9830 *Biochemistry* **2010,** *49,* 9830–9838
DOI: 10.1021/bi101144q

BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Self-Association of the Adenoviral L4-22K Protein[†]

Teng-Chieh Yang and Nasib Karl Maluf*

Department of Pharmaceutical Sciences, School of Pharmacy, University of Colorado Denver, C238-P15, 12700 East 19th Avenue, Aurora, Colorado 80045, United States

Received July 19, 2010; Revised Manuscript Received October 4, 2010

ABSTRACT: Human adenovirus (Ad) is an icosahedral, double-stranded DNA virus that causes infections of the respiratory tract, urinary tract, and gastrointestinal tract. Assembly of virus particles requires condensation and encapsidation of the linear viral genome. This process requires sequence specific binding of two viral proteins, called IVa2 and L4-22K, to a conserved sequence located at the left end of the viral genome, called the packaging sequence (PS). IVa2 and an alternatively spliced form of L4-22K, called L4-33K, also function as transcriptional activators of the major late promoter (MLP), which encodes viral structural and core proteins. IVa2 and L4-33K bind to identical conserved DNA sequences downstream of the MLP, called the downstream element (DE), to activate transcription. To begin to dissect how the IVa2, L4-22K, and L4-33K proteins simultaneously function as transcriptional activators and DNA packaging proteins, we need to understand the thermodynamics of assembly of these proteins on DNA that contains the PS as well as the DE. Toward this end, we have characterized the self-assembly properties of highly purified, recombinant L4-22K protein. We show that L4-22K reversibly assembles into higher-order structures according to an indefinite, isodesmic assembly scheme. We show that the smallest polymerizing unit is likely the L4-22K monomer ($s_{20,w} = 2.16 \pm 0.04$ S) and that the monomer assembles with itself and/or other aggregates with an equilibrium association constant, L, of 112 (102, 124) μM^{-1} (0.1 M NaCl, pH 7, 25 °C). A mechanistic consequence of an isodesmic, indefinite assembly process is that the free concentration of the smallest polymerizing unit cannot exceed $1/L$. We discuss the implications of this observation with respect to the thermodynamics of assembly of L4-22K and IVa2 on the PS.

Human adenovirus (Ad)[1] is a nonenveloped, icosahedral virus that causes respiratory, ocular, urinary tract, and gastrointestinal tract infections (*1*). Ad causes significant levels of morbidity in immunocompromised individuals, such as organ transplant recipients who are taking immunosuppressive drugs, and in some cases can lead to death (*1*). Ad genome encapsidation is the process by which the viral genome is sequestered inside the capsid, eventually leading to production of infectious virus particles (*2*, *3*). Genome encapsidation requires the sequence specific binding of two viral proteins, IVa2 and L4-22K, to the packaging sequence (PS) located at the left end of the linear genome (*4*−*7*). The PS is composed of multiple copies of repeated sequence elements, called A repeats (*8*, *9*), which suggests that cooperative binding of multiple copies of IVa2 and L4-22K on the PS is required to signal the genome encapsidation reaction (*4*, *6*).

In addition to their roles in signaling genome encapsidation, IVa2 and an alternatively spliced variant of L4-22K, called L4-33K, also function to stimulate transcription of the major late promoter (MLP) (*10*), which predominantly controls production of viral structural and core proteins required for construction of infectious virus particles. This is accomplished by sequence specific binding to the so-called DE sequence, which resembles (but is not identical to) a single copy of an A repeat (*10*−*12*). Although the L4-33K protein recognizes the same DNA sequence that L4-22K recognizes, L4-33K does not support viral DNA encapsidation (*5*). Thus, if L4-33K is produced to high levels in the cell, it is expected that it would compete with L4-22K for binding to the PS and subsequently inhibit production of infectious virus particles. Furthermore, recent data suggest that the L4-33K protein, and not the L4-22K protein, is required for stimulation of the MLP (*10*). If this is correct, then production of the L4-22K protein would be expected to compete with that of L4-33K for stimulation of transcription of the MLP.

Transcription from the MLP is required to produce optimal concentrations of viral proteins for construction of infectious virus particles. Once these optimal concentrations have been reached, viral genome encapsidation must ensue, which requires binding of IVa2 and L4-22K to identical sequence elements present in the PS that are present in the DE sequence. An important question that arises from these considerations is how the IVa2, L4-22K, and L4-33K proteins simultaneously control transcription from the MLP while also signaling viral genome encapsidation from the PS, when they use the same DNA sequences for specific binding. To develop a quantitative and predictive model of this process, we must understand the thermodynamics of assembly of IVa2, L4-22K, and L4-33K on the PS and DE sequence. In pursuit of these aims, we have characterized the self-association properties of the highly purified,

[†]This work was supported by start-up funds provided by the Department of Pharmaceutical Sciences, School of Pharmacy, University of Colorado Denver.

*To whom correspondence should be addressed: Department of Pharmaceutical Sciences, School of Pharmacy, University of Colorado Denver, C238-P15, 12700 E. 19th Ave., Aurora, CO 80045. Phone: (303) 724-4036. Fax: (303) 724-2627. E-mail: karl.maluf@ucdenver.edu.

[1]Abbreviations: NLLS, nonlinear least-squares; SSR, sum of squared residuals; SD_{fit}, standard deviation of the fit; Ad, adenovirus; PS, packaging sequence; DE, downstream element; MLP, major late promoter.

 Published on Web 10/06/2010

recombinant, L4-22K protein. We show that L4-22K reversibly assembles into large, higher-order structures, up to ~70-mers at total monomer concentrations approaching 10 μM. We are able to quantitatively model this assembly according to a reversible, indefinite isodesmic assembly model. We discuss some biological implications of macromolecules participating in indefinite assembly reactions.

EXPERIMENTAL PROCEDURES

Reagents. LB broth and LB agar were purchased from Fisher Scientific. Chitin beads were purchased from New England Biolabs. Q-Sepharose FF and SP-Sepharose FF chromatography resins were purchased from GE Healthcare. TCEP-HCl was purchased from Thermo scientific. Urea was purchased from Fisher scientific.

