**Supporting Information**

**Title:** ProTeOn and ProTeOff, new protein devices that inducibly activate bacterial gene expression.

**Authors:** Katherine Volzing, Konstantinos Biliouris, Yiannis N. Kaznessis*

**Author Affiliation:** Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN, USA

## Experimental proTeOn and proTeOff characterization

### *PROTEON and PROTEOFF expression*

Under the control of a LacI repressible T7 promoter, PROTEON and PROTEOFF expression is IPTG inducible. Cells cultured with 0, 0.1, 0.5 and 1 mM IPTG showed increasing cytosolic protein levels with increasing IPTG, as shown in Supporting Figure 2a. Even at high levels, PROTEON is efficiently translated, stable, and soluble in *E. coli*. As PROTEOFF differs from PROTEON by a single amino acid, the same is assumed for PROTEOFF protein.

### *PROTEON and PROTEOFF activation*

proTeOn activity is aTc inducible in *E. coli*. In the absence of aTc, PROTEON is inactive and GFP is expressed at a low basal level. Upon the addition of aTc, PROTEON is activated through a conformational change; it binds the synthetic promoter and upregulates GFP expression. With 10 and 200 ng/ml aTc, GFP is upregulated 10 and 30 - fold respectively by Western blot. Meanwhile, the total cytosolic transactivator levels modestly decrease in the presence of aTc, as shown in Supporting Figure 2b. This decrease can be attributed to the increased demand for the cell's transcription and translation machinery with active PROTEON upregulating GFP.

***proTeOn and proTeOff phenotype analysis***

proTeOn and proTeOff induction and phenotype analysis by flow cytometry was investigated over 20 hours post-induction. Induction experiments were repeated and the behaviors compared across replicates.

The general behavior observed across replicates for proTeOn is in good agreement with the specific results presented in the Results and Discussion. Overall, proTeOn upregulates target genes by one hour post-aTc treatment and achieves 15-fold upregulation through long times. Steady state expression levels are reached by 5 hours and 10 hours post-treatment with low and high (10 and 200 ng/ml) aTc respectively. Expression is generally maintained at 10 and 15-fold above uninduced controls through long times with low and high aTc levels.

When expressed in the presence of aTc, PROTEON achieves target gene upregulation within 2 hours. Overall, steady state is achieved by 5 hours after the transactivator's initial expression, with target protein levels 10-fold over the uninduced controls. proTeOn's behavior is presented in Supplementary Figure 3a.

proTeOff upregulates target gene expression in the absence of aTc. With high (200 ng/ml) aTc, expression is reduced to half that of the untreated samples by 2 hours. In general, with low (10 ng/ml) aTc, this reduction may not be realized until 20 hours post-treatment. Low, steady state expression is achieved by 5 hours post-treatment with high aTc and maintained through long times. Minimum target gene expression levels of one-half and one-fifth that of the untreated samples are observed with low and high aTc respectively.

When expressed in the absence of aTc, transcription upregulation by proTeOff is observed within one hour after PROTEOFF expression is induced. In low and high aTc, reduced proTeOff activity is observed across all times. In general, steady state activity is achieved by 5 hour after PROTEOFF expression is induced. Overall, proTeOff activity is one-fourth that of untreated samples for both 10 and 200 ng/ml aTc through long times. proTeOff's behavior is shown in Supplementary Figure 3b.

## proTeOn and proTeOff stochastic models

### *Model delineation*

Small biological systems, either natural or synthetic, are subject to stochasticity (*1-4*). The behavior of such systems has been accurately described using detailed stochastic modeling (*5-13*). Each step of biology's central dogma, including transcription, translation, degradation, dimerization, repression and induction are represented by biochemical reactions. The reaction network that portrays the behavior of both of our synthetic systems is outlined in Supporting Figure 4a.

