Determining the Identity and Purity of Recombinant Proteins by UV Absorption Spectroscopy

Ultraviolet absorption spectrophotometry can be used to confirm the identity and, to a lesser extent, assess the purity of recombinant proteins and their peptide fragments. The near-UV (250 to 350 nm) absorbance spectrum of a protein is almost entirely a function of its aromatic amino acids—tryptophan, tyrosine, and phenylalanine (Fig. 7.2.1). Because each protein (gene product) has a unique amino acid sequence, the particular aromatic amino acid content of each protein results in a unique spectrum in the near-UV region. The highly specific microenvironment experienced by each aromatic residue in the three-dimensional protein matrix results in fine shifts in a protein’s spectrum, which can be detected (Basic Protocol 1) and analyzed (Support Protocol). Thus, quality UV spectra serve as indicators of the structural integrity of proteins. In contrast to electrophoretic analyses (Chapter 10), UV spectroscopy allows the detection of nucleic acids, certain prosthetic groups, and metal ions which are frequently associated with proteins.

The unique UV spectral properties of proteins can in turn be used to assess their purity. This application is inherent in the use of a diode array detector to monitor the effluent.

**Figure 7.2.1** Normal (zero-order) and second-derivative spectra of aromatic model compounds: N-acetyl-L-tryptophanamide, N-acetyl-L-tyrosineamide, and N-acetyl-phenyl ethyl ester (solid lines) and “average” aromatic amino acid residues as determined using a set of globular, water-soluble proteins (dotted lines). Absorbance and second-derivative values correspond to 1 M solutions, 1-cm path length. The band positions in denatured proteins usually have intermediate positions.
from a high-performance liquid chromatography (HPLC) column. A protocol using this technique to assess the purity of recombinant proteins is presented in Basic Protocol 2.

**ANALYSIS OF PROTEINS USING NEAR-UV SPECTROPHOTOMETRY**

This procedure identifies proteins using near-UV spectrophotometry. Protein solution is placed in a UV-transparent quartz cuvette, absorption of light is measured as a function of wavelength, and the resulting spectrum is transferred either to a displaying device (e.g., CRT or printer) for inspection or to a computer that can employ a spreadsheet or other program for derivative calculations and/or quantitative analyses (see Support Protocol).

**Materials**

- Detergent solution
- Methanol
- 0.1 N hydrochloric acid
- Acid/ethanol cleaning solution (see recipe)
- 10 N NaOH
- 4% holmium oxide (aqueous solution in 10% perchloric acid; see recipe)
- Buffer identical to that containing the protein (reference standard)
- Sample for analysis in solution at an appropriate concentration (e.g., 0.05 to 1 mg/ml; see below)
- Vacuum-powered cuvette washer (NSG Precision Cells, Kontes Glass, or equivalent)
- Two factory-matched synthetic quartz (Suprasil or equivalent) cuvettes (Hellma, NSG Precision Cells, or equivalent; for double-beam spectrophotometers) or a single cuvette (for single-beam use)
- Gel-loading plastic pipet tips (Marsh Biomedical Products; Rainin Instruments) for a pipettor (Pipetman, Eppendorf, or equivalent)

**CAUTION:** Employ appropriate safety precautions when using concentrated acids and bases.

**Ready the spectrophotometer**

1. Turn on the spectrophotometer and the UV lamp; allow 20 min for warm-up.

   *Lamp stability can be monitored by recording a baseline as a function of time. Only the UV (deuterium) lamp need be turned on; it should provide sufficient light intensity up to ~400 nm. Higher wavelengths are needed only for proteins with visible-light-absorbing chromophores, such as heme proteins and metalloproteins. The instrumental wavelength cross-over point (the wavelength at which the light source is switched from the deuterium to the tungsten lamp) must be set ≥ to 360 nm.*

**Prepare the cuvette**

2. Use the vacuum-powered cuvette washer to wash the cuvettes. Apply a detergent solution briefly and wash three or four times with water, waiting each time until all the water is exhausted from the cuvette. Finally wash with methanol to remove water-insoluble impurities and speed up drying. Wipe the outside of the cuvette with a tissue (Kimwipe) or, better, with lens cleaning paper and inspect visually. If not clean, repeat the cleaning procedure starting with 0.1 N HCl instead of detergent. If the cuvette is still not clean, soak it in acid/ethanol solution for several hours. As a last resort, soak briefly (20 min) in 10 N NaOH.

   *Cuvettes should always be cleaned immediately after use. Denatured proteins may often be removed by soaking overnight in a mild pepsin solution. Prolonged contact with the NaOH may partially dissolve quartz faces of the cuvette.*
**Set up appropriate instrument parameters**

3. Identify the cross-sectional area of the light beam (from the instrument’s instruction manual) and determine the correct height to which the cuvettes should be fitted.

   The light beam may be as large as $1 \times 2$ cm. The sample volume must be large enough so that the meniscus is completely above the beam. The light must not impinge on the air-solution interface. If the beam is wider than the portion of the cuvette containing the solution, the remaining portions of the cuvette walls should be masked by a light-absorbing coat; special blackened cuvettes are available for this purpose. In some instruments the height of the beam can also be reduced; do so if sample availability is a problem. The cuvettes must be positioned vertically; any deviations may alter the effective path length and reduce the precision of measurements.

4. Adjust the wavelength (scanning) range to 240- to 360-nm. For diode-array spectrophotometers, use 10-sec measurement times. For conventional scanning spectrophotometers use bandwidths of 1 or 2 nm, time constants (response times) of 2 sec (medium) or less, and scanning speeds of 60 nm/min or less. Set the data point interval to 1 nm (scanning instruments only).

   Use the same settings if spectra are to be compared to previously collected ones.

**Measure the spectrum of a reference standard**

5. Measure the baseline (background) spectrum with no cuvettes in the sample holders. Insert a sealed cuvette containing 4% holmium oxide (aqueous solution in perchloric acid; Weidner et al., 1985) into the sample holder. Measure the spectrum and save it on a disc or other permanent form of storage. Collect a second spectrum using a scanning speed approximately one-half of that initially used.

6. Determine the exact peak positions by calculating the first-derivative spectrum in the 280- to 290-nm range. If the spectrophotometer has a built-in derivative calculation ability, choose derivative order 1, polynomial degree 2, and a window of five data points. If not, transfer the spectra in ASCII (text) format (or type it) into a spreadsheet (Excel, LOTUS-123, or equivalent) and perform the calculation using Equation 7.2.1 (Savitzky and Golay, 1964; Steiner et al., 1972), where \( FD(\lambda) \) is the first-derivative value at the integral wavelength \( \lambda \), \( A(\lambda \pm n) \) is the absorbance values at wavelength \( \lambda + n \) (where \( n = -2, -1, 1, \) and 2), and \( k \) is the data point interval (in nanometers).

   \[
   FD(\lambda) = \frac{-2[A(\lambda - 2)] - A(\lambda - 1) + A(\lambda + 1) + 2[A(\lambda + 2)]}{10 \times k}
   \]

   **Equation 7.2.1**

   The values of \( n \) double if a 2-nm data point spacing is employed.

