Analytical Biochemistry **299,** 136–146 (2001)
doi:10.1006/abio.2001.5411, available online at http://www.idealibrary.com on IDEAL®

Online Size-Exclusion High-Performance Liquid Chromatography Light Scattering and Differential Refractometry Methods to Determine Degree of Polymer Conjugation to Proteins and Protein–Protein or Protein–Ligand Association States

Brent S. Kendrick,*[,1] Bruce A. Kerwin,* Byeong S. Chang,* and John S. Philo†
**Department of Pharmaceutics, Amgen, Inc., Thousand Oaks, California; and*
†*Alliance Protein Laboratories, Thousand Oaks, California*

Received March 7, 2001; published online November 10, 2001

Characterizing the solution structure of protein–polymer conjugates and protein–ligand interactions is important in fields such as biotechnology and biochemistry. Size-exclusion high-performance liquid chromatography with online classical light scattering (LS), refractive index (RI), and UV detection offers a powerful tool in such characterization. Novel methods are presented utilizing LS, RI, and UV signals to rapidly determine the degree of conjugation and the molecular mass of the protein conjugate. Baseline resolution of the chromatographic peaks is not required; peaks need only be sufficiently separated to represent relatively pure fractions. An improved technique for determining the polypeptide-only mass of protein conjugates is also described. These techniques are applied to determining the degree of erythropoietin glycosylation, the degree of polyethylene glycol conjugation to RNase A and brain-derived neurotrophic factor, and the solution association states of these molecules. Calibration methods for the RI, UV, and LS detectors will also be addressed, as well as online methods to determine protein extinction coefficients and *dn/dc* values both unconjugated and conjugated protein molecules.

Key Words: **online SEC-HPLC; light scattering; differential refractometry; pegylation; glycosylation; protein–ligand interactions.**

Significant challenges exist for characterizing the molecular mass of conjugated proteins such as glycosylated and polyethylene glycol (PEG)[2]-modified proteins. Additional challenges are found in characterizing their association states in solution. Empirical methods such as size-exclusion chromatography or dynamic light scattering can be run under physiological conditions, but only measure hydrodynamic radius. Interpreting hydrodynamic radius in terms of molecular mass for conjugated proteins has been elusive (1, 2). Sedimentation equilibrium (SE) ultracentrifugation is another useful technique; however, it relies on analysis of extremely pure samples that are stable for the length of the analysis (i.e., a few days if multiple speeds are used) (1, 3).

Many of the limitations associated with the above techniques have been overcome by using size-exclusion high-performance liquid chromatography (SEC-HPLC) coupled online with a combination of refractive index (RI) detection and ultraviolet (UV) detection (2, 4). Kunitani *et al.* demonstrated how the mass fraction of PEG on PEGylated protein or carbohydrate on a glycoprotein could be determined based on RI and UV peak areas. This analysis is dependent on comparing the RI and UV areas of baseline-resolved peaks to calibration curves for each component. Absorptive losses of any of the components will influence the final

[1] To whom reprint requests should be addressed at current address: Amgen, Inc., 4000 Nelson Road, MS AC-3A, Longmont, CO 80503. Fax: (303) 938-6241. E-mail: kendrick@amgen.com.

[2] Abbreviations used: PEG, polyethylene glycol; SE, sedimentation equilibrium; SEC-HPLC, size-exclusion high-performance liquid chromatography; RI, refractive index; LS, light scattering; BSA, bovine serum albumin; EPO, erythropoietin; CHO, Chinese hamster ovary; BDNF, brain-derived neurotrophic factor; PBS, phosphate-buffered saline.

0003-2697/01 $35.00

molecular mass. Since the change in solution refractive index as a function of solute (i.e., protein and/or polymer) concentration (i.e., dn/dc) is a function of the refractive index of the solvent calibration curves must be made for each mobile phase of interest.

A similar approach to one above has been extensively applied to characterizing detergent–protein association states (5–7) and dextran–protein conjugates (8). The main technical difference is that the UV/RI detector system was calibrated with protein standards of known dn/dc to create a standard curve of $\epsilon_p(\mathrm{RI})(\mathrm{UV})^{-1}$ vs dn/dc (where ϵ_p is the polypeptide extinction coefficient at 280 nm and (RI) and (UV) are the areas under the RI and UV peaks, respectively). This curve is used to determine the dn/dc for the polypeptide component, (dn/dc_p). The derived (dn/dc_p) is then used to determine the fraction of conjugate bound (9).

The methods described above using UV and RI detectors will not yield information on the association state of the protein (i.e., self-association or other protein–protein types of association). However, by utilizing a light scattering detector in conjugation with a RI and UV detector, there exists a method that yields the polypeptide molecular mass across a SEC-HPLC peak (for a review see 9, 10). This method eliminates the contribution of the conjugating polymer to the analysis and must be used in conjunction with the methods of Kunitani *et al.* or Hayashi *et al.* to obtain information on the degree of conjugation. In this technique, standard proteins are usually used as mass standards to generate a calibration curve of (LS) (UV) $(\mathrm{RI})^{-2}$ ϵ_p^{-1} vs molecular mass which is then used to derive the association state of the conjugate either with itself or with protein receptors/ligands ((LS) is the area under the LS peak). Through a self-consistent iterative process the degree of conjugation can also be determined.

While the methods above may be combined to give both the degree of conjugation and the association state of the conjugate, it is a fairly tedious process to go through the various calibrations and iterations. This paper will demonstrate some novel improvements over the current methods for determining degree of conjugation and association state that are facile and rapid, do not require the use of calibration curves, and are unaffected by conditions where protein or polymer–SEC column adsorptive losses may occur.

MATERIALS AND METHODS

Preparation of Proteins

Ribonuclease A (RNase A), protease-free highly purified bovine pancreas, and bovine serum albumin (BSA) were purchased from Calbiochem (La Jolla, CA) and used without further purification. Recombinant *Escherichia-coli* derived erythropoietin (EPO), Chinese hamster ovary cell-derived erythropoietin (CHO EPO), brain-derived neurotrophic factor (BDNF), and PEGylated BDNF (20-kDa PEG) were obtained from Amgen, Inc. and used without further purification.

PEGylation

RNase A (2 mg/ml) in 50 mM BisTricene, pH 8, was mixed with the *N*-hydroxysuccinimide ester of 5-kDa methoxy-PEG (Shearwater Polymers, Inc., Huntsville, AL) at either a 0.5:1 or a 5:1 molar ratio of PEG:RNase A. The reaction was allowed to proceed overnight at 25°C. The resulting mixture was directly injected into the SEC-HPLC LS UV RI system.

