that overall  $^{13}\text{C}$  and  $^{15}\text{N}$  isotope incorporation in the labeled G-CSF is nearly 98% complete (data not shown). The peak of amide I' band shifts from 1653  $\text{cm}^{-1}$  for the unlabeled G-CSF to 1610  $\text{cm}^{-1}$  for  $^{13}\text{C}/^{15}\text{N}$  uniformly labeled protein, which can be assigned to  $\alpha$ -helical conformation in G-CSF (18). For the isotope labeled protein, the weak amide I' band at 1677  $\text{cm}^{-1}$  due to  $\beta$ -turns in G-CSF shifts to 1635  $\text{cm}^{-1}$ . There is also a distinct shoulder component at 1641  $\text{cm}^{-1}$  for the amide I' band of the unlabeled G-CSF, which shifts accordingly to 1599  $\text{cm}^{-1}$  for  $^{13}\text{C}/^{15}\text{N}$  uniformly labeled protein. The observed frequency shifts ( $\sim 40 \text{ cm}^{-1}$ ) of the amide I' bands of  $^{13}\text{C}/^{15}\text{N}$  uniformly labeled G-CSF are consistent with previous reports that  $^{13}\text{C}$  uniform labeling of amide carbonyl (C=O) groups lead to approximately a 30-40  $\text{cm}^{-1}$  red-shift of amide I' band (9, 10, 13, 14). The fact that the observed red-shift (nearly 40  $\text{cm}^{-1}$ ) of the amide I' bands of  $^{13}\text{C}/^{15}\text{N}$  uniformly labeled G-CSF is larger than that ( $\sim 30 \text{ cm}^{-1}$ ) of  $^{13}\text{C}$  uniformly labeled proteins also confirms the contribution from the stretching vibration of amide C-N bond to amide I' band. The amide I' band near 1641  $\text{cm}^{-1}$  in the spectrum of the unlabeled G-CSF (Fig. 1, spectrum A) may be attributed to  $3_{10}$  helix, as suggested in a number of previous studies (19-22). However, it should be noted that proteins containing  $\beta$ -strands and/or irregular structures may also give rise to IR bands from 1620 to 1640  $\text{cm}^{-1}$  (23, 24). The crystal structure of G-CSF reveals essentially no  $\beta$ -strand conformation and four  $\alpha$ -helix stretches connected by loops (25). In the connecting loop between helix A and B there is a short  $3_{10}$  helical region (from His-44 to Ser-54) which begins immediately after the first disulfide bond of G-CSF. This helical region runs almost perpendicular to the four helical bundle and packs in front of the amino terminus of helix D. NMR data of G-CSF in solution (26) also suggest a  $3_{10}$  helical region involving 8 residues in the same location as revealed in the crystal structure. Since the residues assigned to the  $3_{10}$  helix compromise only approximately 5% of the total residues in G-CSF protein, the residues in the loop structures including AB, BC and CD loops may also contribute significantly to

**Table 1**  
**Effect of  $^{13}\text{C}$  Uniform Labeling on Amide I' Bands**

Protein Secondary Structures	Amide I' Bands ( $\text{cm}^{-1}$ )		
	Non-labeled	$^{13}\text{C}$ uniformly labeled	$\Delta$ ( $\text{cm}^{-1}$ )
$\alpha$ -helix	1650-1660	1610-1620	-40
$\beta$ -strand	1620-1640	1570-1590	-50
Irregular (in $\text{D}_2\text{O}$ )	1640-1650	1590-1600	-50
Loops, Turns	1660-1690	1620-1640	-40

Note: The listed isotopic frequency shifts of amide I' bands are taken from previous literature (9-14).

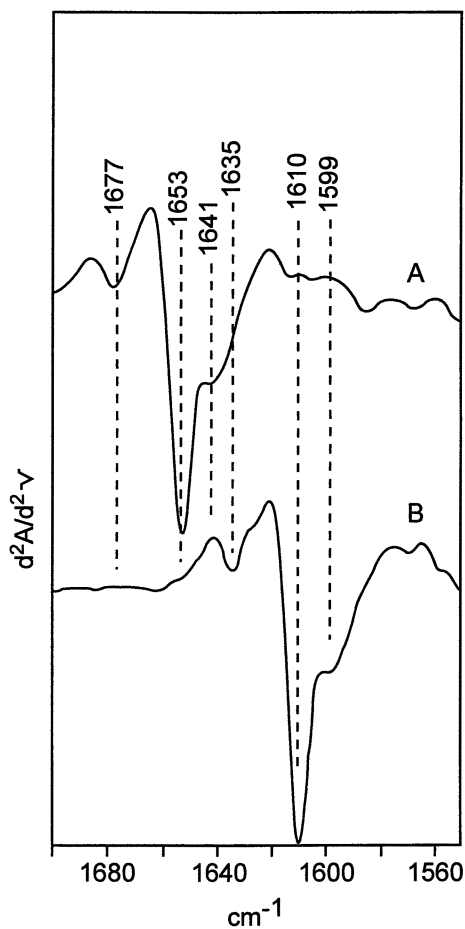


Figure 1. Second derivative FTIR spectra in the amide I' region 1600-1700 cm<sup>-1</sup> of the unlabeled G-CSF (spectrum A) and <sup>13</sup>C/<sup>15</sup>N uniformly labeled G-CSF (spectrum B). The proteins are in D<sub>2</sub>O buffer containing 10 mM sodium phosphate, 50 mM NaCl, pD 7.0. (Adapted from reference 9. Copyright 1992 American Chemical Society.)

the amide I' shoulder component near  $1641\text{ cm}^{-1}$  and to the IR intensities in the broadening feature between  $1620$  and  $1641\text{ cm}^{-1}$  in the amide I' region of the unlabeled G-CSF (Fig. 1, spectrum A). As discussed in a latter section, the isotopic shift of the amide I' band of the labeled G-CSF from  $1653$  to  $1610\text{ cm}^{-1}$  (Fig. 1, spectrum B) eliminates most of the spectral interference from G-CSF to the amide I' bands of the ligand-bound receptor in the complex, which makes it possible to attribute more definitively the spectral changes in those regions to the conformational changes in the receptor.