L4-22K Protein Purification. The L4-22K gene was cloned into pTYB4 (NEB) using a published protocol (*4*) and transformed into *Escherichia coli* BL21(RIPL) (Stratagene). The cells were grown in 2 × 1 L of LB broth to an OD of 0.8 at 37 °C, followed by induction with 0.5 mM IPTG for approximately 20 h at 4 °C, and harvested by centrifugation. The L4-22K protein purification was based on a previous protocol (*4*), with modifications. All purification steps were performed at 4 °C. The cell pellet was resuspended with 50 mL of lysis buffer [750 mM NaCl, 20 mM HEPES (pH 8.0), 0.5 mM EDTA, 0.1% Triton X-100, 1 mM β-me, 20 μM PMSF, and 1 tablet of protease inhibitor (Roche)] and lysed by sonication. The lysed cells were centrifuged at 20000*g* for 30 min, and the supernatant was then added to 3 × 6 mL of 50% chitin resin (NEB) that was equilibrated in lysis buffer. The mixture was rocked at 4 °C for 1 h. The chitin resin was then washed with 5 column volumes of wash buffer 1 [1.25 M NaCl, 20 mM HEPES (pH 8.0), 0.5 mM EDTA, 0.1% Triton X-100, and 1 mM β-me], followed by 3 column volumes of wash buffer 2 [150 mM NaCl, 20 mM HEPES (pH 8.0), 0.5 mM EDTA, 1 mM β-me, and 0.1% Triton X-100]. The protein was first eluted with 6 column volumes of elution buffer [150 mM NaCl, 20 mM HEPES (pH 8.0), 0.5 mM EDTA, 50 mM DTT, and 0.1% Triton X-100] by incubation at 4 °C for at least 20 h, with gentle rocking, and then further eluted with 10 column volumes of elution buffer. The eluted fractions were pooled and loaded onto a 30 mL Q Sepharose (fast flow) column (Pharmacia) that was equilibrated with 10 column volumes of equilibration buffer [150 mM NaCl, 20 mM HEPES (pH 8.0), 0.5 mM EDTA, 5 mM β-Me, and 0.1% Triton X-100]. The protein was eluted with a linear salt gradient from 150 mM to 1 M NaCl over 25 column volumes. The fractions containing the L4-22K protein were collected, and the salt concentration was adjusted to 150 mM NaCl. The L4-22K protein was further purified and concentrated by SP Sepharose chromatography (HI trap SP column, 5 mL, from Pharmacia). The SP column was equilibrated with SP buffer [150 mM NaCl, 20 mM HEPES (pH 8.0), 0.5 mM EDTA, and 5 mM β-Me]. After the sample had been loaded onto the SP column, the column was washed with 5 column volumes of SP buffer then eluted with elution buffer [1 M NaCl, 20 mM HEPES (pH 8.0), 0.5 mM EDTA, and 5 mM β-Me]. The purified protein was dialyzed exhaustively against storage buffer [250 mM NaCl, 40 mM HEPES (pH 7.6 and 25 °C), 50% (v/v) glycerol, 0.5 mM EDTA, and 1 mM TCEP]. The protein concentration was determined by absorbance spectroscopy using an extinction coefficient of 3.24×10^4 M^{-1} cm^{-1}. The L4-22K protein purified by this procedure was more than 95% pure as judged by SDS−PAGE followed by Commassie staining (see Figure 1B). The identity of the L4-22K band was confirmed by trypsin digestion of the SDS−PAGE-purified protein, followed by MALDI-TOF/TOF analysis of the fragments; the peptide fragments were analyzed using MASCOT.

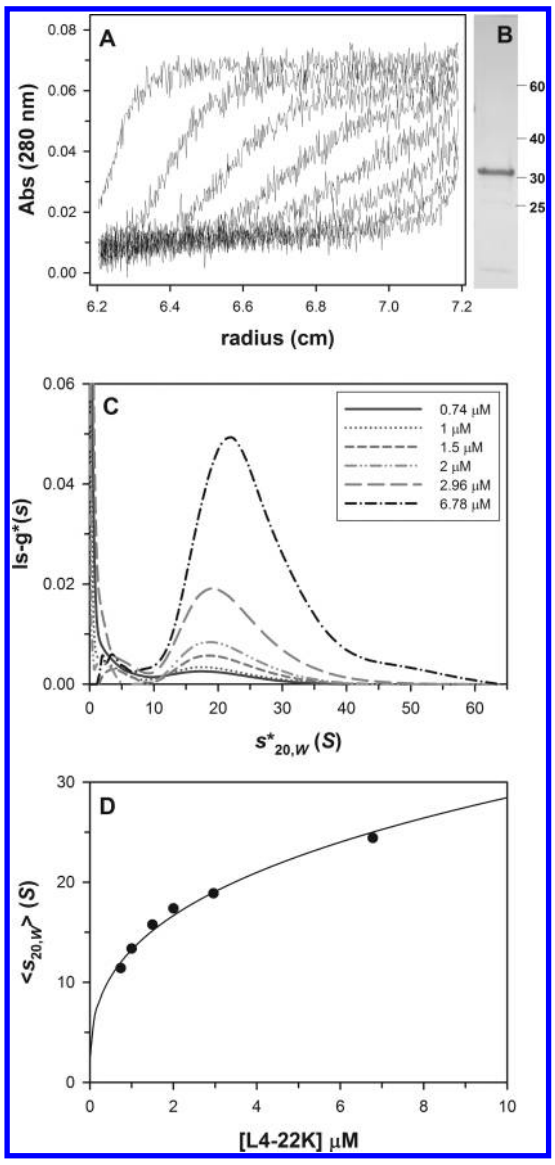

FIGURE 1: Sedimentation velocity analysis of the purified L4-22K protein. (A) A sedimentation velocity experiment was performed at 2 μM L4-22K in buffer P. Data were collected at 280 nm and 30K rpm. Every 10th scan is shown. (B) SDS−PAGE gel showing the purified L4-22K protein. The purified protein runs at ~35 kDa, which is consistent with previous observations (*4*). The identity of this polypeptide was confirmed by mass spectroscopy, as described in Experimental Procedures. The purity of L4-22K was greater than 95%. (C) Effect of concentration on the aggregation state of the protein. The total loading concentrations are indicated in the figure. The data were analyzed according to the ls-g*(s) method using SEDFIT. (D) Weight average sedimentation coefficients $\langle s_{20,w} \rangle$ from the distributions in panel C are plotted as a function of L4-22K loading concentration. The smooth curve is the result of an NLLS analysis using the isodesmic indefinite assembly model (eqs 9 and 10). The resulting best fit value of L was 69 (64, 74) μM^{-1}.

Sedimentation Velocity Experiments. Sedimentation velocity experiments were performed using a Beckman XLA analytical ultracentrifuge. The purified protein was dialyzed into buffer P [100 mM NaCl, 10 mM Na_2HPO_4/NaH_2PO_4 (pH 7.0 and 25 °C), and 1 mM TCEP] and then diluted with the buffer dialysate to

the desired concentration. In general, three protein concentrations were analyzed per run. Protein samples were loaded into Epon charcoal-filled two-sector centerpieces, and experiments were performed at 25 °C and 30K rpm. The samples and instrument were equilibrated at 25 °C for at least 2 h prior to initiation of the experiment, and in general, the protein dilutions were kept on ice for approximately 1 h prior to being loaded into the sample cells. The data were collected at 230 or 280 nm, every 0.002 cm with two averages in the continuous scan mode. The raw sedimentation velocity data were analyzed using SEDFIT (*13*), and the sedimentation coefficients were corrected to standard conditions (20 °C, in H_2O) using SEDNTERP.

Sedimentation Equilibrium Experiments. The purified protein was dialyzed into buffer P [100 mM NaCl, 10 mM Na_2HPO_4/NaH_2PO_4 (pH 7.0 and 25 °C), and 1 mM TCEP]. The samples (110 μL) were loaded into Epon charcoal-filled two-sector or six-sector centerpieces. The sedimentation equilibrium experiments were performed at 25 °C, at the indicated rotor speed. Data were collected at 230 or 280 nm, every 0.001 cm in the step mode, with 20 averages per step. Equilibrium was confirmed using Winmatch (*14*) and determined by the absence of deviations between successive scans.