The first reaction captures the effective rate of aTc penetration into the cell. Reactions 2-9 model the binding of 2 aTc molecules to one PROTEON or PROTEOFF (PROTET) molecule (*14*). In particular, reactions 2-5 capture the binding and unbinding of aTc to PROTET when the latter is free, whereas reactions 6-9 correspond to the same interactions when PROTET is bound to its operator site, *tetO*. Reactions 10-13 represent the binding of PROTET to *tetO* (*14*). PROTET can bind to *tetO* either when free (reactions 10, 11) or bound to aTc (reactions 12, 13). Reactions

14-26 describe *gfp* transcription (*15, 16*). *gfp* transcription occurs in three different ways since RNApol can transcribe *gfp* when *tetO* is: a) free (reactions 14-18) or b) occupied by PROTET:aTc2 (reactions 19-22) or c) occupied by PROTET (reactions 23-26). Three different sets of reactions are therefore used to describe transcription. In these 3 sets, the same approach is used. The first two reactions (14-15, 19-20, 23-24) capture the binding and unbinding of RNApol to pro. The binding of RNApol to pro is followed by the formation of an open complex between RNApol and DNA (reactions 16, 21, 25). Subsequently, RNApol releases the pro (reactions 17, 22, 26), allowing other RNApol to bind, and proceeds transcribing the entire *gfp* (reaction 18). The product of transcription, mRNA(*gfp*), is thereafter translated into GFP protein. This process is coded in reactions 27-29. The time needed for a nucleotide to be transcribed or for a codon to be translated is assumed to be exponential distributed. Consequently, the time required for the entire process of transcription and translation is considered to be gamma distributed (*17*). Finally, reaction 30 captures the GFP maturation process whereas reactions 31 and 32 represent the effective rate at which GFP and mRNA(gfp) are degraded.

The initial conditions regarding the transcription and translation machinery define 100 RNApol and 100 ribosomes at t = 0. In addition, the number of synthetic plasmids in each *E. coli* is set equal to 20, inferring that there are 20 *tetO* and 20 pro sites. The effective amount of PROTET that exists in each cell was considered constant (20 molecules) throughout the simulations since overnight IPTG treatment led to a constant PROTET concentration. The definition of each species as abbreviated in the model is given in Supporting Figure 4b.

***Model parameters***

The conditions of this system (temperature, pH) are assumed to be constant throughout the simulations guaranteeing that the kinetic parameters used do not vary over time. It is postulated that the cell is a homogeneous, yet well stirred reactor. The volume of each *E. coli* cell is equal to $10^{-15}$ L. The volume of the cells increases exponentially over time and is halved during cell division. The cell division time was adjusted relative to experimentally observed behaviors with changes in intracellular aTc and GFP concentrations. Cell division time increases with increasing aTc as it retards cellular processes in higher concentrations (*18*). In addition, cell division time increases with GFP concentration since this increase in protein production taxes the cellular machinery. The cell division rates that were used in our simulations are shown in Supporting Figure 4c. proTeOn and proTeOff are synthetic systems and many of the associated kinetic parameters do not exist in the literature. Thus, most of the parameters have values that are close to parameters of similar, either synthetic or naturally occurring, systems and have been adjusted to fit the experimental results.

The penetration rate of aTc into the cell is considered equal to the penetration rate of tetracycline (*19*). The binding of aTc to PROTET is assumed equal to the binding of aTc to rTetR as there is no available data for PROTET (*14*). The binding of PROTET, either when free or bound to aTc, to the operator site, is considered almost equal to the binding of rTetR to *tetO* (*14*). The affinity of RNApol for the pro, when *tetO* is free, is presumed similar to the affinity of RNApol for the *tet* promoter (*15*). Further, the affinity of RNApol for the pro when it is occupied by PROTET or PROTET:aTc2 was fit, considering RNApol recruitment by LuxRΔN to the

promoter, so that the results match the experimental phenotypes. The rates of transcription and translation do vary significantly throughout the cell population. In our simulations, we hypothesize average transcription and translation rates equal to 30 nucleotides and 100 codons per second. The GFP maturation rate was adjusted such that the simulation results were in agreement with experimental observations. Finally, the GFP degradation rate was adjusted for a 10 min protein half life, whereas mRNA degradation was adjusted to give 20 proteins per mRNA molecule.