### Thermal Stabilities of G-CSF and Its Receptor in Isolated Forms

Structural changes associated with thermal denaturation of proteins in solutions can be conveniently monitored by employing FTIR spectroscopy (3, 27-30). During thermal transition from native to denatured states, many globular proteins in solutions give rise to intense IR amide I/I' bands near  $1620$  and  $1680\text{ cm}^{-1}$ , which are generally attributed to anti-parallel  $\beta$ -sheets present in thermally denatured proteins (3, 27-30). Thermal stabilities of  $^{13}\text{C}/^{15}\text{N}$  uniformly labeled G-CSF and its receptor in  $\text{D}_2\text{O}$  solutions at neutral pH were monitored by using FTIR spectroscopy. Fig. 2 shows the second derivative FTIR spectra of the isotope labeled G-CSF at temperatures from  $20$  to  $90^\circ\text{C}$ . Thermal transition of  $^{13}\text{C}/^{15}\text{N}$  uniformly labeled G-CSF appears to be sharp and to start at around  $40^\circ\text{C}$ , as evidenced by the appearance of the amide I' band at  $1573\text{ cm}^{-1}$  and the decrease in IR intensity at  $1610\text{ cm}^{-1}$ . Considering the fact that there is an approximately  $40\text{ cm}^{-1}$  red-shift for the amide I' bands of  $^{13}\text{C}/^{15}\text{N}$  uniformly labeled G-CSF in comparison to those of the unlabeled protein, the observed spectral changes in the FTIR spectra in Fig. 2 typify the structural changes associated with the thermal denaturation of a globular protein in solution. In addition, there is a significant band broadening in the spectral region  $1580$ - $1610\text{ cm}^{-1}$  at or above  $40^\circ\text{C}$ , suggesting the presence of a significant amount of irregular structures during the thermal transition of G-CSF. The disappearance of the  $1610\text{ cm}^{-1}$  band above  $40^\circ\text{C}$  suggests that most of the  $\alpha$ -helical conformation in G-CSF are converted to intermolecular  $\beta$ -sheet and irregular structures when the temperature is above  $40^\circ\text{C}$ , as indicated by the appearance of the IR bands at  $1573$  and  $1600\text{ cm}^{-1}$ . The thermal unfolding of unlabeled G-CSF exhibits identical spectral changes to that of the labeled protein except for isotopic frequency shifts (data not shown). Thermal stability of G-CSF receptor in isolated form can also be monitored by using FTIR spectroscopy. Fig. 3 shows the second derivative spectra of the isolated G-CSF receptor (ECD) at temperatures from  $20$  to  $90^\circ\text{C}$ . The second derivative FTIR spectra display a large conformation change near  $55^\circ\text{C}$ , as evidenced by the appearance of the  $1624\text{ cm}^{-1}$  band. The melting transition of the free receptor occurs at approximately  $50^\circ\text{C}$  and is marked by the decreasing intensity of the  $1636\text{ cm}^{-1}$  band and the increasing intensity of the  $1624\text{ cm}^{-1}$  band. The temperature dependence of the receptor appears to be more complex than that of G-CSF, which may be expected considering of the multi-domain structures in the receptor (17). Fig. 3 also shows that the  $1686\text{ cm}^{-1}$  band due to  $\beta$ -turns of the receptor is shifted to  $1681\text{ cm}^{-1}$  when the receptor becomes unfolded. These changes in the amide I' bands of the receptor are consistent with the conversion of



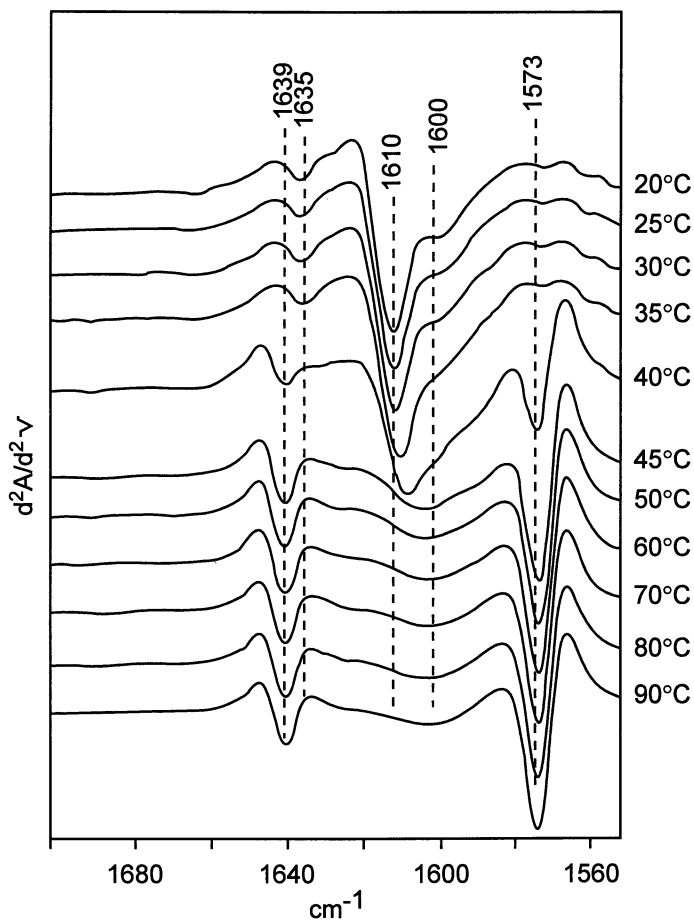


Figure 2. Second derivative FTIR spectra in the amide I' region 1600-1700  $\text{cm}^{-1}$  of  $^{13}\text{C}/^{15}\text{N}$  uniformly labeled G-CSF at temperatures from 20 to 90°C. Other conditions are the same as in Fig. 1. (Adapted from reference 9. Copyright 1992 American Chemical Society.)

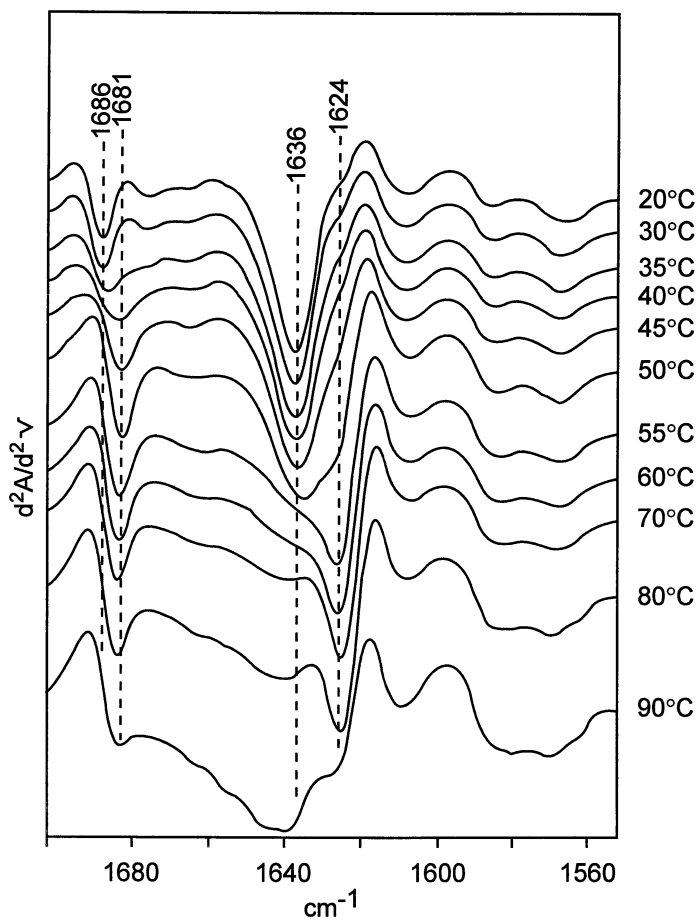


Figure 3. Second derivative FTIR spectra in the amide I' region 1600-1700  $\text{cm}^{-1}$  of G-CSF receptor at temperatures from 20 to 90°C. Other conditions are the same as in Fig. 1.

