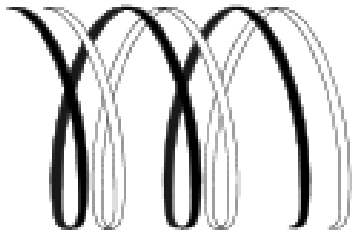
Biophysical modeling of bacterial immune system regulation

Marko Djordjevic

Faculty of Biology, University of Belgrade



Bacterial immune systems

Restriction-modification

restriction enzyme + methyltransferase

Rudimental immune system, restriction enzyme cuts all DNA sequences that are not protected by the antidote (methylase)

CRISPR-Cas

A more sophisticated (adaptive) immune system, based on expression of small RNAs.

Recently theoretically predicted and experimentally confirmed.

General goals

- **Understand how CRISPR-Cas expression is regulated**
- **Use R-M systems to facilitate understanding of CRISPR-Cas**
- **Understand general principles of bacterial immune system functioning**

Overview

Restriction-modification (R-M) systems

- explaining *in-vitro* measurements
- modeling *in-vivo* data
- population dynamics effects
- R-M system design principles

CRISPR-Cas systems

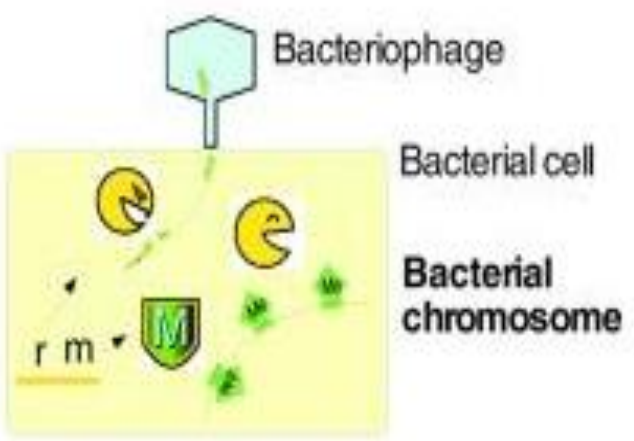
- CRISPR transcript processing
- CRISPR-Cas transcription regulation
- Common design principles with R-M systems
- Contribution of CRISPR-Cas regulatory features to its dynamics

Conclusion

- Common properties in dynamics of CRISPR-Cas and restriction-modification induction/establishment

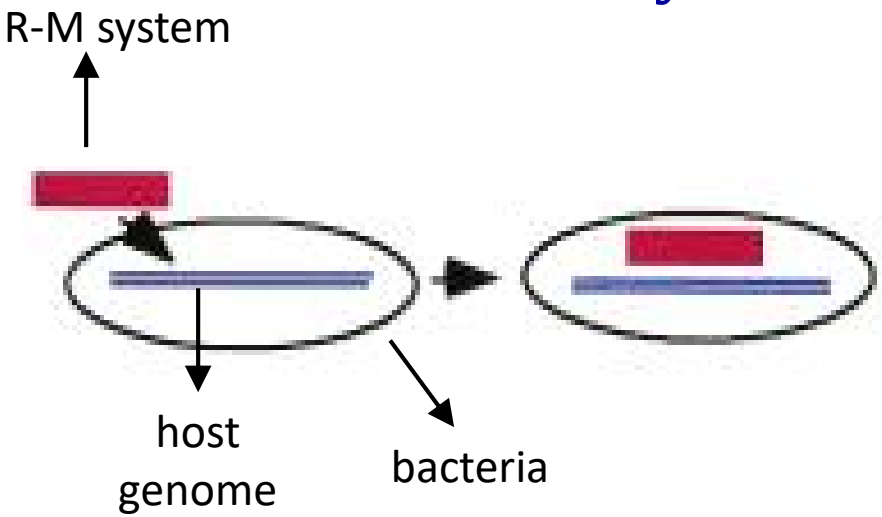
Modeling R-M system establishment

R-M switch is a rudimental “immunological system”



Host DNA is methylated and avoids restriction, but invading foreign DNA (e.g. bacteriophage) is not methylated and is therefore destroyed.

R-M systems are often mobile

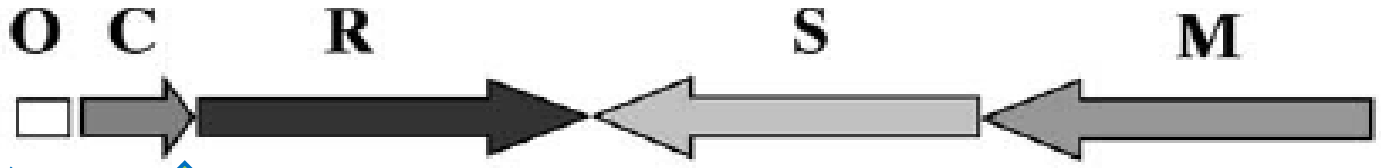


R-M genes and can spread from one bacterial host to another, thus propagating through bacterial populations.

Bacterial Restriction-Modification (R-M) system

Endonuclease (R) recognizes and cuts specific DNA sequences

Methylase (M) methylates the same DNA sequences that are cut by the endonuclease, and consequently protects them from cleaving.



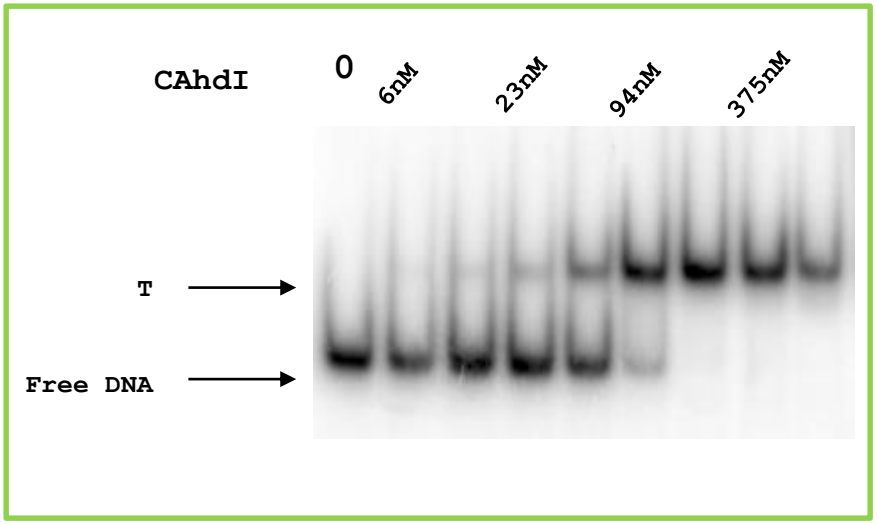
Operator sequence

Control protein (C) coordinates expression of R and M genes by binding to the operator sequence upstream of CR genes and regulating their expression.

Regulation by the control protein

The control protein exhibits very large cooperativity:

- **Only dimmer can bind to DNA**
- **Only tetramer is bound to DNA in the absence of RNA polymerase**

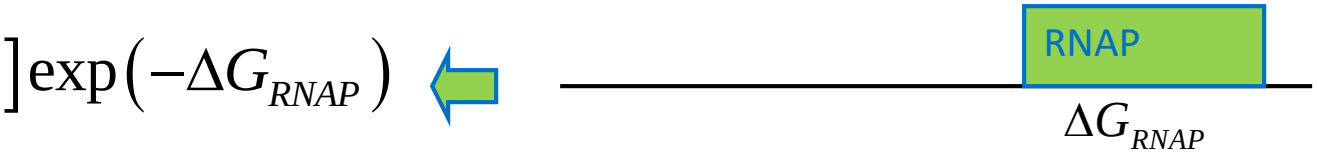


Over the entire range of control protein concentrations two dimmers are simultaniously bound

