Data-driven Modelling of Receptor Tyrosine Kinase Signalling Networks Quantifies Receptor-specific Potencies of PI3K- and Ras-dependent ERK Activation

Murat Cirit and Jason M. Haugh

Supplementary Text S1: Computational Models and Methods

A. Kinetic model of the FGF receptor signalling network

Receptor-mediated recruitment of Ras-GEF and PI3K activities

Fractional recruitment of Ras-GEF enzymatic activity (e_{GEF}) is potentially saturable and assumed to be in quasi-equilibrium with receptor activation, which varies with time according to the kinetics of receptor binding, dimerization, and downregulation. Based on those considerations and minimizing the number of adjustable parameters, we formulated the following algebraic equation.

$$e_{GEF}(t) = \left[\frac{A_{GEF}\left(e^{-k_1 t} - e^{-k_2 t}\right)}{1 + A_{GEF}\left(e^{-k_1 t} - e^{-k_2 t}\right)}\right] f_{GEF}(t). \qquad \text{(Eq. S1)}$$

The phenomenological parameters A_{GEF} (dimensionless), k_1 (min^{-1}), and k_2 (min^{-1}) are determined by data fitting, with different values at each dose of FGF (3 parameters*2 doses = 6 fit parameters here). The fits are constrained so that these parameters are all positive, and $k_1 < k_2$. Roughly speaking, A_{GEF} controls the degree of saturation of the dose response, k_2 characterizes how quickly the system responds, and k_1 characterizes the adaptation of the response, e.g. through receptor downregulation and/or ligand depletion. The variable $f_{GEF}(t)$ is the fraction of Ras-GEF that is freely available, i.e., not desensitized by ERK (see below).

Recruitment of PI3K enzymatic activity is modeled in an analogous fashion as follows.

$$e_{PI3K}(t) = \frac{A_{PI3K}\left(e^{-k_1 t} - e^{-k_2 t}\right)}{1 + A_{PI3K}\left(e^{-k_1 t} - e^{-k_2 t}\right)}. \qquad \text{(Eq. S2)}$$

For each of the two FGF doses, the values of k_1 and k_2 are the same as in Eq. S1, and there are different values of the saturation parameter, A_{PI3K}.

Accumulation of membrane-associated messengers Ras-GTP and 3' PI lipids

The dimensionless densities of Ras-GTP (m_{Ras}) and 3' PI lipids (m_{3PI}) respond to the recruitment of Ras-GEF and PI3K activities, respectively, according to the following differential equations of identical mathematical form.

$$\frac{dm_{Ras}}{dt} = k_{Ras}\left(m_{Ras,\max} e_{GEF} - m_{Ras}\right); \quad m_{Ras}(0) = 0. \qquad \text{(Eq. S3)}$$

$$\frac{dm_{3PI}}{dt} = k_{3PI}\left(m_{3PI,\max} e_{PI3K} - m_{3PI}\right); \quad m_{3PI}(0) = 0. \qquad \text{(Eq. S4)}$$

Relative to our previous PDGF receptor signalling models [1,2], the parameters $m_{Ras,\max}$ and $m_{3PI,\max}$ are new and account for differences in the maximal activation of Ras and PI3K mediated

by FGF versus PDGF receptors in our cells (their values are equal to 1 for PDGF signalling). Another minor difference in the m_{Ras} conservation equation, Eq. S3, is the omission of a saturation parameter, Γ, which had already been set to an arbitrarily low value in previous models to reflect the observation that most of the Ras remains in the inactive, GDP-bound form.

