Assignment 3 Answers – Protein Molecular Weight Determination by Sedimentation Equilibrium

The single domain protein 'X' was analyzed in a sedimentation equilibrium experiment. Details of the experiment are:

rotor speed = 40,000 rpm
Temperature = 20 °C
Buffer density = 1.026 g/mL

Determine the molecular weight of the protein from the experimental data given in AUCdat.opj.

The equilibrium distribution of the protein concentration versus radial position in the sample cell was determined at three different initial (i.e. uniform) protein concentrations (0.2, 0.5 and 1.0 mg/mL). It is advisable to globally fit the data to obtain the molecular weight estimate. You may assume that the protein did not engage in a monomer-dimer equilibrium, and that the protein solutions behave ideally.

The data can be fit to a single exponential function, because the protein is not engaged in a monomer-dimer (or more generally an oligomerization) process that will alter its molecular weight. Also sufficiently dilute solutions of proteins often (but not always) exhibit ideal behavior, *i.e.* excluded volume and charge interactions among proteins do not significantly influence the contribution concentration distribution. The Boltzmann distribution in Berg,

$$C(r) = C(r_0)e^{m'\omega^2\left(r_0^2 - r^2\right)/2kT} \tag{1}$$

(equation 5.14 on page 72) ignores such inter-particle effects. The underlying assumption is that the particles move independently of one another. Inter-particle interactions, such as excluded volume and electrostatic effects, are concentration-dependent and lead to deviations from equation 5.14. The analysis will tell us that inter-particle effects are small (undetectable) in these data.

The steps to determining the molecular weight are:

(1) Plotting concentration versus the radius, the radius-squared or the radius-squared/2.
(2) Excluding the spurious data near the solution meniscus from the fit, *e.g.* data where radius < 24.
(3) Fit the concentration *versus* radius-squared data to an exponential function in Origin:

$$y = C_0 + C_1\exp((x\text{-}rsq_0)/rsq)) \tag{2}$$

to retrieve the coefficient in the exponent, *rsq*. C_0 and C_1 were treated as independent variables in the three data sets (0.1, 0.5 and 1 mg/mL protein 'X'), and were allowed to float during the fit. rsq_0 was fixed at 48.1 and shared among the three sets of data. *rsq* was also shared, but allowed to vary.

(4) A comparison of equations (1) and (2) indicates that

$$rsq = 2k_BT/m'\omega^2 \tag{3}$$

Substitution of $m(1 - \upsilon\rho)$ for m' in equation (3) and solving for m yields:

$$m = \frac{2k_BT}{rsq(1-\bar{\upsilon}\rho)\omega^2} \tag{4}$$

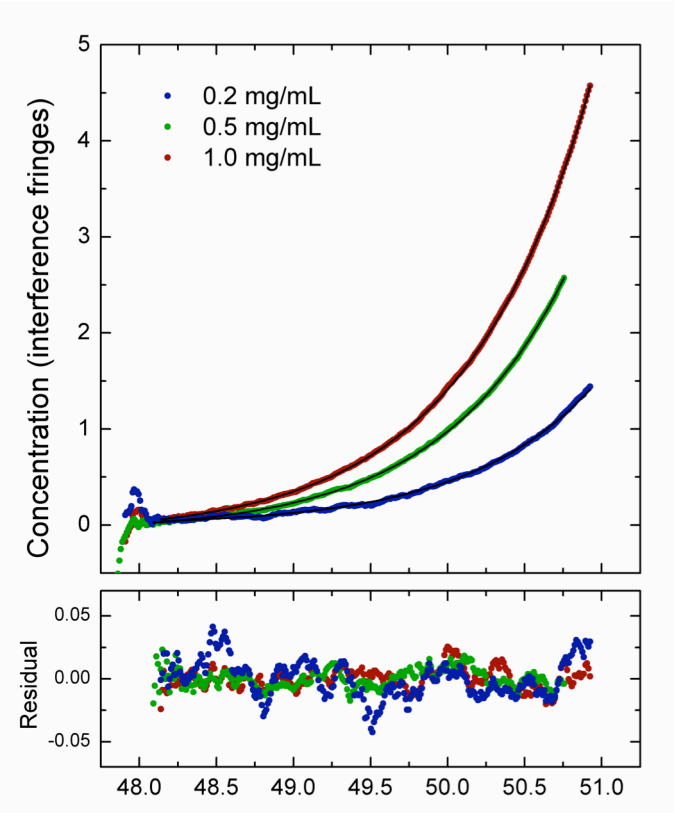
where m will be in grams as long as cgs units are used for k_B, υ, ρ and ω. The molecular weight (MW) in grams per mole is simply N_0m, where N_0 is Avogadro's number. **A molar mass of 13,500** is determined for the protein based on the parameters given in the problem, the values of the constants and the estimate of *rsq* returned from the global fit:

