Supporting text

A. Quantification of experimental data

To establish quantitative dynamic models of biological systems, a reliable method to quantify experimental data is essential. In our experimental setup we analyzed temporal changes in erythropoietin recptor (EpoR) and signal transducer and activator of transcription (STAT)5 tyrosine phosphorylation in response to Epo stimulation. The respective components were sequentially immunoprecipitated from detergent lysates of BaF3-EpoR cells, analyzed by immunoblotting followed by chemiluminescence and Lumilmager detection, and quantified by the LUMIANALYST software. Whereas the linear range of X-ray films conventionally used for detection of immunoblots is limited, Lumilmager facilitates according to the manufacturer measurements linearly related to the input over four orders of magnitude. To confirm this experimentally, a serial dilution of BaF3-EpoR cell lysates was analyzed. Quantification of the immunoblots probed with anti-STAT5 antiserum (Fig.5) revealed that the detected signal strength ranged from 10^5 to 10^7 Boehringer Light Units (BLU), covering the values obtained in our immunoprecipitation experiments that were used for the mathematical modeling. Linear regression of the experimental data (Fig. 5) confirmed that detection with the Lumilmager is over more than two orders of magnitude linear. Therefore, Lumilmager can be used to generate reasonably quantitative data for databased modeling.

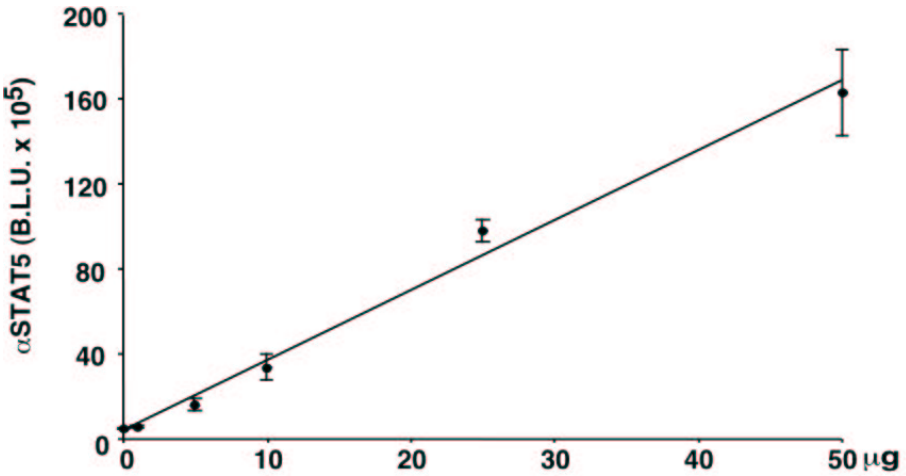

Fig. 5. Experimental validation of the linear detection range of the Lumilmager. Detergent lysates of BaF3-EpoR cells were prepared, and the protein concentration contained was determined by Bradford assay. One to fifty micrograms of protein from total cell lysate was separated by 15% SDS/PAGE and analyzed by immunoblotting using an anti-STAT5 antiserum followed by chemiluminescence detection. After 10min exposure on the Lumilmager, the immunoblot was quantified by using the LumiAnalyst software, applying the single band analysis package with automated lane and band identification, flat background correction, and slant correction. Plotted are the protein concentration of total cell lysates versus the measured Boehringer Light Units (B.L.U.). Per protein concentration the average of three measurements $\pm$ SD is depicted.

B. Modeling of an independent experiment

To validate our mathematical model, quantitative data from an experiment showing atypical EpoR tyrosine phosphorylation were used as input function (Fig. 6A) to predict the development of cytoplasmic STAT5 in this experiment. Using the dynamical parameters $k_1 - k_4$, τ and $x_1(0)$ from the previous experiments, and fitting only the nuisance parameters k_5, k_6, and k_7, a curve was obtained that closely matched experimental data determined for both tyrosine phosphorylated and total STAT5 in the cytoplasm (Fig. 6B), thus demonstrating the predictive power of our mathematical model.