   The derivative calculation is performed as follows: (1) copy the absorbance values of the spectrum into the first column (typically A1:A100 range); (2) fill second column (B1:B100) with the corresponding wavelength values, entering the first wavelength and using Edit/Fill/Series commands; (3) enter the formula in the cell of the third column, which corresponds to \( A(\lambda) \) absorbance value (B3)—e.g., for Equation 7.2.1 and \( k=1 \), cell C3 contains \( =(-2*\text{A1}-\text{A2}+\text{A4}+2*\text{A5})/10 \) (derivative values cannot be calculated for first two and last two data points because a five-data point window is used); (4) copy the content of that cell (the formula) down onto the entire column (B3:B98) using Edit/Copy and Edit/Paste commands; and (5) use the Chart Wizard option to plot the resulting spectrum.

7. Note where the spectra cross the zero line (change sign from positive to negative) and calculate the exact positions of these intercepts using Equation 7.2.2, where \( \lambda \) is the
wavelength prior to the intersection point, \( k \) is the data point interval (in nanometers), and \(|\text{FD}(\lambda)|\) and \(|\text{FD}(\lambda + 1)|\) are the absolute values of the first derivative at the integral wavelengths \( \lambda \) (before the intersection point) and \( \lambda + 1 \) (after the intersection point).

This equation arises from simple geometric interpolation, illustrated for the second-derivative spectra in the inset of Figure 7.2.2B.

| Equation 7.2.2 |

Note that the resulting intersection point is the position of the peak near 287 nm and serves as a reference point for further determination of exact peak positions of the protein spectra.

8. Calculate the intersection point for the second spectrum collected using a slower scanning speed (scanning instruments only); if the position is different, decrease the time constant (response time) or use a slower scanning speed until the positions of the intersections become independent of these parameters.

Test for the proper response time only with the first spectra acquired. Once established, the settings can then be used with confidence. The 287-nm holmium oxide peak is found at 287.18 nm (for 1-nm bandwidths) and at 287.47 nm (for 2- and 3-nm bandwidth) on an instrument extensively verified by measurements of mercury and deuterium emission lines (Weidner et al., 1985). Commercial instruments will deviate somewhat from this value. Correct the wavelength results for proteins for the difference observed. Diode-array instruments (see Basic Protocol 2) will typically require such calibration once a year; scanning instruments should be calibrated frequently depending on use. If a holmium oxide standard is not readily available, Trp (monomeric amino acid) or a related compound can be used for reference purposes.

Measure the protein spectrum

9. Identify the number of light beams by looking at the number of sample holders and their positions in the sample compartment. In a double-beam spectrophotometer, use two matched cuvettes and perform an initial background (baseline) measurement with just buffer (identical to that in which the protein sample is dissolved) in both cuvettes. Then substitute protein solution for the buffer in the sample position cuvette, leaving the reference buffer in the other cuvette. Acquire the sample spectrum and subtract the previously measured background spectrum if necessary. In a single beam spectrophotometer, use a single cuvette. Measure the background with buffer in the cuvette. Then remove the buffer, measure the spectrum of the unfiltered protein solution using the same cuvette in the same orientation, and subtract the background spectrum (baseline).

Sample can be removed from the cuvette using a pipettor with gel-loading plastic pipet tips. Do not scratch the inner optical surface of the cuvette. Many proteins, especially when partially unfolded, strongly adhere to quartz and are not removed by brief detergent or organic solvent washes. If UV absorption peaks at 270- to 280-nm are detected in the baseline, clean the cuvette again (step 2).

10. Note the absorbance at the protein peak centered between 275 and 282 nm. If the absorbance value is >1.0 at the maximum, check the range of linearity of the instrument in use (from the instruction manual). If the absorbance value observed is above this range, dilute the sample or decrease the path length. If a fluorescence-type 2-5 × 10-mm cell is used, turn the cuvette sideways to reduce the path length. A new baseline should be obtained. If the absorbance falls below 0.2, carefully observe the
noise levels and use longer acquisition times (both for the baseline and the actual spectrum).

11. Save the spectrum in a proper format for the particular instrument employed or transfer the spectrum in an ASCII (text) format to a disc for retrieval by a spreadsheet for further analysis. If this is not feasible, print or manually record absorbance values as a function of wavelength.

Most protein solutions containing 1 mg/ml protein have an absorbance near 1.1 ± 0.5 in the 280-nm region. If an estimate of the concentration of the solution is available, the sample can be diluted or concentrated to obtain an absorbance in the 0.2 to 1.0 range. Alternately, make a preliminary measurement at a single wavelength (e.g., 280 nm) to obtain this information. Note that with modern spectrophotometers very low values of absorbance (<0.01) can be accurately measured; however, care must be taken to avoid adsorption artifacts, as a substantial portion of the protein may now reside on the interior surface of the cuvette. This can be checked by obtaining a spectrum of the empty cuvette after carefully removing the protein solution.

INTERPRETING PROTEIN NEAR-UV SPECTRA

Normal (zero-order) spectra are spectra of absorbance versus wavelength as measured by a spectrophotometer without derivative calculation. Different regions of the spectrum contain information about various aspects of the structure of the protein. The 320- to 350-nm region may contain information about aggregation state (due to light scattering) or the presence of UV-absorbing prosthetic groups; the 250- to 300-nm region provides information about the aromatic amino acids and their environment (Fig. 7.2.1 and Table 7.2.1). Numerical calculation of the second derivative of the spectrum minimizes the effect of broad spectral components (such as light scattering or anion absorption), exaggerating fine structure originating from absorption by aromatic rings. This method is particularly useful for precise determination of tertiary structure–dependent band positions as well as for quantitative analysis.

Correct for light scattering

1. Examine the spectral region between 320 and 350 nm (Fig. 7.2.2 contains an example for a typical protein).

Significant “apparent absorbance” of gradually, almost linearly decreasing intensity with increasing wavelength indicates that a portion of the light beam did not reach the detector because it was scattered. Absorbance that falls to zero at 350 nm or does not gradually decline (Fig. 7.2.2) is probably true absorbance arising from prosthetic groups or contaminants.

