BIOTECHNOLOGY: PHARMACEUTICAL ASPECTS

Methods for Structural Analysis of Protein Pharmaceuticals

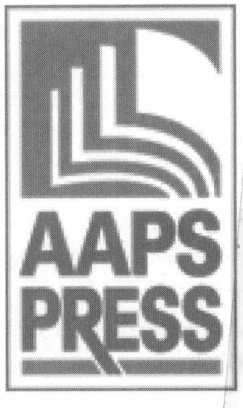
Edited by
Wim Jiskoot
Daan Crommelin

AAPS PRESS

Methods for Structural Analysis of Protein Pharmaceuticals

Wim Jiskoot
Utrecht University, The Netherlands

Daan J.A. Crommelin
Utrecht University, The Netherlands



American Association of Pharmaceutical Scientists
Arlington, VA

Library of Congress Cataloging-in-Publication Data

Jiskoot, Wim.
Methods for structural analysis of protein pharmaceuticals / Wim Jiskoot, Daan J.A. Crommelin.
p. cm. -- (Biotechnology: pharmaceutical aspects ; 3)
Includes index.
ISBN 0-9711767-2-8
1. Protein drugs--Analysis. I. Crommelin, D.J.A. (Daan J. A.) II. Title. III. Biotechnology (Arlington, Va.) ; 3.

RS431.P75J55 2005
615'.1901--dc22

2005041029

Wim Jiskoot
Department of Pharmaceutics
Utrecht Institute for
Pharmaceutical Sciences (UIPS)
Utrecht University
P.O. Box 80082
3508 TB Utrecht
The Netherlands

Daan J.A. Crommelin
Department of Pharmaceutics
Utrecht Institute for
Pharmaceutical Sciences (UIPS)
Utrecht University
P.O. Box 80082
3508 TB Utrecht
The Netherlands

ISBN: 0-9711767-2-8

Executive Director:	John Lisack, Jr.
Project Coordination:	Ericka McIntyre (Custom Editorial Productions Inc.)
Design/Production:	Custom Editorial Productions Inc.
Cover design:	Sitki Kazanci (AAPS)

PREFACE

Protein pharmaceuticals form a fast-growing category in our arsenal of drugs. New findings in molecular biology and in biotechnology provide more and better opportunities to design new therapeutic agents that are based on endogenous compounds such as cytokines and protein hormones or that are based on the concept of the structure and action of antibodies. At this moment, the FDA has approved one hundred eight biotechnology-based medicines, and the majority of these are pharmaceutical proteins (see http://www.phrma.org/newmedicines/biotech/). Moreover, the pipeline for new drugs shows that the number of pharmaceutical proteins will continue to grow: More than 320 biotechnology-based medicines are currently under development in the United States.

In the coming years, the patents on a number of the first-generation pharmaceutical proteins, such as erythropoietin, will expire. This means that so-called biogenerics may be introduced into the market, just as generic, low-molecular-weight drugs are currently available alongside the products of innovators. However, the discussion on how to deal with biogenerics is still ongoing, as there are a number of distinct differences between pharmaceutical proteins and low-molecular-weight drugs. Low-molecular-weight drug products, such as aspirin, can be characterized in minute detail in terms of chemistry and physics. Biopharmaceutical profiling of such drug products has reached a high level of sophistication that ensures bioequivalence between the innovator's product and the generic product. With pharmaceutical protein products, full characterization of the active compound is much less straightforward, as the high-molecular-weight compounds have, besides their chemical structure (primary structure and post-translational modifications), complex secondary, tertiary, and sometimes even quaternary structures that have to be described and defined in detail and, in the case of biogenerics, have to be mimicked.

The leading issue in this book is, how can protein pharmaceuticals be characterized? What are the techniques currently at our disposal? And, what does each of these techniques from our "toolbox" tell us? This question was brought up in an earlier book edited by us together with Dr. James N. Herron in 1995: *Physical Methods to Characterize Pharmaceutical Proteins* (Plenum, New York, 1995). Since then, ten years have passed and a lot of progress, both in terms of basic insights and in terms of technology, has been made. *Methods for Structural Analysis of Protein Pharmaceuticals* can be considered a follow-up of the 1995

publication. The number of chapters has significantly been expanded and a number of additional techniques have been included.