Size-Exclusion Chromatography and Detector Setup

A Hewlett Packard Model 1050 HPLC system (Bellevue, WA) with an HP diode array detector (UV) was connected to a Wyatt Optilab differential refractometer (RI) (Santa Barbara, CA) and a Wyatt DAWN laser light scattering (LS) instrument with an argon–ion laser (Santa Barbara, CA). The detectors were connected in the following order: LS–UV–RI. Both the RI and the LS instruments operated at a wavelength of 488 nm. For BSA, a Tosohaas (Montgomeryville, PA) G3000 SWXL or a G2000 SWXL SEC column was used, and for EPO, CHO EPO, unconjugated RNase A, and unconjugated BDNF a Tosohaas G2000 SWXL SEC column was used with a mobile phase of 25 mM sodium phosphate, 125 mM NaCl, pH 7.0 (PBS), delivered at a flow rate of 0.8 ml/min. A mobile phase of 50 mM sodium phosphate, 250 mM NaCl, 10% ethanol, pH 7.0 (PBS–EtOH), was delivered at a flow rate of 0.5 to 0.8 ml/min over a Tosohaas G3000SWXL SEC column for PEGylated RNase A experiments and PEGylated BDNF experiments. Proteins were dissolved in PBS to a concentration of ca. 1 mg/ml, and injection volumes were between 10 and 50 μl.

Detector Calibration

The light scattering detector was calibrated either with toluene according to the manufacturers' instructions or with BSA as described below. The UV detector output and the RI detector output were connected to the light scattering instrument such that the signals from the three detectors could be simultaneously collected with the Wyatt ASTRA software. The UV detector manufacturer's specifications of 0.200 AU/volt output and a pathlength of 0.600 cm gave the UV calibration constant ($UVCal_{AU}$ = 0.333 AU/volt/cm) for converting the data to absorbance units at a 1-cm pathlength (i.e., for facile incorporation of extinction coefficients which are typically given for a 1-cm path). For simplicity in the equations presented here, all pathlengths will be assumed to be 1 cm. The RI detector was initially calibrated with stock potassium chloride

solutions made with potassium chloride carefully dried in a vacuum oven, weighed into tared volumetric flasks, and filled to the volume mark with Milli-Q water.

dn/dc of BSA Determined by Flow Injection Analysis

The dn/dc of BSA in PBS at 488 nm (the wavelength of our laser and differential refractometer) was determined using stock solutions of BSA brought to dialysis equilibrium with PBS. These were directly injected with sufficient volume to generate a plateau value across the UV and RI detectors. The UV absorbance at 280 nm was also read in a calibrated benchtop diffraction grating-type UV-Vis spectrophotometer, which provided a convenient check of the HPLV UV detector calibration constant. In the case of our HP1050 DAD UV detector, the calibration constant needed to be adjusted to 0.3478 AU/volt/cm based on the stock solution absorbance value of BSA. The difference between this value and the manufacturer's value may be due to the inherent wavelength resolution inaccuracies of diode array detectors. When the calibration constant of another manufacturer's dual wavelength diffraction grating type HPLC detector was checked with BSA, the manufacturer's value and the value based on the stock solution agreed to three significant figures (not shown). The slope of the plot of RI signal against concentration gave a $(dn/dc)_p = 0.187$ ml/g (data not shown), using an extinction coefficient of 0.670 ml/mg/cm (10).

Detector Calibration with BSA

Convenient calibration checks can be made using a protein of known molecular mass, $(dn/dc)_p$, and extinction coefficient. A method found to be equivalent to the above method was to calibrate the RI detector and LS detector chromatographically with BSA. BSA in a mobile phase of PBS (0.5 ml/min) was chromatographed over a Tosohaas G3000SWXL column with the UV, LS, and RI signals collected by the Wyatt software. The monomer peak area was selected and integrated. To get the true mass under the peak through the Wyatt software (which requires the UV signal to be converted into RI units), the UV signal is multiplied by a UV calibration constant specific to the protein through the relationship

$$UVCal_{\mathrm{BSA}} = UVCAL_{\mathrm{AU}} \frac{(dn/dc)_{\mathrm{BSA}}}{1000 \text{ mg/g } (\epsilon)_{\mathrm{BSA}}}. \quad [1]$$

The same area was integrated using the RI signal with an arbitrary RI calibration constant. The true RI calibration constant is then given by

$$RICal_{\mathrm{true}} = RICal_{\mathrm{arb}} \frac{\text{mass by UV area}}{(\text{mass by RI area})_{\mathrm{arb}}}. \quad [2]$$

Since the whole monomer peak area is used in the determination of the detector calibration constants, band broadening artifacts are insignificant and the calibration constant should be equivalent to those determined by traditional means.

It is standard practice in our lab to run a BSA calibration standard for each run. To do this, the $(dn/dc)_p$ of BSA must be known for the mobile phase used (a discussion on why this is necessary is reserved for the theoretical section below). The absolute calibration constants are first determined in a PBS mobile phase delivered at 0.5 ml/min over a Tosohaas G3000SWXL as described above. Now for a different mobile phase, BSA is again run using the same G3000SWXL column. The whole monomer peak is selected, and using $UVCal_{\mathrm{UV}}$ and $RICal_{\mathrm{true}}$, the $(dn/dc)_p$ of BSA can be determined using Eq. [7] (below). This method for adjusting the $(dn/dc)_p$ value for a different mobile phase assumes that the extinction coefficient of BSA is the same in the two mobile phases; it therefore may not be appropriate to apply it to mobile phases with large amounts of organic solvents or in solvents which cause BSA to be substantially unfolded.

The LS calibration constant can also be determined now using either the UV or the RI as the concentration detector since the true molecular mass of BSA is known:

$$LSCal_{\mathrm{true}} = LSCal_{\mathrm{arb}} \frac{\text{true molecular mass of BSA}}{\text{arbitrary molecular mass of BSA}}. \quad [3]$$

A convenient check of the calibration and associated $(dn/dc)_p$ value is comparison of the molecular mass determined using the RI vs the UV detector as the concentration detector for the Wyatt software. In our experience, after this calibration procedure the two masses should differ by $<0.05\%$.