Activation of MEK kinase activities and phosphorylation of Akt

As formulated previously, enzymatic activity directed towards the phosphorylation of MEK is comprised of Ras-dependent (x_1) and PI3K-dependent (x_2) contributions; PI3K-dependent signalling is also responsible for Akt phosphorylation (a_p), which is modeled in order to compare with those measurements. The following conservation equations are taken from our previous model [2].

$$\frac{dx_1}{dt} = k_{d,x1}\left[\frac{(1+K_{x1})m_{Ras}}{1+K_{x1}m_{Ras}} - \frac{x_1}{1+y/K_{M,x11}+y_p/K_{M,x12}}\right]; \quad x_1(0)=0. \qquad \text{(Eq. S5)}$$

$$\frac{dx_2}{dt} = k_{d,x2}\left[\frac{(1+K_{x2})m_{3PI}}{1+K_{x2}m_{3PI}} - \frac{x_2}{1+y/K_{M,x21}+y_p/K_{M,x22}}\right]; \quad x_2(0)=0. \qquad \text{(Eq. S6)}$$

$$\frac{da_p}{dt} = k_{d,a}\left[\frac{(1+K_a)m_{3PI}}{1+K_a m_{3PI}} - a_p\right]; \quad a_p(0)=0. \qquad \text{(Eq. S7)}$$

y and y_p are the unphosphorylated and mono-phosphorylated fractions of total MEK, respectively, which appear in Eqs. S5 and S6 to allow for sequestration of active x_1 and x_2 by their substrates; thus, the rate of MEK kinase deactivation would be correspondingly reduced.

Phosphorylation of MEK and ERK

For MEK (dual phosphorylated MEK fraction defined as y_{pp}), the conservation equations are as follows [2].

$$\frac{dy}{dt} = -\sum_{i=1}^{2}\frac{V_{\max,xi1}x_i f_{xi} y/K_{M,xi1}}{1+y/K_{M,xi1}+y_p/K_{M,xi2}} + \frac{V_{\max,yph1}\,y_p/K_{M,yph1}}{1+y_p/K_{M,yph1}+y_{pp}/K_{M,yph2}}; \quad y(0)=1; \qquad \text{(Eq. S8)}$$

$$\frac{dy_{pp}}{dt} = \sum_{i=1}^{2}\frac{V_{\max,xi2}x_i f_{xi}\,y_p/K_{M,xi2}}{1+y/K_{M,xi1}+y_p/K_{M,xi2}} - \frac{V_{\max,yph2}\,y_{pp}/K_{M,yph2}}{\left(1+z/K_{M,y1}+z_p/K_{M,y2}\right)\left(1+y_p/K_{M,yph1}\right)+y_{pp}/K_{M,yph2}}; \quad y_{pp}(0)=0; \qquad \text{(Eq. S9)}$$

$$y_p = 1-y-y_{pp}. \qquad \text{(Eq. S10)}$$

The variables $f_{xi}(t)$, account for the fractions of x_1 and x_2 that are not desensitized by ERK (see below). MEK phosphatase activity (e.g., PP2A) is taken to be constant (characterized by the parameters $V_{\max,yph1}$, $K_{M,yph1}$, $V_{\max,yph2}$, and $K_{M,yph2}$).

For ERK (non-, mono-, and dual-phosphorylated fractions defined as z, z_p, and z_{pp}, respectively), the conservation equations are as follows.

$$\frac{dz}{dt} = -\frac{V_{\max,y1} y_{pp}\, z/K_{M,y1}}{1 + z/K_{M,y1} + z_p/K_{M,y2}} + \frac{V_{\max,zph1}\, z_p/K_{M,zph1}}{1 + z_p/K_{M,zph1} + z_{pp}/K_{M,zph2}}; \quad z(0) = 1; \qquad \text{(Eq. S11)}$$

$$\frac{dz_{pp}}{dt} = \frac{V_{\max,y2} y_{pp}\, z_p/K_{M,y2}}{1 + z/K_{M,y1} + z_p/K_{M,y2}} - \frac{V_{\max,zph2}\, z_{pp}/K_{M,zph2}}{1 + z_p/K_{M,zph1} + z_{pp}/K_{M,zph2}}; \quad z_{pp}(0) = 0; \qquad \text{(Eq. S12)}$$