2. Use Equation 7.2.3 to calculate optical density (OD) due to light scattering, where \( a \) and \( b \) are constants.

\[
\log \text{OD} = a \log \lambda + b
\]

Equation 7.2.3

If the instrument’s software cannot make such a correction, use a spreadsheet (Microsoft Excel, LOTUS-123, or equivalent; see Basic Protocol 1, step 6 annotation, for an example) to generate the light-scattering curve using optical density values at 320 nm and 350 nm. Calculate the light-scattering correction from Equation 7.2.4, where \( m = 64.32 - 25.67 \log \lambda \).

\[
A_c = 10^{[(m + 1) \times \log \text{OD}_{320} - m \times \log \text{OD}_{350}]} - 1
\]

Equation 7.2.4
If this is not feasible, calculate the light-scattering OD at 260 nm (m = 2.31) and 280 nm (m = 1.50) with a calculator. If significant light scattering is detected, remeasure the sample after filtration through a 0.1 to 0.45-micron diameter cut-off filter. The difference of protein concentration before and after filtration will indicate the fraction of aggregated protein. Compare the second-derivative spectra to determine whether the initial sample was homogenous.

Note that optical density (OD) refers to the total extinction of light produced by all processes (including light scattering) whereas absorbance (A) refers only to the actual absorption process. The coefficient a varies from −4 (the ideal case of scattering from a small particle, not larger than one-tenth the wavelength of the light; this situation is known as Rayleigh scattering) to +2.2 for submicron-sized aggregates (Mie theory, in which the scattering intensity becomes a complex function of molecular size; Timasheff, 1966). Because the light-scattering contribution to the optical density is almost linear in the 300- to 350-nm region, a simple linear fit using a spreadsheet or even a crude graphic estimate can serve as a rudimentary measure of the correction necessary to obtain the correct absorbance in the 250- to 300-nm range.

Visually recognize the presence of the individual types of aromatic residues
3. Inspect the spectrum in the 250- to 300-nm range.

A small inflection point on the declining shoulder of the main peak around 290 nm (Fig. 7.2.2) originates from Trp residues; the sharper the inflection point, the more homogeneous and apolar the average microenvironment of these side chains. The distinctive spectral features of Tyr residues are usually masked and less pronounced as a consequence of the less-sharp fine structure of these residues, their broadening by the medium, and their greater heterogeneity from their usually greater frequency in proteins. The presence of the Tyr residues can be deduced, however, from the deviation of the overall spectral shape from that expected for pure Trp. The presence of Phe residues is usually manifested as a series of weak but distinctive bumps superimposed on the rising 250- to 270-nm region, which is dominated by more strongly absorbing Trp and Tyr side chains.

Table 7.2.1  Spectroscopic Properties of UV-Absorbing Amino Acids, Natural Ligands (Prosthetic Groups), and Nucleic Acids

<table>
<thead>
<tr>
<th>Chromophore</th>
<th>λ (nm)</th>
<th>ε (M⁻¹ cm⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tryptophan (in native proteins)</td>
<td>280</td>
<td>5540</td>
</tr>
<tr>
<td>Tyrosine (in native proteins)</td>
<td>280</td>
<td>1480</td>
</tr>
<tr>
<td>Disulfide bond (glutathione)</td>
<td>280</td>
<td>134</td>
</tr>
<tr>
<td>N-Acetyl-L-tryptophanamide</td>
<td>280</td>
<td>5390</td>
</tr>
<tr>
<td>N-Acetyl-L-tyrosinamide</td>
<td>275</td>
<td>1390</td>
</tr>
<tr>
<td>N-Acetyl-L-tyrosinamide</td>
<td>280</td>
<td>1185</td>
</tr>
<tr>
<td>N-Acetyl-L-phenylalanine ethyl ester</td>
<td>257</td>
<td>195</td>
</tr>
<tr>
<td>Cu²⁺ (azurin)</td>
<td>781</td>
<td>320</td>
</tr>
<tr>
<td>FAD (pyruvic dehydrogenase)</td>
<td>460</td>
<td>1270</td>
</tr>
<tr>
<td>Fe³⁺-heme (cytochrome c, reduced)</td>
<td>550</td>
<td>2770</td>
</tr>
<tr>
<td>FMN (amino acid oxidase)</td>
<td>455</td>
<td>1270</td>
</tr>
<tr>
<td>Retinal-Lys (rhodopsin)</td>
<td>498</td>
<td>4200</td>
</tr>
<tr>
<td>DNA (native, per base)</td>
<td>258</td>
<td>6600</td>
</tr>
<tr>
<td>RNA (per base)</td>
<td>258</td>
<td>7400</td>
</tr>
</tbody>
</table>

aWavelengths listed indicate the peak maxima, except for tyrosine, N-Acetyl-L-tyrosinamide, and L-cystine at 280 nm.
bMolar absorptivity values of tryptophan and tyrosine in native proteins as well as oxidized glutathione (disulfide bond) at 280 nm are used for the calculation of molar extinction coefficients of proteins.
**Estimate the nucleic acid content**

4. Calculate the ratio R of absorbances at 260 nm and 280 nm, $A_{260,c}$ to $A_{280,c}$ after correction for light scattering (Equation 7.2.4). Use Table 7.2.2 to estimate % weight of DNA (%N) in the sample (Glasel, 1995).

**Estimate protein concentration in the sample**

5. If a significant amount of DNA was detected (R > 0.70), calculate the protein absorbance $A_{280,c,p}$ using % weight of protein and DNA found in Table 7.2.2, where $E_{0.1\%}$ is the extinction coefficient of the protein (per 1 mg/ml). If the amino acid composition is known, calculate $E_{0.1\%}$ using Equation 7.2.7 and divide the result ($\epsilon$) by the

---

**Figure 7.2.2**  (A) Normal (zero-order) spectrum of a 100 µM solution of a hypothetical protein containing 1 Trp, 2 Tyr, 3 Phe, and 4 Cys (S-S) residues. The constituent spectral components and a typical light scattering (LS) contribution are also shown. (B) Calculated second-derivative spectra of A. The second-derivative spectra of cystine and light-scattering components are essentially zero. Inset: Enlargement of the marked segment illustrating the interpolation between adjacent second-derivative points that yields the precise position of the intersect with the wavelength axis.
molecular weight of the protein. For a crude estimation, use the average value of $E_{0.1\%} = 1.1$.

$$A_{280,c,p} = \frac{\% P \times E_{0.1\%}}{\% P \times E_{0.1\%} + \% N \times 10}$$

Equation 7.2.5

6. Use the spectral intensity values at 280, 320, and 350 nm to assess the concentration of protein based on the known amino acid composition of the protein using Equation 7.2.6 and Equation 7.2.7 (Mach et al., 1992b), where $C$ is the protein concentration (molar), $\varepsilon$ is the molar extinction coefficient ($\text{M}^{-1} \text{cm}^{-1}$) that would be obtained from a solution of 1 M concentration and a 1-cm path length (L), and $N_{\text{Trp}}, N_{\text{Tyr}},$ and $N_{\text{S-S}}$ are the numbers of Trp, Tyr, and disulfide bonds (cystine residues) in the protein. $A_{280,c}$ is the optical density at 280 nm (OD$_{280}$) corrected for light scattering (see Equation 7.2.4). When sufficient amount of protein is available, the extinction coefficient $\varepsilon$ is usually determined by quantitative amino acid analysis (UNIT 3.2).