**References**


1. Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. (2002) Stochastic gene expression in a single cell, *Science 297*, 1183-1186.
2. McAdams, H. H., and Arkin, A. (1997) Stochastic mechanisms in gene expression, *Proc Natl Acad Sci U S A 94*, 814-819.
3. McAdams, H. H., and Arkin, A. (1999) It's a noisy business! Genetic regulation at the nanomolar scale, *Trends Genet 15*, 65-69.
4. Kaern, M., Elston, T. C., Blake, W. J., and Collins, J. J. (2005) Stochasticity in gene expression: from theories to phenotypes, *Nat Rev Genet 6*, 451-464.
5. Biliouris, K., Daoutidis, P., and Kaznessis, Y. N. Stochastic simulations of the tetracycline operon, *BMC Syst Biol 5*, 9.
6. Kaznessis, Y. N. (2009) Computational methods in synthetic biology, *Biotechnol J 4*, 1392-1405.
7. Ramalingam, K. I., Tomshine, J. R., Maynard, J. A., and Kaznessis, Y. N. (2009) Forward engineering of synthetic bio-logical AND gates, *Biochemical Engineering Journal 47*, 38-47.
8. Sotiropoulos, V., and Kaznessis, Y. N. (2007) Synthetic tetracycline-inducible regulatory networks: computer-aided design of dynamic phenotypes, *BMC Syst Biol 1*, 7.
9. Tuttle, L. M., Salis, H., Tomshine, J., and Kaznessis, Y. N. (2005) Model-driven designs of an oscillating gene network, *Biophys J 89*, 3873-3883.
10. Kaznessis, Y. N. (2007) Models for synthetic biology, *BMC Syst Biol 1*, 47.
11. Kaznessis, Y. N. (2009) Multiscale models for synthetic biology, *Conf Proc IEEE Eng Med Biol Soc 2009*, 6408-6411.
12. Hasty, J., Pradines, J., Dolnik, M., and Collins, J. J. (2000) Noise-based switches and amplifiers for gene expression, *Proc Natl Acad Sci U S A 97*, 2075-2080.
13. Stamatakis, M., and Mantzaris, N. V. (2009) Comparison of deterministic and stochastic models of the lac operon genetic network, *Biophys J 96*, 887-906.
14. Kamionka, A., Bogdanska-Urbaniak, J., Scholz, O., and Hillen, W. (2004) Two mutations in the tetracycline repressor change the inducer anhydrotetracycline to a corepressor, *Nucleic Acids Res 32*, 842-847.

15. Bertrand-Burggraf, E., Lefevre, J. F., and Daune, M. (1984) A new experimental approach for studying the association between RNA polymerase and the tet promoter of pBR322, *Nucleic Acids Res 12*, 1697-1706.
16. Vogel, U., and Jensen, K. F. (1994) The RNA chain elongation rate in Escherichia coli depends on the growth rate, *J Bacteriol 176*, 2807-2813.
17. Gibson, M. A., and Bruck, J. (2000) Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels, *The Journal of Physical Chemistry A 104*, 1876-1889.
18. D.E. Golan, A. H. T., E.J. Armstrong, A.W. Armstrong (2005) *Principles of pharmacology: the pathophysiologic basis of drug therapy* Lippincott Williams and Wilkins, Philadelphia.
19. Sigler, A., Schubert, P., Hillen, W., and Niederweis, M. (2000) Permeation of tetracyclines through membranes of liposomes and Escherichia coli, *Eur J Biochem 267*, 527-534.