Analysis of Sedimentation Velocity and Equilibrium Data According to Indefinite Assembly Models. The indefinite, isodesmic self-assembly model has been discussed in detail (*15*−*19*). This model assumes the molecule of interest can continuously assemble into higher-order structures with no upper bound on their stoichiometries. This process is summarized in eq 1:

$$[A_T] = [A_1] + 2[A_2] + 3[A_3] + ... \quad (1)$$

where $[A_T]$ is the total monomer concentration in molar units and each term on the right-hand side is the molar concentration of each stoichiometric species. The isodesmic model assumes that each successive addition of another monomer to the growing aggregate is governed by the same equilibrium constant, L, such that

$$L = \frac{[A_n]}{[A_{n-1}][A_1]} \quad (2)$$

where n is the species stoichiometry and $[A_1]$ is the concentration of the free monomer. Upon substitution of eq 2 into eq 1, one obtains eq 3:

$$[A_T] = [A_1] + 2L[A_1]^2 + 3L^2[A_1]^3 + ... \quad (3)$$

This infinite series can be expressed in a closed form solution (*17*):

$$[A_T] = \frac{[A_1]}{(1 - L[A_1])^2} \quad (4)$$

where it is implicitly understood that the product $L[A_1]$ must be less than 1 (*17*). If this was not the case, the infinite sum in eq 3 would be guaranteed to diverge. To use this result to analyze our sedimentation velocity data, we followed the approach outlined by Na et al. (*16*). The weight average sedimentation coefficient is calculated as follows:

$$\langle s_{20,w} \rangle = \frac{s_1[A_1] + 2s_2[A_2] + 3s_3[A_3] + ...}{[A_T]} \quad (5)$$

The sedimentation coefficient of each species is assumed to follow a general scaling law (*16*), as follows:

$$s_n = s_1 n^m \quad (6)$$

Upon substitution of eqs 2 and 6 into eq 5, one obtains (*16*)

$$\langle s_{20,w} \rangle = \frac{s_1 \sum_{n=1}^{\infty} n^{1+m} L^{n-1} [A_1]^n}{[A_T]} \quad (7)$$

In general, it is not possible to obtain a closed form solution for the infinite sum in eq 7. However, eq 7 can still be useful for NLLS fitting procedures. For any particular total concentration, $[A_T]$, the fraction of each stoichiometric species is given by

$$f_n = \frac{nL^{n-1}[A_1]^n}{[A_T]} \quad (8)$$

Because convergence of the infinite sum shown in eq 3 requires that $L[A_1]$ be less than 1, we can see that as n increases toward infinity, eventually f_n must approach zero [because $(L[A_1])^n$ will decrease exponentially while n will increase linearly]. This indicates that even though the system can continue to assemble indefinitely, for a fixed $[A_T]$, only a finite fraction of these species will actually exist at appreciable levels in solution. Thus, use of eq 7 in practice simply requires replacing ∞ in the summation with a sufficiently large finite number, Q, that ensures f_Q is essentially equal to zero (*16*):

$$\langle s_{20,w} \rangle = \frac{s_1 \sum_{n=1}^{Q} n^{1+m} L^{n-1} [A_1]^n}{[A_T]} \quad (9)$$

In our hands, and for our particular system, Q values on the order of 300−400 were more than sufficient to ensure the summation had converged. This was determined by checking that the final calculated value of $[A_T]$ was essentially equal to the known $[A_T]$, i.e., $[A_T]_{calc}$ was at least 99.9% of the known $[A_T]$ (*16*). This was conducted using the SUMMATION command provided in the SCIENTIST (MicroMath) NLLS software package. In eq 9, $[A_1]$ is calculated for each $[A_T]$ by solving eq 4 for $[A_1]$ (*15*), which yields [after taking the (−) root, which guarantees $[A_T] > [A_1]$]

$$[A_1] = \frac{1}{2L^2[A_T]} + \frac{1}{L} - \frac{\sqrt{1 + 4L[A_T]}}{2L^2[A_T]} \quad (10)$$

Following the same approach, we obtained expressions for $[A_1]$ and $\langle s_{20,w} \rangle$ for two additional indefinite assembly cases. In the first case, which has been considered previously (*15*), it is assumed that successive additions of a monomer to the growing aggregate become increasingly more favorable (weak positive cooperativity) according to the equation $L_n = L(n - 1)/n$. In this model, L_n does not grow without bound (which would obviously result in a divergent infinite series) but rather approaches L as n approaches ∞. The expression for the total monomer concentration is given by (*15*)

$$[A_T] = \frac{[A_1]}{(1 - L[A_1])} \quad (11)$$

which can be solved for $[A_1]$:

$$[A_1] = \frac{[A_T]}{(1 + L[A_T])} \quad (12)$$

The resulting expression for $\langle s_{20,w} \rangle$ is

$$\langle s_{20,w} \rangle = \frac{s_1 \sum_{n=1}^{Q} n^m L^{n-1} [A_1]^n}{[A_T]} \quad (13)$$

In eq 13, the resulting species fraction distributions are shifted to higher values, which necessitates inclusion of a larger number of terms (higher value of Q) in the infinite summation. However, in our hands, and for our particular data set, we were unable to ensure convergence during NLLS analysis, even upon setting Q equal to 2000. Values of Q that were larger than 2000 resulted in the failure of the Scientist fitting software.

In the second case, it is assumed that successive additions of the monomer to the growing aggregate become less favorable, according to the equation $L_n = Ln/(n-1)$, where again L_n approaches L as n approaches ∞, and $n > 1$. We are unaware of this case being treated previously in the literature and therefore briefly describe the derivation of the expression for $[A_T]$ here. Upon substitution of the equation $L_n = Ln/(n-1)$ into eq 1, we obtain

$$[A_T] = [A_1](1 + 2^2x + 3^2x^2 + 4^2x^3 + ...) \quad (14)$$

where $x \equiv L[A_1]$. The infinite sum in eq 14 can be expressed as the derivative of another infinite sum, for which a closed form solution is known (*15*):

$$\frac{d}{dx}(x + 2x^2 + 3x^3 + 4x^4 + ...) = 1 + 2^2x + 3^2x^2 + 4^2x^3 + ... \quad (15)$$

Because $x + 2x^2 + 3x^2 + 4x^2 + ... = x/(1-x)^2$, we need to differentiate the right-hand side of this expression with respect to x and substitute the result into eq 14 to obtain our final expression for $[A_T]$:

$$[A_T] = \frac{[A_1](1 + L[A_1])}{(1 - L[A_1])^3} \quad (16)$$

Explicitly solving this expression for $[A_1]$ requires solving a cubic polynomial. Rather than do this, we simply solved this equation numerically in the NLLS analysis using SCIENTIST with the constraint that $0 < L[A_1] < 1$. The final expression for $\langle s_{20,w} \rangle$ is given by

$$\langle s_{20,w} \rangle = \frac{s_1 \sum_{n=1}^{Q} n^{2+m} L^{n-1} [A_1]^n}{[A_T]} \quad (17)$$

For this model, Q values greater than 100 were more than sufficient to ensure convergence of the summation.

