EGFR binding data reported by Pike et al. clearly depends on receptor concentration in addition to ligand concentration. This is because EGFR is a dimerizing system, and the dimer has different binding affinity for EGF than the monomer [ref]. Pike explains the binding behavior in the context of a linked system with negative cooperativity. This means that dimerization is not governed only by a dimerization equilibrium constant, but that the binding of ligands also affects the distribution of monomeric and dimeric forms of the receptor, and that the binding of a ligand to the first site on the dimer diminishes the affinity of the second binding site.

However, Pike's criteria for negative cooperativity is not strict, and her data is not fit to simpler models for comparison. Visual inspection of the data does not preclude the possibility that cooperativity is not present as the ligand binding intervals (fractional saturation = 0.1-0.9) fall within ~2 log units [ref].

By showing that the model invoked by Pike is of the general form

(1) $(m)(fsat_{mon})+(1-m)(fsat_{dim})$

where m is the fraction of the receptor population that is in monomeric form, 1-m is the fraction in dimeric form, and $fsat_{mon}$ and $fsat_{dim}$ are general models for one and two site binding equations, then we can establish strict criteria for cooperativity and use simulation to support or refute the presence of cooperativity.

The general models for one and two site binding are given below.

One site:

(2) $$fsat=\frac{k_1[L]}{1+k_1[L]}$$

Two site:

(3) $$fsat=\frac{k_1[L]+2k_1k_2[L]^2}{2(1+k_1[L]+k_1k_2[L]^2)}$$

The two site model can account for different affinities at each site, and since EGFR forms a homodimer, then differences in the site can be attributed to cooperativity. The difference between k_1 and k_2 required for cooperativity are defined below.

Consider a system of two distinct binding sites:

(4) $$\frac{k_a[L]}{1+k_a[L]}+\frac{k_b[L]}{1+k_b[L]}$$

This cross-multiplies and factors to:

(5) $$\frac{(k_a+k_b)[L]+2k_ak_b[L]^2}{1+(k_a+k_b)[L]+k_ak_b[L]^2}$$

If we assign

$$k_a + k_b = k_1 \quad \text{and} \quad k_a k_b = k_1 k_2$$

then (5) takes on the same form as (3), and if the two sites are equivalent ($k_a = k_b = k_{actual}$), then $k_1 = 2k_{actual}$, $k_1 k_2 = k_{actual}^2$ and $k_1 = 4k_2$.

Thus (5) collapses to a single site model (2) where $k_1 = k_{actual}$.

When $k_2 < k_1/4$, the binding sites are distinct with one site binding at a lower affinity than the other. In the case of a homodimer, where the sites are chemically identical, then this defines negative cooperativity – binding at the first site diminishes the binding affinity of the second site. The opposite case is true for positive cooperativity.

We can use this criteria to test for cooperativity if we can show that the model used by Pike is of the form described in (1).

Derivation of Pike's model

Pike's model was previously derived and introduced for a dimerizing system by Wyman and Gill in *Binding and Linkage: Functional Chemistry of Biological Macromolecules* pp. 208-210

L = free EGF
R = free EGFR monomer
RR = free EGFR dimer
RL = bound monomer
RRL = singly bound dimer
$RRLL$ = fully bound dimer

Equilibrium equations (using Pike's association constant notation):

$[RR] = L_{20}[R]^2$
$[RL] = K_{11}[R][L]$
$[RRL] = K_{21}[RR][L]$
$[RRLL] = K_{22}[RR][L]^2$

Fractional saturation is defined as the concentration of bound ligand divided by the total concentration of receptor.

$$(6) \quad fsat = \frac{[RL] + [RRL] + 2[RRLL]}{[R] + [RL] + 2[RR] + 2[RRL] + 2[RRLL]}$$

Substitution with the equilibrium equations and factoring out [R] yields the following equation which was used by Pike to fit the data. It is presented in her papers in a slightly more factored form but the equations are the same.

$$(7) \quad fsat = \frac{K_{11}[L] + K_{21}L_{20}[R][L] + 2K_{21}K_{22}L_{20}[R][L]^2}{1 + K_{11}[L] + 2L_{20}[R] + 2K_{21}L_{20}[R][L] + 2K_{21}K_{22}L_{20}[R][L]^2}$$

Note that (2) depends on the concentration of free receptor monomer and the concentration of free ligand. The concentration of free ligand is known or approximated, but the concentration of free receptor monomer is not known and must be calculated from the total receptor concentration.