In this book, the nature of the different analytical techniques and their contribution to define pharmaceutical proteins is described by a group of leading experts. In our 1995 publication, the last paragraph of the preface made the important point that, in general, even in a concerted approach, all the sophisticated techniques described in the different chapters could not fully define the protein structures in their pharmaceutical formulations as can be done with formulations of low-molecular-weight drugs. Ten years later, in spite of all the progress we made, this statement still holds true. The quality of pharmaceutical proteins is based on the combination of a strictly controlled production procedure and specifications based on a set of relevant analytical techniques.

This book deals with analytical techniques that are relevant for characterizing pharmaceutical proteins. Therefore, it should serve the analytical chemist who is responsible for therapeutic protein analysis in the industry. However, the impact of the information goes beyond the analytical chemist. The formulation scientist should also be aware of the principles, the potential, and the limitations of the contents of our analytical toolbox, because the formulation scientist has the responsibility to design a robust formulation of a pharmaceutical protein that can be monitored during production and over (storage) time. Furthermore, our target audience is represented by Ph.D. students, postdoctoral fellows, and research scientists who are working in the academic world and in the pharmaceutical/biotech industry on basic aspects of proteins.

The editors hope that this book will contribute to the rational design of high-quality pharmaceutical protein formulations and to the selection of relevant quality control tests to guarantee product quality when future protein pharmaceuticals are produced and marketed.

Wim Jiskoot and Daan J.A. Crommelin

CONTENTS

1

Ultraviolet Absorption Spectroscopy

LISA A. KUELTZO

Department of Pharmaceutical Sciences, University of Colorado Health Sciences Center, Denver, CO

C. RUSSELL MIDDAUGH

Department of Pharmaceutical Chemistry, University of Kansas, Lawrence, KS

History

More than fifty years ago, it was proposed that changes in the ultraviolet absorbance spectrum of a protein could be correlated with its structural perturbations (Beaven and Holiday, 1952). A variety of studies were conducted at that time to explore this potential application, as described in comprehensive reviews by Beaven and Holiday (1952) and Wetlaufer (1962). In the following decades, however, other spectroscopic techniques, including circular dichroism, intrinsic fluorescence, nuclear magnetic resonance (NMR), and Fourier transform infrared (FTIR) spectroscopies, as well as x-ray crystallographic methods, became the methods of choice for characterization of protein structure. Absorbance spectroscopy, although still commonly used in everyday laboratory practices, was primarily relegated to concentration measurements, turbidity assays, and enzymatic activity studies.

At periodic intervals, however, interest in the structural implications of the protein absorption spectrum has been renewed, especially regarding the information contained within the derivative spectrum of the region from 250 to 300 nm—the absorbance region of the three aromatic amino acid residues. It was observed that not only could these signals be employed to quantify the aromatic content of a protein (Balestrieri et al., 1978), but that the difference spectrum (Donovan, 1973), as well as a ratio of the magnitudes of the various derivative peaks

(Ragone et al., 1984), could be used to monitor subtle structural changes, providing information regarding the exposure of the aromatic residues to solvent. Subsequent work has established that more dramatic structural changes (e.g., unfolding) can also easily be detected by these methods. More recently, advances in instrumentation technology, specifically the development of photodiode array detectors, have permitted highly accurate and reproducible measurement of minute shifts in derivative peak positions, which can be directly correlated to small changes in the environment of the aromatic chromophores (Mach et al., 1991b).

Today, interest in ultraviolet absorption spectroscopy as a probe of protein structure is experiencing a minor renaissance. The cost of high-resolution instrumentation has decreased, the availability of photodiode array detectors has expanded the application of absorbance analysis to on-line evaluation of chromatographic separations, and advances in software and algorithm design have increased our ability to obtain the maximum amount of data from spectra. Currently, absorbance spectroscopy offers a fast, nondestructive, high-resolution, and inexpensive alternative to other commonly used spectroscopic techniques, as described in more detail below.