**Supporting Figures and Legends**

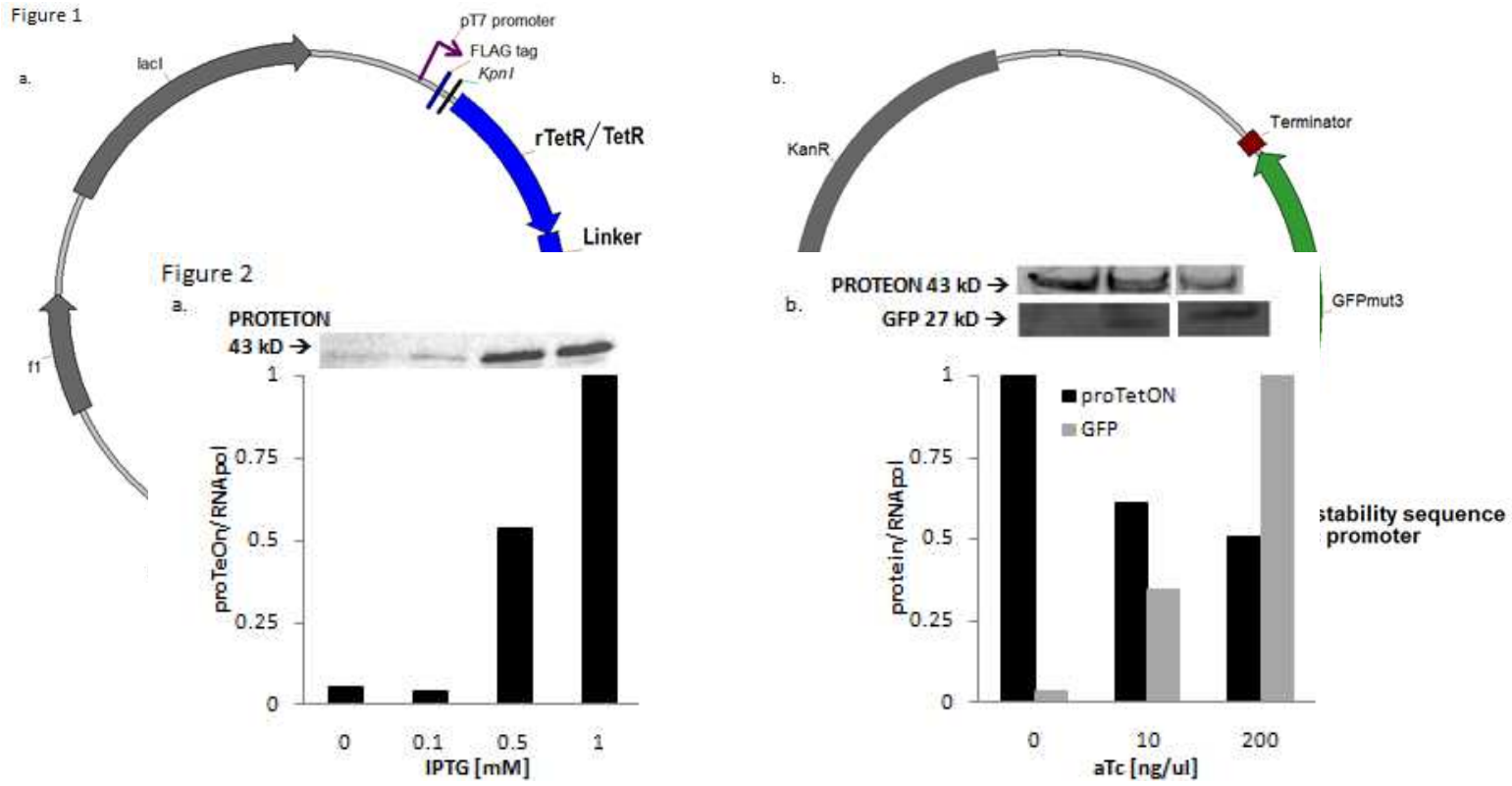
Supporting Figure 1. proTeOn and proTeOff system vectors

a. PROTEON and PROTEOFF. Both synthetic proteins, rTetR-LuxRΔN and TetR-LuxRΔN, are under the control of a LacI repressible T7 promoter on low-copy plasmid, pT7-FLAG1 (Sigma).

b. proTeOn and proTeOff synthetic promoter. The synthetic promoters and GFPmut3 gene were synthesized by GENEART on, pMK, a pUC19 derived expression vector that's compatible in *E. coli*, high copy, and kanamycin resistant.

Supporting Figure 2. System induciblity

a. PROTEON expression. Relative to uninduced conditions, PROTEON monomer (43 kD) levels are 10-fold higher upon induction with 0.5 mM IPTG and 20-fold higher with 1.0 mM IPTG. At 0.1 mM IPTG expression is not induced.




b. PROTEON activation. In the absence of aTc, PROTEON is inactive and GFP (27 kD) is expressed at a low basal level. Upon the addition of 10 and 200 ng/ml aTc, GFP expression increases 10 and 30-fold respectively. Total cytosolic transactivator levels modestly decrease in the presence of aTc.

