Analytical Biochemistry **270,** 286–295 (1999)
Article ID abio.1999.4095, available online at http://www.idealibrary.com on IDEAL®

Molecular Weights of CTLA-4 and CD80 by Sedimentation Equilibrium Ultracentrifugation

Robert Fairman,*[,1] William Fenderson,† Mark E. Hail,‡ Youling Wu,‡ and Shyh-Yu Shaw†[,2]

**Department of Biology, Haverford College, 370 Lancaster Avenue, Haverford, Pennsylvania 19041; and †Division of Macromolecular Structure and ‡Analytical Research & Development, Bristol-Myers Squibb Pharmaceutical Research Institute, P.O. Box 4000, Princeton, New Jersey 08543-4000*

Received December 29, 1998

Proteins that are heavily glycosylated pose unique challenges in their biophysical characterization. In particular, molecular weight analysis is exacerbated by such glycosylation. For example, glyoproteins are refractory to careful mass spectrum analysis and often give anomalous retention times using size exclusion chromatography. We combine several approaches to characterize the molecular weights of the extracellular domains of the glycoproteins CTLA-4 and CD80 using carbohydrate analysis, electrospray mass spectrometry, size exclusion chromatography, and analytical ultracentrifugation. In addition, we have applied a method described previously, using sedimentation equilibrium analysis to calculate the contribution of carbohydrates to the molecular masses of CTLA-4 and CD80. It is important to understand the oligomeric states of these protein domains because the interaction between these lymphocyte receptors plays an important costimulatory role in the T_h-cell antigenic response. It is thought that extracellular interactions between these receptors may regulate both the self-association of these receptor proteins and the oligomeric state of the heterocomplex; this regulation has important consequences for potentiating the signaling mechanism between T_h-cells and antigen-presenting cells.

CTLA-4 and CD80 (also named B7-1) are T-cell and B-cell receptors, respectively (2–4), that induce a costimulatory response (5, 6) for the activation of immunocompetent cells in the immune response. Such interactions act to modulate the primary interactions that occur between T-cells and B-cells via presentation of antigen by the MHC class II receptor to the T-cell receptor complex. Interactions between CTLA-4 (or its homolog, CD28) and CD80 (or its homolog, CD86) can stimulate T-cell proliferation as well as negatively regulate T-cell activation, depending on which particular set of receptors are engaged with one another (7–9).

CTLA-4 exists as a disulfide-linked dimer on the cell surface; this covalent dimerization is important for high binding avidity. Evidence suggests that each monomeric unit of CTLA-4 can bind its coreceptors, CD80 or CD86, independently (10). An understanding of the structural determinants for this binding has come from a recent NMR structure of a monomeric, soluble fragment of CTLA-4 (11). Its structure is related to the immunoglobulin superfamily variable domains. Two N-linked glycosylation sites at Asn 78 and Asn 111 were observed showing varying degrees of interactions with the protein; however, neither appear to be important for the structure or stability of this domain from CTLA-4 (Fairman, unpublished results). No detailed structural information is yet available for CD28, CD80, or CD86.

Of critical importance in understanding the interaction of these receptors is their oligomeric states, namely their ability to self-associate because this directly impacts the magnitude of the binding avidity. There is some evidence that extracellular interactions with CTLA-4 drive oligomerization of CD80, thus creating the potential to regulate intracellular signal transduction pathways (10). It is important to understand the oligomeric states of these protein–protein interaction domains to fully understand their roles in signal transduction and subsequent regulation of the immune response.

[1] To whom correspondence and reprint requests should be addressed. Fax: 610-896-4963. E-mail: rfairman@haverford.edu.

[2] Present address: Department of Chemistry, National Cheng-Kung University, One University Road, Tainan, Taiwan, R.O.C.

0003-2697/99 $30.00

As a prerequisite to determining the nature of the heteromeric complexes formed between CTLA-4 and CD80, it is first necessary to establish the native oligomeric states of these proteins by themselves. The approach we have taken is to combine several analytical methods with ultracentrifugation analysis to determine the native molecular weights of the glycosylated extracellular domains of CTLA-4 and CD80 (12).

Glycoproteins provide a significant challenge in determining accurate molecular weights by sedimentation equilibrium because the partial specific volumes of such proteins are not readily determined and this is an essential parameter in sedimentation equilibrium analysis. If an accurate mass for the linked carbohydrate is not known (and this is often the case for glycoproteins), it is impossible to determine the correct partial specific volume to be used. This problem is compounded because the average partial specific volume for polypeptide chains (0.73 mL g^{-1}) is significantly different from that for carbohydrates (0.63 mL g^{-1}). One simple method for experimentally determining the partial specific volume of a protein is to compare the sedimentation in solvents of different densities such as H_2O and D_2O (13); however, this method cannot be applied to heterogeneous systems. Classical methods to measure this parameter, such as pycnometry, suffer from the requirement for large amounts of protein. Instead, we use a methodology first described by Shire (1) to empirically determine accurate molecular weights without knowing the partial specific volume *a priori*. Here we show a test of this methodology using CTLA-4, where we have detailed information about carbohydrate identity and content and then apply this to CD80 for which only limited information is available about its associated carbohydrates.

MATERIALS AND METHODS

Preparation of Protein

Protein was expressed and purified as described in Metzler *et al.* (11). The concentrations of the proteins were determined by UV absorption in 6 M GuCl by the method of Edelhoch (14).

Size Exclusion Chromatography

Both CTLA-4 and CD80 were analyzed by size exclusion chromatography using a Bio-Sil SEC250-5 column (Bio-Rad) running on a Pharmacia FPLC system. The column was equilibrated with phosphate-buffered saline and run using a flow rate of 0.5 mL/min in the same buffer. Apparent molecular weights were determined using a standard curve derived from size exclusion chromatography standards from Bio-Rad.