For analysis of our sedimentation equilibrium data, we substituted eq 18

$$[A_1] = [A_{1,ref}] \times \exp\left[\frac{M(1 - \bar{v}\rho)\omega^2}{RT} \frac{r^2 - r_{ref}^2}{2}\right] \quad (18)$$

which describes the sedimentation equilibrium distribution of an L4-22K monomer, into either eq 4, 11, or 16, depending upon which indefinite assembly scheme was to be evaluated. This allows us to express the total L4-22K monomer concentration, $[A_T]$, as a function of both the free monomer concentration, $[A_1]$, and the radial position within the cell, r. In eq 18, r is the radial position relative to the center of rotation of the rotor, M is the monomer molecular mass, $\bar{v}$ is the partial specific volume of the protein, ρ is the buffer density, ω is the rotor speed, R is the gas constant, and T is the absolute temperature. $[A_{1,ref}]$ refers to the monomer concentration at a reference radial position, r_{ref}.

The dependence of the weight average stoichiometry of the L4-22K protein on the total monomer concentration, $[A_T]$, was calculated for the isodesmic model using (*20*)

$$\langle N_{wt} \rangle = \frac{d(\ln[A_T])}{d(\ln[A_1])} = \frac{1 + L[A_1]}{1 - L[A_1]} \quad (19)$$

where $\langle N_{wt} \rangle$ is given by $\langle M_{wt} \rangle / M_1$, where $\langle M_{wt} \rangle$ is the weight average molecular mass, and M_1 is the molecular mass of the L4-22K monomer (21924 Da). To quantify the spread of the calculated species distributions as a function of $[A_T]$, we also calculated the concentration dependence of the standard deviation of the species distributions, according to

$$\sigma_{wt} = \sqrt{\langle N_{wt}^2 \rangle - \langle N_{wt} \rangle^2} \quad (20)$$

Using the same strategy that was employed in ref *20* to derive the expression for $\langle N_{wt} \rangle$ for the isodesmic model (eq 19), we obtain the following expression for $\langle N_{wt}^2 \rangle$:

$$\langle N_{wt}^2 \rangle = \frac{1}{[A_T]}\left[\frac{d^2[A_T]}{d(\ln[A_1])^2}\right] = \frac{1 + 4L[A_1] + L^2[A_1]^2}{(1 - L[A_1])^2} \quad (21)$$

Combining eqs 19−21 yields

$$\sigma_{wt} = \frac{\sqrt{2L[A_1]}}{1 - L[A_1]} \quad (22)$$

where the subscript wt indicates that the standard deviation was calculated for a weight average probability distribution (as opposed to a number average, etc.).

Hydrodynamic Calculations. The frictional coefficient ratio, f/f_0, was calculated using (*21*)

$$\frac{f}{f_0} = \left[\frac{M^2(1 - \bar{v}\rho)^3}{162\pi^2 (s_{20,w})^3 \eta^3 N_{av}^2 (\bar{v} + \delta \bar{v}^0_{H_2O})}\right]^{1/3} \quad (23)$$

where M is the molecular mass of the protein species of interest, η is the viscosity of pure water at 20 °C, N_{av} is Avogadro's number, ρ is the density of pure water at 20 °C, $\bar{v}$ is the partial specific volume of the L4-22K protein at 20 °C, $\bar{v}^0_{H_2O}$ is the partial specific volume of pure water at 20 °C, and δ is the protein hydration, in grams of water bound per gram of macromolecule. For the L4-22K protein, δ was calculated on the basis of its primary sequence, using SEDNTERP, according to the method of Kuntz (*22*), which yielded a value of 0.4552 g/g.

The predicted $s_{20,w}$ for an unfolded, random coil polypeptide chain was calculated using the following equation (*21*, *23*):

$$s_{20,w} = 0.296M^{0.457} \quad (24)$$

where M (kilodaltons) is the molecular mass of the polypeptide chain.

Calculation of Uncertainties in NLLS Parameters. All reported NLLS uncertainties correspond to 68.3% confidence intervals calculated manually using the F statistics approach (*24*).

RESULTS

Sedimentation Velocity Studies of the Ad L4-22K Protein. Purified recombinant L4-22K was dialyzed into buffer P. The protein was then diluted to the desired concentration and subjected to a sedimentation velocity experiment. These experiments were conducted at 30K rpm, at 25 °C, and were monitored

by absorbance using either 230 or 280 nm. The results of a representative experiment are shown in Figure 1A. We analyzed the primary data using SEDFIT according to the ls-g*(s) approach (*25*, *26*) and show the results in Figure 1C. This analysis clearly shows the protein is highly aggregated and exists in solution as a broad distribution of aggregated structures. On the basis of this, we also investigated the effect of loading concentration on the aggregation state of the protein. The highest loading concentration we were able to achieve after dialysis was 6.8 μM, and the lowest concentration we were able to detect accurately by absorbance at 230 nm was 0.74 μM. The resulting ls-g*(s) concentration dependencies obtained over this concentration range are shown in Figure 1C, and all of these data were collected over a 2 day period in which each concentration originated from the identical, dialyzed stock concentration of 6.8 μM, which was kept at 4 °C. Inspection of these traces shows that as the total loading concentration is decreased, the peak positions in the ls-g(s) traces shift toward lower values, indicating a mass action aggregation process. To quantify this effect, the weight average sedimentation coefficient, $\langle s_{20,w} \rangle$, was calculated for each loading concentration, and the results are plotted in Figure 1D. Even over this relatively narrow loading concentration range, we see that $\langle s_{20,w} \rangle$ decreases from ~25 S at 6.8 μM to ~12 S at 0.74 μM. Control experiments show that $\langle s_{20,w} \rangle$ is independent of the amount of time (from 2 to 5 h) the sample is incubated at 25 °C prior to initiation of the sedimentation velocity run.

To investigate the molecular nature of these aggregates, we looked at their stability in the presence of increasing urea concentrations by sedimentation velocity. The L4-22K protein was dialyzed into buffer P, in the presence of the indicated urea concentration. Next, the protein was diluted to 3 μM and subjected to a sedimentation velocity experiment. The primary data were analyzed using SEDFIT (*13*), according to the *c*(*s*) approach, and the results are shown in Figure 2A. In the absence of added urea, the *c*(*s*) traces show no apparent species sedimenting below ~10 S. However, as the urea concentration is increased, a peak appears at ~2.2 S, and its area increases with an increase in urea concentration. Furthermore, the peak position of the aggregated material decreases from ~17.5 S in the absence of urea to ~12.5 S in the presence of 3 M urea. The dependence of $\langle s_{20,w} \rangle$ on urea concentration is shown in Figure 2B and confirms that an increasing urea concentration disrupts these higher-order aggregates. The simultaneous decrease in the apparent peak position of the higher-order aggregates with the increase in the area of the ~2.2 S peak suggests that the increasing urea concentration is freeing the smallest polymerizing unit from the aggregates.