$$C = \frac{A_{280,c}}{\varepsilon L}$$

Equation 7.2.6

$$\varepsilon = 5540N_{\text{Trp}} + 1480N_{\text{Tyr}} + 134N_{\text{S-S}}$$

Equation 7.2.7

### Table 7.2.2  Theoretical R(260/280) Values vs %P and %N in Mixtures of Nucleic Acids and Proteins$^a$

<table>
<thead>
<tr>
<th>%P$^b$</th>
<th>%N$^b$</th>
<th>R(260/280)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0</td>
<td>0.57</td>
</tr>
<tr>
<td>99</td>
<td>1</td>
<td>0.70</td>
</tr>
<tr>
<td>98</td>
<td>2</td>
<td>0.81</td>
</tr>
<tr>
<td>97</td>
<td>3</td>
<td>0.90</td>
</tr>
<tr>
<td>96</td>
<td>4</td>
<td>0.99</td>
</tr>
<tr>
<td>95</td>
<td>5</td>
<td>1.06</td>
</tr>
<tr>
<td>90</td>
<td>10</td>
<td>1.32</td>
</tr>
<tr>
<td>80</td>
<td>20</td>
<td>1.59</td>
</tr>
<tr>
<td>60</td>
<td>40</td>
<td>1.81</td>
</tr>
<tr>
<td>40</td>
<td>60</td>
<td>1.91</td>
</tr>
<tr>
<td>20</td>
<td>80</td>
<td>1.97</td>
</tr>
<tr>
<td>0</td>
<td>100</td>
<td>2.00</td>
</tr>
</tbody>
</table>

$^a$Amount of DNA as small as 1% can be detected. This is due to the approximately tenfold higher values of the DNA extinction coefficient (i.e., $E_{0.1\%}$ of 10 and 20 at 280 nm and 260 nm, respectively). Average value of R for 12 globular proteins was 0.57 +/- 0.06. The empirical equation for this table is: $%N=(11.16\times R−6.32)/(2.16−R)$ (Glasel, 1995; see also Manchester, 1995, for further comments).

$^b$Abbreviations: %N, percent weight of DNA in the sample; %P, percent weight of protein in the sample.
Based on studies of a series of proteins (Equation 7.2.7), the above relationships can be used to determine the extinction coefficients of proteins and, consequently, their aqueous concentration to an accuracy generally within 2%.

**Calculate and visually inspect the second-derivative spectra**

7. Calculate the second-derivative spectrum. If the instrument’s software has derivative computation capability, choose derivative order 2, polynomial order 2, and a window of five data points. Otherwise use Equation 7.2.8 (Savitzky and Golay, 1964; Steiner et al., 1972), where SD(\(\lambda\)) is the second-derivative value at integral wavelength \(\lambda\), \(A(\lambda \pm n)\) is the absorbance value at \(\lambda + n\) wavelength (where \(n = -2, -1, 0, 1\) and 2), and \(k\) is the data point interval (in nanometers).

\[
SD(\lambda) = \frac{2[A(\lambda - 2)] - A(\lambda - 1) - 2[A(\lambda) - A(\lambda + 1)] + 2[A(\lambda + 2)]}{7 \times k}
\]

Equation 7.2.8

See Basic Protocol 1, step 6 annotation, for an example. Note that the values of \(n\) double for 2-nm spectral bandwidth.

8. Inspect the features of the resultant second-derivative spectrum.

The negative peak centered near 290 nm (Fig. 7.2.2) originates from the Trp inflection point observed in the zero-order spectrum. Although its position is partially influenced by the overlapping Tyr spectrum, the position of the negative 290-nm peak reflects the average polarity of Trp microenvironments: the more hydrophobic the environment, the higher the wavelength (i.e., an apolar environment produces a “red shift” in this peak; see Fig. 7.2.1). The magnitude of this negative peak is diminished by a Tyr band of opposite (positive) sign. In contrast, Trp and Tyr bands of the same sign (both negative) amplify the composite second-derivative peak near 280 nm. The positive peak centered at \(-285\) nm is composed mostly of the Trp component, as the second-derivative spectrum of Tyr crosses the zero axis in this region. In some proteins individual Tyr residues have microenvironments different enough to produce 2- to 3-nm shifts and band overlap sufficient to obviate individual peaks in this region. The alternating positive and negative peaks between 250 and 270 nm are the result of the wave-like fine structure from Phe side chains seen in the zero-order spectrum and are affected only marginally by Trp and Tyr counterparts.

The magnitudes of the second-derivative peaks are proportional to the concentration of aromatic amino acids, revealing both the relative content of each type of aromatic residue and the protein concentration. The relative heights of the second-derivative peaks constitute a characteristic “fingerprint” by which purity and structural integrity of a protein can be recognized. Some instruments (e.g., Hewlett-Packard 8452) have software capable of comparing two spectra using linear regression and providing a quantitative “match factor.” Unlike that for holmium oxide, the first derivative cannot be used for protein solutions because of residual nonzero values originating from light scattering, cystine, and other amino acid absorbance below 270 nm.

9. Calculate the positions of the intersections of the second derivative spectrum with the wavelength axis in the 285- to 295-nm region. Use Equation 7.2.9, where \(\lambda\) is the wavelength prior to the intersection point, \(k\) is the data point interval (in nanometers), and \(|SD(\lambda)|\) and \(|SD(\lambda + 1)|\) are the absolute values of the second derivative at the integral wavelengths \(\lambda\) (before the intersection point) and \(\lambda + 1\) (after the intersection point).

\[
\text{peak position (nm)} = \lambda + k \frac{|SD(\lambda)|}{|SD(\lambda)| + |SD(\lambda + 1)|}
\]

Equation 7.2.9
These values can be used as probes of protein structural integrity. In future manipulations or purification of the same protein, identical values of these parameters will indicate that the structure of the protein did not significantly change. The values of intersection points are typically reproducible within a 0.05-nm error margin using diode-array spectrophotometers. Scanning spectrophotometers require frequent wavelength calibration using standardized spectra to obtain this precision. The intersections of the second-derivative spectra of the holmium oxide standard determined using Equation 7.2.9 provide suitable reference points with which to compare results obtained at different times or using different instruments.

ANALYSIS OF PROTEINS USING AN HPLC WITH A DIODE-ARRAY DETECTOR

Purifying recombinant proteins relies on large-scale chromatographic procedures that generally lack the resolution of the analytical methods on which they are based. A common problem in the large-scale process is that a second protein (or some other compound, e.g., a detergent) may coelute with the protein of interest, contaminating the sample. High-performance liquid chromatography (HPLC) with a diode-array detector is a good method to determine the purity of a sample. By exquisitely regulating the pumps, this chromatography system minimizes flow problems that can otherwise interfere with peak detection, yielding highly reproducible gradients. Thus, HPLC provides the best chromatographic resolution of a protein.