Liquid Chromatography/Electrospray Ionization Mass Spectrometry (LC/ESI/MS)

All CTLA-4 samples were subjected to on-line micro-HPLC to remove nonvolatile salts and to separate CTLA-4 from other proteins present in the samples. The micro-HPLC and columns were obtained from Michrom BioResources (Auburn, CA). Samples containing approximately 100 pmol of CTLA-4 were injected onto a protein trap column (1 × 10 mm PLRP-s 4000 Å). The trap column was flushed using a syringe containing 100 μL of H_2O/acetonitrile/acetic acid (78/20/2), allowing salts to be flushed to waste, while retaining protein on the trap column. After the washing step, the trap column was switched in line and samples were separated by passing a gradient through the trap and a second analytical column (0.5 mm i.d. × 50 mm PLRP-s 1000 Å). A gradient from 20 to 95% solvent B in 22 min at a flow rate of 20 μL/min was used. Solvent A contained H_2O/acetonitrile/acetic acid (96/2/2) and solvent B contained acetonitrile/*n*-propanol/H_2O/acetic acid (73/15/10/2). Mass spectrometry was conducted using a Finnigan MAT TSQ 7000 (San Jose, CA) equipped with an electrospray ion source. The Finnigan TSQ7000 mass spectrometer has a dual mass range (*m*/*z* 10–2500 or 20–4000). Most proteins can be analyzed successfully by scanning a range of *m*/*z* 500–2000; however, for CTLA-4, no detectable ions were obtained for CTLA-4 in this range. The Finnigan TSQ7000 was therefore operated in the high mass range with scanning from *m*/*z* 1500 to 3800. The instrument *m*/*z* scale was calibrated during direct infusion of a 10 μM solution of lysozyme (Sigma Chemical Co.), which gave ion signals out to *m*/*z* 3578 (the +4 charge state). During LC/MS, the instrument was scanned using the first quadrupole in the positive ion mode. Data were acquired in the profile acquisition mode using a scan time of 5 s. ESI mass spectra were processed using the Finnigan MAT BIOMASS deconvolution software which converts the multiply charged ion mass spectra into a zero-charge plot indicating molecular mass.

Analytical Ultracentrifugation

Samples were dialyzed extensively against 10 mM Na phosphate, pH 7.0, 150 mM NaF. The loading concentrations of CTLA-4 and CD80 were 3.7 and 2.9 μM, respectively.

Experiments were performed at 4°C in a Beckman Model XLA ultracentrifuge using an An 60 Ti rotor. Data were collected at 238 or 239 nm (except for the 8.7 μM CD80 sample, which was collected at 275 nm) using six-channel Epon, charcoal-filled centerpieces with a 12-mm path length containing 110 μL samples and 125 μL buffer references. At intervals of 2 h, 20 successive radial scans were averaged using a

0.001-cm step size and equilibrium was assumed if no change in distribution was observed between sets of scans.

Data analysis was performed using both the HID program obtained from the Analytical Ultracentrifugation Facility at the University of Connecticut and software running under Igor (Wavemetrics, Lake Oswego, OR) and incorporating the algorithm of M. L. Johnson (15) obtained as a gift from P. Hensley (Pfizer, Inc., Groton, CT). The data were fitted with a single species model using the equation

$$c_r = c_m \exp[M(1 - \bar{v}\rho)\omega^2(r^2 - r_m^2)/2RT], \qquad [1]$$

where c_r and c_m are the concentrations of the protein at radial position r and at some reference position m (often the meniscus), respectively, M is the weight-average molecular weight, $\bar{v}$ is the partial specific volume, ρ is the solvent density, ω is the angular velocity, r is the radial distance from the center of rotation, r_m is the distance from the center of rotation to the reference position, R is the gas constant, and T is the absolute temperature in Kelvin. The fitting routine used by the HID program does not require *a priori* knowledge of solvent density or partial specific volumes as it fits data using the parameter, σ. This parameter is related to the buoyant molecular weight as

$$\sigma = \frac{M(1 - \bar{v}\rho)\omega^2}{RT}, \qquad [2]$$

where $M(1 - \bar{v}\rho)$ is the buoyant molecular weight.

The densities of the solvents used were measured gravimetrically ($\rho = 1.0013 \pm 0.0014$ g/mL for the standard solvent and $\rho = 1.1112 \pm 0.0008$ g/mL for the solvent containing 5 M GuHCl. The partial specific volumes ($\bar{v}$) of the polypeptide chains were calculated as described in Laue *et al.* (16) and are 0.723 mL g^{-1} for CTLA-4 and 0.725 mL g^{-1} for CD80 after correction for temperature using the equation from Laue *et al.* (16)

$$\bar{v}_T = \bar{v}_{25} + 4.25 \times 10^{-4}(T - 298.15), \qquad [3]$$

where $\bar{v}_{25}$ is the value calculated at 25°C. A table of amino acid partial specific volumes has more recently been reported by Kharakoz (17) that may replace the table presented by Laue *et al.* (16) as originally reported by Cohn and Edsall (18) in calculating protein partial specific volumes. Using this newer table as a reference, the partial specific volumes obtained for CTLA-4 and CD80 are both 0.732 mL g^{-1} after applying the temperature correction as shown in Eq. [3]. Since this difference in partial specific volumes as calculated by the two methods is only on the order of 1%, the values used in this work are based on the information taken from Laue *et al.* (16).

A temperature-corrected carbohydrate partial specific volume of 0.641 mL g^{-1} was calculated based on information from carbohydrate analysis of CTLA-4 (16). This same value was assumed for CD80 carbohydrates. For 5 M GuHCl, ϕ_2' was used in place of $\bar{v}$ to correct for GuHCl effects using the equation from Laue *et al.* (16)

$$\phi_2' = \bar{v} + \left\{\frac{1}{\rho} - \bar{v}_3\right\}(\delta_3 - g_3\delta_1), \qquad [4]$$

where the density of 5 M GuHCl, $\rho = 1.1112 \pm 0.0008$ g mL^{-1}, the partial specific volume of 5 M GuHCl, $\bar{v}_3 = 0.758$ mL g^{-1}, the degree of hydration of GuHCl, $g_3 = 1.007$ g GuHCl/g H_2O, and the degree of protein hydration (using the value for collagen), $\delta_1 = 0.48$ g H_2O/g protein. δ_3 is calculated using the relationship

$$\delta_3 = \frac{M_d}{M_p}\left\{\frac{N_t - 1}{2} + N_{aromatic} + N_{carbohydrate}\right\}, \qquad [5]$$

where the molecular weight of the denaturant $M_d = 95.54$, the molecular weight of CD80 $M_p = 46{,}400$, the number of amino acid residues $N_t = 216$; the number of aromatic residues $N_{aromatic} = 28$, and the number of carbohydrates (estimated from Table 3) $N_{carbohydrate} = 141$.

