Orthogonal optogenetic triple-gene control in mammalian cells

Konrad Müller[1], Raphael Engesser[2,3], Jens Timmer[2,3,4], Matias D. Zurbriggen[1,3] & Wilfried Weber[1,3,4,*]

[1] Faculty of Biology, University of Freiburg, Schänzlestrasse 1, 79104 Freiburg, Germany

[2] Institute of Physics, University of Freiburg, Hermann-Herder-Str. 3, 79104 Freiburg, Germany

[3] BIOSS Centre for Biological Signalling Studies, University of Freiburg, Schänzlestrasse 18, 79104 Freiburg, Germany

[4] Freiburg Centre for Biosystems Analysis (ZBSA), University of Freiburg, Habsburgerstrasse 49, 79104 Freiburg, Germany

* Corresponding author: Fax: +49 761 203 97660; Tel: +49 761 203 97654; E-mail: wilfried.weber@biologie.uni-freiburg.de

Supporting Information

Figure S1	Multi-chromatic multi-gene expression control in a single cell culture.
Figure S2	Titration of the 660-nm intensity in the presence of 450-nm illumination.
Table S1	Expression vectors and oligonucleotides designed and used in this study.
Math. Model S1	Modeling of the LightON system.
Math. Model S2	Modeling of the rapidly-reversible blue light system.

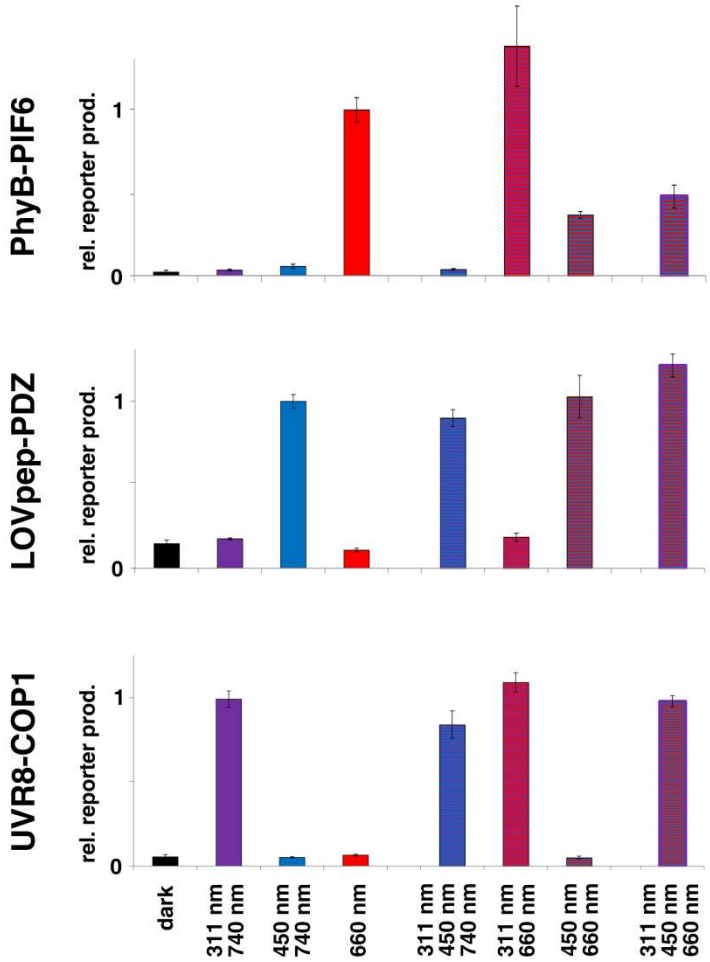


Figure S1 Multi-chromatic multi-gene expression control in a single cell culture. CHO-K1 cells were simultaneously transfected for red light-inducible SEAP production (pKM301, pMF199), blue light-responsive FLuc expression (pKM516, pFR-Luc) and UV-B light-controlled angiopoietin 1 (Ang1) production (pKM279, pKM172). Twenty-four hours post-transfection, the culture medium was replaced with PCB-supplemented medium. After 1 h in the dark, the cells were either incubated in the dark or illuminated as indicated for 24 h before reporter quantification. Illumination conditions: 311 nm (2 µmol m^{-2} s^{-1}, 2 min ON/28 min OFF), 450 nm (10 µmol m^{-2} s^{-1}), 660 nm (2 µmol m^{-2} s^{-1}), 740 nm (20 µmol m^{-2} s^{-1}). Data are means ± SD (n=4).

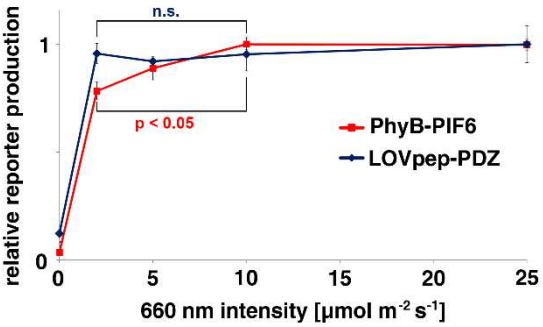


Figure S2 Titration of the 660-nm intensity in the presence of 450-nm illumination. CHO-K1 cells were simultaneously transfected for blue light-inducible FLuc expression (pKM516, pFR-Luc) and for red/far-red light-switchable SEAP production (pKM301, pMF199). Twenty-four hours post-transfection, the culture medium was replaced with fresh PCB-containing medium. After incubation in the dark for 1 h, the cells were illuminated with 450-nm light (10 µmol m^{-2} s^{-1}) that was supplemented with 660-nm light of increasing intensities for 24 h prior to the quantification of reporter production. Data are means ± SD (n=4). Statistics were performed by the two-tailed t-test n.s., not significant.

Table S1 Expression vectors and oligonucleotides designed and used in this study.

