ANALYTICAL BIOCHEMISTRY **182**, 319–326 (1989)

Calculation of Protein Extinction Coefficients from Amino Acid Sequence Data

Stanley C. Gill and Peter H. von Hippel

Institute of Molecular Biology and Department of Chemistry, University of Oregon, Eugene, Oregon 97403

Received June 12, 1989

Quantitative study of protein–protein and protein–ligand interactions in solution requires accurate determination of protein concentration. Often, for proteins available only in "molecular biological" amounts, it is difficult or impossible to make an accurate experimental measurement of the molar extinction coefficient of the protein. Yet without a reliable value of this parameter, one cannot determine protein concentrations by the usual uv spectroscopic means. Fortunately, knowledge of amino acid residue sequence and protomer molecular weight (and thus also of amino acid composition) is generally available through the DNA sequence, which *is* usually accurately known for most such proteins. In this paper we present a method for *calculating* accurate (to ±5% in most cases) molar extinction coefficients for proteins at 280 nm, simply from knowledge of the amino acid composition. The method is calibrated against 18 "normal" globular proteins whose molar extinction coefficients are accurately known, and the assumptions underlying the method, as well as its limitations, are discussed.

Accurate determination of protein concentrations lies at the heart of all quantitative measurements of biochemical interactions. To establish molecular mechanisms one must know the binding affinities and stoichiometries of the species involved. Binding constant measurements require a reasonable degree of accuracy in the determination of protein concentrations, but stoichiometry determinations are *critically* dependent on the accuracy of the underlying concentration measurements. The literature is rife with examples of studies where errors of the order of 10 to 25% or more in protein concentrations have lead to totally erroneous calculations of stoichiometry, especially for oligomeric systems containing large numbers (and often different kinds) of subunits. The accuracy of other types of experimental parameters, including the specific activity of enzymes, the specific radioactivity of labeled proteins, etc., can also be severely compromised by relatively small errors in protein concentration measurement.

The simplest and most accessible way to determine the concentration of a protein in a fairly well-defined solution is to use ultraviolet–visible spectrophotometry, combined with a knowledge of the extinction coefficient of the protein involved. The spectrophotometric measurement can generally be made with a standard error of ±1–3%. To obtain this level of accuracy in the protein concentration determination itself, however, requires a precise knowledge of the extinction coefficient of the protein, and this parameter is usually much less accurately known.

Many methods exist for the determination of extinction coefficients of proteins; these include dry weight, nitrogen determination (1,2), and spectral methods (3,4) for pure samples, as well as the Bradford (5) and Lowry (6) colorimetric techniques. However, few of these methods are useful to biochemists and molecular biologists, either because of the large quantities of protein required or because of the large degrees of error associated with their use.

Clearly what is needed is a method to obtain extinction coefficients of high accuracy *without* the need to squander large amounts of protein. Currently proteins are usually sequenced at the DNA level long before they are purified; thus an accurate determination of the amino acid composition of a given protein is often available *de novo*. In this paper we show that protein extinction coefficients can be *calculated* with considerable accuracy from such amino acid composition data.

0003-2697/89 $3.00

TABLE 1

Molar Extinction Coefficients of Model Compounds[a]

Model compound	Extinction coefficient at 276 nm	278 nm	279 nm	280 nm	282 nm
N-Acetyl-L-tryptophanamide	5400	5600	5660	5690	5600
Gly–L–Tyr–Gly	1450	1400	1345	1280	1200
Cystine	145	127	120	120	100

[a] In 6.0 M guanidinium hydrochloride, 0.02 M phosphate buffer, pH 6.5. Values are taken from Table IV or extrapolated from Fig. 1 of Edelhoch (11). Extinction coefficients are in units of M^{-1} cm^{-1}.