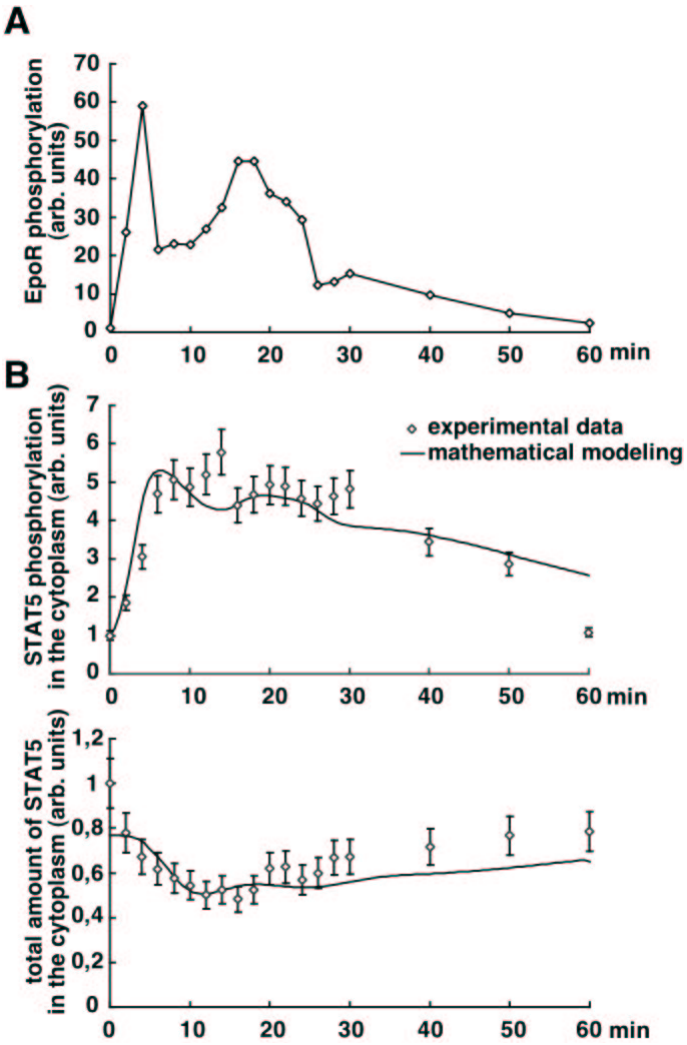

Fig. 6. Predicting the behavior of the STAT5 nucleocytoplasmic cycle based on the dynamical parameters determined in the previous experiments. (A) Time course of EpoR tyrosine phosphorylation was used as input function to model (B) STAT5 tyrosine phosphorylation in the cytoplasm ($Upper$) and the total amount of cytoplasmic STAT5 ($Lower$) in an independent experiment. Points with error bars indicate the experimental data whereas solid lines represent the mathematical modeling. The indicated error was determined based on duplicated measurements.

C. Investigation of additional modeling alternatives

To select an appropriate model describing the dynamics of the STAT5 signaling pathway, it is necessary to go through a stepwise process of testing various models suggested by biochemical knowledge. We started with the assumption that the signaling pathway represents a feed-forward cascade (model 1) delivering a signal from cell surface receptors to the nucleus but, as described in the manuscript, this approach has to be discarded. To improve this model, we first included the possibility that STAT5 dimers possess a certain instability and disintegrate to the monomeric form. Such backreactions can occur, but it is *a priori* not clear to what extent they influence the dynamical behavior of this signaling pathway. Therefore, we investigated whether a feed-forward model with backreaction (model 2) is able to describe the experimental data satisfactorily. In comparison to the simple feed-forward model, the differential equations are extended by one additional term describing the backreaction:

$$\begin{aligned}
\dot{x}_1 &= -k_1 x_1 \mathrm{EpoR_A} & y_1 &= k_5(x_2 + 2x_3) \\
\dot{x}_2 &= +k_1 x_1 \mathrm{EpoR_A} - k_2 x_2^2 \; + 2k_3' x_3 & y_2 &= k_6(x_1 + x_2 + 2x_3) \\
\dot{x}_3 &= -k_3 x_3 + 0.5 k_2 x_2^2 \; - k_3' x_3 & y_3 &= k_7 \mathrm{EpoR_A} \\
\dot{x}_4 &= +k_3 x_3.
\end{aligned}$$

However, the fit of this model to the experimental data is similar to the fit of the simple cascade (see Fig. $7A$), and therefore this model is also not able to describe the data.

