Investigating dynamics of inhibitory and feedback loops in ERK signalling using power-law models

Julio Vera[1,&,*], Oliver Rath[2,&], Eva Balsa-Canto[3,&], Julio R. Banga[3], Walter Kolch[2, 4, 5] and Olaf Wolkenhauer[1]

[1]Systems Biology and Bioinformatics Group, University of Rostock, 18051 Rostock, Germany
[2]The Beatson Institute for Cancer Research, Cancer Research UK Garscube Estate Bearsden, Glasgow G61 1BD, United Kingdom
[3]Process Engineering Group, IIM-CSIC, Eduardo Cabello 6, 36208 Vigo, Spain
[4]Sir Henry Wellcome Functional Genomics Facility, University of Glasgow Glasgow G12 8QQ, United Kingdom
[5]Current address: Systems Biology Ireland, Conway Institute, University College Dublin, Dublin 4, Ireland
[&]These authors made equal contributions

Supplementary material

Abstract

Background. The investigation of the structure and dynamics of signal transduction systems through data-based mathematical models in ordinary differential equations or other paradigms has proved to be a successful approach in recent times. Extending this concept, we here analysed the use of kinetic models based on power-law terms with non-integer kinetic orders in the validation of hypotheses concerning regulatory structures in signaling systems.

Methods. We integrated pre-existent biological knowledge, hypotheses and experimental quantitative data into a power-law model to validate the existence of certain regulatory loops in the Ras/Raf-1/MEK/ERK pathway, a MAPK pathway involved in the transduction of mitogenic and differentiation signals. Towards this end, samples of a human mammary epithelial cell line (MCF-10A) were used to obtained time-series data characterising the behaviour of the system after epidermal growth factor stimulation in different scenarios of expression for the critical players of the system regarding the investigated loops (e.g., the inhibitory protein RKIP). The mathematical model was later calibrated using a computational procedure that included: analysis of structural identifiability; global ranking of parameters to detect the most sensitivity ones towards the experimental setup; model calibration using global optimization methods to find the parameter values that better fit the data; and practical identifiability analysis to estimate the confidence in the estimated values for the parameters**.** The obtained model was used to performed computational simulations concerning the role of the investigated regulatory loops in the time response of the signaling pathway.

Results. Our findings suggest that the special regularity in the structure of the power-law terms make them suitable for a data-based validation of regulatory loops in signalling pathways. Using this approach, we could perform a quantitative data based validation of the existence of two regulatory loops in the pathway. The model proved to be useful to highlight the important consequences of phosphorylation/deactivation of RKIP during the transient stimulation of the signalling pathway.

[*] **Corresponding author:** Julio Vera. Junior Research Group on Systems Biology and Cancer. Chair in Systems Biology and Bioinformatics. Department of Computer Science, University of Rostock. 18051 Rostock, Germany.
E-mail: olaf.wolkenhauer@uni-rostock.de Web: www.sbi.uni-rostock.de
Tel.: +49 381 498 75 77 Fax: +49 381 498 76 81

SM0. Power law equations

Figure SM0.1 shows how a unique power-law rate equation can represent completely different biological behaviours by only changing the values of the kinetic orders.