The method we use to determine protein extinction coefficients is not new. It is based on combining amino acid composition information with spectral data on proteins and model compounds measured in both native and denaturing solvents and has been employed by a number of workers including Elwell (7), Butler *et al.* (8), and others. What *is* new is the development of cloning and DNA sequencing methodology, which has transformed the determination of the amino acid composition and the subunit molecular weight of a new protein from the most tedious and imprecise to the simplest and most accurate step in the analysis.

These developments make it possible, in principle, to determine extinction coefficients with great accuracy by spectral techniques, and, in particular, to determine whether the assumptions underlying the *calculation* of extinction coefficients from amino acid composition data are valid. By surveying a number of proteins for which extinction coefficients have been accurately determined, we conclude that the underlying assumptions *are* valid within very acceptable limits of error, and thus that the method of calculation we present here can be applied with considerable confidence to newly isolated proteins.

MATERIALS

Proteins and chemicals. *Escherichia coli* rho protein was isolated and purified to >95% purity (estimated by Coomassie-stained gels) as described by Geiselmann *et al.* (9). *E. coli* NusA protein was obtained as described by Gill *et al.* (10) and purified to >98% (based on Coomassie-stained gels). Ultrapure guanidine hydrochloride was purchased from American Research Products Co.

Preparation of solutions and spectral measurements. Rho and NusA protein solutions were prepared by gravimetric dilution, based on the densities of protein stock solutions determined with an Anton-Parr oscillating densitometer (standardized against air and water).

Spectral measurements were made using a Hewlett-Packard 8450 uv–vis spectrophotometer thermostatted at 20°C. The optical densities of solutions of native rho protein were corrected for light scattering by determining correction parameters over the 320–380 nm spectral range and extrapolating these parameters into the uv.

METHODS AND RESULTS

Determination of extinction coefficients. The molar extinction coefficient of a denatured protein in 6 M Gdn[1]·HCl is calculated from the number of tryptophan, tyrosine, and cysteine residues per molecule it contains, using the molar extinction coefficients of the appropriate model compounds in this solution. The molar extinction coefficients of these model compounds were measured by Edelhoch (11) and are listed in Table 1 as a function of wavelength. Edelhoch showed that these residues are the only ones that contribute significantly to the measured optical density of a denatured protein over the 276–282 nm range. Using these data, the extinction coefficient of a denatured protein in 6 M Gdn·HCl can then be calculated using

$$\epsilon_{M,Gdn\cdot HCl} = a\epsilon_{M,Tyr} + b\epsilon_{M,Trp} + c\epsilon_{M,Cys}, \quad [1]$$

where ϵ_{Tyr}, ϵ_{Trp}, and ϵ_{Cys} are the molar extinction coefficients of tyrosine, tryptophan, and cysteine residues at the wavelength used (see Table 1), and a, b, and c are the number of each type of residue per molecule of protein.

To determine the molar extinction coefficient of the *native* protein, the absorbance spectra of the native and the denatured (in 6 M Gdn·HCl) protein are measured at identical protein concentrations. Utilizing Beer's law, we may write

$$Abs_{Gdn\cdot HCl}/\epsilon_{M,Gdn\cdot HCl} = C_{den}, \quad [2]$$

where $Abs_{Gdn\cdot HCl}$ is the measured optical density of the sample of denatured protein in 6 M Gdn·HCl, $\epsilon_{M,Gdn\cdot HCl}$ is the molar extinction coefficient of the denatured protein in this solvent, and C_{den} is the concentration of the denatured protein in the solution (in mol/liter). The equivalent equation for the native protein sample is

$$Abs_{nat}/\epsilon_{M,nat} = C_{nat}. \quad [3]$$

Since C_{nat} was originally set equal (experimentally) to

[1] Abbreviation used: Gdn, guanidine.

