

# Optimal experimental design in an epidermal growth factor receptor signalling and down-regulation model

F.P. Casey, D. Baird, Q. Feng, R.N. Gutenkunst, J.J. Waterfall, C.R. Myers, K.S. Brown, R.A. Cerione and J.P. Sethna

**Abstract:** We apply the methods of optimal experimental design to a differential equation model for epidermal growth factor receptor signalling, trafficking and down-regulation. The model incorporates the role of a recently discovered protein complex made up of the E3 ubiquitin ligase, Cbl, the guanine exchange factor (GEF), Cool-1 ( $\beta$ -Pix) and the Rho family G protein Cdc42. The complex has been suggested to be important in disrupting receptor down-regulation. We demonstrate that the model interactions can accurately reproduce the experimental observations, that they can be used to make predictions with accompanying uncertainties, and that we can apply ideas of optimal experimental design to suggest new experiments that reduce the uncertainty on unmeasurable components of the system.

## 1 Introduction

The epidermal growth factor receptor (EGFR) is a transmembrane tyrosine kinase receptor which becomes activated upon binding of its ligand, epidermal growth factor (EGF) and signals via phosphorylation of various effectors [1]. Besides sending signals to downstream effectors, the activated EGFR also will initialise endocytosis which is followed by either degradation or recycling of the receptor. These are the normal receptor down-regulation processes. Persistence of activated receptor on the cell surface can lead to aberrant signalling and transformation of cells [2]. In addition, a variety of tumour cells exhibit overexpressed or hyperactivated EGF receptor [3, 4], indicative of the failure of normal receptor down-regulation.

We concern ourselves with building a mathematical model of the receptor endocytosis, recycling, degradation and signalling processes that can reproduce experimental data and incorporates the effects of regulating proteins that themselves become active after EGF stimulation. The schematic for the model is shown in Fig. 1. In particular,

we examine the roles of the GEF, Cool-1 and the GTPase, Cdc42, that have recently been discovered to be important for EGFR homeostasis [5,6] through their interaction with the E3 ubiquitin ligase, Cbl.

There is evidence for two interaction mechanisms which disrupt the normal receptor down-regulation. The first mechanism involves the formation of a complex between active Cool-1, active Cdc42 and Cbl. After activation of the receptor, Cool-1 becomes phosphorylated through a Src-FAK phosphorylation cascade. Phosphorylated Cool-1 has GEF activity and in turn activates Cdc42 by catalysing the exchange of guanosine diphosphate (GDP) for guanosine triphosphate (GTP). Unlike other GEFs, however, activated Cool-1 can remain bound to its target, Cdc42, [6] and can then form a complex with Cbl (mediated through Cool-1 binding), effectively sequestering Cbl from the receptor. Therefore the internalisation and degradation of the receptor is inhibited and its growth signal is maintained. (We use the ERK pathway as a readout on the receptor mitogenic signal.) The second mechanism is based on the findings of [5] that activated Cool-1 can directly bind to Cbl on the receptor and block endocytosis in a manner we hypothesise to be analogous to the action of Sprouty2 [7].

To maintain normal receptor signalling, we postulate it is crucial that deactivation of Cool-1 and subsequent dissociation of the Cbl, Cool-1 and Cdc42 complex occur. Then Cbl can induce receptor internalisation and ubiquitin tag it for degradation in the lysosome. Internalised receptor lacking ubiquitin moieties can be returned to the cell surface from the early endosome via the recycling pathway.

The role of Cbl in the degradation mechanism for the receptor has been understood for some time [8–10]. However, its function in mediating endocytosis still remains controversial (e.g. [11–15]) as the receptor can be internalised through more than one endocytic pathway. We do not address that issue here but rather we assume in our model that Cbl association and activation is necessary for endocytosis, whether through a CIN85-endophilin interaction [16] or through ubiquitination of the receptor [15] and therefore we do not include a separate Cbl-independent endocytosis pathway. The overall set of these protein–protein interactions is summarised in Fig. 1 (we also incorporate phosphatases in the model to act on the various

© The Institution of Engineering and Technology 2007

doi:10.1049/iet-syb:20060065

Paper first received 15th September 2006 and in revised form 14th February 2007

F.P. Casey was with the Center for Applied Mathematics, Cornell University, Ithaca, NY 14853, USA, and is now with UCD Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Belfield, Dublin 4, Ireland

D. Baird is with the Department of Cellular and Molecular Medicine and the Howard Hughes Medical Institute, University of California at San Diego, La Jolla, CA 92093, USA

Q. Feng and R.A. Cerione are with the Department of Molecular Medicine, College of Veterinary Medicine, Cornell University, NY 14853, USA

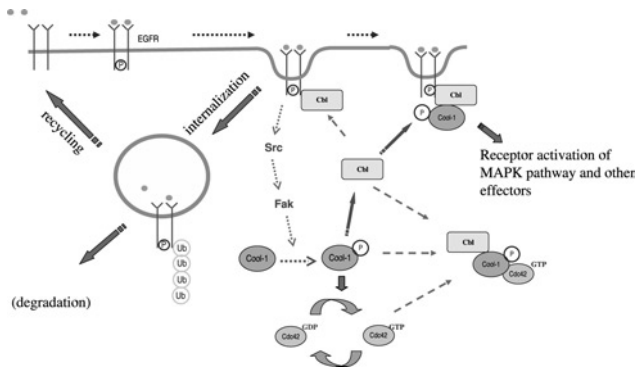
R.N. Gutenkunst, J.J. Waterfall and J.P. Sethna are with the Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, NY 14853, USA

J.J. Waterfall was with the Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, NY 14853, USA, and is now with the Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA

C.R. Myers is with the Cornell Theory Center, Cornell University, Ithaca, NY 14853 USA

K.S. Brown is with the Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA

E-mail: ferg@cam.cornell.edu



**Fig. 1** Schematic diagram showing the set of interactions in the model of EGFR signalling, endocytosis and down-regulation (see also [5])

Phosphatases are not shown

phosphorylated species, but this is not shown in the network figure). There is a significant overlap between our model and previous models of EGF receptor signalling and/or trafficking [17–20]. Since we wish to focus on the role of the Coxs-1/Cdc42 proteins within the network and to demonstrate the utility of optimal experimental design, we leave out some of the known intermediate reactions involved in the MAPK and EGFR-Src activation pathways, preferring a ‘lumped’ description which is more computationally manageable.

The goals of this manuscript are to demonstrate how a modelling approach can be used to

- (a) refine the necessary set of interactions in the biological network,
- (b) make predictions on unmeasured components of the system with good precision, and
- (c) reduce the prediction uncertainty on components that are difficult to measure directly, by using the methods of optimal experimental design.

## 2 Methods

### 2.1 Mathematical model, parameter and prediction uncertainties

Before we introduce the algorithms needed to address the design question, we define the model and data in more detail. Our differential equation model for EGFR signalling and down-regulation contains 56 unknown biochemical constants: 53 unknown rate and Michaelis Menten constants (where they can be found, initial estimates were drawn from the literature), and three unknown initial conditions which we found useful to vary. The dynamical variables comprised 41 separate chemical species, including complexes. The data consist entirely of time series in the form of Western blots. (The data both come from the lab of the co-authors and from the literature, see supplementary information for details.) We have been careful to select data only on NIH-3T3 cells, and in experimental conditions where the cell has been serum-starved prior to EGF stimulation, to prevent activation events not related to the EGFR ligand binding. Most of the time series data are over a period less than a few hours which allows us to ignore transcriptional processes.

Since we have no information on most of the biochemical constants, we must infer them from the data. Therefore we optimise a cost function which measures the discrepancy of

simulated data from the real data

$$C(\theta) = \sum_{\alpha=1}^D \sum_{i=1}^{m_{\alpha}} \left( \frac{y_{\alpha}(t_{\alpha i}, \theta) - d_{\alpha i}}{\sigma_{\alpha i}} \right)^2 \quad (1)$$

where  $\alpha$  is an index on the  $D$  measured species,  $m_{\alpha}$  is the number of time points on species  $\alpha$ ,  $y$  is the trajectory of the differential equation model,  $\theta$  is a vector of the logarithm of the biochemical constants,  $d_{\alpha i}$  is the measured value at time  $t_{\alpha i}$  for species  $\alpha$  and  $\sigma_{\alpha i}$  is the error on the measured value. In other words, we have a standard weighted least squares problem to reduce the discrepancy of the model output to the data by varying  $\theta$ . (We use the logarithm of the biochemical constants as it allows us to apply an unconstrained optimisation method while maintaining the positivity constraint and it removes the discrepancies between biochemical values that have naturally different scales in the problem). As absolute numbers of proteins in the network cannot be accurately measured, data sets measuring activities of proteins are fit up to an arbitrary multiplicative scale factor, which adds parameters to the model not of direct inferential interest (nuisance parameters). Where the relative quantity of a species can be measured (normalised by the level before EGF stimulation e.g.), the output of the differential equations is similarly scaled by an appropriate common factor.