$N_0k_B = R = 8.315 \times 10^7$ g cm^2/s^2/mol/deg
$(1 - \upsilon\rho) = (1 - (0.73\ cm^3/g)(1.026\ g/cm^3) = 0.25102$
$\omega^2 = (2\pi*40{,}000/60)^2 = 1.755 \times 10^7\ s^{-2}$
$rsq = 0.819$ +/- 0.001 cm^2

$$MW = \frac{2RT}{rsq(1-\bar{\upsilon}\rho)\omega^2} = \frac{2\left(8.315x10^7\ g\ cm^2 s^2 mol^{-1}K^{-1}\right)(293\ K)}{(0.819\ cm^2)(0.251)\left(1.755x10^7\ s^{-2}\right)} = 13{,}500\ g\ mol^{-1}$$

The appearance of the fit and residuals plotted below the frame below the fit indicate that the single-mass model adequately explains the data. There is no justification for invoking either nonideal behavior, or polydispersity in the sample molecular weight.

The uncertainty in the molecular weight is given by the uncertainty in the estimate for ***rsq*** *(0.81951 ± 0.00108). This leads to a range in* ***rsq*** *of 0.81843 to 0.82059, which corresponds to a molecular weight range of 13,515 to 13,480, or ~ 13,498 ± 18.*



The data were fit to $y = C0 + C1*\exp((x-rsq0)/rsq)$;

where:
y is the protein concentration (in fringes) as a function of radial position
***C*0** allows for a signal 'background', which offsets the data from zero. (adjusted, not shared)
C*1** is the protein concentration at ***rsq0. (adjusted, not shared)
rsq0 is the radius near the meniscus. (fixed)
rsq is the exponential growth constant that contains information about the molecular weight (shared)
x is the independent variable of radius squared (cm^2)

$y = C0 + C1*\exp((x-rsq0)/rsq)$)

Initial C (mg/mL)	***First Point***	***Last Point***	***C0***	***C1***	***rsq0***	***Rsq***
1.0	25	311	-0.09603	0.14853	48.1	0.81951 ± 0.00108
0.5	25	299	-0.07863	0.10372	48.1	0.81951 ± 0.00108
0.2	25	311	-0.0077	0.0452	48.1	0.81951 ± 0.00108

*fixed at stated value

When the data are fit to a model with fewer variables, *e.g. C*0 is fixed equal to 0, *C*1 is not shared and adjusted, and *rsq* is shared among the data sets and adjusted, the following parameters are returned in the fit:

$y = C1*\exp((x-rsq0)/rsq)$)

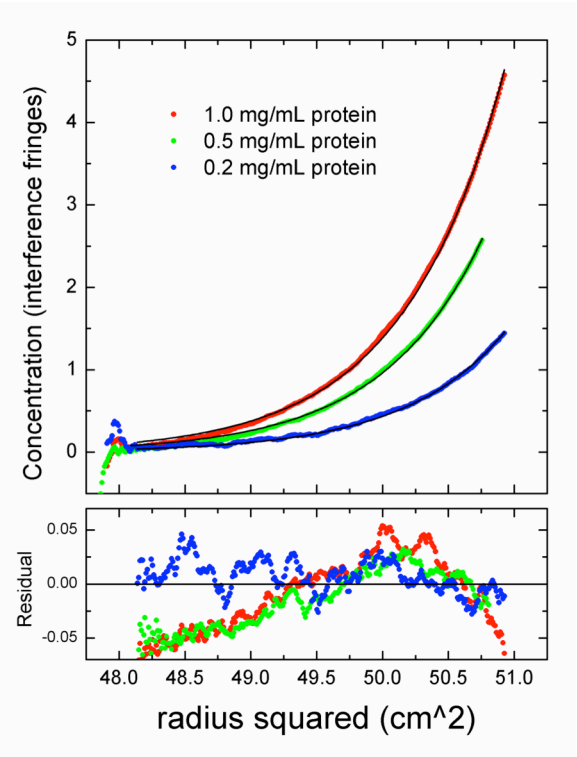
Initial C (mg/mL)	***First Point***	***Last Point***	***C0****	***C1***	***rsq0****	***Rsq***
1.0	25	311	0	0.05493	47.5	0.76552 ± 0.0013
0.5	25	299	0	0.03824	47.5	0.76552 ± 0.0013
0.2	25	311	0	0.01722	47.5	0.76552 ± 0.0013

*fixed at stated value

The molecular weight estimated with these conditions was 14,450 g/mol, which is similar in value, although the residuals in the fit indicate systematic deviation (See plot next page.)

Summary

- *The data are well-described a model for sedimentation equilibrium involving single domain protein with a molecular weight of 13,500 Da, which exhibits ideal behavior.*



Fit of Data according to $y = C1*\exp((x - rsq0)/rsq)$, where *rsq0* is fixed at 47.5 cm² in all sets, and *rsq* is shared, but adjusted.