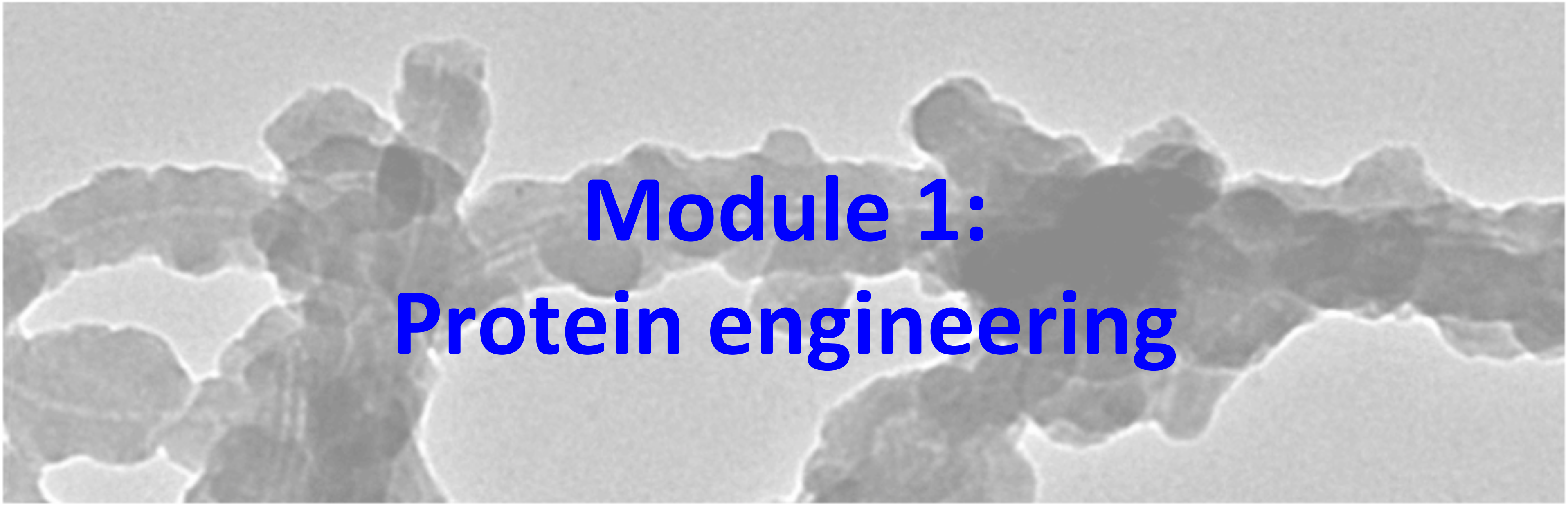


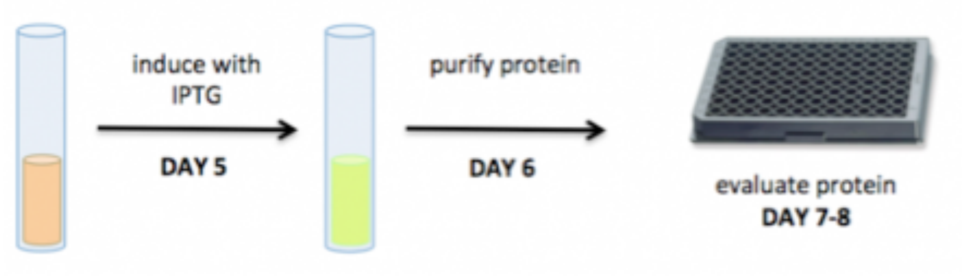
Module 1: Protein engineering

I. Binding analysis

II. MATLAB basics

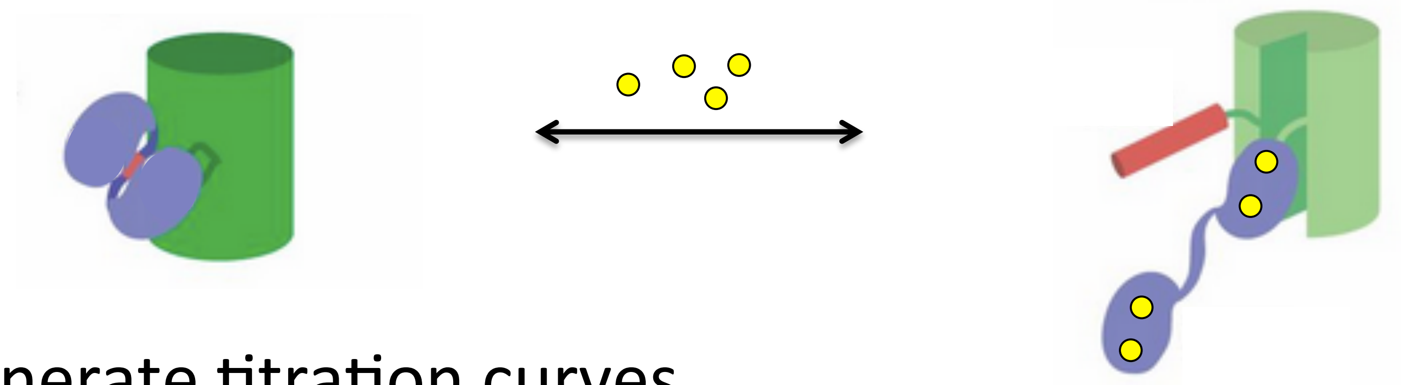
03/03/2016

How will we evaluate our mutant IPC?



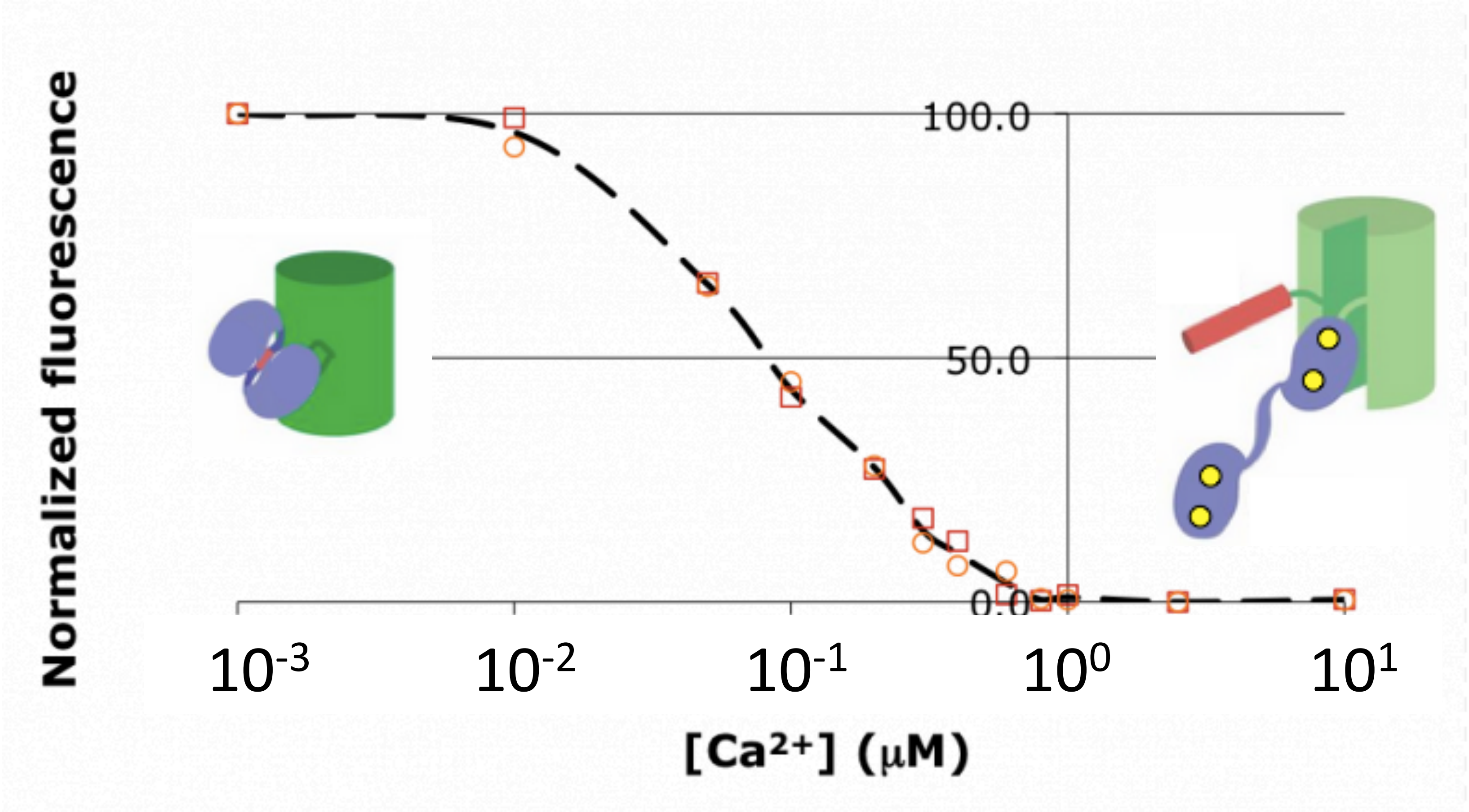
?

- Use solutions of known [Ca^{2+}] calcium concentration
- Measure binding = fluorescence signal



- Generate titration curves

How does your mutation alter IPC-Ca^{2+} binding?



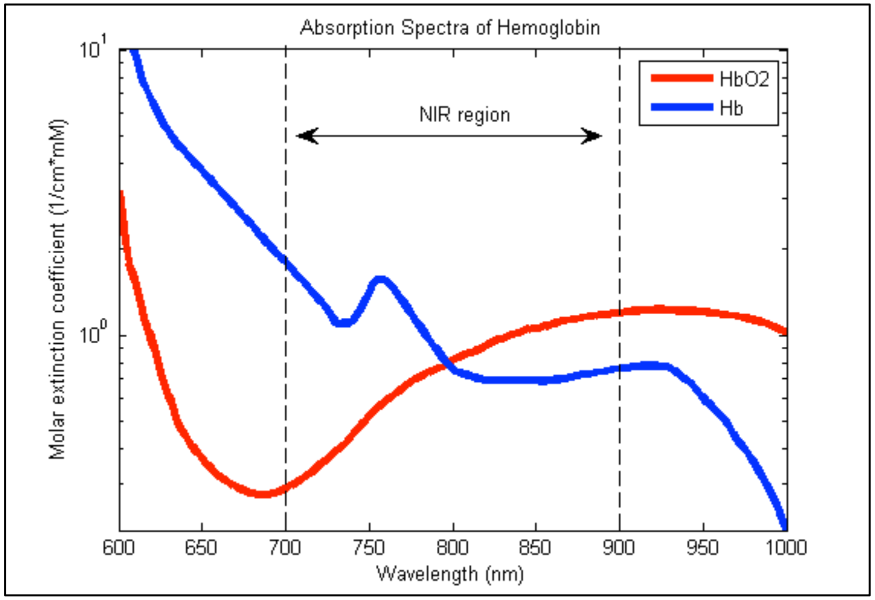
Binding may be quantified using methods other than fluorescence