Figure 2C shows the peak position of the apparent ~2.2 S species plotted as a function of added urea concentration, which shows the $s_{20,w}$ value of this apparent species is independent of urea concentration. The upper dashed line is the expected $s_{20,w}$ value for an L4-22K monomeric sphere (calculated by setting f/f_0 to 1 in eq 23 and solving for $s_{20,w}$). The lower dashed line is the expected $s_{20,w}$ for the fully unfolded protein (eq 24). This result indicates that at these intermediate urea concentrations, the observed peak corresponds to a folded L4-22K structure. Averaging the peak values shown in Figure 2B yields 2.16 ± 0.04 S. If we assume this peak corresponds to a monomer, we calculate a frictional ratio, f/f_0, for this species of 1.18 ± 0.02. If we assume this peak corresponds to an L4-22K dimer, we calculate an f/f_0 of 1.88 ± 0.03. An f/f_0 of 1.88 corresponds to an axial ratio for

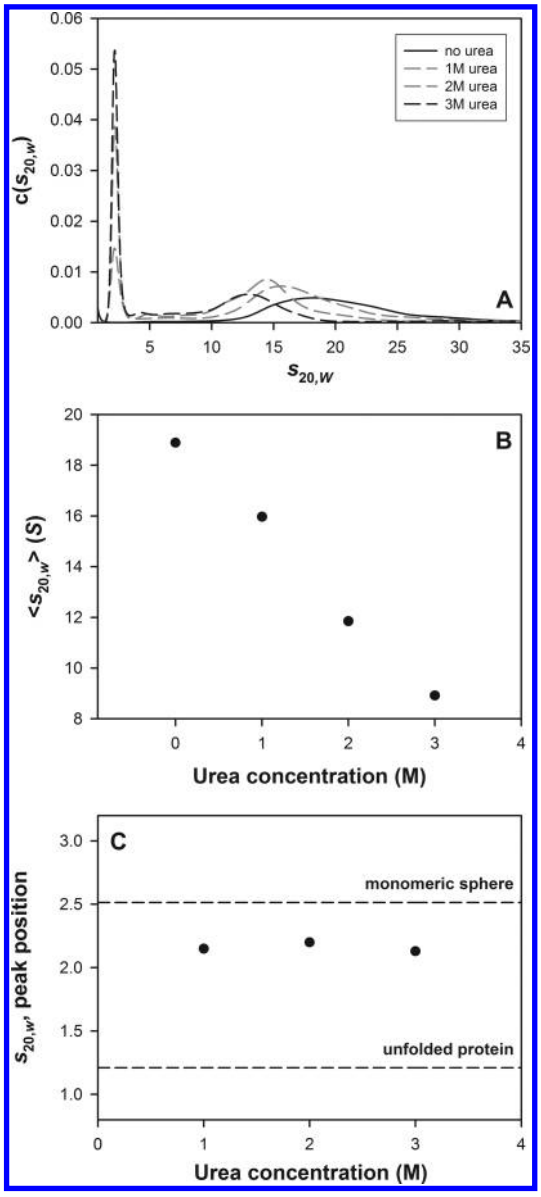

FIGURE 2: Sedimentation velocity analysis of the L4-22K aggregates as a function of urea concentration. (A) The purified L4-22K protein was dialyzed into buffer P along with the indicated urea concentration and then diluted to 3 μM for analysis by sedimentation velocity. The data were collected at 280 nm and 30K rpm and analyzed according to the *c*(*s*) method using SEDFIT. (B) Weight average sedimentation coefficients $\langle s_{20,w} \rangle$ from the distributions in panel A plotted as a function of urea concentration. (C) Peak positions of the ~2.2 S apparent species observed in panel A plotted vs urea concentration. The average value for these peak positions is 2.16 ± 0.04 S. The dashed lines correspond to the expected $s_{20,w}$ values for a spherical L4-22K monomer or a completely unfolded polypeptide of identical molecular weight.

a prolate ellipsoid of revolution of ~17. On the basis of the molecular mass of a presumed L4-22K dimer, its partial specific volume, and its estimated degree of hydration, the width of this prolate ellipsoid (at its widest point) is calculated to be ~2.1 nm, which is smaller than two α-helices placed side by side (2 × 1.2 nm = 2.4 nm). Because of the extreme asymmetry implied by these calculations, we suggest this peak most likely corresponds to an L4-22K monomer.

Using the largest $\langle s_{20,w} \rangle$ (~25 S) we observed in our sedimentation velocity experiments, we calculate an absolute lower limit for an average aggregate stoichiometry of 31 monomers, assuming

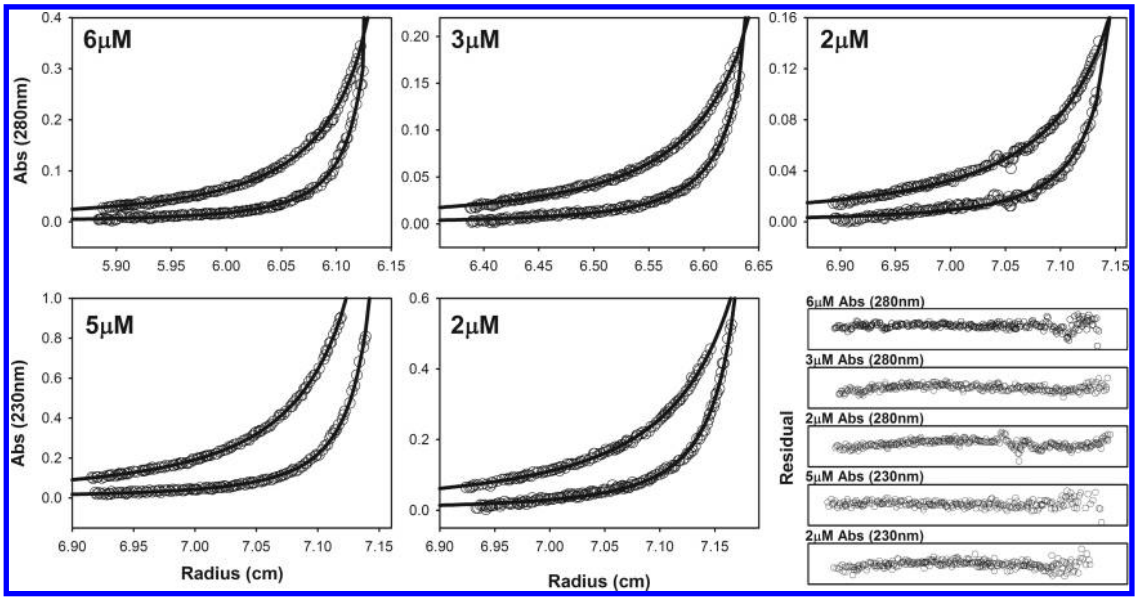

FIGURE 3: Sedimentation equilibrium analysis of L4-22K aggregation. Sedimentation equilibrium experiments were performed at the indicated loading concentrations of L4-22K in buffer P, at 25 °C. The rotor speeds used were 4.5K and 7.2K rpm. The data were collected at 230 and 280 nm and analyzed by NLLS according to an indefinite, isodesmic assembly model using eqs 4 and 18. The smooth curves are the results of the NLLS analysis of the data. The best fit value of L from this analysis was 112 (102, 124) μM^{-1}.

the aggregate adopts a spherical shape. On the basis of the large value of this lower limit, we consider that L4-22K may be undergoing indefinite, reversible self-association. To explore this possibility, we analyzed our $\langle s_{20,w} \rangle$ data using the isodesmic indefinite assembly model (eqs 9 and 10), following the method described by Na et al. (*16*). The smooth curve in Figure 1D is the result of an NLLS analysis of these data assuming (1) the smallest polymerizing unit is the L4-22K monomer, (2) each addition of monomer to the growing aggregate can be described by an identical equilibrium constant, L, and (3) the sedimentation coefficients for each stoichiometry will scale according to the equation $s_n = s_1 n^m$. In the NLLS analysis, we fixed s_1 and m at 2.16 ± 0.04 S and $^2/_3$, respectively, while allowing only L to float. The resulting best fit value for L was 69 (64, 74) μM^{-1}. We note that fixing m at $^2/_3$ is equivalent to assuming only that each aggregate has the same frictional ratio, not that each aggregate necessarily has a spherical shape (*18*).