RESULTS

To determine the native molecular weights of the glycoproteins CTLA-4 and CD80, it is necessary first to define their covalent molecular weights as accurately as possible. SDS–PAGE analysis under nonreducing conditions for these two proteins suggests that their molecular weights are about 40,000 for CTLA-4 and 45,000 for CD80 (see Table 4). These weights, when compared to the expected polypeptide molecular weights of 27,716 for the disulfide-crosslinked dimer of CTLA-4 and 24,639 for CD80, indicate a large discrepancy. This must be due, in part, to significant levels of glycosylation of these proteins, although, in general, proteins can have aberrant mobility as measured by SDS–PAGE due to other effects.

We analyzed more rigorously the molecular weight of CTLA-4 by electrospray ionization mass spectrometry (ESI/MS) (Fig. 1). At least 10 peaks are evident, arising from heterogeneous glycosylation *in vivo*. CTLA-4 has two N-linked glycosylation sites at asparagine residues 78 and 111. The major species at Asn 78 is a disialylated biantennary core-fucosylated *N*-oligosaccharide as shown below (Wu, unpublished):

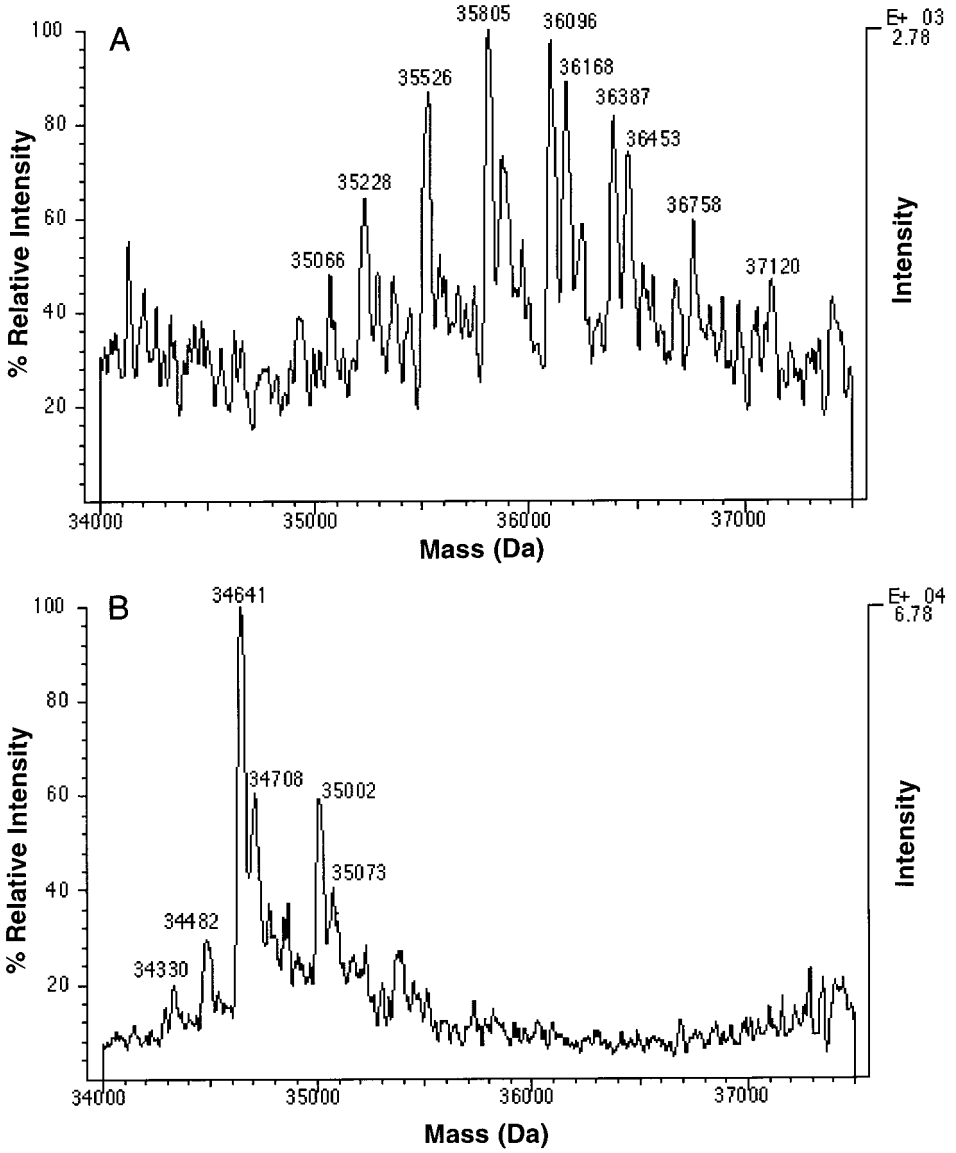
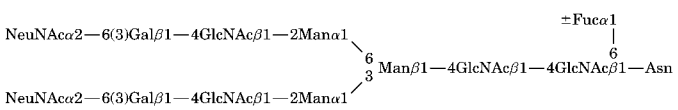


FIG. 1. Deconvoluted electrospray mass spectra of (A) native CTLA-4, and (B) CTLA-4 after treatment with neuraminidase.