TABLE 2

Comparison of Calculated Molar Extinction Coefficients for Denatured Proteins with Measured Molar Extinction Coefficients for Native Proteins

Protein	Calculated[a] $\epsilon_{M,Gdn\cdot HCl}$	Measured $\epsilon_{M,nat}$	$\epsilon_{M,nat} - \epsilon_{M,Gdn\cdot HCl}/\epsilon_{M,nat}$
T4 Lysozyme	24,990	24,170	−3.4%
E. coli rho	14,770	14,980	+1.4%
E. coli NusA	29,760	27,200	−9.4%

[a] References for the amino acid sequence data used to calculate ϵ_M are listed in Table 3, footnote b. Extinction coefficients are in units of $M^{-1}\,cm^{-1}$.

C_{den}, we can combine Eqs. [2] and [3] to obtain the molar extinction coefficient of the native protein:

$$\epsilon_{M,nat} = (Abs_{nat})(\epsilon_{M,Gdn\cdot HCl})/(Abs_{Gdn\cdot HCl}). \quad [4]$$

We have determined the molar extinction coefficients of several proteins using this technique, and the results are summarized in Table 2. Table 2 shows that for these proteins there is a relatively small difference between the calculated value of $\epsilon_{M,Gdn\cdot HCl}$ and the measured value of $\epsilon_{M,nat}$; i.e., there is a minimal change in the hypo- or hyperchromicity of the protein at 280 nm on denaturation. This suggested that the molar extinction coefficients of native proteins at 280 nm might be directly calculated from amino acid composition information by applying the Edelhoch (6 M Gdn·HCl) spectral parameters to the tyrosine, tryptophan, and cysteine residues within the protein and *assuming* that $\epsilon_{M,Gdn\cdot HCl} = \epsilon_{M,nat}$.

To further test the validity of this approach we have compared the *calculated* molar extinction coefficients for the denatured protein ($\epsilon_{M,Gdn\cdot HCl}$) with measured values of the native molar extinction coefficients ($\epsilon_{M,nat}$) of a number of proteins taken from the literature. These extinction coefficients are mostly based on careful dry weight measurements; the data are presented in Table 3.

It is clear from the data summarized in Table 3 (for the 18 proteins investigated) that the native molar extinction coefficients calculated on this basis fall very close to the measured values, with an *average* standard deviation of ±3.8%, and a maximum deviation of 14.9%. Since the general accuracy of dry weight determinations is unlikely to exceed ±5%, this result suggests that for most experiments the molar extinction coefficient of a protein can be calculated with good accuracy directly from the amino acid composition data. Of course, for studies in which plenty of protein is available and results of the highest precision are required, the technique outlined above should be applied to actually *determine* the effects of hypo- or hyperchromic contributions of the tyrosine, tryptophan, and cystine residues to the native protein absorbance at 280 nm.

DISCUSSION

As a consequence of the widespread use of restriction mapping and DNA sequencing, the most accurate information available in preliminary studies of a newly defined protein is often the amino acid sequence and the protomer molecular weight. This information, coupled with the fact that certain regions of a protein spectrum contain contributions from relatively few specific amino acid residues, suggested that a general procedure for the calculation of the molar absorbance of a protein could be developed. The spectral region from 275 to 290 nm seemed appropriate, since only tyrosine, tryptophan, and cystine residues absorb appreciably in this range.

Wetlaufer (12), using the procedure of Fromageot and Schnek (13), first attempted to calculate molar extinction coefficients at 280 nm for 11 common proteins, using molar extinction coefficients for Tyr, Trp, and Cys measured with the free amino acids. A comparison of the calculated molar extinction coefficients with the experimentally determined values yielded an average standard deviation of ±11% and a maximum deviation of 37%. This level of error was obviously too large to make this an acceptable procedure for measuring protein concentrations, although the results did confirm the potential validity of such an approach. (We would guess, based on the much more acceptable limits of error of our comparable determinations, that Wetlaufer's results suffered mostly from an absence of appropriate model chromophores for the absorbing residues of the proteins, together with the large errors in measured Tyr, Trp, and Cys residue contents and protein molecular weights that characterized that era.)

