Biological Chemistry Laboratory
Biology 3515/Chemistry 3515
Spring 2018

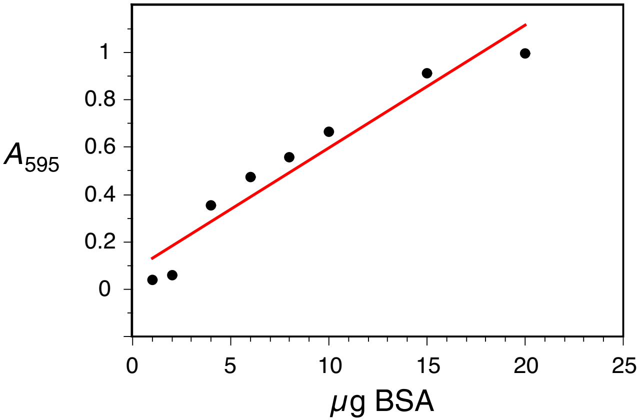
Lecture 7:

Curve Fitting, Part II, and Overlapping Spectra

30 January 2018

University of Utah
goldenberg@biology.utah.edu

A Linear Least-squares Fit to Bradford Calibration Data



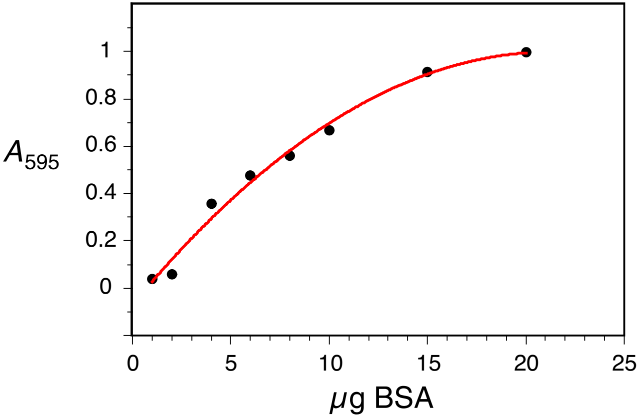
- The estimated parameters for $y = mx + b$:

  $m = 0.052 \pm 0.006$

  $b = 0.08 \pm 0.06$

  $R^2 = 0.93$

A 2nd-order Polynomial Least-squares Fit to Bradford Calibration Data



- For 2nd-order polynomial fit:

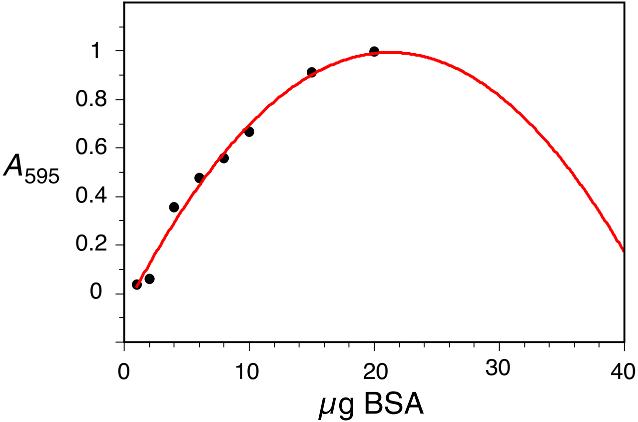
$$\chi^2 = 0.01$$
$$R^2 = 0.988$$

- For linear fit:

$$\chi^2 = 0.062$$
$$R^2 = 0.93$$

- Increasing the number of parameters almost always improves the fit!
- Is it justified here?

Does the Fit Function Make Sense Physically?