Calibration constants may have to be determined by selecting limited regions of the BSA peak. When samples are not homogeneous and the peak of interest is not well resolved from other peak(s), selecting the top one-half to one-third of the peak will usually give sufficient purity for accurate results. It is important to note that for the $LSCal$ value to be accurate for protein analysis, the user must be consistent with the relative peak regions selected for analysis and the relative peak regions selected for calibration. Due to differences in the geometry of the UV or RI detectors (which "see" a plug of solution the length of the detector cell) and the LS detector (which "sees" a pinpoint of solution) and

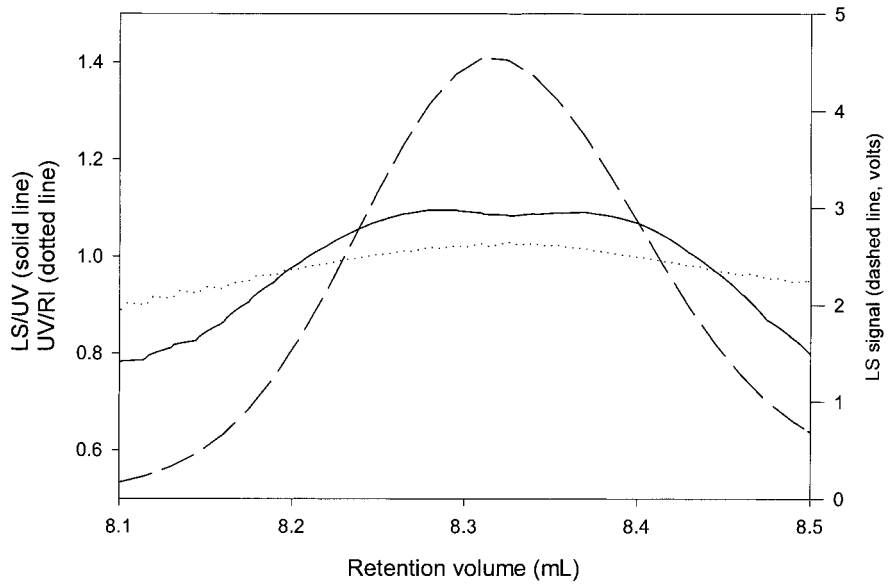

FIG. 1. Illustration of band-broadening effect between LS, UV, and RI detectors for BSA chromatographed with a Tosohaas G3000 SWXL column and a PBS mobile phase. SEC LS chromatogram of BSA (dashed line). A plot of the ratio of LS/UV signals (solid line) and the ratio of UV/RI signal (dotted line) vs retention time for BSA. With no band broadening, the ratio plots would be flat with a slope of 1.0.

band-broadening artifacts, even with the most homogeneous molecular mass sample the LS, RI, and UV traces will not overlay when normalized to a common area. This is illustrated in Fig. 1 with a ratio plot of LS/UV signals and a ratio plot of UV/RI signals (peak areas were normalized to a common peak area for comparison purposes). If there were no band-broadening effects, these ratio plots would result in a flat line with a ratio value of 1.0. Significant band-broadening effects are seen between the LS detector and the UV detector, and lesser effects are seen between the UV and RI detectors. Thus, when the selected region is narrow around the peak center the light scattering signal is artificially high; conversely, at the leading and trailing edges of the peak the light scattering signal is artificially low. A similar pattern occurs for the UV signal relative to the RI signal. If the entire peak can be selected (i.e., if the protein can be completely resolved from other impurities) the band-broadening effects are no longer significant and the true *LSCal* (from toluene) can be effectively used (data not shown). The band-broadening artifacts are dependent on the region selected within the eluting peak, the column type, and mobile phase, but independent of protein (data not shown). It is for these reasons that a protein standard is useful in determining effective calibration constants for a given chromatographic setup.

The calibration procedure using a specific region of the monomer peak is similar to that used for the full monomer peak as described above. First the $(dn/dc)_p$ of BSA is determined in the mobile phase of interest using the full monomer peak. Then the top one-third of the region is selected for analysis and the RI (through Eq. [2]) and LS (through Eq. [3]) detectors can be calibrated relative to the UV detector. Calibration using this approach corrects for the band-broadening artifacts between all the detectors and can capture peak areas sufficiently pure for analysis.

THEORETICAL

Relationship of dn/dc Values and Molecular Mass for Conjugated Protein–Polymers

The refractive index of a substance is a function of its chemical groups and independent of higher order structures. Thus, the dn/dc for a protein conjugate species $(dn/dc)_{cp}$ is (2, 9)

$$\left(\frac{dn}{dc}\right)_{cp} = \frac{M_p}{M_{cp}}\left(\frac{dn}{dc}\right)_p + \frac{M_c}{M_{cp}}\left(\frac{dn}{dc}\right)_c, \qquad [4]$$

where M is molecular mass (Da), n is the refractive index, c is concentration in g/ml, and the subscripts p, c, and cp refer to the protein, conjugating species, and conjugated protein, respectively. Using the conservation of mass equation, $M_c = M_{cp} - M_p$, and substituting into Eq. [4] gives upon rearrangement.

$$M_{cp} = M_p \frac{\left(\frac{dn}{dc}\right)_p - \left(\frac{dn}{dc}\right)_c}{\left(\frac{dn}{dc}\right)_{cp} - \left(\frac{dn}{dc}\right)_c}. \qquad [5]$$

The values for $(dn/dc)_p$ and $(dn/dc)_c$ to be used in this equation can be determined from independent experiments on the protein before conjugation and on the pure conjugating species. The change in dn/dc of the protein or conjugating polymer due to the conjugation chemistry (e.g., dehydration) will normally be insignificant due to the minor contribution a few atoms in the functional endgroups will have on the overall dn/dc of the large polymers. Given the importance of the dn/dc value in this type of analysis, it is useful to review what parameters may affect dn/dc. First, $(dn/dc)_p$ values are a function of amino acid composition (11). Second, dn/dc values are dependent on wavelength and must be known at the wavelength of the light scattering instrument for the proper interpretation of light scattering data (12). Protein dn/dc values are also affected by interactions between the solvent and the protein (13). Finally, dn/dc values are a strong function of the refractive index of the reference. The approximate relationship between dn/dc and solvent refractive index is given by (13)

$$\left(\frac{dn}{dc}\right)_p = \left(\frac{dn}{dc}\right)_{p,ref} - \bar{\nu}_p(n - n_{ref}), \qquad [6]$$

where ref refers to a reference solution in which $(dn/dc)_p$ is known and $\bar{\nu}_p$ is the partial specific volume of the protein (usually ca. 0.73 ml/g). Thus, a 1% change in solution refractive index can lead to a 5% change in dn/dc. These considerations illustrate the importance of independently determining dn/dc values for a given reference solvent and HPLC-SEC light scattering system. A convenient method is using the online LS UV RI system to chromatograph the unconjugated protein using the same mobile phase as will be used for the conjugated protein analysis.

dn/dc of Unconjugated Proteins

Inspection of Eq. [5] shows that $(dn/dc)_p$ must be determined to calculate the mass of conjugated proteins. $(dn/dc)_p$ is also needed to interpret light scattering for unconjugated proteins, and the following description applies to those cases as well. Using the calibration constants $UVCal_{AU}$ to convert the UV signal to AU, and $RICal$ to convert the RI signal to Δn, $(dn/dc)_p$ can be determined for a given peak area selection from

$$\left(\frac{dn}{dc}\right)_p = \epsilon_p\left(\frac{dn}{d(A)}\right)_p = \epsilon_p \frac{\text{mass by RI area}}{\text{mass by UV area}}. \qquad [7]$$

Equation [7] is valid even though the measured mass by RI area and UV area are dependent on (dn/dc), since (dn/dc) cancels out in the ratio of the two.