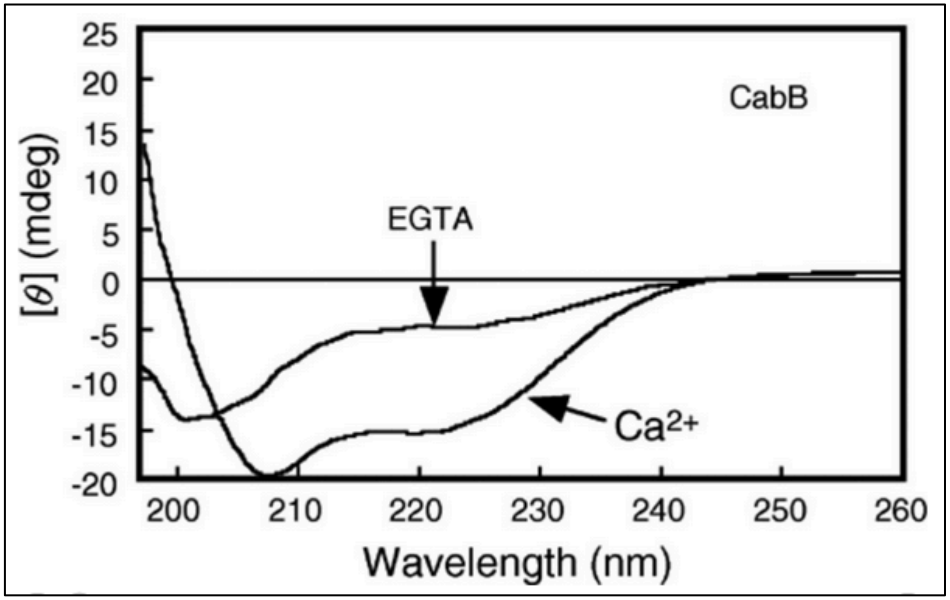
- absorbance spectroscopy

  e.g. hemoglobin binding to O_2

- circular dichroism

  e.g. Ca^{2+} binding to CabB



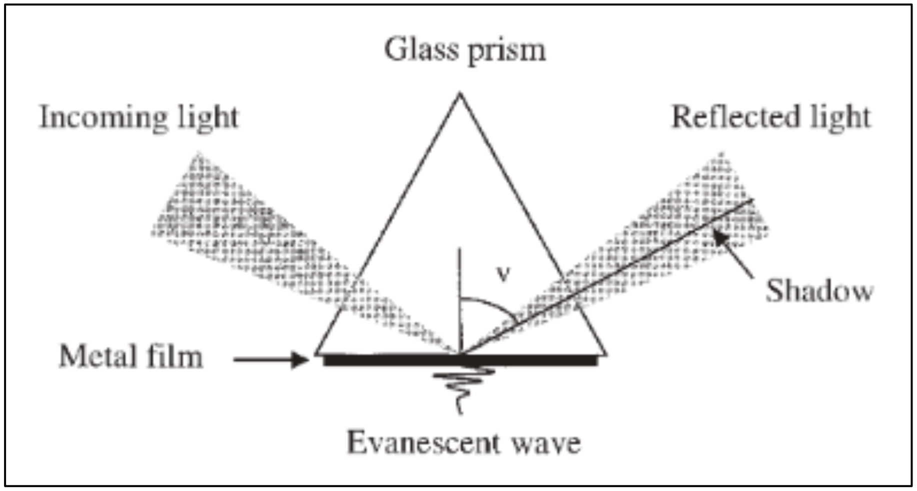


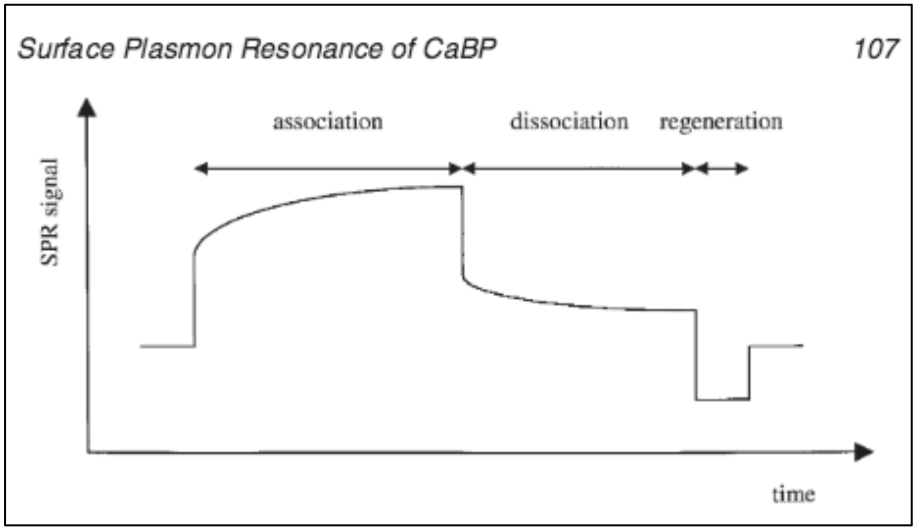
oxymetry.org

Yonekawa *et al*. (2005) *FEMS Microbio Lett* **244**: 315-321

- surface plasmon resonance

  e.g. Ca^{2+} binding to CaM

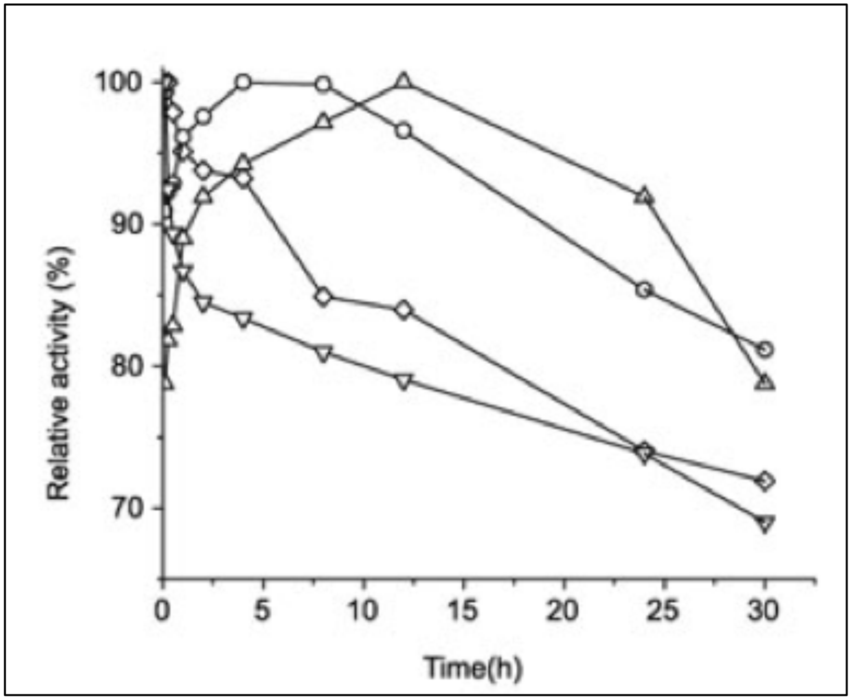




Julenius (2002) *Methods Mol Biol* **173**: 103-111

- enymatic activity

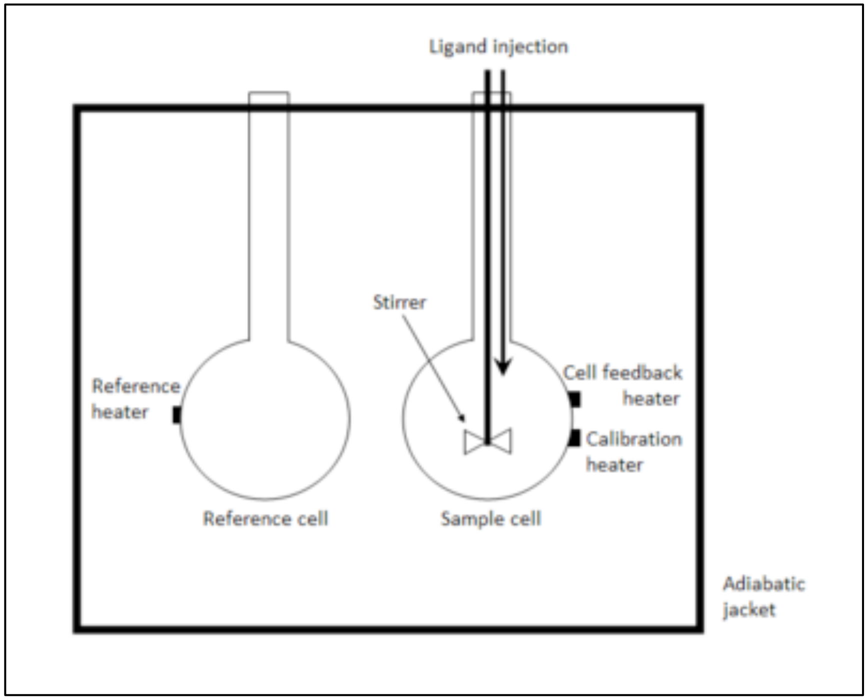
  e.g. [Ca^{2+}] effect on lipase

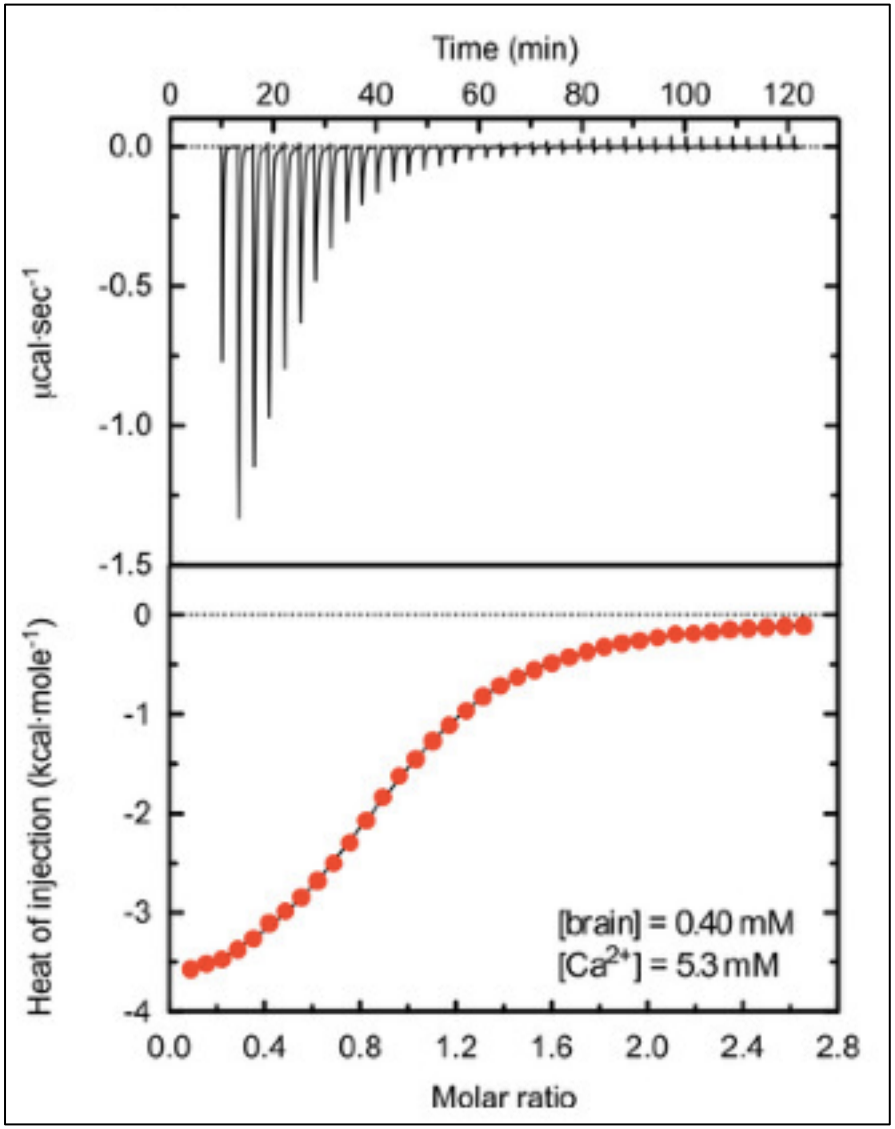


Barbosa *et al.* (2012) *Química Nova* **35**

- isothermal titration calorimetry

  e.g. Ca^{2+} binding to α-actinin

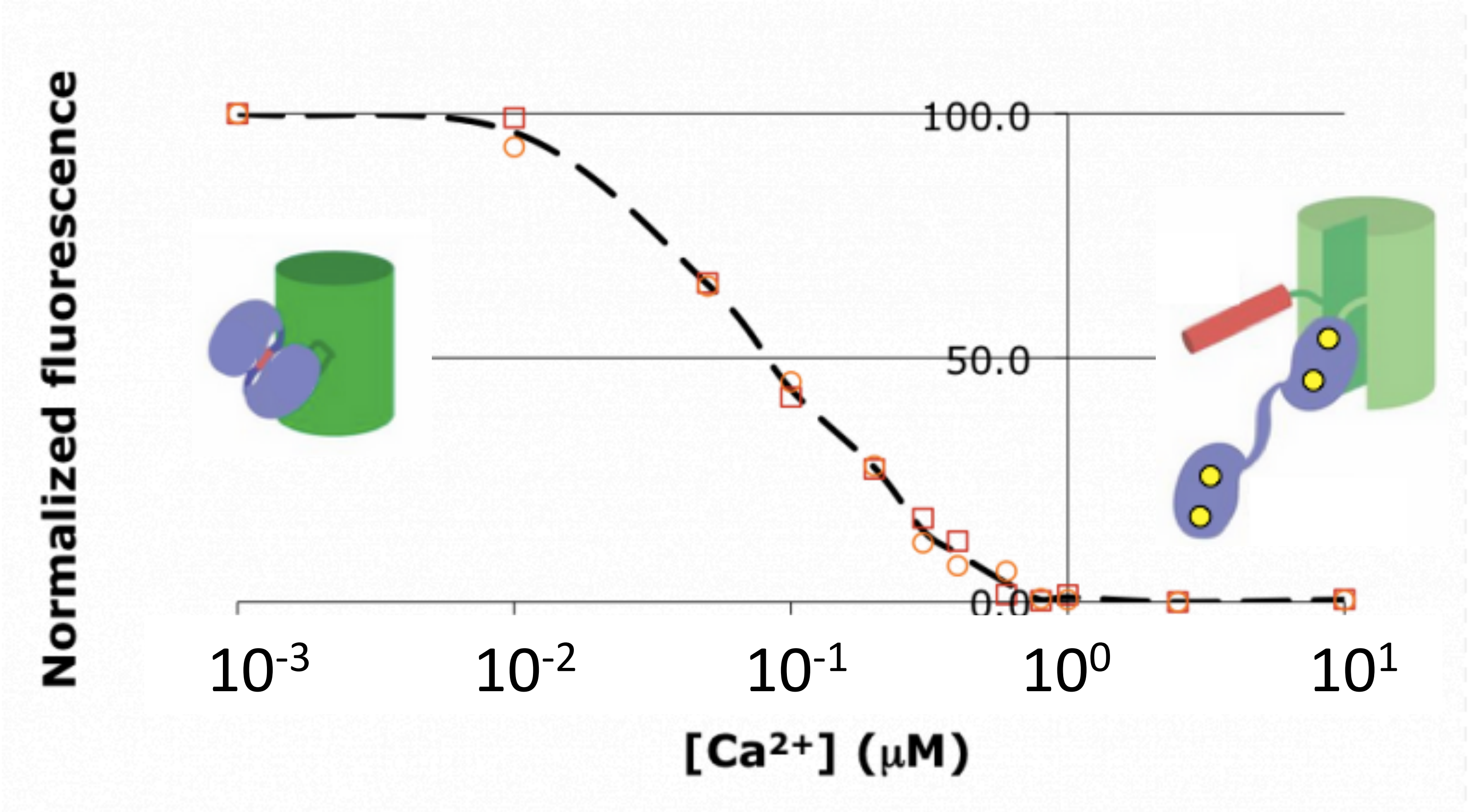




$$\Delta G = -RT \ln K_a = \Delta H - T\Delta S$$

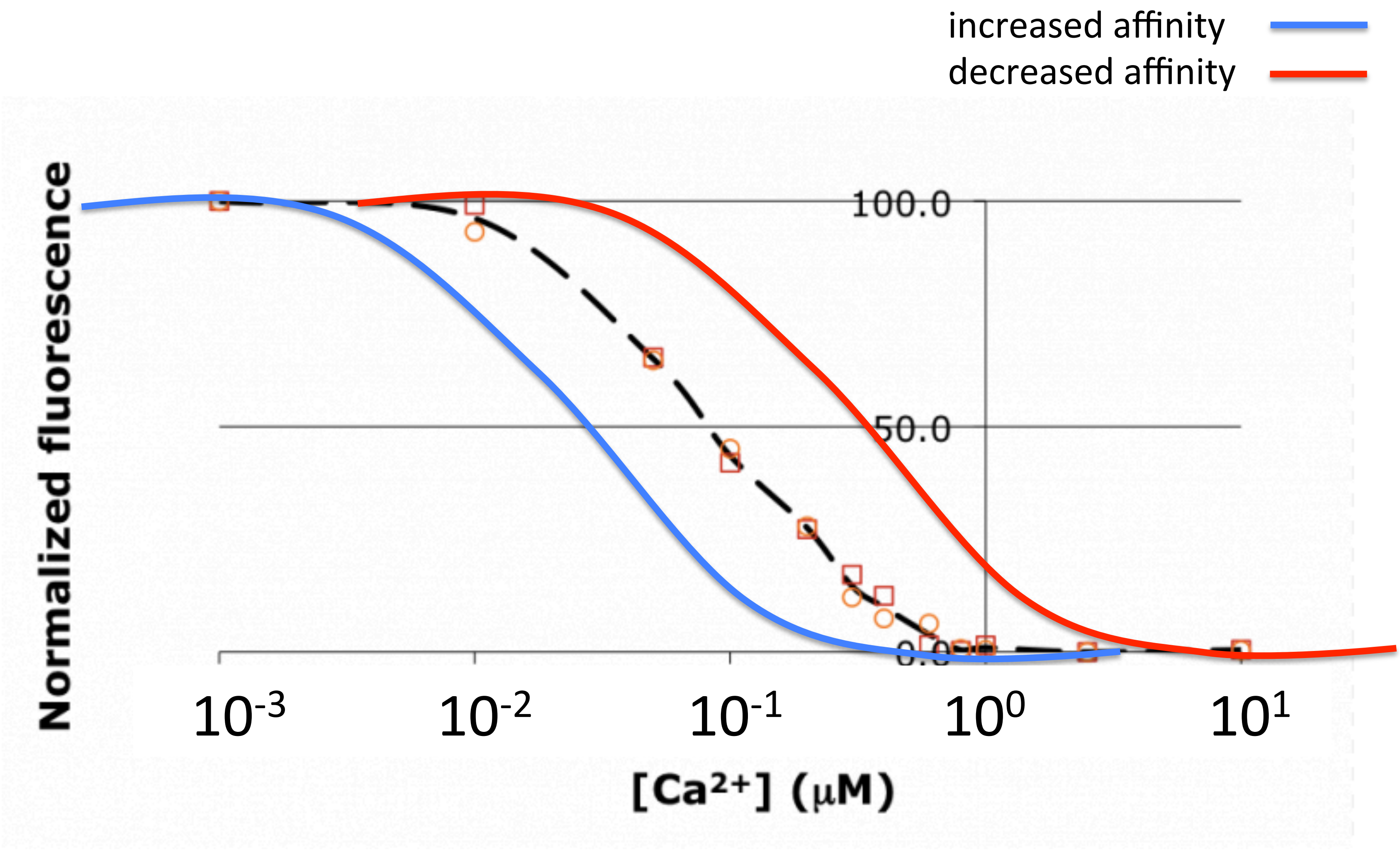
Backman (2015) *PeerJ* **3**: e944

How does your mutation alter IPC-Ca^{2+} binding?