A diode-array detector can be coupled to virtually any type of chromatographic separation, including reversed-phase, ion exchange (UNIT 8.2), and gel filtration (UNIT 8.3). The method assesses the purity of proteins or peptides equally well. In this protocol, the purity of a protein is determined using a C4 reversed-phase column. The same protocol can be adapted to peptides by substituting a C18 reversed-phase column.

Materials

- Buffer A: 0.1% (v/v) trifluoroacetic acid (TFA; reagent grade) in water (Milli-Q or equivalent purity)
- Buffer B: 0.09% (v/v) TFA in 80% acetonitrile (HPLC grade)
- C4 reversed phase column and guard: (e.g., Vydac C4, 250 × 4.6-mm-i.d. column, 5-µm particle size, 300-Å pore size for samples in the 0.5 to 5 nmol range or a 250 × 2.1-mm-i.d. column for samples in the 100 to 500 pmol range)
- HPLC with a diode-array detector (e.g., Hewlett-Packard, Waters, Beckman)

1. Set scan limits on the diode array detector to cover the 200- to 300-nm region.

   Peptide bonds have strong absorption in the 200- to 230-nm region (wavelength maximum = 188 nm) and aromatic amino acids have specific maxima in the 240- to 300-nm region.

2. Set the primary sample wavelength at 215 nm, with secondary tracings being made at 254 nm and 280 nm. Set bandwidth at 4 nm.

   The sample wavelength is the wavelength at which absorbance is measured as a function of time for the whole run. The bandwidth can be set from 4- to 20-nm for a sample. A bandwidth of 4 nm for the primary wavelength of 215 nm means that the absorbance measurement will extend between 213 and 217 nm.

   It is important to set the primary wavelength to measure peptide bonds. The peak absorbance (188 nm) cannot be used, however, because many of the HPLC reagents also absorb strongly in this region and obscure the signal from the protein. At 215 nm, the solvents do not interfere with the protein.

   The secondary tracings cover the absorption of aromatic amino acids. Although having
the whole chromatogram at these secondary wavelengths is not essential for this method, it can be useful (see Anticipated Results).

3. Set reference wavelength at 550 nm with a bandwidth of 100 nm.

The reference wavelength is generally set outside the range of protein absorption. Because proteins do not absorb at this wavelength, a wide bandwidth is recommended to avoid spurious readings. The absorbance of the reference wavelength is subtracted from the absorbance of the sample wavelength to give the absorbance of the signal for the protein.

4. Determine the points at which diode-array spectra will be taken and saved. Set acquisition and storage of spectrum to “peak controlled.”

This means that diode-array spectra will be taken at a point on the upslope, at the apex, and at a point on the downslope of a peak. The information obtained will be used to assess peak purity. If a peak has shoulders or is asymmetrical, spectra will automatically be obtained at additional points using the peak purity program.

5. Set threshold to define height of signal expected for the smallest peak from 0.1 to 999 milliabsorbance units (mAU).

A threshold of 5 mAU is suggested for a sample of 50 pmol protein.

This parameter is relevant to the size of the record that will be made and saved during the run. Diode-array spectra will be saved only for peaks that are higher than the threshold height defined, not for peaks that are lower.

6. Set peakwidth to equal the half-height width expected for the narrowest peak, with limits from 0.007 to 10 min.

A suggested starting value for this parameter is 0.15 min. Together with the threshold, the peakwidth determines the size of the record made for the run. Diode-array spectra will be saved only for peaks displaying a wider half-height width than the value chosen, not for peaks narrower than this width. This eliminates saving spectra for peaks that are due to noise.

7. Set stop time for acquisition of spectra.

This value controls the diode-array detector only; it can be set to stop independently from the run, earlier than the run, or automatically with the run.

8. Set a gradient rate of 1% buffer B/min and a flow rate of 1 ml/min for the 4.6-mm-i.d. column, 200 µl/min for the 2.1-mm-i.d. column. Set the amount of buffer A plus the amount of buffer B to equal 100%.

HPLC pumps are controlled by the B pump, which can vary from 0% to 100%. The amount of buffer A plus the amount of buffer B must equal 100%. Thus 0% buffer B means that the B pump is off and only the A pump is working (e.g., 100% A), and 100% buffer B means that the A pump is off and only the B pump is working. To achieve the same chromatographic separation on a 2.1-mm-i.d. column as obtained with a 4.6-mm-i.d. column, adjust the flow rate according to the ratio of the cross-sectional areas of the two columns. A 4.6-mm-i.d. column is roughly five times the cross-sectional area ($\pi r^2$) of a 2.1-mm-i.d. column.

Gradient rate defines the change in the gradient over time. The gradient should not be started until the sample has had time to flow onto the column. For a sample 100 µl in volume and a flow rate of 1 ml/min, it will take 0.1 min to load the sample onto the column. The following table illustrates typical gradient conditions for a 100-µl sample and a flow rate of 1 ml/min. This information would be entered into a liquid chromatography (LC) time table. The time required for the run should be determined and entered at the appropriate place in the program. This time limit should be long enough to allow for all of the events specified in the LC time table.
<table>
<thead>
<tr>
<th>Time (min)</th>
<th>% B</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>51</td>
<td>60</td>
</tr>
<tr>
<td>56</td>
<td>80</td>
</tr>
<tr>
<td>66</td>
<td>10</td>
</tr>
</tbody>
</table>

*After a trial run, if peaks of interest are too close to each other, the gradient rate can be made shallower (e.g., 0.5% buffer B/min) to spread out the gradient or steeper (e.g., 1.5% buffer B/min) to sharpen well-resolved peaks.*

9. Equilibrate the column with 10% buffer B by running a volume equivalent to at least 3 column volumes.

10. Prepare experimental and control samples in 10% buffer B, keeping the volume to a minimum and making all samples the same volume.

   *A good control sample will contain everything present in the experimental sample except the protein. The following table illustrates a sample preparation:*

<table>
<thead>
<tr>
<th>Control</th>
<th>Experimental</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein sample — 90 µl</td>
<td>— 90 µl</td>
</tr>
<tr>
<td>Sample buffer 90 µl —</td>
<td>90 µl —</td>
</tr>
<tr>
<td>HPLC buffer B 10 µl —</td>
<td>10 µl 10 µl</td>
</tr>
</tbody>
</table>

*Quantities of protein in the sample can range from 0.5 to 5 nmol for a 4.6-mm-i.d. column and from 100 to 500 pmol for a 2.1-mm-i.d. column. For reversed phase chromatography, sample volumes of 20 to 100 µl will work.*

11. Filter samples through a low-protein-absorption 0.21-µm pore size filter or microcentrifuge 5 min at maximum speed to remove particulates.