$$z_p = 1 - z - z_{pp}. \qquad \text{(Eq. S13)}$$

Relative to the previous models [1,2], the only change here is that the ERK phosphatase activity (dual-specificity phosphatases) is taken to be constant (characterized by the parameters $V_{\max,zph1}$, $K_{M,zph1}$, $V_{\max,zph2}$, and $K_{M,zph2}$). This is a simplification of the previous models, in which ERK phosphatase activity was potentially affected by ERK-dependent feedback. The justification for this simplification is two-fold: the best fit of the previous PDGF receptor network model was achieved with constant ERK phosphatase activity, and experimentally we found no relationship between the expression levels of MKP1 and MKP3 (dual-specificity phosphatases that respond in different ways to growth factor stimulation) and ERK phosphorylation [2].

Regulation of the network by ERK-dependent negative feedback

There are two distinct layers of negative feedback in the model. The first affects Ras-GEF activity through the aforementioned variable f_{GEF}, and the second affects MEK kinase activities through the aforementioned variables f_{x1} and f_{x2}. These quantities decrease in response to ERK phosphorylation according to the following, quasi-mechanistic differential equations [2].

$$\frac{df_{GEF}}{dt} = -k_{d,fG}\left[K_{fG} z_{pp} f_{GEF} - (1 - f_{GEF})\right]; \quad f_{GEF}(0) = 1. \qquad \text{(Eq. S14)}$$

$$\frac{df_{x1}}{dt} = -k_{d,fx1}\left[K_{fx1} z_{pp} f_{x1} - (1 - f_{x1})\right]; \quad f_{x1}(0) = 1. \qquad \text{(Eq. S15)}$$

$$\frac{df_{x2}}{dt} = -k_{d,fx2}\left[K_{fx2} z_{pp} f_{x2} - (1 - f_{x2})\right]; \quad f_{x2}(0) = 1. \qquad \text{(Eq. S16)}$$

Summary of model parameters and global fitting to FGF data

The equations formulated above invoke a total of 44 constant parameters, of which 4 are assigned fixed values (k_{Ras}, k_{3PI}, $k_{d,a}$, and K_a; see the previous publications) and 40 were subjected to a global fit to the available data set, which included the following 81 measurements for FGF stimulation: Ras-GTP loading kinetics, with and without MEK inhibitor (18 conditions; Fig. 3a); MEK phosphorylation kinetics, with and without PI3K inhibitor (22 conditions; Fig. 3b); ERK phosphorylation kinetics, with and without PI3K inhibitor (22 conditions; Fig. 3c) and with and without phorbol ester (8 conditions; Supplementary Fig. S2); and Akt phosphorylation kinetics (11 conditions). The latter do not significantly affect the fitting of the rest of the data and parameters, because the low levels of Akt phosphorylation stimulated by FGF are scaled relative to those stimulated by PDGF (Fig. 1b); in other words, for the overall fit it is important only that the Akt phosphorylation levels produced by the model are sufficiently low.

Calculating the dynamic MEK activation comparator (dMAC)

The dMAC is a time-dependent quantity that compares the relative contributions of PI3K- and Ras-dependent signalling converging on MEK [2]. For a given experimental

condition (here, a particular dose of FGF) and for each of the 10,000 selected parameter sets, the model is run with the Ras pathway silenced ($m_{Ras} = 0$; PI3K-dependent activation of MEK) and then with the PI3K pathway silenced ($m_{3PI} = 0$; Ras-dependent activation of MEK). The quantity $y_{pp}/(1 - y_{pp})$, reflecting the rate of MEK activation normalized by the amount of inactive MEK available, is calculated as a function of time, and the dMAC is calculated as the ratio of PI3K-dependent to Ras-dependent $y_{pp}/(1 - y_{pp})$ values.