(8) $[R]_{total}=[R]+[RL]+2[RR]+2[RRL]+2[RRLL]$

(9) $[R]_{total}=[R](1+K_{11}[L])+[R]^2 L_{20}(1+K_{21}[L]+K_{21}K_{22}[L]^2)$

[R] can be found quadratically, and the equation for the positive root is simultaneously fit to the data with equation (2) in Pike's work. The only known values in this case are [L] and [R]$_{total}$. The values for K_{nn} and L_{20} are those of best fit. Pike uses "Global Fitting" by which she means nonlinear least squares fitting of each data set for each unique receptor concentration but constraining the values of K_{nn} and L_{20} to be the same for each fit. This means that the shift in binding curve is primarily driven by [R] and the saturation primarily driven by [L].

Showing Pike's model is of the general form described in (1)

Fractional saturation of monomer and dimer populations **(note that these forms are equivalent to the one and two site models (2) and (3))**

(10) $fsat_{mon}=\frac{[RL]}{[R]+[RL]}=\frac{K_{11}[L]}{1+K_{11}[L]}$

(11) $fsat_{dim}=\frac{[RRL]+2[RRLL]}{2[RR]+2[RRL]+2[RRLL]}=\frac{1}{2}\frac{K_{21}[L]+2K_{21}K_{22}[L]^2}{1+K_{21}[L]+K_{21}K_{22}[L]^2}$

Fraction of population that is monomeric (when substituted with equ. eqns. it depends on K_{11}, K_{21}, K_{22}, and L_{20})

(12) $m=\frac{[R]+[RL]}{[R]+[RL]+2[RR]+2[RRL]+2[RRLL]}$

Fraction of population that is dimeric (when substituted with equ. eqns. it depends on K_{11}, K_{21}, K_{22}, and L_{20})

(13) $1-m=\frac{2[RR]+2[RRL]+2[RRLL]}{[R]+[RL]+2[RR]+2[RRL]+2[RRLL]}$

When eqns. (10)-(13) are combined as $(m)(fsat_{mon})+(1-m)(fsat_{dim})$ they yield equation (7). This proves equivalence of forms and lets us use the criteria we established to test for cooperativity, i.e., negative cooperativity is present if $K_{22} < 4K_{21}$.

Comparison to a model without linkage

Pike's model enforces linkage between ligand binding and dimerization by way of eqn. (9). Although there is evidence of ligand-induced dimerization in EGFR studies [refs], it is useful to rule out the

simpler model where the population of monomers and dimers are governed only by receptor concentration and the dimerization constant. For this we can retain $fsat_{mon}$ and $fsat_{dim}$ as defined in eqns (10) and (11), but replace m with

$$(14) \quad m = 1 - \frac{1}{4[R_{total}]}\left(\left(L_{20}^{-1} + 4[R_{total}]\right) - \sqrt{L_{20}^{-1}}\sqrt{L_{20}^{-1} + 8[R_{total}]}\right)$$

This can be derived from the equilibrium equation

$$L_{20}^{-1} = \frac{\left([R_{total}] - [RR]\right)^2}{[RR]}$$

and the definition of the fraction of R_{total} that is dimeric

$$(1 - m) = \frac{2[RR]}{[R_{total}]}$$

NOTE: results of fitting this simpler model and comparison with Pike's model deem it to be insufficient to explain the data so it is not considered in the rest of this discussion. I'm leaving details out now for brevity but will include in future.

Extraction of data

Data for binding of EGF to Wild Type EGFR were extracted from Pike et al. 2008 (fig 3), 2009 (fig 1), and 2011 (fig 3b)

Data for binding of EGF to kinase-impaired mutants were extracted from Pike et al. 2008 (fig 5b), 2009 (fig 3a and 3b).

Extraction was done by enlarging pdfs of the publication and creating images of the figures. Those images were uploaded to WebPlotDigitizer (http://arohatgi.info/WebPlotDigitizer) and the data were digitally manually extracted. All data were scaled to nM concentrations so that extracted values for free EGF concentration ranged approximately from 0.01-10.

Data from wild type and kinase-impaired mutants were chosen for comparison because their binding isotherms shift in opposite directions, and they fit markedly different association constants, but both exhibit negative cooperativity as determined by Pike. The kinase-impaired mutants differed in mutation, but their data was similar and gave similar association constants so they were pooled for comparison (see Figure 3 – colors correspond to different mutants). Wild type data were also pooled.