The major oligosaccharide at Asn 111 is a monosialylated biantennary core-fucosylated *N*-oligosaccharide. Based on the known structures of the oligosaccharides on CTLA-4, it was possible to assign the majority of the peaks observed in the ESI/MS data. These data are summarized in Table 1. A large contribution to the heterogeneity results from the variable number, ranging from 2 to 7, of attached sialic acids to the carbohydrate structures. From the data in Fig. 1A and Table 1, it is evident that the structures containing between 4 and 5 sialic acids are the most abundant, consistent with the carbohydrate analysis. Additional heterogeneity also arises from variability in the N-terminal amino acid. The most abundant signals in the ESI data indicate that the N-terminal alanine is missing from both CTLA-4 chains. However, another series of peaks is evident which is shifted higher by 71 Da, indicating that a smaller percentage of CTLA-4 retains one of the N-terminal alanine residues.

Differential sialylation represents a major component of this heterogeneity as shown by mass analysis after treatment with neuraminidase, an enzyme that specifically removes sialic acid (Fig. 1B). The overall mass of the glycoprotein has been reduced, with the

TABLE 1

Interpretation of Results from LC/ESI/MS Analysis of Native and Neuraminidase-Treated CTLA-4 Samples

Sample	Experimental mass (Da)	Calculated mass (Da)	Proposed structures
Native CTLA-4	35228	35218	−2 Ala, −6 NeuNAc
	35525	35509	−2 Ala, −5 NeuNAc
	35803	35800	−2 Ala, −4 NeuNAc
	35870	35871	−1 Ala, −4 NeuNAc
	36096	36091	−2 Ala, −3 NeuNAc
	36167	36162	−1 Ala, −3 NeuNAc
	36387	36382	−2 Ala, −2 NeuNAc
	36451	36453	−1 Ala, −2 NeuNAc
	36758	36745	−1 Ala, −1 NeuNAc
Desialylated CTLA-4	34483	34474	−2 Ala, −8 NeuNAc, −Gal
	34640	34636	−2 Ala, −8 NeuNAc
	34707	34707	−1 Ala, −8 NeuNAc
	35004	34998	−1 Ala, −7 NeuNAc
	35070	35069	−7 NeuNAc

Note. NeuNAc, sialic acid; Gal, galactose. $-x$ NeuNAc indicates that x sialic acids have been removed from the fully sialylated structure. $-x$ Ala indicates whether one or both N-terminal alanine residues are missing.

predominant species being 34,461. This represents a loss of four sialic residues (M_r 291 per saccharide) from the predominant peak noted in Fig. 1A (M_r 35,805). The remaining heterogeneity in mass must be due to variation in addition of other sugars and N-terminal heterogeneity as described above.

Using the major peak from Fig. 1A as representative of the average mass of the attached oligosaccharides, we can derive the molecular weight of the carbohydrates by a simple difference from the expected polypeptide molecular weight. This difference is 8,089 or 23% by weight.

The native molecular weight of CTLA-4 was studied initially by size exclusion chromatography (Fig. 2; Table 4). The retention time compared to several protein molecular weight standards gave an apparent molecular weight of 73,800, close to that expected for a dimer. However, it is well known that retention time by size exclusion chromatography is both strongly shape dependent and influenced disproportionately by carbohydrate content.

A more accurate approach of measuring the native molecular weight is by sedimentation equilibrium ultracentrifugation. Figure 3 shows ultracentrifugation data for CTLA-4 collected at three different speeds. The data are shown fit with a single species model using the buoyant molecular weight as an adjustable parameter. The single species model shows a reasonably good fit to the data as judged by the randomness of the residuals. It is unlikely that the small nonrandom component observed in the residuals from this fit is caused by carbohydrate molecular weight heterogeneity. This heterogeneity is only on the order of ±1000 (as seen by ESI/MS in Fig. 1) in the background of a covalent molecular weight of approximately 36,000. This represents only a 3% variation in molecular weight, well within the 10% error typically seen for ultracentrifugation sample columns of 2 mm in height.

Therefore, the nonrandom variations in the residuals in Fig. 3A must be due to other processes such as dynamic self-association or aggregration. Additional evidence for this nonideal behavior comes from the apparent decrease in molecular weight as a function of rotor speed (Table 2). Further support for aggregation comes from the analysis of data collected at different protein concentrations. As the protein concentration is increased, the apparent molecular weight increases as well (data not shown).

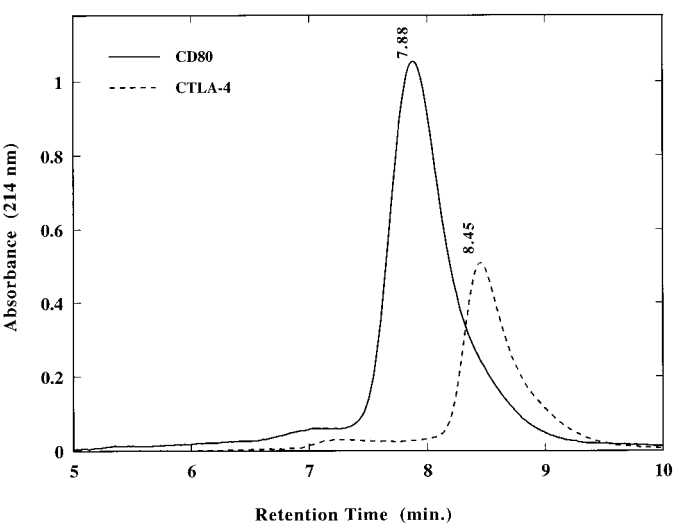


FIG. 2. Size exclusion chromatography analysis of CTLA-4 and CD80. The retention times (RTs) for CTLA-4 and CD80 are 8.45 and 7.88 min, respectively. Retention times of the standards are as follows: thyroglobulin (M_r 670,000; RT = 5.81 min); bovine γ-globulin (M_r 158,000; RT = 7.65 min); chicken ovalbumin (M_r 44,000; RT = 8.69 min); equine myoglobin (M_r 17,000; RT = 10.25 min).