Next, we tested a model including the possibility that STAT5 is exported from the nucleus and reactivated in the cytoplasm. This model with nucleocytoplasmic cycling (model 3) leads to a satisfactory fit and is therefore described in detail in the manuscript (see Fig. $2D$).

To compare two models statistically, we compute twice the difference of the log-likelihood of both models, which leads to likelihood ratio tests (1). When comparing model 2 with model 3, we obtain the test statistic $L = 2(LR_3 - LR_2) = 830.2$, which corresponds to a p-value of $< 10^{-5}$, where LR_i refers to the log-likelihood of model i. Hence, the null hypothesis that the dynamic parameter responsible for cycling k_4 is zero has to be rejected.

It is noteworthy that the distribution of the test statistic does not follow the conventional single χ^2-distribution but a mixture of χ^2-distributions, because the estimated parameter under the alternative is on the boundary of the parameter space, i.e., may not be negative (2). The critical level of the 5% significance niveau is 2.71.
Having established that a model with nucleocytoplasmic cycling is able to describe the dynamic behavior of the system, we investigated whether the inclusion of the backreaction into the model further improves the mathematical description. To test the improvement of the inclusion of the backreaction in model 3, we extended our dynamical model accordingly, resulting in model 4. The new differential equations are now more complex, because we introduce one additional parameter k_3' describing the backreaction:

$$\begin{aligned}
\dot{x}_1 &= -k_1 x_1 \mathrm{EpoR_A} \; + 2k_4 x_3(t-\tau) & y_1 &= k_5(x_2 + 2x_3) \\
\dot{x}_2 &= +k_1 x_1 \mathrm{EpoR_A} - k_2 x_2^2 \; + 2k_3' x_3 & y_2 &= k_6(x_1 + x_2 + 2x_3) \\
\dot{x}_3 &= -k_3 x_3 + 0.5 k_2 x_2^2 \; - k_3' x_3 & y_3 &= k_7 \mathrm{EpoR_A} \\
\dot{x}_4 &= +k_3 x_3 \; - k_4 x_3(t-\tau).
\end{aligned}$$

The resulting fit is displayed in Fig. 7B. Again, to assess the significance of this new parameter, we calculate the test statistic $LR = 2(LR_4 - LR_3) = 0.70$. This value corresponds to a p-value of 0.20, which indicates that this is not a significant improvement of the model.