<u>Supporting Figure 3. proTeOn and proTeOff phenotype analysis</u>

Mean GFP expression was analyzed by flow cytometry 1, 2, 5, 10 and 20 hours post-treatment for both experimental set-ups as described. Induction experiments were repeated for all

inducer concentrations and time points, and the reported trends were observed across the replicates

a. proTeOn. Experiment A: With low and high (10 and 200 ng/ml) aTc concentrations, proTeOn upregulates GFP one hour post-induction with aTc. Steady state expression is achieved by 5 hours and 10 hours with low and high aTc respectively. These steady states are maintained through 20 hours. Maximum overexpression is 10 and 15-fold above the uninduced controls with low and high aTc concentrations respectively. Experiment B: Upon expression in the presence of low and high aTc concentrations, PROTEON significantly upregulates GFP by 2 hours after PROTEON expression is induced. Overall, steady state is reached by 5 hours and maintained through 20 hours. Maximum upregulation is 10-fold above uninduced controls with both aTc concentrations.
b. proTeOff. Experiment A: proTeOff upregulates GFP expression in the absence of aTc. With high (200 ng/ml) aTc, expression is reduced to half that of the untreated samples by 2 hours. In general, with low (10 ng/ml) aTc, this reduced expression may not be realized until 20 hours post-treatment. Low, steady state expression is achieved by 5 hours with high aTc and maintained through 20 hours. Overall minimum expression is one-half and one-fifth that of the untreated sample with low and high aTc concentrations respectively. Experiment B: In the absence of aTc, proTeOff activity is observed by one hour after PROTEOFF expression is induced. In both aTc concentrations, reduced proTeOff activity is observed across all times. Generally, steady state activity is achieved by 5 hours after PROTEOFF expression is induced. At steady state, proTeOff activity is one-fourth of untreated samples for both aTc concentrations overall and maintained at this low level through 20 hours.

Supporting Figure 4. proTeOn and proTeOff stochastic simulations

a. proTeOn and proTeOff reaction network. The biochemical reactions and their kinetic constants are provided for all the processes considered by our model. The kinetic constants that appear in red differentiate the behavior of the two systems. Values that were fit to match the experimental results have a reference denoted with *. Values derived from the literature, "i," and then fit to match the experimental observations, have a reference denoted with " i*."

b. Definition of each reaction network species.

c. Cell division times for proTeOn and proTeOff when 0, 10 and 200 ng/ml aTc are administered.