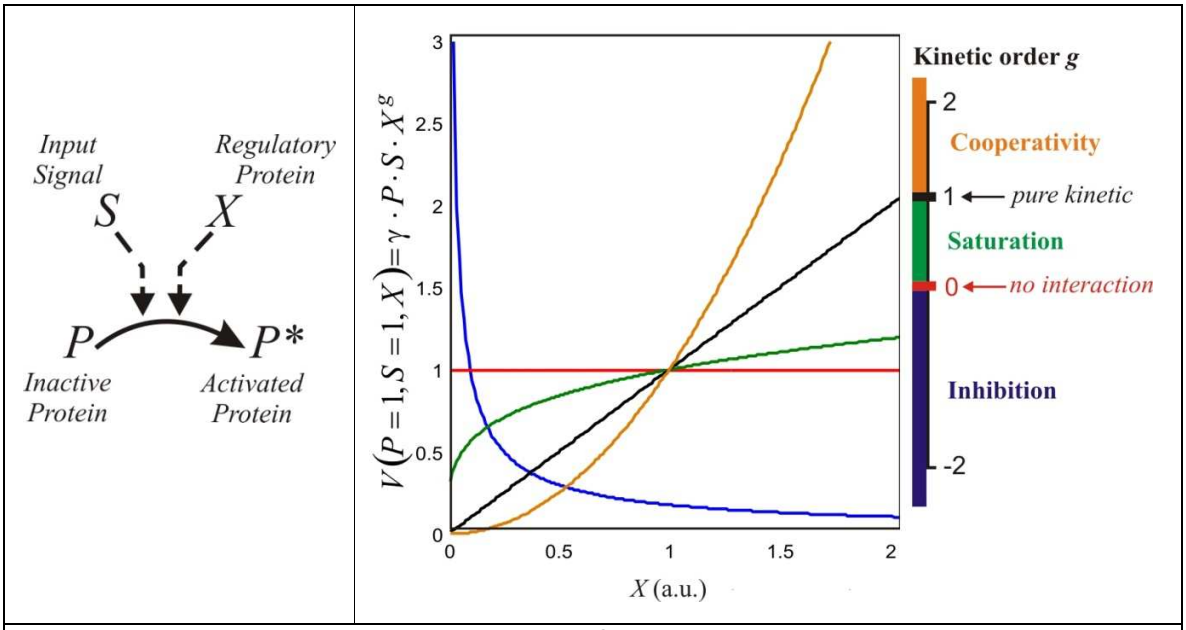

Figure 1. Power-law models can describe a variety of enzymatic reactions and their control by variation of the kinetic order. For illustrative purposes, we discuss activation of protein *P*, a process that is mediated by the input signal *S* and potentially regulated by a second interacting partner called *X* (left-hand side of the figure). In order to investigate the regulatory dynamics induced by *X* we derive the following simplified power-law model:

$$-\frac{dP}{dt}=\frac{dP^*}{dt}=V(P,S,X)=\gamma \cdot P^h \cdot S^j \cdot X^g$$

In cases where no kinetic information on the effect of *X* on *P* is available, the value of the kinetic order *g* must be assigned using specifically designed experiments. The bar on the right-hand side depicts the values for the signal rate *V*(*P*,*S*,*X*) when the same values are assumed for *P* (P=1) and *S* (S=1) and the parameters other than *g* (*h*=1, *j*=1and γ=1). We compute the values of *V* for an interval of values in *X* using different representative values for *g*: i) inhibition (blue), *g*=-0.51; ii) no-interaction (red), *g*=0; iii) saturation (green), *g*=0.26; iv) linear kinetic (black), *g*=1.0; v) cooperativity (yellow), *g*=1.51). For simplifying purposes, in case of inhibition (*g*=-0.51) the variable *X* was normalised around *X*=0.05. As we see here, one and the same mathematical structure for the signal rate can display rather different biological dynamics by just tuning the value of critical kinetic orders, the property of power-law models that we analyse and explore here to elucidate cell signal regulation.

SM1. Input signal EGFR*

In the calibration process we model the input signal with the following differential equation:

$$\frac{dEGFR^*}{dt} = Inp(t) - k_{dEGFR} \cdot EGFR^*$$
$$Inp(t): EGFR_*\left(t = 0^-\right) = 0.0 \wedge EGFR_*\left(t = 0^+\right) = 1.0$$

Where *Inp(t)* is the function accounting for the stimulation of epidermal growth factor receptor activation and k_{dEGFR} accounts for the process of deactivation of EGFR*. We here assume a quick activation of the receptor that is modelled by an input function Inp(t) with the structure defined above. Furthermore and in agreement with previous experimental measurements under similar experimental conditions, we assume that the activated epidermal growth factor receptor (EGFR) have a half-life approximately 45 minutes (Honegger et al. 1990*, Kholodenko 2006). Under this assumption, k_{dEGFR}=0.015 (min^{-1}) and after integration the input function of the system during the perturbatory experiments becomes: $EGFR_*\left(t = 0^-\right) = 0.0$ $EGFR^*(t) = e^{-0.0154t}$ (Figure SM1.1).

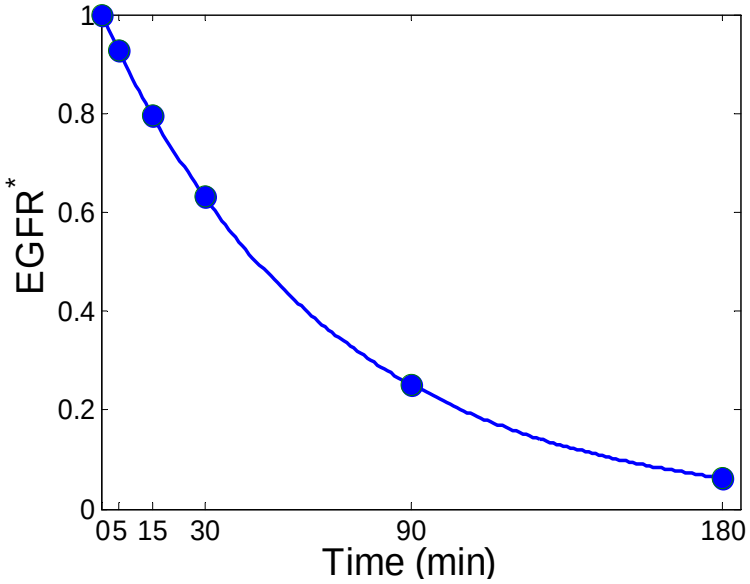

Figure SM1.1. Simulated input signal of the system used in the model calibration.