If ϵ_p is unknown, but the true molecular mass of the protein is known, the data from the LS UV RI system allow a rapid way to determine both ϵ_p and $(dn/dc)_p$ as follows. The simplified classical light scattering equation has the form $(Kc/R \approx 1/M)$, where K is a constant proportional to $(dn/dc)^2$, and R is the excess intensity of light scattered over that of the solvent (14). From the RI data, c is determined by $c = (n - n_0)/(dn/dc)_p$, where n is the RI signal for a given position in the eluting peak and n_0 is the RI signal due to solvent alone. Substituting these relationships into the light scattering equation results in the observation that the product $M \cdot (dn/dc)$ is a constant dependent on a given protein when M is calculated by the analysis software and the RI signal is used as the concentration detector. Thus any dn/dc value, $(dn/dc)_{arbitrary}$, can be input into the light scattering calculations to give an arbitrary M_p for a selected chromatographic peak. The product of these two values results in the α value in

$$M_{obs,ls} * (dn/dc)_{obs,ls} = M_{true} * (dn/dc)_{true} = \alpha_p. \qquad [8]$$

This equation can be applied to determine the $(dn/dc)_p$ when the extinction coefficient is unknown, but the true molecular mass (M_{true}) is known. $M_{true} = NM_p$, where N is the stoichiometry of self-association and M_p is determined from the sequence molecular mass. Using Eq. [8] to solve for $(dn/dc)_p$ requires having a homogeneous species eluting in the peak selected; if impurities or various levels of rapidly reversible self-association occur in the peak region, erroneous values will result. If self-associated species (oligomers) exist and can be adequately separated into pure peaks (as is the case with BSA), M_{true} for a given peak can be determined as follows. Substitute NM_p for M_{true} in Eq. [8], and solve for N

$$N = \frac{\alpha_p}{M_p(dn/dc)_p}. \qquad [9]$$

As proteins typically have $(dn/dc)_p$ values between 0.18 and 0.20 ml/g, substitution of 0.19 for $(dn/dc)_p$ will yield an N value that can be rounded to the nearest whole number to yield the degree of association and the resulting M_{true} value. Then Eq. [9] can be turned around and solved for $(dn/dc)_p$ or ϵ_p as follows:

$$(dn/dc)_p = \frac{\alpha_p}{M_{true}} = \epsilon_p \frac{dn}{dA} \qquad [10]$$

or

$$\epsilon_p = \frac{\alpha_p}{M_{true}(dn/dA)}. \qquad [11]$$

The dn/dc of the conjugating species $(dn/dc)_c$ can be either determined experimentally using stock solutions of known polymer concentration or obtained from the literature. As with proteins, literature values should be used only if they were determined using a similar solvent system and wavelength (12, 13).

Determination of the Association State of the Conjugated Protein

To determine the degree of conjugation and the mass of the conjugated protein, the association state (degree of oligomerization) of the protein must be known. A so-called "three detector method" in which the polypeptide mass of conjugated proteins is calculated through algebraic cancellation of all contributions of the conjugate has been developed by the groups of T. Takagi and T. Arakawa and is now used by many laboratories (9, 10). However, since this method is not directly implemented in the instrument software, in general its use has involved calibration using protein mass standards and exporting data to other software or manually reading values from graphs. Hence often the mass is calculated at only a single elution position (peak height analysis) rather than averaging results over a region of the chromatogram. We have now developed an easier, alternative implementation of this approach that uses the standard "absolute" detector calibrations and values averaged over regions of the chromatogram by the standard instrument software, as outlined below.

The basis for this new algorithm is that a relationship similar to Eq. [8] can be written for conjugated proteins as

$$M_{cp,obs} * (dn/dc)_{cp,obs} = M_{cp,true} * (dn/dc)_{cp,true} = \alpha_{cp}. \quad [12]$$

Then through application of Beer's law

$$M_{cp,true} * (dn/dc)_{cp,true} = M_{cp,true} * \epsilon_{cp}(dn/dA)_{cp,true}. \quad [13]$$

Incorporating the following relationship

$$\epsilon_{cp} = \frac{M_{p,true}}{M_{cp,true}} * \epsilon_p \quad [14]$$

yields the true polypeptide mass by LS, RI, and UV data:

$$M_{p,ls} = \frac{\alpha_{cp}}{(dn/dA)_{cp} * \epsilon_p}, \quad [15]$$

where $(dn/dA)_{cp}$ is calculated in a manner analogous to that for the unconjugated protein (again using $UVCal_{AU}$ in the software), through the relationship

$$\left(\frac{dn}{dA}\right)_{cp} = \frac{\text{mass by RI area}}{\text{mass by UV area}}. \quad [16]$$

It can be shown that Eq. [15] is mathematically equivalent to the standard three detector method. The association state will be given by dividing the resulting $M_{p,ls}$ by the sequence molecular mass. If ϵ_p is affected by conformational effects resulting from conjugation, this effect should be minor and the association state should be obvious (i.e., round the ratio to the nearest whole number). Once the association state is known, a more accurate value of ϵ_p can be determined by rearranging Eq. [15] using the true peptide molecular mass of the associated species.

$$\epsilon_p = \frac{\alpha_{cp}}{(dn/dA)_{cp} * M_{p,true}}. \quad [17]$$

Equation [17] should be applied only to well-resolved peaks containing a pure species; having multiple molecular species present in the region analyzed will lead to an erroneous ϵ_p.