Figure 3

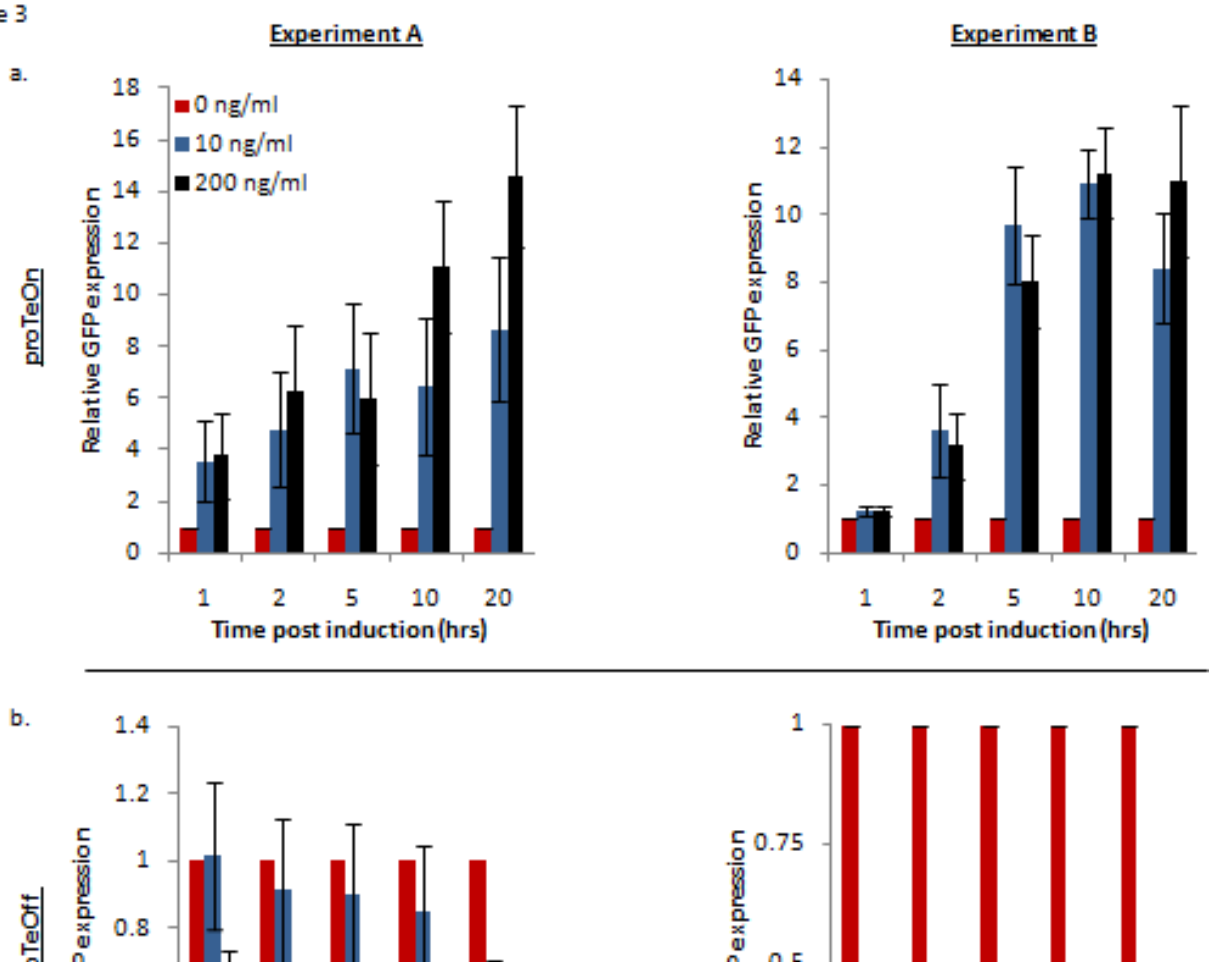


Figure 4

**a.**

| # | Biochemical Reaction | Kinetic constant (proTeOn/proTeOff) | Reference |
|---|---|---|---|
| | **Anydrotetracycline administration** | | |
| 1. | aTcEx ⟶ aTc | 3.3E-4 | 35 |
| | **Interactions between PROTET and aTc** | | |
| 2. | PROTET + aTc ⟶ PROTET:aTc | 2.76E+8 | 31 |
| 3. | PROTET:aTc ⟶ PROTET + aTc | 1 | 31 |
| 4. | PROTET:aTc + aTc ⟶ PROTET:aTc2 | 1.38E+8 | 31 |
| 5. | PROTET:aTc2 ⟶ PROTET:aTc + aTc | 2 | |
| 6. | PROTET:*tetO* + aTc ⟶ PROTET:*tetO*:aTc | 2.76E+8 | 31 |
| 7. | PROTET:*tetO*:aTc ⟶ PROTET:*tetO* + aTc | 1 | 31 |
| 8. | PROTET:*tetO*:aTc + Tc ⟶ PROTET:*tetO*:aTc2 | 1.38E+8 | 31 |
| 9. | PROTET:*tetO*:aTc2 ⟶ PROTET:*tetO*:aTc + aTc | 2 | 31 |
| | **Interactions between PROTET and *tetO*** | | |
| 10. | PROTET + *tetO* ⟶ PROTET:*tetO* | 4E+3 / 4E+8 | 31* |
| 11. | PROTET:*tetO* ⟶ PROTET + *tetO* | 0.1 | 31* |
| 12. | PROTET:aTc2 + *tetO* ⟶ PROTET:*tetO*:aTc2 | 4E+8 / 4E+3 | 31* |
| 13. | PROTET:*tetO*:aTc2 ⟶ PROTET:aTc2 + *tetO* | 0.1 / 100 | 31* |
| | **Transcription of *gfp* gene (when *tetO* is free)** | | |
| 14. | RNApol + pro + *tetO* ⟶ RNApol:pro:*tetO* | 270 | 32* |
| 15. | RNApol:pro:*tetO* ⟶ RNApol + pro + *tetO* | 1E-6 | 32* |
| 16. | RNApol:pro:*tetO* ⟶ RNApol*:pro:*tetO* | 0.013 | 32 |
| 17. | RNApol*:pro:*tetO* ⟶ RNApol*:DNA(*gfp*) + pro + *tetO* | 30 | 33 |
| 18. | RNApol*:DNA(*gfp*) ⟶ RNApol + mRNA(*gfp*) | 30, 714 | 33 |
| | **Transcription of *gfp* gene (when *tetO* is occupied by PROTET:aTc2)** | | |
| 19. | RNApol + pro + PROTET:aTc2:*tetO* ⟶ RNApol:pro:PROTET:aTc2:*tetO* | 6E+3 / 3.9E+3 | * |
| 20. | RNApol:pro:PROTET:aTc2:*tetO* ⟶ RNApol + pro + PROTET:aTc2:*tetO* | 1E-6 | * |
| 21. | RNApol:pro:PROTET:aTc2:*tetO* ⟶ RNApol*:pro:PROTET:aTc2:*tetO* | 0.013 | 32 |
| 22. | RNApol*:pro:PROTET:aTc2:*tetO* ⟶ RNApol*:DNA(*gfp*) + PROTET:aTc2:*tetO* + pro | 30 | 33 |
| | **Transcription of *gfp* gene (when *tetO* is occupied by PROTET)** | | |
| 23. | RNApol + pro + PROTET:*tetO* ⟶ RNAp:pro:*tetO*:PROTET | 6E+3 / 3.9E+3 | * |
| 24. | RNApol:pro:*tetO*:PROTET ⟶ RNAp + pro + PROTET:*tetO* | 1E-6 | * |
| 25. | RNApol:pro:*tetO*:PROTET ⟶ RNAp*:pro:*tetO*:PROTET | 0.013 | 32 |
| 26. | RNApol*:pro:*tetO*:PROTET ⟶ RNAp*:DNA(*gfp*) + pro + PROTET:*tetO* | 30 | 33 |
| | **Translation of *gfp* mRNA** | | |
| 27. | Rib + mRNA(*gfp*) ⟶ Rib:mRNA(*gfp*) | 1E+5 | 17 |
| 28. | Rib:mRNA(*gfp*) ⟶ Rib*:mRNA(*gfp*) + mRNA(*gfp*) | 100 | 34 |
| 29. | Rib*:mRNA(*gfp*) ⟶ Rib + GFP | 100, 238 | 34 |
| | **GFP protein maturation** | | |
| 30. | GFP ⟶ GFP* | 1E-3 | * |
| | **Degradation** | | |
| 31. | GFP ⟶ ø | 1.2E-3 | 17 |
| 32. | mRNA(*gfp*) ⟶ ø | 2E-3 | 17 |