Clearly the central requirement in developing an acceptable calculation procedure of this sort is to have available an accurate set of amino acid compositions and molecular weights for the calibrating proteins, together with a basis set of appropriate and carefully measured model compounds for the absorbing residues. Then, ideally, one must work under conditions where an equimolar solution of the model compounds exactly matches the absorbance spectrum of the protein under study. Edelhoch (11) developed model compounds that meet this criteria by showing that he could exactly match the spectra of several denatured proteins in 6 M Gdn·HCl in this way. We have used the Edelhoch model compound extinction coefficients (Table 1) in our calculations of the denatured protein extinction coefficients presented in Tables 2 and 3.

In Tables 2 and 3 we also list the measured extinction coefficients for these proteins. The results suggest that

TABLE 3

Molar Extinction Coefficients Calculated from Amino Acid Composition Data

Protein	Mol Wt.	Trp	Tyr	Cys	λ (nm)	$\varepsilon_{1\%}$ expt.	ε_M expt.	ε_M calc.	Δ (%)[c]	Ref.[a,b]
Aldolase	38,994	3	12	8	280	9.1	35,480	33,310	- 6.1%	1
(rabbit muscle)					280	9.38	36,580		- 8.9%	2
					280	8.4	32,750		+ 1.7%	3
					280	9.1	35,485		- 6.1%	4
Alcohol	36,712	5	14	8	280	12.1	44,420	47,330	+ 6.5%	5
dehydrogenase					280	14.6	53,600		-11.7%	6
(yeast)						12.6	46,260		+ 2.2%	7
Carboxypeptidase A	34,414	7	19	2	278	18.8	64,698	66,050	+ 2.1%	8
(bovine)										
Carboxypeptidase B	34,617	8	22	7	280	21.0	72,696	74,520	+ 2.4%	9
(bovine)										
Chymotrypsinogen A	25,670	8	4	10	282	20.3	52,110	50,600	- 2.9%	10
(Beef Pancreas)					280	20.0	51,340	51,840	+ 1.0%	11
Glyceraldehyde-	35,606	3	11	2	280	9.08	32,330	31,390	- 2.9%	12
3-phosphate-						8.6	30,621		+ 2.5%	13
dehydrogenase						9.4	33,470		- 6.2%	14
(yeast)[d]						8.94	31,832		- 1.4%	15
						8.6	30,621		+ 2.5%	16
Glutamate	55,755	4	18	6	280	9.3	51,852	46,520	-10.3%	17
dehydrogenase						8.9	49,622		- 6.3%	17
(bovine)						9.5	52,967		-12.2%	18
Insulin	5,734	0	4	6	280	10.0	5,734	5,840	+ 1.9%	19
(bovine)					278	10.6	6,078	6,362	+ 4.7%	20
					280		5,220	5,840	+11.9%	21
lac Repressor	34,612	2	8	3	280		22,500	21,980	- 2.3%	22
(*E. coli*)							23,880		- 8.0%	23
α-Lactalbamin	14,186	4	4	8	280	20.1	28,510	28,840	+ 1.1%	24
(bovine)						20.5	29,081		- 0.8%	25
β-Lactoglobulin	18,285	2	4	5	278	9.6	17,550	17,435	- 0.7%	26
(bovine)					280	9.7	17,740	17,100	- 3.6%	27
					280	9.5	17,371	17,100	- 1.6%	28
					278.5	9.66	17,663	17,350	- 1.8%	29
Lysozyme	14,314	6	3	8	280		37,932	38,940	+ 2.7%	30
(hen egg white)							37,717		+ 3.2%	31
Lysozyme (T_4)	18,700	3	6	2	280		23,900	24,990	+ 4.6%	32
Ovalbumin (Chicken)	42,756	3	10	6	280	7.01	29,972	30,590	+ 2.1%	33
Papain	23,426	5	19	7	278	25.0	58,570	55,490	- 5.3%	34
Ribonuclease A	13,693	0	6	8	278	7.38	10,105	9,416	- 6.8%	35
(Beef Pancreas)					280	6.95	9,508	8,640	- 9.1%	36
					277.5	7.2	9,859	9,220	- 6.5%	37
Serum Albumin	66,296	2	20*	35	278	6.58	43,623	43,645	+ 0.1%	38
(bovine)						6.68	44,300		- 1.5%	39
Serum Albumin	66,470	1	18	35	280	5.8	38,553	32,810	-14.9%	40
(human)					280	5.31	35,296	32,810	- 7.0%	41
					277.5	5.03	33,434	35,446	+ 6.0%	42
Trypsinogen	23,998	4	10	12	280		33,357	37,000	+10.9%	43
(bovine)										