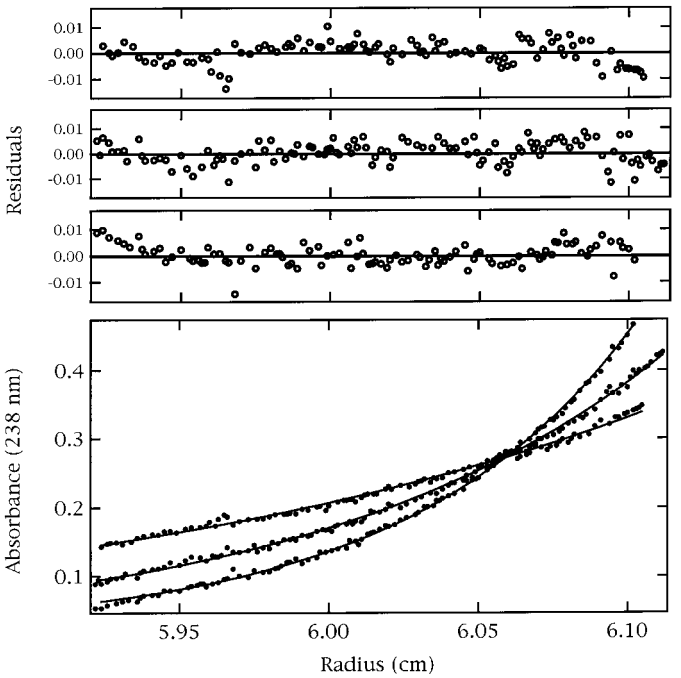


FIG. 3. Sedimentation equilibrium analysis of CTLA-4. A 3.7 μM sample of CTLA-4 was run at 4°C in 10 mM Na phosphate, pH 7.0, 150 mM NaF. The data were collected at rotor speeds of 8000, 12,000, and 16,000 rpm. These data are shown fit globally with a model describing a single ideal species and allowing the molecular weight to vary. The residuals from the curve fits are shown as well.

Careful analysis of the SE data suggests that a limited aggregation state is present. This conclusion is based on following the change in the square root of the variance (SQOV) from the fit to the data as a function of the model being used (Fig. 4A). The model that best minimizes the SQOV is a monomer–*n*-mer model where the *n* state is an aggregation state of 5. In this model no more than 5% of the protein is aggregated (Fig. 4B). NMR data (11) for CTLA-4 are consistent with a monomeric state being the major species in solution.

On a more technical note, it is often possible that single species analysis using global fits can show nonideal behavior owing to signficant problems with baseline absorbance. This can be assessed by depleting the meniscus of protein at high rotor speeds and evaluating the residual absorbance. If the residual absorbance is a significant fraction of the overall starting absorbance, then the uncertainty in this parameter will significantly affect the accuracy of molecular weight determinations. Such problems are often associated with incomplete dialysis of samples. The data for CTLA-4 were evaluated for residual absorbance at high rotor speeds. The residual absorbance was found to be sufficiently low so that it did not impact the conclusions drawn above. At present, quantitative methods for assessing the contribution of ill-determined baseline absorbances are not readily available.

Shire (1) describes a method to determine the molecular weight of a glycoprotein when its partial specific volume is not known. First, the data must fit reasonably well to a single species (as described above) as judged by the buoyant molecular weight, $M(1 - \bar{\nu}\rho)$. Having satisfied this requirement, the buoyant molecular weight can be converted to an absolute molecular weight by making certain assumptions about the partial specific volume. The partial specific volume of the polypeptide chain can be calculated from the amino acid composition as described by Laue *et al.* (16) and has been shown to agree well with experimentally determined values; the same can be done for carbohydrates. In the case of CTLA-4, the polypeptide partial specific volume is 0.723 mL g^{-1} (see Materials and Methods) and the carbohydrate partial specific volume is 0.641 mL g^{-1}. Using these values, along with the experimentally determined buoyant molecular weight, the carbohydrate content, f_{carb}, can be computed as

$$f_{carb} = \frac{M_{gp,buoy} - M_{p,buoy}}{M_{gp,buoy} + (\bar{\nu}_p - \bar{\nu}_{carb})\rho M_p}, \qquad [6]$$

where $M_{gp,buoy}$ is the experimentally determined buoyant molecular weight of the glycoprotein, $M_{p,buoy}$ is the computed buoyant molecular weight for the polypeptide chain, M_p is the molecular weight of the polypeptide chain, and $\bar{\nu}_p$ and $\bar{\nu}_{carb}$ are the partial specific volumes for the polypeptide and the carbohydrate components, respectively. This computation gives a value of 27.4% carbohydrate content by weight for CTLA-4 (Table 3). The average mass percentage as determined by ESI/MS is 23%, which is 4% smaller than that measured by SE. This discrepancy in mass can be accounted for by the observation of a small amount of aggregation for CTLA-4. Having determined the percentage carbohydrate in this fashion, it is then

TABLE 2

Speed Dependence of Apparent Molecular Weights of CTLA-4 and CD80

	Speed (rpm)		
Sample	12,000	16,000	20,000
CTLA-4[a]	46,800 ± 16 %	41,300 ± 6.0%	35,800 ± 5.6%
CD80[a]			
Exp 1	57,400 ± 11 %	54,000 ± 4.2%	45,700 ± 3.2%
Exp 2	51,400 ± 6.7%	48,500 ± 2.2%	44,200 ± 1.5%
8.7 μM	46,000 ± 6.7%	45,600 ± 2.7%	43,300 ± 3.1%
+5 M GuHCl	55,900 ± 12 %	51,500 ± 4.0%	46,200 ± 2.1%

[a] The partial specific volumes used for CTLA-4 and CD80 are 0.700 and 0.685 mL g^{-1}, respectively, as determined in Table 3.