Determining the Mass of a Conjugated Protein–Polymer

Utilizing Beer's law and assuming that the absorbance detector is used at a wavelength where the conjugating species has a negligible absorbance, the equation for $(dn/dc)_{cp}$ can be written as

$$\left(\frac{dn}{dc}\right)_{cp} = \left(\frac{dn}{d\left(\frac{A}{\epsilon}\right)}\right)_{cp} = \epsilon_p \frac{M_p}{M_{cp}} \left(\frac{dn}{dA}\right)_{cp}. \quad [18]$$

Substituting this expression into Eq. [5] and solving for M_{cp} gives

$$M_{cp} = M_p \frac{\epsilon_p\left(\frac{dn}{dA}\right)_{cp} - \left(\frac{dn}{dc}\right)_p + \left(\frac{dn}{dc}\right)_c}{\left(\frac{dn}{dc}\right)_c}. \quad [19]$$

In this expression, M_p is the polypeptide molecular mass of the associated state based on sequence, and only the RI and UV detectors are being used to determine M_{cp}. Alternatively, Eq. [19] can be written as

$$M_{cp,ls} = \frac{\alpha_{cp} - M_{p,ls}\left(\frac{dn}{dc}\right)_p + M_{p,ls}\left(\frac{dn}{dc}\right)_c}{\left(\frac{dn}{dc}\right)_c}. \quad [20]$$

This formulation uses only LS and RI data and is not directly dependent on the protein extinction coefficient (although ϵ_p may have been used in determining $(dn/dc)_p$). In Eq. [20] it is probably best to use $M_{p,ls}$ as determined by Eq. [15] rather than M_{true} since it has the same dependence on the detector calibrations as α_{cp}.

It can be shown that Eq. [20] is mathematically equivalent to an iterative, self-consistent procedure for calculating M_{cp} using only LS and RI detectors that we and others have employed previously. In that procedure initially the dn/dc is set to $(dn/dc)_p$ and the apparent mass of the peak for the conjugate, $M_{cp,ls}$, is calculated using the standard software. For a conjugated protein this returned value of $M_{cp,ls}$ will of course exceed M_p, so the excess mass is assigned to M_c, and a new estimate of $(dn/dc)_{cp}$ is calculated using Eq. [4]. This new estimate of $(dn/dc)_{cp}$ is then used to produce a better estimate of $M_{cp,ls}$, and the whole process is iterated several times until $M_{cp,ls}$ and $(dn/dc)_{cp}$ converge to constant values. Equation [20] gives the equivalent result through a rapid direct algebraic approach rather than this lengthy iterative one.

It is also possible to utilize the data from all three detectors simultaneously and directly calculate the total mass. In essence, this approach uses the RI and UV detectors to calculate $(dn/dc)_{cp}$, in a manner analogous to Eq. [19] and then incorporates this value in a standard analysis of the LS data using the RI as the concentration detector. By rearranging Eqs. [5] and [18] we find

$$\left(\frac{dn}{dc}\right)_{cp} = \frac{\epsilon_p\left(\frac{dn}{dA}\right)_{cp}\left(\frac{dn}{dc}\right)_c}{\epsilon_p\left(\frac{dn}{dA}\right)_{cp} - \left(\frac{dn}{dc}\right)_p + \left(\frac{dn}{dc}\right)_c}, \quad [21]$$

where again the $(dn/dA)_{cp}$ values are determined as described previously using Eq. [16]. The $(dn/dc)_{cp}$ value is then combined with the LS data (most conveniently using the LS software provided by the instrument manufacturer) to calculate the molecular mass of the species in the peak of interest. Combining this information with the sequence molecular mass of the monomer/receptor/ligand, the degree of association and the degree of conjugation are directly obtained.

RESULTS

Glycosylated-monomeric EPO, various PEGylated fractions of monomeric RNase A, and a dimeric PEGylated BDNF protein were selected to illustrate the utility of the methods described above. The typical approach we use is to first determine $(dn/dc)_p$, $(dn/dc)_c$ using unconjugated species, followed by an SEC-LS/RI/UV chromatographic run of the conjugated molecule to determine α (Eq. [12]) and $(dn/dA)_{cp}$ (Eq. [16]). The molecular mass of the protein, $M_{p,ls}$, is then determined using Eq. [15] and compared to the sequence molecular mass to determine association state. Finally any of Eqs. [19]–[21] can be used to determine the overall molecular mass as described under Theoretical.

Glycosylated EPO

In order to determine the $(dn/dc)_p$ of the CHO EPO protein an online experimental determination was made for the nonglycosylated form of EPO expressed in *E. coli*. A single injection of 0.25 mg of *E. coli* EPO gave a pure baseline-resolved main peak (data not shown). A $(dn/dc)_p$ value of 0.191 ml/g was determined via Eq. [7]. To estimate the $(dn/dc)_c$ of the carbohydrate in CHO EPO, stock solutions of β-cyclodextrin (M = 1100 Da) at known concentrations (by weight) were directly injected into the differential refractometer until a plateau of constant Δn was reached. The slope from a plot of Δn vs concentration gave a $(dn/dc)_c$ of 0.145 ml/g for β-cyclodextrin in PBS. With $(dn/dc)_p$ and $(dn/dc)_c$ parameters now known, analysis of the glycosylated EPO was able to proceed. A single injection of 0.5 mg of CHO EPO was analyzed with LS UV RI detection (Fig. 2A). Applying Eqs. [12] and [16] resulted in α = 5440 and $(dn/dA)_{cp}$ = 0.230, respectively. The molecular mass of the peptide as determined via Eq. [15] is $M_{p,ls}$ = 18,200 Da, which agrees well with the sequence molecular mass of 18,236 Da and confirms that CHO EPO is monomeric. A CHO EPO peptide extinction coefficient ϵ_p of 1.24 ml/mg/cm was used in Eq. [15] based on dry weight analysis previously published (15), which agrees reasonably well with ϵ_p = 1.30 ml/mg/cm as determined via Eq. [17]. Application of Eq. [19] using ϵ_p = 1.30 ml/mg/cm derived here (which should be valid since a pure peak was obtained) and using M_p = 18,236 based on sequence results in a total molecular mass of 31,800 Da for the glycosylated protein, a 42.8% degree of glycosylation (Table 1). Application of Eq. [20] using $M_{p,ls}$ = 18,200 based on Eq. [15] results in a total molecular mass of 31,700 Da. From Eq. [21] we derive $(dn/dc)_{cp}$ = 0.171 and applying this value with the standard light scattering software (with RI as the concentration detector) gives a total molecular mass of 31,800 Da. The results of the various analyses are summarized in Table 1.