Plasmid	Description	Ref. or source
pDS221	Vector encoding ePDZb1-mCherry under control of P_{TEF} (P_{TEF}-ePDZb1-mCherry-pA)	(*1*)
pDS271	Vector encoding Mid2-GFP-LOVpep[T406A,T407A,V416I] und control of P_{TEF} (P_{TEF}-Mid2-GFP-LOVpep[T406A, T407A, V416I])	(*1*)
pFR-LUC	Plasmid encoding FLuc under control of P_{Gal4} ($(UAS_G)_5$-TATA-FLuc-pA)	Agilent
pKM001	Vector encoding SEAP under the control of P_{Tet} harboring 8 repeats of the etr operator site between the heptameric tetO operator and the minimal promoter ($tetO_7$-etr_8-$P_{hCMVmin}$-SEAP-pA)	(*2*)
pKM006	Vector encoding SEAP under the control of a modified P_{Tet} harboring a 422 bp spacer between the 13-mer tetO operator and the minimal promoter ($tetO_{13}$-422bp-$P_{hCMVmin}$-SEAP-pA)	(*2*)
pKM018	Vector encoding P_{SV40}-driven expression of PhyB(1-650)-VP16-NLS (P_{SV40}-PhyB(1-650)-VP16-NLS)	(*2*)
pKM022	Bicistronic vector encoding PhyB(1-650)-VP16-NLS and TetR-PIF6(1-100)-HA under control of P_{SV40} (P_{SV40}-PhyB(1-650)-VP16-NLS-$IRES_{PV}$-TetR-PIF6(1-100)-HA-pA)	(*2*)
pKM081	Vector encoding SEAP under control of a modified P_{ETR} (etr_8-$P_{hCMVmin}$-SEAP-pA)	(*3*)
pKM083	Vector encoding GLuc under control of P_{Gal4} ($(UAS_G)_5$-TATA-GLuc-pA) GLuc with secretion signal was amplified using oligos oKM213 (5'-tgactaggtaccgttcgagatctgcgatctaagtaagcttggccaccatgggagtcaaagttctgtttgccctgatctgcatcgctgtggccgaggccAAGCCCACCGAGAACAACGAAG-3') and oKM214 (5'-tgactaccaaaggatgggcggccgcttaGTCACCACCGGCCCCC-3'), digested (*Kpn*I/*Pfl*MI) and ligated (*Kpn*I/*Pfl*MI) into pFR-LUC.	This work
pKM084	Vector encoding SEAP under control of P_{Gal4} ($(UAS_G)_5$-TATA-SEAP-pA) SEAP was amplified from pKM001 using oligos oKM215 (5'-tgactaggtaccgttcgagatctgcgatctaagtaagcttggccaccATGCTGCTGCTGCTGCTGC-3') and oKM216 (5'-tgactagcggccgcTTAACCCGGGTGCGCGGC-3'), digested (*Kpn*I/*Not*I) and ligated into pKM083 (*Kpn*I/*Not*I).	This work
pKM085	Vector encoding Gal4(65)-VVD-p65 under control of $P_{EF1\alpha}$ ($P_{EF1\alpha}$-Gal4(65)-VVD-p65-pA)	(*3*)
pKM115	Vector encoding COP1(WD40)-VP16 under control of P_{SV40} (P_{SV40}-COP1(WD40)-VP16-pA)	(*3*)
pKM168	Vector encoding E-UVR8(12-381) under control of P_{SV40} (P_{SV40}-E-UVR8(12-381)-pA)	(*3*)
pKM172	Vector encoding Ang1 under control of a modified P_{ETR} (etr_8-$P_{hCMVmin}$-Ang1-pA)	(*3*)
pKM248	Vector encoding COP1(WD40)-p65 under control of P_{SV40} (P_{SV40}-COP1(WD40)-p65-pA) p65 was amplified from pKM085 using oligos oKM340 (5'-caagtcgcgcgcttGAGTTCCAGTACCTGCCTGAC-3') and oKM341 (5'-caagtcggatccTCATTTGTCGTCGTCGTCCTTG-3'), digested (*Bss*HII/*Bam*HI) and ligated (*Bss*HII/*Bam*HI) into pKM115.	This work
pKM279	Bicistronic vector encoding E-UVR8(12-381) and COP1(WD40)-p65 under control of P_{SV40} (P_{SV40}-E-UVR8(12-381)-$IRES_{PV}$-COP1(WD40)-p65-pA) COP1(WD40)-p65 was excised (*Not*I/*Mfe*I) from pKM248, while $IRES_{PV}$ was excised (*Hin*dIII/*Not*I) from pKM022. Both fragments were ligated (*Hin*dIII/*Mfe*I) into pKM168.	This work
pKM290	Vector encoding ePDZb1-VP16-NLS under control of P_{SV40} (P_{SV40}-ePDZb1-VP16-NLS-pA) ePDZb1 was amplified from pDS221 using oligos oKM396 (5'-caagtcgcggccgccccaccATGCCAGAACTTGGATTTAGCATATC-3') and oKM397 (5'-caagtcgcgcgccggcgccGGTACGGTAGTTAATCGAGATTGG-3'), digested (*Not*I/*Bss*HII) and ligated (*Not*I/*Bss*HII) into pKM018.	This work
pKM291	Vector encoding P_{SV40}-controlled expression of Gal4BD-LOVpep[T406A,T407A] (P_{SV40}-Gal4BD-LOVpep[T406A,T407A]-pA) Gal4BD was amplified using oligos oKM398 (5'-tcttttatttcaggtcccggatcgaattgcggccgcGAATTCGCCACCATGAAGCTAC-3') and oKM399 (5'-agccaaacttccacctgaacctccagatccGCCGGTACCCGATACAGTC-3'), while LOVpep[T406A,T407A] was amplified from pDS271 using oligos oKM500 (5'-gggtaccggcggatctggaggttcaggtggaagtTGGCTGCTGCACTTGAACGTATTGAGAAGAACTTTgTCATTAC-3') and oKM501 (5'-tgtctggatcgaagcttgggctgcaggtcgactctagaggatccTTACACCCAGGTATCCACCGC-3'). Both fragments were fused with *Eco*RI/*Bam*HI-digested pSAM200 by Gibson cloning.	This work