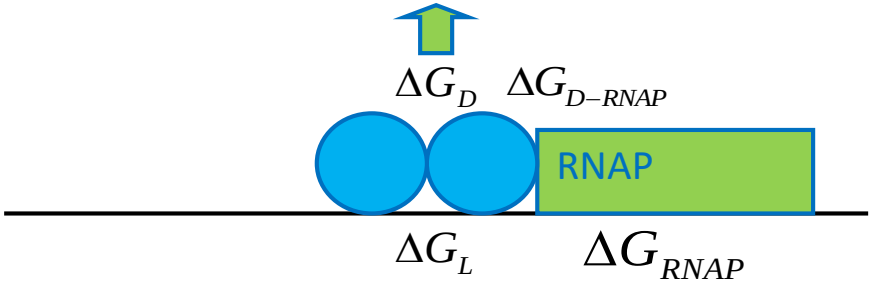
Large cooperativity in binding

Quantitative model

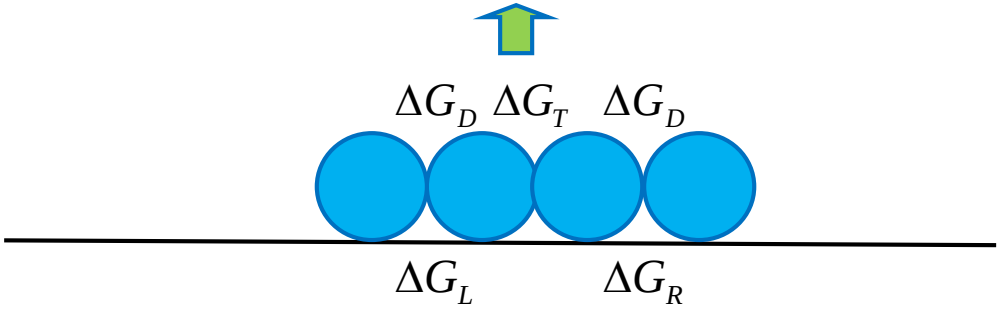
$$Z_{RNAP} = K[RNAP]\exp(-\Delta G_{RNAP})$$



$$Z_{D-RNAP} = K^2[M]^2[RNAP]\exp(-\Delta G_D - \Delta G_L - \Delta G_{D-RNAP} - \Delta G_{RNAP})$$



$$Z_T = K^4[M]^4[RNAP]\exp(-\Delta G_L - \Delta G_R - \Delta G_T - 2\Delta G_D)$$



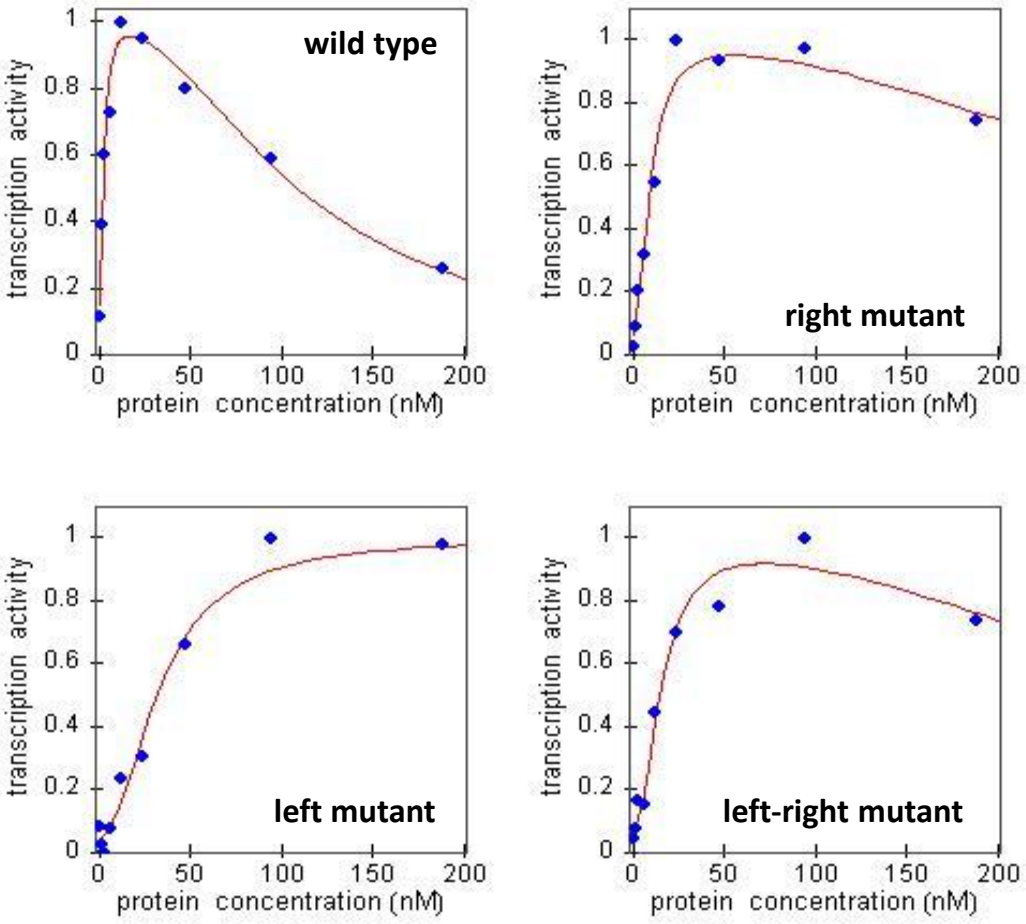
$$\varphi = \frac{Z_{RNAP} + Z_{D-RNAP}}{1 + Z_{RNAP} + Z_{D-RNAP} + Z_T} = \frac{a + b[M]^2}{1 + a + b[M]^2 + c[M]^4}$$

transcription activity

a, b and c are directly related with the biophysical properties of the switch.

Comparison with the experiment

E. Bogdanova, M.D. *et al.*, NAR 36,1429 (2008)



The model agrees well with the measured transcription activities in the case of both wild-type and mutant promoter sequences.