Theoretical Background

Light is traditionally represented as a wave composed of both an electric and a magnetic field component. From a quantum mechanics perspective, light can be described as particles (photons) of energy (E), as defined by Planck's law

$$E = h\nu = hc/\lambda \tag{1}$$

where h is Planck's constant, ν and λ are the frequency and wavelength of the light, respectively, and c is the speed of light. Molecules residing in a ground energy state under normal conditions can be raised to a higher, excited energy state upon interaction with a photon that has an energy equaling the difference between the ground and excited energy states. The molecule can then return to the ground state by emission of the energy by either radiative (as observed in fluorescence and phosphorescence) or nonradiative processes, as seen in electronic absorption. The transition from a ground state to an excited state is customarily defined in terms of the configurations of the molecular orbitals in the ground and excited states; the most commonly observed transitions in protein absorption spectra are the $n \leftrightarrow \pi^*$ and $\pi \leftrightarrow \pi^*$ transitions, where * denotes the excited state.

Each electronic state contains a number of underlying vibrational and rotational energy states. Transition to an excited state can therefore occur at any of a number of potential vibrational energy levels (Figure 1). Collisions between molecules, with solvent molecules under solution conditions, as well as small differences in the absorbance of identical

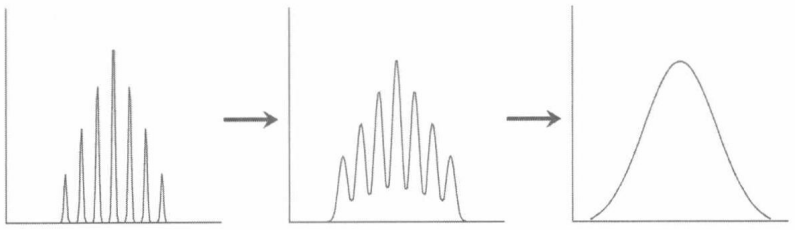


Figure 1 Representation of a vibronic absorption spectrum. The energy of the fine bands is smeared by collisions with other molecules (e.g., solvent molecules), resulting in the traditionally observed broad electronic absorption spectrum.

chromophores under different conditions within the system (e.g., multiple tyrosine residues within a protein), produce a smearing of the energy of the vibronic bands into the broad peak that is normally observed in solution electronic absorption spectra.

Chromophores

The Peptide Group

The most prevalent chromophore in a protein molecule is the amide group of the protein backbone. Its absorbance spectrum contains two main transitions, the first absorbing strongly at 195 nm ($\pi \leftrightarrow \pi^*$) with a second, weaker transition occurring at ~220 nm ($n \leftrightarrow \pi^*$). The peptide group has traditionally been of interest spectroscopically due to the information it can provide regarding the secondary structure content of a protein. Although distinctly different characteristics of the far-UV transition have been observed for model helical, β-sheet, and random coil peptides (Wetlaufer, 1962), interpretation of signals in this region is complicated by the absorbance of common solution components (e.g., inorganic ions and dissolved oxygen) below 200 nm (Wetlaufer, 1962). Additionally, because techniques such as circular dichroism and FTIR spectroscopy possess greater sensitivity to and resolution of secondary structure composition and changes, the application of simple absorption spectroscopic methods to this region has been relatively uninvestigated. Instead, interest has focused on the near-UV region between 250 and 320 nm, which contains the primary absorption from the aromatic amino acid side-chains.