12. Choose an appropriate label for the control run. Inject the control sample into the HPLC apparatus.

   *The control sample is used to control for spectral impurities in buffers. The control should be run before the experimental sample to permit careful inspection of the entire gradient. The HPLC starts acquiring data on injection. If a rheodyne injector is used, leave the sample loop on line long enough to allow the whole sample to be applied to the column.*

13. Choose an appropriate label for the experimental run. Inject the experimental sample into the HPLC.

   *As above, the injection of the sample starts the acquisition of data. The data is stored in a computer file under the given label.*

**Data analysis**

14. Subtract the UV spectrum of the control from that for the protein to take into account impurities found in the sample buffer.

15. Run the peak purity program, which compares spectra for points at the apex and on the upslope and downslope of each peak.

16. Compare normalized overlaid spectra.
REAGENTS AND SOLUTIONS

Use Milli-Q-purified water or equivalent for the preparation of all buffers. For common stock solutions, see APPENDIX 2E; for suppliers, see SUPPLIERS APPENDIX.

Acid/ethanol cleaning solution (3 N HCl/50% ethanol)
Add 50 ml of 12 N HCl to 50 ml deionized water. Add the resulting solution to 100 ml ethanol. Store several months at room temperature.

CAUTION: Never pour water or ethanol into the concentrated acid. Always pour the acid into water or ethanol. Wear appropriate personal protection.

Holmium oxide reference solution
4% holmium oxide in aqueous solution of 10% (v/v) perchloric acid sealed in a nonfluorescent fused silica cuvette of optical quality (National Institute of Standards and Technology, Standard Reference Materials catalog; see SUPPLIERS APPENDIX). Store at room temperature for several months.

COMMENTARY

Background Information
Ultraviolet absorption spectroscopy played a key role in early investigations of protein structural transitions (Wetlaufer, 1962). The explosive growth of molecular biology and expanded use of more powerful spectroscopic techniques such as nuclear magnetic resonance (NMR), X-ray crystallography, circular dichroism, and fluorescence relegated this technique to use in simple assays and concentration determinations. During this period, however, electronic and optical systems improved considerably, allowing extraction of much more detailed information from simple near-UV spectra. It was soon shown that the magnitude of the near-UV second derivative peaks can be used for quantitative determination of the content of aromatic amino acids in proteins (Ichikawa and Terada, 1977; Servillo et al., 1982, Levine and Federici, 1982; Nozaki, 1990). These methods employ either peak-to-peak distances or least-squares multicomponent analysis and are capable of resolving all three components simultaneously. In fact, the spectrum of a mixture of three proteins can be easily resolved by this method if the spectra of the individual proteins are known (Mach et al., 1989). The composition of simple protein mixtures and complexes can then be monitored without actual chromatographic separation.

Generally, the spectra of the individual aromatic amino acids become red-shifted (move to higher wavelengths) upon incorporation into protein structures (see Fig. 7.2.1). These shifts can be as large as 5 nm for Tyr, 4 nm for Trp, and 2 nm for Phe residues (Mach and Middaugh, 1994). Of course, many side chains naturally exist in quite polar microenvironment, so these shifts are not always so large. Unfolding of a protein usually results in some increased solvent exposure of aromatic side chains and, consequently, in blue shifts (toward lower wavelengths). If the spectral signal near 295 nm (Fig. 7.2.2) shifts left even a fraction of a nanometer, the absorbance value at any given wavelength in this vicinity may change significantly because of the steepness of the slope. This phenomenon can be effectively utilized to follow the unfolding of a protein by measuring the spectral intensity at this wavelength as a function of a structure-perturbing variable (e.g., temperature, pH, protein concentration, or urea concentration) or by measuring the difference spectra between two proteins of identical concentration, one of which is subject to structural perturbation. The resulting values, however, are relative. Use of the intersection of a portion of the second-derivative spectrum with the wavelength axis is therefore recommended.

Application of more complex fitting routines and use of systematically shifted standard spectra allow the individual spectra of Trp, Tyr, and Phe to be resolved from the composite protein spectrum (Mach and Middaugh, 1994). This allows the fractional exposure of aromatic residues to be assessed using additions of apolar compounds as perturbants. The observed shifts are compared to the shifts in model, solvent-exposed compounds (e.g., acetylated and amidated aromatic amino acid derivatives), and the solvent exposure is estimated from a comparison of these values. In addition to proteins, DNA and RNA possess strongly absorbing chromophores (purine and pyrimidine bases) in the near-UV (Table 7.2.1), resulting in...
in the well-known featureless spectrum of nucleic acids with a peak near 260 nm. The molar absorptivity (extinction coefficient) of nucleotides is strongly structure dependent and this is often utilized in monitoring thermal structural transitions (i.e., in melting experiments). In addition, DNA and RNA have well-defined, albeit weak, second-derivative spectra that can be employed to determine the amount of protein in the presence of DNA (Mach et al., 1992a).

Very strong absorbance is also produced by all proteins in the far-UV range (180 to 250 nm). This is primarily the result of the intense absorption by the peptide bond in the far-UV. Because of amide absorption near 188 to 190 nm, a peak is actually present, but the intense absorption by oxygen as well as most solutes in this region makes it difficult to resolve experimentally. In some spectrophotometers, this peak can be resolved by flushing the instrument with N₂ and by the use of UV-transparent anions (e.g., phosphate or perchlorate), but this is rarely done today. Although the structure and position of this peak contain secondary-structure information as well as side-chain contributions, the peak is rarely used experimentally because of the superiority of circular dichroism measurements in the same region. It is occasionally employed, however, for concentration measurements or simple protein detection during chromatography by monitoring wavelengths on the slope of the peak (e.g., 210 and 215 nm). If the protein or peptide studied lacks aromatic amino acids, or the concentration of the sample does not produce sufficient signal, its concentration can be calculated by employing a set of average extinction coefficient values in the far-UV range that were obtained based on the analysis of twelve proteins (Scopes, 1974):

(1) \( E(1 \text{ mg/ml, 1 cm}) \) at 205 nm = 31 with stated error of 5%  
(2) \( E(1 \text{ mg/ml, 1 cm}) \) at 205 nm = 27.0 + 120 \( A_{280}/A_{205} \) with stated error of 2%.

Extreme care should be taken when correcting for the many buffer components absorbing in this region.