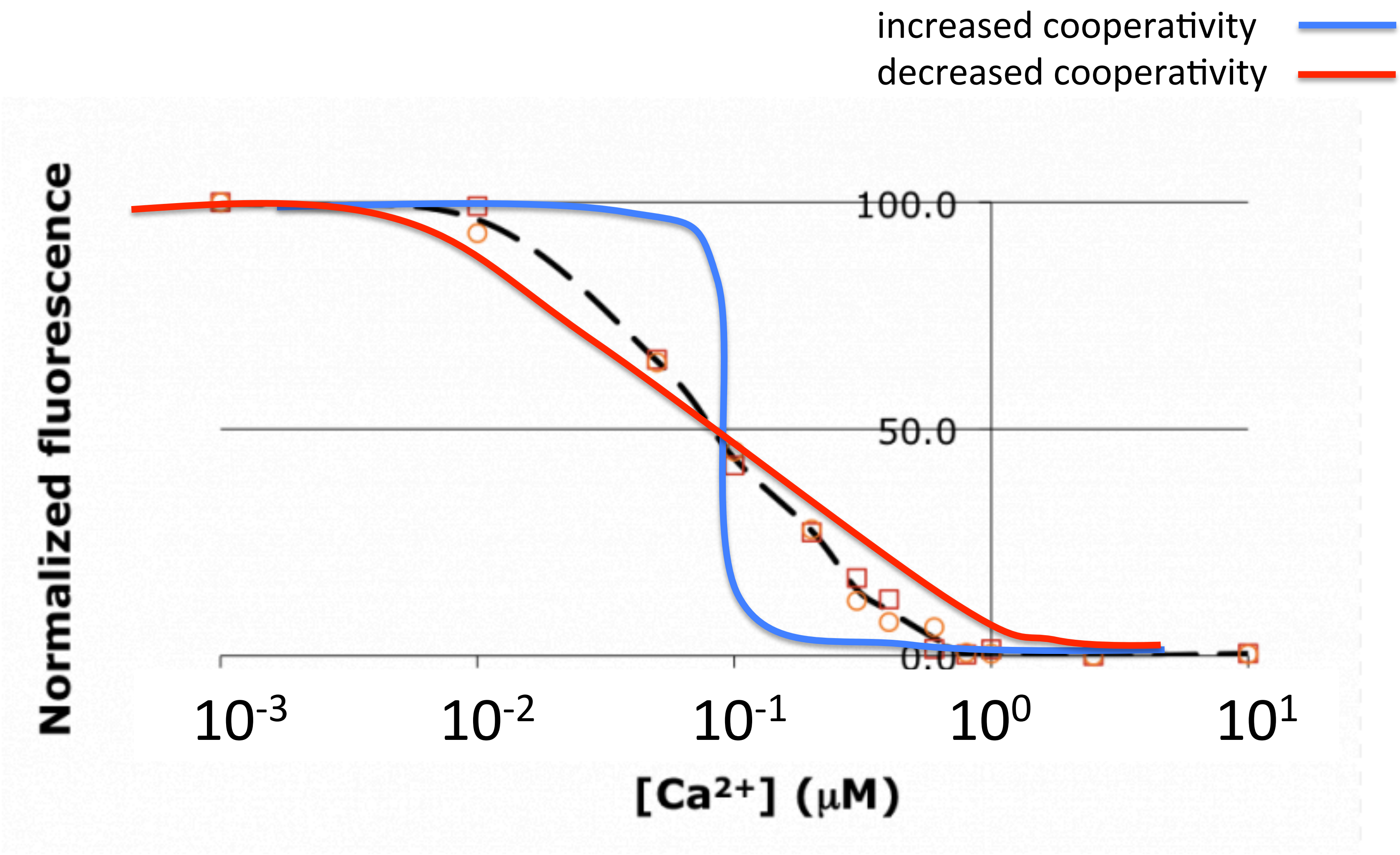


- What parameters are we assessing?

Does your mutation affect affinity?



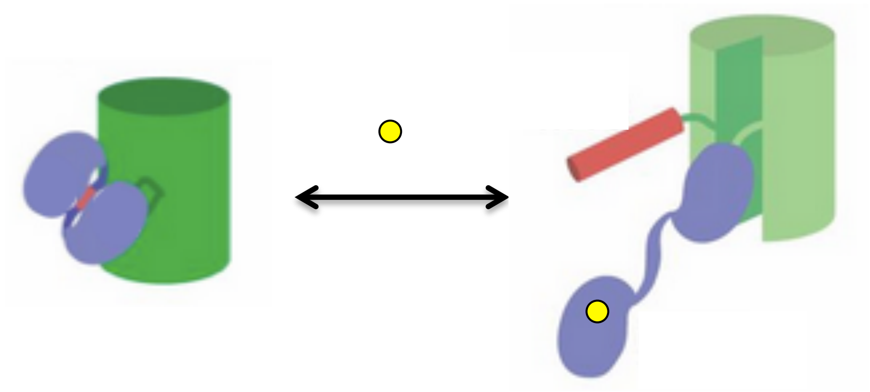
(and/or) does your mutation change cooperativity?



First-order kinetics, single ligand case

$$P_{unbound} + L \leftrightarrow P_{bound}$$

$$K_d = \frac{[P_{unbound}][L]}{[P_{bound}]}$$



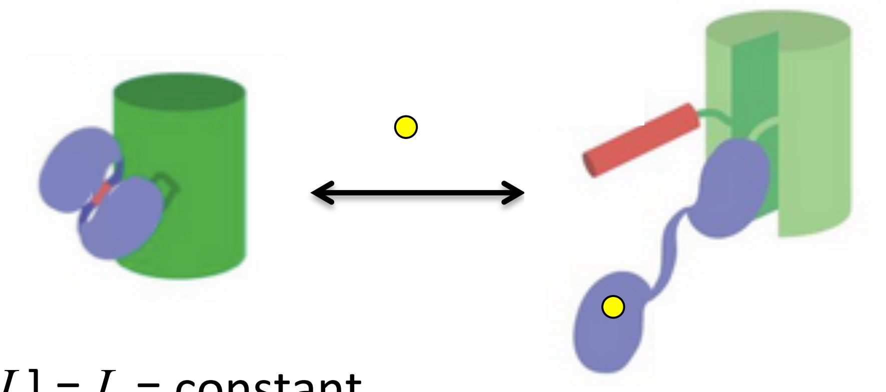
dissociation constant K_d = ligand concentration at which ½ of binding sites occupied

$$y = \frac{[P_{bound}]}{[P_{bound}]+[P_{unbound}]} = \frac{\dfrac{[P_{unbound}][L]}{K_d}}{\dfrac{[P_{unbound}][L]}{K_d}+[P_{unbound}]}$$, so $$\boxed{y = \frac{[L]}{[L]+K_d}}$$

First-order kinetics, single ligand case

$$P_{unbound} + L \leftrightarrow P_{bound}$$

$$y = \frac{[L]}{[L] + K_d}$$

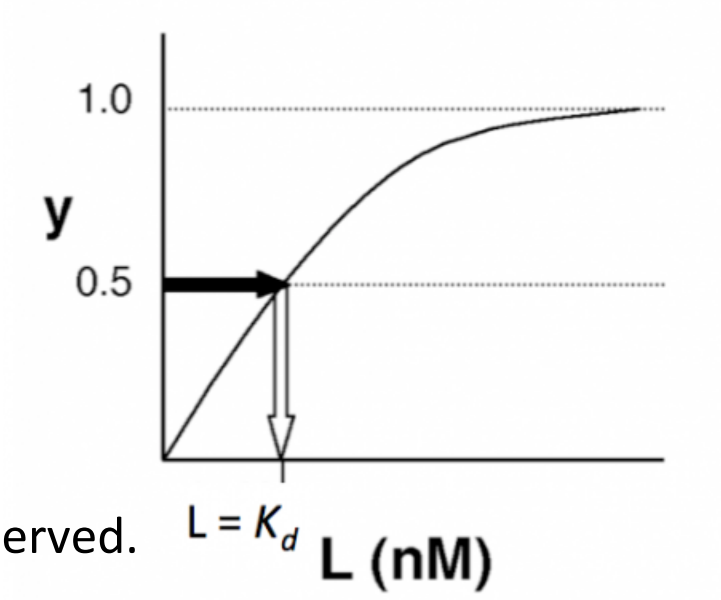


If L in excess (buffered solution), and $[L] = L$ = constant

- if $L << K_d$ then $y \approx \frac{[L]}{K_d}$ (linear)
- if $L >> K_d$ then $y \approx 1$ (saturation)
- at $L = K_d$ $y = 0.5$

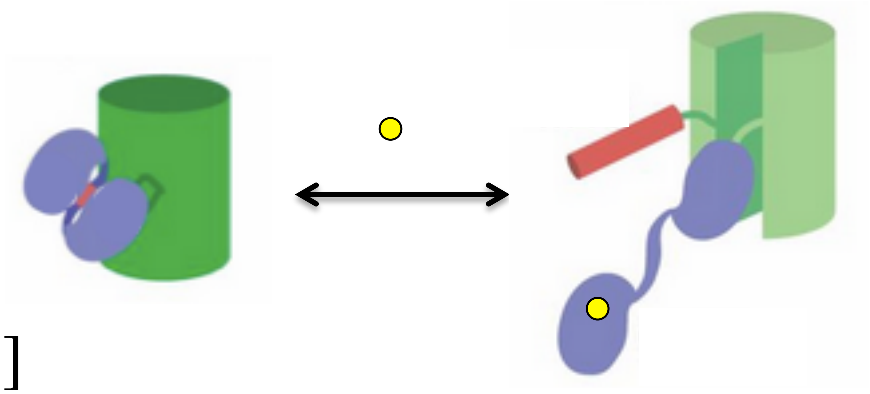
and $\boldsymbol{K_d = EC_{50}}$

EC_{50} = ligand concentration at which ½ of maximum response observed.



Measuring K_d from fluorescence signal

$$P_{unbound} + L \leftrightarrow P_{bound}$$

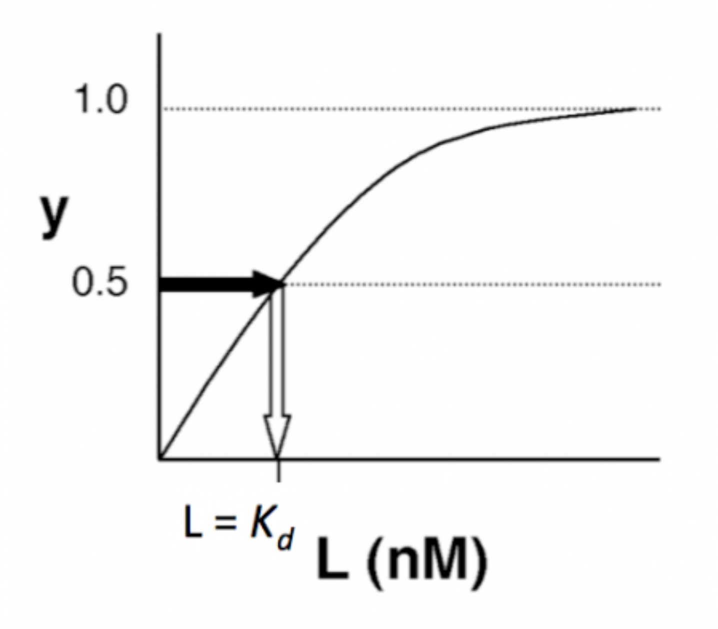


- $P_{unbound}$ and P_{bound} are both fluorescent, to different degrees

$$F = F_{unbound}[P_{unbound}] + F_{bound}[P_{bound}]$$

- Define y as **fractional saturation** of fluorescence signal:
  - note: $F_{max} \equiv$ all unbound
    $F_{min} \equiv$ all bound

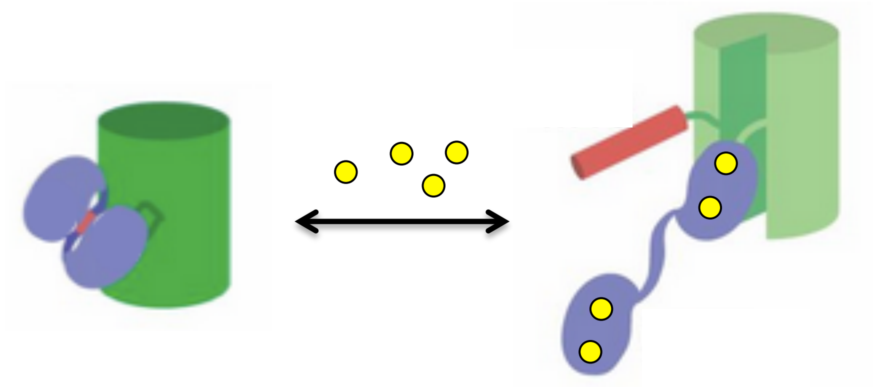
$$y = \frac{[P_{bound}]}{[P_{bound}]+[P_{unbound}]} = \frac{F_{max} - F}{F_{max} - F_{min}}$$