We have also attempted to fit our data to an indefinite assembly model that assumes each successive addition of monomer occurs with a decreasing equilibrium constant, according to the prescription $L_n = (Ln)/(n - 1)$ (eqs 16 and 17). This model can be considered a test case to see if introducing a small amount of negative cooperativity might improve agreement between the model and the experimental data. This model could not fit the data when m was fixed at $^2/_3$. However, when both L and m were allowed to float, this model could fit the data as well as the isodesmic model and returned best fit values of 24 (17, 50) μM^{-1} and 0.95 (0.87, 1.0), respectively. Even though we cannot differentiate between these two models on the basis of fitting statistics alone, we can exclude the decreasing equilibrium constant model on the basis that an m value of 0.95 is physically unrealistic; e.g., using an m of 0.95, we calculate for our hypothetical 31-mer an absolute upper limit for the frictional ratio of 0.45, which would indicate our 31-mer would have to sediment faster than a compact sphere, which is physically impossible. In fact, any m value of greater than $^2/_3$ is physically unrealistic by the same argument. Furthermore, we note that, for the purposes of calculating the fitting uncertainties (using F statistics), an m value of >1 is mathematically impossible, because the infinite summation in eq 17 is guaranteed to diverge when $m > 1$. Therefore, we fixed the upper mathematical limit for m at 1 and used this value to determine the lower limit for L.

We also attempted to fit our data to a model that assumed successive equilibrium constants increased according to the equation $L_n = L(n - 1)/n$ (i.e., weak positive cooperativity), which has been discussed previously (*15*). However, in our hands, we were unable to include a sufficient number of terms in the summation (see Experimental Procedures) to ensure accurate calculation of $\langle s_{20,w} \rangle$ in the fitting routine. However, we were able to assess the feasibility of this model when analyzing our sedimentation equilibrium data, which are discussed below.

Sedimentation Equilibrium Studies of the Ad L4-22K Protein. To further study the size and distribution of the L4-22K aggregates and to determine if they exist in solution in a true thermodynamic equilibrium, we conducted a series of sedimentation equilibrium experiments, under conditions identical to those used for the sedimentation velocity studies. Dialyzed L4-22K protein was diluted to the desired concentration and then sedimented to equilibrium at 4.5K and 7.2K rpm. All concentrations reached sedimentation equilibrium within ~30 h (loading volumes of 110 μL), which was determined via comparison of multiple scans taken 3 h apart using Winmatch.

We initially analyzed these data by fitting to a single, ideal species model (*27*) to calculate an approximate average molecular mass of the sample. At a loading concentration of 2 μM (units of total monomer concentration), the apparent cell average molecular mass suggested a stoichiometry of roughly 46 monomers, while at a loading concentration of 6 μM, this value increased to roughly 60 monomers (not shown), consistent with a mass action aggregation process. Next, we used the indefinite isodesmic aggregation model (eqs 4 and 18) to globally analyze each channel of data. This analysis included six independent loading concentrations (6, 3, and 2 μM, collected at 280 nm, and 5 and 2 μM, collected at 230 nm), each collected at two rotor speeds (4.5K and 7.2K rpm). The results are shown in Figure 3 and show that the isodesmic model can accurately describe our sedimentation

equilibrium data. The best fit value we obtained for L was 112 (102, 124) μM^{-1}, which is similar to the value we obtained from analysis of the $\langle s_{20,w} \rangle$ data according to the isodesmic model, which was $L = 69$ (64, 74) μM^{-1} when fixing m at $^2/_3$. While these values are similar, they are outside of the fitting uncertainties. Therefore, this may indicate that m is less than $^2/_3$. If we reanalyze the $\langle s_{20,w} \rangle$ data by fixing L at 112 μM^{-1} and allowing only m to float, we obtain an m of 0.622 (0.615, 0.629), consistent with this possibility.

Next, we attempted to analyze our data using the increasing equilibrium constant model (eqs 11 and 18). The resulting standard deviation of the fit (SD_{fit}) was ±0.00482 OD, compared to ±0.00401 OD for the isodesmic assembly model, which corresponds to sum of squared residual (SSR) values of 0.0314 and 0.0217, respectively. To address which model fits the experimental data better, we used the F statistics approach described by Johnson and Straume (*24*). In this approach, we are asking if we can be 95% confident that the variances that result from these two models are significantly different. The calculated SSR_1/SSR_2 ratio is 1.094 (with the total degrees of freedom for each fit equal to 1352, and the probability equal to 0.95). The ratio of the two SSR values derived from the NLLS analyses is 1.447 (0.0314/0.0217). Because 1.447 > 1.094, we can be >95% confident these two models produce variances that are significantly different. From this analysis, we conclude that the isodesmic assembly model fits our sedimentation equilibrium data significantly better than the increasing equilibrium constant model.

We also analyzed our sedimentation equilibrium data using the decreasing equilibrium constant model (eqs 16 and 18), where L_n decreases according to the equation $L_n = (Ln)/(n - 1)$. In this case, the F statistics analysis described above indicated that the resulting SD_{fit} (±0.00392 OD, with an SSR of 0.0207) was statistically indistinguishable, at 95% confidence, from that of the isodesmic model; however, we can rule this model out on the basis of (1) the inability of this model to describe the $\langle s_{20,w} \rangle$ data with a physically realistic value of m and (2) the fact that even if we ignore the previous point, the resulting best fit value of L from analysis of the sedimentation equilibrium data was 1790 (1580, 2000) μM^{-1}, which does not agree with the value obtained from the velocity experiments, which was 24 (17, 50) μM^{-1}. In contrast to the decreasing equilibrium constant model, when we use the isodesmic model, we obtain good agreement for the L values estimated from both the sedimentation velocity data [$L = 69$ (64, 74) μM^{-1}] and the sedimentation equilibrium data [$L = 112$ (102, 124) μM^{-1}]. Finally, we also tried the decreasing equilibrium constant model presented in ref *28* to analyze our sedimentation equilibrium data. In this model, L_n decreases with increasing aggregate stoichiometry according to the equation $L_n = L/n$, which occurs at a much more rapid rate than the $L_n = (Ln)/(n - 1)$ model we used above. The $L_n = L/n$ model was unable to fit our data, likely indicating that if any negative cooperativity does indeed exist upon continued assembly, it is relatively small.