Modeling restriction enzyme dynamics

$$\frac{dc}{dt} = \gamma\left(C_C\right) - \lambda c$$

generation of RC transcripts

Transcription activity versus C protein concentration

$$\frac{dC_C}{dt} = \alpha_C c - \beta C_C$$

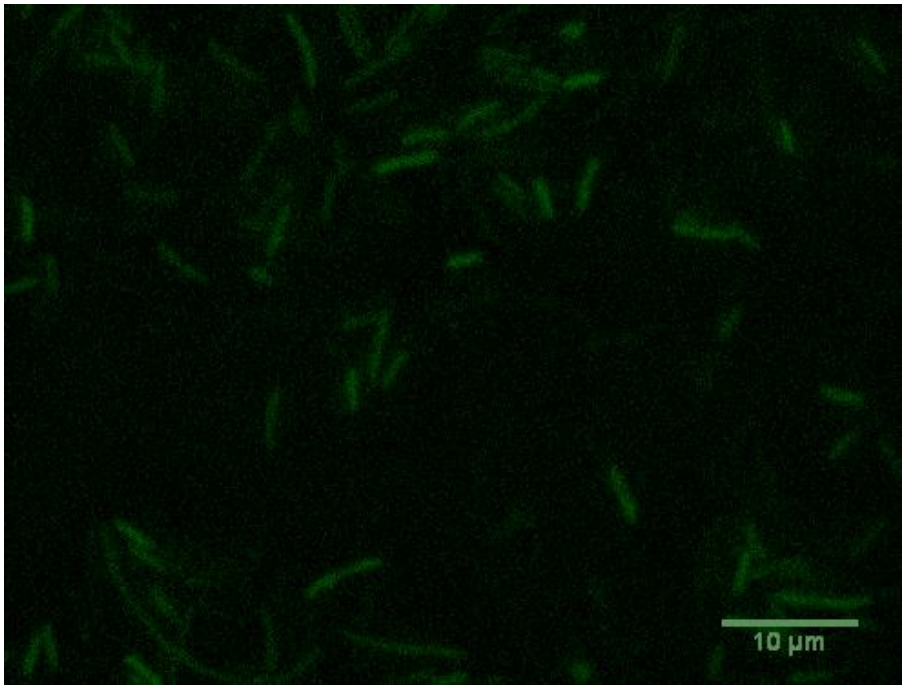
synthesis of C protein

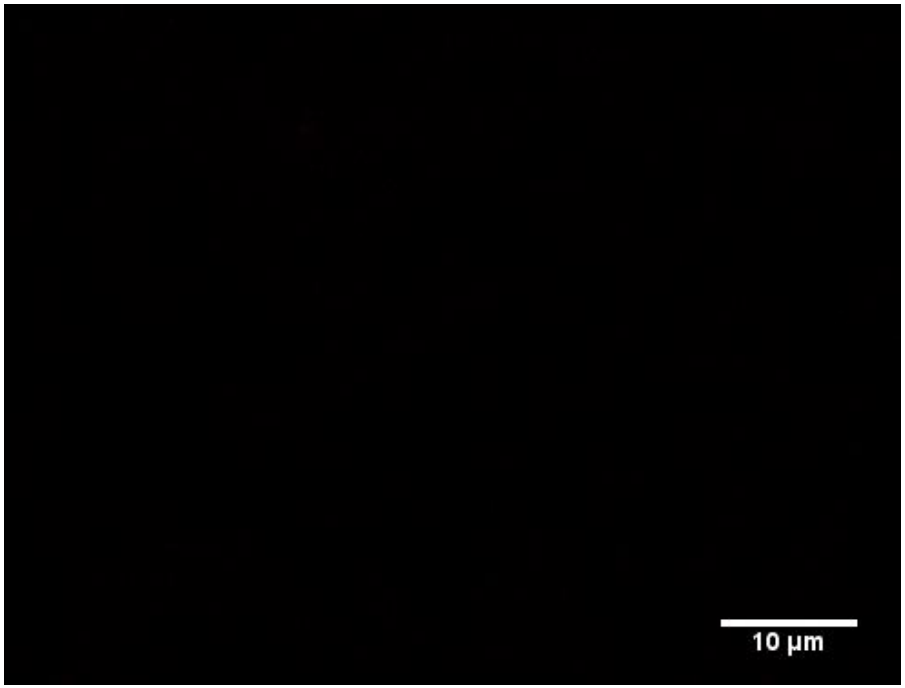
$$\frac{dC_R}{dt} = \alpha_R c - \beta C_R$$

synthesis of R protein

In-vivo R-M expression

N. Morozova, A.R., M.D. et al., *Nucleic Acids Res* 44, 790 (2016)





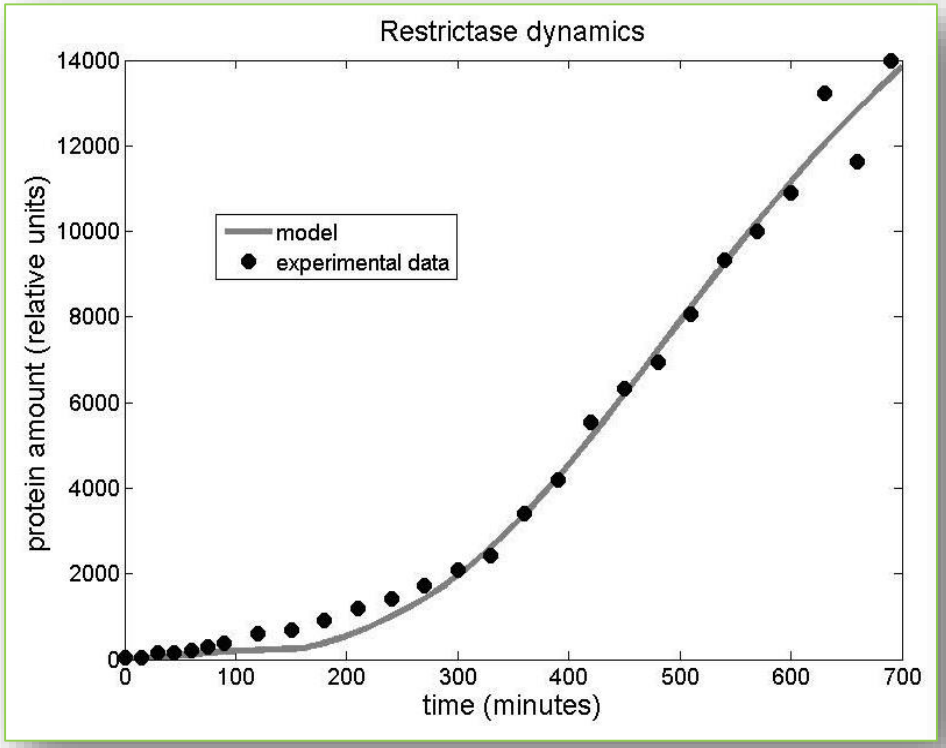
methyltransferase

restriction endonuclease

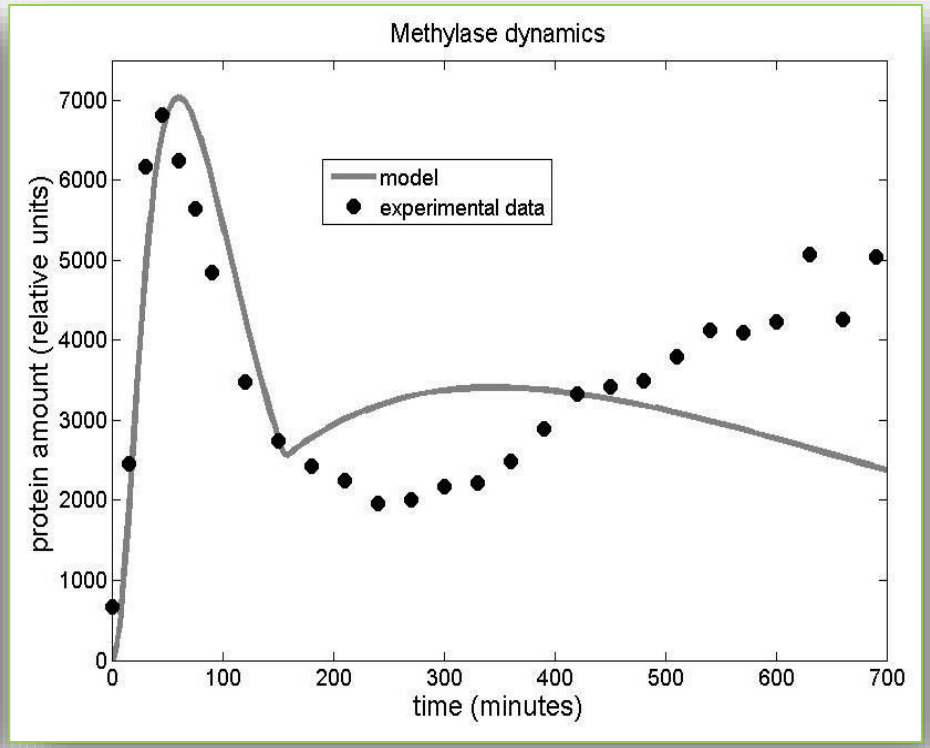
Culture developing from a single transformed cell

R and M expression dynamics

restriction endonuclease



methyltransferase



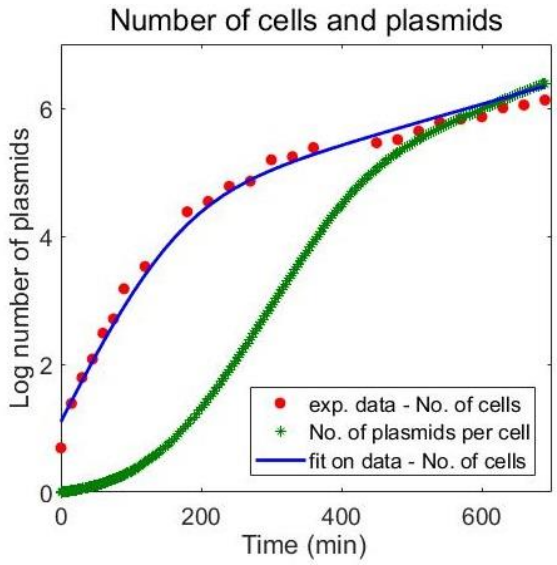
Delayed R expression

Early M accumulation

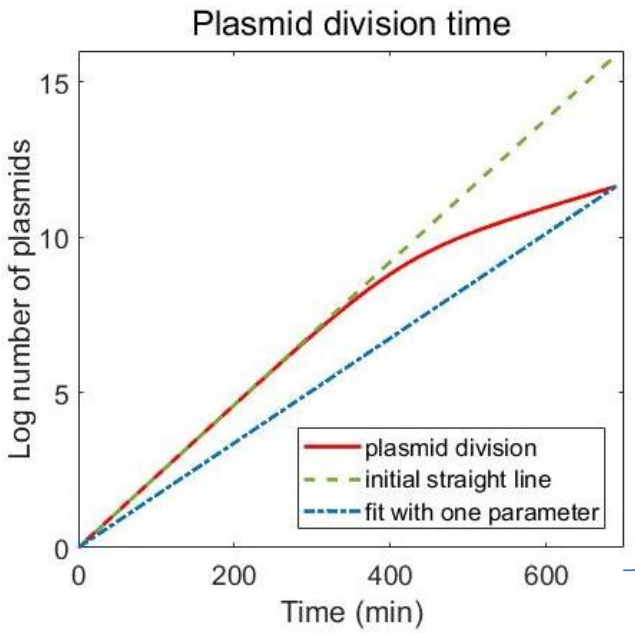
N. Morozova, A.R., M.D. et al., *Nucleic Acids Res* , 44,790 (2016)