A

Model	SQOV (x10^3)
single species	4.18
monomer	4.39
dimer	14.7
1<-->2	4.12
1<-->3	4.07
1<-->4	4.05
1<-->5	4.04
1<-->6	4.05

B

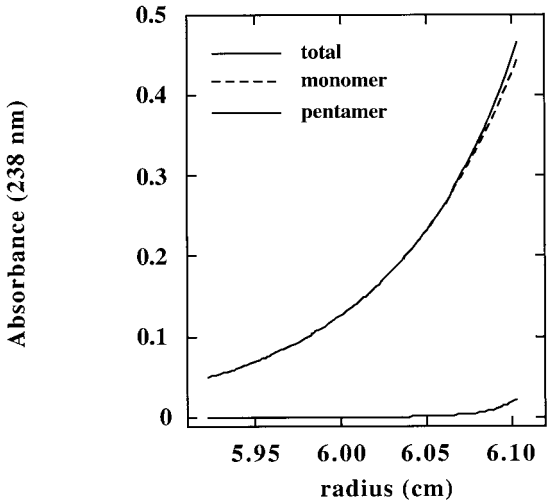


FIG. 4. Model analysis of sedimentation equilibrium data for CTLA-4 (A) Square root of variance reported for data fitted with various models. The analysis uses global fits to 8000, 12,000, 16,000, and 20,000 rpm data. The single species model treats the molecular weight as a varying parameter. Monomer and dimer models are single-species models where the monomer molecular weight is based on the dominant molecular weight as determined by ESI/MS. The remaining models represent monomer–*n*-mer fits. (B) Partition of the total absorbance for the 20,000 rpm data for CTLA-4 between a monomer and a pentamer species.

straightforward, in general, to compute the expected partial specific volume for a glycoprotein. For CTLA-4, this computes to 0.700 mL g^{-1}; this value can now be used to compute the apparent molecular weight from the buoyant molecular weight and a value of 38,200 is obtained (Table 3). This molecular weight is consistent with a monomeric state for this protein.

As an aside, if one uses the partial specific volume for the polypeptide chain as computed from the Kharakoz table (17) for this analysis, a molecular weight of 41,000 is obtained, predicting that the carbohydrate is 30% by weight, thus being further removed from agreement with the ESI/MS data. It is unclear what might be the cause of this further discrepancy.

As Shire points out (1), even if the carbohydrate composition is unknown, and $\bar{\nu}_{carb}$ cannot be calculated, using an estimate of 0.63 mL g^{-1} for this parameter does not seriously compromise the ability to determine the molecular weight by this method. Table 3 shows how the apparent molecular weight for CTLA-4 varies by using a reasonable range of values of 0.63 and 0.65 for $\bar{\nu}_{carb}$.

The validation of this SE method using CTLA-4 encouraged us to apply this method to CD80 where the carbohydrate content is not known. We were not able to obtain sufficiently resolved peaks in the ESI/MS data for CD80 due to the increased level of glycosylation observed in this protein. First, the CD80 data were analyzed using an ideal single species model. The data shown in Figs. 5a and 5b are fit well using a single species model. However, as with CTLA-4 some decrease in molecular weight was noted as a function of increasing rotor speed. In the case of CD80, this effect is probably not caused by

TABLE 3

Effect of Assumed Carbohydrate Partial Specific Volume on Biophysical Parameters

Protein[a]	$\bar{\nu}_{carb}$[b]	$\bar{\nu}_{gp}$[c]	M_r	% carb[d]
CTLA-4	0.63	0.698	37,800	26.8
	0.641[e]	**0.700**	**38,200 ± 7.5%**	**27.4**
	0.65	0.703	38,400	27.9
CD80				
(1)[f]	0.63	0.679	47,800	48.5
	0.64[g]	**0.683**	**48,500 ± 5.2%**	**49.2**
	0.65	0.688	49,200	49.9
(2)	0.63	0.681	45,500	45.9
	0.64[g]	**0.685**	**46,000 ± 3.0%**	**46.5**
	0.65	0.690	46,700	47.2
8.7 μM	**0.64**[g]	**0.687**	**44,800 ± 4.5%**	**45.0**
5 M GuHCl	**0.64**[g]	**0.688**	**46,000 ± 4.3%**	**46.5**

[a] Protein loading concentration is 2.9 μM.
[b] Carbohydrate partial specific volume.
[c] Glycoprotein weight averaged partial specific volume as calculated from the % contributions from carbohydrate (%carb) and protein content.
[d] % carbohydrate as calculated using Eq. [6].
[e] Numbers in bold are computed based on Eq. [6]. Nonbold numbers represent the effect of using test values of 0.63 and 0.65 mL g^{-1} for the carbohydrate partial specific volume.
[f] (1) and (2) represent two independent experiments.
[g] Assumed values for carbohydrate partial specific volume based on value for CTLA-4.