After the model has been successfully fit to the experimental data, we have a parameter estimate  $\hat{\theta}$  which in general will have large covariances, approximated by the inverse of the Fisher information matrix (FIM). The FIM is defined as

$$M = E[\partial^2 C / \partial \theta^2] \quad (2)$$

$$= \sum_{\alpha=1}^D \sum_{i=1}^{m_{\alpha}} \frac{1}{\sigma_{\alpha i}} \frac{\partial y_{\alpha}(t_{\alpha i}, \theta)}{\partial \theta} \bigg|_{\hat{\theta}} \frac{1}{\sigma_{\alpha i}} \frac{\partial y_{\alpha}(t_{\alpha i}, \theta)}{\partial \theta} \bigg|_{\hat{\theta}} \quad (3)$$

$$= J^t J \quad (4)$$

where the expectation is over the distribution of errors in the data, which are assumed to be Gaussian. The expression for the FIM above is exact when the model fits perfectly, that is at the best fit, the expectation of the residuals is zero,  $E[y_{\alpha}(t_{\alpha i}) - d_i] = 0$ . The  $i$ th parameter uncertainty is given by the square root of the  $i$ th diagonal element of the inverse FIM.  $J = (1/\sigma_{\alpha i}) \partial y_{\alpha}(t_{\alpha i}, \theta) / \partial \theta |_{\hat{\theta}}$  is the sensitivity matrix of residuals with respect to parameters at the best fit and is the analog to the design matrix in a linear regression setting. The design space is the range of species  $\alpha$  and of time points  $t_{\alpha i}$  for which measurements could be taken. ( $\alpha i$  is the row index of  $J$ .)

We can also make predictions on components of the trajectory (measured or unmeasured),  $\hat{y}_{\beta}(t) = y_{\beta}(t, \hat{\theta})$ . The variances on these quantities are given by

$$\text{Var}(\hat{y}_{\beta}(t)) \simeq \frac{\partial y_{\beta}(t, \theta)}{\partial \theta} \bigg|_{\hat{\theta}} M^{-1} \frac{\partial y_{\beta}(t, \theta)}{\partial \theta} \bigg|_{\hat{\theta}} \quad (5)$$

The form of (5) can be thought of as a combination of the underlying parameter uncertainty, quantified by  $M^{-1}$ , and the linear response of the system to the parameter uncertainty, quantified by the sensitivities. Note that  $M$  is also computed using the sensitivities of the trajectory of the differential equations, which we obtain by implementing the forward sensitivity equations [21]. In practice,  $M$  is close to singular if we do not include some prior information on parameter ranges. Therefore we assume a Gaussian prior on the parameters centred on the best fit values, and with a standard deviation of  $\log(1000)$ .

(This corresponds to an approximately 1000-fold increase or decrease in the non-logarithmic best fit biochemical values.)

We recognise that there can be other sources of uncertainty in predictions, for example if the dynamics of the system are modelled stochastically or if there is model uncertainty that needs to be taken into account. The former is not relevant here as the measurements we fit are not on the single cell level, but rather the average of large populations of cells. The latter is certainly of interest but we choose an approach where model errors are corrected during the fitting and validation process, rather than included a priori in the model definition.

Given the approximate nature of variance estimates derived from the FIM and the linearised model response, we supplemented these calculations with a computationally intensive Bayesian Markov chain Monte Carlo (MCMC) method to compute credible intervals for the predictions we make on the model (see supplementary material). The estimates from the Bayesian MCMC approach are in sufficient agreement with the linearised error analysis results that we believe that the optimal experimental design algorithms introduced below are justifiably aimed at reducing the approximate uncertainties of (5). Using MCMC for error estimates within the framework of the optimal design algorithms would be computationally infeasible.

## 2.2 Optimal experimental design

Optimal experimental design is a technique for deciding what data should be collected from a given experimental system such that quantities we wish to infer from the data can be done so with maximum precision. Typically the network as shown in Fig. 1 has components that can be measured (e.g. total levels of active Cdc42, total levels of surface receptor etc.) and components that are not directly measurable (e.g. levels of the triple complex comprising Cool-1, Cdc42 and Cbl). Therefore we can pose the question of how to minimise the average prediction uncertainty on some unmeasurable component of interest by collecting data on measurable components of the system (we will use the term unmeasurable loosely for the remainder to describe species that are between difficult and impossible to measure by standard methods). Minimising the average variance in predictions is a design criterion called V-optimality in the literature [22]. Other authors [23–25] have focused on reducing parameter uncertainty but we believe that complex biological models, even with large amounts of precise time series data, have intrinsically large parameter uncertainty [26–28]. On the other hand, even with no extra data collection, the uncertainty on unmeasured time trajectories in these biological systems can be surprisingly small despite the large parameter uncertainty [27].

By altering the form of the matrix  $J$  in (4), through measurements on different species at different times, we have the possibility of reducing the average variance of  $\hat{y}_\beta$ , which is an integral over time of the quantity defined in (5). We discuss two types of design, continuous design and sequential design.

Continuous design refers to the selection of a design measure,  $\eta$ , which is equivalent to a probability density over the design space. For a linear model described by  $y = f(t)^t \theta + \epsilon$  where  $f(t) \in \mathbf{R}^N$  and  $\epsilon$  is an error term, the FIM is

$$M(\eta) = \int_{\tau} f(t)f(t)^t \eta(x) dt$$

by definition of the design measure,  $\eta$ . However,  $M$  is a symmetric  $N \times N$  matrix made up of a convex combination of the rank one symmetric matrices,  $f(t)f(t)^t$ . Therefore it can be represented by a convex combination of at most  $N(N+1)/2$  design points (from Caratheodory's Theorem)  $x_1, \dots, x_{N(N+1)/2}$ , that is as a convex combination of delta function probability measures on those points. In other words even continuous optimal designs for linear models have only a finite number of design support points [29]. In one of the approaches that follows, we will attempt to find a continuous design by approximating the design measure by a number of finely spaced measurement points with weights associated with each one, and we will see that a near optimal design is in fact only supported on a small subset of those points.

Sequential designs are more relevant to the situation we consider here: experimental data have already been collected and the model has already been fit. Therefore we can get an initial estimate for the parameters in the system and we can evaluate the FIM. Suppose that the current design already has  $n$  points and the current FIM is  $M_n = J_n^t J_n$ . The effect of adding the  $(n+1)$ -th design point (e.g.  $y_\alpha$  at time point  $t_{\alpha i}$ ) merely adds a single row to  $J_n$ . Therefore the new FIM is the old FIM plus a rank one update

$$M_{n+1} = J_{n+1}^t J_{n+1} = J_n^t J_n + \frac{\partial y_\alpha(t_{\alpha i})}{\partial \theta} \left| \frac{\partial y_\alpha(t_{\alpha i})}{\partial \theta} \right|_{\hat{\theta}}$$

The new inverse FIM is also a sum of terms (by applying the Sherman-Woodbury-Morrison formula [30]): one involving the inverse of the old FIM and the other involving the sensitivity vector at the new point,  $\partial y_\alpha(t_{\alpha i})/\partial \theta|_{\hat{\theta}}$ , so evaluating (5) for a large number of proposed measurements is computationally inexpensive.

We take an approach which is a combination of continuous design and sequential design: assume that some initial experiments have already been carried out and we have an FIM for the system. We will then define a cost function  $K(\alpha, t_{\alpha i})$  based on the integral of (5) and minimise it with respect to  $\alpha$  and  $t_{\alpha i}$ . Initially the minimisation looks for the best single data point to reduce the uncertainty (a sequential design method). Once we know for which species the data needs to be collected, we can then place many potential measurements on that species with associated weights and minimise over the weights (to mimic continuous design methods where the set of weights is the approximate design measure).

## 3 Results

### 3.1 Model refinements

The model was fit to 11 data sets, all Western blot data that describe various signalling, internalisation and degradation events that are triggered after receptor activation by ligand (see supplementary information for the full set of fitted time series and description of experiments).

During the iterative process of fitting and model refinement we discovered certain interactions and model parameters had to be adjusted to be consistent with the experimentally observed behaviour. These can be viewed as predictions of putative interactions that emerge from the modelling process and are incorporated into the final set of dynamical equations listed in the Appendix. We briefly summarise these adjustments below.

1. It appears necessary to incorporate an interaction to allow the triple complex to be dissociated by a dephosphorylation reaction. In particular, a reaction was needed whereby Cool-1 within the complex could be inactivated by its own dedicated phosphatase (a possible candidate already present in the system is SHP-2, which has been shown to dephosphorylate the related Sprouty protein [31]). Without this effect, we would not observe the complete deactivation of Cool-1 as it would be 'protected' within the triple complex. Additionally, a sensitivity analysis to determine dominant reactions in the model identified phosphatase reactions as important (see supplementary material).

2. Interestingly, there is an important balance between the level of receptor and Cool-1 in the system to maintain the correct dynamics: if the level of receptor greatly exceeds the Cool-1 level, then the activated receptor will lead indirectly to phosphorylation of Cool-1 which in turn sustains the level of signalling receptor before significant amounts can be endocytosed.

3. The F28L fast cycling (hyperactive) mutant of Cdc42 has the ability to delay endogenous receptor down-regulation for many hours beyond wild type cells (see experiment 5 in the supplementary information). This is only possible if the binding affinity of active Cdc42 to the Cool-1-Cbl complex is strong enough to deplete the levels of the latter and force the forward binding reaction of Cbl to activated Cool-1. This provides a mechanism to sequester more of the Cbl protein (in both the triple complex and the Cool-1-Cbl complex) than would otherwise be possible.