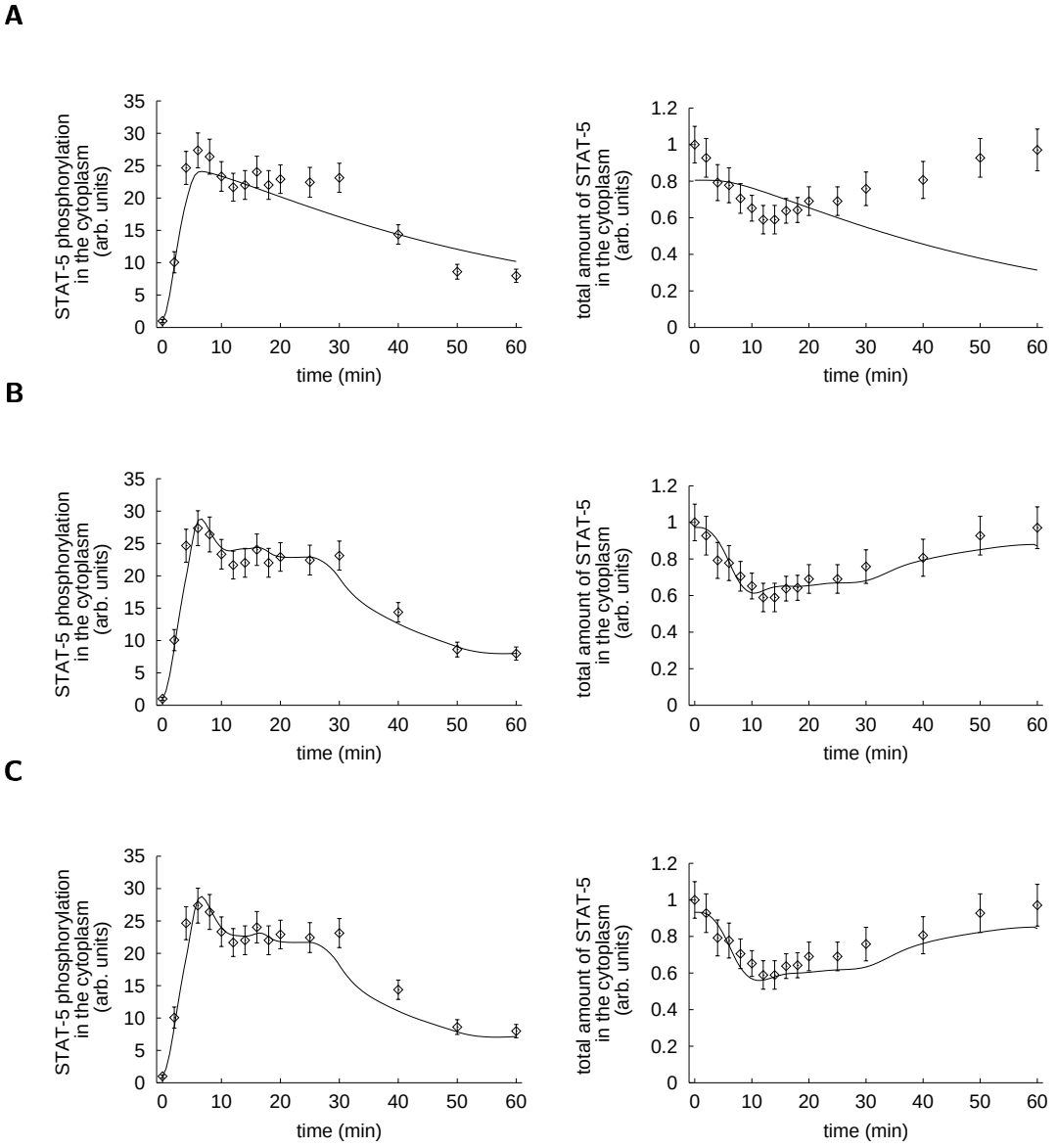

Fig. 7. Time course (points with error bars) and mathematical modeling (solid lines) of the STAT5 signaling pathway. (A) Fit with model 2, the feed-forward model with backreaction. (B) Fit with model 4, the proposed model (see text) with additional backreaction. (C) Fit with model 5, the proposed model (see text) with additional delay-distribution. The experimental data and the corresponding fits are shown for cytoplasmic tyrosine phosphorylated STAT5 ($Left$) and total cytoplasmic STAT5 ($Right$).

We also compared model 3 to a model that describes the nuclear cycling by a compartment model. Because these models are not nested, a bootstrap procedure was applied. As a result, model 3 was preferred to the compartment model. For a detailed discussion of the derivation of the test statistic, see ref 3.

Finally, we relaxed our assumption of a fixed sojourn time τ for STAT5 in the nucleus. We investigated whether an introduction of a distribution of delays (model 5) improves the description of the experimental data significantly. The delay distribution was realized by folding the time course of $x_3(t)$ with a Gaussian kernel with mean delay time τ_0 and window width w. The new time course of STAT5 leaving the nucleus $x_3^G(t)$ is computed as

$$x_3^G(t) = \frac{1}{\sqrt{2\pi}w} \int_{-\infty}^{+\infty} exp\left(-\frac{(t-t'-\tau_0)^2}{2w^2}\right) x_3(t')dt'$$

and the new differential equations are:

$$\begin{aligned}
\dot{x}_1 &= -k_1 x_1 \mathrm{EpoR_A} \quad + 2k_4 x_3^G(t) & y_1 &= k_5(x_2 + 2x_3) \\
\dot{x}_2 &= +k_1 x_1 \mathrm{EpoR_A} - k_2 x_2^2 & y_2 &= k_6(x_1 + x_2 + 2x_3) \\
\dot{x}_3 &= -k_3 x_3 + 0.5 k_2 x_2^2 & y_3 &= k_7 \mathrm{EpoR_A} \\
\dot{x}_4 &= +k_3 x_3 \quad - k_4 x_3^G(t). &&
\end{aligned}$$