Introducing population dynamics

- Cells change their division rate with time:



Plasmid numbers in cell increase for ~ two orders of magnitude



Assume that plasmids also change in time analogously to cells

Introduce population dynamics in the model

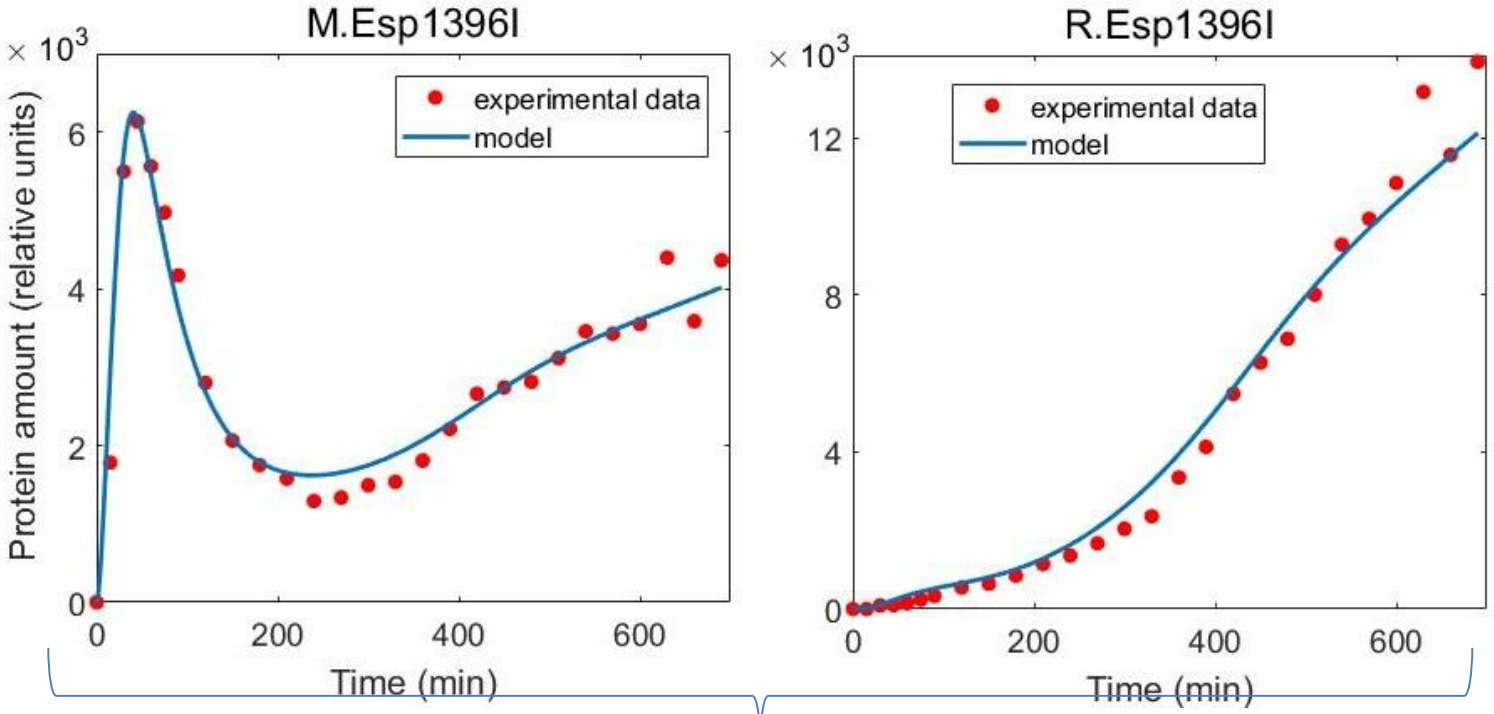
$$\frac{dc(t)}{dt} = \phi(t)\gamma\left(C_C(t)\right) - \left(\lambda_c + \lambda_d\right)c(t)$$

$\phi(t)$ Plasmid division rate

λ_d Effective decay due to cell division

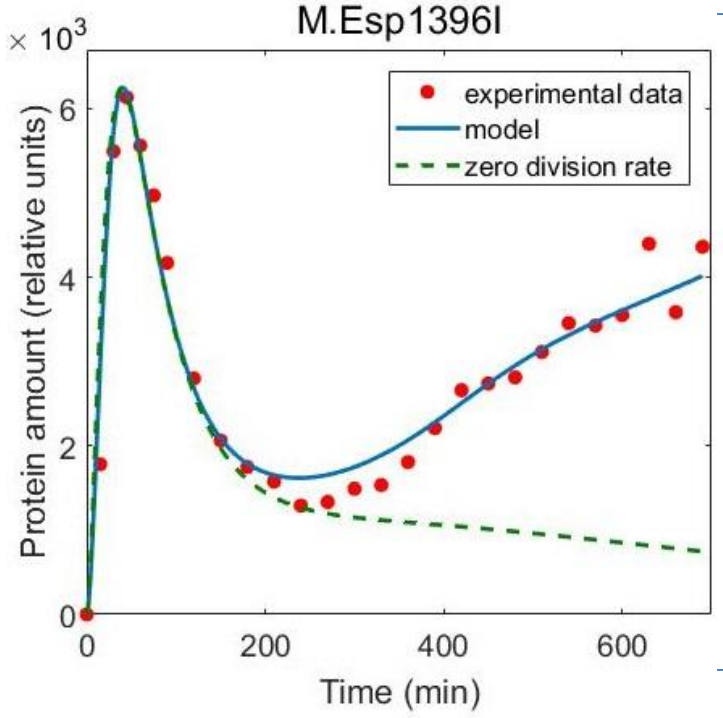
Technically, population dynamics can significantly increase dimensionality of the parameter inference, effectively coupling regulatory independent species (in our case 18 dim. problem).

Developed an iterative, “mean-field like” procedure where we empirically estimate dynamics of one species, than estimate population parameters from the second species, than going back to the first species…



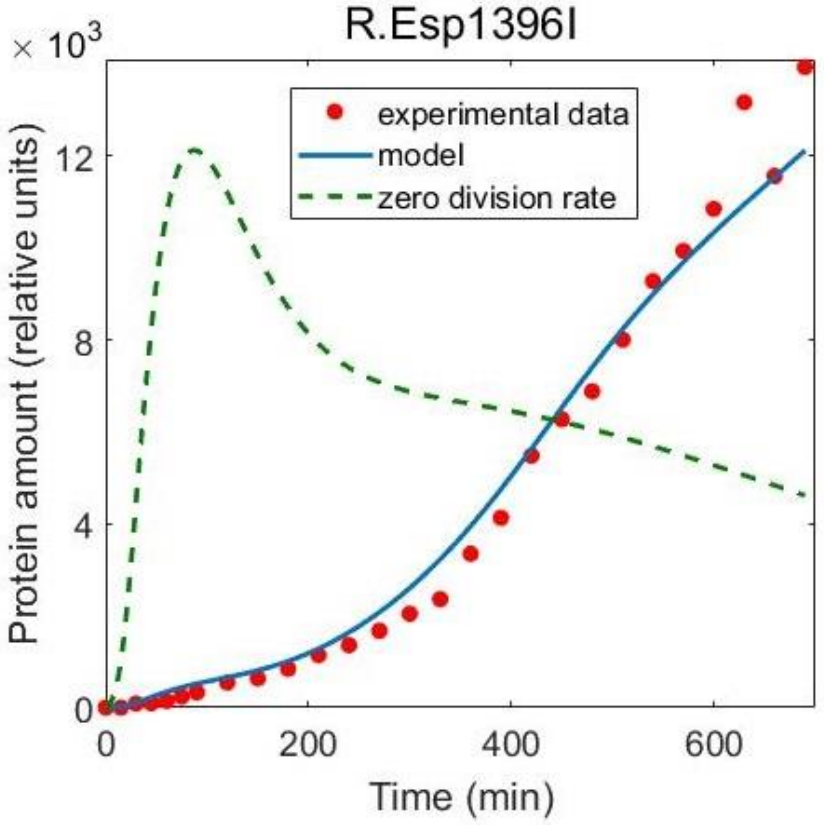
A very good agreement of the model with the data

Identifying false regulatory mechanisms



Experimental data suggest false activation at high C protein concentration that otherwise does not exist.