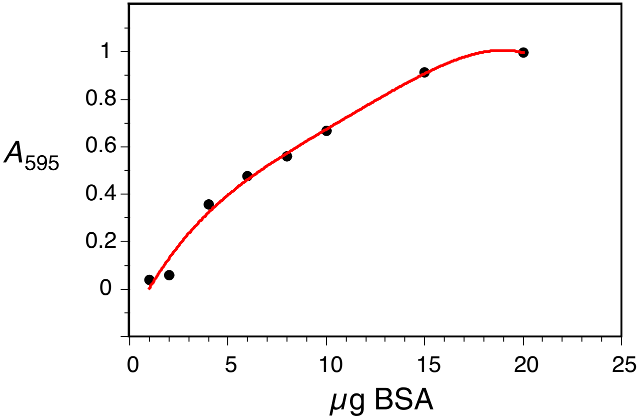


- Should the absorbance decrease as the amount of BSA increases beyond 20 μg?

  Probably not!

- The function serves as a calibration curve over the range used to fit it, but not beyond.

A 4^{th}-order Polynomial Least-squares Fit to Bradford Calibration Data



- For 4^{th}-order polynomial fit:

$$\chi^2 = 0.01$$
$$R^2 = 0.991$$

- For 2^{nd}-order polynomial fit:

$$\chi^2 = 0.012$$
$$R^2 = 0.988$$

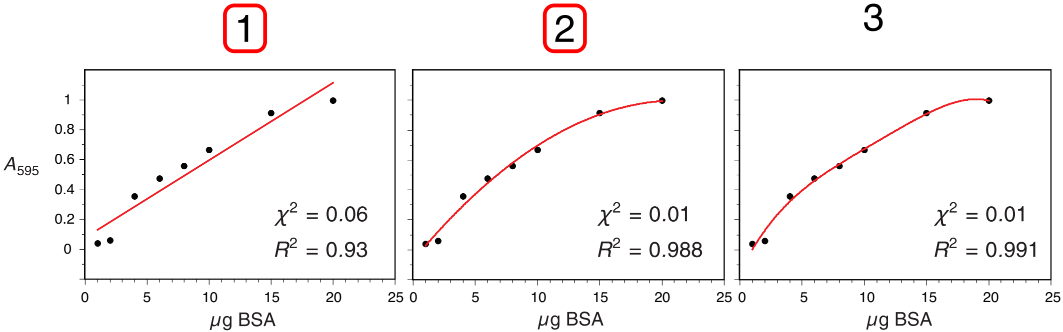
- For linear fit:

$$\chi^2 = 0.062$$
$$R^2 = 0.93$$

- Have we gone to far?

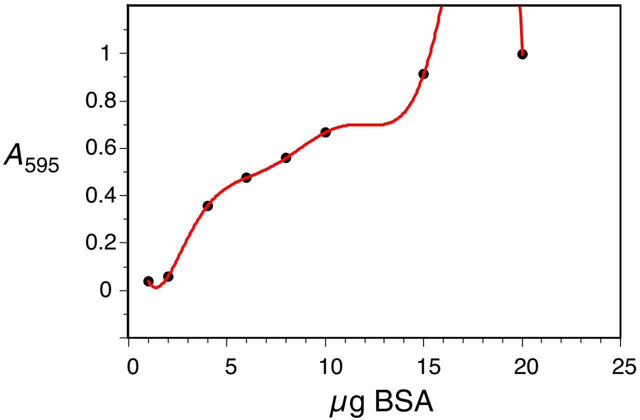
Clicker Question #1

Which is the most reasonable fit?



All answers count (for now)!

A 7th-order Polynomial Least-squares Fit to Bradford Calibration Data



- For 7th-order polynomial fit:

  $$\chi^2 = 0$$

  $$R^2 = 1$$

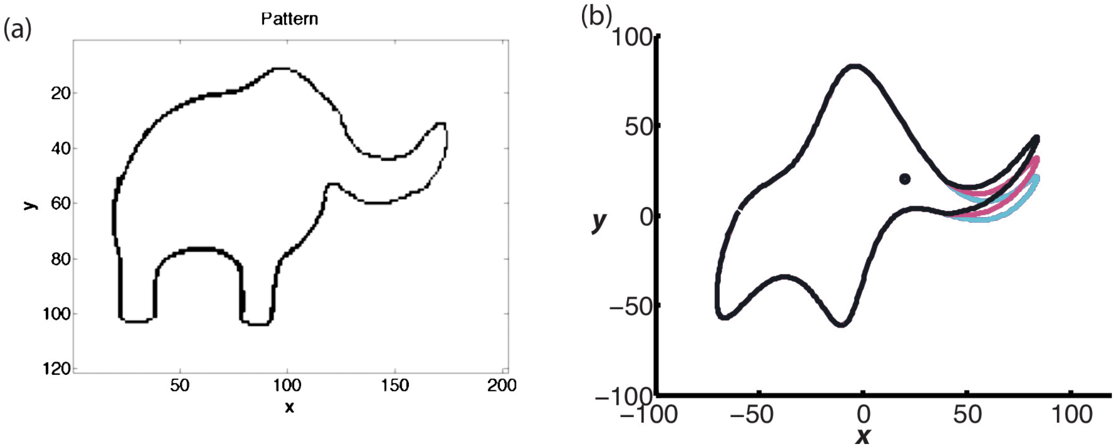
  A perfect fit!

  Or, perfectly absurd?

"With four parameters I can fit an elephant, and with five I can make him wiggle his trunk"

John von Neumann, according to Enrico Fermi, as quoted by Freeman Dyson. *Nature* (2004) **427**, 297

Fitting an Elephant



Mayer, J., Khairy, K. & Howard, J. (2010). Drawing an elephant with four complex parameters. *Am. J. Phys.*, 78, 648–649.
http://dx.doi.org/10.1119/1.3254017

Another Interesting Function

$$y = \frac{ax}{b + x}$$

- When $x \ll b$

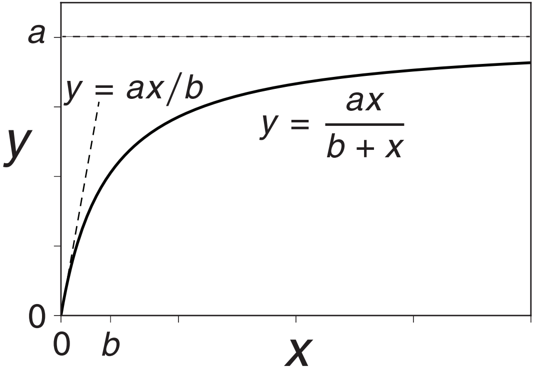
$$y = \frac{ax}{b + x} \approx \frac{ax}{b}$$

A line through the point $(0, 0)$, with slope a/b.

- When $x \gg b$

$$y = \frac{ax}{b + x} \approx \frac{ax}{x} = a$$

A constant, a.



"Linear" versus "Non-linear" Curve Fitting

- In the context of curve-fitting, a polynomial

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots + a_n x^n$$

  is said to be a "linear" function in the sense that y is a linear function of each of the fit parameters, a_i (even if it isn't linear with respect to x).

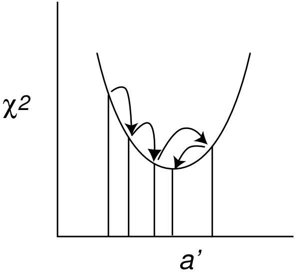
- Equations of this type can be fit to data relatively easily using equations like those shown for the straight line fit.
- The equation for a rectangular hyperbola:

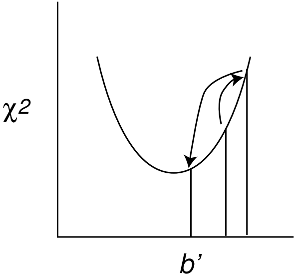
$$y = \frac{a \cdot x}{b + x}$$

  is *not* linear with respect to the parameter b.

- For non-linear equations, least-squares fitting usually must be done iteratively.