SM2. Total concentration of proteins

In the model the total concentrations of MEK and ERK were assumed to be constant in the considered time scale (10^2 minutes). This assumption is supported by the quantitative measurements for the total concentration of both proteins performed in parallel to the

* Honegger, A.M., Schmidt, A., Ullrich, A., Schlessinger, J. (1990) Separate endocytic pathways of kinase-defective and -active EGF receptor mutants expressed in same cells. J. Cell Biol. 110(5):1541-8.

measurements of phosphorylated fractions. In the following figures we show how the total concentration for both proteins remains approximately constant and similar during both experiments. The agreement between our assumption and the data is especially strong in case of ERK (Figure SM2.A) and acceptable for MEK (Figure SM2.B). We obtained similar results for RKIP during the experiments, which showed a rather stable total amount during the perturbative experiments (data not shown).

Figure SM2.A

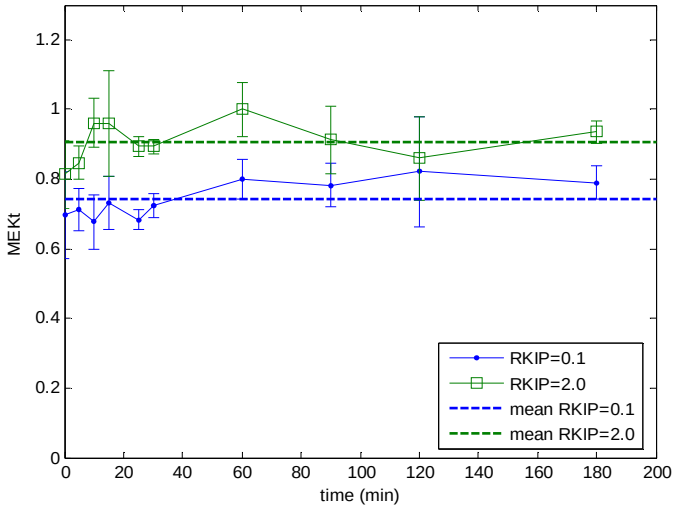

Figure SM2.B

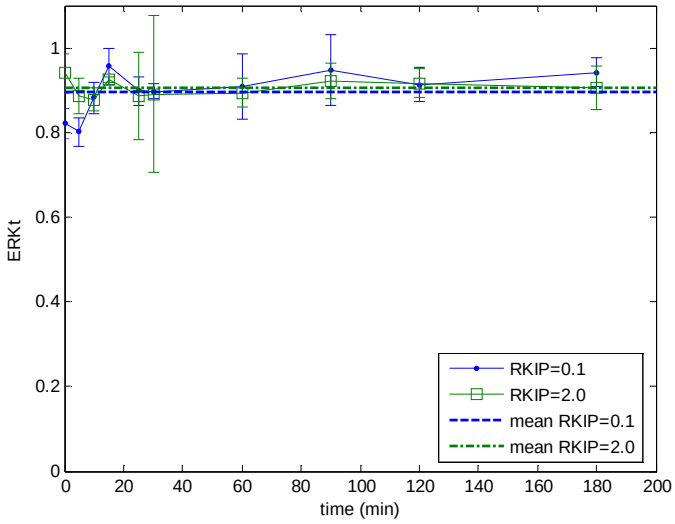

Figure SM2.2. Total amount of proteins MEK and ERK. A) Measurement of the total amount of MEK during the experiments performed in our investigation (solid lines). Data were normalised in a way maximum concentration of MEK measured was considered equal to one. In the figure we also included the computed mean value for the concentration of MEK in both experiments (dashed lines). **B)** Measurement of the total amount of ERK during the experiments performed in our investigation (solid lines). As in Figure SM2.A, data were normalised in a way maximum concentration of ERK measured was considered equal to one. Mean value for ERK total amount in both experiments was computed and included in the figure (dashed lines).