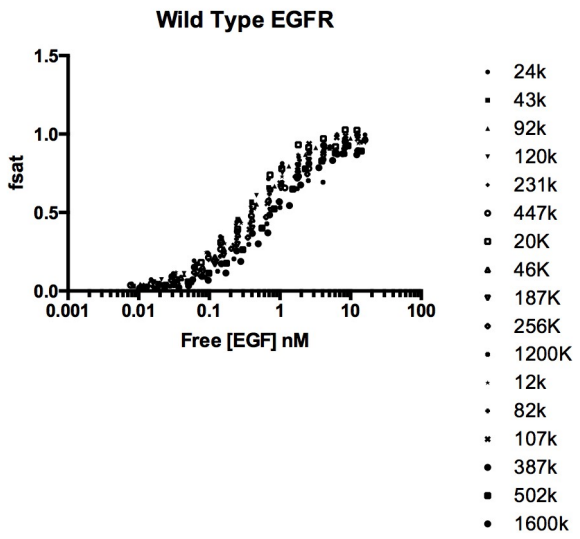


Figure 2

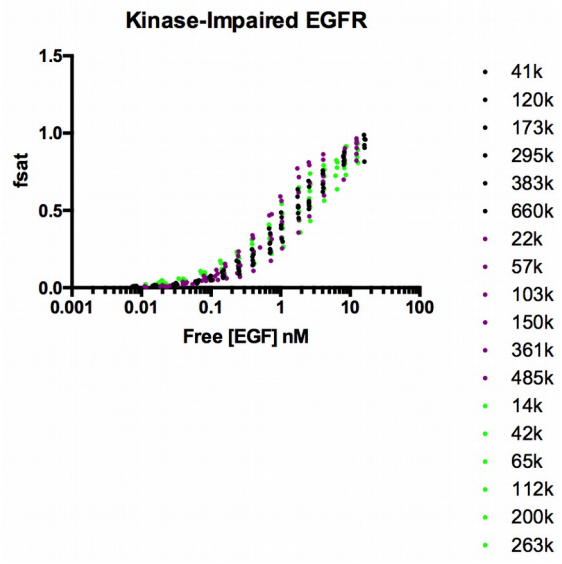


Figure 3

The extracted data sets were reverse fit with the model described by Pike using the fitted association constant values. This was done to determine the total receptor concentration, which were not reported in the publications. The concentrations were checked with estimates using the receptor/cell numbers that were reported. The fitted numbers were within a factor of 2 on average and considered valid.

Approach for model analysis and simulation

We can constrain the full model (7) such that $K_{21} = 4K_{22}$ to create a nested model that cannot account for cooperativity. The models are nested because they are identical minus a degree of freedom. We can statistically compare fits of a nested model using the extra sum of squares test to determine if the full model invoking cooperativity is a more accurate description of the system. The fits can be used to simulate data sets that assume an underlying mechanism that is not cooperative, and we can determine the effects of random variation on the chance that a non-cooperative system would be better fit by a model invoking cooperativity.

Results of fitting (units are nM^{-1}):

	Wild Type COOP			Wild Type NO-COOP		
	mean	95% conf. Int.		mean	95% conf. Int.	
K11	3.787	3.352	4.269	4.081	3.601	4.707
L20	302.6	162	520.5	475.8	267.5	792.6
K21	1.807	1.489	2.159	1.403	1.194	1.634
K22	0.1134	0.01375	0.2175	0.3507	0.2985	0.4085

	Kinase-impaired COOP			Kinase-impaired NO-COOP		
	mean	95% conf. Int.		mean	95% conf. Int.	
K11	0.1716	0.06574	0.2536	0.1339	0.02578	0.2209
L20	48.75	15.66	108	88.34	41.39	164.6
K21	5.068	3.418	8.344	2.996	2.645	3.48
K22	0.6599	0.5691	0.7879	0.7491	0.6613	0.87

These results are fits of the full (cooperative) and constrained (non-cooperative) models. I will have a much more developed discussion here but the critical points are the following:

In all cases, the raw data is best fit by the full model when compared to the constrained model using the extra sum of squares test. The fitted values are similar to Pike's but likely better estimates as the data was pooled before global fitting. The full model is given a statistical certainty of being correct with $p < 0.0001$. However, the constrained model still gives a reasonably good fit, it just describes less of the variation in the data. We can use the fitted values from the non-cooperative model to simulate data sets with noise and then compare fits between the full and constrained models to get a range on expected values of parameters if the underlying mechanism is non-cooperative and the cooperative model is describing noise. We can then compare that to the opposite case where the underlying model is truly cooperative and understand what the expected range of values for parameters are.