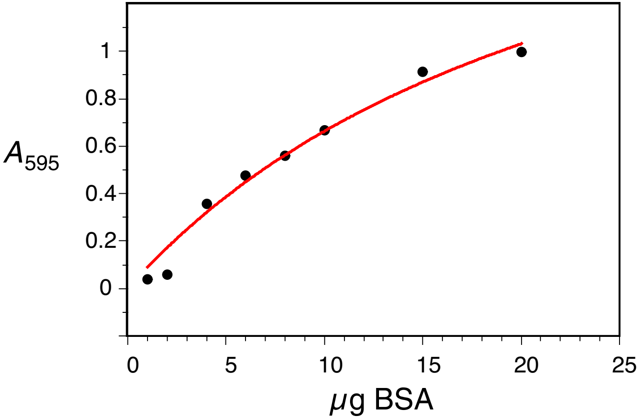
An Iterative Method to Minimize χ^2





1. Make initial estimates of parameters *a* and *b*
2. Calculate χ^2
3. Change the parameters a little bit and recalculate χ^2
4. If χ^2 decreases, change the parameters some more in the same direction, otherwise change the parameters in the opposite direction.
5. Repeat until χ^2 does not decrease further.

A Rectangular Hyperbola Fit to Bradford Calibration Data



- For fit to rectangular hyperbola:

  $$\chi^2 = 0.02$$
  $$R^2 = 0.977$$

  With only two parameters!

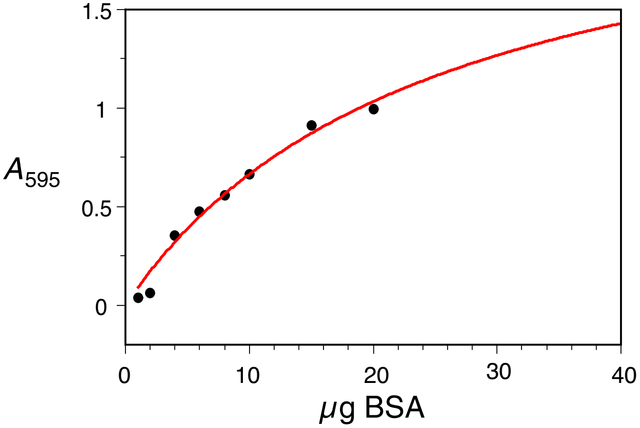
- For 2nd-order polynomial fit:

  $$\chi^2 = 0.01$$
  $$R^2 = 0.988$$

- For linear fit:

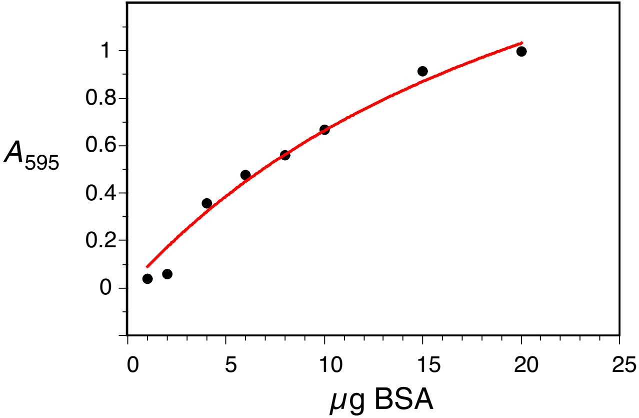
  $$\chi^2 = 0.062$$
  $$R^2 = 0.93$$

Does the Fit Function Make Sense Physically?



- Does the extrapolation look plausible?
- Is the curvature real?
- How could we find out?
- Why might the Bradford calibration curve have this shape?

A Rectangular Hyperbola Fit to Bradford Calibration Data



- Fit function:

$$y = \frac{ax}{b + x}$$

- Fit parameters:

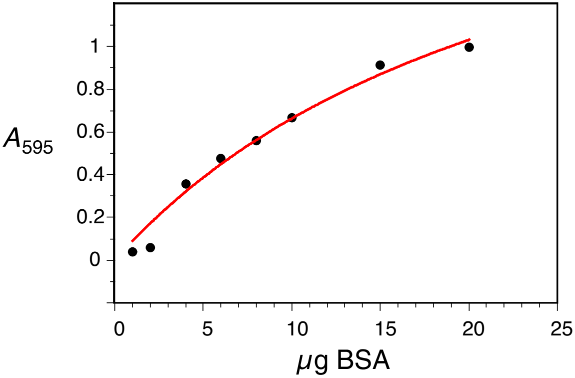
$$a = 2.32 \pm 0.53$$
$$b = 24.9 \pm 6.6$$

- What are the units for the parameters?