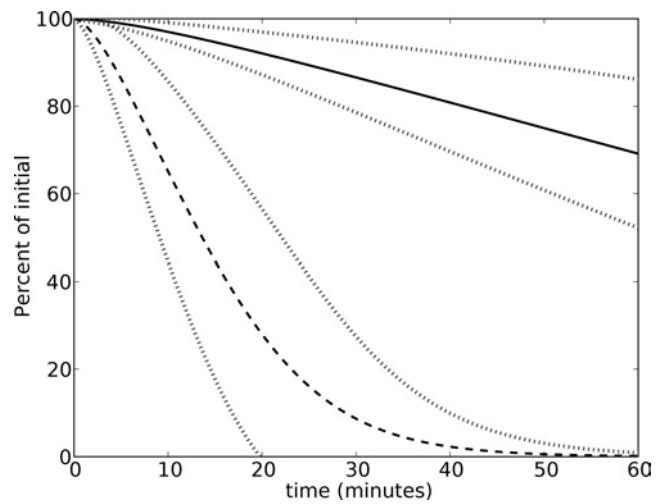
In addition to the above adjustments, we made the following observations relating to the network dynamics and structure.

We find that given these experimental data sets, a single endocytosis mechanism which is Cbl dependent and solely acts on activated receptors is sufficient to describe the available data on EGFR trafficking in NIH-3T3 cells. We acknowledge that there is much controversy in the literature as to the dominant endocytosis mechanisms and required regulators, in general.

Despite the apparently earlier activation of Cdc42 than its putative GEF, active Cool-1 (see experiments 10 and 11 in supplementary information), the data still support a mechanism whereby Cdc42 activation only occurs through Cool-1. The explanation of this effect is that the level of Cool-1 is significantly higher than Cdc42. Then, although only a fraction of Cool-1 is being activated at early times, it is still sufficient to induce substantial activation of Cdc42. In particular, we found that there was no need to invoke another parallel activation mechanism for Cdc42 (through Vav e.g.) as initially might have been assumed.

### 3.2 Predictions

Once we have a model which reproduces the experimental observations, we would like to make predictions on unmeasured or unmeasurable components of the system. The motivation is twofold. Firstly, if we make a prediction on a currently unmeasured component of the system which is subsequently measured, we have an opportunity to test the validity of our model. Secondly, if we are confident in the model, we may want to test a hypothesis about the role of an unmeasurable component in the system. If that unmeasurable component has large uncertainties, we then need to apply the methods of experimental design to improve the situation. We will discuss these issues in what follows.



**Fig. 2** Total surface receptor numbers after EGF stimulation in stably expressing v-Src cells

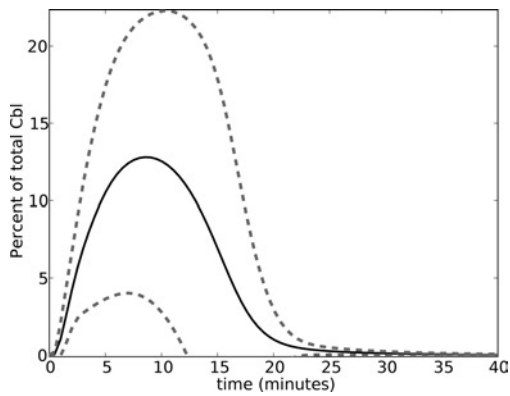
Endogenous levels of Cool-1 (dashed curve) or overexpressed Cool-1 (solid curve). The dotted lines show the uncertainties in each of the best fit predictions

**3.2.1 Model validation:** To first give an example of model validation, consider the qualitative observation in [5] that in stably expressing v-Src cells, in conditions where Cool-1 is overexpressed, ligand-induced receptor internalisation is blocked compared to an endogenous Cool-1 control, for at least 60 min. The model is adjusted to simulate the conditions of these v-Src cells by making all Src in its active form, switching off Src inactivation and increasing the initial amounts ten fold to mimic the stable transfection. We then predict the total surface receptor number under the two conditions and assign uncertainties using (5). The results are shown in Fig. 2.

The qualitative observation of strong inhibition of internalisation under conditions of overexpressed Cool-1 is verified by the model. Note that in this case the uncertainties are small enough that we can confidently predict a large difference in the fraction of receptors on the cell surface after 60 min under the two conditions. Interestingly, the model also predicts that this inhibition is much weaker in cells that are not stably expressing v-Src, essentially because the Cool-1 is not 'pre-activated' and endocytosis of significant numbers of receptors can occur before the pool of Cool-1 can become phosphorylated.

**3.2.2 Optimal design for the triple complex:** Another question of interest is whether the triple complex, which appears to be responsible for sequestering Cbl and blocks receptor down-regulation when Cdc42(F28L) is expressed, also forms in appreciable amounts in wild type cells. Since the triple complex is an example of a species that is very difficult to obtain an accurate set of measurements for, we can test a hypothesis about its formation in wild type cells by looking at its predicted time course, Fig. 3.

The relative amount of the triple complex is shown in Fig. 3, where the number of molecules of the triple complex has been scaled relative to the total level of Cbl. Relative levels of complexes and the times of formation/dissociation are more meaningful quantities than absolute numbers of molecules, which are merely rough estimates used to initialise the simulations. The best fit trajectory for the triple complex suggests that at a maximum over 12% of Cbl is sequestered in the complex which represents a significant proportion. However the uncertainty bounds are too large to make this assertion; at the level of the



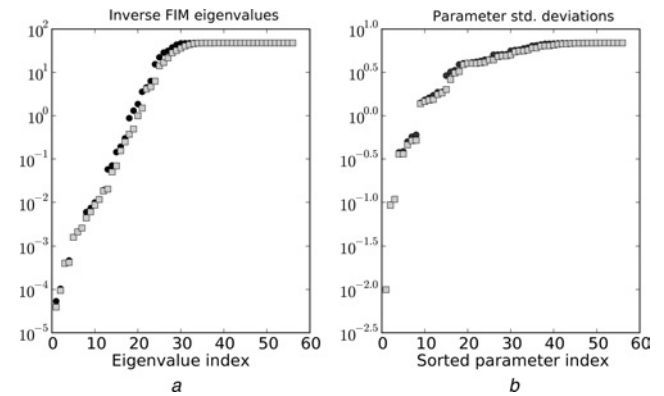
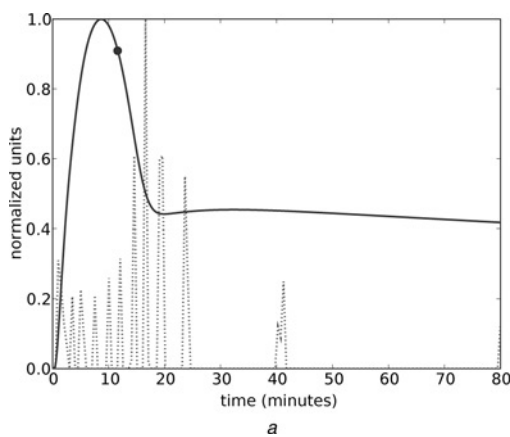
**Fig. 3** Predictions with uncertainty on the time course of the triple complex consisting of active Cool-1, Cbl and active Cdc42  
The quantity plotted is the percentage of total Cbl that is bound in the triple complex

lower bound, less than 4% of Cbl is sequestered at a maximum, and the triple complex dissociates within 15 min. This motivates the need for an optimal design approach. We define a criterion which is the average uncertainty in the prediction on the triple complex. We then optimise this quantity using a sequential design approach (therefore we need to perform only line minimisations in the time coordinate for each of the 11 measurable species in the system) and follow up by finding an approximate optimal continuous design on that species. The results of such an analysis are shown in Fig. 4.

The most striking features of the optimal design results are that

1. a single measurement on total active Cdc42 can significantly reduce the variance we see in the prediction on the triple complex, as in Fig. 4b;
2. ‘even though the approximate continuous design allows for 160 hypothetical measurements on the activity of Cdc42, the optimal design weights are concentrated to just a dozen early time points. That is, by just taking a few measurements we can get a design very close to the optimal continuous design for measuring total active Cdc42.

It is worth noting here that these extra measurements have little effect on the parameter uncertainty as shown in Fig. 5. Conversely, if we include hypothetical measurements on the



**Fig. 5** Inverse FIM eigenvalues and parameter standard deviations

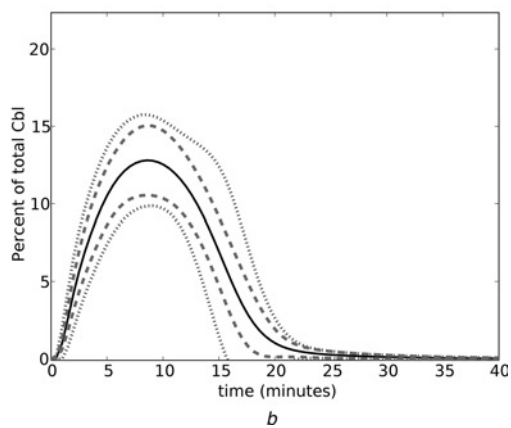
*a* Eigenvalues of the approximate parameter covariance matrix,  $M^{-1}$ , with (light squares) and without (dark circles) the optimally designed data to reduce the uncertainty of the triple complex trajectory

*b* Individual parameter standard deviations, sorted from smallest to biggest with (light squares) and without (dark circles) the optimally designed data. Note that the cutoff in the spectrum of eigenvalues is because of the prior information assumed on parameters ranges. Even with prior information, 40 of the 60 parameters have uncertainties corresponding to a greater than 20-fold increase or decrease in their non-logarithmic values

binding and unbinding constants involved in forming the triple complex, we find only a negligibly small decrease in the uncertainty in the prediction of the triple complex (see supplementary information). This is not so surprising when we understand that the uncertainty arises from the uncertainties in components of the system upstream of the triple complex; using parameter measurements alone, almost every rate constant in the system would have to be measured accurately to constrain the prediction [27].