Calmodulin has 4 calcium binding sites

$$P_{unbound} + 4L \leftrightarrow P_{bound}$$

$$K_d = \frac{[P_{unbound}][L]^4}{[P_{bound}]}$$

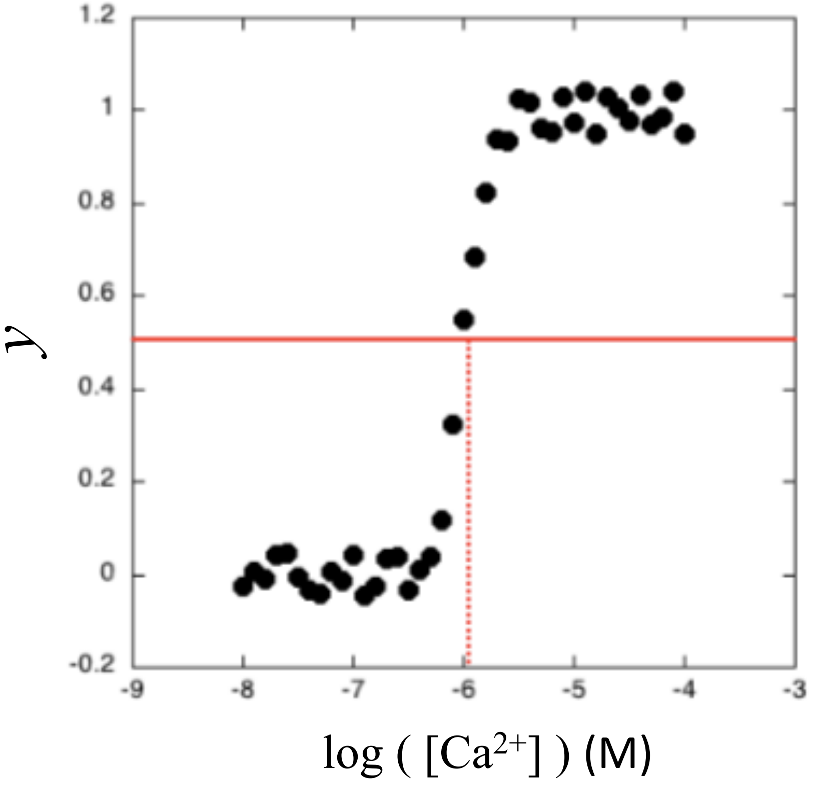


$K_d^{1/4}$ = ligand concentration at which ½ of IPC is bound to calcium

$$y = \frac{F_{\max} - F}{F_{\max} - F_{\min}} = \frac{[L]^4}{[L]^4 + K_d} = \frac{[L]^4}{[L]^4 + (EC_{50})^4}$$

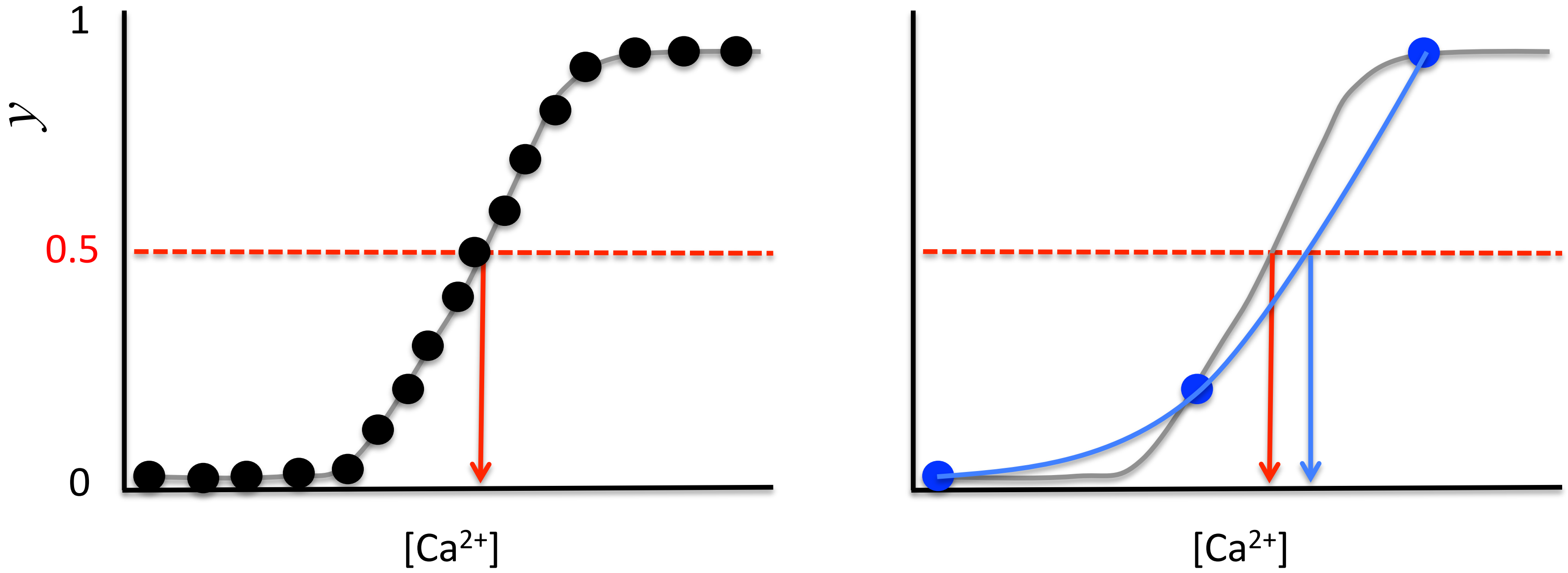
Determine K_d and EC_{50} from binding curves

1. Look at the mid-point of the fluorescence change



(simulated data)

On the importance of having dense data point collection for [Ca^{2+}] ~ EC_{50}



Each point represents a fluorescence measurement at a known [Ca^{2+}]

Determine K_d and EC_{50} from binding curves

2. Curve fitting (with MATLAB)

Part 1: fit apparent K_d

$$y = \frac{L}{K_d + L}$$

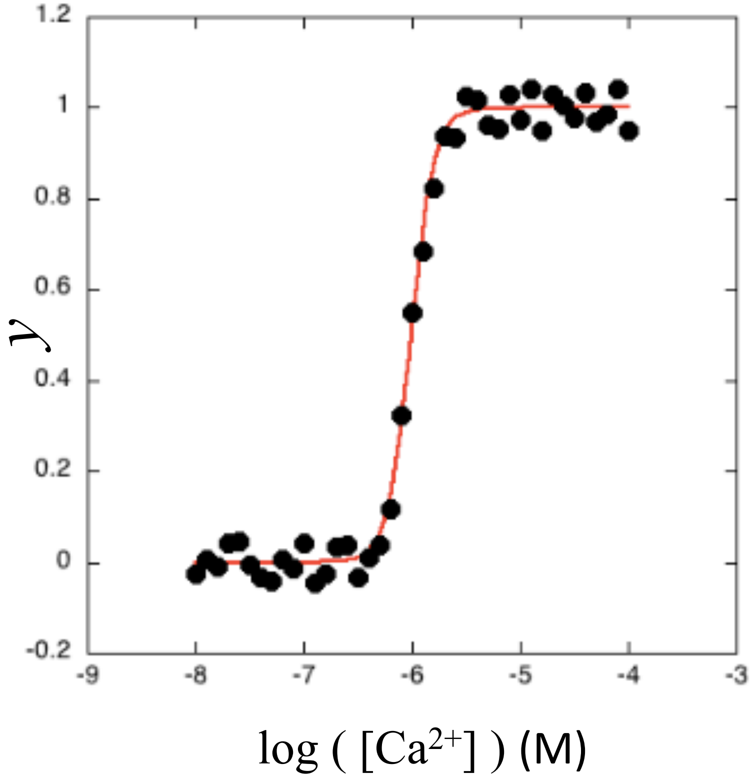
fit assuming

- first-order kinetics
- with 4 ligands

$$y = \frac{L^4}{K_d + L^4}$$

Part 2: fit K_d and n

$$y = \frac{L^n}{K_d + L^n}$$



Determine K_d and EC_{50} from binding curves

3. Hill analysis

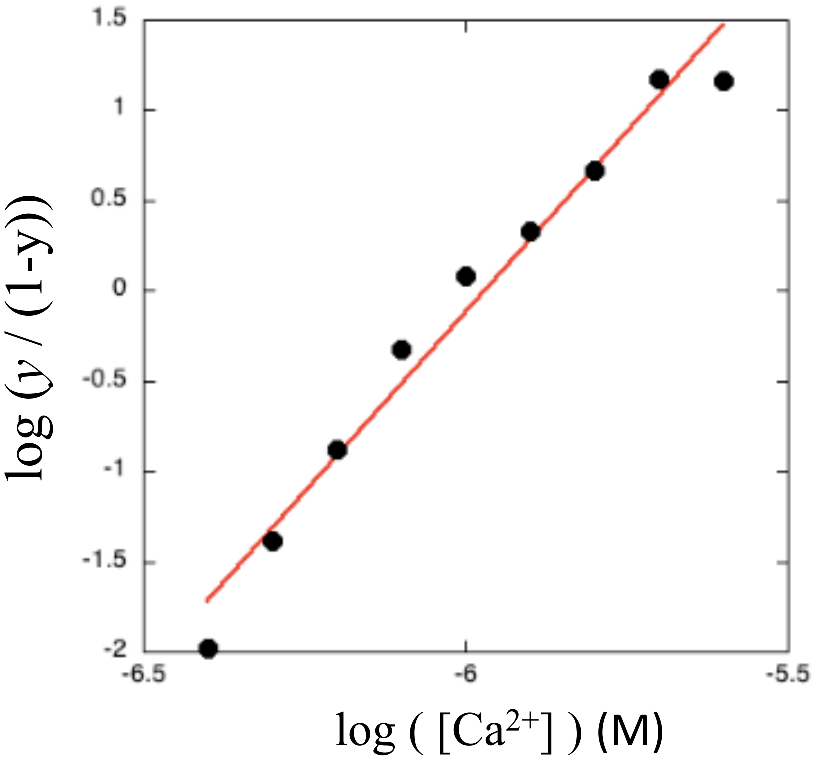
$$y = \frac{L^n}{K_d + L^n}$$

$$1 - y = 1 - \frac{L^n}{K_d + L^n} = \frac{K_d}{K_d + L^n}$$

$$\boxed{\log\left(\frac{y}{1-y}\right) = n\log(L) - n\log(K_d)}$$

- slope = n
  - **Hill coefficient**
  - indicative of cooperativity
- x-intercept = $\log K_d$

Part 3: plot $\log\left(\frac{y}{1-y}\right)$ *vs.* $\log([\mathrm{Ca}^{2+}])$

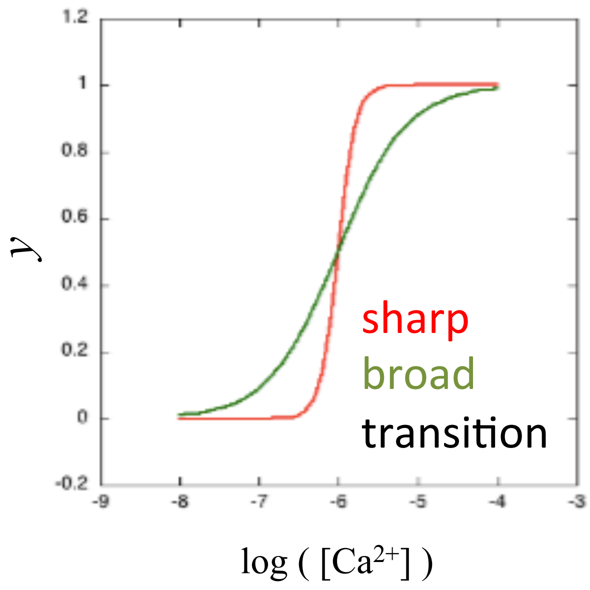


Cooperativity effects on titration curves

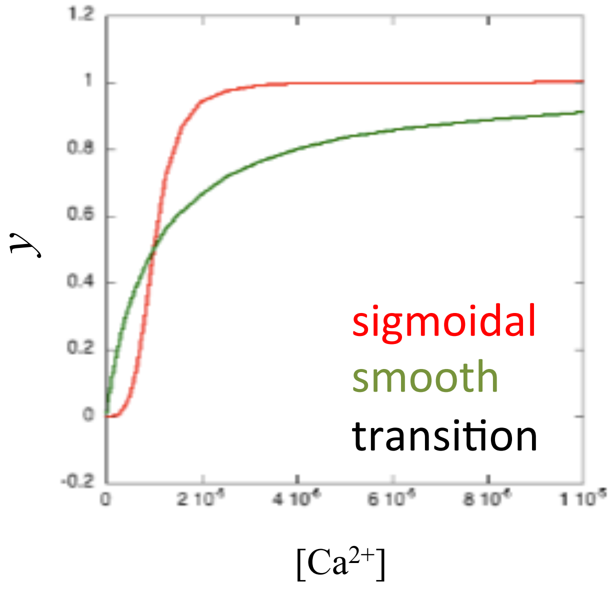
- Hill coefficient n reflects cooperativity
  - positive cooperativity ($n > 1$): binding to one site promotes binding to other sites
  - negative cooperativity ($n < 1$): depresses

- Pericam shows positive cooperativity in calcium binding. BAPTA shows none.