Clicker Question #2

What are the units for the parameter b?

$$y = \frac{ax}{b + x}$$

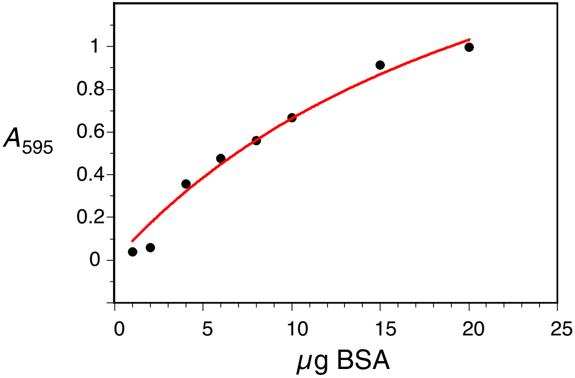


1. μg
2. None
3. μg^{-1}

Clicker Question #3

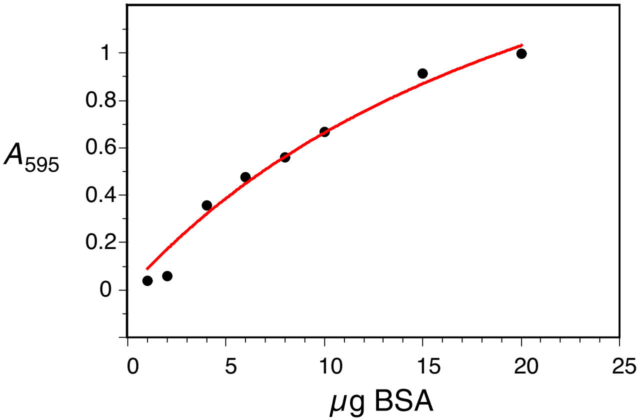
What are the units for the parameter a?

$$y = \frac{ax}{b + x}$$



1. μg
2. None
3. μg^{-1}

A Rectangular Hyperbola Fit to Bradford Calibration Data



- Fit function:

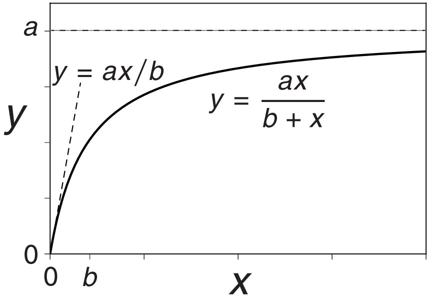
$$y = \frac{ax}{b + x}$$

- Fit parameters:

$$a = 2.32 \pm 0.53$$
$$b = 24.9 \pm 6.6$$

- Why are the uncertainties so large?

Why Are the Uncertainties So Large?



- To determine both a and b, we need data over a range that includes values that are less than b and values that are greater than b.
- Good data analysis requires good experimental design! (And, good data!)

- When x is small relative to b:

  $$y = \frac{ax}{b+x} \approx \frac{ax}{b}$$

  A line through the point $(0, 0)$, with slope a/b.

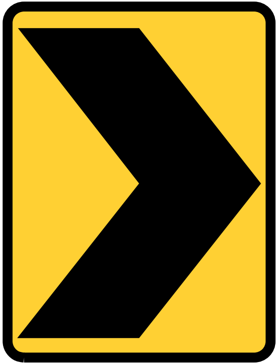
  If we only have data in this region, the slope, a/b, is well defined, but lots of pairs of a and b will fit the data well.

- When x is large relative to b:

  $$y = \frac{ax}{b+x} \approx \frac{ax}{x} = a$$

  If we only have data in this region, what will happen to our fit?

Warning!