(Adapted from reference 9. Copyright 1992 American Chemical Society.)

intramolecular  $\beta$ -strands in native proteins to extended intermolecular  $\beta$ -sheets in thermally denatured proteins (27-30).

**Analysis of G-CSF/Receptor Interactions** Fig. 4 compares the second derivative FTIR spectra of the isolated G-CSF receptor (spectrum A), the unlabeled G-CSF/receptor complex (spectrum B) and the isotope labeled G-CSF/receptor complex (spectrum C). As shown in the FTIR spectrum of the complex containing the unlabeled G-CSF (Fig. 4, spectrum B), there is significant spectral overlap between amide I' band of the unlabeled G-CSF ( $1653\text{ cm}^{-1}$ ) and that of the receptor ( $1636\text{ cm}^{-1}$ ), which partially obscures the  $1653\text{ cm}^{-1}$  band of G-CSF and makes it difficult to determine quantitatively the spectral changes in amide I' bands of the complex. As a result, it is difficult to draw any meaningful conclusion regarding to the effect of complex formation on the structures of either G-CSF or the receptor. However, the amide I' band at  $1610\text{ cm}^{-1}$  due to the isotope labeled G-CSF is well resolved from that of the receptor in the second derivative FTIR spectrum of the complex containing  $^{13}\text{C}/^{15}\text{N}$  uniformly labeled G-CSF (Fig. 4, spectrum C), with the main amide I' band of the labeled G-CSF appearing at  $1610\text{ cm}^{-1}$  and well separated from that of the receptor.

Since  $^{13}\text{C}/^{15}\text{N}$  uniformly labeled G-CSF makes negligible spectral contributions to the amide I' region  $1620\text{-}1700\text{ cm}^{-1}$  in comparison to that of the receptor, it is now possible to compare the spectrum of the isolated receptor (Fig. 4, spectrum A) to that of the receptor in the complex (Fig. 4, spectrum C) with little interference from the amide I' band of the labeled G-CSF. There is no major difference between the amide I' bands of the ligand-bound receptor and the free receptor, i.e., both exhibit intense amide I' band at  $1636\text{ cm}^{-1}$  and weak one at  $1686\text{ cm}^{-1}$ , which are consistent with intramolecular  $\beta$ -strands as the dominant secondary structure. This is in agreement with the recent CD data showing that the dominant secondary structure of G-CSF receptor is  $\beta$ -strand (17). However, there is a weak but significant intensity at  $1652\text{ cm}^{-1}$  in the second derivative FTIR spectrum of the ligand-bound receptor (Fig. 4, spectrum C) whereas that of free receptor (Fig. 4, spectrum A) shows minimal intensity at the same frequency. Two possible explanations for the appearance of the weak IR band at  $1652\text{ cm}^{-1}$  in the spectrum of the labeled G-CSF/receptor complex would be due to either some residual unlabeled G-CSF or a slight increase in  $\alpha$ -helical content of the receptor. Since the unlabeled G-CSF (Fig. 1, spectrum A) gives rise to a peak IR intensity around  $1653\text{ cm}^{-1}$ , residual unlabeled G-CSF bound to the receptor would generate a weak IR band near  $1653\text{ cm}^{-1}$  in the spectrum of the complex. However, we think this scenario is less likely than the possibility of a real conformational change in the ligand-bound receptor for the following reasons. First, the second derivative FTIR spectrum of  $^{13}\text{C}/^{15}\text{N}$  uniformly labeled G-CSF (Fig. 1 spectrum B) exhibits no IR intensity at  $1653\text{ cm}^{-1}$ , while the spectrum of the complex shows a sharp and discrete band near the same frequency (Fig. 4, spectrum C). Secondly, the amide I' bands of G-CSF are only approximately one-third as intense as those of the receptor (Fig. 4, spectra B and C) because of their relative molar ratio in the complex (e.g.,  $\sim 18\text{ kDa}$  for G-CSF versus  $\sim 68\text{ kDa}$  for the receptor considering

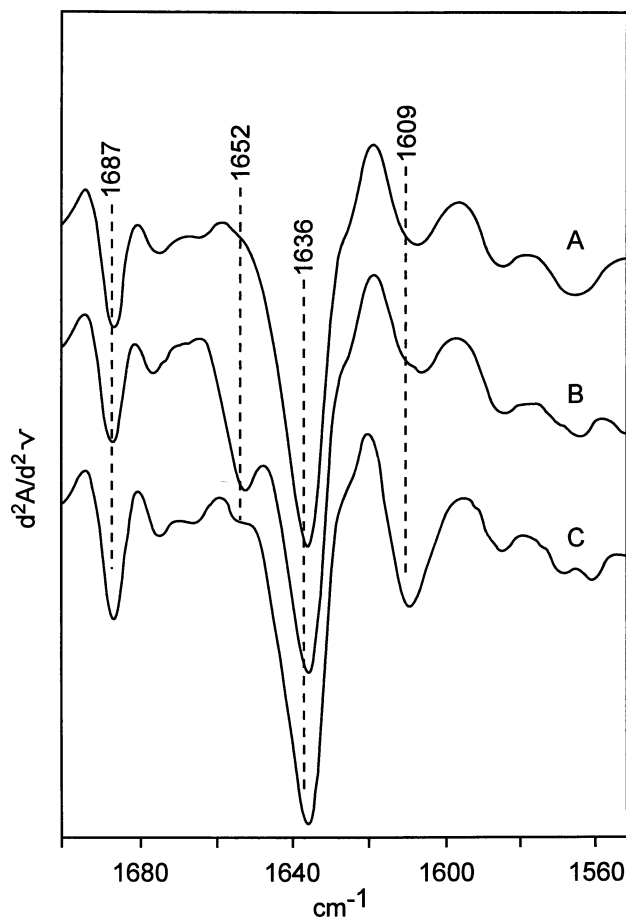


Figure 4. Second derivative FTIR spectra in the amide I' region 1600-1700 cm<sup>-1</sup> of the isolated G-CSF receptor (spectrum A), the unlabeled G-CSF/receptor complex (spectrum B) and <sup>13</sup>C/<sup>15</sup>N uniformly labeled G-CSF/receptor complex (spectrum C). Other conditions are the same as in Fig. 1.