pKM292	**Vector encoding P_{SV40}-controlled expression of Gal4BD-LOVpep[T406A,T407A,I532A] (P_{SV40}-Gal4BD-LOVpep[T406A,T407A,I532A]-pA)** Gal4BD was amplified using oligos oKM398 (5'-tcttttatttcaggtcccggatcgaattgcggccgcGAATTCGCCACCATGAAGCTAC-3') and oKM399 (5'-agccaaacttccacctgaacctccagatccGCCGGTACCCGATACAGTC-3'), while LOVpep[T406A,T407A,I532A] was amplified from pDS271 using oligos oKM500 (5'-gggtaccggcggatctggaggttcaggtggaagtTTGGCTGCTGCACTTGAACGTATTGAGAAGAACTTTgTCATTAC-3') and oKM502 (5'-tgtctggatcgaagcttgggctgcaggtcgactctagaggatccTTACACCCAGGTATCCACCGCTTTATCAATCTCTTCTGC AGTTTTCTTAgcCAGCATG-3'). Both fragments were fused with *Eco*RI/*Bam*HI-digested pSAM200 by Gibson cloning.	This work
pKM293	Vector encoding P_{SV40}-controlled expression of Gal4BD-LOVpep (P_{SV40}-Gal4BD-LOVpep-pA) Gal4BD was amplified using oligos oKM398 (5'-tcttttatttcaggtcccggatcgaattgcggccgcGAATTCGCCACCATGAAGCTAC-3') and oKM399 (5'-agccaaacttccacctgaacctccagatccGCCGGTACCCGATACAGTC-3'), while LOVpep was amplified from pDS271 using oligo oKM503 (5'-gggtaccggcggatctggaggttcaggtggaagtTTGGCTaCTaCACTTGAACGTATTGAGAAGAACTTTgTCATTAC-3') and oKM501 (5'-tgtctggatcgaagcttgggctgcaggtcgactctagaggatccTTACACCCAGGTATCCACCGC-3'). Both fragments were fused with *Eco*RI/*Bam*HI-digested pSAM200 by Gibson cloning.	This work
pKM296	Vector encoding PDZ-VP16-NLS under control of P_{SV40} (P_{SV40}-PDZ-VP16-NLS-pA) PDZ was amplified from pDS221 using oligos oKM396 (5'-caagtcgcggccgccaccATGCCAGAACTTGGATTTAGCATATC-3') and oKM504 (5'-caagtcgcgcgccggcgccAACTTCTCGTACAATGATGAGTTCAAC-3'), digested (*Not*I/*Bss*HII) and ligated (*Not*I/*Bss*HII) into pKM018.	This work
pKM297	**Vector encoding ePDZb-VP16-NLS under control of P_{SV40} (P_{SV40}-ePDZb-VP16-NLS-pA)** The N-terminal part of ePDZb was amplified from pDS221 using oligos oKM396 (5'-caagtcgcggccgccaccATGCCAGAACTTGGATTTAGCATATC-3') and oKM505 (5'-GTTACCACCGGTTTCACCGTACGTGATACGGTAATAACTAACatgagagtcgtaATACGCATCCCAGC-3' digested (*Not*I/*Age*I) and ligated (*Not*I/*Age*I) into pKM290.	This work
pKM298	Vector encoding P_{SV40}-controlled expression of Gal4BD-LOVpep[V529N] (P_{SV40}-Gal4BD-LOVpep[V529N]-pA) Gal4BD-LOVpep[V529N] was amplified from pKM293 using oligos oKM398 (5'-tcttttatttcaggtcccggatcgaattgcggccgcGAATTCGCCACCATGAAGCTAC-3') and oKM506 (5'-TCTAGAGGATCCTTACACCCAGGTATCCACCGCTTTATCAATCTCTTCTGCAGTTTTCTTAATCAGCATGttTGCCTCTC TCTCGGC-3'), digested (*Eco*RI/*Bam*HI) and ligated (*Eco*RI/*Bam*HI) into pSAM200.	This work
pKM301	Bicistronic vector encoding PhyB(1-650)-VP16-NLS and PiP-PIF6(1-100)-HA under control of P_{SV40} (P_{SV40}-PhyB(1-650)-VP16-NLS-$IRES_{PV}$-PiP-PIF6(1-100)-HA-pA)	(*4*)
pKM516	**Bicistronic vector encoding Gal4BD-LOVpep[T406A,T407A,I532A] and ePDZb-VP16-NLS under control of P_{SV40} (P_{SV40}-Gal4BD-LOVpep[T406A,T407A,I532A]-$IRES_{PV}$-ePDZb-VP16-NLS-pA** ePDZb-VP16-NLS-pA was excised (*Not*I/*Mfe*I) from pKM297, while $IRES_{PV}$ was excised (*Hin*dIII/*Not*I) from pKM022. Both fragments were ligated (*Hin*dIII/*Mfe*I) into pKM292.	This work
pMF199	Vector encoding SEAP under control of P_{PIR3} (PIR_3-$P_{HSP70min}$-SEAP-pA)	(*5*)
pSAM200	Constitutive TetR-VP16 expression vector (P_{SV40}-TetR-VP16-pA)	(*6*)

Ang1, angiopoietin 1; COP1, E3 ubiquitin ligase COP1; E, macrolide-responsive repressor protein; etr, operator sequence binding E; ePDZb/ePDZb1, enhanced PDZ domains engineered by fusion to the fibronectin FN3 domain and affinity maturation; FLuc, firefly luciferase; Gal4(65), amino acids 1-65 of the Gal4 DNA binding domain; Gal4BD, Gal4 DNA binding domain; GFP, green fluorescent protein; GLuc, gaussia luciferase; HA, human influenza hemagglutinin-derived epitope tag; $IRES_{PV}$, polioviral internal ribosome entry site; LOV, light-oxygen-voltage photoreceptor domain; LOVpep, engineered LOV2 domain from *Avena sativa* phototropin 1 with an epitope tag fused to its C-terminal Jα-helix;