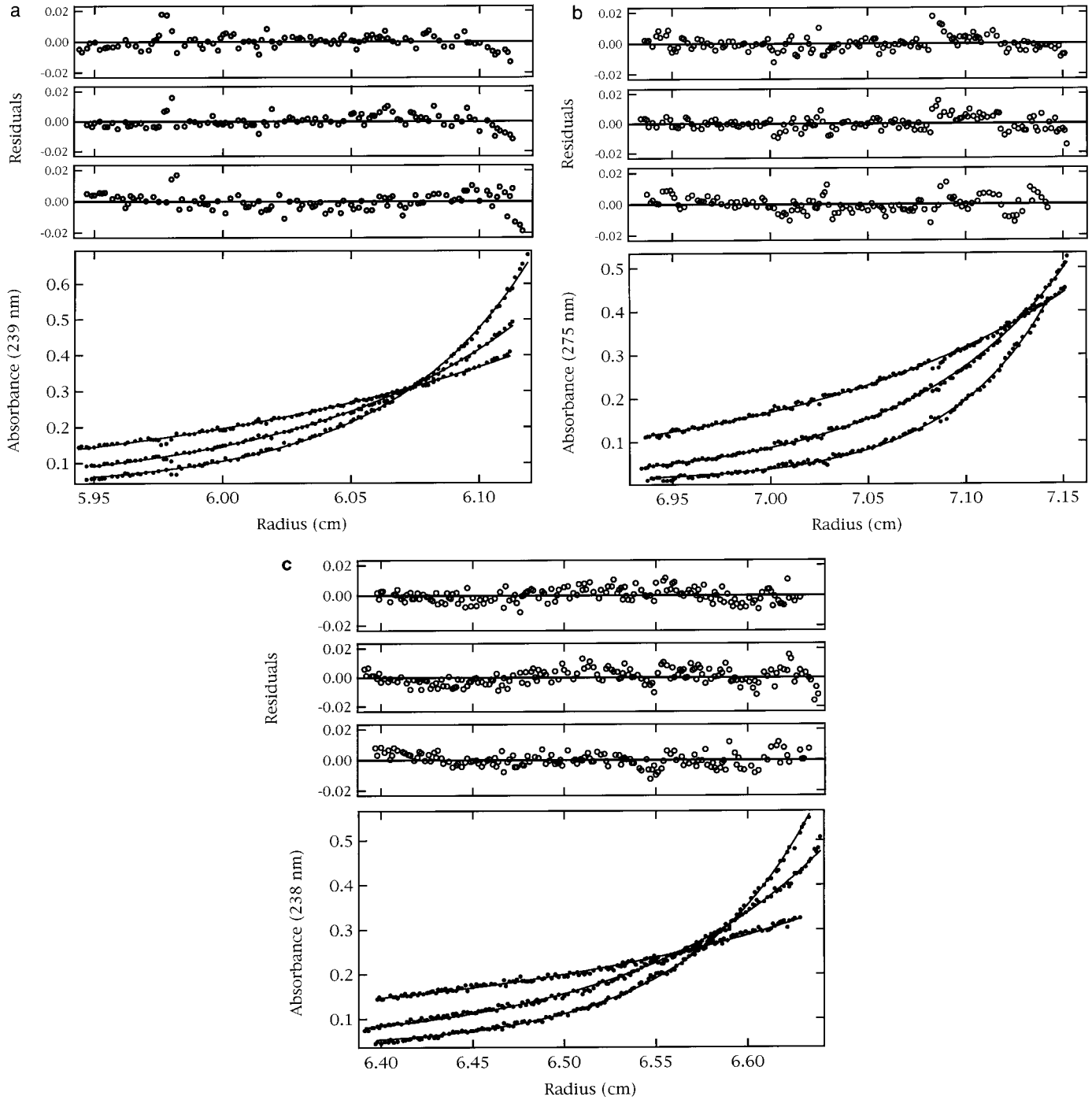


FIG. 5. Sedimentation equilibrium analysis of CD80. The same three speeds shown in Fig. 3 were used in the analysis of CD80 at (a) 2.9 μM CD80, (b) 8.6 μM CD80, and (c) 2.9 μM CD80 in 5 M GuHCl.

problems with self-association because the molecular weight does not increase as a function of protein concentration (Table 2). In fact, this trend in decreasing molecular weight persists even when measured in 5 M GuHCl (Table 2), a concentration of denaturant that fully unfolds CD80 (data not shown). Therefore, in the case of CD80, the apparent heterogeneity as judged by the effects of rotor speed on molecular weight must be due to heterogeneity in the covalent molecular weight (i.e., differential gly-

TABLE 4

Comparison of Molecular Weights for CTLA-4 and CD80

Protein	Polypeptide	SDS–PAGE[a]	Gel filtration	Mass spectrometry	Analytical ultracentrifugation
CTLA-4	27,716[b]	40,000	73,800	35,805	38,200
CD80	24,639	45,000	126,400	—	48,500

[a] Samples for SDS–PAGE were prepared under nonreducing conditions.
[b] Molecular weight of disulfide-crosslinked species.

cosylation). This type of heterogeneity that is not in chemical equilibrium is often referred to as polydispersity.

Using Eq. [6], we obtain a value of 49.2% carbohydrate by weight for CD80. Further computation, using this percentage, yields a partial specific volume of 0.683 mL g^{-1} and a native molecular weight of 46,400. This analysis assumes that CD80 is a single ideal species in solution. Based on these data collected under native conditions, it is not possible to distinguish between a highly glycosylated monomer state or a stable unglycosylated dimer state since the expected polypeptide molecular weight is 24,639, or about half the experimentally determined value. To rule out the dimer model, CD80 was sedimented in 5 M GuHCl, correcting for GuHCl effects as described under Materials and Methods (Fig. 5c). The molecular weight as measured in GuHCl agrees well with that determined by SDS–PAGE (Table 4) and is the same as that measured by SE under native conditions (Table 3), confirming that CD80 behaves as a monomer in solution. This can be contrasted to the behavior observed by size exclusion chromatography where the calculated molecular weight is 126,400 (Fig. 2, Table 4), a size nearly three times the covalent molecular weight of CD80.

DISCUSSION

Determination of accurate molecular weights for heavily glycosylated proteins can be a difficult task. Addition of carbohydrates typically results in aberrantly high molecular weights as determined by size exclusion chromatography (Fig. 2). Even determining the covalent molecular weight can be challenging. Heavily glycosylated proteins typically are highly diffuse in SDS–PAGE analysis and mass spectrometry often fails to provide resolution owing to the heterogeneity of glycosylation, in both molecular weight and charge. Therefore, often the only recourse in determining native molecular weights is by analytical ultracentrifugation. Sedimentation velocity experiments suffer from similar constraints such as size exclusion chromatography, particularly with respect to hydrodynamic radius and shape dependence. In contrast, sedimentation equilibrium experiments overcome these limitations and, with some effort, significant information regarding native molecular weights can be extracted for heavily glycosylated proteins.

Problems associated with glycoproteins that are inherent in sedimentation equilibrium include polydispersity due to charge heterogeneity (owing to the negative charges associated with addition of the carbohydrate, sialic acid) and size heterogeneity. It is most challenging to distinguish models of self-association from polydispersity because these phenomena have similar hallmarks in their sedimentation equilibrium behavior. Distinguishing these two phenomena requires multiple experiments to study the protein loading concentration dependence and behavior under denaturing conditions. CTLA-4 provides a good test of the sedimentation equilibrium methodology because we were able to obtain accurate molecular weight information via ESI/MS. Further, these studies using CTLA-4 provide an added level of confidence in our sedimentation equilibrium analysis of CD80 for which we were not able to get accurate ESI/MS data.