The Aromatic Side-Chains

Tryptophan

The strongest absorbing of the three aromatic amino acids is tryptophan ($\varepsilon_{280nm} = 5540$). As shown in Figure 2, the maximum absorbance of tryptophan (Trp) occurs at ~280 nm, with a less-intense transition observed as a shoulder near 292 nm (both corresponding to $\pi \leftrightarrow \pi^*$

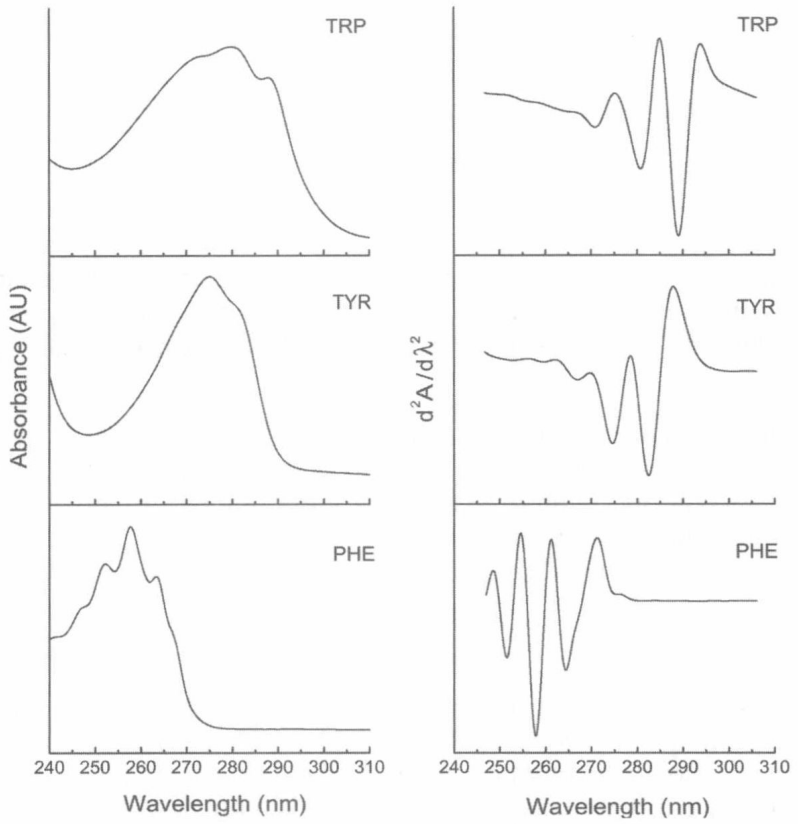


Figure 2 Zero-order and second-derivative absorption spectra of *N*-acetyl-X-ethyl ester derivatives of tryptophan (TRP), tyrosine (TYR), and phenylalanine (PHE). The ordinate axis is presented in arbitrary units.

transitions). The latter is responsible for the shoulder frequently seen in the protein spectrum. Although the fine structure is not well defined in the zero-order spectrum, the derivative shows three distinct minima. Covalent modification of tryptophan residues is occasionally observed, primarily reflecting oxidation of the indole ring (Zhang et al., 2003). Additionally, tryptophan can be modified for investigational purposes (e.g., *N*-bromosuccinimide–induced oxidation) (Wetlaufer, 1962); such modifications change the absorbance properties of the chromophore, and their potential presence always needs to be considered.

Tyrosine

The absorption of tyrosine (Tyr), although less intense (ε_{280} = 1480), overlaps significantly with that of tryptophan, with a maximum absorbance occurring at ~276 nm ($\pi \leftrightarrow \pi^*$). The zero-order spectrum consists of a broad absorbance band with two small shoulders, reflecting the fine structure that is observed once again in the multiple

peaks of the derivative spectrum (Figure 2). Of the aromatic amino acids, tyrosine is the most susceptible to structural modifications including nitration, iodination, phosphorylation, and acetylation, as well as ionization of the tyrosine hydroxyl group (see "Effect of pH," below; Creighton, 1993). Such modifications can significantly alter both the intensity and the position of the absorption transitions (Table 1) and complicate spectral interpretation.

Phenylalanine

In contrast to the previous two aromatic side-chains, phenylalanine (Phe) absorbs only weakly (ε_{258nm} = 197) in the 240- to 270-nm region ($\pi \leftrightarrow \pi^*$). Significant fine structure is observed in both the zero order and derivative spectra (Figure 2), with an average of five peaks observable in various environments. In zero-order protein absorption spectra, this fine structure is usually visible only as subtle inflection points (bumps), due to the dominance of the spectrum by the more intense signals of Tyr and Trp.