NLS, nuclear localization signal from simian virus 40 large T antigen; p65, transactivation domain from nuclear factor of activated B cells (NF-κB); pA, polyadenylation signal; PDZ, Erbin PDZ domain; $P_{EF1\alpha}$, human elongation factor 1α promoter; P_{ETR}, macrolide-responsive promoter; P_{Gal4}, Gal4-responsive promoter; $P_{hCMVmin}$, minimal human cytomegalovirus immediate early promoter; $P_{HSP70min}$, minimal heat-shock protein 70 promoter from Drosophila; PhyB, Phytochrome B; PhyB(1-650), N-terminus of Phytochrome B with amino acids 1-650; PIF6, Phytochrome-interacting-factor 6; PIF6(1-100), N-terminus of Phytochrome-interacting-factor 6 with amino acids 1-100; PiP, pristinamycin-induced protein; PIR, operator sequence binding PiP; P_{SV40}, simian virus 40 early promoter; P_{Tet}, tetracycline-responsive promoter; P_{TEF}, translation elongation factor 1α promoter from yeast; SEAP, human placental secreted alkaline phosphatase; tetO, operator sequence binding TetR; TetR, tetracycline repressor protein; UAS_G, Gal4 binding site; UVR8, UV-resistance protein UVR8; UVR8(12-381), core domain of UVR8 consisting of amino acids 12-381; VP16, *Herpes simplex* virus-derived transactivation domain; VVD, vivid; WD40; WD40 domain.

Uppercase in oligos, annealing sequence; underlined sequence, restriction site. Bold, components of the optimized blue light-inducible LOVpep-PDZ-based gene switch.

Mathematical Model S1 Modeling of the LightON system.

1 Development of a mathematical model

In order to construct an orthogonal blue light system and find targets to modify the LightON system (*3, 7*), we developed a mathematical model describing the LightON system. This model is fitted to the intensity response data shown in Figure 1B. The LightON system is based on inactive monomers M which are homodimerized by light:

$$\mathrm{M} + \mathrm{M} \underset{k_{off}}{\overset{k_{on}\, I}{\rightleftharpoons}} \mathrm{C} \qquad \text{(R1)}$$

The light-induced homodimerization rate is proportional to the light intensity I with the parameter k_{on}. The complex C dissociates with the constant rate k_{off}. The homodimer C activates the expression of the target gene G. The reaction (R1) is fast compared to the time scale of the gene expression. Therefore it is in a chemical equilibrium with an equilibrium concentration $[C]_{eq} = [C]_{eq}(I)$ depending on the light intensity I.

According to the law of mass action it is

$$\frac{[C]_{eq}}{[M]_{eq}^2} = I\,\frac{k_{on}}{k_{off}} := I\,K_b.$$

With mass conservation $[C]_{eq} = (M_{total} - [M]_{eq})/2$ the equilibrium concentration of the active complex C depending on the light intensity I is

$$[C]_{eq}(I) = \frac{M_{total}}{2} + \frac{1}{8K_bI} - \sqrt{\frac{1}{(8K_bI)^2} + \frac{M_{total}}{8K_bI}}.$$

The expression of the target gene is described by the ordinary differential equation (ODE)

$$\frac{d[G](t,I)}{dt} = k_{expression}[C]_{eq}(I), \qquad \text{with } [\mathrm{G}](0) = 0.$$

Since $[C]_{eq}(I)$ does not depend on time this equation can easily be integrated:

$$[G](I,t) = k_{expression}[C]_{eq}(I)\ t$$

We measured the response to different light intensities at $t = 24h$. The concentration of the target gene for a fixed time point depends only on the light intensity I:

$$[G](I, t = 24h) = k_{expression}[C]_{eq}(I)\ 24h = \alpha[C]_{eq}(I) = [G](I)$$

with $\alpha = k_{expression}\ 24h$.

This leads to

$$[G](I) = \alpha\left(\frac{M_{total}}{2} + \frac{1}{8K_bI} - \sqrt{\frac{1}{(8K_bI)^2} + \frac{M_{total}}{8K_bI}}\right) = \left(\frac{M'_{total}}{2} + \frac{1}{8K'_bI} - \sqrt{\frac{1}{\left(8K'_bI\right)^2} + \frac{M'_{total}}{8K'_bI}}\right)$$

with $K'_b = \alpha^{-1}K_b$ and $M'_{total} = \alpha M_{total}$.

To model the measurement error we used an error model with a constant Gaussian error $\epsilon \sim N(0, \sigma^2)$ with variance σ^2 for all data points.

2 Fitting the model to the experimental data

The three unknown parameters $\boldsymbol{\theta} = (K'_b, M'_{total}, \sigma)$ were fitted to the data shown in Figure 1B. The data can be written as the vector

$$\boldsymbol{y} = (y_{I_1}, y_{I_2}, \dots, y_{I_{n_{data}}})$$

Were n_{data} is the number of data points y_{I_k} of the target gene for different light intensities I_k. The fitting was performed by maximizing the likelihood function

$$L(\boldsymbol{y}|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sigma}\prod_{k=1}^{n_{data}} e^{-\frac{(y_{I_k} - [G](I_k))^2}{2\,\sigma^2}}$$

with respect to the parameter vector $\boldsymbol{\theta}$. Instead of maximizing $L(\boldsymbol{y}|\boldsymbol{\theta})$ it is equivalent to minimize $-2\log(L)$. In our case for Gaussian distributed errors $-2\log(L)$ is the sum of the squared weighted residuals $\chi^2 = \sum_{k=1}^{n_{data}} \frac{(y_{I_k} - [G](I_k))^2}{\sigma^2}$ with a second term due to the error model

$$-2\log(L) = \chi^2 + 2\, n_{data} \log\left(\sqrt{2\pi}\sigma\right) =: \chi^2_{mod}.$$

The optimal $\boldsymbol{\theta_{opt}}$ parameter set is obtained by

$$\boldsymbol{\theta_{opt}} = \arg\min_{\boldsymbol{\theta}} \chi^2_{mod}\,(\boldsymbol{\theta}, \boldsymbol{y}).$$

The optimization was performed with a trust region algorithm implemented in MATLAB (lsqnonlin) (*8*). The trust region algorithm is a local optimization method. To find the global minimum we performed multiple optimization runs were the initial parameters were sampled largely over the parameter space. To obtain a good coverage of the parameter space we used a latin hypercube sampling.