SM3. Calibration of the first candidate model

SM3.1. First candidate model calibration. Each experiment in Figure 3 was done in three replicates thus allowing computation of the mean and the standard deviation for each measurement (Figure 6). In this way we could introduce experimental error information in the log-likelihood function used for parameter estimation (Equation 3). Based on the structure of the mathematical model discussed before, we first derived a simplified version in terms of number of kinetic orders to be computed and assumed all g values were fixed to 1 excepting the ones corresponding to the feedback loops affecting MEK activation that we tried to validate (e.g. g_2 and g_3). This will be regarded as first candidate model. The calibration of this first candidate model requires the estimation of 9 global parameters (the two cited kinetic orders and seven rate constants) plus two additional local (experiment dependent) unknowns, namely ERK_T and pERK(0). The final computational problem has 9 global and 18 local parameters, which makes a total of 27 unknown parameters. The metaheuristic SSm [40] was used to estimate the unknowns from the experimental data in the following hyperrectangle:

$$g_2, g_3 \in [-4, 0];$$

$$\gamma_{i=1,..,7} \in [0, 10];$$

$$ERK_T \in [0.1, 1];$$

$$pERK(t = 0) \in [0, 1];$$

The best fits, corresponding to a value of J_{llk}=339.5, are shown in the Supplementary material. Figures 6 a) and b) present two illustrative examples: (a) presents results for experiment RKIP wt/ ERK o, which corresponds to the case that reported the minimum mean residual; and (b) presents results for experiment RKIP o/ ERK d which corresponds to the case reporting the minimum maximum residual.

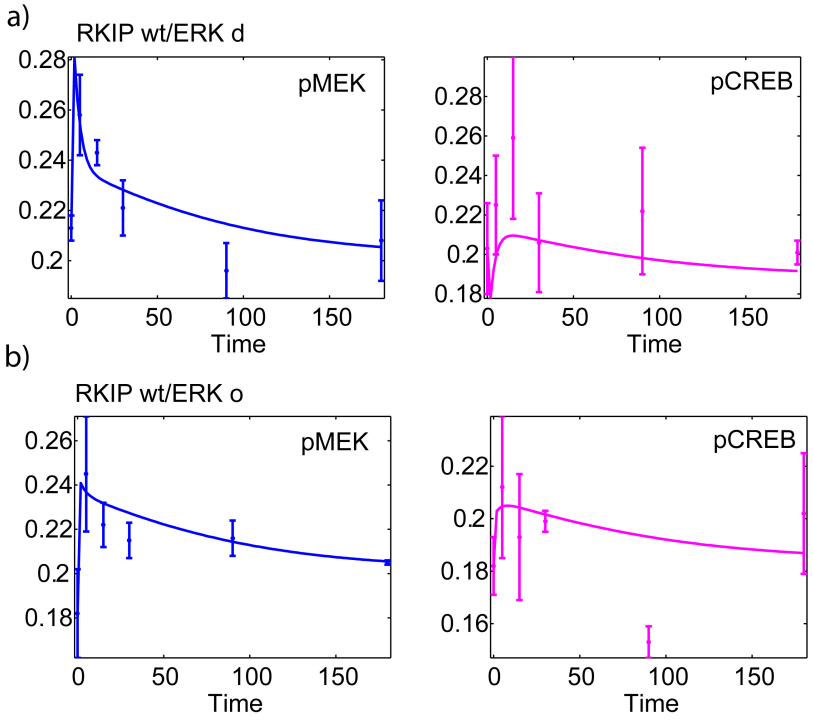

Figure 6. Illustrative examples of the best fits achieved for the first candidate model corresponding to a J_{lik}=339.5. a) Experiment RKIP wt/ ERK d, corresponds to the experiment which reported the minimum maximum residual in average for both observables, b) experiment RKIP wt/ ERK o, reported the minimum mean residual for both observables. Predicted kinetics are shown as solid lines, experimentally measured results as data points with standard deviation. In these figures, time is measured in minutes and the phosphorylation of the different proteins is in arbitrary units

It should be noted at this point that even though the sensitivity analysis showed maximum sensitivity for the experiment RKIP d/ ERK d, and thus one would expect to obtain the best fit for that experiment, this is not the case here (see Supplementary Material SM3). In addition, and attending to the sensitivity of pCREB to the parameters, one would also expect that predictions for pCREB would be better than those for pMEK. Both behaviours may be explained attending to the experimental error. The data for pCREB are subject, in general, to more experimental error than those for pMEK, thus the log-likelihood is giving more confidence to the experimental data related to pMEK, and thus the quality of fit is better for pMEK. This is particularly true for experiment RKIP d/ ERK d for which the maximum mean experimental error is reported for pCREB.

The model is able to capture the tendency of the data, however for some of the experiments prediction errors are rather significant (see Supplementary Material SM3). We notice that the computed solution reports the presence of the regulatory loops that we were trying to validate (e.g. RKIP and ERK inhibition of MEK activation, represented by the kinetic orders g_2=-0.86±0.05, and g_3=-3.9±0.8 respectively). Since the experiments were designed to analyse in detail the possible presence of feedback loops confidence intervals for the parameter estimates are very reasonable.