Standard Deviation of Δ = 3.8%

TABLE 3—*Continued*

[a] **References for Protein Extinction Coefficients**

Aldolase

1 Baranowski, T., and Niederland, T.R., J. Biol. Chem. 180, 543 (1949)
2 Donovan, J. W., Biochemistry, 3, 67 (1964)
3 Sine, H. E., and Hass, L. F., J. Biol. Chem., 244, 430 (1969)
4 C. L. Sia & Horecker, B. L., Arch. Biochem. Biophys., 123, 186 (1968)

Alcohol Dehydrogenase

5 Negelein, E., and Wulff, H. J., Biochem. Z., 293, 351 (1937)
6 Buhner, M., and Sund, H., Eur. J. Biochem., 11,73 (1969)
7 Hayes, J. E., and Velick, S. F., J. Biol. Chem., 207, 225 (1954)

Carboxypeptidase A

8 Barzetzi, J. P., Sampath Kumarjk, S. V., Cox, S. J., Walsh, K. A. and Neurath, H. Biochemistry, 2, 1468 (1963)

Carboxypeptidase B

9 Cox, Wintersberger, and Neurath, H., Biochemistry, 1, 1078 (1962)

Chymotrypsinogen A

10 Nichol, J. C., J. Biol. Chem., 243, 4065 (1968)
11 Guy, O., Gratecos, D., Rovery, M., and Desnuell, P., Biochem. Biophys., 115, 404 (1966)

Glyceraldehyde - 3 phosphate dehydrogenase

12 Krebs, E. G., Methods Enzymol., 1, 407 (1955)
13 Jaenick, R., Schmid, D., Knof, S., Biochem. 7, 919 (1968)
14 Warburg, O., Christian, W., Biochem. Z., 303, 40 (1939)
15 Kirschner, V. K., and Voigt, B., Hoppe-Seylers Z. Physio. Chem., 349, 632 (1968)
16 Jaenicke, R., in Pyridine Nucleotide Dependent Dehydrogenases, Sund, Ed, Springer-Verlag Berlin, p. 70 (1970)

Glutamate Dehydrogenase

17 Egan, R. R., and Dalziel, K., Biochem. Biophys. Acta., 250, 47 (1971) 1474[2]
18 Sund, H., and Akeson, A., Biochem. Z., 340, 421 (1964)

Insulin

19 Porter, R. R., Biochem. J., 53, 320 (1953)
20 Weil, L., Seibles, T. S., and Herskovitz, T. T., Arch. Biochem. Biophys., 111, 308 (1965)
21 M. Praissmah and Rupley, J. A., Biochemistry 7, 2431 (1968)

lac Repressor (*E. coli*)

22 Butler, A. P., Revzin, A., and von Hippel, P. H., Biochemistry, 16, 4757 (1977)
23 Barth, G., Bunnengerg, E., and Djerassi, C., Anal. Biochem., 48, 471 (1972)

α–Lactalbumin

24 Schmidt, D. V., and Ebner, K. E., Biochem. Biophys. Acta. 243, 273 (1971)
25 Krigbaum, W. R., and Kugler, F. R., Biochem. 9, 1216 (1970)