The Disulfide and Others

Although cysteine does not absorb significantly in the 250- to 300-nm wavelength region, the disulfide-bridged form of the residue, cystine, does show some weak absorbance. The preponderance of disulfide bridges in some proteins can result in a significant contribution to the overall extinction coefficient (see "Protein Concentration Determination and Extinction Coefficients," below). Although methionine and histidine also show some weak absorbance, these residues absorb in the same region as the peptide bond (185 to 220 nm) (Wetlaufer, 1962) and therefore are not usually resolvable in the absorption spectrum of most proteins.

Extrinsic Chromophores

In addition to the intrinsic protein chromophores described above, the absorption spectrum of a protein may include absorbance peaks from a wide variety of other species. Although the absorbance of these extrinsic chromophores is most commonly monitored in the visible wavelength region, in some cases strong and potentially useful signals in the UV or near-UV region are also present. These additional spectral components are sensitive to both the environment as well as the oxidation state of the cofactor and thus can often provide useful information about overall protein conformation, characteristics of the ligand binding site, complex stoichiometry, and reaction kinetics (e.g., cofactor conversion).

The most common extrinsic chromophores are divalent metal cations. Although some cations, such as zinc and magnesium, are optically transparent, others, such as iron, copper, or non-native metals such as cobalt, cadmium, and terbium, possess intense absorbance

Chromophore	λ_{max} 1	λ_{max} 2	λ_{max} 3	λ_{max} 4	ε (λ_{max} 2)
Amide bond	195[a]	220[a]	—	—	—
Tryptophan (in proteins)	—	—	—	—	5540[e]
N-Ac-L-Tryptophan-NH_2	219[b]	280[c]	288[d]	—	5390[f]
Tyrosine (in proteins)	—	—	—	—	1480 (at 280 nm)[e]
N-Ac-L-Tyrosine-NH_2	224[b]	275[d]	[280][f]	—	1390[f]
∈-Iodotyrosine	—	283[c]	—	—	2750[c]
∈-Nitrotyrosine	—	360[c]	—	—	2790[c]
∈-Aminotyrosine	—	275[c]	—	—	1600[c]
O-Acetyltyrosine	—	363[c]	—	—	262[c]
N-Ac-L-Phenylalanine-O-Me	208[b]	252[b]	258[b]	264[b]	197 (at 258 nm)[b]
Cystine	—	[250][b]	[260][b]	[280][b]	134 (at 280 nm)[e]

Brackets denote points of inflection.

[a]van Holde et al. (1998); [b]Fasman (1992); [c]Creighton (1993); [d]Beaven and Holiday (1952); [e]Mach et al. (1992); [f]Mach et al. (1995).

Table 1 Absorption Characteristics of Selected Intrinsic Protein Chromophores

bands (Vergani et al., 2003). These signals can often provide information about the metal binding site, such as binding capacity (Vergani et al., 2003), oxidation state (Favilla et al., 2002), identity of binding site residues, and coordination geometry (Burke et al., 1992). In addition to divalent cations, a variety of enzyme cofactors, such as NADH (Luong and Kirsch, 1997; Piersma et al., 1998) and flavin adenine dinucleotide (FAD; Gadda and Fitzpatrick, 1998), exhibit intense absorbance signals in the near-UV and visible ranges; changes in these signals with oxidation state are often used in the analysis of enzyme kinetics. In a very familiar example, the ferroprotoporphyrin, or heme, group possesses a strong absorbance signal that is frequently monitored in structural studies of heme-containing proteins (e.g., myoglobins, hemoglobins, and various cytochromes), as its absorbance characteristics are often perturbed during functional structural alterations of these proteins.

Instrumentation

A variety of UV-visible spectrophotometers are commercially available, ranging widely in both cost and capability. Regardless of design, however, each spectrophotometer can be broken down into four basic components: the light source, sample compartment, light dispersion device, and the detector. An ideal light source should possess low noise, good intensity over the desired wavelength range, and stability over an extended period of time (Owen, 1996). Most current UV-visible spectrophotometers employ a combination of two sources: a deuterium lamp, possessing good intensity over UV wavelengths, and a tungsten-halogen lamp for the visible region. The instrument then either switches between the lamps during spectral collection, or the two signals are combined into one broad-range light source, depending on the design. Broad-range xenon lamps are also available but are in general a less desirable choice due to their significantly higher noise levels compared to the deuterium-tungsten-halogen combination.