### 3.2.3 New measurements on total active Cdc42:

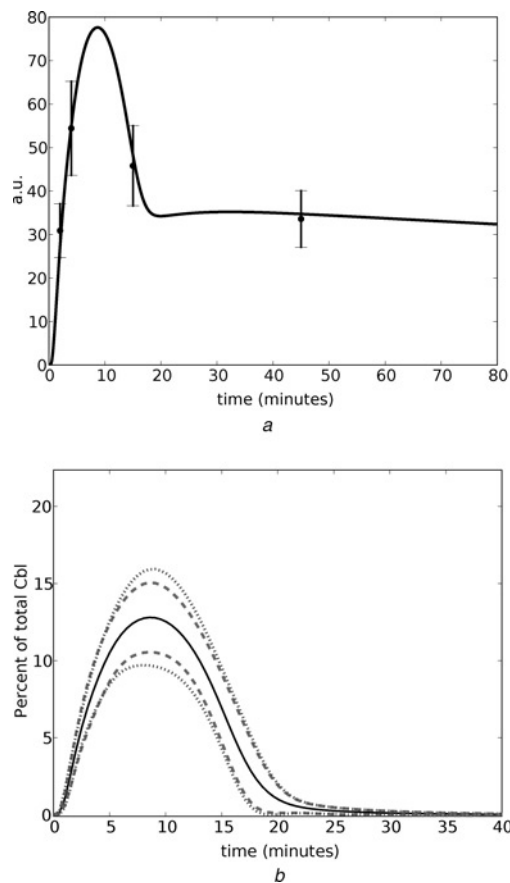
Further measurements were made on total activated Cdc42 in the lab by Western blotting and with no refitting, our model was able to match the new data using a scale factor alone, see Fig. 6a. (However, we cannot consider this as a validation of our model, since prior to the inclusion of the new data, the uncertainties on total activated Cdc42 were very large. Any experimental observations within the uncertainty bounds would be consistent with the model.) The uncertainties of the triple complex time



**Fig. 4** Optimal designs points on total active Cdc42 and reduced prediction uncertainty for the triple complex

*a* Trajectory of total active Cdc42 (solid line) with single sequential design measurement (marked with a dot) and approximate continuous design weights (dotted line) to reduce the average variance of the prediction on the Cool-1, Cbl, Cdc42 complex. The weights are optimised over 160 uniformly spaced hypothetical measurements placed between 0 and 80 min on Cdc42

*b* Shows the reduction in the original uncertainty bounds resulting from the single measurement (dotted line) and the approximate continuous design measurement (dashed line) in *a*. Compare with Fig. 3 before the addition of new measurements



**Fig. 6** Fit to total active Cdc42 and reduced prediction uncertainty for the triple complex.

*a* Without refitting to the new total active Cdc42 data, our prediction matches the data using only a single multiplicative factor. a.u. = arbitrary units

*b* Reduced uncertainty on the time course of the active Cool, Cbl and active Cdc42 complex for the optimal set of design points (dashed line) (same as Fig. 4b) and for the real data (dotted line)

course, given the real data and the optimally weighted data, is shown in Fig. 6b. Importantly, given that the measured activities of total Cdc42 were consistent with the trajectory for the optimised set of parameters, the reduction in uncertainty of the triple complex for the real data is comparable to that for the optimally selected data and we can make a firm conclusion that the triple complex does sequester significant amounts of the Cbl protein even in wild type cells after EGF stimulation. Therefore it appears that the complex plays a part under normal conditions in the EGFR homeostasis. (Note that if the new data collected showed a very different time course than in Fig. 6a, an additional re-optimisation step would need to be performed before we could assess the prediction and uncertainties for the triple complex.)

#### 4 Discussion

We have demonstrated that by quantitatively modelling the dynamics of EGFR signalling and down-regulation in a mammalian cell line, we are led to incorporate interactions and modify existing reactions in order to reproduce the experimental observations. Note that these interactions are not directly tested by experiments, but we can infer them from the existing data. This refinement of an existing model of network interactions and parameters is one important aspect of the modelling effort and gives insight into the underlying dynamics. Of course, we recognise that the model as it stands will only explain the behaviours observed

in the data sets we have chosen. The addition of new experiments that test for receptor signalling from early endosomes [32], alternative endocytic mechanisms [15], autocrine signalling [33, 34] or the interactions between members of the erb-B family [35], for example, will require appropriate extensions of the mathematical model. We have also relied extensively on the Michaelis Menten approximation for reaction kinetics, which we do not believe fundamentally alters our results, but whose appropriateness for large interconnected networks is examined elsewhere [36].

The second part of the process is to make predictions on the unmeasured or unmeasurable species of the system, assuming that the model has been suitably refined. We suggest that for testable predictions to be made, uncertainty estimates need to be attached to them [26]. In some cases the prediction uncertainties are rather small, despite large parameter uncertainty. On the other hand, if some predictions show large uncertainty, and involve species that are not directly measurable, we may then define a suitable design criterion and suggest new experimental measurements that need to be taken to reduce that uncertainty. The results of such an analysis are promising, in that we find a rather small number of measurements (realistic to perform with standard molecular biology techniques) need to be taken to begin to make predictions with good precision. Given such measurements on the EGFR system, we see that the triple complex of active Cool-1, Cbl and active Cdc42 does indeed form in appreciable quantities in wild type cells and we also get an estimate for the time of formation and dissociation.

More generally, we believe that experimental design for reducing prediction uncertainties can play an important role in the iterative process of model refinement and validation and can be used in the testing of biological hypotheses.

#### 5 References

- 1 Carpenter, G.: 'The egf receptor: a nexus for trafficking and signaling', *BioEssays*, 2000, **22**, pp. 697–707
- 2 Wells, A., Welsh, J.B., Lazar, C.S., Wiley, H.S., Gill, G.N., and Rosenfeld, M.G.: 'Ligand-induced transformation by a non-internalizing epidermal growth factor receptor', *Science*, 1990, **247**, pp. 962–964
- 3 Hirsch, F.R., Varella-Garcia, M., Bunn, P.A. Jr., Maria, M.V. D. Veve, R., Bremmes, R.M., Baron, A.E., Zeng, C., and Franklin, W.A.: 'Epidermal growth factor receptor in non-small-cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis', *J. Clin. Oncol.*, 2003, **20**, pp. 3798–3807
- 4 DiGiovanna, M.P., Stern, D.F., Edgerton, S.M., Whalen, S.G. II, Moore, D., and Thor, A.D.: 'Relationship of epidermal growth factor receptor expression to erbB-2 signaling activity and prognosis in breast cancer patients', *J. Clin. Oncol.*, 2005, **23**, pp. 1152–1160
- 5 Feng, Q., Baird, D., Peng, X., Wang, J., Ly, T., Guan, J-L., and Cerione, R.A.: 'Cool-1 functions as an essential regulatory node for egf receptor-and src-mediated cell growth', *Nature Cell Biology*, 2006, pp. 945–956
- 6 Wu, W.J., Tu, S., and Cerione, R.A.: 'Activated cdc42 sequesters c-cbl and prevents egf receptor degradation', *Cell*, 2003, **114**, pp. 715–725
- 7 Wong, E.S.M., Fong, C.W., Lim, J.Y.P., Low, B.C., Langdon, W.Y., and Guy, G.R.: 'Sprouty2 attenuates epidermal growth factor receptor ubiquitylation and endocytosis, and consequently enhances ras/erk signalling', *EMBO J.*, 2002, **21**, pp. 4796–4808
- 8 Levkowitz, G., Waterman, H., Ettenberg, S.A., Katz, M., Tsygankov, A.Y., Alroy, I., Lavi, S., Iwai, K., Reiss, Y., Ciechanover, A., Lipkowitz, S., and Yarden, Y.: 'Ubiquitin ligase activity and tyrosine phosphorylation underlie suppression of growth factor signaling by c-cbl/sli-1', *Mol. Cell*, 1999, **4**, pp. 1029–1040
- 9 Ettenberg, S.A., Magnifico, A., Cuello, M., Nau, M.M., Rubinstein, Y.R., Yarden, Y., Weissman, A.M., and Lipkowitz, S.: 'Cbl-b-dependent coordinated degradation of the epidermal growth factor receptor signaling complex', *J. Biol. Chem.*, 2001, **276**, pp. 27677–27684