(Adapted from reference 9. Copyright 1992 American Chemical Society.)

amino acid only). As shown in the second derivative spectrum of the complex (Fig. 4, spectrum C), however, the relative intensity ratio of the amide I' band at  $1652\text{ cm}^{-1}$  to that of the band at  $1609\text{ cm}^{-1}$  ( $I_{1652}/I_{1609}$ ) exceeds significantly that of the labeled G-CSF alone (Fig. 1, spectrum B), which suggests that the amide I' band at  $1652\text{ cm}^{-1}$  in the spectrum of the labeled G-CSF/receptor complex is due to mostly the receptor rather than to residual unlabeled G-CSF. Therefore, it is likely that binding of G-CSF may induce the formation of a small amount of  $\alpha$ -helix in the receptor. This conclusion is also consistent with the published CD data suggesting a small increase in  $\alpha$ -helical conformation in the receptor upon binding of G-CSF (17).

To examine the structural features of G-CSF in free and complex forms, both deconvolved (Fig. 5, left panel) and second derivative (Fig. 5, right panel) FTIR spectra of the labeled G-CSF in isolated form are compared with those of the labeled G-CSF in the complex. As shown in Fig. 5 (left panel, spectra C and D), the receptor exhibits nearly no IR intensity at  $1598\text{ cm}^{-1}$  where the amide I' band of the labeled G-CSF in isolated form displays a shoulder component (Fig. 5, left panel, spectrum A). Though the presence of an IR band near  $1610\text{ cm}^{-1}$  due to aromatic (Tyr and Phe) residues of the receptor increases the overall intensity of the amide I' band of the labeled G-CSF in the complex, it should not affect the overall shape of the shoulder component at  $1598\text{ cm}^{-1}$  because of the frequency separation between the two bands. Consequently, the IR band at  $1610\text{ cm}^{-1}$  due to the receptor should have little impact on the shoulder component at  $1598\text{ cm}^{-1}$  of the labeled G-CSF in the complex. Consistent with this notion, additional IR data demonstrate that non-specific interactions of G-CSF with another protein (murine tumor necrosis factor, mTNF) containing mostly  $\beta$ -strands and generating an IR band at  $1610\text{ cm}^{-1}$  do not affect the shoulder component at  $1598\text{ cm}^{-1}$  of the amide I' band of the labeled G-CSF (data not shown). Therefore, the absence of the shoulder component at  $1598\text{ cm}^{-1}$  in the spectra of the receptor-bound G-CSF, as shown in both deconvolved (Fig. 5, left panel, spectrum B) and second derivative spectra (Fig. 5, right panel, spectrum B), likely results from conformational changes in G-CSF rather than from any spectral interference of the receptor in the complex. As discussed in the previous section, the shoulder component at  $1598\text{ cm}^{-1}$  of the amide I' band of the labeled G-CSF is probably due to mostly the  $3_{10}$ -helix and the long loops connecting the four  $\alpha$ -helices. The disappearance of this shoulder component at  $1598\text{ cm}^{-1}$  may indicate conformational changes involving the  $3_{10}$ -helix and the surrounding loop residues of G-CSF upon complex formation. The amide I' band of the receptor-bound G-CSF (Fig. 5, left panel, spectrum B and right panel, spectrum B) also becomes broader and is approximately  $2\text{ cm}^{-1}$  lower in frequency than that of the free G-CSF (Fig. 5, left panel, spectrum A; right panel, spectrum A). Furthermore, the spectral features in the region  $1550\text{--}1585\text{ cm}^{-1}$  of the two complexes containing unlabeled and isotope labeled G-CSF proteins are very similar (Figure 4, spectra B and C), which indicates that the conversion of the loops involving the  $3_{10}$  helix to disordered structures or  $\beta$ -strands is unlikely. Therefore, it is likely that binding of G-CSF to the receptor may lead to the loss of the  $3_{10}$  helix in favor of  $\alpha$ -helix, and that some adjacent loop residues

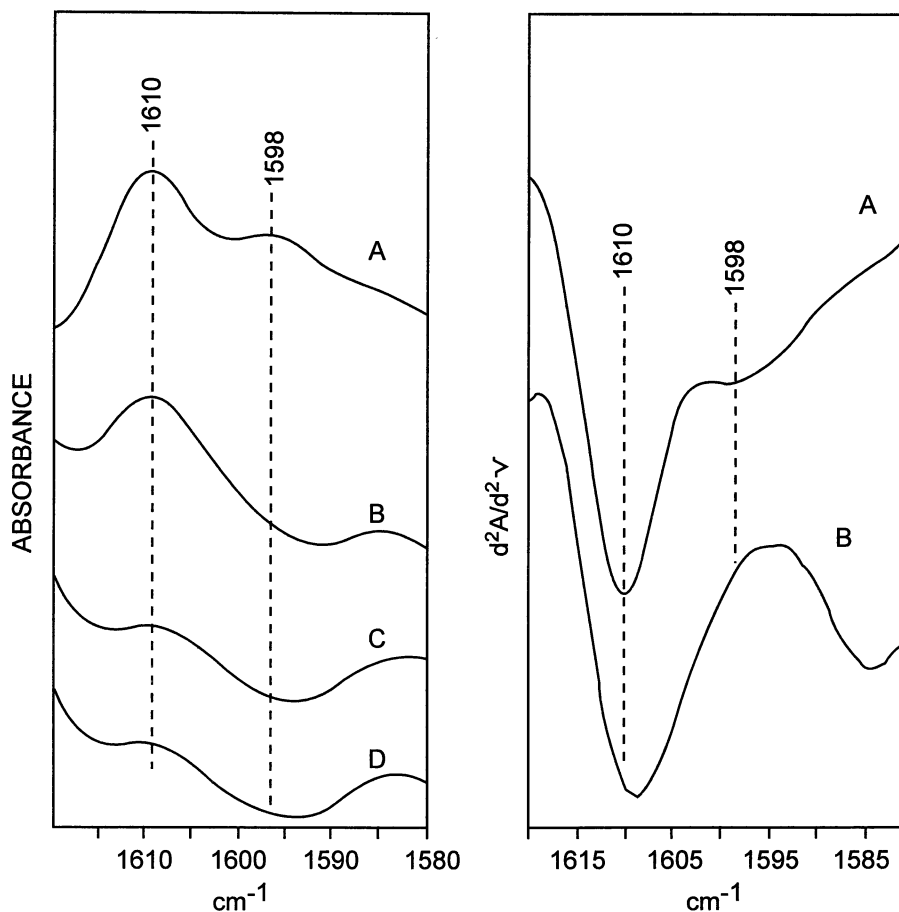


Figure 5. Comparison of the amide I' band of  $^{13}\text{C}/^{15}\text{N}$  uniformly labeled G-CSF in isolated form with that of the labeled G-CSF in the complex. Left panel: deconvoluted FTIR spectra in the region 1580-1620  $\text{cm}^{-1}$  of the labeled G-CSF in isolated form (spectrum A), the labeled G-CSF in the complex (spectrum B), the isolated receptor (spectrum C) and the complex containing the unlabeled G-CSF (spectrum D). Right panel: Second derivative FTIR spectra in the region 1580-1620  $\text{cm}^{-1}$  of the labeled G-CSF in isolated form (spectrum A) and in the complex (spectrum B). Other conditions are the same as in Fig. 1.