To estimate the parameter uncertainty in terms of 95% confidence intervals we calculated the profile likelihood (PL) for each parameter θ_i (*9*). The PL is defined by

$$PL(\theta_j) = \min_{\boldsymbol{\theta_{i\neq j}}} \chi^2_{mod}\,(\boldsymbol{\theta}, \boldsymbol{y}).$$

It gives information about the identifiability of each parameter. The optimization and uncertainty analysis was done with the Data2Dynamics software (*10*).

3 Results

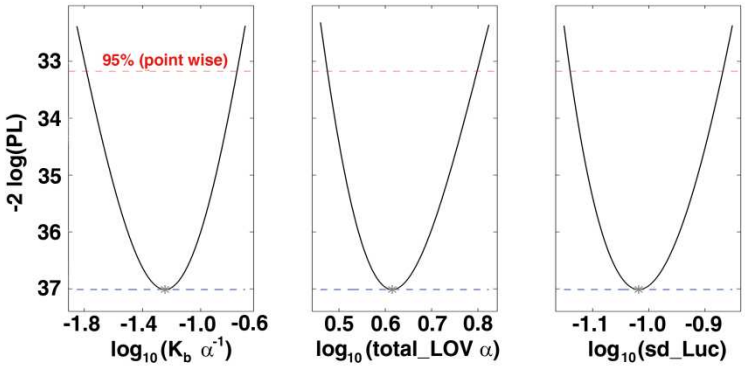
The model with three unknown parameters was fitted to 20 data points. As optimum we found $\chi^2_{mod} = -2\log(L) = -37.01$. The estimated parameter set is shown in the table below.

Fitted model parameters of the LightOn system obtained by a maximum likelihood estimation.

Parameter	$\boldsymbol{\theta_{opt,i}}$	Unit
$K_b' = \alpha^{-1} K_b$	5.68×10^{-2}	RRP^*
$M_{total}' = \alpha M_{total}$	4.12	$(RRP \cdot \mu mol\, m^{-2}\, s^{-1})^{-1}$
σ	9.59×10^{-2}	RRP

* RRP = relative reporter production

To scan many orders of magnitude of the parameter space the latin hypercube sampling was performed on a logarithmic scale from -6 to +4 corresponding to 10 orders of magnitude. We performed 100 optimization runs. The best parameter set was found over 77 % which is a strong hint that we found the global optimum. The uncertainty analysis by exploiting the profile likelihood suggests that all parameters are identifiable. The profile likelihood is shown below.



Profile likelihood of the estimated model parameters for the intensity-dependent activation of the LightON system. The solid lines indicate the profile likelihood. The optimal parameter set is marked with a grey star and the red dashed line marks the 95 % confidence level.

Simulations with the parameterized model are demonstrated that the sensitivity of the system to the light intensity is controlled by the binding constant K_b (Figure 1B). The LightOn system is already activated at low intensities in the same range as the UVR8-COP1 system. To construct a system which is orthogonal to the UVR8-COP1 system one has to shift the sensitivity to higher intensities. We realized this by a system with a much faster dark conversion rate k_{off}, resulting in a decrease of K_b. This system is analyzed in Mathematical Model S2.

Mathematical Model S2 Modeling of the rapidly-reversible blue light system.

1 Development of a mathematical model

The first component of the rapidly-reversible blue light system (Figure 2A) is Gal4(BD)-LOVpep which can bind to the DNA upstream of a minimal promoter. The concentration of DNA binding sites with a bound Gal4(BD)-LOVpep protein is assumed to be in a chemical equilibrium and denoted by LOV_{DNA}. The transactivator VP16 is fused to the PDZ interaction domain. Illumination with blue light of the intensity I induces the binding of PDZ-VP16 with LOV:

$$LOV_{DNA} + VP16 \underset{k_{off}}{\overset{k_{on}\, I}{\rightleftharpoons}} LOVVP16 \qquad \text{(R2)}$$

This reaction was modeled with mass action kinetics with the binding rate $k_{on}\, I$ and a linear dark revision rate k_{off}. According to (*1*) the active state has a half-life time of 17 s which corresponds to $k_{off} = 1.47 \times 10^2\ h^{-1}$. The resulting complex LOVVP16 triggers the expression of the target gene SEAP. Since we have time course and intensity dose response data we modeled this system with ordinary differential equations. For future applications with combinations with the UVR8-COP1 and the red/far-red light system the complexity of the model was chosen on the same level as in the other two systems (*2, 3*). The full ODE model is:

$$\frac{d[VP16](t)}{dt} = k_{off}[LOVVP16] - k_{on}\, I\, [LOV_{\mathrm{DNA}}]\, [VP16] - k_{growth}\, [VP16] \qquad (1)$$

$$\frac{d[LOV_{\mathrm{DNA}}](t)}{dt} = k_{off}[LOVVP16] - k_{on}\, I\, [LOV_{\mathrm{DNA}}]\, [VP16]\ - k_{growth}\, [LOV_{\mathrm{DNA}}] \qquad (2)$$

$$\frac{d[LOVVP16](t)}{dt} = -k_{off}[LOVVP16] + k_{on}\, I\, [LOV_{\mathrm{DNA}}]\, [VP16] \qquad (3)$$

$$\frac{d[mRNA_{nuc}](t)}{dt} = k_{basal,mRNA_{nuc}} + k_{prod,mRNA_{nuc}}[LOVVP16] - k_{mRNA,nuc2cyt}\,[mRNA_{nuc}] \quad (4)$$

$$\frac{d[mRNA_{cyt}](t)}{dt} = k_{nuc2cyt}\,[mRNA_{nuc}] - k_{deg,mRNA_{nuc}}[mRNA_{cyt}] \quad (5)$$

$$\frac{d[SEAP](t)}{dt} = k_{transl,SEAP}\,\left[mRNA_{cyt}\right] N \quad (6)$$

$$\frac{dN(t)}{dt} = k_{growth}\,N \quad (7)$$

The LOVVP16 complex directly activates the transcription of the mRNA. The basal mRNA activation is included by the parameter $k_{basal,mRNA_{cyt}}$. Like in the UVR8-COP1 model we included a second mRNA compartment in the cytoplasm where it is linearly degraded with the rate $k_{deg,mRNA_{nuc}}$. The cytoplasmic mRNA induces the production of the target gene SEAP. Since SEAP is secreted to the medium its production is proportional to the number of cells N. The cells are growing with a growth rate k_{growth}. Like in (*3*) we assumed a doubling time of 14 h which corresponds to $k_{growth} = 4.95 \times 10^{-2} h^{-1}$. The plasmids are transfected transiently into the cells and are therefore not replicated during cell division. This means the concentration of VP16 and LOV_{DNA} is decreasing with time with the rate of the cell growth.