Predicting the effect of siRNA knockdown

This calculation was performed as described previously [2]. Defining δ_z as the fractional knockdown of ERK1/2 (e.g., $\delta_z = 0.8$ corresponds to 80% reduction of intracellular ERK1/2), the prediction is implemented by multiplying or dividing the values of the following parameters by the factor, $(1-\delta_z)$, according to how they are scaled by the intracellular concentration of total ERK:

$$\text{Multipled by } (1-\delta_z)\text{: } K_{fG},\ K_{fx1},\ K_{fx2}.$$
$$\text{Divided by } (1-\delta_z)\text{: } V_{max,y1},\ K_{M,y1},\ V_{max,y2},\ K_{M,y2},\ V_{max,zph1},\ K_{M,zph1},\ V_{max,zph2},\ K_{M,zph2}.$$

B. Co-alignment of PDGF and FGF receptor signalling network models

Model and data set for the PDGF receptor network

The model of PDGF receptor-mediated ERK activation described in detail previously [2], comprised of 57 constant parameters (14 fixed, 43 fit), was used with the following simplification: the modulation of ERK phosphatase activity was neglected, such that ERK species evolve according to Eqs. S11-S13 given above. Together with the modifications prescribed in Eqs. S3 and S4 (removal of the fixed parameter Γ and addition of parameters $m_{Ras,\max}$ and $m_{3PI,\max}$, which are fixed to values of 1 for PDGF), this results in the elimination of 11 fit parameters and the addition of 1 fixed parameter, leaving a total of 47 constant parameters (32 fit).

The data used for aligning the PDGF receptor network model were presented previously [1,2] and include the following 209 measurements: Ras-GTP loading kinetics, with PI3K inhibited, MEK inhibited, or control (21 conditions); MEK phosphorylation kinetics, PI3K inhibited versus control (42 conditions) and with dominant-negative (S17N) Ras versus control (42 conditions); ERK phosphorylation kinetics, PI3K inhibited versus control (42 conditions), dominant-negative (S17N) Ras versus control (42 conditions), and with and without phorbol ester (20 conditions). As explained in the main text of the paper, the PDGF and FGF data are co-normalized with each other based on the side-by-side kinetic data presented in Fig. 1 and have common conversion factors for alignment of each experimental readout to the corresponding model output (see also Section C below).

Model variations used to simultaneously fit FGF and PDGF stimulation data

In the FGF receptor network model outlined in Section A, Eqs. S1 and S2 and their 8 adjustable parameters are unique to FGF receptor-mediated signalling, as are the FGF-specific values of $m_{Ras,\max}$ and $m_{3PI,\max}$ in Eqs. S3 and S4, respectively. The remaining 34 parameters in the FGF receptor network model, of which 30 are fit, are common with the PDGF receptor

signalling model. Taken together, the minimum number of adjustable parameters needed to align both FGF and PDGF stimulation data is 57 (42 fit).

Variation 1: All common parameters the same. The simplest and most restrictive model variation is one in which all 30 of the common fit parameters alluded to above are constrained to have the same values for FGF and PDGF signalling; each parameter set is forced to best reconcile both sets of data.

Variation 2: Distinct MEK activation kinetics. All parameters are constrained to be the same for FGF and PDGF signalling except the following 8 MEK phosphorylation parameters: $V_{max,x11}$, $K_{M,x11}$, $V_{max,x12}$, $K_{M,x12}$, $V_{max,x21}$, $K_{M,x21}$, $V_{max,x22}$, and $K_{M,x22}$. This allows the Ras- and PI3K-dependent inputs to MEK to be more or less potent depending on where and how they are activated in response to FGF versus PDGF.

Variation 3: Distinct MEK and ERK activation kinetics. Here, there are 12 parameters that are allowed to vary between FGF- and PDGF-stimulated signalling: the 8 parameters listed under Variation 2 and the 4 ERK phosphorylation parameters ($V_{max,y1}$, $K_{M,y1}$, $V_{max,y2}$, and $K_{M,y2}$). This is the most complicated model variation in terms of adjustable parameters and considers the possibility that the two receptors have differential accessibility to scaffold proteins that hold MEKK, MEK, and ERK species in the same complex, for example.

Variation 4: Distinct ERK activation kinetics. Here, only the 4 ERK phosphorylation parameters listed under Variation 4 are allowed to vary between FGF- and PDGF-stimulated signalling. This model variation is the least plausible but serves as a control of sorts for the comparisons among Variations 1-3.