The spectrophotometer sample compartment typically contains one or more aligned cuvette holders. In choosing an instrument, the availability of sample accessories, which expand the experimental capabilities of the instrument, should also be considered. Many systems today offer optional temperature-controlled sample holders (either Peltier-based or water jacketed), multi-cell devices, and stopped-flow and titration accessories. Additionally, some systems are based on a double-beam design, in which the sample and reference scans are collected concurrently from two independent, matched sample cells. Although this design was considered optimal in the past, as it minimizes errors produced by fluctuations in lamp intensity from the time difference between reference and sample spectra collection, the stability of current light sources has made single-beam instruments a viable option.

Scanning Monochromator Based

One of the most significant differences among current spectrophotometer designs is the nature of the light dispersion/detector system. This can be especially crucial in high-resolution derivative absorbance studies (discussed in "High-Resolution Derivative Analysis," below). In a monochromator-based system, the light from the source enters the monochromator through an entrance slit that focuses it on a dispersion device, most commonly a holographic grating, although prisms are still occasionally employed. A narrow band of the dispersed light is then selected by the exit slit of the monochromator and passed through the sample to the detector, most commonly a photomultiplier tube (PMT). In scanning instruments, the grating (or prism) physically moves to produce the desired wavelength range; although these instruments do produce high-quality spectra, the movement of the monochromator can introduce some error into the wavelength reproducibility of repeated measurements.

Diode Array Based

In a diode array instrument, the entire spectrum of the light source is focused on the sample. The light transmitted through the sample passes into the entrance slit of the polychromator and through the dispersion element to the detector. The detector consists of an array of photodiodes set at regular intervals (typically 0.5- to 8-nm apart) that can simultaneously detect the full spectrum of light transmitted (Owen, 1996). Although the resolution of the collected spectrum is limited by the spacing of the photodiodes, simple interpolation algorithms, often included in the spectrophotometer operating software, can mathematically increase the effective resolution of spectra down to 0.01 nm. Offsetting the requirement for interpolation is the very high wavelength reproducibility of the collected data, as the detectors in these systems contain no moving parts.

Data Analysis

Zero-Order Spectra

The raw data collected with a spectrophotometer is referred to as the "zero-order" spectrum. In the case of a protein with intrinsic chromophores, this typically consists of a strong absorbance from the peptide bond at ~200 nm and a weaker, rather broad absorbance between 250 and 300 nm. The weaker absorbance is the additive signal of the overlapping phenylalanine, tyrosine, and tryptophan residues, as described above. The zero-order spectrum has commonly been used for concentration determination, examination of gross spectral characteristics, and detection of major structural changes such as extensive unfolding. Although many of the applications described in this chapter

focus on the derivative form of the spectrum, the zero-order spectrum should always be examined before performing other forms of data analysis. This examination provides an estimate of the overall quality of the sample with regard to uniformity compared to previous samples, impurities, as well as the presence of particulates that can result in light scattering, which can distort the pure absorptive spectrum.

Light-Scattering Correction

Interpretation of zero-order absorbance spectra can be significantly complicated by the presence of light scattering. Solute molecules whose size begins to approach 1/20 to 1/50 of the wavelength of the incident light (e.g., protein multimers, aggregates) will elastically scatter the light, preventing it from reaching the detector and artificially inflating absorbance values. The scattering effect can be defined (Timasheff, 1966) as

$$\ln(\text{OD}) = a \ln \lambda + b \tag{2}$$

When scattering is present, the raw spectrum should be represented as "optical density" (OD) on the ordinate rather than absorbance, as both absorbance and scattering (turbidity) components are included. To obtain an accurate measurement of sample absorbance, the light-scattering contribution, if any, must be eliminated.