As initial conditions we used:

$$[VP16](0) = init_{VP16}$$

$$[LOV_{\mathrm{DNA}}](0) = 1$$

$$[LOVVP16](0) = 0$$

$$[mRNA_{nuc}](0) = \frac{k_{basal,mRNA_{nuc}}}{k_{mRNA,nuc2cyt}}$$

$$\left[mRNA_{cyt}\right](0) = \frac{k_{basal,mRNA_{nuc}}}{k_{deg,mRNA_{nuc}}}$$

$$[SEAP](0) = init_{SEAP}$$

$$N(0) = 1$$

The measured output of the system is the SEAP concentration in the medium. The concentration scale of VP16, LOV_{DNA} and LOVVP16 is not assessable, therefore we could model VP16 and LOV_{DNA} in relative units. This is done by setting the initial concentration of LOV_{DNA} to one. This means VP16 and LOVVP16 is measured in relative amounts of the initial LOV_{DNA} concentration. The initial concentration of VP16 is a free parameter. There is no LOVVP16 complex in the beginning since the binding is only induced by light. The initial mRNA concentration is set to the steady state without light input and with no cell growth. The SEAP concentration in the beginning is a free parameter. It is sufficient to consider the relative cell growth, therefore the initial cell number is set to one.

The parameters $k_{nuc2cyt}$ and $k_{deg,mRNA_{nuc}}$ were taken from the UVR8-COP1 model (*3*):

$$k_{nuc2cyt} = 0.1597\ h^{-1}$$

$$k_{deg,mRNA_{nuc}} = 0.5312\ h^{-1}$$

2 Fitting the model to the experimental data

This model was fitted to the data shown in Figure 2C and Figure 2D. In the time course experiment in Figure 2C we measured the SEAP concentration for three different illumination conditions. The cells were either illuminated with (i) 450-nm light with the intensity of 10 µmol m^{-2} s^{-1}, (ii) 6 h with the same intensity and then incubated in the dark or (iii) incubated for 24 h in the dark. These experimental conditions were modeled with:

(i) $I(t) = 10\ \mu mol\ m^{-2}s^{-1}$

(ii) $I(t) = 10\ \mu mol\ m^{-2}s^{-1}$ for $t < 6\ h$

$I(t) = 0\ \mu mol\ m^{-2}s^{-1}$ for $t > 6\ h$

(iii) $I(t) = \quad 0\ \mu mol\ m^{-2}s^{-1}$

The model was linked with the data by the function

$$\left[SEAP_{t,meassured}\right] = [SEAP](t).$$

The scale of SEAP was defined by this experiment.

In the intensity dose response experiment (Figure 2D) the cells were illuminated with different light intensities I. After 24 h the SEAP concentration was measured. The data of this experiment were linked to the model with the function

$$\left[SEAP_{I,meassured}\right] = scale_{dose\_resp}\ [SEAP](t = 24\ h, I).$$

The measurement error of the two experiments was modeled with an independent constant Gaussian error $\epsilon_i \sim N(0, \sigma_i^2)$ for each experiment $i = 1{,}2$.

The fitting was done by the same methods described in Text E1, section 2.