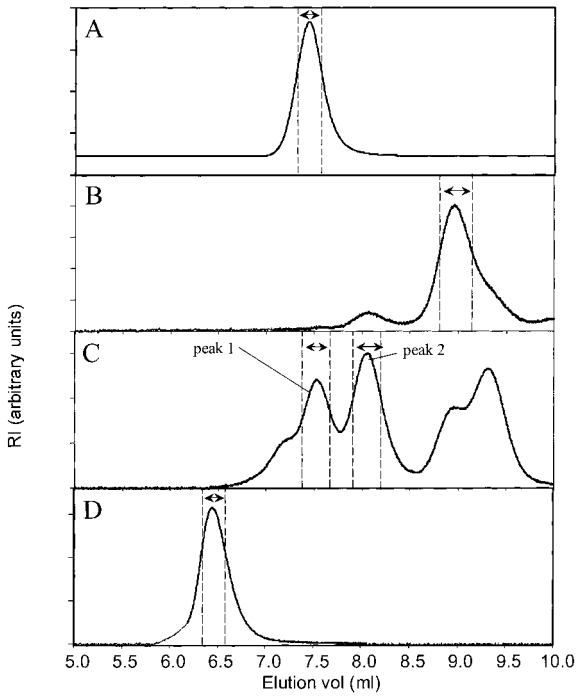

FIG. 2. SEC chromatograms of glycosylated and PEGylated proteins. (A) CHO EPO (monomer, molecular mass 31,800 Da, 42.8% glycosylated). (B) 0.5:1 PEG:RNase A reaction mixture (monomer, molecular mass 18,300, mono-PEGylated with 5000 Da PEG). (C) 5:1 PEG:RNase A reaction mixture (peak 1, monomer, molecular mass 28,600, tri-PEGylated with 5000-Da PEGs; peak 2, monomer, molecular mass 23,900, di-PEGylated with 5000-Da PEGs). (D) Purified PEG:BDNF (dimer, molecular mass 62,100, each subunit mono-PEGylated with 20,000-Da PEG). Dashed vertical lines represent regions selected for analysis.

Determination of the Degree of PEGylation of RNase A and Its Association State

Native RNase A was used to determine $(dn/dc)_p$ in PBS–EtOH. Analysis of a single injection of 0.05 mg via Eq. [7] gave $(dn/dc)_p = 0.197$ ml/g. Application of Eq. [11] gave $\epsilon_p = 0.715$, in good agreement with the published value of 0.712 (which was in pH 5 acetate buffer) (16). The $(dn/dc)_c$ of PEG was determined to be 0.133 ml/g by injection of stock solutions of known PEG concentrations in PBS–EtOH into the differential refractometer as described above for β-cyclodextrin.

PEGylated RNase A from a solution reacted with a ratio of 0.5:1 PEG:RNase A elutes as shown in Fig. 2B. The extended trailing edge of the peak was not observed in the UV trace and is presumably unreacted PEG. The protein was found to be monomeric with an $M_{p,ls}$ (Eq. [15]) of 12,800 Da (the sequence mass is 13,682 Da). Due to the impure nature of the peaks, $\epsilon_p = 0.715$ was used based on the unPEGylated RNase value rather than trying to calculate an experimental value via Eq. [17]. Applying Eq. [19] yields a total molecular mass of 18,300 Da for the PEGylated RNase A peak. Subtracting this value from M_{true} gives the PEG molecular mass of 4620 Da, consistent with a mono-PEGylated product. Equations [20] and [21] give slightly lower masses (Table 1), but are also quite consistent with a mono-PEGylated product.

RNase A was also PEGylated using a reaction ratio of 5:1 PEG:RNase A. The resulting SEC chromatogram for this reaction (Fig. 2C) indicates multiple PEGylated species. The large peak eluting at 9.3 ml is unreacted PEG based on an injection of PEG alone (data not shown). The results of the molecular mass analyses are given in Table 1 and are consistent with a monomeric species that is tri-PEGylated in peak 1 ($M_{cp} = 28{,}600$ Da) and di-PEGylated in peak 2 ($M_{cp} = 23{,}900$ Da).

Determination of the Degree of BDNF PEGylation and Its Association State

Native BDNF before conjugation with 20-kDa PEG gave a $(dn/dc)_p$ value equal to 0.204 ml/g in PBS–

TABLE 1

Molecular Mass Determination of Protein Component and Conjugated Protein for CHO EPO, PEG–RNase A, and PEG–BDNF

Protein	$M_{p,ls}$ Eq. [15]	M_p based on sequence	Association state	M_{cp} Eq. [19] ($M_c = M_{cp} - NM_p$)	M_{cp} Eq. [20] ($M_c = M_{cp} - M_{p,ls}$)	M_{cp} based $(dn/dc)_{cp}$ and LS data ($M_c = M_{cp} - M_{p,ls}$)	Degree of glycosylation/ PEGylation (ratio of data in column 2)
CHO EPO (Fig. 2A)	18,200	18,236	Monomer	31,800 (13,600)	31,700 (13,500)	31,800 (13,600)	42.8%
5-kDa PEG–RNase A (Fig. 2B peak)	12,800	13,682	Monomer	18,300 (4,620)	17,100 (4,300)	17,200 (4,400)	Mono-PEG
5-kDa PEG–RNase A (Fig. 2C Peak 1)	12,900	13,682	Monomer	28,600 (14,900)	27,000 (14,100)	27,000 (14,100)	Tri-PEG
5-kDa PEG–RNase A (Fig. 2C Peak 2)	12,400	13,682	Monomer	23,900 (10,200)	21,700 (9,300)	21,700 (9,300)	Di-PEG
20-kDa PEG–BDNF (Fig. 2D peak)	26,000	13,513	Dimer	62,100 (35,100)	59,000 (33,000)	59,800 (33,800)	Mono-PEG (on each monomer)

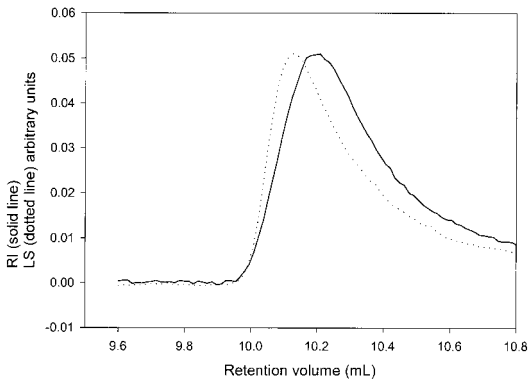

FIG. 3. SEC chromatogram of unPEGylated BDNF illustrating the distribution of dimer/monomer across the peak. This type of rapidly reversible equilibrium is an example of where Eq. [8] in the text cannot be applied, since a pure species of a single M_{true} is not adequately resolved.