β–Lactoglobulin

26 Townend, R., Winterbotton, R. J., and Timasheff, S. N., J. Amer. Chem. Soc., 82, 3161 (1960)
27 Polis, D., Schmukler, H. W., Custer, J. H., and McMeekin, T. L., J. Amer. Chem. Soc., 72, 4965 (1950)
28 Wetlaufer, D. B., and Lovrien, R., JBC, 239, 596 (1964)
29 Baker, H. P., and Saroff, H. A., Biochemistry, 4, 1670 (1965)

Lysozyme (Hen)

30 Steiner, R. F., Biochem. Biophys. Acta., 79, 51 (1964)
31 Sophianopoulus, A. J., Rhodes, C. K., Holcomb, D. N., and Van Holde, K. E., J. Biol. Chem. 237, 1107 (1962)

Lysozyme T_4

32 Isugira, A., Inouye, M., Terzaghi, E., and Streisinger, G., J. Biol. Chem., 243, 391 (1968)

Ovalbumin

33 Weintraub, M. S., and Schlamowitz, M., Comp., Biochem., Physiol., 38B, 513 (1971)

Papain

34 Glazer, A. N., and Smith, E. L., J. Biol. Chem., 236, 2948 (1961)

TABLE 3—*Continued*

Ribonuclease A

35 Scott, R. A. and Scheraga, H. A., J. Amer. Chem. Soc., 85, 3866 (1963)
36 Sherwood, L. M., and Potts, J. T. Jr., J. Biol. Chem., 240, 3799 (1965)
37 Sela, M., Anfinsen, C. B. and Harrington, W. F., Biochim. Biophys. Acta, 26, 502 (1957)

Serum Albumin (bovine)

38 Noelken, M. E., and Timasheff, S. N., J. Biol. Chem., 242, 5080 (1967)
39 Foster, J.F. and Sterman, M.D., J. Amer. Chem. Soc., 78, 3656 (1956)

Serum Albumin (human)

40 Schoenenberger, V. M., Z. Naturforsch, 10B, 474 (1955)
41 Hunter, M. J., and McDuffie, F. C., J. Amer. Chem. Soc., 81, 4100 (1959)
42 Lerner, A. B., and Barnum, C. P., Arch. Biochem. Biophys., 10, 417 (1946)

Trypsinogen

43 Davie, E. W., and Neurath, H., J. Biol. Chem., 212, 515 (1955)

[b] References for Amino Acid Sequence Data

Aldolase	Dayhoff, M. O., ed. Atlas of Protein Sequence and Structure, vol. 5, suppl. 2, p 99, The National Biomedical Research Foundation, Baltimore, MD (1976)
Alcohol Dehydrogenase	Jomvall, H., Eur. J. Biochem., 72, 425, (1977)
Carboxypeptidase A	Bradshaw, R. A., Walsh, K. A., and Neurath, H., Biochemistry, 10, 938, 951 ,961 (1971)
Carboxypeptidase B	Titani, K., Ericsson, L. H., Walsh, K. A., and Neurath, H., Proc. Nat. Acad. Sci., U.S.A., 72, 1666 (1975)
Chymotrypsinogen A	Brown, J. R., and Hartley, B. S., Biochem. J., 101, 214 (1966)
Glyceraldehyde 3-Phosphate-Dehydrogenase (different genes)	Holland, J. P., and Holland, M. J., J. Biol. Chem. 254, 9839 (1979) Holland, J. P., and Holland, M. J., J. Biol. Chem. 255, 2596 (1980) Holland, J. P., Labieniec, L., Swimmer, C., and Holland, M. J., J. Biol. Chem., 258, 5291 (1983)
Glutamate Dehydrogenase	Moon, K., and Smith, E. L., J. Biol. Chem., 248, 3082 (1973) Witzemann, V., et al., Eur. J. Biochem., 43, 319 (1974)
Insulin (bovine)	Dayhoff, M. O., ed. Atlas of Protein Sequence and Structure, vol. 5, p. D-209, The National Biomedical Research Foundation, Baltimore, MD (1972)
lac Repressor (*E. coli*)	Dayhoff, M. O., ed. (1978) Atlas of Protein Sequence and Structure, vol. 5, suppl. 3, p. 253, The National Biomedical Research Foundation, Baltimore, MD.
α-Lactalbumin	Brew, K., Castellino, F. J., Vanaman, T. C., Hill, R. L., J. Biol. Chem., 245, 4570 (1970)
β-Lactoglobulin	Braunizer, G., Chen, R., Schrank, B.,, and Stangl, A., Hoppe-Seylers Z. Physiol. Chem., 354, 867 (1973)
Lysozyme (hen's egg)	Jung, A., Sipel, A. E., Grez, M., and Schutz, G., Proc. Nat. Acad. Sci., USA, 77, 5759 (1980)
Lysozyme (T_4)	Tsugita, A., and Inouye, M., J. Mol. Biol. 37, 201 (1968)
NusA Protein (*E. coli*)	Ishii, S., Ihara, M., Naekawa, T., Nakamura, Y. Uchida, H., and Imamoto, F., Nucl. Acids. Res. 12, 3333 (1984) Saito, M., Tsugawa, A., Egawa, K., and Nakamura, Y., Mol. Gen. Genet. 205, 380 (1986)
Ovalbumin	Wog, S. L. C., Beattie, W. G., Catterall, J. F. Duguiczyk, A., Stoden, R., Brownlee, G. G., O'Malley, B. W., Biochemistry, 20, 6437 (1981)
Papain	Dayhoff, M. O., ed. (1972) Atlas of Protein Sequence and Structure, vol. 5, p. D-121, The National Biomedical Research Foundation, Baltimore, MD.
Rho Protein (*E. coli*)	Pinkham, J. L., and Platt, T., Nuc. Acids Res. 11, 3531 (1983)