- 10 Grovdal, L.M., Stang, E., Sorkin, A., and Madshus, I.H.: 'Direct interaction of cbl with ptyr 1045 of the egf receptor (egfr) is required to sort the egfr to lysosomes for degradation', *Exp. Cell Res.*, 2004, **300**, pp. 388–395
- 11 Huang, F., and Sorkin, A.: 'Growth factor receptor binding protein 2-mediated recruitment of the ring domain of cbl to the epidermal growth factor receptor is essential and sufficient to support receptor endocytosis', *Mol. Biol. Cell*, 2005, **16**, pp. 1268–1281
- 12 Jiang, X., and Sorkin, A.: 'Epidermal growth factor receptor internalization through clathrin-coated pits requires cbl ring finger and proline-rich domains but not receptor polyubiquitylation', *Traffic*, 2003, **4**, pp. 529–543
- 13 Mosesson, Y., Shtiegman, K., Katz, M., Zwang, Y., Vereb, G., Szollosi, J., and Yarden, Y.: 'Endocytosis of receptor tyrosine kinases is driven by monoubiquitylation, not polyubiquitylation', *J. Biol. Chem.*, 2003, **278**, pp. 21323–21326
- 14 Duan, L., Miura, Y., Dimri, M., Majumder, B., Dodge, I.L., Reddi, A.L., Ghosh, A., Fernandes, N., Zhou, P., Mullane-Robinson, K., Rao, N., Donoghue, S., Rogers, R.A., Bowtell, D., Naramura, M., Gu, H., Band, V., and Band, H.: 'Cbl-mediated ubiquitylation is required for lysosomal sorting of epidermal growth factor receptor but is dispensable for endocytosis', *J. Biol. Chem.*, 2003, **278**, pp. 28950–28960
- 15 Sigismund, S., Woelk, T., Puri, C., Maspero, E., Tacchetti, C., Transidico, P., Fiore, P.P.D., and Polo, S.: 'Clathrin-independent endocytosis of ubiquitinated cargos', *PNAS*, 2005, **102**, pp. 2760–2765
- 16 Szymkiewicz, I., Kowanetz, K., Soubeyran, P., Dinarina, A., Lipkowitz, S., and Dikic, I.: 'Cin85 participates in cbl-b-mediated downregulation of receptor tyrosine kinases', *J. Biol. Chem.*, 2002, **277**, pp. 39666–39672
- 17 Schoeberl, B., Eichler-Jonsson, C., Gilles, E.D., and Muller, G.: 'Computational modeling of the dynamics of the map kinase cascade activated by surface and internalized egf receptors', *Nat. Biotechnol.*, 2002, **20**, pp. 370–375
- 18 Kholodenko, B.N., Demin, O.V., Moehren, G., and Hoek, J.B.: 'Quantification of short term signaling by the epidermal growth factor receptor', *J. Biol. Chem.*, 1999, **274**, pp. 30169–30181
- 19 Resat, H., Ewald, J.A., Dixon, D.A., and Wiley, H.S.: 'An integrated model of epidermal growth factor receptor trafficking and signal transduction', *Biophys. J.*, 2003, **85**, pp. 730–743
- 20 Blinov, M.L., Faeder, J.R., Goldstein, B., and Hlavacek, W.S.: 'A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity', *BioSystems*, 2006, **83**, pp. 136–151
- 21 Serban, R., and Hindmarsh, A.C.: 'Cvodes: an ode solver with sensitivity analysis capabilities'. LLNL Technical Report UCRL-MA-148813, Lawrence Livermore National Labs, 2002
- 22 Atkinson, A.C., and Donev, A.N.: 'Optimum experimental design' (Oxford University Press, New York, 1992, 1st edn.)
- 23 Faller, D., Klingmuller, U., and Timmer, J.: 'Optimal experimental design in systems biology', *Simulation*, 2003, **79**, p. 717
- 24 Rodriguez-Fernandez, M., Mendes, P., and Banga, J.R.: 'A hybrid approach for efficient and robust parameter estimation in biochemical pathways', *BioSystems*, 2006, **83**, pp. 248–265
- 25 Kutalik, Z., Cho, K-H., and Wolkenhauer, O.: 'Optimal sampling time selection for parameter estimation in dynamic pathway modelling', *BioSystems*, 2004, **75**, pp. 43–55
- 26 Brown, K.S., and Sethna, J.P.: 'Statistical mechanics approaches to models with many poorly known parameters', *Phys. Rev. E.*, 2003, **68**, p. 021904
- 27 Gutenkunst, R.N., Waterfall, J.J., Casey, F.P., Brown, K.S., Myers, C.R., and Sethna, J.P.: 'Universally sloppy parameter sensitivities in systems biology', 2007, Submitted for publication
- 28 Waterfall, J.J., Casey, F.P., Gutenkunst, R.N., Brown, K.S., Myers, C.R., Brouwer, P.W., Elser, V., and Sethna, J.P.: 'The sloppy model universality class and the vandermonde matrix', *Phys. Rev. Lett.*, 2006, **97**, p. 150601
- 29 Silvey, S.D.: 'Optimal design: an introduction to the theory for parameter estimation' (Chapman and Hall, London, 1980, 1st edn.)
- 30 Golub, G.H., and Loan, C.F.V.: 'Matrix computations' (The Johns Hopkins University Press, Baltimore, 1996, 3rd edn.)
- 31 Hanafusa, H., Torii, S., Yasunaga, T., Matsumoto, K., and Nishida, E.: 'Shp2, an sh2-containing protein-tyrosine phosphatase, positively regulates receptor tyrosine kinase signaling by dephosphorylating and inactivating the inhibitor sprouty', *J. Biol. Chem.*, 2004, **279**, pp. 22992–22995
- 32 Burke, P., Schooler, K., and Wiley, H.S.: 'Regulation of epidermal growth factor receptor signaling by endocytosis and intracellular trafficking', *Mol. Biol. Cell*, 2001, **12**, pp. 1897–1910
- 33 Shvartsman, S.Y., Hagan, M.P., Yacoub, A., Dent, P., Wiley, H.S., and Lauffenburger, D.A.: 'Autocrine loops with positive feedback enable context-dependent cell signaling', *Am. J. Physiol. Cell Physiol.*, 2002, **282**, C545–C559
- 34 Shvartsman, S.Y., Wiley, H.S., Deen, W.M., and Lauffenburger, D.A.: 'Spatial range of autocrine signaling: Modeling and computational analysis', *Biophys. J.*, 2001, **81**, pp. 1854–1867
- 35 Hendriks, B.S., Cook, J., Burke, J.M., Beusmans, J.M., Lauffenburger, D.A., and de Graaf, D.: 'Computational modelling of erbb family phosphorylation dynamics in response to transforming growth factor alpha and heregulin indicates spatial compartmentation of phosphatase activity', *IEE Proc., Syst. Biol.*, 2006, **1**, pp. 22–33
- 36 Aldridge, B.B., Burke, J.M., Lauffenburger, D.A., and Sorger, P.K.: 'Physicochemical modelling of cell signalling pathways', *Nature Cell Bio.*, 2006, **8**, pp. 1195–1203

## 6 Appendix

The differential equations for the EGFR down-regulation model and the best fit parameter values are included here.

There are 41 dynamical variables and 11 'assignment' rules which are derived from the dynamical variables. The assigned variables are functions of the dynamical variables (e.g. total amount of protein) which correspond to experimentally measured quantities. The abbreviations used to shorten the names of the dynamical or assigned variables are:

Lyso = Lysosomal, Endo = Endosomal, PMEGFR = Plasma membrane EGFR, Ubi = Ubiquitinated, PPase = Phosphatase.

### Assignment rules

$$\begin{aligned}
 [\text{TotalPMEGFR}] &= [\text{PMEGFR}] \\
 &+ [\text{PMBoundEGFR}] \\
 &+ [\text{PMBoundEGFRCbl}] \\
 &+ [\text{PMBoundEGFRCblActive}] \\
 &+ [\text{PMBoundEGFRCool ActiveCbl}]
 \end{aligned}$$

$$\begin{aligned}
 [\text{TotalPMBoundEGFR}] &= [\text{PMBoundEGFR}] \\
 &+ [\text{PMBoundEGFRCbl}] \\
 &+ [\text{PMBoundEGFR CblActive}] \\
 &+ [\text{PMBoundEGFR CoolActiveCbl}]
 \end{aligned}$$

$$\begin{aligned}
 [\text{TotalInternalEGFR}] &= [\text{EndosomalEGFR}] \\
 &+ [\text{EndosomalEGFRCbl}] \\
 &+ [\text{EndosomalEGFR CblActive}] \\
 &+ [\text{EndosomalEGFR CoolActiveCbl}] \\
 &+ [\text{UbiEndosomal EGFRCblActive}] \\
 &+ [\text{UbiEndosomalEGFR}] \\
 &+ [\text{LysoEGFRCblActive}] \\
 &+ [\text{LysoEGFR}]
 \end{aligned}$$

$$\begin{aligned}
 [\text{TotalInternalEGF}] &= [\text{EndosomalEGF}] \\
 &+ [\text{LysoEGF}]
 \end{aligned}$$

$$\begin{aligned}
 [\text{TotalEGFR}] &= [\text{TotalPMEGFR}] \\
 &+ [\text{TotalInternalEGFR}]
 \end{aligned}$$

$$\begin{aligned}
 [\text{TotalUbiEGFR}] &= [\text{UbiEndosomalEGFR CblActive}]
 \end{aligned}$$

$$\begin{aligned}
& + [\text{UbiEndosomalEGFR}] \\
& + [\text{LysoEGFR}] \\
& + [\text{LysoEGFRCblActive}]
\end{aligned}$$

$$\begin{aligned}
[\text{TotalCdc42GDP}] &= [\text{Cdc42GDP}] \\
& + [\text{CoolActiveCdc42GDP}] \\
& + [\text{CoolActiveCblCdc42GDP}]
\end{aligned}$$

