Biological Chemistry Laboratory
Biology 3515/Chemistry 3515
Spring 2017

Lecture 7:

Quiz Questions and More on Curve Fitting

31 January 2017

University of Utah
goldenberg@biology.utah.edu

A Quiz Question from 2015

- The stated question:
  - 1.3 mg/mL protein solution
  - Molecular mass = 37,000 Da
  - $A_{280} = 0.93$ in a 1-cm cuvette

  Calculate the molar extinction coefficient. Be sure to express your answer with the correct units.

- Some “hidden” questions:
  - What is this question about?
  - What will the answer look like?

Clicker Question #1

What are the correct units for the answer?

1. M
2. $M \cdot cm$
3. $M^{-1}cm^{-1}$
4. $M \cdot cm^{-1}$

Another Part of the Quiz Question

- The stated question:

  It turns out that the sample you were given is contaminated with 0.01 mg/mL DNA. But, the original estimate of the protein concentration, 1.3 mg/mL is, in fact, correct.

  Explain how the contaminating DNA will affect your estimate of the extinction coefficient.

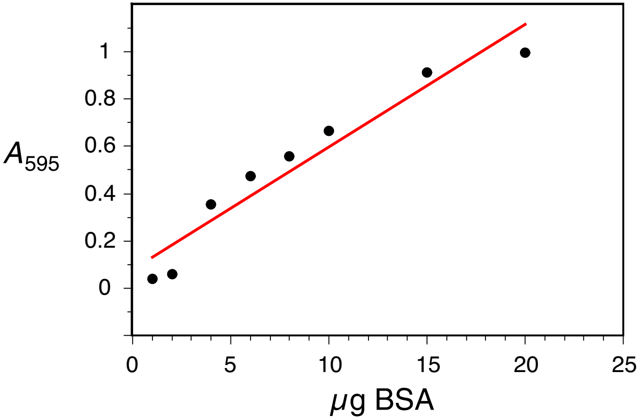
- What is this question about?

Clicker Question #2

How will the presence of contaminating DNA affect the estimated extinction coefficient of the protein?

1. Make the extinction coefficient too low.
2. Not affect the extinction coefficient.
3. Make the extinction coefficient too high.

A Linear Least-squares Fit to Bradford Calibration Data



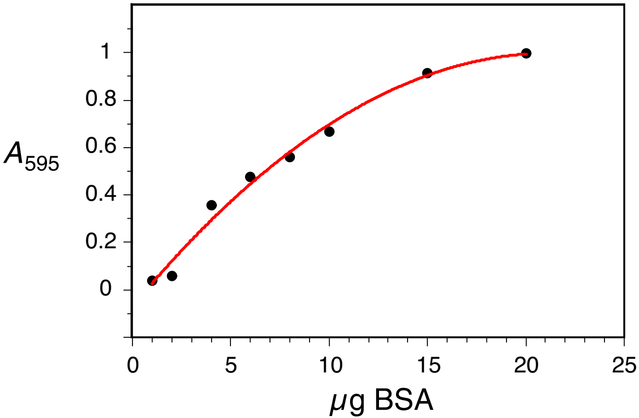
- The estimated parameters for $y = mx + b$:

  $m = 0.052 \pm 0.006$

  $b = 0.08 \pm 0.06$

  $R^2 = 0.93$

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



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

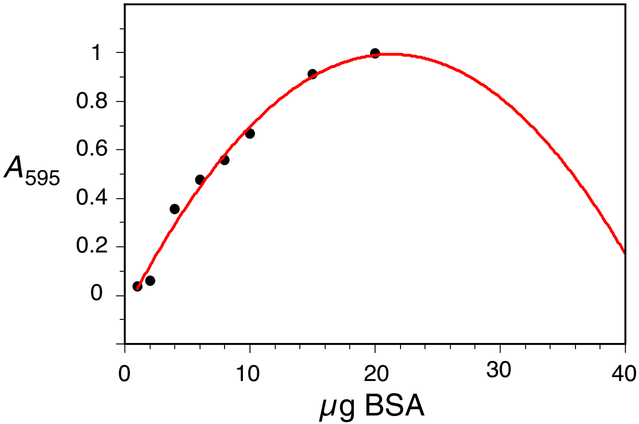
$$\chi^2 = 0.012$$
$$R^2 = 0.988$$

- For linear fit:

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

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

Does the Fit Function Make Sense Physically?