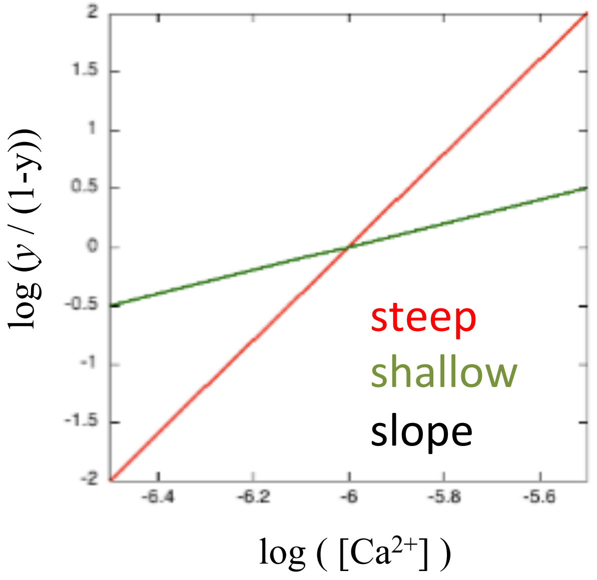
saturation *vs.* log $[Ca^{2+}]$



saturation *vs.* $[Ca^{2+}]$

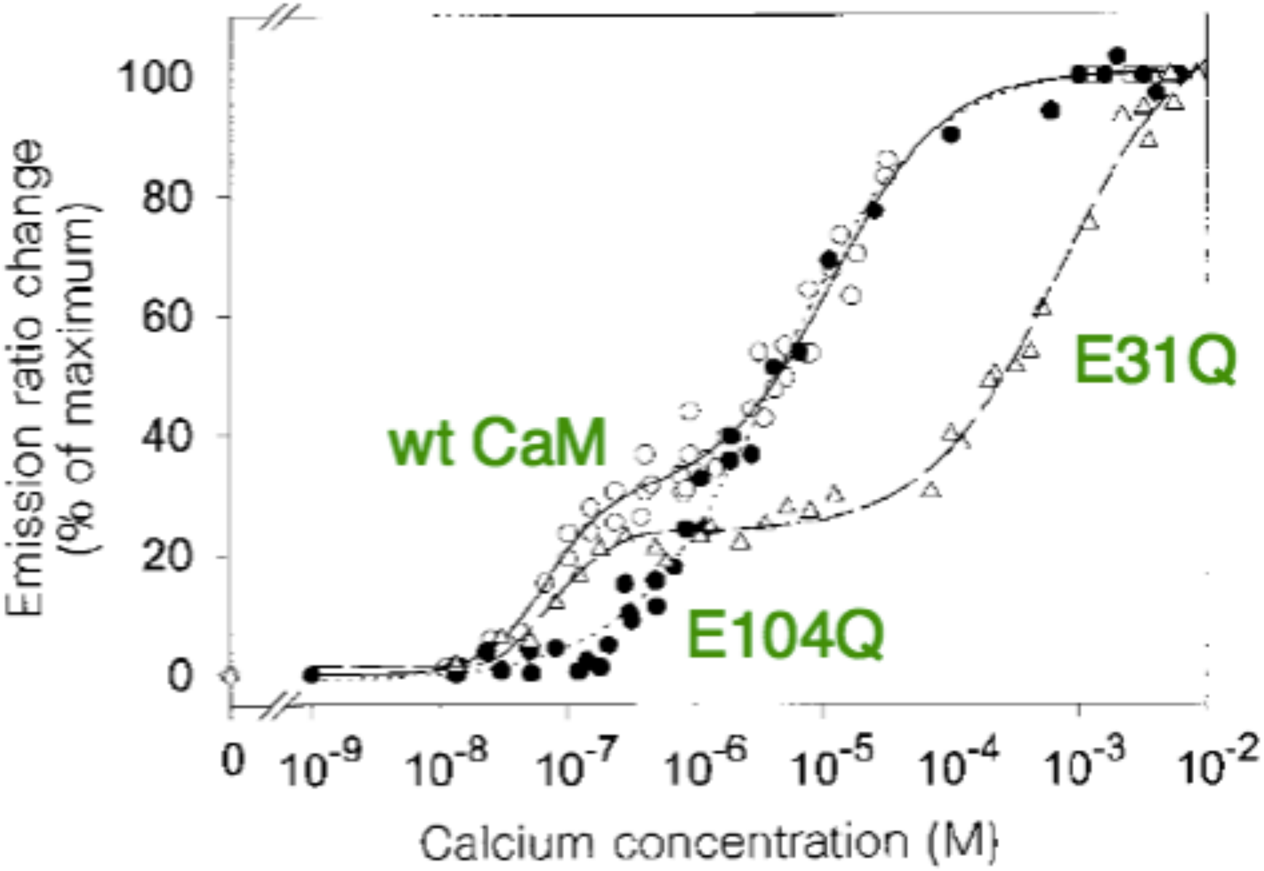


Hill plot



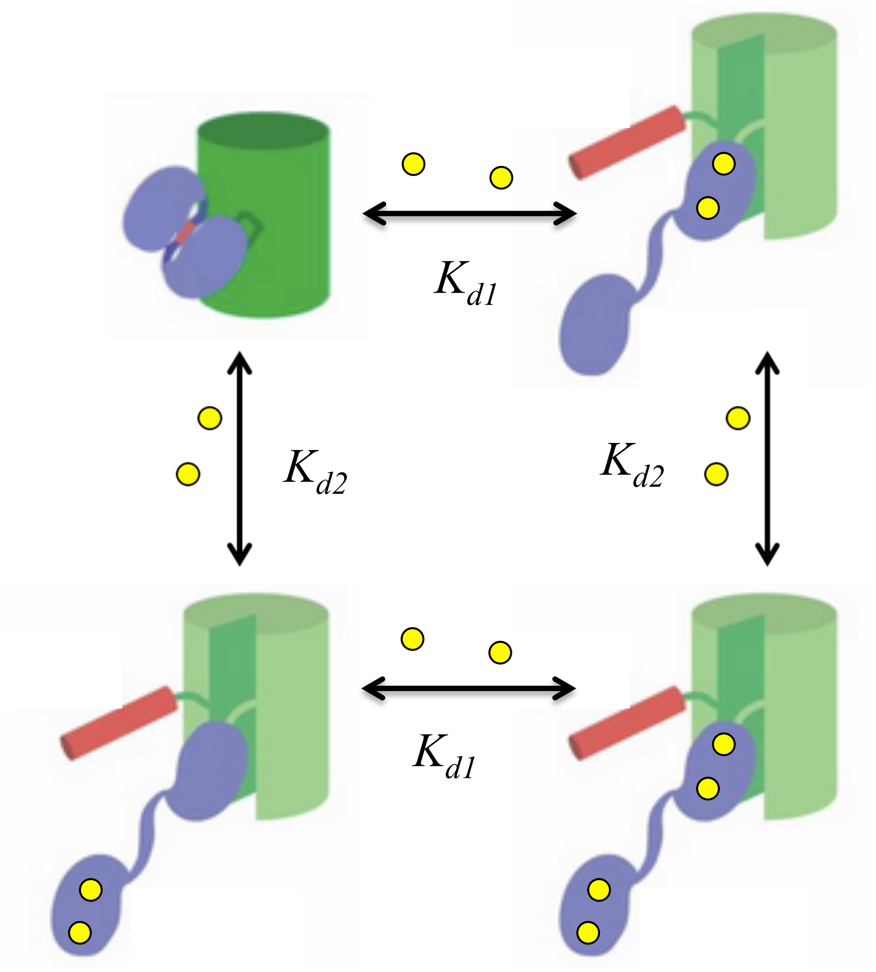
* BAPTA: (1,2-bis(o-aminophenoxy)ethane-N,N,N',N'-tetraacetic acid)

Complete cooperativity assumes that all 4 Ca2+ ions bind at the same time



Miyawaki A, Griesbeck O, Heim R, Tsien RY (1999). "Dynamic and quantitative Ca2+ measurements using improved Cameleons.". *Proc Natl Acad Sci USA* **96** (5): 2135–40.

Independent pairwise calcium binding model may better approximate CaM data



$$y = f_1 \frac{L^{n1}}{K_{d1} + L^{n1}} + f_2 \frac{L^{n2}}{K_{d2} + L^{n2}}$$

Can you think of structural reasons to justify the validity of this model?

Will your mutation exacerbate / re-establish this two-step transition?

M1D8 in lab

- Analyze data with Excel
- Analyze data with MATLAB

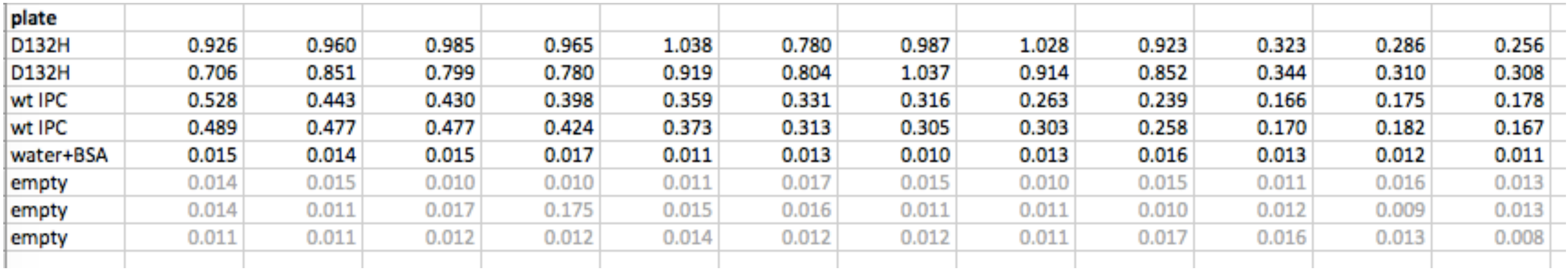
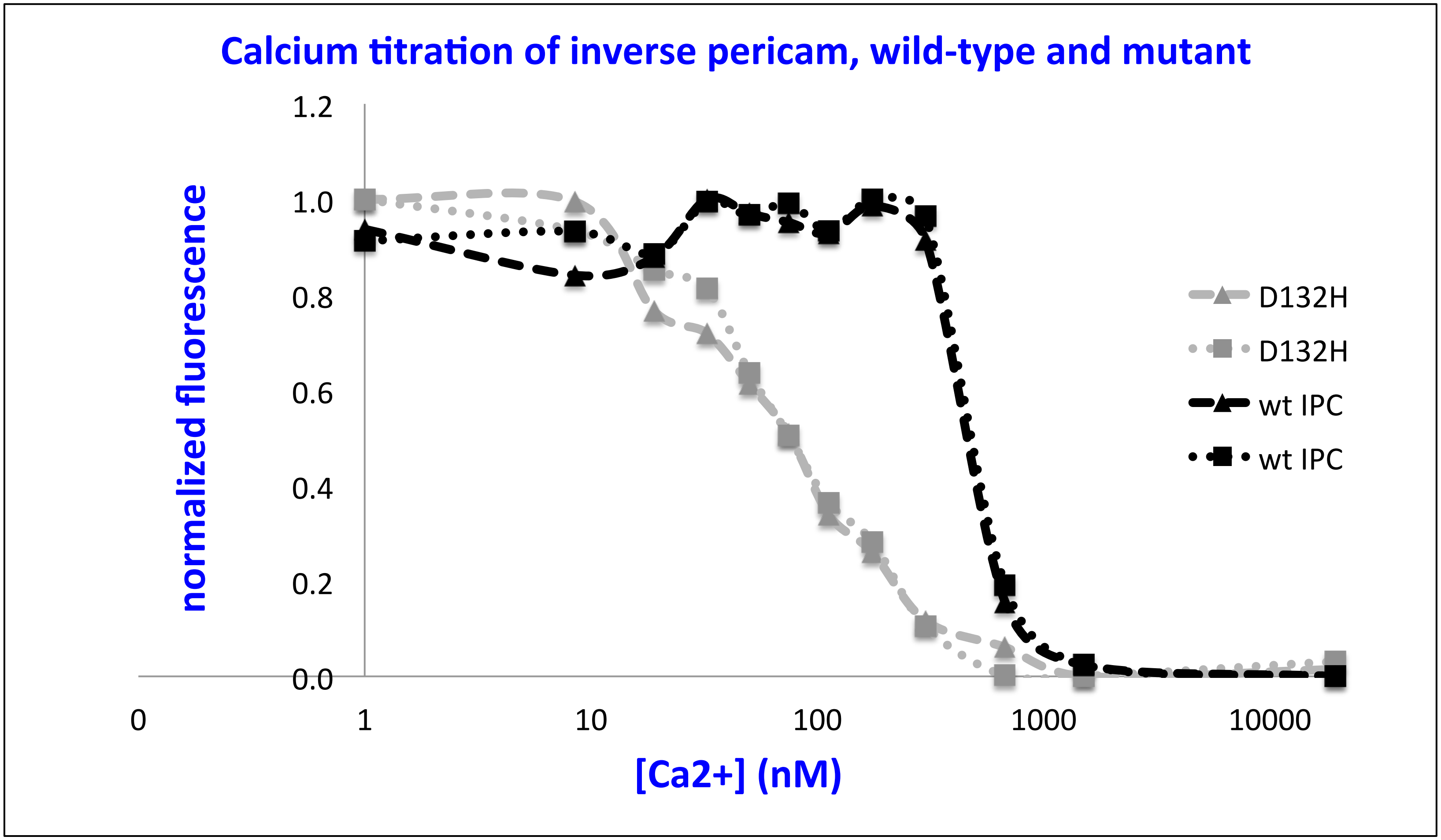


plate												
D132H	0.926	0.960	0.985	0.965	1.038	0.780	0.987	1.028	0.923	0.323	0.286	0.256
D132H	0.706	0.851	0.799	0.780	0.919	0.804	1.037	0.914	0.852	0.344	0.310	0.308
wt IPC	0.528	0.443	0.430	0.398	0.359	0.331	0.316	0.263	0.239	0.166	0.175	0.178
wt IPC	0.489	0.477	0.477	0.424	0.373	0.313	0.305	0.303	0.258	0.170	0.182	0.167
water+BSA	0.015	0.014	0.015	0.017	0.011	0.013	0.010	0.013	0.016	0.013	0.012	0.011
empty	0.014	0.015	0.010	0.010	0.011	0.017	0.015	0.010	0.015	0.011	0.016	0.013
empty	0.014	0.011	0.017	0.175	0.015	0.016	0.011	0.011	0.010	0.012	0.009	0.013
empty	0.011	0.011	0.012	0.012	0.014	0.012	0.012	0.011	0.017	0.016	0.013	0.008

Plot your IPC-calcium titration data in Excel

- Normalize data (or average of 2 data sets): $S = \dfrac{F - F_{\min}}{F_{\max} - F_{\min}}$