$$\begin{aligned}
[\text{TotalCdc42GTP}] &= [\text{Cdc42GTP}] \\
& + [\text{CoolActiveCdc42GTP}] \\
& + [\text{CoolActiveCblCdc42GTP}]
\end{aligned}$$

$$\begin{aligned}
[\text{TotalCoolActive}] &= [\text{CoolActive}] \\
& + [\text{PMBoundEGFRCoolActiveCbl}] \\
& + [\text{EndosomalEGFRCoolActiveCbl}] \\
& + [\text{CoolActiveCdc42GDP}] \\
& + [\text{CoolActiveCdc42GTP}] \\
& + [\text{CoolActiveCblCdc42GDP}] \\
& + [\text{CoolActiveCblCdc42GTP}] \\
& + [\text{CoolActiveCbl}]
\end{aligned}$$

$$\begin{aligned}
[\text{TotalCblActive}] &= [\text{CblActive}] \\
& + [\text{PMBoundEGFRCblActive}] \\
& + [\text{EndosomalEGFRCblActive}] \\
& + [\text{UbiEndosomalEGFRCblActive}] \\
& + [\text{LysoEGFRCblActive}]
\end{aligned}$$

$$\begin{aligned}
[\text{TotalCbl}] &= [\text{TotalCblActive}] \\
& + [\text{Cbl}] \\
& + [\text{PMBoundEGFRCbl}] \\
& + [\text{PMBoundEGFRCoolActiveCbl}] \\
& + [\text{EndosomalEGFRCbl}] \\
& + [\text{EndosomalEGFRCoolActiveCbl}] \\
& + [\text{CoolActiveCblCdc42GDP}] \\
& + [\text{CoolActiveCblCdc42GTP}] \\
& + [\text{CoolActiveCbl}]
\end{aligned}$$

#### Differential equations

$$\begin{aligned}
\frac{d[\text{EGF}]}{dt} &= ku\text{EGF} \cdot [\text{PMBoundEGFR}] \\
& - kb\text{EGF} \cdot [\text{EGF}] \cdot [\text{PMEGFR}]
\end{aligned}$$

$$\begin{aligned}
\frac{d[\text{PMEGFR}]}{dt} &= ku\text{EGF} \cdot [\text{PMBoundEGFR}] \\
& + k\text{Rec} \cdot [\text{EndosomalEGFR}] \\
& + k\text{tEGFR} \\
& - kb\text{EGF} \cdot [\text{EGF}] \cdot [\text{PMEGFR}] \\
& - kd\text{EGFR} \cdot [\text{PMEGFR}]
\end{aligned}$$

$$\frac{d[\text{PMBoundEGFR}]}{dt} = kb\text{EGF} \cdot [\text{EGF}] \cdot [\text{PMEGFR}]$$

$$\begin{aligned}
& + \frac{k\text{PPaseCool} \cdot [\text{PPase}] \cdot [\text{PMBoundEGFR}]}{(\text{KmPPaseCool} + [\text{PMBoundEGFR} \cdot \text{CoolActiveCbl}])} \\
& + ku\text{CoolCblEGFR} \cdot [\text{PMBoundEGFRCoolActiveCbl}] \\
& + ku\text{Cbl} \cdot [\text{PMBoundEGFRCbl}] \\
& + ku\text{Cbl} \cdot [\text{PMBoundEGFRCblActive}] \\
& - ku\text{EGF} \cdot [\text{PMBoundEGFR}] \\
& - kb\text{CoolCblEGFR} \cdot [\text{CoolActiveCblCdc42GTP}] \cdot [\text{PMBoundEGFR}] - kb\text{CoolCblEGFR} \cdot [\text{CoolActiveCbl}] \cdot [\text{PMBoundEGFR}] \\
& - kb\text{Cbl} \cdot [\text{PMBoundEGFR}] \cdot [\text{Cbl}] \\
& - kb\text{Cbl} \cdot [\text{PMBoundEGFR}] \cdot [\text{CblActive}]
\end{aligned}$$

$$\frac{d[\text{PMBoundEGFRCbl}]}{dt} = kb\text{Cbl} \cdot [\text{PMBoundEGFR}]$$

$$\begin{aligned}
& + \frac{k\text{PPaseCbl} \cdot [\text{PPase}] \cdot [\text{PMBoundEGFR} \cdot \text{CblActive}]}{([\text{PMBoundEGFR} \cdot \text{CblActive}] + \text{KmPPaseCbl})} \\
& - ku\text{Cbl} \cdot [\text{PMBoundEGFRCbl}] \\
& - k\text{Cbl} \cdot [\text{PMBoundEGFRCbl}]
\end{aligned}$$

$$\frac{d[\text{PMBoundEGFRCblActive}]}{dt} = kb\text{Cbl} \cdot [\text{PMBoundEGFR}]$$

$$\begin{aligned}
& \cdot [\text{CblActive}] + k\text{Cbl} \cdot [\text{PMBoundEGFRCbl}] \\
& - k\text{Endo} \cdot [\text{PMBoundEGFRCblActive}] - ku\text{Cbl} \cdot [\text{PMBoundEGFRCblActive}] \\
& + \frac{k\text{PPaseCbl} \cdot [\text{PPase}] \cdot [\text{PMBoundEGFR} \cdot \text{CblActive}]}{([\text{PMBoundEGFR} \cdot \text{CblActive}] + \text{KmPPaseCbl})}
\end{aligned}$$

$$\frac{d[\text{PMBoundEGFRCoolActiveCbl}]}{dt} = kb\text{CoolCblEGFR}$$

$$\begin{aligned}
& \cdot [\text{CoolActiveCblCdc42GTP}] \cdot [\text{PMBoundEGFR}] + kb\text{CoolCblEGFR} \cdot [\text{CoolActiveCbl}] \cdot [\text{PMBoundEGFR}]
\end{aligned}$$



$$\begin{aligned} & \frac{kPPaseCool \cdot [PPase] \cdot [PMBoundEGFRCoolActiveCbl]}{(KmPPaseCool + [PMBoundEGFRCoolActiveCbl])} \\ & - kuCoolCblEGFR \cdot [PMBoundEGFRCoolActiveCbl] \\ \frac{d[Cbl]}{dt} = & \frac{kPPaseCool \cdot [PPase] \cdot [CoolActiveCbl]}{(KmPPaseCool + [CoolActiveCbl])} \\ & + \frac{kPPaseCool \cdot [PPase] \cdot [PMBoundEGFRCoolActiveCbl]}{(KmPPaseCool + [PMBoundEGFRCoolActiveCbl])} \\ & + \frac{kPPaseCool \cdot [PPase] \cdot [CoolActiveCblCdc42GTP]}{(KmPPaseCool + [CoolActiveCblCdc42GTP])} \\ & + kuCool42_{Cbl} \cdot [CoolActiveCblCdc42GTP] \\ & + kuCoolCbl \cdot [CoolActiveCbl] \\ & + kuCbl \cdot [PMBoundEGFRCbl] \\ & + \frac{kPPaseCbl \cdot [PPase] \cdot [CblActive]}{([CblActive] + KmPPaseCbl)} \\ & + kuCbl \cdot [EndosomalEGFRCbl] \\ & - kbCool42_{Cbl} \cdot [CoolActiveCdc42GTP] \cdot [Cbl] \\ & - kbCoolCbl \cdot [CoolActive] \cdot [Cbl] \\ & - kbCbl \cdot [PMBoundEGFR] \cdot [Cbl] \\ & - kbCbl \cdot [EndosomalEGFR] \cdot [Cbl] \end{aligned}$$

$$\begin{aligned} \frac{d[CblActive]}{dt} = & kuCbl \cdot [PMBoundEGFRCblActive] + kuCbl \cdot [EndosomalEGFRCblActive] + kuCblUbi \cdot [UbiEndosomalEGFRCblActive] - kbCbl \cdot [PMBoundEGFR] \cdot [CblActive] \\ & - \frac{kPPaseCbl \cdot [PPase] \cdot [CblActive]}{([CblActive] + KmPPaseCbl)} \\ & - kbCbl \cdot [EndosomalEGFR] \cdot [CblActive] \\ & - kbCblUbi \cdot [UbiEndosomalEGFR] \cdot [CblActive] \end{aligned}$$

$$\begin{aligned} \frac{d[Cool]}{dt} = & \frac{kPPaseCool \cdot [PPase] \cdot [CoolActive]}{([CoolActive] + KmPPaseCool)} \\ & + \frac{kPPaseCool \cdot [PPase] \cdot [CoolActiveCbl]}{(KmPPaseCool + [CoolActiveCbl])} \\ & + \frac{kPPaseCool \cdot [PPase] \cdot [PMBoundEGFRCoolActiveCbl]}{(KmPPaseCool + [PMBoundEGFRCoolActiveCbl])} \\ & + \frac{kPPaseCool \cdot [PPase] \cdot [CoolActiveCblCdc42GTP]}{(KmPPaseCool + [CoolActiveCblCdc42GTP])} \end{aligned}$$

$$\begin{aligned} & + \frac{kPPaseCool \cdot [PPase] \cdot [CoolActiveCdc42GTP]}{(KmPPaseCool + [CoolActiveCdc42GTP])} \\ & + \frac{kPPaseCool \cdot [PPase] \cdot [CoolActiveCdc42GDP]}{(KmPPaseCool + [CoolActiveCdc42GDP])} \\ & - \frac{kFAKCool \cdot [FAKActive] \cdot [Cool]}{([Cool] + KmFAKCool)} \end{aligned}$$