In conclusion, when we attempt to analyze both our sedimentation velocity and equilibrium data using indefinite assembly models that incorporate weak negative or weak positive cooperativity, they fail. On the other hand, both the sedimentation velocity and equilibrium data can be analyzed adequately according to the indefinite, isodesmic assembly model. On this basis, we conclude the isodesmic model is appropriate for a quantitative description of the self-association properties of the L4-22K protein.

DISCUSSION

In this work, we have characterized the self-association properties of a recombinant version of the Ad L4-22K protein. Quantitative knowledge of the self-association properties of L4-22K is required to understand how it assembles, along with the Ad IVa2 protein, onto the packaging sequence within the Ad genome, ultimately signaling the onset of viral DNA encapsidation. We find that highly purified L4-22K self-assembles reversibly into a broad distribution of higher-order aggregation states. We were able to accurately model this process using an isodesmic, indefinite aggregation model and were able to determine the equilibrium constant, L, associated with this process, which is 112 (102, 124) μM^{-1}, which corresponds to a dissociation constant of 8.9 (8.1, 9.8) nM (0.1 M NaCl, pH 7, 25 °C). This equilibrium constant refers to the affinity of addition of a single L4-22K monomer to either another monomer or any other aggregate. To visualize how these species are distributed in solution, we have plotted the fraction of each species (calculated using eq 8) calculated over a range of total L4-22K concentrations and have also plotted the average and standard deviation of the aggregate stoichiometries (Figure 4). These calculations show that at total L4-22K concentrations (in total monomer units) lower than ~10 nM, the largest fraction of species present in solution is the monomer. Total monomer concentrations of ≥100 nM are required before the majority of the protein exists in an aggregated state in solution. Furthermore, as the total monomer concentration is increased, the breadth of the species distribution is increased.

What are the biological implications of these observations? An interesting mathematical consequence of the isodesmic aggregation model arises from the requirement that the product $L[A_1]$ must be less than one, to ensure the infinite species summation converges. This requirement places an absolute upper limit on the value of the free monomer concentration, $[A_1]$, that can exist in solution, which is $1/L$ (*17*). This upper limit can also be obtained by taking the limit of eq 10 when $[A_T]$ goes to infinity, which yields $1/L$, as expected. This implies that a consequence of a biological macromolecule taking part in an indefinite aggregation process is that the free monomer concentration will not increase without bound upon increasing the total concentration of that macromolecule in the cell. This is in contrast to what is observed for macromolecular assembly processes that terminate at a finite stoichiometry, e.g., a monomer−dimer system or a monomer−dimer−tetramer system, etc. In these cases, it is straightforward to show that as the total monomer concentration is increased without bound, the free monomer concentration also increases without bound; even though the fraction of the monomeric species decreases with an increasing total monomer concentration, the absolute concentration of the free monomer species continues to increase as the total monomer concentration increases. Therefore, an immediate consequence of an indefinite aggregation process is that the free monomer concentration can be "buffered" through equilibration with its higher-order aggregates. To illustrate this, we have plotted the free L4-22K monomer concentration, $[A_1]$, as a function of the total monomer concentration and show the results in Figure 4D. We can see from this plot that for total concentrations of L4-22K greater than ~250 nM, $[A_1]$ is essentially invariant. This provides the cell with a mechanism for fixing the free L4-22K monomer concentration at a bit less than $1/L$, by simply expressing L4-22K to levels sufficient to approach closely this asymptote. Because it is the

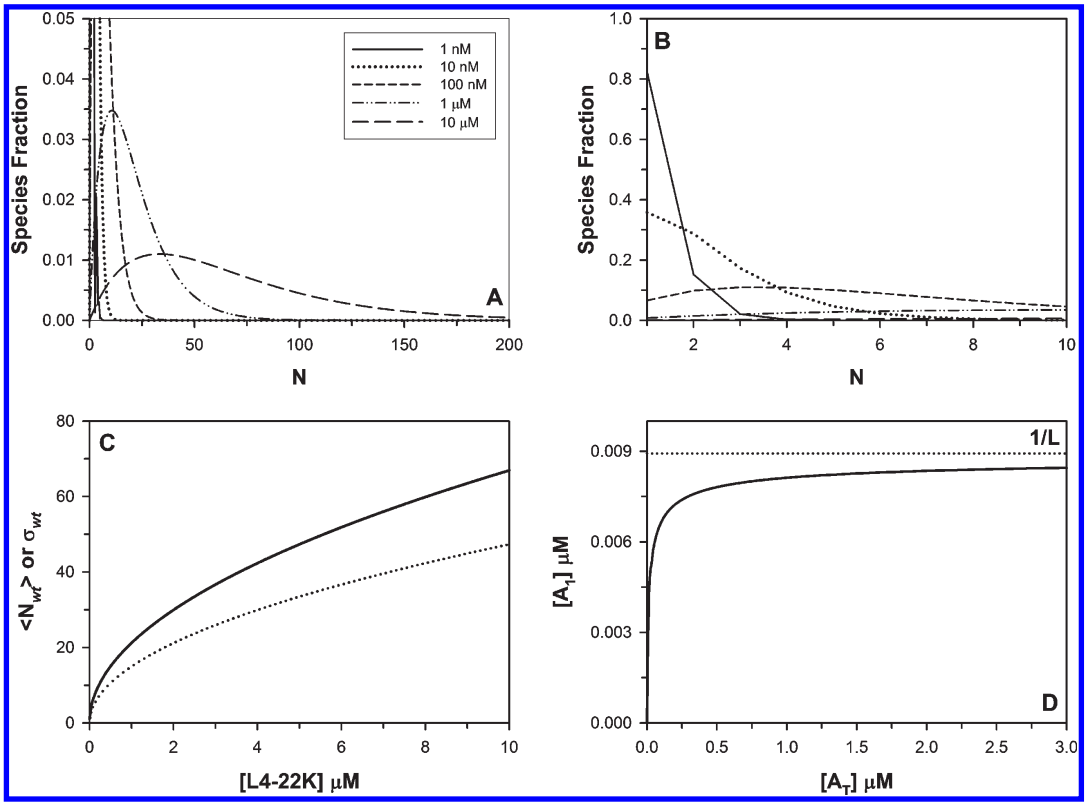

FIGURE 4: (A) L4-22K species distributions were calculated as a function of total L4-22K monomer concentration using an L value of 112 μM^{-1} according to eq 8. (B) The distributions in panel A are shown over a narrower range of N. (C) Weight average stoichiometry (—; $\langle N_{wt} \rangle$) and weight average standard deviation (– – –; σ_{wt}) of the L4-22K species distributions calculated as a function of total L4-22K monomer concentration using eqs 19 and 22, respectively. (D) Free monomer concentration, $[A_1]$, calculated as a function of the total L4-22K concentration, $[A_T]$. $[A_1]$ cannot exceed $1/L$.

free ligand concentration that uniquely determines the fractional saturation of a ligand binding site in the cell (*29*), this implies cellular control mechanisms may involve linkage with indefinite assembly processes.

Another physical mechanism that can control free ligand (or free monomer) concentrations is the linkage of ligand binding with phase transitions, called polyphasic linkage (*30*). For example, in sickle-cell hemoglobin, the deoxgenated protein undergoes a cooperative phase transition to a solid, gel-like state (*30*, *31*). Thus, at sufficiently low oxygen concentrations and at sufficiently high protein concentrations, the cooperative assembly of sickle-cell hemoglobin will result in a constant free hemoglobin concentration in solution.