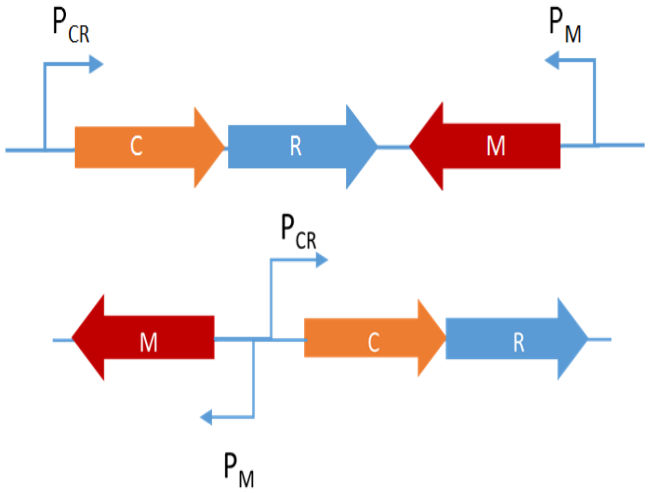
This is entirely a consequence of plasmid number increase, as seen from the model.



Experimental data suggest only activation, or no regulation with stable proteins and transcripts (leading to ~C^2 dependence). The late dynamics is however a serendipitous interplay of repression at high C concentration and increase in the plasmid numbers.

RM system design

- **Different promoter architectures:**
  - **Convergent RM system**
  - **Divergent RM system**



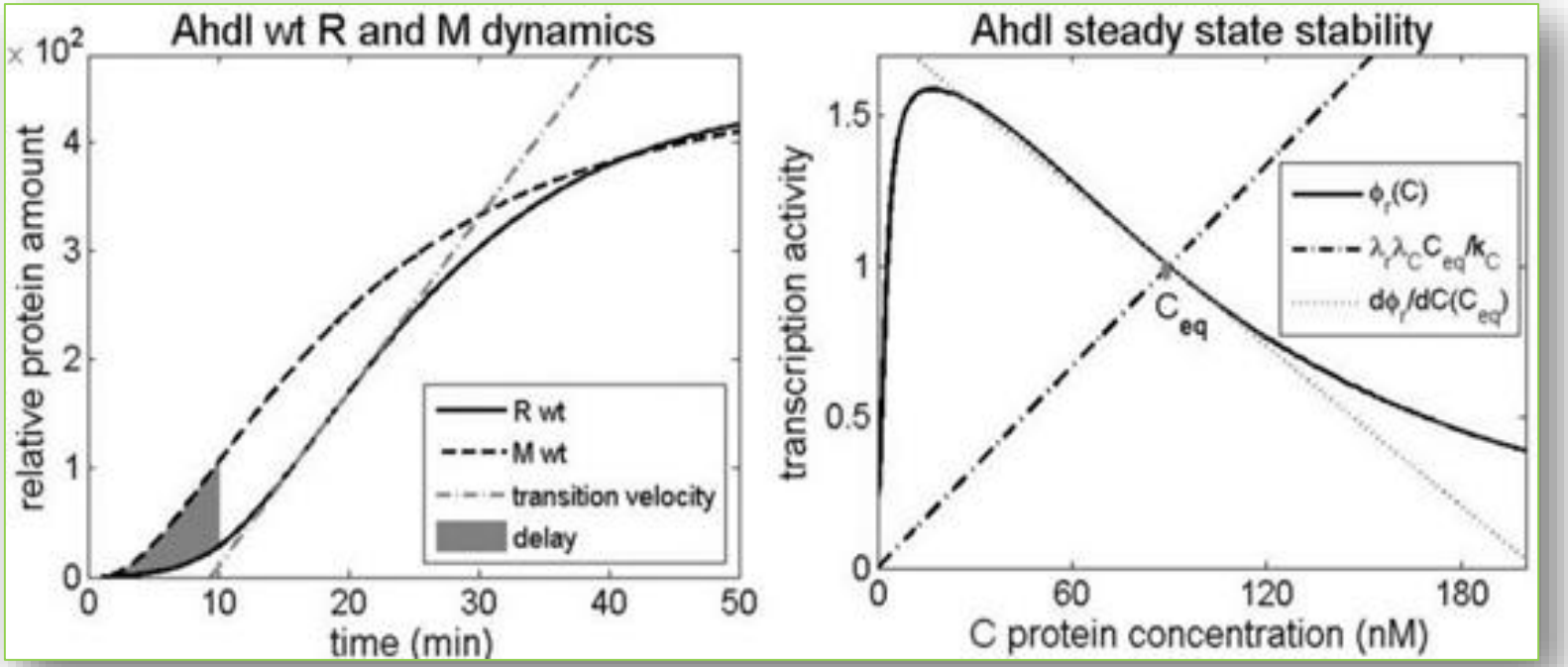
- **Different regulatory features**

(e.g. cooperativity, equilibrium dissociation constants, translation rates…)

Despite all differences, all RM systems should exhibit the same general dynamical properties that ensure safe and efficient RM system establishment.

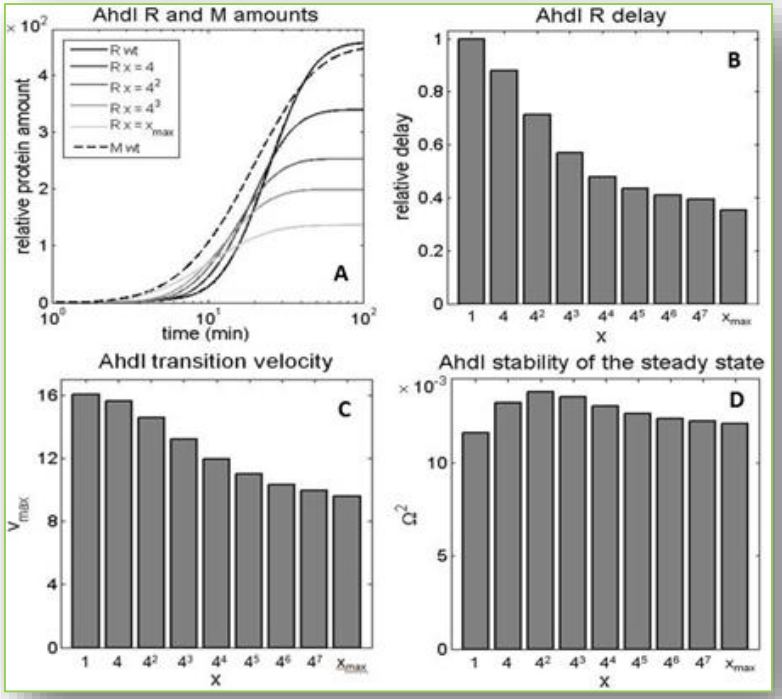
Dynamical property observables

A. Rodic, B. Blagojevic, E. Zdobnov, M. Djordjevic and MD, *BMC Sys Biol* 11, 377 (2017).

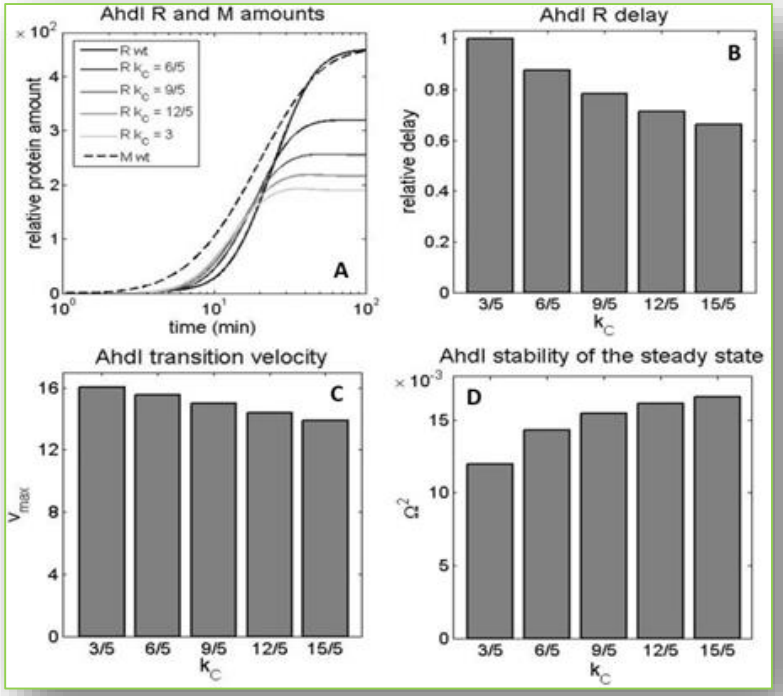


1. **Time delay:** **Ratio of shaded areas in perturbed and in wild-type system**
2. **Transition velocity from "OFF" to "ON" state:** **Maximal slope of R curve**
3. **R steady state stability**