$$\begin{aligned} \frac{d[CoolActive]}{dt} = & kuCool42 \cdot [CoolActiveCdc42GDP] + kuCool42GTP \cdot [CoolActiveCdc42GTP] \\ & + \frac{kFAKCool \cdot [FAKActive] \cdot [Cool]}{([Cool] + KmFAKCool)} \\ & + kuCoolCbl \cdot [CoolActiveCbl] - kbCool42 \cdot [CoolActive] \cdot [Cdc42GDP] - kbCool42GTP \cdot [CoolActive] \cdot [Cdc42GTP] \\ & - \frac{kPPaseCool \cdot [PPase] \cdot [CoolActive]}{([CoolActive] + KmPPaseCool)} \\ & - kbCoolCbl \cdot [CoolActive] \cdot [Cbl] \end{aligned}$$

$$\frac{d[EndosomalEGF]}{dt} = kEndo \cdot [PMBoundEGFRCblActive] - kLyso \cdot [EndosomalEGF]$$

$$\begin{aligned} \frac{d[EndosomalEGFR]}{dt} = & kuCoolCblEGFR \cdot [EndosomalEGFRCoolActiveCbl] + kuCbl \cdot [EndosomalEGFRCbl] + kuCbl \cdot [EndosomalEGFR \cdot CblActive] - kbCoolCblEGFR \cdot [CoolActiveCblCdc42GTP] \cdot [EndosomalEGFR] - kbCbl \cdot [EndosomalEGFR] \cdot [Cbl] - kbCbl \cdot [EndosomalEGFR] \cdot [CblActive] - kRec \cdot [EndosomalEGFR] \end{aligned}$$

$$\begin{aligned} \frac{d[EndosomalEGFRCbl]}{dt} = & \frac{kPPaseCbl \cdot [PPase] \cdot [EndosomalEGFRCblActive]}{([EndosomalEGFRCblActive] + KmPPaseCbl)} \\ & + kbCbl \cdot [EndosomalEGFR] \cdot [Cbl] - kuCbl \cdot [EndosomalEGFRCbl] - kCbl \cdot [EndosomalEGFRCbl] \end{aligned}$$

$$\begin{aligned} \frac{d[EndosomalEGFRCblActive]}{dt} = & kEndo \cdot [PMBoundEGFRCblActive] + kbCbl \cdot [EndosomalEGFR] \cdot [CblActive] + kCbl \cdot [EndosomalEGFRCbl] \end{aligned}$$



$$\begin{aligned}
& - k_{\text{Cool42GTP}} \\
& \cdot [\text{CoolActiveCblCdc42GTP}] \\
& - k_{\text{Cool42}_{\text{Cbl}}} \\
& \cdot [\text{CoolActiveCblCdc42GTP}] \\
& - \frac{k_{\text{GAP}} \cdot [\text{Cdc42GAP}] \cdot [\text{CoolActiveCblCdc42GTP}]}{([\text{CoolActiveCblCdc42GTP}] + K_{\text{mGAP}})} \\
& - k_{\text{bCoolCblEGFR}} \\
& \cdot [\text{CoolActiveCblCdc42GTP}] \\
& \cdot [\text{PMBoundEGFR}] - k_{\text{bCoolCblEGFR}} \\
& \cdot [\text{CoolActiveCblCdc42GTP}] \\
& \cdot [\text{EndosomalEGFR}]
\end{aligned}$$

$$\begin{aligned}
\frac{d[\text{CoolActiveCbl}]}{dt} = & k_{\text{Cool42GTP}} \\
& \cdot [\text{CoolActiveCblCdc42GTP}] \\
& + k_{\text{Cool42}} \\
& \cdot [\text{CoolActiveCblCdc42GDP}] \\
& + k_{\text{bCoolCbl}} \cdot [\text{CoolActive}] \\
& \cdot [\text{Cbl}] + k_{\text{CoolCblEGFR}} \\
& \cdot [\text{PMBoundEGFRCoolActiveCbl}] \\
& + k_{\text{CoolCblEGFR}} \\
& \cdot [\text{EndosomalEGFRCoolActiveCbl}] \\
& \frac{k_{\text{PPaseCool}} \cdot [\text{PPase}] \cdot [\text{CoolActiveCbl}]}{(K_{\text{mPPaseCool}} + [\text{CoolActiveCbl}])} \\
& - k_{\text{bCool42GTP}} \cdot [\text{CoolActiveCbl}] \\
& \cdot [\text{Cdc42GTP}] - k_{\text{CoolCbl}} \\
& \cdot [\text{CoolActiveCbl}] - k_{\text{bCoolCblEGFR}} \\
& \cdot [\text{CoolActiveCbl}] \cdot [\text{PMBoundEGFR}]
\end{aligned}$$

$$\begin{aligned}
\frac{d[\text{UbiEndosomalEGFR}]}{dt} = & k_{\text{CblUbi}} \\
& \cdot [\text{UbiEndosomalEGFRCblActive}] \\
& - k_{\text{bCblUbi}} \cdot [\text{UbiEndosomalEGFR}] \\
& \cdot [\text{CblActive}] - k_{\text{Lyso}} \\
& \cdot [\text{UbiEndosomalEGFR}]
\end{aligned}$$

$$\begin{aligned}
\frac{d[\text{UbiEndosomalEGFRCblActive}]}{dt} = & k_{\text{Ubi}} \cdot [\text{EndosomalEGFRCblActive}] \\
& + k_{\text{bCblUbi}} \cdot [\text{UbiEndosomalEGFR}] \\
& \cdot [\text{CblActive}] - k_{\text{CblUbi}} \\
& \cdot [\text{UbiEndosomalEGFRCblActive}] \\
& - k_{\text{Lyso}} \cdot [\text{UbiEndosomalEGFRCblActive}]
\end{aligned}$$

$$\frac{d[\text{recycledEGFR}]}{dt} = k_{\text{Rec}} \cdot [\text{EndosomalEGFR}]$$

$$\begin{aligned}
\frac{d[\text{LysoEGFR}]}{dt} = & k_{\text{Lyso}} \cdot [\text{UbiEndosomalEGFR}] \\
& - k_{\text{Prot}} \cdot [\text{LysoEGFR}]
\end{aligned}$$

$$\begin{aligned}
\frac{d[\text{LysoEGFRCblActive}]}{dt} = & k_{\text{Lyso}} \\
& \cdot [\text{UbiEndosomalEGFRCblActive}] \\
& - k_{\text{Prot}} \cdot [\text{LysoEGFRCblActive}]
\end{aligned}$$

$$\begin{aligned}
\frac{d[\text{LysoEGF}]}{dt} = & k_{\text{Lyso}} \cdot [\text{EndosomalEGF}] \\
& - k_{\text{ProtEGF}} \cdot [\text{LysoEGF}]
\end{aligned}$$

$$\begin{aligned}
\frac{d[\text{Src}]}{dt} = & \frac{k_{\text{PPase}} \cdot [\text{PPase}] \cdot [\text{SrcActive}]}{([\text{SrcActive}] + K_{\text{mPPase}})} \\
& \cdot k_{\text{EGFRSrc}} \\
& - \frac{[\text{TotalPMBoundEGFR}] \cdot [\text{Src}]}{([\text{Src}] + K_{\text{mEGFRSrc}})} \\
& \cdot k_{\text{EGFRSrc}} \\
& - \frac{[\text{TotalInternalEGFR}] \cdot [\text{Src}]}{([\text{Src}] + K_{\text{mEGFRSrc}})}
\end{aligned}$$

$$\begin{aligned}
\frac{d[\text{SrcActive}]}{dt} = & \frac{k_{\text{EGFRSrc}} \cdot [\text{TotalPMBoundEGFR}] \cdot [\text{Src}]}{([\text{Src}] + K_{\text{mEGFRSrc}})} \\
& + \frac{k_{\text{EGFRSrc}} \cdot [\text{TotalInternalEGFR}] \cdot [\text{Src}]}{([\text{Src}] + K_{\text{mEGFRSrc}})} \\
& - \frac{k_{\text{PPase}} \cdot [\text{PPase}] \cdot [\text{SrcActive}]}{([\text{SrcActive}] + K_{\text{mPPase}})}
\end{aligned}$$

$$\begin{aligned}
\frac{d[\text{FAK}]}{dt} = & \frac{k_{\text{PPase}} \cdot [\text{PPase}] \cdot [\text{FAKActive}]}{([\text{FAKActive}] + K_{\text{mPPase}})} \\
& - \frac{k_{\text{SrcFAK}} \cdot [\text{SrcActive}] \cdot [\text{FAK}]}{([\text{FAK}] + K_{\text{mSrcFAK}})}
\end{aligned}$$

$$\begin{aligned}
\frac{d[\text{FAKActive}]}{dt} = & \frac{k_{\text{SrcFAK}} \cdot [\text{SrcActive}] \cdot [\text{FAK}]}{([\text{FAK}] + K_{\text{mSrcFAK}})} \\
& - \frac{k_{\text{PPase}} \cdot [\text{PPase}] \cdot [\text{FAKActive}]}{([\text{FAKActive}] + K_{\text{mPPase}})}
\end{aligned}$$