C. Review of Monte Carlo algorithm

A large ensemble of parameter sets was obtained using a modified simulated annealing algorithm described in detail previously [1,2]. The acquisition and use of the ensemble is reviewed here.

1) Initial guesses of the parameter values were chosen. We confirmed that different initial guesses did not qualitatively change the ultimate parameter value distributions.

2) Given an array of parameters $\mathbf{k}_i$ for iteration i, the differential equations were solved numerically using the stiff solver ode15s, generating the kinetics for all variables as a function of time.

3) Using a branch-and-bound subroutine, we estimated a factor that converts the model output to the arbitrary experimental units for each readout j (Ras-GTP, ppMEK, ppERK, and pAkt), such that the sum of squared deviations between measured and calculated values at each data point, SSD_{ij}, is minimized. As done previously, the PDGF stimulation data for the different readouts were renormalized so that the means of the values for the 1 nM PDGF, control (DMSO and empty vector) time courses are all equal to 1, setting the arbitrary units of the different readouts on a common scale. As stated above, the FGF stimulation data were renormalized so as to be consistent with side-by-side comparisons between 1 nM PDGF-BB and 1 nM FGF-2. The subroutine iteratively subdivides the range of possible values until each SSD_{ij} can no longer be reduced by more than 0.1%.

4) The cumulative sum of squared deviations, $cSSD_i$, was calculated:

$$cSSD_i = \sum_j w_j SSD_{ij}\,.$$

Since the data types were already normalized in a consistent way, we used $w_j = 1$.

5) Each parameter k_i was updated according to

$$k_{i+1} = k_i\left(1 + \alpha\,\mathrm{randn}\right),$$

where randn is a random number drawn for each parameter from a standard normal distribution. Thus, the value of α determines how much the values of the parameters tend to change from one iteration to the next; a value of $\alpha = 0.05$ was used here. If any of the new parameter values was below 10^{-4} or greater than 10^4, the new value was thrown out, and another value was drawn based on the previous value.

6) For the new set of parameters $\mathbf{k}_{i+1}$, steps 2-4 were repeated to obtain $cSSD_{i+1}$.

7) If $cSSD_{i+1} < cSSD_i$ (improved fit), the new set of parameters $\mathbf{k}_{i+1}$ was accepted, and i incremented; otherwise, it was accepted with probability

$$p_{i+1} = \exp\left[-\frac{\left(cSSD_{i+1} - cSSD_i\right)}{T_i}\right];$$

T_i is called the "temperature" for iteration i; it determines how forgiving the algorithm is if the fit does not improve. As explained previously [2], our modification of the standard simulated annealing approach is to tie T_i to the current error metric,

$$T_i = \beta \cdot cSSD_i.$$

Thus, once the value of $cSSD$ approaches its minimum value, the algorithm operates at approximately constant temperature; a value of $\beta = 0.01$ was used. If the new parameter set was rejected, the algorithm proceeded with the previous parameter set $\mathbf{k}_i$.

8) Steps 5-7 were repeated until at least 50,000 parameter sets were accepted in total, and the best 10,000 of these (those with the lowest $cSSD$ values) were identified as the parameter set ensemble used to generate computational results. These parameter sets and their associated conversion factors were saved in a matrix for further analysis.

9) For the model variables compared to the data, the aligned model output was recomputed for each of the 10,000 parameter sets in the ensemble, and the ensemble mean and standard deviation were calculated for each time point (1-minute intervals).

D. References

1. Wang C-C, Cirit M, Haugh JM (2009) PI3K-dependent crosstalk interactions converge with Ras as quantifiable inputs integrated by Erk. Mol Syst Biol 5: article no. 246.
2. Cirit M, Wang C-C, Haugh JM (2010) Systematic quantification of negative feedback mechanisms in the extracellular signal-regulated kinase (ERK) signaling network. J Biol Chem 285: 36736-36744.