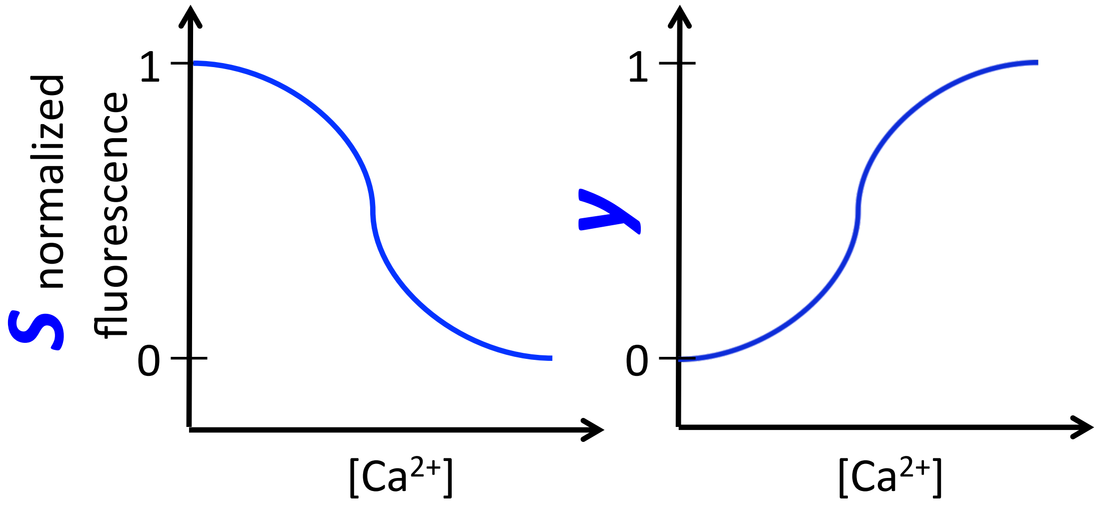


MATLAB code analyzes data along 3 models

- Fractional saturation formalism:

$$y = 1 - S$$

$$y = \frac{F_{\max} - F}{F_{\max} - F_{\min}}$$



Part 1: fit apparent K_d

$$y = \frac{L}{K_d + L}$$

Part 2: fit K_d and n

$$y = \frac{L^n}{K_d^n + L^n}$$

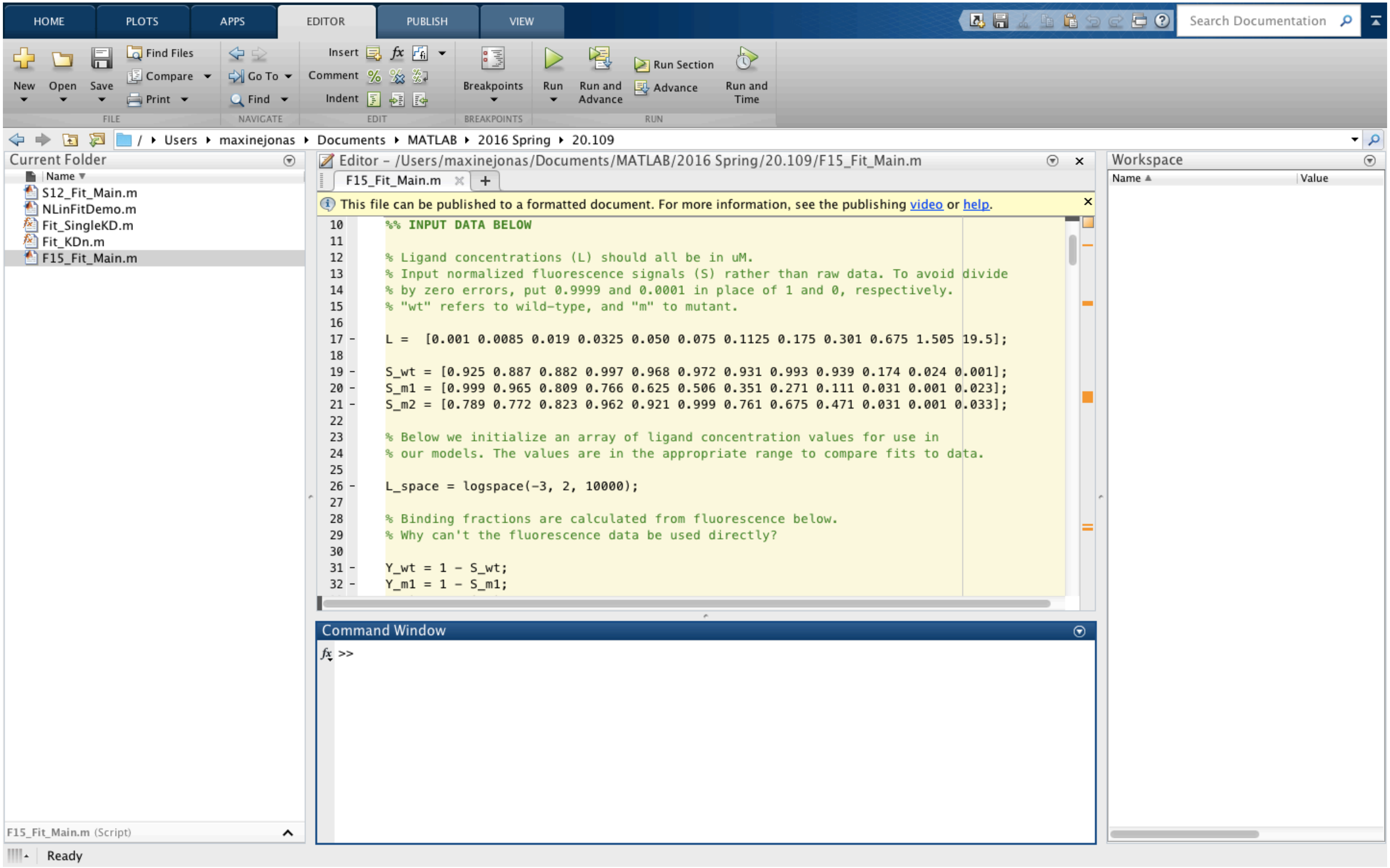
Part 3: fit K_d and n

Hill analysis

$$\log\left(\frac{y}{1-y}\right) = n\log(L)$$

$$-n\log(K_d)$$

Welcome to MATLAB!



This file can be published to a formatted document. For more information, see the publishing video or help.

```
%% INPUT DATA BELOW

% Ligand concentrations (L) should all be in uM.
% Input normalized fluorescence signals (S) rather than raw data. To avoid divide
% by zero errors, put 0.9999 and 0.0001 in place of 1 and 0, respectively.
% "wt" refers to wild-type, and "m" to mutant.

L =  [0.001 0.0085 0.019 0.0325 0.050 0.075 0.1125 0.175 0.301 0.675 1.505 19.5];

S_wt = [0.925 0.887 0.882 0.997 0.968 0.972 0.931 0.993 0.939 0.174 0.024 0.001];
S_m1 = [0.999 0.965 0.809 0.766 0.625 0.506 0.351 0.271 0.111 0.031 0.001 0.023];
S_m2 = [0.789 0.772 0.823 0.962 0.921 0.999 0.761 0.675 0.471 0.031 0.001 0.033];

% Below we initialize an array of ligand concentration values for use in
% our models. The values are in the appropriate range to compare fits to data.

L_space = logspace(-3, 2, 10000);

% Binding fractions are calculated from fluorescence below.
% Why can't the fluorescence data be used directly?

Y_wt = 1 - S_wt;
Y_m1 = 1 - S_m1;
```

```
>>
```

Analyze data further in MATLAB

1. Enter your data:
   - `L` = [ligand] = [Ca^{2+}] in μM
   - `S_wt`: signal wild-type IPC
   - `m1` is *your* mutant, `m2` is another team's

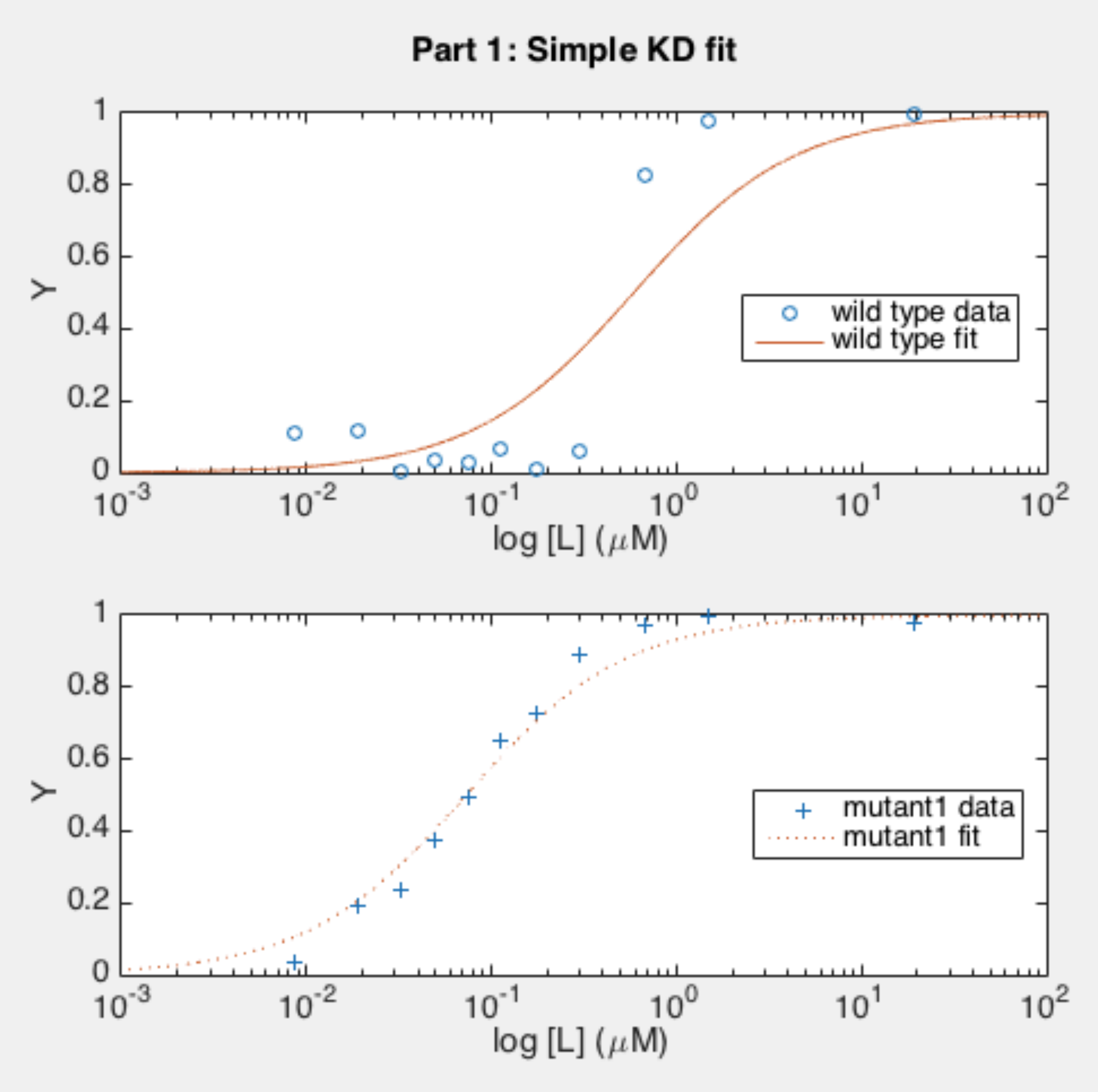
2. `logspace` (a, b, N)
   - generates a row vector of N logarithmically equally spaced points between decades 10^a and 10^b.
   - choose a = -3 , b = 2, and N = 10,000

3. `A ./ B`
   - divides element by element

$$\begin{bmatrix} 2 & 4 & 6 \\ 3 & 6 & 9 \\ 4 & 8 & 12 \end{bmatrix} ./ \begin{bmatrix} 2 & 2 & 2 \\ 3 & 3 & 3 \\ 4 & 4 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}$$

Part 1: fit apparent K_d

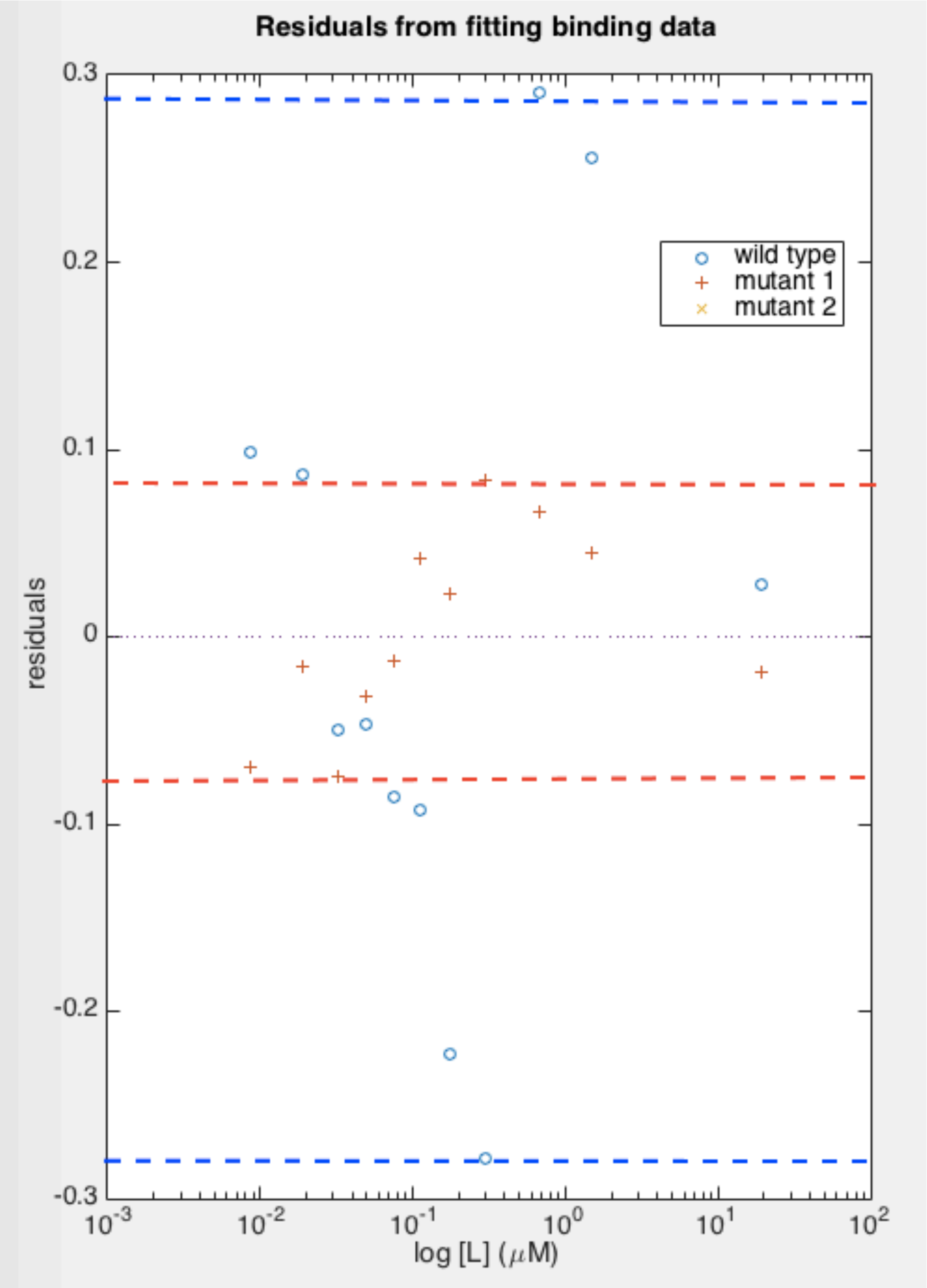
$$y = \frac{L}{K_d + L}$$



```
KD1_wt = 0.5858 µM

KD1_m1 = 0.0729 µM
```

- How good is the fit?
  - for wt-IPC?
  - for mutant?