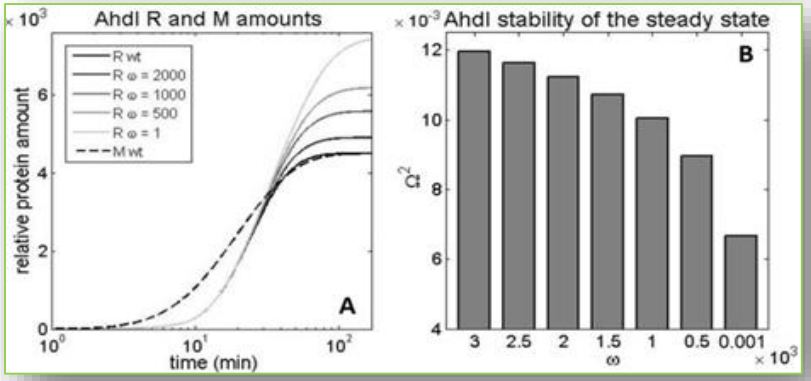
AhdI: Increasing C translation rate



AhdI: decreasing C dimerization constant



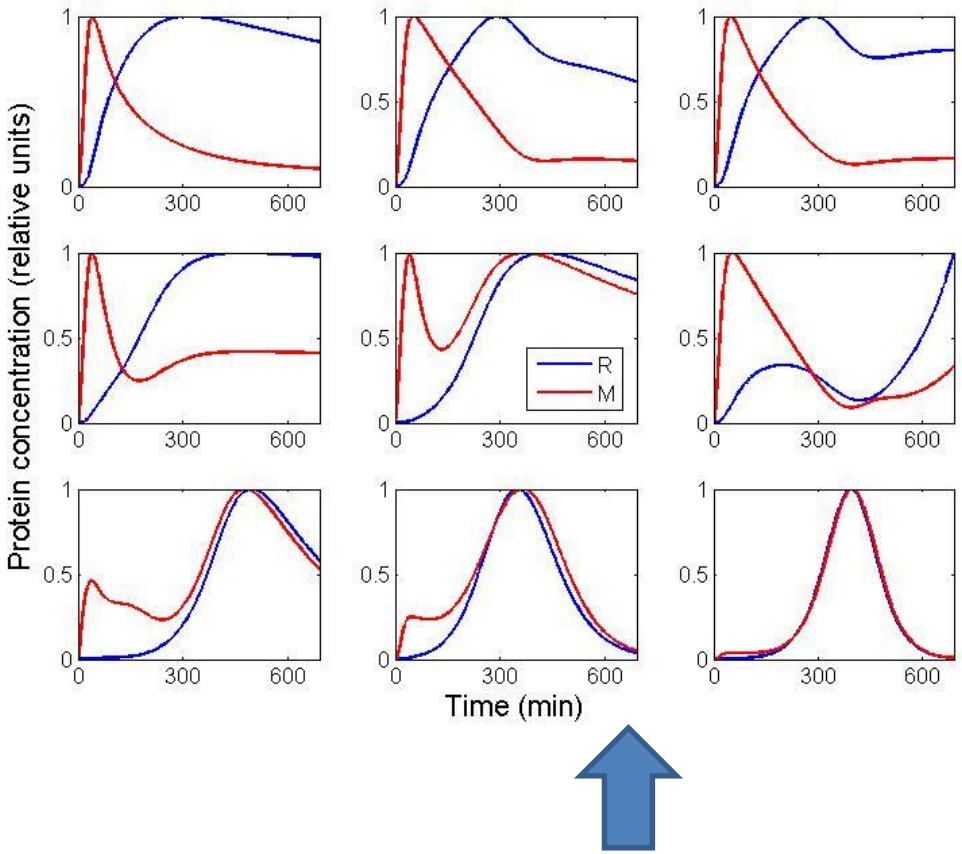
AhdI: Decreasing C dimer binding cooperativity



Perturbing any of the characteristic system features, diminishes some of the dynamical property observables.

A. Rodic, B. Blagojevic, E. Zdobnov, M. Djordjevic and MD, *BMC Sys Biol* 11, 377 (2017).

Perturbing population dynamics



Perturbing the population parameters
(due to e.g. changing physiological conditions)
significantly distorts the resulting dynamics,
leading to very different conclusions on the underlying regulation.

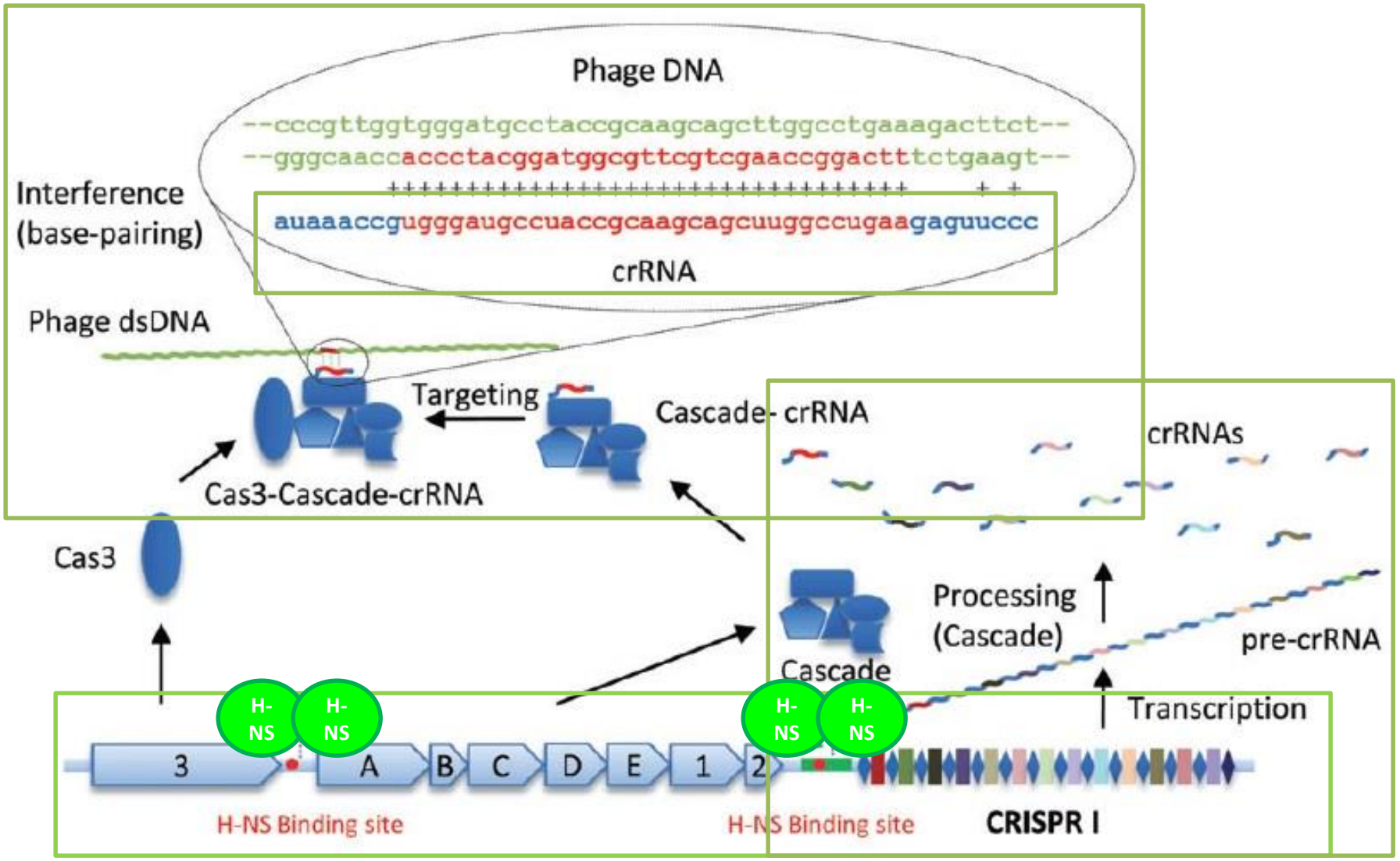
However, we see that the main dynamical constraint (expression of methyltransferase before restrictase) remains robust.