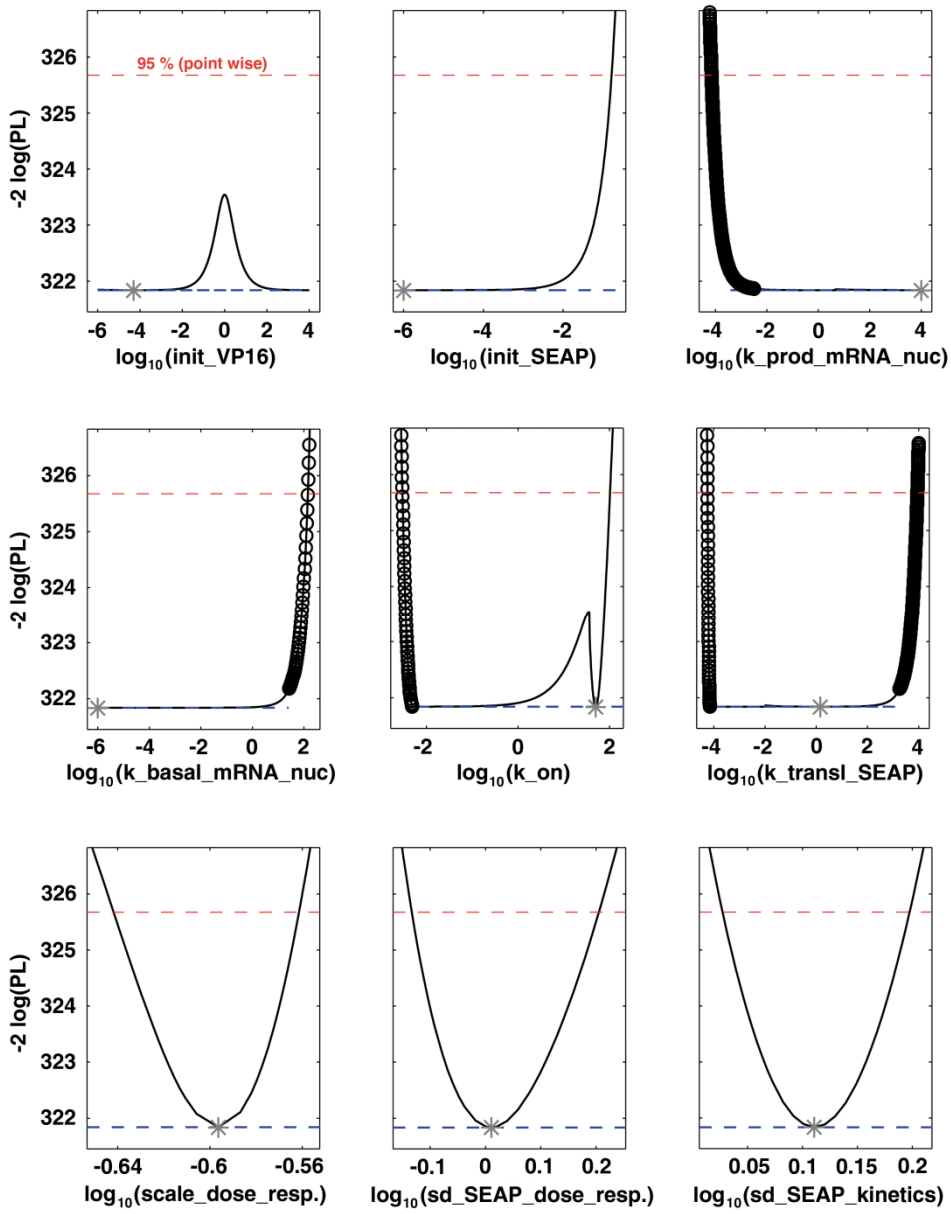
3 Results

The model with 9 unknown parameters was fitted to 100 data points. As optimum we found $\chi^2_{mod} = -2\log(L) = 321.8$. The estimated parameter set is shown in in the table below.

Fitted model parameters for the rapidly-reversible blue light system obtained by a maximum likelihood estimation.

Parameter	$\boldsymbol{\theta_{opt,i}}$	Unit	Comment
k_{on}	5.00×10^{-2}	$\left(h \cdot \mu mol\, m^{-2}\, s^{-1} \cdot\ [LOV_{\mathrm{DNA}}](0)\right)^{-1}$	fitted
k_{off}	0.637	h^{-1}	(*1*)
$k_{transl,SEAP}$	0.717	$U/L\ \cdot [LOV_{\mathrm{DNA}}](0)^{-1} \cdot h^{-1}$	fitted
$init_{VP16}$	1.00×10^{4}	$[LOV_{\mathrm{DNA}}](0)$	fitted
$sd_{SEAP,dose\_resp}$	1.01	U/L	fitted
$sd_{SEAP,kinetics}$	1.27	U/L	fitted
$scale_{dose\_resp}$	0.253	1	fitted
$init_{SEAP}$	0	U/L	according PL
$k_{basal,mRNA_{nuc}}$	0	$[LOV_{\mathrm{DNA}}](0)$	according PL
$k_{prod,mRNA_{nuc}}$	1	h^{-1}	after rescaling [mRNA]
k_{growth}	4.95×10^{-2}	h^{-1}	(*3*)
$k_{nuc2cyt}$	0.160	h^{-1}	(*3*)
$k_{deg,mRNA_{nuc}}$	0.531	h^{-1}	(*3*)

To scan many orders of magnitude of the parameter space the latin hypercube sampling was performed on a logarithmic scale from -6 to +4 corresponding to 10 orders of magnitude. We performed 100 optimization runs. The best parameter set was found over 81 % which is a strong hint that we found the global optimum. The profile likelihood functions are shown below. According to the profiles multiple parameters are not identifiable.



Profile likelihood of the estimated model parameters for the LOVpep-PDZ-based gene expression system. The solid lines indicate the profile likelihood. The optimal parameter set is marked with a grey star and the red dashed line marks the 95 % confidence level. The circles are indicating that a refitted parameter hit the boundary of the allowed parameter space which was set between 10^{-6} and 10^{4}.

The profile of the parameter $init_{SEAP}$ is suggesting that the initial SEAP concentration is negligible:

$$init_{SEAP} = 0 \tag{8}$$

This is consistent with the parameter profile of the parameter $k_{basal,mRNA_{nuc}}$ which shows the same behavior. After setting

$$k_{basal,mRNA_{nuc}} = 0 \tag{9}$$

the equations (4-6) read:

$$\frac{d[mRNA_{nuc}](t)}{dt} = k_{prod,mRNA_{nuc}}[LOVVP16] - k_{mRNA,nuc2cyt}\,[mRNA_{nuc}] \tag{4'}$$

$$\frac{d[mRNA_{cyt}](t)}{dt} = k_{nuc2cyt}\,[mRNA_{nuc}] - k_{deg,mRNA_{nuc}}[mRNA_{cyt}] \tag{5'}$$

$$\frac{d[SEAP](t)}{dt} = k_{transl,SEAP}\,\left[mRNA_{cyt}\right] N \tag{6'}$$

A transformation of the mRNA units by a factor α

$$\left([mRNA_{nuc}], \left[mRNA_{cyt}\right], k_{prod,mRNA_{nuc}}, k_{transl,SEAP}\right) \longrightarrow$$

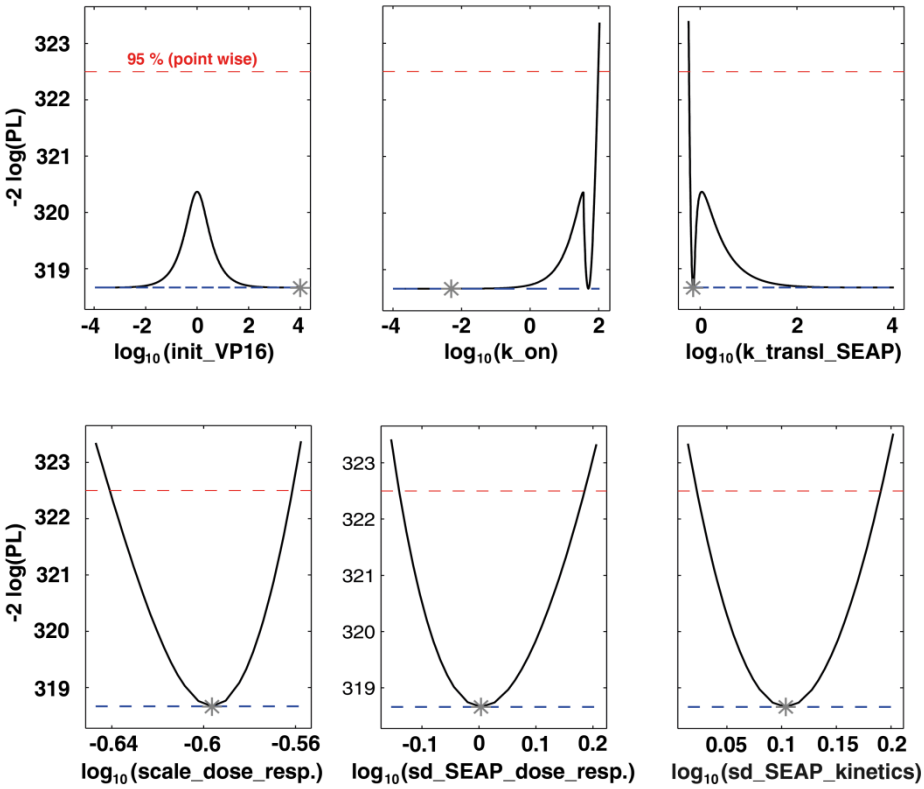
$$(\alpha[mRNA_{nuc}], \alpha\left[mRNA_{cyt}\right], \alpha\, k_{prod,mRNA_{nuc}}, \alpha^{-1}\, k_{transl,SEAP})$$