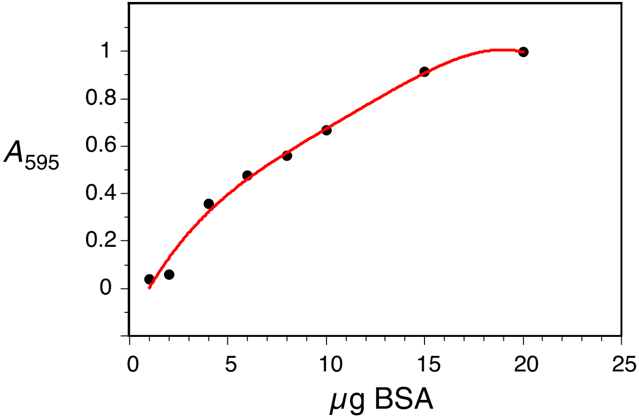


- Should the absorbance decrease as the amount of BSA increases beyond 20 μg?

  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:

  $$\chi^2 = 0.015$$
  $$R^2 = 0.991$$

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

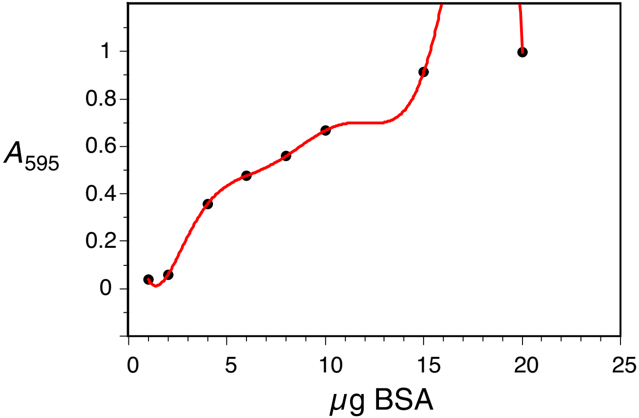
  $$\chi^2 = 0.012$$
  $$R^2 = 0.988$$

- For linear fit:

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

- Have we gone to far?

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



- For 7^{th}-order polynomial fit:

$$\chi^2 = 0$$
$$R^2 = 1$$

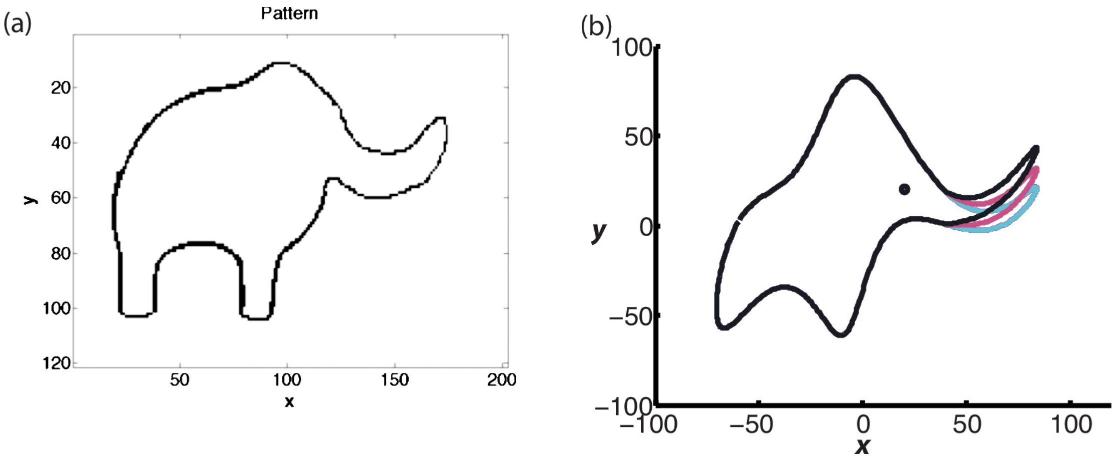
A perfect fit!

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 $x \ll b$

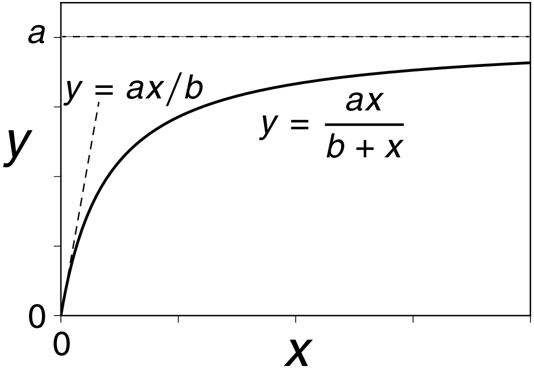
$$y = \frac{ax}{b + x} \approx \frac{ax}{b}$$