(Adapted from reference 9. Copyright 1992 American Chemical Society.)

near the two ends of the  $3_{10}$  helix are most likely involved as well. The crystal structure of G-CSF reveals a number of residues before and after the  $3_{10}$  helix not incorporated in either helix A or B, and these residues could be the potential candidates for the extension of  $\alpha$ -helix in the receptor-bound G-CSF. Conformational changes in this region suggest that the AB loop region containing the  $3_{10}$  helix may be involved in specific interactions with the receptor. The crystal structure of G-CSF shows that at the beginning of the  $3_{10}$  helix there is a potential salt bridge between Glu-47 and His-44. The  $3_{10}$  helical region is potentially stabilized by the interaction of these two residues. However, this salt bridge can be broken as the carboxylate group of Glu-47 becomes protonated at low pH or by interacting with the receptor, which may in turn destabilize the  $3_{10}$  helix in favor of a regular  $\alpha$ -helix. Interestingly, the conformational changes of G-CSF observed here share somewhat similar features with interferon- $\gamma$  upon receptor binding. The crystal structure of interferon- $\gamma$ /receptor complex has revealed that the AB loop residues of interferon- $\gamma$  undergo conformational change from disordered structures to  $3_{10}$  helix upon complex formation (31).

**Effect of Complex Formation on Protein Stability** To determine the structural impact of complex formation on G-CSF and the receptor, we have collected FTIR spectra of G-CSF/receptor complex containing the labeled G-CSF in the temperature range from 20 to 90°C. Fig. 6 shows temperature dependence of the second derivative FTIR spectra of the complex containing the labeled G-CSF and the receptor at temperatures from 20 to 90°C. As discussed previously, the melting transitions of both the receptor and the labeled G-CSF in the complex can be monitored by the intensities of amide I' bands at 1624  $\text{cm}^{-1}$  and 1573  $\text{cm}^{-1}$ , which indicate the thermal denaturation of the receptor and the labeled G-CSF, respectively (see also Figs. 2 and 3). The frequency shift of the 1686  $\text{cm}^{-1}$  band to 1681  $\text{cm}^{-1}$  can also be used to monitor the thermal unfolding of the receptor in the complex since the isotope labeled G-CSF gives no spectral interference in this region (see also Fig. 1, spectrum B and Fig. 2). As shown in Fig. 6, there is no significant increase in the intensity of the 1573  $\text{cm}^{-1}$  band at temperatures up to 65°C for G-CSF in the complex and the median transition point appears to be in the range of 70-75°C. In comparison to the free G-CSF (Fig. 2), it is clear that thermal denaturation of G-CSF in the complex is delayed by approximately 30°C. The  $\alpha$ -helical conformation in G-CSF appears to be entirely stabilized by receptor binding, since the amide I' band centered at 1609  $\text{cm}^{-1}$  of the labeled G-CSF in the complex shows little changes in terms of intensity and band width from 20 to 65°C (Fig. 6). This may also suggest that residues of G-CSF involved in specific interactions with its receptor could play vital roles in stabilizing the structural integrity of G-CSF. Furthermore, Fig. 6 shows that the thermostability of the receptor in the complex is also increased in comparison to that of the free receptor (Fig. 3). Though the intensity increase of the 1624  $\text{cm}^{-1}$  band due to the receptor in the complex exhibits similar temperature profile as that of the free receptor, we believe this is due to the melting of the receptor domains which do not interact with the ligand. This proposal is supported by the following observations. First, the peak frequency of the 1636  $\text{cm}^{-1}$  band indicative of native  $\beta$ -strands of the

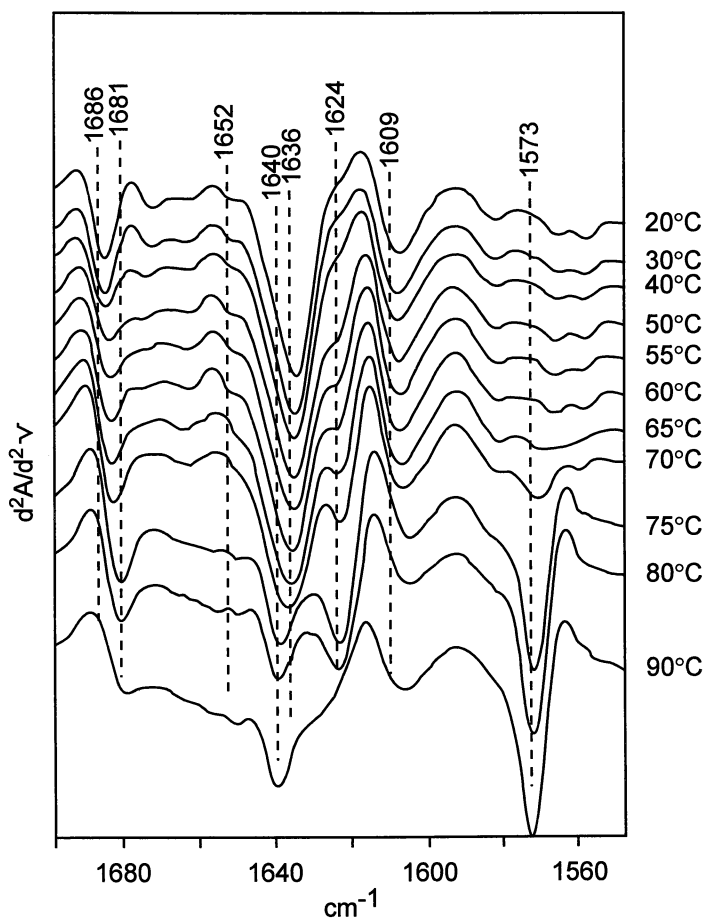


Figure 6. Second derivative FTIR spectra in the amide I' region 1540-1700  $\text{cm}^{-1}$  of  $^{13}\text{C}/^{15}\text{N}$  uniformly labeled G-CSF/receptor complex at temperatures from 20 to 90°C. Other conditions are the same as in Fig. 1. (Adapted from reference 9. Copyright 1992 American Chemical Society.)



ligand-bound receptor remains unchanged even at 65°C (Fig. 6) while for the free receptor it is nearly completely shifted to 1624 cm<sup>-1</sup> at around 55°C (Fig. 3). This suggests that only part of the native  $\beta$ -structure of the ligand-bound receptor becomes unfolded at 65°C but the ligand-free receptor unfolds entirely at this temperature. Secondly, the temperature profile of the 1686 cm<sup>-1</sup> band of the ligand-bound receptor is markedly different from that of the ligand-free receptor. While the peak of 1686 cm<sup>-1</sup> band is completely shifted to 1681 cm<sup>-1</sup> at 55°C for the ligand-free receptor (Fig. 3), it is gradually red-shifted with increasing temperature and is completely shifted to 1681 cm<sup>-1</sup> at 75°C for the ligand-bound receptor (Fig. 6). This is consistent with native  $\beta$ -strands being more thermal stable in the complex than in the free receptor. Thus, the melting temperatures of both G-CSF and the receptor in the complex have been dramatically increased in comparison to those of each subunit in isolated form.