does not change the output SEAP of the system (4'-6'). Since α is a free parameter we can set $\alpha = 1/k_{prod,mRNA_{nuc}}\ h$. Applying this scaling we obtain:

$$k_{prod,mRNA_{nuc}} = 1\ h^{-1}. \tag{10}$$

The mRNA concentration has now the same unit than LOV, VP16 and LOVVP16 and is measured in relative amounts of the initial LOV concentration.

After reduction of the model by fixing the parameters (8-10) we recalculated the profile likelihood functions for the remaining parameters:



The profile likelihood for each parameter of the reduced rapidly-reversible blue light model. The solid lines indicate the profile likelihood. The optimal parameter set is marked with a grey star and the red dashed line marks the 95 % confidence level.

The profile of $init_{VP16}$ is flat to the right side meaning that the model only can predict that $init_{VP16} \gg LOV_{\mathrm{DNA}}(0)$. This prediction is reasonable, since LOV_{DNA} is the amount of binding sites on the transfected plasmids and VP16 is a protein which is available in much higher concentrations.

Since equation (1)-(3)

$$\frac{d[VP16](t)}{dt} = k_{off}[LOVVP16] - k_{on}\, I\, [LOV_{\mathrm{DNA}}]\, [VP16] - k_{growth}\, [VP16] \quad (1)$$

$$\frac{d[LOV](t)}{dt} = k_{off}[LOVVP16] - k_{on}\, I\, [LOV_{\mathrm{DNA}}]\, [VP16]\ - k_{growth}\, [LOV_{\mathrm{DNA}}] \quad (2)$$

$$\frac{d[LOVVP16](t)}{dt} = -k_{off}[LOVVP16] + k_{on}\, I\, [LOV_{\mathrm{DNA}}]\, [VP16] \quad (3)$$

are symmetric in LOV and VP16 the profile of $init_{VP16}$ is also symmetric around $init_{VP16} = LOV_{\mathrm{DNA}}(0) = 1$ (note that the profile is on a logarithmic scale). This means according to the model it is also possible that $init_{VP16} \ll LOV_{\mathrm{DNA}}(0)$. Being on the left or right side of the profile likelihood of

$init_{VP16}$ only changes the scale of LOVVP16. This can be compensated by a scaling of k_{on} and $k_{transl,SEAP}$ whose units both depend on the scale of LOV_{DNA}.

Supporting References

1. Strickland, D., Lin, Y., Wagner, E., Hope, C. M., Zayner, J., Antoniou, C., Sosnick, T. R., Weiss, E. L., and Glotzer, M. (2012) TULIPs: tunable, light-controlled interacting protein tags for cell biology, *Nat. Methods 9*, 379-384.
2. Muller, K., Engesser, R., Metzger, S., Schulz, S., Kampf, M. M., Busacker, M., Steinberg, T., Tomakidi, P., Ehrbar, M., Nagy, F., Timmer, J., Zubriggen, M. D., and Weber, W. (2013) A red/far-red light-responsive bi-stable toggle switch to control gene expression in mammalian cells, *Nucleic Acids Res. 41*, e77.
3. Muller, K., Engesser, R., Schulz, S., Steinberg, T., Tomakidi, P., Weber, C. C., Ulm, R., Timmer, J., Zurbriggen, M. D., and Weber, W. (2013) Multi-chromatic control of mammalian gene expression and signaling, *Nucleic Acids Res. 41*, e124.
4. Muller, K., Siegel, D., Rodriguez Jahnke, F., Gerrer, K., Wend, S., Decker, E. L., Reski, R., Weber, W., and Zurbriggen, M. D. (2014) A red light-controlled synthetic gene expression switch for plant systems, *Mol. Biosyst. 10*, 1679-1688.
5. Fussenegger, M., Morris, R. P., Fux, C., Rimann, M., von Stockar, B., Thompson, C. J., and Bailey, J. E. (2000) Streptogramin-based gene regulation systems for mammalian cells, *Nat. Biotechnol. 18*, 1203-1208.
6. Fussenegger, M., Moser, S., Mazur, X., and Bailey, J. E. (1997) Autoregulated multicistronic expression vectors provide one-step cloning of regulated product gene expression in mammalian cells, *Biotechnol. Prog. 13*, 733-740.
7. Wang, X., Chen, X., and Yang, Y. (2012) Spatiotemporal control of gene expression by a light-switchable transgene system, *Nat. Methods 9*, 266-269.
8. Coleman, T. F., and Li, Y. Y. (1996) An interior trust region approach for nonlinear minimization subject to bounds, *Siam J Optimiz 6*, 418-445.
9. Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmuller, U., and Timmer, J. (2009) Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood, *Bioinformatics 25*, 1923-1929.
10. Raue, A., Schilling, M., Bachmann, J., Matteson, A., Schelke, M., Kaschek, D., Hug, S., Kreutz, C., Harms, B. D., Theis, F. J., Klingmuller, U., and Timmer, J. (2013) Lessons learned from quantitative dynamical modeling in systems biology, *PloS one 8*, e74335.