R-M system summary

- **Thermodynamical and dynamical system modeling well explains both *in-vitro* and in *in-vivo* R-M measurements.**
- **Cell population dynamics (dividing cells and plasmids) is necessary to explain data, otherwise leading to identifying false regulatory mechanisms.**
- **Key system regulatory features are different, but can be explained in terms of few simple design principles:**
  - **i) Ensuring a time delay in R with respect to M**
  - **ii) Rapid transition from OFF to ON state**
  - **iii) Increased steady state stability**
- **Key system dynamical constraint is also robust with respect to changing population dynamics, e.g. related with changing physiological conditions.**

Modeling CRISPR-Cas activation

CRISPR/Cas system



F.J.M. Mojica *et al.*, *Mol Microbiol*, 77, 1341 (2010)

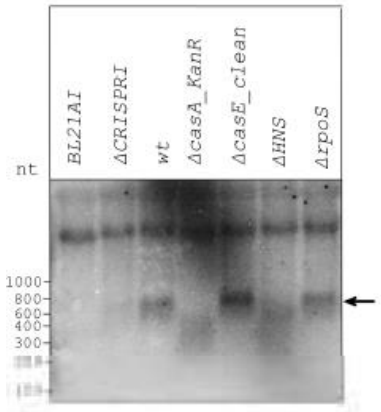
How is the system induced?

- **Artificially…**
  - **– by Cas overexpression**
  - **– by removing promoters repression by H-NS**

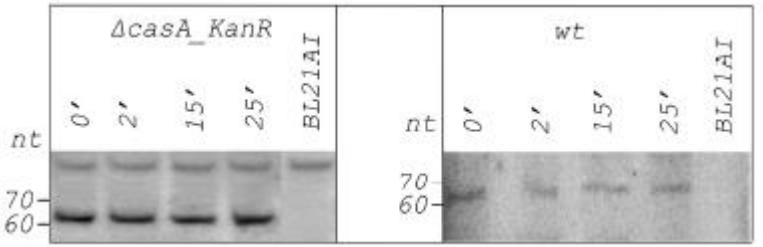
CRISPR transcript processing

Cas gene overexpression

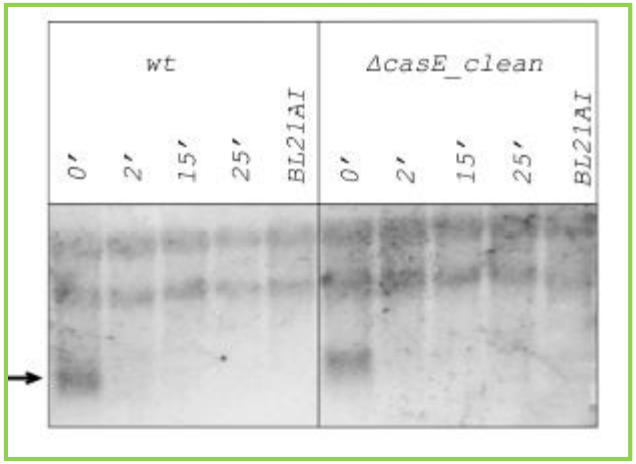
Less than tenfold decrease of pre-crRNAs



A much larger (more than two orders of magnitude) increase of crRNA.



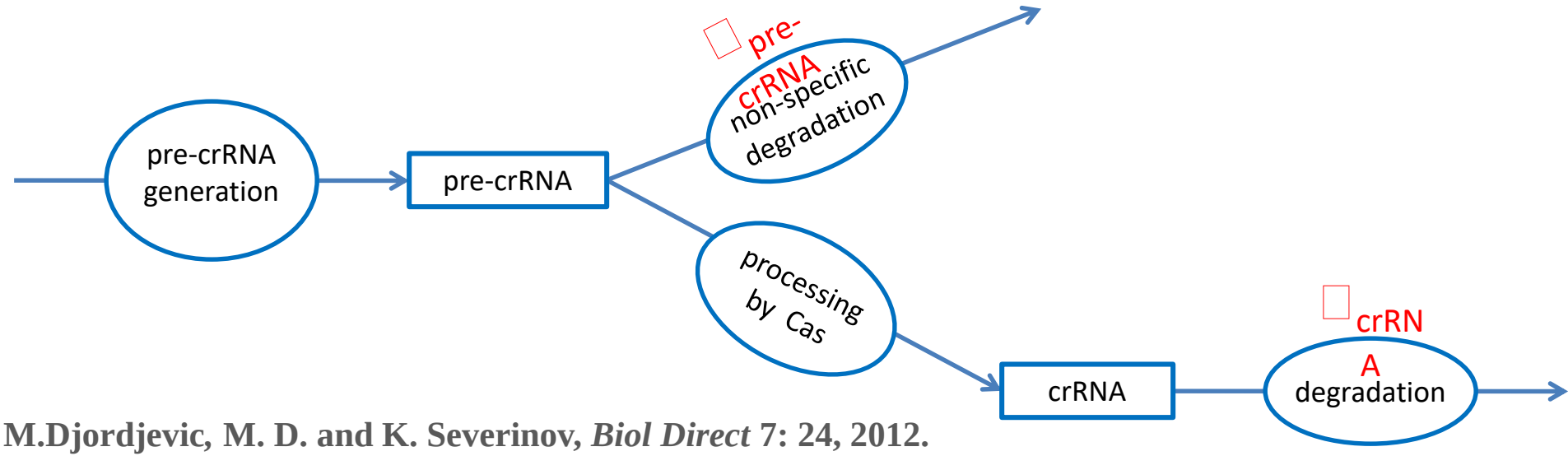
crRNAs are very stable!



pre-crRNA decays fast, which is due to non-specific degradation by an unidentified nuclease.

Pougach et. al., *Mol Microbiol* 77:1367, 2010

Transcript processing model



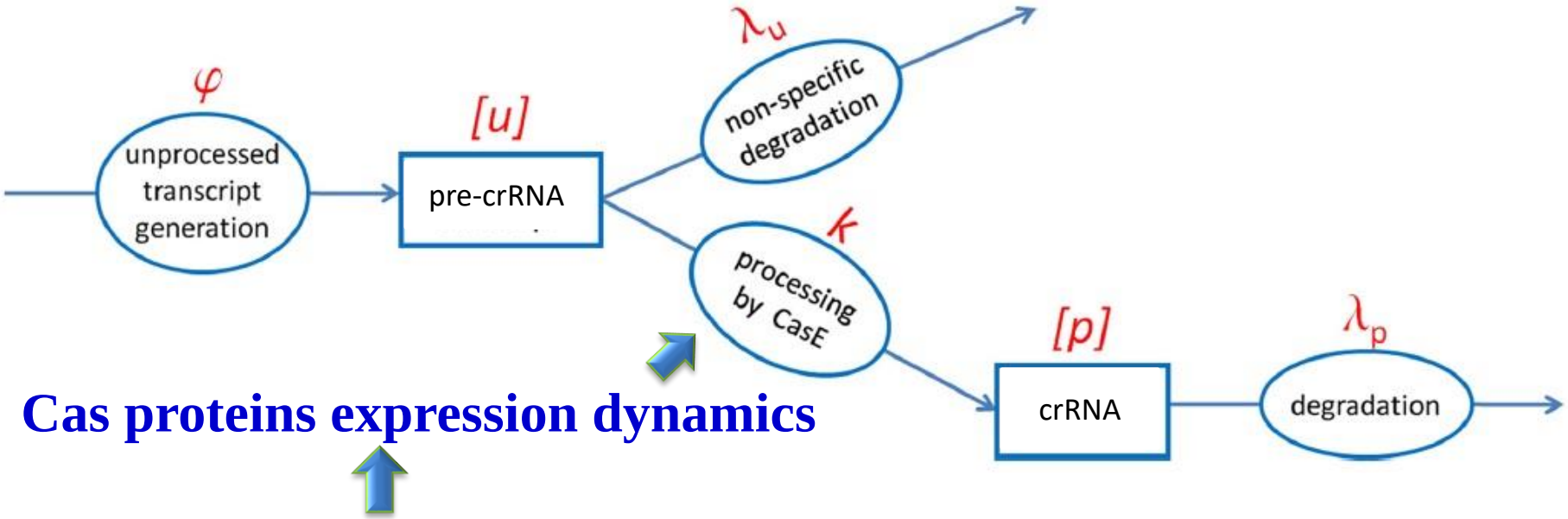
M.Djordjevic, M. D. and K. Severinov, *Biol Direct* 7: 24, 2012.