TABLE 3—*Continued*

Ribonuclease A	Smyth, D. G., Stein, W. H., and Moore, S., J. Biol. Chem., 238, 227 (1963)
Serum Albumin (bovine)	Brown, J. R., in Albumin Structure, Function and Uses, Rosender, V. M., Oratz, M., and Rothschild, M. A., eds. pp. 27-51, Pergamon Press, Oxford (1977)
	Levine, R. L., and Federici, M. M., Biochemistry 21, 2600 (1982)
Serum Albumin (human)	Dayhoff, M. O., ed. Atlas of Protein Sequence and Structure, vol 5, suppl. 3, p. 306, The National Biomedical Research Foundation, Baltimore, MD. (1978)
Trypsinogen	Mikes, O., Heleysovsky, V., Tomasek, V., and Sorm, F., Biochem. Biophys. Res. Commun., 24, 346 (1966)

[c] $\Delta = [\varepsilon_{M,\ expt.} - \varepsilon_{M,\ calc.}] \ / \ [\varepsilon_{M,\ expt.}]$

[d] Glyceraldehyde-3-phosphate-dehydrogenase is encoded by several non-tandem genes that vary in tyrosine content from 10 to 12 residues per protein molecule. We have used tyr = 11 as an average value for these calculations.

[e] We have used a value of 20 tyrosine residues for bovine serum albumin, instead of the 19 tyrosine residues reported from amino acid sequence, based on the arguments in Levine and Federici (1982).

the molar extinction coefficient for an average (as defined by these tables) protein may indeed be calculated from amino acid composition information, with an average standard deviation of ±5% from the experimentally determined value.

The central assumption in such an approach is, of course, that the spectral contributions of the tyrosine, tryptophan, and cystine residues that dominate the native protein spectrum around 280 nm are not significantly shifted in the native protein, relative to their contributions to the denatured protein spectrum. We know that this is not always true, since protein spectral peaks are often shifted in position and intensity upon denaturation, relative to the spectra for the native moieties. However, these changes rarely represent as much as 10% of the total absorbance of the protein solution from 275 to 280 nm, and the results in Table 2, as well as similar measurements on other proteins (see Table 3, below), show that accurate molar extinction coefficients *can* be obtained by this procedure.