Simulation details

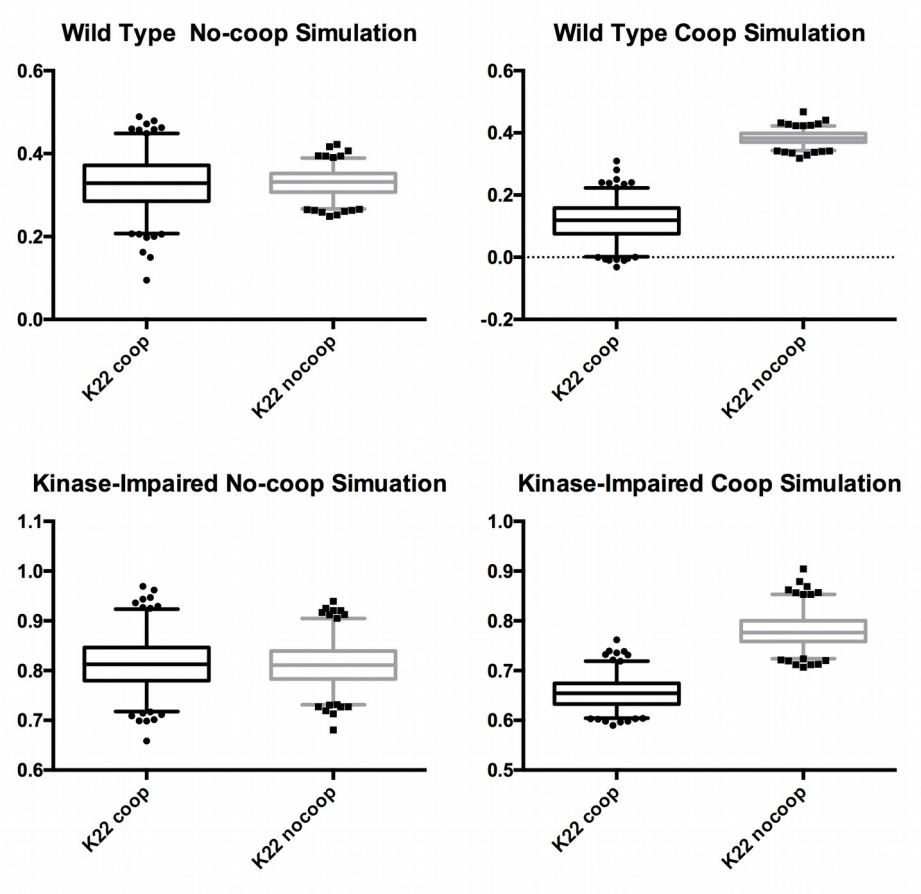
Data sets were generated using the full and constrained models across receptor concentration values of 0.001, 0.002, 0.005, 0.008, 0.01, 0.02, and 0.05 nM with 5% relative standard deviation randomly applied (chosen based on standard deviation values plotted on Pike's data). Ligand concentration ranged from 0.01-10 nM as in the actual data.

1000 data sets were generated for each model using the mean values of the four sets of parameters above. The data sets were then fit with the full and constrained models and the fits were compared with the extra sum of squares test. The fitted values were tabulated to determine distributions for comparison.

Simulation results (this is a truncated set only looking at the most important parameter for cooperativity, K_{22})

Below are plots of the distribution of fitted values for K_{22} for the different simulations. The box and whiskers represent the 2.5-97.5 percentile, and the dots are any outliers. Panel 1 contains the results of simulating a system that does not have cooperativity and then fitting with the full and constrained model. What we observe is a larger spread of values for K_{22} taken on when fitting with the full model (cooperativity allowed), but the median value is the same as that for fitting with the non-cooperative model from which the data were simulated. This means that if cooperativity was absent from the

system and given random scatter at 5% relative SD, then the mean fitted value of K_{22} would be close to $K_{21}/4$, and we would only expect to get a value of $K_{22} << K_{21}/4$ (of the magnitude observed when fitting the real data) in rare cases – statistically less than 2.5% by chance. However, in the real data, we get a value of $K_{22} << K_{21}/4$, and this is statistically likely to arise from an underlying mechanism that is negatively cooperative as we can see in panel 2, which contains the results of simulating a system that does have cooperativity and fitting it with both the full and constrained model. In that case the fitted value for K_{22} using the full model is significantly different than that obtained by fitting with the constrained (cooperativity forbidden) model. The same is seen for the kinase-impaired system.



Conclusions

Based on extra sum of squares test and simulations, the data reported by Pike does suggest the cooperativity is a salient aspect of EGF-EGFR binding and the model used is appropriate for description of the system