## Conclusions

In this chapter, we discussed the usefulness of FTIR spectroscopy in the study of protein-protein complexes such as a cytokine-receptor complex. It was shown that <sup>13</sup>C/<sup>15</sup>N uniform isotope labeling of G-CSF can lead to significant frequency separation between the amide I' bands of the isotope labeled protein and those of the unlabeled receptor in the FTIR spectrum of their complex. Consequently, definitive spectral assignments can be achieved and meaningful spectral interpretations regarding the secondary structures of individual subunits in a protein-protein complex are made possible. It is also noted that thermal denaturation of both the isotope labeled subunit and the unlabeled one in a protein/protein complex can be simultaneously monitored by FTIR spectroscopy because of the large frequency separation between amide I/I' bands of the labeled subunit and those of the unlabeled one. It is therefore possible to access the effect of complex formation on the structural integrity of protein subunit by employing isotope edited FTIR spectroscopy.

## References

1. Byler, D. M. & Susi, H. *Biopolymers* **1986**, *25*, 469-487.
2. Byler, D. M. & Susi, H. *J. Industrial Microbiol.* **1988**, *3*, 73-84.
3. Arrondo, J. L. R., Muga, A., Castresana, J., & Goni, F. M. *Prog. Biophys. Mol. Biol.* **1993**, *59*, 23-56.
4. Prestrelski, S. J., Tedeschi, N., Arakawa T. and Carpenter, J. F. *Biophys. J.* **1993**, *65*, 661-671.
5. Dong, A., Prestrelski, S. J., Allison, S. D. and Carpenter, J. F. *J. Pharm. Sci.* **1995**, *84*, 415-424.
6. Jackson, M. & Mantsch, H. H. *Critical Rev. Biochem. Mol. Biol.* **1995**, *30*, 95-120.
7. Haris, P. I. & Chapman, D. *Biopolymers* **1995**, *37*, 251-263.
8. Prestrelski, S. J., Pikal, K. A. and Arakawa, T. *Pharm. Res.* **1995**, *12*, 1250-1259.
9. Haris, P. I., Robillard, G. T., van Dijk, A. A., & Chapman, D. *Biochemistry* **1992**, *31*, 6279-6284.

10. Zhang, M., Fabian, H., Mantsch, H. H., & Vogel, H. J. *Biochemistry* **1994**, *33*, 10883-10888.
11. Ludlam, C. F. C., Sonar, S., Lee, C.-P., Coleman, M., Herzfeld, J., RajBhandary, U. L., & Rothschild, K. J. *Biochemistry* **1995**, *34*, 2-6.
12. Hubner, W., Mantsch, H. H., & Casal, H. L. *Appl. Spectrosc.* **1990**, *44*, 732-734.
13. Tadesse, L., Nazarbachi, R., & Walters, L. *J. Am. Chem. Soc.* **1991**, *113*, 7036-7037.
14. Martinez, G. V., Fiori, W. R., & Millhauser, G. *Biophys. J.* **1994**, *66*, A65.
15. Domanski, P., Nadean, O. W., Platanius, L. C., Fish, E., Kellum, M., Pitha, P., & Colmanici, O. R. *J. Biol. Chem.* **1998**, *273*, 3144-3147.
16. Li, T., Horan, T., Osslund, T., Stearns, G. & Arakawa, T. *Biochemistry* **1997**, *36*, 8849-8857.
17. Horan.T., Wen, J., Narhi, L. O., Parker, V., Arakawa, T. *Biochemistry* **1996**, *35*, 4886-4896.
18. Surewicz, W. K., Mantsch, H Tadesse, L., Nazarbachi, R., & Walters, L. *J. Am. chem. Soc.* **1991**, *113*, 7036-7037.
19. Dwivedi, A. M., Krimm, S., & Malcolm, B. R. *Biopolymers* **1984**, *23*, 2026-2065.
20. Holloway, P. & Mantsch, H. H. *Biochemistry* **1989**, *28*, 931-935.
21. Prestrelski, S. J., Byler, D. M., & Thompson, M. P. *Int. J. Peptide Protein Res.* **1991**, *37*, 508-512.
22. Miick, S. M., Martinez, G. V., Fiori, W. R., Todd, A. P. & Millhauser, G. L. *Nature* **1992**, *359*, 653-655.
23. Krimm, S. & Bandekar, J. *Adv. Prot. Chem.* **1986**, *38*, 181-364.
24. Susi, H. & Byler, D. M. *Arch. Biochem. Biophys.* **1987**, *258*, 465-469.
25. Hill, C. P., Osslund, T. D., & Eisenberg, D. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5167-5171.
26. Werner, J. M., Breeze, A. L., Kara, B., Rosenbrock, G., Boyd, J., Soffe, N., & Campbell, I. D. *Biochemistry* **1994**, *33*, 7184-7192.
27. Clark, A. H., Saunderson, D. H. & Suggett, A. *Int. J. Pept. Prot. Res.* **1981**, *17*, 353-364.
28. Yang, P. W., Matsh, H. H., Arrondo, J. L., Saint-Girons, I., Guillou, Y., Cohen, G. N., & Barzu, O. *Biochemistry* **1987**, *26*, 2706-2711.
29. Seshadri, S., Oberg, K. A. , & Fink, A. L. *Biochemistry* **1994**, *33*, 1351-1355.
30. van Stokkum, I. H. M., Linsdell, H., Hadden, J. M., Haris, P. I., Chapman, D., & Bloemendal, M. *Biochemistry* **1995**, *34*, 10508-10518.
31. Walter, M. R., Windsor, W. T., Nagabhushan, T. L., Lundell, D. J., Lunn, C. A., Zauodny, P. J. & Narula, S. K. *Nature* **1995**, *376*, 230-235.