In Ad, the IVa2 and L4-22K proteins both assemble at the packaging sequence, and this assembly is cooperative and is required for the initiation of viral DNA encapsidation (*4*−*6*) in the nucleus of the cell. While IVa2 can bind to the PS in the absence of L4-22K, sequence specific binding of L4-22K to the PS is absolutely dependent upon the presence of the IVa2 protein (*4*). A consequence of the observation that L4-22K assembles indefinitely is that it is not possible to drive saturation of the PS by only increasing the total L4-22K concentration while keeping the IVa2 concentration fixed. If L4-22K did not assemble indefinitely, then saturation of the PS or the DE sequence, at any fixed IVa2 concentration, could be achieved by simply increasing the cellular L4-22K concentration. Considering the likelihood that saturation of the PS will be tightly regulated to prevent inappropriate signaling of the viral DNA encapsidation process and that the IVa2 protein does not self-associate (*21*), this implies that at saturating L4-22K concentrations (to ensure $[A_1] \sim 1/L$), the IVa2 concentration will control occupancy of the PS. Further experiments are required to test the biological relevance of these in vitro observations.

ACKNOWLEDGMENT

We thank Drs. David Bain and Aaron Lucius for insightful discussions.

REFERENCES

1. Ljungman, P. (2004) Treatment of adenovirus infections in the immunocompromised host. *Eur. J. Clin. Microbiol. Infect. Dis. 23*, 583–588.
2. McConnell, M. J., and Imperiale, M. J. (2004) Biology of adenovirus and its use as a vector for gene therapy. *Hum. Gene Ther. 15*, 1022–1033.
3. Ostapchuk, P., and Hearing, P. (2005) Control of adenovirus packaging. *J. Cell. Biochem. 96*, 25–35.
4. Ewing, S. G., Byrd, S. A., Christensen, J. B., Tyler, R. E., and Imperiale, M. J. (2007) Ternary complex formation on the adenovirus packaging sequence by the IVa2 and L4 22-kilodalton proteins. *J. Virol. 81*, 12450–12457.
5. Ostapchuk, P., Anderson, M. E., Chandrasekhar, S., and Hearing, P. (2006) The L4 22-kilodalton protein plays a role in packaging of the adenovirus genome. *J. Virol. 80*, 6973–6981.
6. Tyler, R. E., Ewing, S. G., and Imperiale, M. J. (2007) Formation of a multiple protein complex on the adenovirus packaging sequence by the IVa2 protein. *J. Virol. 81*, 3447–3454.
7. Zhang, W., and Imperiale, M. J. (2003) Requirement of the adenovirus IVa2 protein for virus assembly. *J. Virol. 77*, 3586–3594.
8. Grable, M., and Hearing, P. (1990) Adenovirus type 5 packaging domain is composed of a repeated element that is functionally redundant. *J. Virol. 64*, 2047–2056.
9. Schmid, S. I., and Hearing, P. (1997) Bipartite structure and functional independence of adenovirus type 5 packaging elements. *J. Virol. 71*, 3375–3384.

10. Ali, H., LeRoy, G., Bridge, G., and Flint, S. J. (2007) The adenovirus L4 33-kilodalton protein binds to intragenic sequences of the major late promoter required for late phase-specific stimulation of transcription. *J. Virol. 81*, 1327–1338.
11. Lutz, P., and Kedinger, C. (1996) Properties of the adenovirus IVa2 gene product, an effector of late-phase-dependent activation of the major late promoter. *J. Virol. 70*, 1396–1405.
12. Tribouley, C., Lutz, P., Staub, A., and Kedinger, C. (1994) The product of the adenovirus intermediate gene IVa2 is a transcriptional activator of the major late promoter. *J. Virol. 68*, 4450–4457.
13. Schuck, P. (2003) On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal. Biochem. 320*, 104–124.
14. Ucci, J. W., and Cole, J. L. (2004) Global analysis of non-specific protein-nucleic interactions by sedimentation equilibrium. *Biophys. Chem. 108*, 127–140.
15. Martin, R. B. (1996) Comparisons of Indefinite Self-Association Models. *Chem. Rev. 96*, 3043–3064.
16. Na, G. C., and Timasheff, S. N. (1980) Stoichiometry of the vinblastine-induced self-association of calf brain tubulin. *Biochemistry 19*, 1347–1354.
17. Oosawa, F., and Kasai, M. (1962) A theory of linear and helical aggregations of macromolecules. *J. Mol. Biol. 4*, 10–21.
18. Sontag, C. A., Stafford, W. F., and Correia, J. J. (2004) A comparison of weight average and direct boundary fitting of sedimentation velocity data for indefinite polymerizing systems. *Biophys. Chem. 108*, 215–230.
19. Tang, L. H., Powell, D. R., Escott, B. M., and Adams, E. T., Jr. (1977) Analysis of various indefinite self-associations. *Biophys. Chem. 7*, 121–139.
20. Cantor, C. R., and Schimmel, P. R. (1980) Biophysical chemistry, Vol. II, W. H. Freeman, San Francisco.
21. Yang, T. C., Yang, Q., and Maluf, N. K. (2009) Interaction of the adenoviral IVa2 protein with a truncated viral DNA packaging sequence. *Biophys. Chem. 140*, 78–90.
22. Kuntz, I. D. (1971) Hydration of macromolecules. IV. Polypeptide conformation in frozen solutions. *J. Am. Chem. Soc. 93*, 516–518.
23. Uversky, V. N. (2002) What does it mean to be natively unfolded? *Eur. J. Biochem. 269*, 2–12.
24. Johnson, M. L., and Straume, M. (1994) Comments on the analysis of sedimentation equilibrium experiments. In Modern Analytical Ultracentrifugation, pp 37−65, Birkhauser, Boston.
25. Schuck, P., and Rossmanith, P. (2000) Determination of the sedimentation coefficient distribution by least-squares boundary modeling. *Biopolymers 54*, 328–341.
26. Stafford, W. F., III (1994) Boundary analysis in sedimentation velocity experiments. *Methods Enzymol. 240*, 478–501.
27. Laue, T. M. (1995) Sedimentation equilibrium as thermodynamic tool. *Methods Enzymol. 259*, 427–452.
28. Morris, M., and Ralston, G. B. (1989) A thermodynamic model for the self-association of human spectrin. *Biochemistry 28*, 8561–8567.
29. Lohman, T. M., and Bujalowski, W. (1991) Thermodynamic methods for model-independent determination of equilibrium binding isotherms for protein-DNA interactions: Spectroscopic approaches to monitor binding. *Methods Enzymol. 208*, 258–290.
30. Wyman, J., and Gill, S. J. (1980) Ligand-linked phase changes in a biological system: Applications to sickle cell hemoglobin. *Proc. Natl. Acad. Sci. U.S.A. 77*, 5239–5242.
31. Gill, S. J., Benedict, R. C., Fall, L., Spokane, R., and Wyman, J. (1979) Oxygen binding to sickle cell hemoglobin. *J. Mol. Biol. 130*, 175–189.