The resulting fit is displayed in Fig. 7*C*. Using the same analysis as above, we compute the test statistic as $LR = 2(LR_5 - LR_3) = 0.55$, corresponding to a p-value of 0.23. Thus we conclude that this model also implies an insignificant enlargement.

In summary, we conclude that the model with nucleocytoplasmic cycling but with neither the backreaction nor the delaydistribution is the appropriate model. It represents a good tradeoff between the complexity of the mathematical description and its ability to represent the experimental data.

D. Parameter identifiability

Due to the observation function of the system, which allows only limited measurements of the dynamical behavior, not all dynamical and nuisance parameters as well as the initial value $x_1(0)$ can be extracted from the data. This implies that the system is not fully identifiable (4). Therefore, it is possible to obtain a dynamical system with less parameters consisting of the same observation function. The following transformation of variables leads to such an identical representation:

$$\begin{aligned}
v_i &= k_2 x_i & r_1 &= \frac{k_1}{k_7} & r_3 &= k_3 \\
r_4 &= k_4 & r_5 &= \frac{k_5}{k_2} & r_6 &= \frac{k_6}{k_2}.
\end{aligned}$$

Therefore we are only able to extract single dynamical parameters k_3, k_4, and τ and the parameter combinations $k_2 x_1(0), k_1/k_7, k_5/k_2$, and k_6/k_2 from the measured data. The new system of differential equations reads as follows:

$$\begin{aligned}
\dot{v}_1 &= -r_1 v_1 D \quad + 2r_4 v_3(t-\tau) & y_1 &= r_5(v_2 + 2v_3) \\
\dot{v}_2 &= +r_1 v_1 D \quad - 0.5 v_2^2 & y_2 &= r_6(v_1 + v_2 + 2v_3) \\
\dot{v}_3 &= -r_3 v_3 \quad + v_2^2 & y_3 &= k_7 \mathrm{EpoR_A} = D \\
\dot{v}_4 &= +r_3 v_3 \quad - r_4 v_3(t-\tau). &&
\end{aligned}$$

It is noteworthy that the most interesting parameters, k_3, k_4, and τ, which determine the cycling behavior, are not affected by this transformation and may be extracted with the help of the measured data, and that the non-identifiability of the parameters does not affect the model selection procedure described above.

References

1. Cox, D. & Hinkley, C. (1974) *Theoretical Statistics* (Chapman & Hall, London).
2. Self, S. G. & Liang, K. Y. (1987) *J. Am. Stat. Assoc.* **82**, 605-610.
3. Timmer, J., Müller,T.G., Swameye,I., Sandra, O. & Klingmüller, U. (2003) *Int. J. Bif. Chaos*, in press.
4. Ljung, L. (1999) *System Identification: Theory For The User* (Prentice-Hall, Englewood Cliffs, NJ).

Addition to methods

Preparation of nuclear extracts

For nuclear extracts cells were lysed in hypotonic buffer (20 mM Hepes, pH 7.9/10 mM KCl/0.1 mM Na_3VO_4/1 mM EDTA/0.2% Nonidet P-40/10% glycerol/1 mM DTT/1 mM PMSF). After centrifugation the nuclei were resuspended in high salt buffer (420 mM NaCl/20% glycerol/20 mM Hepes, pH 7.9/10 mM KCl/0.1 mM Na_3VO_4/1 mM EDTA/1 mM DTT/0.1% Nonidet P-40/1 mM PMSF). After three freezethaw cycles, the samples were centrifuged, and the supernatant was used for immunoprecipitation with anti-STAT5 antiserum.