SM3.2. Practical identifiability analysis for the first candidate model. The confidence intervals for the parameters computed by means of the Fisher Information Matrix reveal that the experimental scheme is suitable to compute g_2 and g_3 with reasonable confidence. Taking into account that the maximum relative experimental error is around the 30% and 7% in mean, the maximum relative confidence interval of the 20% obtained for g_3 is reasonable. The correlation matrix (Figure 7) reveals that g_2 is, in general, not substantially correlated to the other parameters whereas g_3 is slightly more correlated. The rate constants are in some cases highly correlated thus being more complicated to accurately estimate their values. This could be anticipated in view of the low sensitivity of the observables with respect to for example γ_1, γ_2 or γ_3 (Figures 4 and 5) under the current experimental scheme. Further optimally designed experiments would be required to improve confidence for rate constants [34].

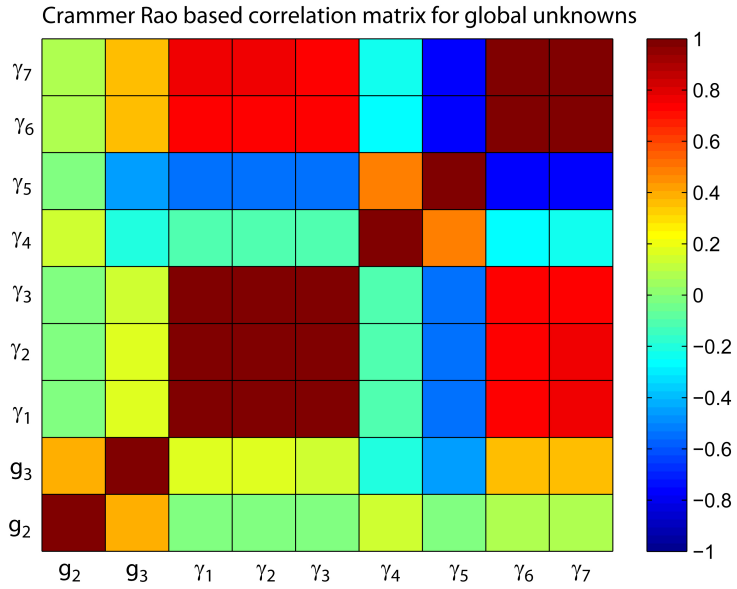

Figure 7. Correlation matrix of the global unknowns corresponding to the best solution for the first candidate model. Figure reveals that the kinetic orders g_2 and g_3 are almost fully uncorrelated whereas all combinations of $\gamma_1, \gamma_2, \gamma_3$, or γ_6, γ_7, are highly correlated, which introduces large difficulties at the time of estimating those parameters.

To complement the information provided by the correlation matrix, Figures 8a-d show the projection of the maximum likelihood function over four pairs of parameters in the vicinity, within the 25% error, of the solution: a) presents an illustrative example of (almost) uncorrelated parameters γ_1 - g_2 , b) presents an example of low correlation between parameters (γ_4 - g_3) and c) and d) present examples of highly correlated parameters γ_5 - γ_2 and γ_7 - γ_6.