**b.**

| Species | Definition |
|---|---|
| aTcEx | Extracellular anhydrotetracycline |
| aTc | Intracellular anhydrotetracycline |
| PROTET | PROTEON or PROTEOFF dimer |
| PROTET:aTc | PROTET bound to one aTc molecule |
| PROTET:aTc2 | PROTET bound to two aTc molecules |
| PROTET:*tetO* | PROTET bound to *tetO* |
| PROTET:*tetO*:aTc | PROTET bound to *tetO* and one aTc molecule |
| PROTET:*tetO*:aTc2 | PROTET bound to *tetO* and two aTc molecules |
| *tetO* | TetR and rTetr operator sites |
| RNApol | RNA polymerase |
| pro | Promoter site |
| RNApol:pro:*tetO* | Open complex of RNApol and pro when *tetO* is free |
| RNApol*:pro:*tetO* | Close complex of RNApol and pro when *tetO* is free |
| RNApol*:DNA(*gfp*) | RNApol bound to *gfp* |
| RNApol:pro:PROTET:aTc2:*tetO* | Open complex of RNApol and pro when *tetO* is occupied by PROTET:aTc2 |
| RNApol*:pro:PROTET:aTc2:*tetO* | Close complex of RNApol and pro when *tetO* is occupied by PROTET:aTc2 |
| RNApol:pro:*tetO*:PROTET | Open complex of RNApol and pro when *tetO* is occupied by PROTET |
| RNApol*:pro:*tetO*:PROTET | Close complex of RNApol and pro when *tetO* is occupied by PROTET |
| mRNA(*gfp*) | *gfp* message/mRNA |
| Rib | Ribosome |
| Rib:mRNA(*gfp*) | Open complex of Rib and mRNA(*gfp*) |
| Rib*:mRNA(*gfp*) | Close complex of Rib and mRNA(*gfp*) |
| GFP | Unfolded GFP protein |
| GFP* | Folded GFP protein |
| *gfp* | gfp gene |

**c.**

| aTc Concentration [ng/ml] | Cell division time (±10 min) | |
|---|---|---|
| | PROTEON | PROTEOFF |
| 0 | 32 | 57 |
| 10 | 59 | 35.5 |
| 200 | 72.5 | 39 |