In principle, methods such as SDS-polyacrylamide gel electrophoresis (SDS-PAGE; UNITS 10.1 & 10.8) are more effective than UV spectroscopy in assessing the purity of a protein sample, based on detection of the presence and absence of additional bands. Such methods, however, may fail to ascertain contaminating proteins of similar molecular weight, unless isoelectric focusing (IEF) gels are used (UNITS 10.2 & 10.8). In contrast to electrophoretic (Chapter 10) or chemical (Chapter 11) analyses, the UV absorption technique is both nondestructive and rapid. Moreover, only a modest amount of sample is required.

This unit describes recombinant protein analyses using the two general types of UV spectrophotometers (Fig. 7.2.3): conventional scanning instruments, in which light absorption is measured by mechanically moving a monochromator allowing data to be accumulated temporally as a function of wavelength (Basic Protocol 1 and Support Protocol) and diode-array spectrophotometers (Basic Protocol 2), in which a light beam containing all frequencies is simultaneously split onto an array of detectors (diodes) after passing through the sample. Diode-array spectrophotometers (Hewlett-Packard 8450 to 8453 series or Beckman DU 7000) are simpler and consequently have excellent wavelength reproducibility. A single spectrum is typically acquired in 0.1 sec; longer acquisition times provide an average of a number of measurements. Spectral bandwidth is fixed at 1 or 2 nm. The small cross-sectional area of the light beam (−1×5 mm for the Hewlett-Packard diode-array instrument) allows small sample volumes and free choice of cuvette format. Such instruments can be easily used on-line to collect and analyze spectra of protein fractions leaving a chromatographic column—e.g., from gel filtration (UNIT 8.8), ion-exchange (UNIT 8.2), and reversed-phase media. In contrast, conventional scanning spectrophotometers require longer acquisition times, since at any given time only one wavelength is read.

**Critical Parameters**

The spectrophotometers contain moving internal parts that require frequent wavelength calibration. Of great importance when using scanning spectrophotometers is the setting of proper detector response time relative to scanning speed. A decrease in absorbance values, broadening of bands, and shifts are often observed in such cases.

Although the intrinsic absorbance of the aromatic amino acids varies little with temperature, it is advisable in many cases to control the temperature of the sample. Many processes destructive to proteins (e.g., oxidation, deamidation, unfolding, and aggregation) are accelerated at elevated temperature (Volkin and Middaugh, 1992). Under most conditions of humidity, however, a temperature of 15°C should be regarded as minimal. Decreasing the
temperature below this level may result in water condensing on the optical surface of the cuvette. The power of deuterium lamps used in UV spectrophotometers is usually not sufficient to cause photooxidation of aromatic residues. However, if the fluorescence spectra were acquired using the same sample, some of the tryptophan residues may have been oxidized to N-formylkynurenine and kynurenine by light from more powerful xenon lamps. This condition can be recognized by the appearance of a broad spectral band extending over 350 nm (similar to light scattering) and can be identified by comparing spectra as a function of irradiation time in a fluorometer or by direct observation of photoproduct fluorescence bands at \( \sim 400 \) nm (with excitation at 320 nm). Solvent-accessible tryptophan side chains are particularly prone to photooxidation, with half-lives as short as several minutes (Mach et al., 1995).

An aqueous buffer in which the protein of interest is stable should always be used. For example, phosphate-buffered saline, pH 7.4 (PBS; Appendix 2E), or an equivalent buffer containing 10 mM sodium phosphate and 120 to 150 mM sodium chloride, pH 7.0 to 7.4, is often employed. Choice of pH is important, because the spectral contribution of absorbing residues may be directly or indirectly (effect on protein conformation) pH dependent. Whereas tyrosine ionization (pK=10) results in an entirely different spectrum (\( \lambda_{\text{max}} = 294 \) nm), ionization of nonabsorbing or weakly absorbing side chains (e.g., lysine, pK\( \sim \)10; aspartic and glutamic acids, pK\( \sim \)4.5; histidine, pK\( \sim \)6.5; or cysteine, pK\( \sim \)8.5) may produce subtle shifts if the microenvironments of UV-absorbing chromophores or the overall conformation are altered. The pH should be sufficiently different from the isoelectric point to ensure good solubility. Note that many agents used to improve protein solubility (e.g., surfactants, denaturing agents, and sugars; Middaugh and Volk, 1992) may affect spectral band positions.

For HPLC analysis, it is important that HPLC-grade solvents, free of UV-absorbing contaminants, be used. Similarly, the sample

---

Figure 7.2.3  Basic scheme for spectrophotometric measurements: lamp, sample and detector; b (path) and db are shown in the sample and I and I\(_0\) are denoted (see below). (A) Diode-array and (B) conventional scanning spectrophotometers. The relative decrease of the intensity of the light beam after passing through the sample is proportional to the distance (db) and the concentration of the absorbing component: \(-\text{db}/b = k\times\text{db}\), where k is an absorption coefficient. Upon integration, the result is \(\ln(I/I_0) = -kl\). The fraction of light transmitted through the sample is known as the transmittance T = I/I\(_0\). Absorbance is defined as A = log I\(_0\)/I = \(\varepsilon cb\), where \(\varepsilon\) is a molar absorptivity (extinction coefficient), c is the molar concentration, and b is the path length (in centimeters). Alternatively, weight extinction coefficients E\(_1\) cm, 0.1% expressed per 1 mg/ml (0.1%) concentration are used.
Determining the Proteins by UV Spectroscopy

Troubleshooting

A few recurring problems are encountered in the analysis of proteins by UV spectrophotometry. A major, often unrecognized difficulty is the phenomenon known as absorption flattening. Imagine a large number of protein molecules aggregating so that some incident light is able to pass “in between” them without encountering an absorbing entity. However, because of the high concentration of chromophores in the aggregated particle, photons that happen to encounter the aggregate are unlikely to emerge on the other side. As a result, the intensity of such spectra are significantly reduced and often distorted. If this problem is suspected, the length of the effective light path should be changed. If absolute values of protein concentration are found to be pathway dependent, this phenomenon may be present.

The presence of anomalies such as significant curvature, negative values, or “bumps” in the 300- to 350-nm spectral region indicates either technical problems or the presence of unknown absorbing components, perhaps due to contaminating solutes. The spectra of any suspected absorbing impurities can be measured and compared to the experimental spectra to help determine if impurities are present. Detergents are often found to produce such extraneous absorbance, and their potential contribution should be evaluated if they are employed at any stage in the purification process.

If the second-derivative spectra do not resemble the examples shown in Figures 7.2.1 and 7.2.2, the most probable reason is too wide a bandwidth. Typically, a 2-nm bandwidth is the largest value permitting accurate derivative calculation. If the derivative spectrum is calculated by the instrument’s software, probably more than five data points (e.g., 7 to 21) are used for calculation of each value or the polynomial order is different from 2. If the calculated second-derivative spectra resemble random noise, the signal-to-noise ratio of the normal (zero-order) spectrum is probably too low to allow this type of analysis. This problem may be overcome by increasing spectral acquisition times or averaging a number of spectra.