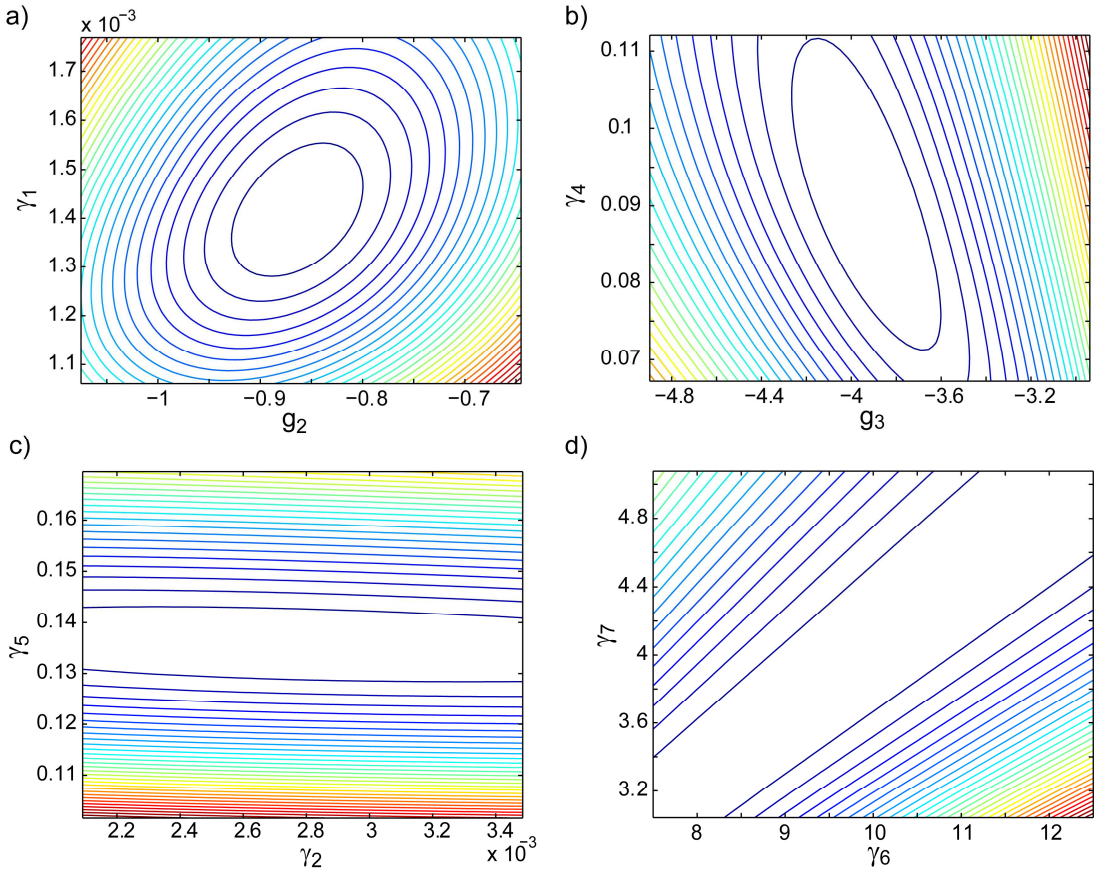

Figure 8. Illustrative examples of the projection of the log-likelihood over pairs of parameters in the vicinity of the global solution: a) presents an illustrative example of (almost) uncorrelated parameters, b) presents an example of low correlation between parameters, c) and d) present examples of highly correlated parameters. These plots give an intuition of the difficulties to compute an unique value for the parameters.

SM3.3. Experimental scheme. A total of 9 experiments were performed under different combinations of RKIP and ERK conditions (over expression, wild type and down regulation). In all experiments 6 measurements were performed for pERK, pCREB and total MEK at the following sampling times: 0, 5, 15, 30, 90, 180. Three replicates are available for each case thus allowing to compute the mean and the standard deviation for each measurement. This allowed to introduce experimental error information in the log-likelihood function.

SM3.4. Definition of unknowns. For the first approximation to the problem only g2 and g3 were to be estimated together with the kinetic rates whereas remaining kinetic orders were fixed to one. The initial condition for pMEK and pCREB was assumed to be known and equal to the mean measured value and the initial condition for pERK was estimated for each experiment (local unknown). In addition the value of total ERK was estimated for each experiment (local uknown).

The following bounds were established for all unknowns:

$g_2, g_3 \in [-4, 0]$;

$\gamma_{i=1,..,7} \in [0,10]$;

$ERK_T \in [0.1,1]$;

$pERK(t=0) \in [0,1]$;

SM3.5. Solution with a global optimization method. The model calibration problem was solved by means of SSm (Egea et al., 2007). Results achieved for the parameter values are:

```
g2 = -8.6019e-001  ±  4.5641e-002;
g3 = -3.9178e+000  ±  5.3651e-001;
γ1 =  1.4150e-003  ±  9.6817e-002;
γ2 =  2.7879e-003  ±  1.9082e-001;
γ3 =  9.1702e+000  ±  6.2354e+002;
γ4 =  8.9643e-002  ±  2.6275e-002;
γ5 =  1.3561e-001  ±  2.5110e-002;
γ6 =  1.0000e+001  ±  8.4764e+001;
γ7 =  4.0620e+000  ±  3.4663e+001;
```

It should be noted that confidence intervals for the kinetic orders are reasonable whereas for the kinetic rates confidence is low. This may be explained considering that the design of experiments was performed to analyse the possible presence of feedback loops. Further experiments specifically designed for the purpose of kinetic rates estimation would possibly improve the quality of the results. Figure SM3.1 shows corresponding fits.