The method we describe here also involves other assumptions. (i) We assume that the protein contains no chromophores (other than Tyr, Trp, and Cys) that absorb at 280 nm. This means that the concentration of conjugated proteins (e.g., catalase, hemoglobin, or peroxidase) that contain prosthetic groups absorbing in the near uv and visible portions of the spectrum cannot be analyzed by this approach. (ii) It is also assumed that the amino acid composition data used in the calculation are correct. Obviously errors in determining the number of Tyr, Trp, and Cys residues per protein molecule can result in large errors in the calculated molar extinction coefficient of the protein. Finally, (iii) since we assume that the amino acid composition data used will generally be derived from DNA sequencing results, we cannot know whether Cys residues will appear as cysteine or as cystine in the final native protein. The assumption we have made in our treatment is that all Cys residues appear as half cystines, which do contribute to the 280-nm absorbance of proteins (cysteine residues do not absorb appreciably at wavelengths > 260 nm; see Beaven and Holiday, (14)). However, this potential source of error should not have a large effect in most proteins, since the molar absorbance of the Cys residue at 280 nm is *much* smaller than those of the Tyr and Trp residues (see Table 1).

In summary, by using proper precautions to assess the validity of these assumptions, it appears that one can use the procedures described in this paper to obtain a calculated molar extinction coefficient for an unknown protein with a high degree of accuracy. This calculation approach is likely to yield much more accurate protein concentrations than experimental measurements with the inadequate amounts of incompletely purified protein that are generally available for newly isolated proteins of molecular biological interest. Of course, for critical determinations of (e.g.) subunit stoichiometries of large complexes, and comparably concentration-sensitive studies, results using calculated extinction coefficients should be confirmed with carefully measured parameters if at all possible. On the other hand, history suggests that if the measured extinction coefficient of a protein deviates very much from its value calculated as described here, the *measured* value is likely to be wrong.

ACKNOWLEDGMENTS

The research described in this paper was supported, in part, by USPHS Research Grants GM-15792 and GM-29158 (to P.H.v.H.), by a

grant to the University of Oregon from the Lucille P. Markey Charitable Trust, and by a traineeship (to S.C.G.) on USPHS Institutional Training Grant GM-07759. P.H.v.H. is an American Cancer Society Research Professor of Chemistry. We thank Professor John A. Schellman for useful discussions of the calculation procedures used in this paper.

REFERENCES

1. Johnson, M. J. (1941) *J. Biol. Chem.* **137,** 575–586.
2. Yeh, C. S. (1966) *Microchem. J.* **11,** 229–236.
3. Whitaker, J. R., and Granum, P. R. (1980) *Anal. Biochem.* **109,** 156–159.
4. Scopes, R. K. (1974) *Anal. Biochem.* **59,** 277–282.
5. Bradford, M. M. (1976) *Anal. Biochem.* **72,** 248–254.
6. Lowry, O. H., Rosebrough, N. J., Farr, A. L., and Randall, R. J. (1951) *J. Biol. Chem.* **193,** 265–275.
7. Elwell, M. L. (1976) Ph.D. thesis, Department of Chemistry, University of Oregon.
8. Butler, A. P., Revzin, A., and von Hippel, P. H. (1977) *Biochemistry* **16,** 4757–4768.
9. Geiselmann, J., Yager, T. D., Gill, S. C., and von Hippel, P. H. (1989) Manuscript in preparation.
10. Gill, S. C., Yager, T. D., and von Hippel, P. H. (1989) *J. Mol. Biol.,* submitted.
11. Edelhoch, H. (1967) *Biochemistry* **6,** 1948–1954.
12. Wetlaufer, D. B. (1962) *Adv. Protein Chem.* **17,** 375–378.
13. Fromageot, C., and Schnek, G. (1950) *Biochim. Biophys. Acta* **6,** 113–122.
14. Beaven, G. H., and Holiday, E. R. (1952) *Adv. Protein Chem.* **7,** 319–387.