Problems in obtaining a stable baseline may arise from a number of factors. If the incorrect lamp was turned on—the tungsten-halogen (or “visible”) instead of the deuterium (or “UV”) lamp—spectra will not be obtainable. Failure of the UV lamp will usually manifest itself in an inability to obtain zero absorbance values during a “blank” measurement after collecting the baseline with both cuvettes containing the buffer. Many instruments contain an indicator of this condition. Care should be taken not to directly inspect the deuterium lamp for ignition, as UV irradiation can seriously damage the retina of the eye. The cuvette itself should also be reexamined. Only synthetic quartz provides sufficient transparency over the UV region. The inadvertent use of a glass cuvette will block passage of light through the system. If spectra cannot be obtained, the cuvette may have been positioned incorrectly, with opaque or black walls perpendicular to the light beam.

Instability of the absorbance signal can occur for a number of other reasons, such as liquid smears, water vapor condensation or fingerprints on the optical surfaces, or air bubbles or other suspended or sedimenting particles in the test solution. As indicated previously, the meniscus should be above the upper edge of the beam and no light should be able to bypass the sample. In some cases, the problem of insufficient volume of sample available can be solved by inserting some space-filling material beneath the cuvette.

In diode-array detection schemes, it is important to run a control sample consisting of the solution in which the protein is dissolved, to detect possible contaminants. Threshold and
peak width values should be set correctly to ensure that all peaks of interest are seen by the diode-array detector. Otherwise, small contaminating peaks may be skipped.

**Anticipated Results**

Figure 7.2.2A illustrates a typical example of a near-UV protein spectrum. Despite differences in aromatic amino acid compositions in proteins, spectra are usually qualitatively similar, with a maximum falling between 276 and 282 nm and only marginal absorbance above 300 nm. Small soluble monomeric proteins do not usually produce any light scattering. The second-derivative pattern should have the distinct features shown in Figure 7.2.2B characteristic of tryptophan, tyrosine, and phenylalanine residues, if these are present in the sequence.

The position of the intersect with the wavelength axis of the declining second-derivative region near 290 nm is usually found at 289.41 ± 0.61 nm using 1-nm data point spacing (based on the average from 14 tryptophan-containing proteins). This value should be ~0.5 nm lower when the 2-nm data point interval is used. The reproducibility should be within 0.05 nm for pure and structurally intact proteins.

The diode-array detector obtains spectra of very high quality and precision, enabling the operator to check the purity of a protein by comparing normalized spectra from different points of the eluted peak. It uses information within the protein (i.e., its amino acid composition) to check purity. Proteins of differing composition will have different spectral absorptions, particularly if they vary widely in aromatic amino acid content. Thus, if the retention times of two proteins are close enough, the proteins appear to be in the same peak. Based on their absorption characteristics the diode-array detector can distinguish them by comparing normalized spectra at the front end of the peak on the upward slope, at the back end of the peak on the downward slope, and at the apex. If the protein is pure, the spectra collected from these three points will superimpose exactly. If the protein is impure, the spectra collected from the three points will deviate, aside from the rare case in which the impurity and the sample exactly coelute from the column. The general rule is that small-to-moderate levels of impurity can be detected in a contaminated sample, provided the spectral properties of the contaminant vary from the recombinant protein.

The examples in Figures 7.2.4 and 7.2.5 illustrate pure and impure samples. Figures 7.2.4A and 7.2.5A show the main chromatography tracings at 215 nm. The first visible peak in each chromatogram is the injection peak and the second is the protein of interest. Figures 7.2.4B and 7.2.5B show the stored diode-array spectra for points collected on the upswing, the apex, and the downswing of the protein peak of interest. A pure protein is characterized by identical spectra at these points, whereas an impure protein will have different spectra at each point. In Figure 7.2.4B, the spectra are close to identical, indicating that the protein is fairly pure. In Figure 7.2.5B, on the other hand, the spectra vary widely. Diode-array spectra for points 1 and 2 were collected early in the peak. These spectra have little absorbance in the region from 200 to 220 nm, indicating that a nonproteinaceous component is eluting at the front end of the BSA peak. Peaks 3, 4, and 5 correspond to the upswing, the apex, and the downswing, respectively, of the main peak. The diode-array spectra of these points are different enough to indicate contamination. Figures 7.2.4C and 7.2.5C show the superposition of tracings measured at the primary, secondary, and tertiary wavelengths. Note that these closely superimpose for the pure sample (Fig. 7.2.4C) but not for the impure sample (Fig. 7.2.5C).

**Time Considerations**

The entire procedure for measurement and analysis of protein UV spectra can usually be completed within 30 min, of which 20 min is needed for lamp warm-up and sample preparation. This time should be shorter if the instrument is already turned on. Diode-array spectrophotometers interfaced to computers equipped with data processing programs are especially time efficient when a larger number of samples is analyzed. Characteristically, even the most complex analysis routines, when implemented within the existing software, usually require only seconds for completion. In the case of conventional scanning instruments, calibration (measurements of standards) combined with slower scanning rates may require some additional time.

An HPLC run takes ~2 hr, including the column equilibration with buffer A. Examination of the diode-array detector information should take at most 1 additional hour for the novice and 30 min for the experienced operator.
**Figure 7.2.4** Spectra of a pure protein sample (carbonic anhydrase). (A) Main chromatography tracings at 215 nm. Arrow indicates retention time of the component analyzed. (B) Stored diode-array spectra for points collected on the upswing, the apex, and the downswing of the protein peak. (C) Superposition of tracings measured at the primary, secondary, and tertiary wavelengths.
Figure 7.2.5  Spectra of an impure protein sample (bovine serum albumin). (A) Main chromatography tracings at 215 nm. Arrow indicates retention time of the component analyzed. (B) Stored diode-array spectra for points collected on the upswing, the apex, and the downswing of the protein peak. (C) Superposition of tracings measured at the primary, secondary, and tertiary wavelengths.
Determining the Identity and Purity of Recombinant Proteins by UV Absorption Spectroscopy

7.2.20

**Literature Cited**


**Key References**


Provides theoretical background and practical considerations in absorption spectroscopy.


Provides comprehensive discussion of factors involved in analysis of proteins using UV spectroscopy.


Provides formulas for smoothing and differentiation of data.
Steiner, J. et al. 1974. See above.
*Provides formulas for smoothing and differentiation of data.*

*Describes an application of second-derivative spectroscopy to analyze aromatic amino acid content employing a diode-array spectrophotometer.*

*Describes calculation of protein extinction coefficients based on sequence and provides a formula for light-scattering correction at 280 nm.*

Contributed by Henryk Mach and C. Russell Middaugh
Merck Research Laboratories
West Point, Pennsylvania

Nancy Denslow
University of Florida
Gainesville, Florida