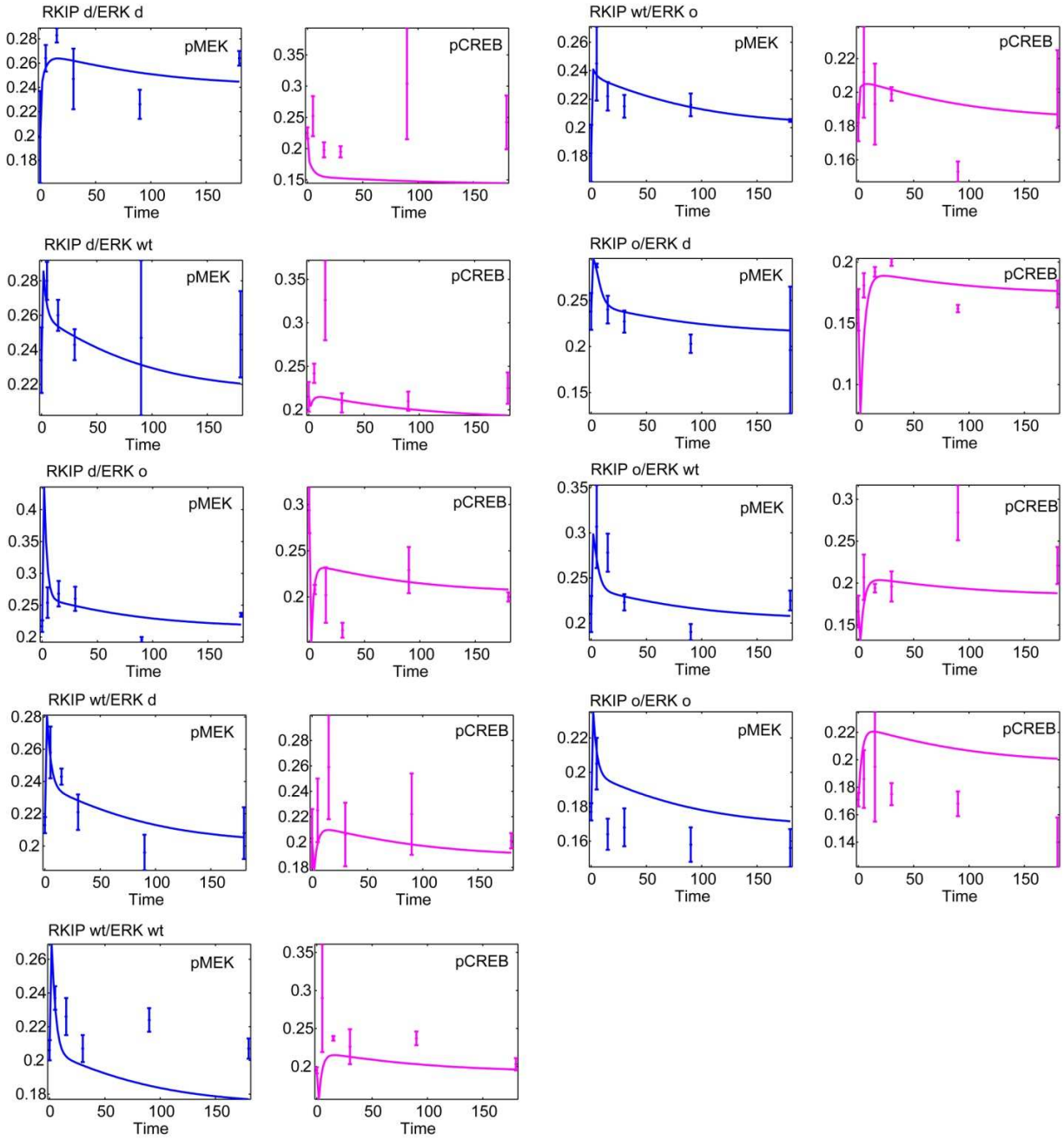

Figure SM3.1: Best fits for the first candidate model: J_{llk}=339.5

SM4. Details on the calibration of the second candidate model

SM4.1. Experimental scheme. A total of 9 experiments were performed under different combinations of RKIP and ERK conditions (over expression, wild type and down regulation). In all experimeevants 6 measurements were performed for pERK, pCREB and total MEK at the following sampling times: 0, 5, 15, 30, 90, 180. Three replicates are available for each case thus

allowing to compute the mean and the standard deviation for each measurement. This allowed to introduce experimental error information in the log-likelihood function.

SM4.2. Definition of unknowns. For the first power-law model g2 and g3 were to be estimated together with the most influencing kinetic orders as detected by the global sensitivity analysis (g_{11}, g_9, g_8 and g_6) and the kinetic rates. As in previous case the initial condition for pERK and the amount of total ERK were estimated for each experiment (local unknowns). The following bounds were established for all unknowns:

$g_{11,9,8,6} \in [0,4];$

$g_2, g_3 \in [-4,0];$

$\gamma_{i=1,..,7} \in [0,10];$

$ERK_T \in [0.1,1];$

$pERK(t=0) \in [0,1];$

SM4.3. Solution with a global optimization method. The model calibration problem was solved by means of SSm (Egea et al., 2007). Results achieved for the parameter values are:

```
g2 = -1.3972e+000  ±  1.1426e-001;
g3 = -1.2336e+000  ±  2.1159e-001;
g6 =  2.9594e+000  ±  3.7278e-001;
g8 =  6.4457e-002  ±  5.5995e-002;
g9 =  1.6221e-002  ±  9.0505e-002;
g11=  1.8035e-002  ±  1.4964e+000;
γ1 =  3.0692e-005  ±  2.6387e-004;
γ2 =  4.3944e-003  ±  1.0394e-002;
γ3 =  2.0728e-001  ±  4.3802e-001;
γ4 =  1.1914e-001  ±  6.8618e-002;
γ5 =  8.6303e-004  ±  1.1744e-004;
γ6 =  7.6618e+000  ±  4.8271e+002;
γ7 =  5.8766e+000  ±  3.6732e+002;
```

Figure SM4.1 shows corresponding fits.

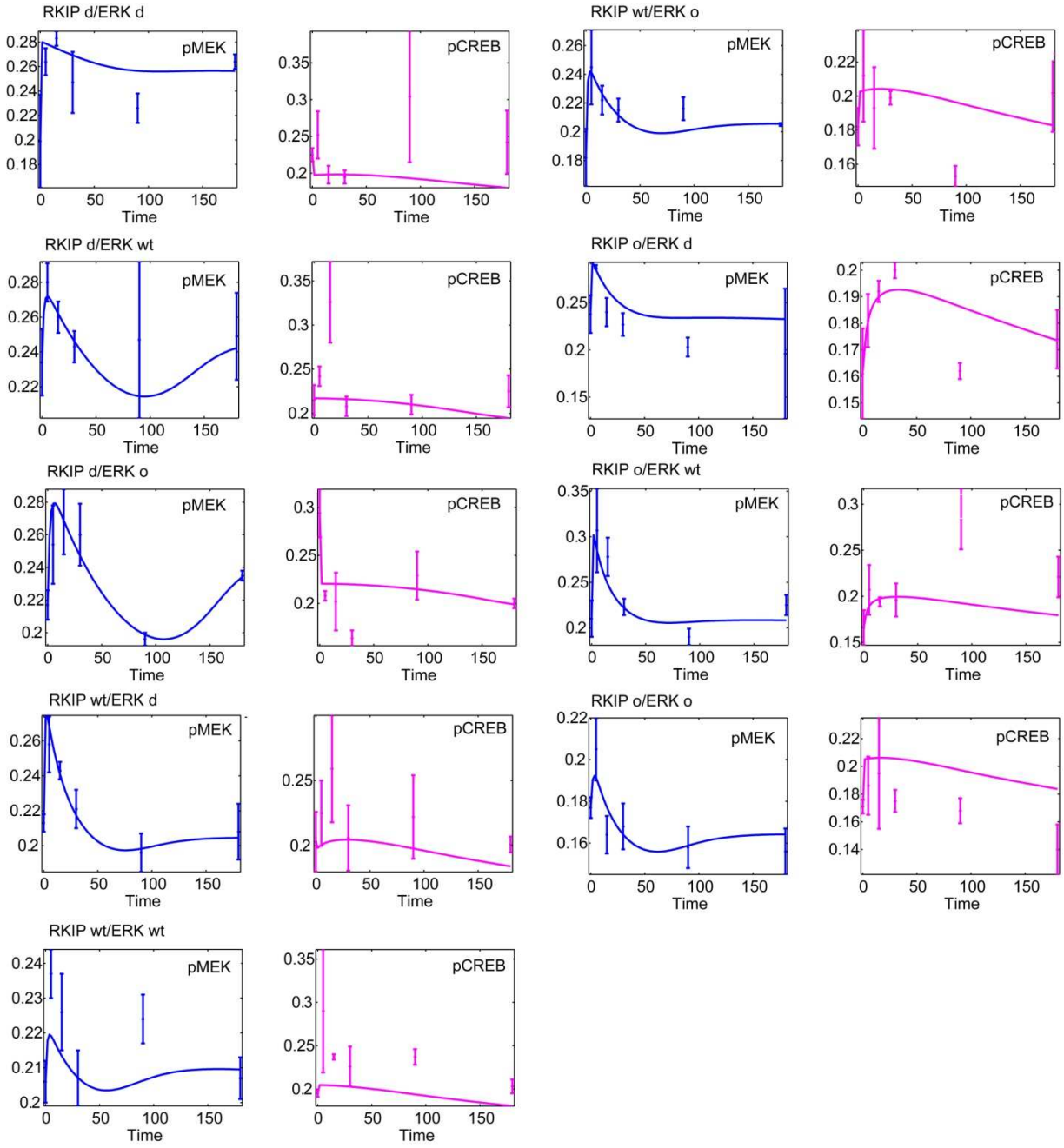

Figure SM4.1: Best fits for the second candidate model: J_{llk}=239.4