EtOH via Eq. [7]. The sequence molecular mass is 13,513 Da, and the extinction coefficient is 1.76 ml/mg/cm (determined by amino acid analysis—data not shown). Equation [11] could not be applied to determine ϵ_p due to the presence of unresolved dimer and monomer species in the main peak (Fig. 3).

The results for the 20-kDa PEGylated BDNF are presented in Table 1 and Fig. 2D. Application of Eq. [15] results in $M_{p,ls}$ = 26,000, which is consistent with a dimeric species. Equation [19] gives M_{cp} = 62,100, which is consistent with a dimeric species with two PEG molecules attached.

DISCUSSION

The methods presented here allow a reliable means to determine the degree of conjugation as well as the association state of conjugated proteins. CHO EPO was chosen as a model protein in this study because it has been well characterized (17–19), allowing direct comparison of previous results to our method. A molecular mass of 31,800 Da by Eq. [19] (Table 1) is in good agreement with our previous report of 30,390 Da by sedimentation equilibrium (19). The single peak observed in the chromatogram is indicative of a pure protein sample.

Unpurified reaction mixtures composed of either 0.5:1 or 5:1 ratios of PEG to RNase A resulted in chromatograms containing overlapping peaks (Figs. 2B and 2C). The RI UV detection method of Kunitani *et al.* requires integration of fairly pure baseline-resolved peaks and would not be able to characterize these species without an additional purification step (2). Similarly, sedimentation equilibrium would require purified fractions in order to determine the degree of PEGylation. In contrast, the methods presented here can determine the degree of PEGylation and the association state directly from a single injection of the reaction mixture (Table 1).

One note about highly impure samples is that while the LS data can be used in conjunction with the α value to determine the association state of the protein, caution should be used in applying Eq. [17] to determine extinction coefficients. It was apparent that the peptide extinction coefficients based on Eq. [17] (not shown) were artificially low compared to the value found for the unPEGylated proteins. This may possibly be due to compounding of artifacts from impurities in the peak and band-broadening/detector geometry effects, which are much larger for the LS detector relative to the differences between the UV and RI detectors (Fig. 1). For these reasons, the sequence molecular mass (adjusted for association state) should probably be used in Eq. [19]. However, for light scattering analysis using Eq. [20] and for the combined $(dn/dc)_{cp}$ LS data approach, it is probably more appropriate to use $M_{p,ls}$ from Eq. [15], which keeps the molecular mass dependencies on the light scattering data self-consistent. By inspection of Table 1, the results closest to the expected values comes from Eq. [19], possibly due to the similar flow cell and detector geometries of the UV and RI detectors. Overall, then, our preferred strategy is to use Eq. [15] to confirm the state of association, but then use the appropriate integer multiple of the sequence mass in calculating the total mass via Eq. [19]. It should also be noted that a highly accurate value for ϵ_p is generally not necessary to obtain the correct association state (and thus a value calculated based on amino acid composition will usually suffice).

Native BDNF has been previously shown to exist predominantly as a dimer in solution (20, 21). After conjugation with a 20-kDa PEG, the analysis clearly indicates a dimeric species (Table 1), and the total mass data are consistent with addition of two PEG molecules per dimer, as expected. Presumably the PEG molecules are attached one per BDNF monomer, but it should be noted the light scattering analysis could not distinguish that situation from one with both PEG molecules on the same BDNF monomer.

Different mobile phase conditions (and thus different solvent refractive indices) are typically required to provide optimal peak resolution, good sample recovery, good reproducibility, and/or reasonable column life. It is important to remember that dn/dc values are a function of the refractive index of the solvent and interactions between the solvent and protein (12, 13). Thus, it is important to measure the dn/dc values for the unconjugated polymer and protein, and the dn/dA for the conjugated molecule, in the same solvent that is used for the chromatography. The RI/UV method (Eq. [19]) and the LS/RI/UV total mass method (through Eq. [21]) also depend upon having an accurate extinction coefficient, which can be elusive with impure samples

or when limited quantities of purified protein are available. Often, the extinction coefficient is calculated based on amino acid sequence, but this estimate can be unreliable, especially when the tryptophan content is very low or the number of Cys residues participating in disulfide bonds is unclear, and can be in error by 5 to 15% in some cases (22). Thus the alternative method based on Eq. [20], that is independent of extinction coefficient, may be preferred.

There may be circumstances where measuring the dn/dc values of the unconjugated molecule is not possible. Isolating an intact deglycosylated protein may be difficult due to stability or solubility problems during or after the deglycosylation procedure. In this case, $(dn/dc)_p$ can be estimated by measurements on a protein standard in the reference solvent, realizing that the best results will be obtained if the standard has a similar amino acid composition (11). The exact extent of protein-to-protein variation in dn/dc is difficult to assess because of the difficulties in measuring concentration accurately, but it is probably fairly small (a few percent). Indeed, within the analytical ultracentrifugation community it has been common practice for 50 years to assume that dn/dc is the same for all proteins. Thus in many cases (and particularly where the accuracy of an extinction coefficient is questionable), it may be preferable to use an estimated $(dn/dc)_p$ together with Eq. [20] for calculating total mass. Based on Eqs. [6]–[8], a 5% error in $(dn/dc)_p$ translates to approximately a 5% error in M_{cp}.

In the case of carbohydrate dn/dc values, it is convenient to use model carbohydrates such as dextran or β-cyclodextrin (10), due to the difficulties in purifying the glycation moieties from the glycosylated protein. Fortunately, errors in $(dn/dc)_c$ have a less dramatic effect on M_{cp} than errors in $(dn/dc)_p$. For example, a 5% error in $(dn/dc)_c$ only results in a 1 to 2% error in M_{cp}. Furthermore, Kunitani *et al.* demonstrated the similarity of $(dn/dc)_p$ values for various glycation moieties (4).

For determining the degree of PEGylation, the importance in having precise dn/dc values is driven by the PEG conjugate molecular mass. For example, with a 5-kDa PEG conjugation on a 15-kDa protein, the correct degree of conjugation can be obtained with up to a 12% error in $(dn/dc)_p$ or a 34% error in $(dn/dc)_c$. For a 20-kDa PEGylation, a 75% error in $(dn/dc)_p$ or a 84% error in $(dn/dc)_c$ will still yield the correct degree of PEGylation.

Although complete resolution of individual peaks is not required for analysis, the presence of shoulders or overlapping peaks will influence the derived M_{cp}. This situation was encountered with the unpurified RNase A–5-kDa PEG reaction mixture (Figs. 2B and 2C), for which various PEGylated species were incompletely resolved in the chromatograms. Selecting a narrow region within each peak for analysis minimized the contribution of the adjacent peaks, allowing a more accurate determination of the degree of PEGylation (Fig. 2 and Table 1). Like all online SEC-HPLC light scattering techniques, the most accurate results will be obtained with broad peaks and baseline resolution.