$$\begin{aligned}
\frac{d[\text{ERK}]}{dt} = & \frac{k_{\text{dErk}} \cdot [\text{PPase}] \cdot [\text{ErkP}]}{([\text{ErkP}] + K_{\text{mdErk}})} \\
& - \frac{k_{\text{Erk}} \cdot [\text{RasGTP}] \cdot [\text{ERK}]}{([\text{ERK}] + K_{\text{mErk}})}
\end{aligned}$$

$$\begin{aligned}
\frac{d[\text{RasGDP}]}{dt} = & \frac{k_{\text{RasGAP}} \cdot [\text{RasGAP}] \cdot [\text{RasGTP}]}{([\text{RasGTP}] + K_{\text{mRasGAP}})} \\
& + \frac{k_{\text{ErkRas}} \cdot [\text{ErkP}] \cdot [\text{RasGTP}]}{([\text{RasGTP}] + K_{\text{mErkRas}})} \\
& \cdot k_{\text{Ras}} \\
& \cdot [\text{TotalPMBoundEGFR}] \\
& \cdot [\text{RasGDP}] \\
& - \frac{[\text{RasGDP}]}{([\text{RasGDP}] + K_{\text{mRas}})} \\
& \cdot k_{\text{Ras}} \cdot [\text{TotalInternalEGFR}] \\
& - \frac{[\text{RasGDP}]}{([\text{RasGDP}] + K_{\text{mRas}})}
\end{aligned}$$

$$\frac{d[\text{RasGAP}]}{dt} = 0$$

$$\frac{d[\text{ErkP}]}{dt} = \frac{k_{\text{Erk}} \cdot [\text{RasGTP}] \cdot [\text{ERK}]}{([\text{ERK}] + K_{\text{mErk}})} - \frac{k_{\text{dErk}} \cdot [\text{PPase}] \cdot [\text{ErkP}]}{([\text{ErkP}] + K_{\text{mErk}})}$$

$$\frac{d[\text{RasGTP}]}{dt} = \frac{k_{\text{Ras}} \cdot [\text{TotalPMBoundEGFR}] \cdot [\text{RasGDP}]}{([\text{RasGDP}] + K_{\text{mRas}})} + \frac{k_{\text{Ras}} \cdot [\text{TotalInternalEGFR}] \cdot [\text{RasGDP}]}{([\text{RasGDP}] + K_{\text{mRas}})} - \frac{k_{\text{RasGAP}} \cdot [\text{RasGAP}] \cdot [\text{RasGTP}]}{([\text{RasGTP}] + K_{\text{mRasGAP}})} - \frac{k_{\text{ErkRas}} \cdot [\text{ErkP}] \cdot [\text{RasGTP}]}{([\text{RasGTP}] + K_{\text{mErkRas}})}$$

$$\frac{d[\text{PPase}]}{dt} = 0$$

$$\frac{d[\text{DegradedEGFR}]}{dt} = k_{\text{Prot}} \cdot [\text{LysoEGFRCblActive}] + k_{\text{Prot}} \cdot [\text{LysoEGFR}]$$

$$\frac{d[\text{DegradedEGF}]}{dt} = k_{\text{ProtEGF}} \cdot [\text{LysoEGF}]$$

Abbreviations for rate constant names in Table 1 are as follows: all rate constants begin with 'k', Michaelis-Menten constants with 'Km', 'ku' is an unbinding rate, 'kb' is a binding rate, 'kt' is transcription/translation rate, 'kd' is a degradation rate. The specific reactions that each biochemical parameter describes are shown in the Table. Note that in the context of the parameter uncertainty shown in Fig. 5, referring to a single best set of parameter values is inaccurate, as there exists an entire family of parameter sets with essentially equivalent model outputs. However, the parameters shown in the table are the 'best fit' parameters which provide the fits shown in the supplementary material.

**Table 1: Best fit parameter values. Unless otherwise stated, the parameter is a rate constant. Molecular amounts are measured by molecular number, and time is measured in minutes**

<i>CblIC</i>	35 824.1567 916	Cbl initial number
<i>CoolIC</i>	171 986.913 584	Cool initial number
<i>Cdc42IC</i>	6287.20 482 709	Cdc42 initial number
<i>kbEGF</i>	$1.27\ 007\ 412\ 549 \times 10^{-8}$	EGF binding
<i>kuEGF</i>	0.00 308 907 282 474	EGF unbinding
<i>kbCoolCbl</i>	$2.9696\ 348\ 704 \times 10^{-6}$	Cool Cbl binding
<i>kuCoolCbl</i>	0.0139 521 257 042	Cool Cbl unbinding
<i>kbCoolCblEGFR</i>	0.00 236 902 387 151	Cool Cbl complex binding to receptor
<i>kuCoolCblEGFR</i>	0.977 152 064 172	Cool Cbl complex unbinding from receptor
<i>kbCoolI42</i>	0.00 201 364 864 644	Cool Cdc42GDP binding
<i>kuCoolI42</i>	0.0222 192 029 238	Cool Cdc42GDP unbinding
<i>kbCoolI42<sub>cbl</sub></i>	0.0516 455 719 396	Cool Cdc42 complex binding to Cbl
<i>kuCoolI42<sub>cbl</sub></i>	0.00 682 821 254 319	Cool Cdc42 complex unbinding to Cbl
<i>kbCoolI42GTP</i>	0.0847 193 655 231	Cool Cdc42GTP binding
<i>kuCoolI42GTP</i>	0.00 671 059 371 127	Cool Cdc42GTP unbinding
<i>kGAP</i>	0.0188 624 358 733	Cdc42 deactivation constant
<i>KmGAP</i>	7086.09 430 902	Michaelis-Menten Cdc42 deactivation constant
<i>kEndo</i>	11.1665 976 526	endocytosis rate constant
<i>kLyso</i>	14.5741 564 171	rate constant for transfer to lysosome
<i>kProt</i>	19.7749 621 241	proteolysis rate constant for receptor
<i>kProtEGF</i>	0.00 841 869 306 691	proteolysis rate constant for EGF
<i>kbCbl</i>	$3.4519\ 451\ 8304 \times 10^{-5}$	Cbl receptor binding
<i>kuCbl</i>	0.00 554 568 738 575	Cbl receptor unbinding
<i>kbCblUbi</i>	0.00 303 250 359 176	Cbl ubiquitinated receptor binding
<i>kuCblUbi</i>	143.533 297 692	Cbl ubiquitinated receptor unbinding
<i>kCbl</i>	14.9383 334 083	Cbl activation by receptor
<i>kUbi</i>	12.8941 557 064	ubiquitination of receptor by bound active Cbl
<i>kCoolI42</i>	25.8977 169 465	Cdc42GDP activation by Cool
<i>kRas</i>	0.130 528 255 097	Ras activation
<i>KmRas</i>	2571.6114 437	Ras Michaelis-Menten activation constant
<i>kRasGAP</i>	0.0259 233 381 417	Ras deactivation by GAP
<i>KmRasGAP</i>	6947.94 735 971	Ras Michaelis-Menten deactivation constant
<i>kErk</i>	0.78 289 949 978	Erk activation by RasGTP

(Table continued)

**Table 1: Continued**


---

<i>KmErk</i>	2653.56 615 157	Erk Michaelis-Menten activation constant
<i>kdErk</i>	5.51 257 830 148	Erk deactivation by phosphatase
<i>KmdErk</i>	11.1526 801 363	Erk Michaelis-Menten deactivation constant
<i>kErkRas</i>	1.17 574 704 673	Ras deactivation by activated Erk (negative feedback)
<i>KmErkRas</i>	342.874 428 432	Ras deactivation Michaelis-Menten constant
<i>kEGFRSrc</i>	0.4439 199 291	Src activation
<i>KmEGFRSrc</i>	4220.0030 575	Src Michaelis-Menten activation constant
<i>kSrcFAK</i>	8.01 255 586 265	FAK activation
<i>KmSrcFAK</i>	2569.69 644 914	FAK Michaelis-Menten activation constant
<i>kFAKCool</i>	78.0319 081 233	Cool activation
<i>KmFAKCool</i>	121 085.033 481	Cool Michaelis-Menten activation constant
<i>ktEGFR</i>	0.00 270 489 867 598	receptor transcription
<i>kdEGFR</i>	0.00 023 499 407 766	receptor (non-specific) degradation
<i>kRec</i>	0.0108 367 146 003	recycling rate constant
<i>kPPase</i>	63.9789 592 234	general dephosphorylation
<i>KmPPase</i>	2083.79 578 915	general dephosphorylation Michaelis-Menten constant
<i>kPPaseCbl</i>	0.0234 779 943 608	Cbl dephosphorylation rate
<i>KmPPaseCbl</i>	3107.44 359 155	Cbl dephosphorylation Michaelis-Menten constant
<i>kPPaseCool</i>	146.955 731 588	Cool dephosphorylation rate
<i>KmPPaseCool</i>	3096.00 331 638	Cool dephosphorylation Michaelis-Menten constant
<i>PMEGFRIC</i>	15 564.3598 552	total initial number surface EGFR (experiments 6, 7 and 10 only)
<i>kF28</i>	0.32 481 948 655	Cdc42 F28L mutant activation rate
<i>OccupiedEGFRNoEGF3T3</i>	38 194.6447 963	initial occupied number of surface EGFR (experiment 4 only)

---