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

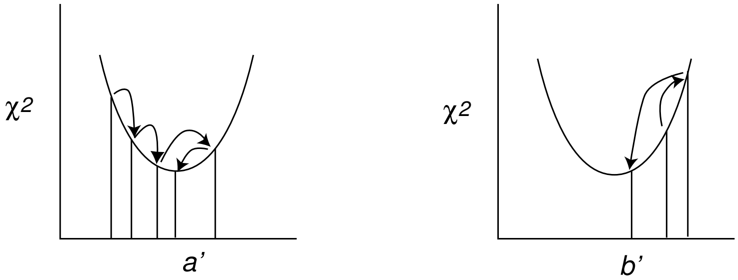
$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots + a_n x^n$$

  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:

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

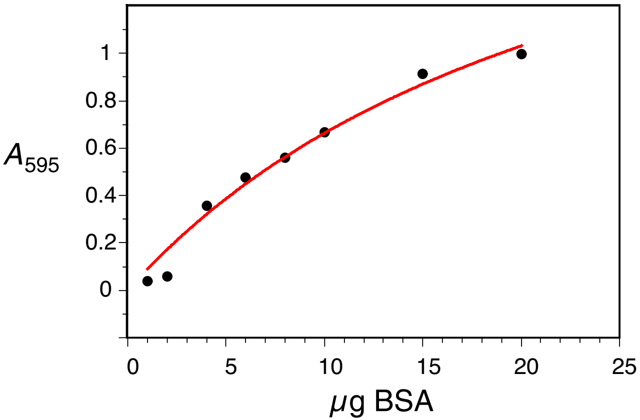
  is *not* linear with respect to the parameter b.
- For non-linear equations, least-squares fitting usually must be done iteratively.

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!

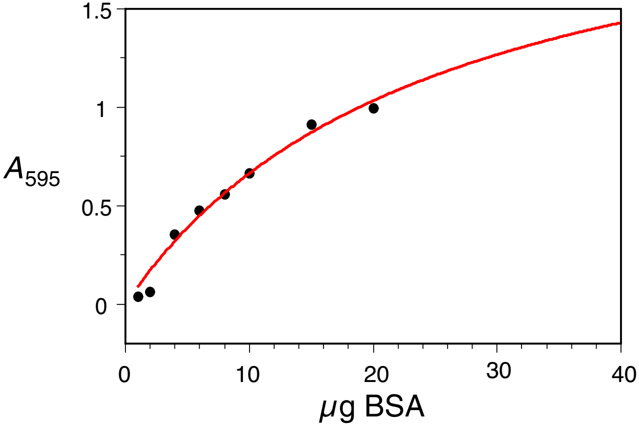
- For 2nd-order polynomial fit:

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

- For linear fit:

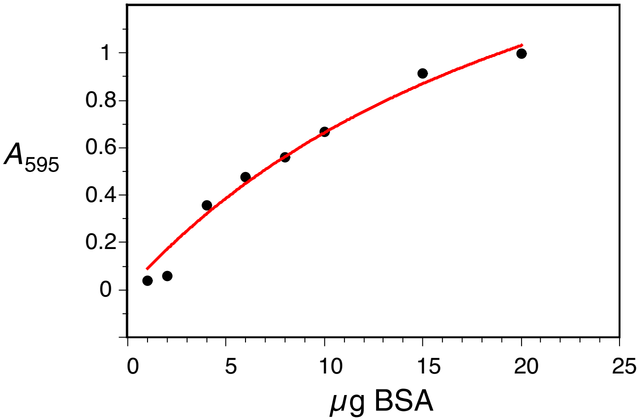
  $$\chi^2 = 0.062$$
  $$R^2 = 0.93$$

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:

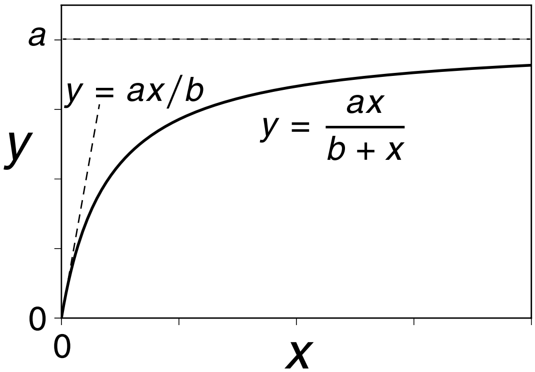
$$y = \frac{ax}{b + x}$$

- Fit parameters:

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

- What are the units for the parameters?
- Why are the uncertainties so large, relative to the parameter values?

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

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

  A constant, a.

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