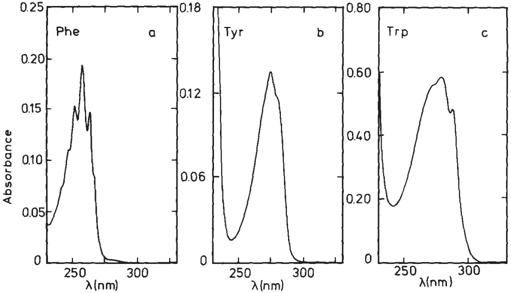


Direction Change

Back to Spectrophotometry

What if a Solution Contains Multiple Compounds that Absorb Light?

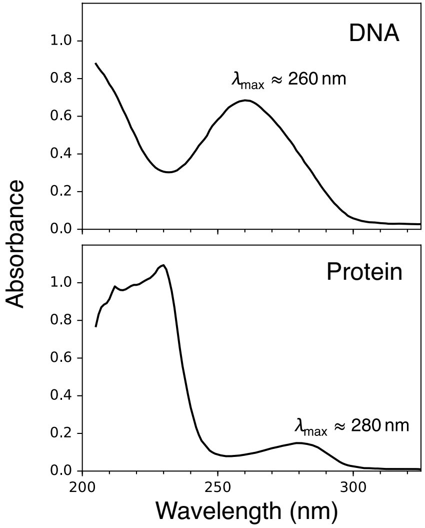
- Peaks in UV-visible absorption spectra are quite broad:



- Peaks from different compounds often overlap.
- Absorption at a given wavelength may contain contributions from multiple compounds.

Spectra from: Schmid, F. X. (1997). In *Protein Structure: A Practical Approach* (Creighton, T. E., ed.), pp. 261-297. IRL Press, Oxford. (A good introduction to optical spectroscopy of proteins)

UV Absorption Spectra of Proteins and DNA



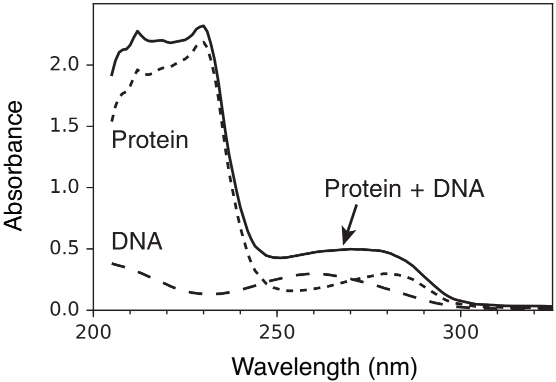
- DNA spectra do not depend much on sequence.
- Protein spectra do depend on amino acid composition, and a bit on three-dimensional structure.
- DNA and protein spectra, between 250 and 300 nm overlap extensively.
- Concentrations:

$$[\text{DNA}] \approx 0.03\,\text{mg/ml}$$

$$[\text{Protein}] \approx 0.16\,\text{mg/ml}$$

Spectra adapted from Spectrophotometry Handbook (2012) GE LifeScience

Spectra of DNA, Protein and a Mixture



- Absorbances of different components add.
- Assumes components don't interact.
- Can we interpret absorbance of mixtures?

Estimating Concentrations of Protein and DNA in a Mixture

- Between 250 and 300 nm

  For Protein: $\lambda_{max} \approx 280$ nm

  For DNA: $\lambda_{max} \approx 260$ nm

- At 260 nm (assuming 1-cm cuvette):

$$A_{260} = [\text{Protein}] \cdot \epsilon_{260}^{\text{Protein}} + [\text{NA}] \cdot \epsilon_{260}^{\text{NA}}$$

- At 280 nm:

$$A_{280} = [\text{Protein}] \cdot \epsilon_{280}^{\text{Protein}} + [\text{NA}] \cdot \epsilon_{280}^{\text{NA}}$$

- If all four extinction coefficients are known, and we measure A_{260} and A_{280}, we have two equations in two unknowns.

  Solve for [Protein] and [NA].

- What could go wrong?