# Author Index

- Arakawa, Tsutomu, 167  
Byler, D. M., 145  
Cai, S., 117  
Chang, Byeong, 167  
Fink, A. L., 132  
Gerwert, Klaus, 159  
Goormaghtigh, E., 96  
Haris, Parvez, 54  
Horan, Thomas P., 167  
Khurana, R., 132  
Krimm, Samuel, 38  
Lee, D. L., 145  
Li, Tiansheng, 167  
Oberg, K. A., 132  
Randall, C. S., 145  
Ruysschaert, J.-M., 96  
Seshadri, S., 132  
Singh, Bal Ram, 2, 117

# Subject Index

## A

$\alpha$ -helical  $\beta$ -sheet proteins, analysis using IR spectroscopy, 19,27–28

$\alpha$ -helical protein, analysis using IR spectroscopy, 19,26*f*

### Advantages

Fourier-transform IR spectroscopy, 54–92

partial least-squares method, 117–128

Aggregation of peptides, problem, 71–80

### Amide I bands

analysis of peptides and proteins, 3–34

characteristics of peptide group modes, 40–44

description, 146

partial least-squares method for secondary structure determination of proteins, 117–128

Amide II bands, characteristics of peptide group modes, 44

### Amide III bands

analysis of peptides and proteins, 3–34

characteristics of peptide group modes, 44–46

partial least-squares method for secondary structure determination of proteins, 117–128

Amide V, characteristics of peptide group modes, 46

Amide A, characteristics of peptide group modes, 39–40

Amide hydrogen/deuterium exchange kinetics, tertiary structural changes during ligand binding, 104–108

Amyloid fibrils, determination of secondary structures, 139–141,142*f*

Analysis, Fourier-transform IR spectroscopy, 56–59

Apolipoprotein III  $\alpha$ -helices, alignment in complexes with dimyristoylphosphatidylcholine, 100–102

### Apoprotein E

function, 103

low-density lipoprotein receptor active conformation, 103–104

Applications, studies, 132–181

Attenuated total reflectance accessory critical angle of incidence, 6

crystal, 6

depth of penetration, 6

description, 4,6,7*f*

Attenuated total reflectance Fourier-transform IR spectroscopy for secondary structure determination in protein aggregates

advantages, 142–143

amyloid fibrils, 139–141,142*f*

cell designs, 134

data analysis, 135–136

folding aggregates, 137–139

inclusion bodies, 137–139

partially folded intermediates as aggregation precursors, 136

procedure, 134–135

protein aggregation

mechanism model, 141–142

ordered and disordered deposits, 133–134

problems, 133–134

secondary structure induction via partially folded intermediate aggregation, 137

Attenuated total reflection IR spectroscopy

advantages, 97,114

orientation determination in proteins and lipids

low-density lipoprotein receptor

- active conformation of apolipoprotein E, 103–104
  - potential, 100–102
  - principles, 97–99, 101*f*
  - tertiary structural changes
    - class III and IV exchange kinetics, 111–113
    - exchanging proton assignment to secondary structure
      - class I and II exchange kinetics, 108, 110–111
      - exchange kinetics comparison at secondary and molecular level, 113–114, 115*f*
    - ligand binding, 104–108
- B**
- $\beta$ -sheet proteins, analysis using IR spectroscopy, 16, 18–25
  - Bacteriorhodopsin, time-resolved Fourier-transform IR spectroscopy of molecular reaction mechanisms, 159–165
  - BPTI
    - amide I region after deconvolution, 108, 109*f*
    - class I and II exchange kinetics, 108, 110–111
    - class III and IV exchange kinetics, 111–113
    - comparison of exchange kinetics at secondary and molecular level, 113–114, 115*f*
- C**
- Chromophores, Fourier-transform IR spectroscopy, 79, 81–82
  - $\alpha$ -Chymotrypsin, IR spectroscopy, 4, 5*f*
  - Circular dichroism, limitations, 96–97
  - Classic least-squares method, description, 119–120
  - Complex formation, role in protein stability, 179–181
  - Critical angle of incidence, definition, 6

Cytokine–receptor interactions, isotope-edited Fourier-transform IR spectroscopy, 167–181

## D

- Deconvolution, data analysis, 135–136
- Depth of penetration, definition, 6
- Developments, studies, 132–181
- Dichroic ratio, definition, 97–98
- Differential scanning calorimetry, thermal denaturation of elastase with guanidinium chloride, 145–156
- Dimyristoylphosphatidylcholine, alignment in complexes with apolipoprotein III  $\alpha$ -helices, 100–102
- Disadvantages
  - Fourier-transform IR spectroscopy 54–92
  - partial least-squares method, 117–128

## E

- Elastase
  - reasons for interest, 147
  - thermal denaturation with guanidinium chloride, 145–156
- Environment, role in structural analysis of peptides using Fourier-transform IR spectroscopy, 85–89
- Exchanging protons, assignment to secondary structure, 107–115
- Experimental directions, interpretation of IR spectra, 48–51

## F

- Factor analysis, data analysis, 135–136
- Folding aggregates, secondary structure determination, 137–139
- Force fields, role in interpretation of IR spectra, 47–48
- Fourier-transform IR spectroscopy
  - advantages, 54–55
  - applications, 167–168
  - cytokine–receptor interactions, *See*

Isotope-edited Fourier-transform IR spectroscopy of cytokine–receptor interactions

proteins and peptides, 3–34

structural analysis of peptides

  complexity of membrane protein spectra

  comparison of membrane protein–amide I bands, 89–90

  proteolysis for spectra

    simplification, 90–91

environmental effect

  amide I frequency for viral coat protein, 87,88*f*

  coil–helix transition, 85,87

  proteins in crystalline and solution states, 87,89

future prospects, 91–92

hydrogen–deuterium exchange analysis

  amide I bands, 84–85,86*f*

  conformational change detection, 83–84

  description, 83

  membrane proteins, 84

measurement and analysis of spectra

  impurity detection, 58–59

  qualitative and quantitative analysis techniques, 57–58

  recording using transmission vs. attenuated total reflectance spectroscopy, 56–57

  water and water vapor, 55–56

peptide aggregation problem

  example, 71,72*f*

  ion channel peptides, 73,76–80

  viral coat proteins, 71,73,74–75*f*

peptide secondary structure analysis

  absorption

$\alpha$ -helical structures, 60–69

$\beta$ -sheet structure, 68,70

$\beta$ -turn and random coil structure, 70–71

  amide band–protein secondary structure relationship, 59–60

polarized IR spectroscopy, 91

side chains and chromophores, 79,81–82

## G

Granulocyte colony stimulating factor–receptor interactions, isotope-edited Fourier-transform IR spectroscopy, 167–181