The intensity of scattered light is approximately proportional to the inverse of the fourth power of the wavelength of the light in the Rayleigh regime (λ less than 1/20 incident light). Scattering is usually observed in absorption spectra as a negative slope in a region of the spectrum where the sample does not absorb light (greater than 320 nm in most protein samples), as shown in Figure 3. Once detected, scattering can be eliminated by either physical (e.g., removing the larger material by centrifugation or filtration) or mathematical methods. The scattering component within the absorbing region of the spectrum can be extrapolated mathematically through the absorbing region by fitting the nonabsorbing region of the spectrum (e.g., 320 to 500 nm) to Equation 2 and then subtracting the scattering signal to isolate the pure absorbance signal (Figure 3). Most commercial spectrophotometer software packages now include this type of correction in their basic analysis package. Alternatively, the OD contribution at any wavelength can be calculated from the OD values at 320 and 350 nm as follows (Mach and Middaugh, 1993):

$$\text{OD}_{(\lambda)} = 10^{(m+1) * \log(\text{OD } 320) - m \log(\text{OD } 350)} \tag{3}$$

where $m = 64.32 - 25.67 * \log \lambda$ ($m = 1.5$ at 280 nm). The absorbance component at the desired wavelength can then be calculated. It is important to remember that these corrections take into account only

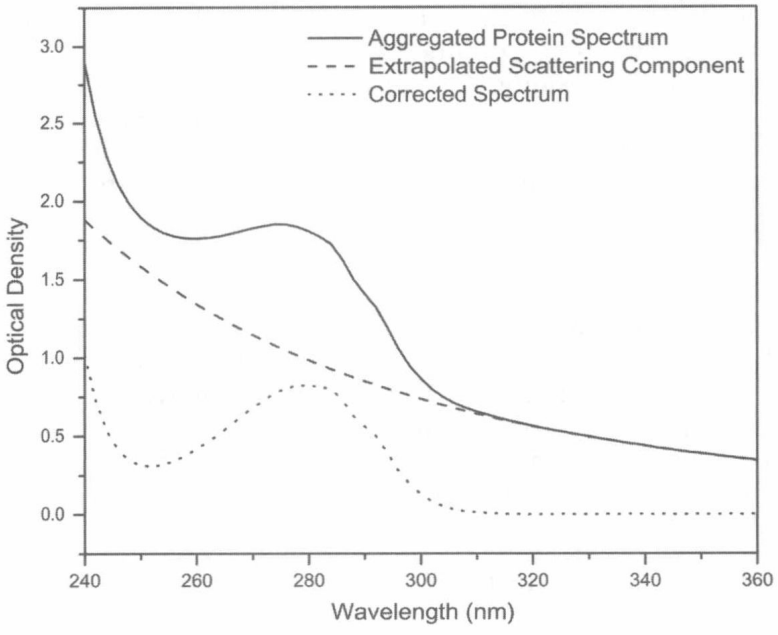


Figure 3 Light-scattering correction of an aggregated protein sample. The light-scattering component is extrapolated through the protein absorbing region (less than 300 nm) and subtracted to obtain the corrected zero-order spectrum.

absorbance and Rayleigh light-scattering components. As particles become larger, absorption flattening, a phenomenon where the chromophores within a particle obscure one another, may also occur (see "Spectral Artifacts and Complications," below) and is not correctable by these methods. Furthermore, correction of scattering by larger particles (e.g., Mie and multiple scattering) requires much more complex methods and is not generally feasible.

Derivative Analysis

Aromatic Amino Acid Quantitation

One of the first applications of derivative absorbance analysis to proteins investigated the quantitation of the aromatic amino acid content, although advances in protein and gene sequencing techniques have made this approach less common in current practices. Calculation of the first through fourth derivatives of a protein's absorption spectrum resolves the overlapping absorbance contributions of Phe, Tyr, and Trp, permitting selective analysis of the individual residues. Balestrieri and colleagues (1978) demonstrated that, once resolved, the intensities of the derivative peaks can be directly correlated to the number of Phe and Trp residues present in a protein. This approach was expanded by Servillo and colleagues (1982), who determined the relationship