Surprisingly, there is little literature on determining molecular weights of glycoproteins using sedimentation equilibrium ultracentrifugation, probably because of the difficulties in measuring the partial specific volumes correctly and accurately. Because of sample limitations, we could not use pycnometry to directly measure partial specific volumes. We attempted to determine partial specific volumes for CTLA-4 and CD80 by using the well-established H_2O/D_2O method (13). In this method, since the densities of H_2O and D_2O are significantly different, the difference in the buoyant molecular weights in these two solvent systems often is sufficiently great to provide reasonably accurate partial specific volumes. Using this method, we measured partial specific volumes of 0.716 and 0.711 mL g^{-1} for CTLA-4 and CD80, respectively. We then used these partial specific volumes to calculate molecular weights and obtained values of 45,200 and 61,000. The molecular weight for CTLA-4 is in poor agreement with that determined by ESI/MS and appears to be a consequence of a gross overestimation of

the partial specific volume. It is clear that accuracy of this experimental method is highly dependent on the degree of heterogeneity of a system and is inadequate for the systems in consideration here.

Instead, we apply a method first introduced by Shire (1) where the partial specific volume can be determined empirically. In addition to determining the oligomeric state of the protein of interest, one can also establish the mass fraction of bound carbohydrate. A prerequisite for this empirical method is that one's data can be treated adequately using a single species model.

While we show that CTLA-4 is heterogeneous in its glycosylation, we can still treat the data reasonably well with a single-species model since the degree of mass heterogeneity (3%) is within the accuracy of our measurements of weight (10%). We also find that a small fraction of the protein is aggregated and thus contributes to the small amount of nonrandomness that we observe in the residuals from the curve fits. This fraction of aggregated material probably leads to the overestimation of the molecular weight and carbohydrate content when compared to the ESI/MS data.

Despite the minor problems owing to aggregation, the relatively good agreement in molecular weights obtained for CTLA-4 by SDS–PAGE, ESI/MS, and SE analytical ultracentrifugation provided us with a sufficient level of confidence to determine the oligomeric state of the extracellular domain of CD80. Applying the same SE methodology described for CTLA-4, we show that CD80 is a monomer in solution with approximately 49% of its mass being contributed by carbohydrates.

Finally, having firmly established the molecular weights of these extracellular domains of CTLA-4 and CD80 in solution using an emerging technology, it will be possible to determine the heteromeric molecular weight of a CTLA-4/CD80 complex. Critical analysis of the heteromeric complex should lead to an understanding of the role that this protein–protein interaction plays in the potential to induce self-association and subsequent activation of signal transduction mechanisms.

ACKNOWLEDGMENTS

We thank Gary Matsueda and Thom Lavoie for many invaluable discussions and support throughout the course of this work.

REFERENCES

1. Shire, S. J. (1994) *in* Modern Analytical Ultracentrifugation (Schuster, T. M., and Laue, T. M., Eds.), pp. 261–297, Birkhäuser, Boston.
2. Brunet, J.-F., Denizot, F., Luciani, M.-F., Roux-Dosseto, M., Suzan, M., Mattei, M.-G., and Golstein, P. (1987) *Nature* **328,** 267–270.
3. Freeman, G. J., Freedman, A. S., Segil, J. M., Lee, G., Whitman, J. F., and Nadler, L. M. (1989) *J. Immun.* **143,** 2714–2722.
4. June, C. H., Ledbetter, J. A., Linsley, P. S., and Thompson, C. B. (1990) *Immunol. Today* **11,** 211–216.
5. Schwartz, R. H. (1992) *Cell* **71,** 1065–1068.
6. June, C. H., Bluestone, J. A., Nadler, L. M., and Thompson, C. B. (1994) *Immunol. Today* **15,** 321–332.
7. Linsley, P. S., Brady, W., Grosmaire, L., Aruffo, A., Damle, N. K., and Ledbetter, J. A. (1991) *J. Exp. Med.* **173,** 721–730.
8. Linsley, P. S., Brady, W., Urnes, M., Grosmaire, L., Damle, N. K., and Ledbetter, J. A. (1991) *J. Exp. Med.* **174,** 561–569.
9. Linsley, P. S., and Ledbetter, J. A. (1993) *Annu. Rev. Immunol.* **11,** 191–212.
10. Linsley, P. S., Nadler, S. G., Bäjorath, J., Peach, R., Leung, H. T., Rogers, J., Bradshaw, J., Stebbins, M., Leytze, G., Brady, W., Malacko, A. R., Marquardt, H., and Shaw, S.-Y. (1995) *J. Biol. Chem.* **270,** 15417–15424.
11. Metzler, W. J., Bajorath, J., Fenderson, W., Shaw, S.-Y., Constantine, K. L., Naemura, J., Leytze, G., Peach, R. J., Lavoie, T. B., Mueller, L., and Linsley, P. S.. (1997) *Nat. Struct. Biol.* **4,** 527–531.
12. Linsley, P. S., Wallace, P. M., Johnson, J., Gibson, M. G., Green, J. L., Ledbetter, J. A., Singh, C., and Tepper, M. A. (1992) *Science* **257,** 792–795.
13. Edelstein, S. J., and Schachman, H. K. (1973) *Methods Enzymol.* **27,** 82–98.
14. Edelhoch, H. (1967) *Biochemistry* **6,** 1948–1954.
15. Johnson, M. L., Correia, J. J., Yphantis, D. A., and Halverson, H. R. (1981) *Biophys. J.* **36,** 575–588.
16. Laue, T. M., Shah, B. D., Ridgeway, T. M., and Pelletier, S. L. (1992) *in* Analytical Ultracentrifugation in Biochemistry and Polymer Science (Harding, S. E., Rowe, A. J., and Horton, J. C., Eds.), pp. 90–125, R. Soc. Chem., Cambridge.
17. Kharakoz, D. P. (1997) *Biochemistry* **36,** 10276–10285.
18. Cohn, E. J., and Edsall, J. T. (1943) Proteins, Amino Acids and Peptides as Ions and Dipolar Ions, Reinhold, New York.