Guanidinium chloride, thermal denaturation of elastase, 145–156

Guanidinium ion, use for unfolding and denaturing of proteins, 147

## H

Hydrogen–deuterium exchange, structural analysis of peptides, 83–86

## I

Inclusion bodies, secondary structure determination, 137–139

Infrared spectroscopy

  potential, 38

  use for protein conformational analysis, 54–55

  peptides and proteins

    advantages, 14

    applications

      protein secondary structure by amide I vs. amide III

      critical assessment, 30–32

      description, 16

      examples

$\alpha$ -helical  $\beta$ -sheet proteins, 19,27–28

$\alpha$ -helical protein, 19,26*f*

$\beta$ -sheet proteins, 16,18–25

      surface protein adsorption analysis, 15–16,17*f*

$\alpha$ -chymotrypsin, 4,5*f*

  data collection

    attenuated total reflectance accessory, 4,6,7*f*

- experimental conditions, 6,8
    - difficulties in amide I vibration region analysis, 4
    - disadvantages, 14
    - interpretation of spectra, 38–51
    - modern advanced techniques, 32,34
    - previous studies, 3–4
    - sensitivity, 9–14
    - spectral processing, 8–9
  - Intensity predictions, role in interpretation of IR spectra, 48,49*f*
  - Interpretation of IR spectra of peptides and proteins
    - approach, 38–39
    - characteristics of peptide group modes
      - amide I, 40–44
      - amide II, 44
      - amide III, 44–46
      - amide V, 46
      - amide A, 39–40
    - background, 39
    - future directions
      - experimental directions
        - isotopic substitution, 48–50
        - solution studies, 50–51
      - protocol, 46–47
      - theoretical directions
        - improved force fields, 47–48
        - intensity predictions, 48,49*f*
  - Ion channel peptides, aggregation problem, 73,76–80
  - Isotope-edited Fourier-transform IR spectroscopy of cytokine–receptor interactions
    - complex formation effect on protein stability, 179–181
    - experimental procedure, 168
    - granulocyte colony stimulating factor–receptor interaction analysis, 175–179
    - previous studies, 167–168
    - protein secondary structure analysis, 168–172
    - thermal stabilities of granulocyte colony stimulating factor and receptor in isolated forms, 172–175
  - Isotope-edited IR spectroscopy, analysis of peptides and proteins, 32,34
  - Isotopic substitution, role in interpretation of IR spectra, 48–50
- L**
- Ligand binding, tertiary structural changes, 104–108
  - Lipid(s), orientation determination, 97–104
  - Lipid bilayer, orientation and tertiary structural changes of proteins and peptides, 104–115
  - Low-density lipoprotein receptor active conformation, apoprotein E, 103–104
- M**
- Macromolecular attributes, influencing factors, 145
  - Measured order parameter, definition, 98–99,101*f*
  - Measurement, Fourier-transform IR spectra, 56–59
  - Membrane proteins, problems with structure determination, 96
  - Molecular reaction mechanisms of proteins, time-resolved Fourier-transform IR spectroscopy, 159–165
- N**
- Nuclear magnetic resonance spectroscopy
    - analysis of protein structure, 117–118
    - limitations, 96
- O**
- Orientation, determination in proteins and lipids, 97–104

Orientational order parameter,  
calculation, 98

## P

Partial least-squares method for  
secondary structure determination of  
proteins from amide I and III IR  
bands

advantages, 118,125

disadvantages, 125,128

experimental materials, 119

experimental procedure, 119

method description, 119–121

secondary structure content  
prediction, 121–124,126–127

studies, 118–119

Partially folded intermediates,  
aggregation precursors, 136

Peptide(s)

interpretation of IR spectra, 38–51

structural analysis using Fourier-  
transform IR spectroscopy, 54–92

technique and applications of IR  
spectroscopy, 3–34,36

Peptide group modes, characteristics,  
39–46

P-glycoprotein

description, 105

function, 105

tertiary structure changes during  
ligand binding, 105–108

Polarized IR spectroscopy  
analysis of peptides and proteins,  
32,34

structural analysis of peptides, 91

Principles, studies, 2–128

Protein(s)

conformations, 146

interpretation of IR spectra, 38–51

IR spectrum, 159,160*f*

orientation determination, 97–104

secondary structure determination  
from amide I and III IR bands using  
partial least-squares method, 117–  
128

technique and applications of IR  
spectroscopy, 3–34

time-resolved Fourier-transform IR  
spectroscopy of molecular reaction  
mechanisms, 159–165

vibrational spectra, 146

Protein aggregates

attenuated total reflectance Fourier-  
transform IR spectroscopy for  
secondary structure determination,  
132–143

problem, 132

Protein aggregation

mechanism model, 141–142

ordered and disordered deposits, 133–  
134

problems, 133–134

Protein–protein interactions, isotope-  
edited Fourier-transform IR  
spectroscopy, 167–181

Protein secondary structure analysis,  
use of amide I vs. amide III bands,  
16–32

Protein stability, role of complex  
formation, 179–181

Protein structure analysis, methods,  
3,117–118,132,145–146

Protons, exchanging, assignment to  
secondary structure, 107–115

## R

Raman spectroscopy, limitations, 97

Receptor–cytokine interactions,  
isotope-edited Fourier-transform IR  
spectroscopy, 167–181

## S

Second-derivative IR spectroscopy  
identification of protein bands, 8–9  
thermal denaturation of elastase with  
guanidinium chloride, 145–156

Secondary structure analysis  
Fourier-transform IR spectroscopy,  
54–92



- protein aggregates, 132–143
  - use of amide I vs. amide III bands, 16–32, 117–128
  - Self-deconvolution, identification of protein bands, 8–9
  - Side chains, Fourier-transform IR spectroscopy, 79, 81–82
  - Solid-state NMR spectroscopy, limitations, 96–97
  - Solution studies, role in interpretation of IR spectra, 50–51
  - Spectral processing
    - curve analysis methods
      - second derivative, 8–9
      - self-deconvolution, 8–9
      - statistical analysis of individual bands, 9
    - fitting process parameters, 8
  - Statistical analysis of individual bands, identification, 9
  - Structural analysis of peptides, Fourier-transform IR spectroscopy, 54–92
  - Surface protein adsorption analysis, IR spectroscopy, 15–16, 17*f*
- T**
- Techniques, protein structure analysis, 3
  - Tertiary structural changes, proteins and peptides, 104–115
  - Theoretical directions, interpretation of IR spectra, 47–48, 49*f*
  - Theory, studies, 2–128
  - Thermal denaturation of elastase with guanidinium chloride
    - differential scanning calorimetric measurements, 149–151
    - experimental materials, 147
    - experimental procedure
      - data collection parameters, 148–149
      - differential scanning microcalorimetry, 147
  - IR spectroscopy, 147–148
  - IR spectroscopy
    - amide I region changes, 151, 152*f*
    - band parameters vs. temperature, 151–152
    - irreversibility, 153–154
    - mechanism, 154–156
    - normalization of IR intensities, 149
    - previous studies, 146–147
    - Tilt of helix axis, definition, 99
  - Time-resolved Fourier-transform IR
    - difference spectroscopy applications, 160–161
    - molecular reaction mechanisms, 159
  - Time-resolved Fourier-transform IR spectroscopy of molecular reaction mechanisms of proteins
    - analysis of peptides and proteins, 32, 34
    - bacteriorhodopsin
      - light-induced bacteriorhodopsin–intermediate transition, 162–163
      - proton pump mechanism model, 164–165
      - thermal reactions, 163–164
    - experimental setup, 161
    - previous studies, 160
  - Two-dimensional IR spectroscopy, analysis of peptides and proteins, 32, 34
- U**
- Urea, use for unfolding and denaturing of proteins, 147
- V**
- Viral coat proteins, aggregation problem, 71, 73, 74–75*f*
- X**
- X-ray crystallography, analysis of protein structure, 117–118