Part 1

$$y = \frac{L}{K_d + L}$$

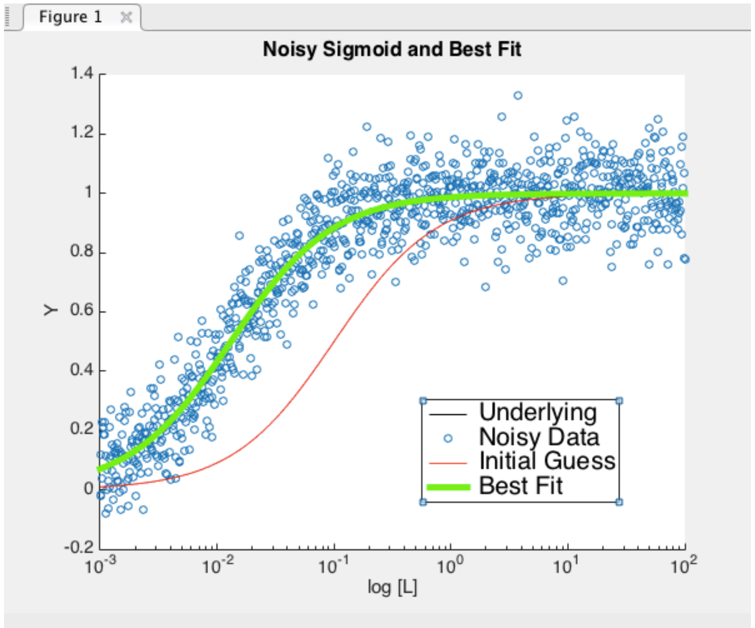
- How good is the fit?
  - for wt-IPC?
  - for mutant?

➢ Quantify *residuals*: distribution and amplitude

Nonlinear regression is at the core of the MATLAB code

- `nlinfit( X, Y, @model, initialGuess )`
  - X (predictors): calcium concentrations
  - Y (responses): fluorescence signal
  - model: `Fit_SingleKD`

    ```
    x ./ ( KD + x );
    ```
  - initialGuess: starting value for `KD`



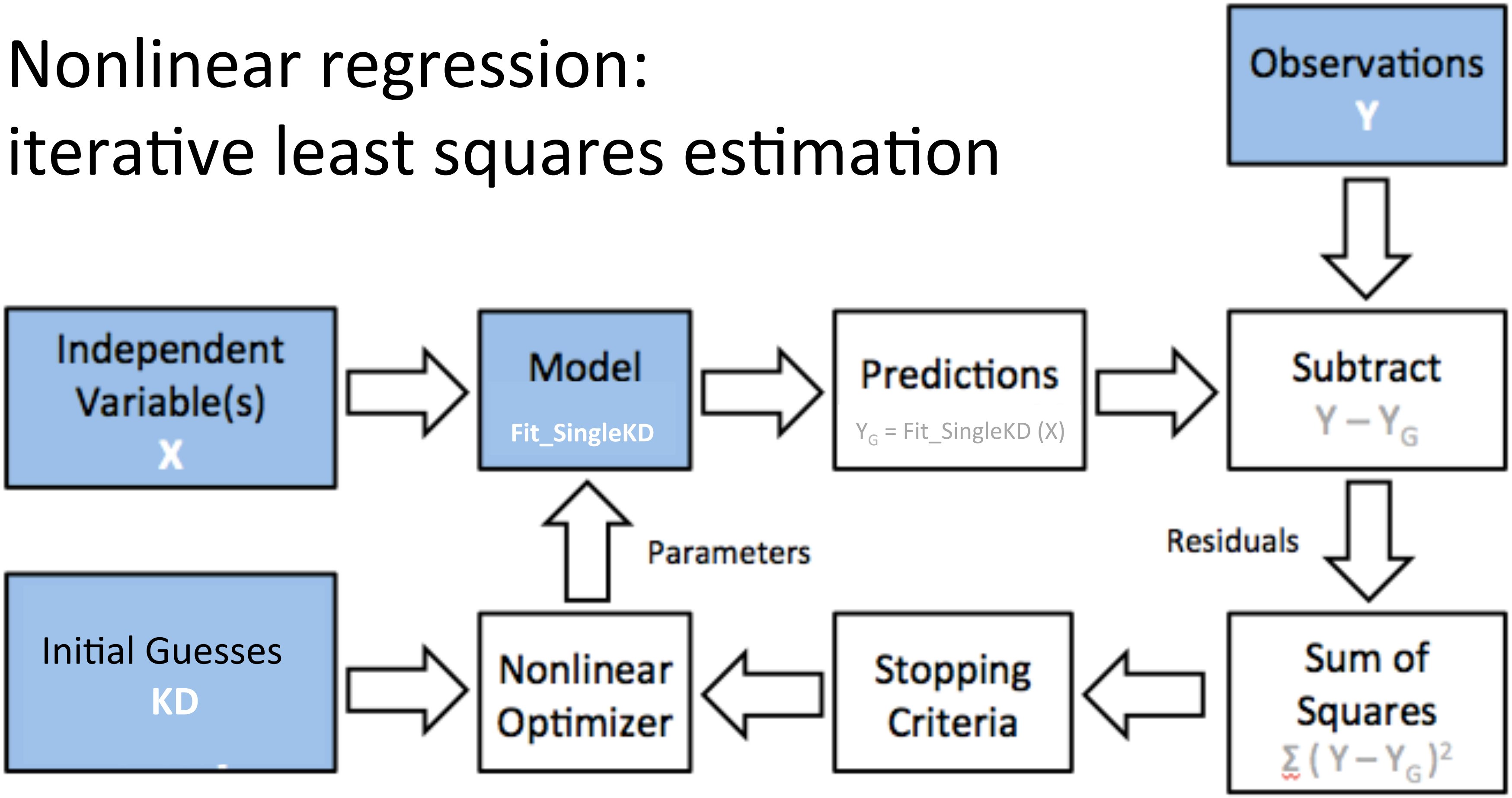
- Find parameters that can explain

  `Y = model( Parameters, X )`

  and start your search with

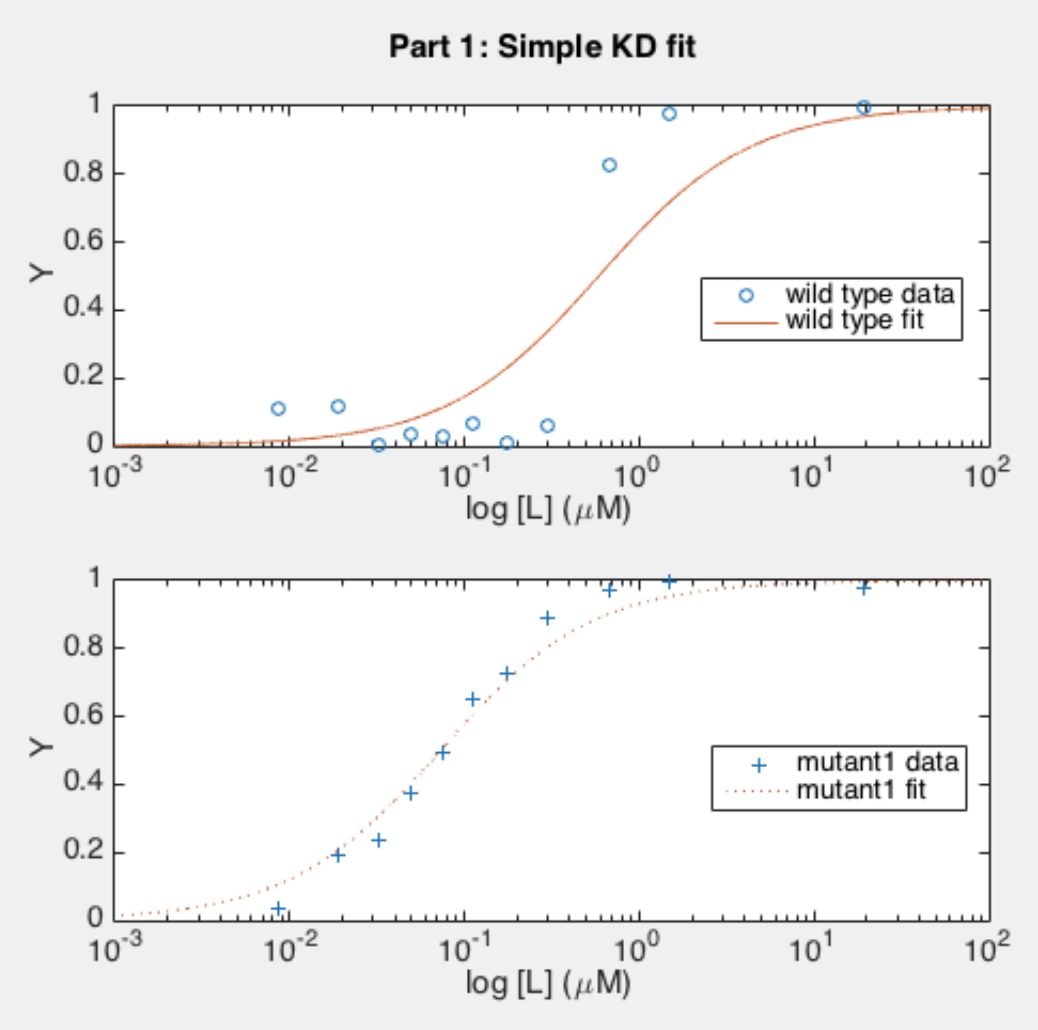
  `parameters0 = initialGuess`

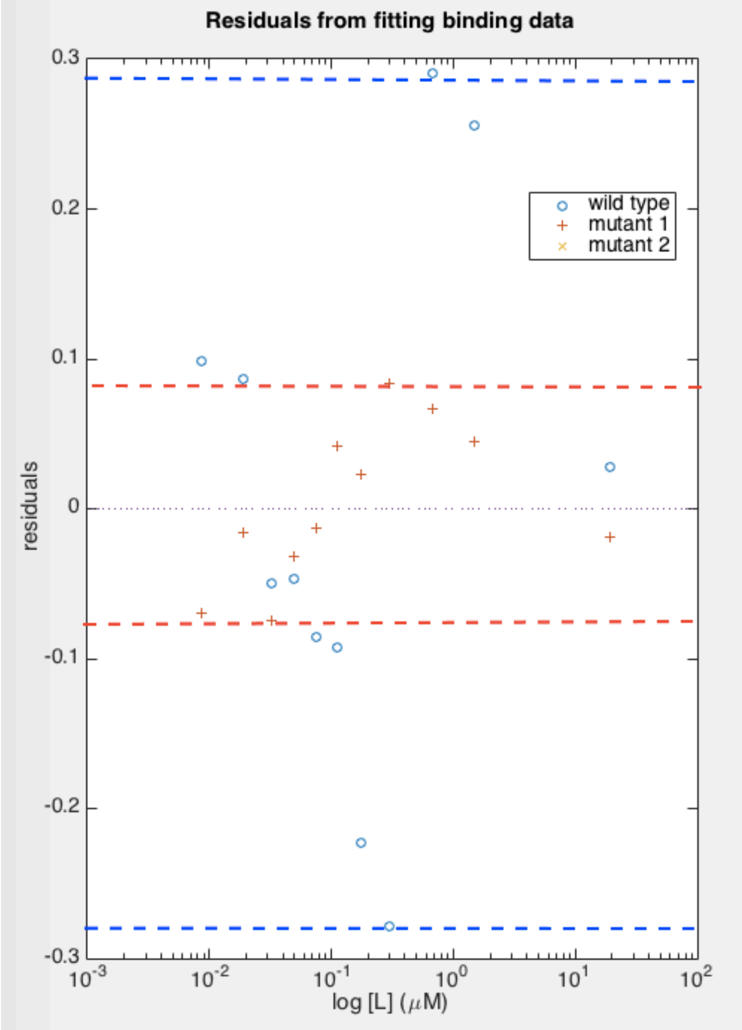
Nonlinear regression:
iterative least squares estimation



- Optimum reached = changing any of the parameters will result in a higher residual sum of squares.
- Optimizer stops when parameters or sum of squared residuals changes less than tolerance, or when maximum number of iterations reached.

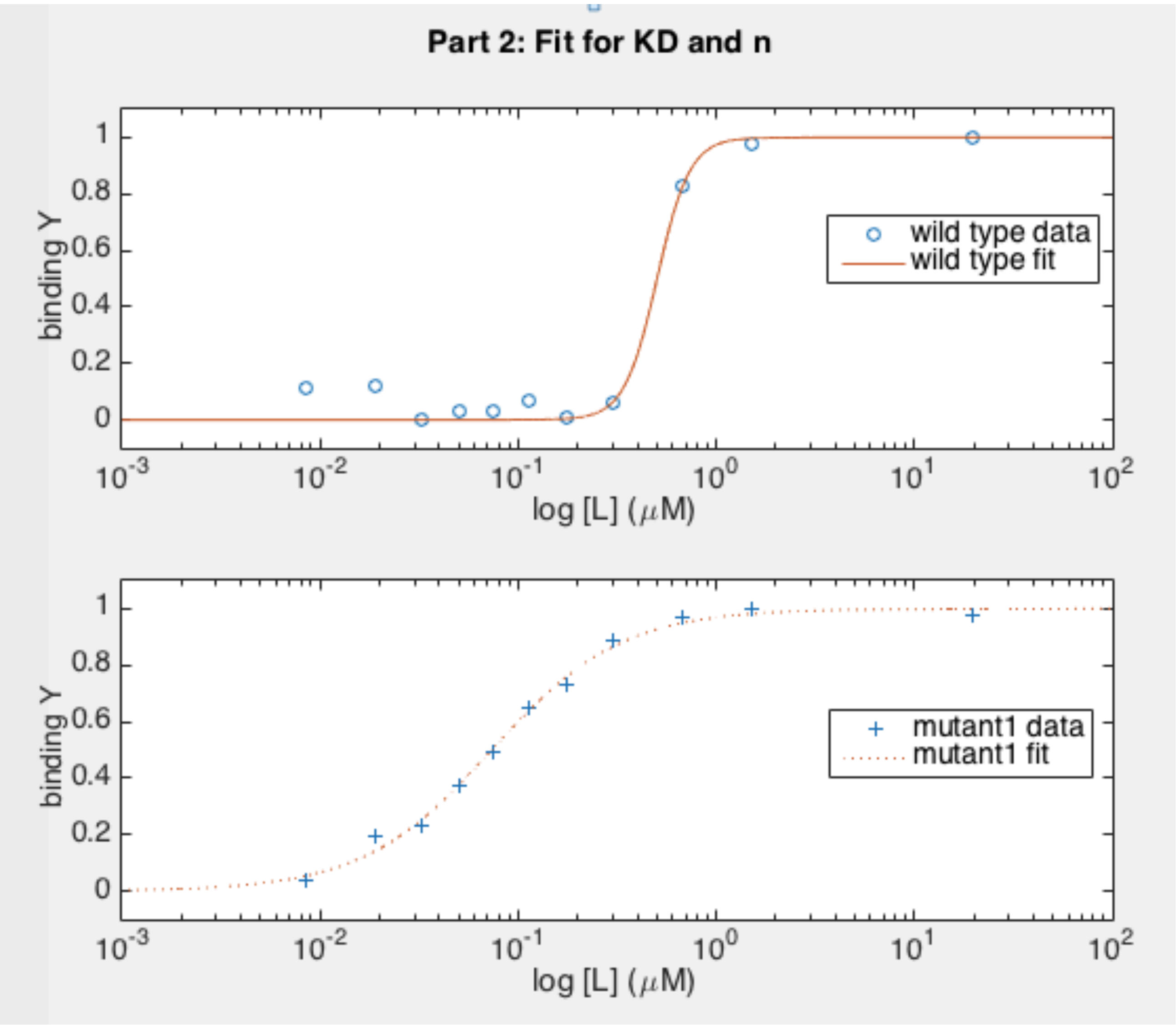
… and this is why residuals $y - y_{model}$ provide qualitative and quantitative goodness of fit!