Finally, since Eq. [21] can be used to directly calculate total mass without knowing sequence masses or the state of association, should this not be the method of choice? Not necessarily. This method depends explicitly on extinction coefficients, and one drawback of requiring data from all three detectors is greater opportunity for errors (calibration, baseline drift, band spreading, etc.). On the other hand, in the RI/UV approach (Eq. [19]) or the LS/RI approach (Eq. [20]), when the polypeptide mass is based on sequence mass this can in essence recalibrate the data and serve to eliminate or minimize some sources of error (provided, of course, that the sample truly is purely this sequence mass). Thus it is not straightforward to say which of these approaches is "best" or most accurate. It may often be advisable to try more than one of these approaches and compare the results (which also can provide an estimate of the uncertainty), and hopefully the new approaches outlined here will make it quicker and easier to apply more than one calculation method.

REFERENCES

1. Fairman, R., Fenderson, W., Hail, M. E., Wu, Y., and Shaw, S. Y. (1999) Molecular weights of CTLA-4 and CD80 by sedimentation equilibrium ultracentrifugation. *Anal. Biochem.* **270,** 286–295.
2. Kunitani, M., Dollinger, G., Johnson, D., and Kresin, L. (1991) On-line characterization of polyethylene glycol-modified proteins. *J. Chromatogr.* **588,** 125–137.
3. Shire, S. J. (1994) *in* Modern Analytical Ultracentrifugation (Schuster, T. M., and Laue, T. M., Eds.), pp. 261–297, Birkhauser, Boston.
4. Kunitani-M, K.-L. (1993) High-performance liquid chromatographic analysis of carbohydrate mass composition in glycoproteins. *J. Chromatogr.* **632,** 19–28.
5. Takagi, T., Maezawa, S., and Hayashi, Y. (1987) Determination of subunit molecular weights of canine renal (Na+, K+)-ATPase by low-angle laser light scattering coupled with high performance gel chromatography in the presence of sodium dodecyl sulfate. *J. Biochem. (Tokyo)* **101,** 805–811.
6. Maezawa, S., Hayashi, Y., Nakae, T., Ishii, J., Kameyama, K., and Takagi, T. (1983) Determination of molecular weight of membrane proteins by the use of low-angle laser light scattering combined with high-performance gel chromatography in the presence of a non-ionic surfactant. *Biochim. Biophys. Acta* **747,** 291–297.
7. Watanabe, Y., Kijima, Y., Kadoma, M., Tada, M., and Takagi, T. (1991) Molecular weight determination of phospholamban oligomer in the presence of sodium dodecyl sulfate: Application of low-angle laser light scattering photometry. *J. Biochem. (Tokyo)* **110,** 40–45.
8. Kato, A., Kameyama, K., and Takagi, T. (1992) Molecular weight determination and compositional analysis of dextran–protein conjugates using low-angle laser light scattering technique com-

bined with high-performance gel chromatography. *Biochim. Biophys. Acta* **1159,** 22–28.

9. Hayashi, Y., Matsui, H., and Takagi, T. (1989) Membrane protein molecular weight determined by low-angle laser light-scattering photometry coupled with high-performance gel chromatography. *Methods Enzymol.* **172,** 514–528.
10. Wen, J., Arakawa, T., and Philo, J. (1996) Size-exclusion chromatography with on-line light scattering, absorbance, and refractive index detectors for studying proteins and their interactions. *Anal. Biochem.* **240,** 155–166.
11. McMeekin, T., Wilensky, M., and Groves, M. (1962) Refractive indices of proteins in relation to amino acid composition and specific volume. *Biochem. Biophys. Res. Commun.* **7,** 151–156.
12. Eisenberg, H., Josephs, R., and Reisler, E. (1977) Scattering correction to the absorbance, wavelength dependence of the refractive index increment, and molecular weight of the bovine liver glutamate dehydrogenase oligomer and subunits. *Biopolymers* **16,** 2773–2783.
13. Eisenberg, H. (1992) Halophilic malate dehydrogenase—A case history of biophysical investigations: Ultracentrifugation, light-, X-ray- and neutron scattering. *Biochem. Soc. Symp.* **58,** 113–125.
14. Wyatt, P. J. (1993) Light scattering and the absolute characterization of macromolecules. *Anal. Chim. Acta* **272,** 1–40.
15. Davis, J. M., Arakawa, T., Strickland, T. W., and Yphantis, D. A. (1987) Characterization of recombinant human erythropoietin produced in Chinese hamster ovary cells. *Biochemistry* **26,** 2633–2638.
16. Xie, G., and Timasheff, S. N. (1997) Mechanism of the stabilization of ribonuclease A by sorbitol: Preferential hydration is greater for the denatured then for the native protein. *Protein Sci.* **6,** 211–221.
17. Rush, R. S., Derby, P. L., Smith, D. M., Merry, C., Rogers, G., Rohde, M. F., and Katta, V. (1995) Microheterogeneity of erythropoietin carbohydrate structure. *Anal. Chem.* **67,** 1442–1452.
18. Hokke, C. H., Bergwerff, A. A., Van Dedem, G. W., Kamerling, J. P., and Vliegenthart, J. F. (1995) Structural analysis of the sialylated N- and O-linked carbohydrate chains of recombinant human erythropoietin expressed in Chinese hamster ovary cells: Sialylation patterns and branch location of dimeric *N*-acetyllactosamine units. *Eur. J. Biochem.* **228,** 981–1008.
19. Philo, J. S., Aoki, K. H., Arakawa, T., Narhi, L. O., and Wen, J. (1996) Dimerization of the extracellular domain of the erythropoietin (EPO) receptor by EPO: one high-affinity and one low-affinity interaction. *Biochemistry* **35,** 1681–1691.
20. Philo, J. S., Rosenfeld, R., Arakawa, T., Wen, J., and Narhi, L. O. (1993) Refolding of brain-derived neurotrophic factor from guanidine hydrochloride: Kinetic trapping in a collapsed form which is incompetent for dimerization. *Biochemistry* **32,** 10812–10818.
21. Radziejewski, C., Robinson, R. C., DiStefano, P. S., and Taylor, J. W. (1992) Dimeric structure and conformational stability of brain-derived neurotrophic factor and neurotrophin-3. *Biochemistry* **31,** 4431–4436.
22. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **4,** 2411–2423.