Part 2: fit K_d and n

$$y = \frac{L^n}{K_d^n + L^n}$$



```
KD2_wt = 0.5025 μM
n_wt = 5.2508


KD2_m1 = 0.0737 μM
n_m1 = 1.3250
```

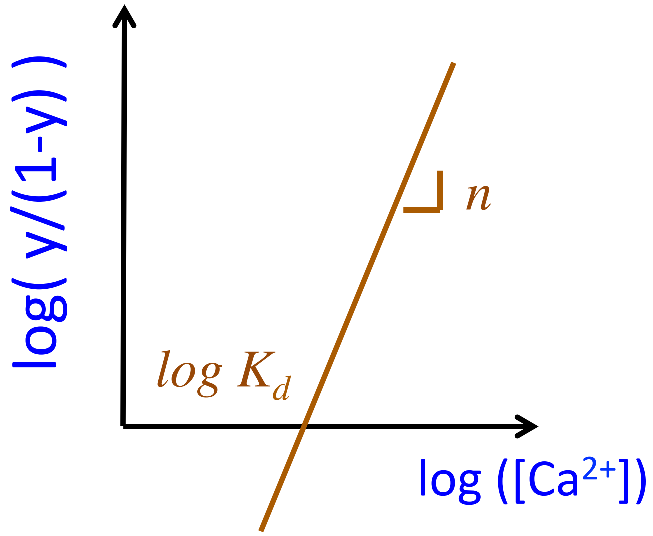
- Is the fit any better?

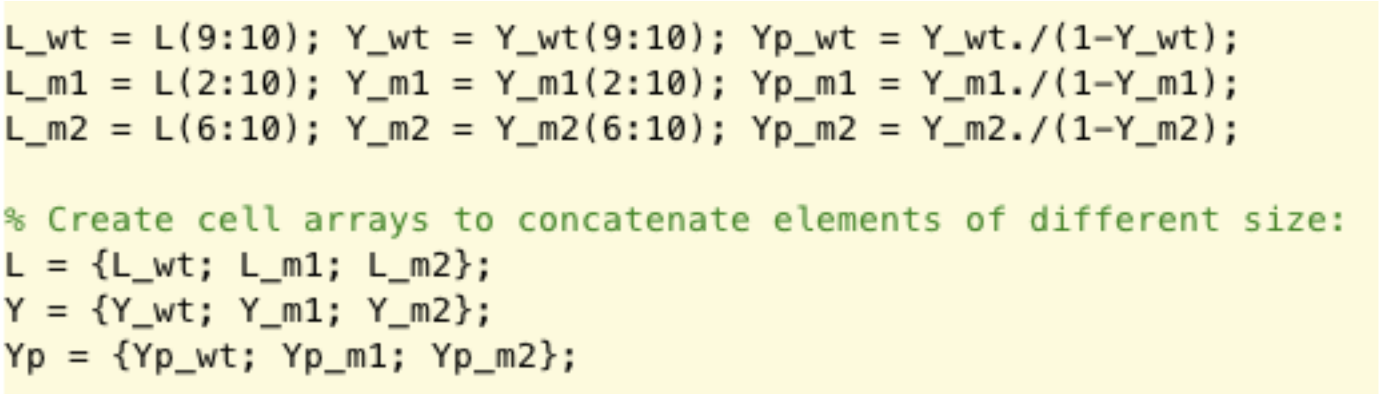
Part 3: fit K_d and n by Hill analysis

$$\log\left(\frac{y}{1-y}\right) = n\log(L) - n\log(K_d)$$

- Work only on linear transition region
  - linear fit (polynomial of degree 1)
  - x-intercept = log (K_d)
  - slope = n

- Will need to change indexes in MATLAB algorithm
  - then work with *cell arrays* to parallelize analysis

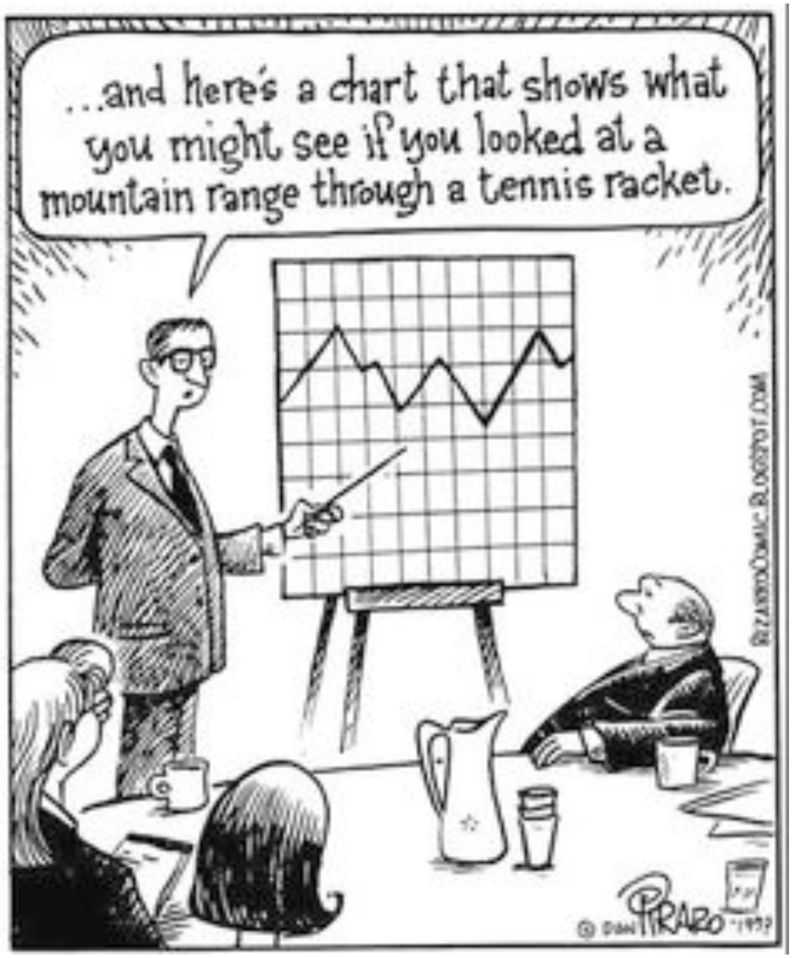


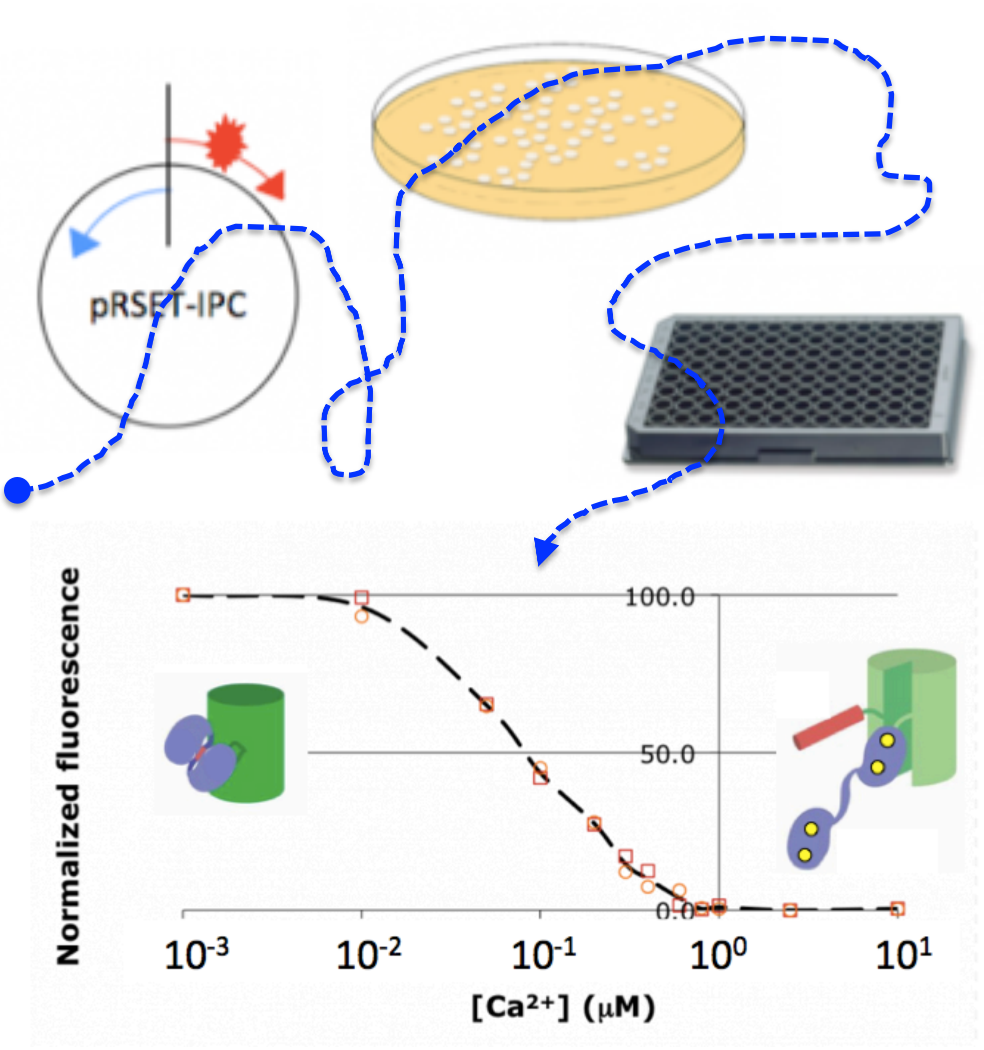


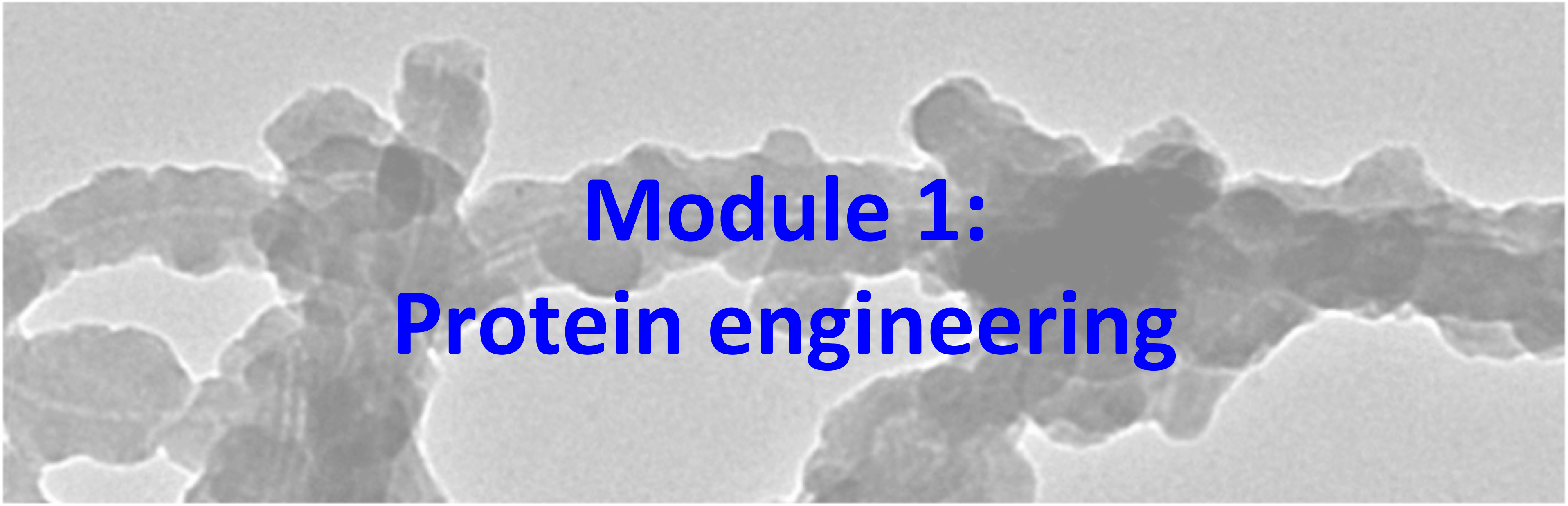
```
L_wt = L(9:10); Y_wt = Y_wt(9:10); Yp_wt = Y_wt./(1-Y_wt);
L_m1 = L(2:10); Y_m1 = Y_m1(2:10); Yp_m1 = Y_m1./(1-Y_m1);
L_m2 = L(6:10); Y_m2 = Y_m2(6:10); Yp_m2 = Y_m2./(1-Y_m2);

% Create cell arrays to concatenate elements of different size:
L = {L_wt; L_m1; L_m2};
Y = {Y_wt; Y_m1; Y_m2};
Yp = {Yp_wt; Yp_m1; Yp_m2};
```

Make a story out of your M1 results


...and here's a chart that shows what you might see if you looked at a mountain range through a tennis racket.
BIZARROCOMIC.BLOGSPOT.COM
© Dan PIRARO 1997





Module 1:
Protein engineering

I. Binding analysis

II. MATLAB basics

03/03/2016