Overexpression of *cas* genes leads to strong linear amplification of transcripts:

$$\Delta[crRNA] = -\frac{\lambda_{pre-crRNA}}{\lambda_{crRNA}}\Delta[pre-crRNA]$$

Interestingly, this strong amplification crucially depends on fast non-specific degradation of the substrate (pre-crRNA).

How to model transcription regulation?



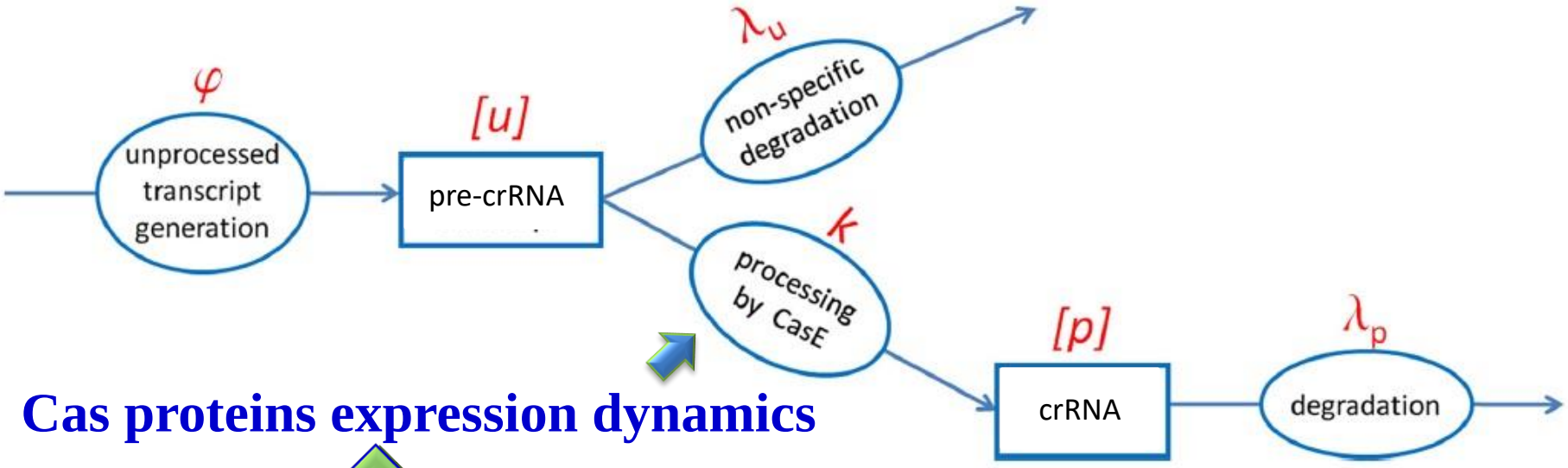
Cas proteins expression dynamics

cas promoter transcription regulation	≈	RM system transcription regulation

- H-NS binds to the promoter in a highly cooperative manner
- H-NS is likely displaced by an activator

- C dimers bind cooperatively to the promoter
- RNAP itself acts as an activator, displacing C dimer

How to model transcription regulation?



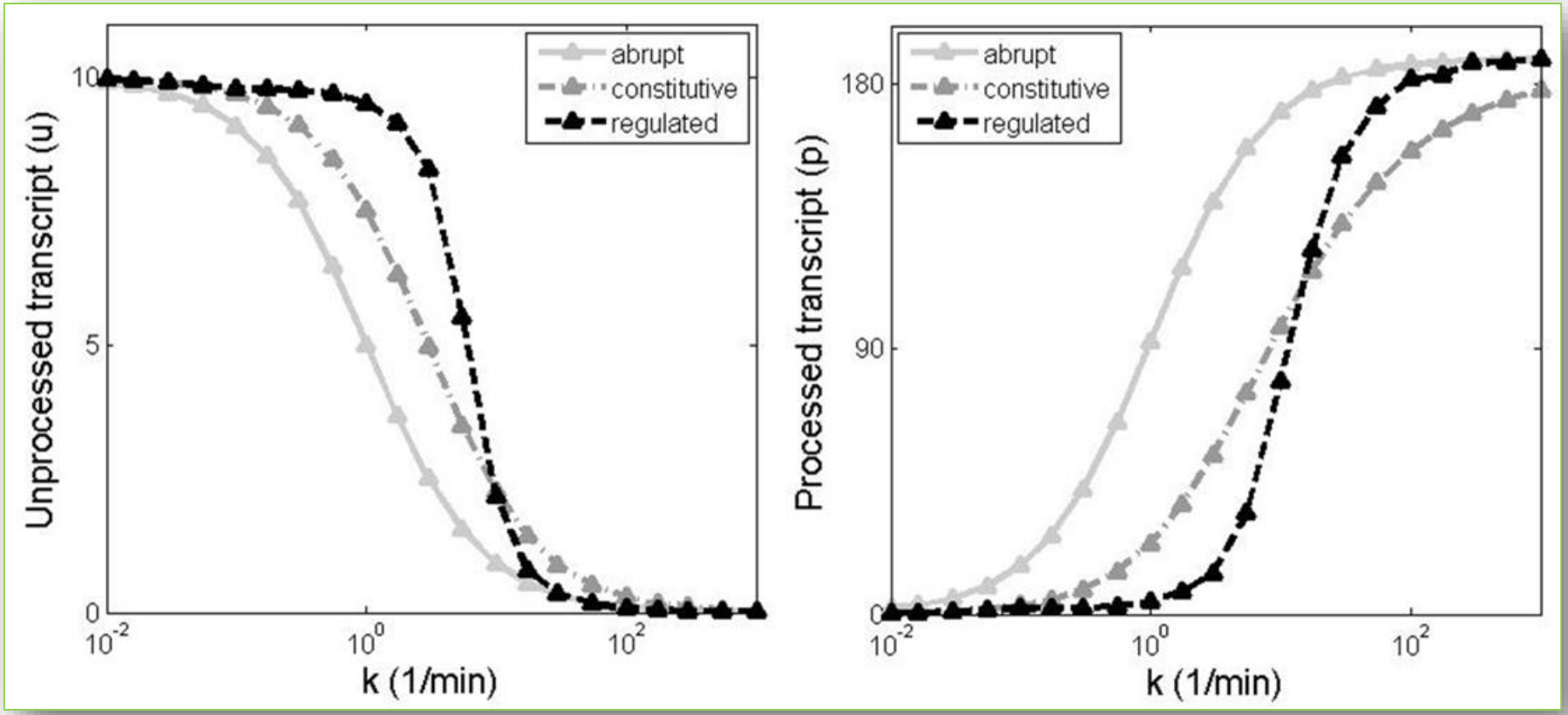
Cas proteins expression dynamics

cas promoter transcription regulation ≈ **RM system transcription regulation**

- **H-NS binds to the promoter in a highly cooperative manner**
- **H-NS is likely displaced by an activator**

- **C dimers bind cooperatively to the promoter**
- **RNAP itself acts as an activator, displacing C dimer**

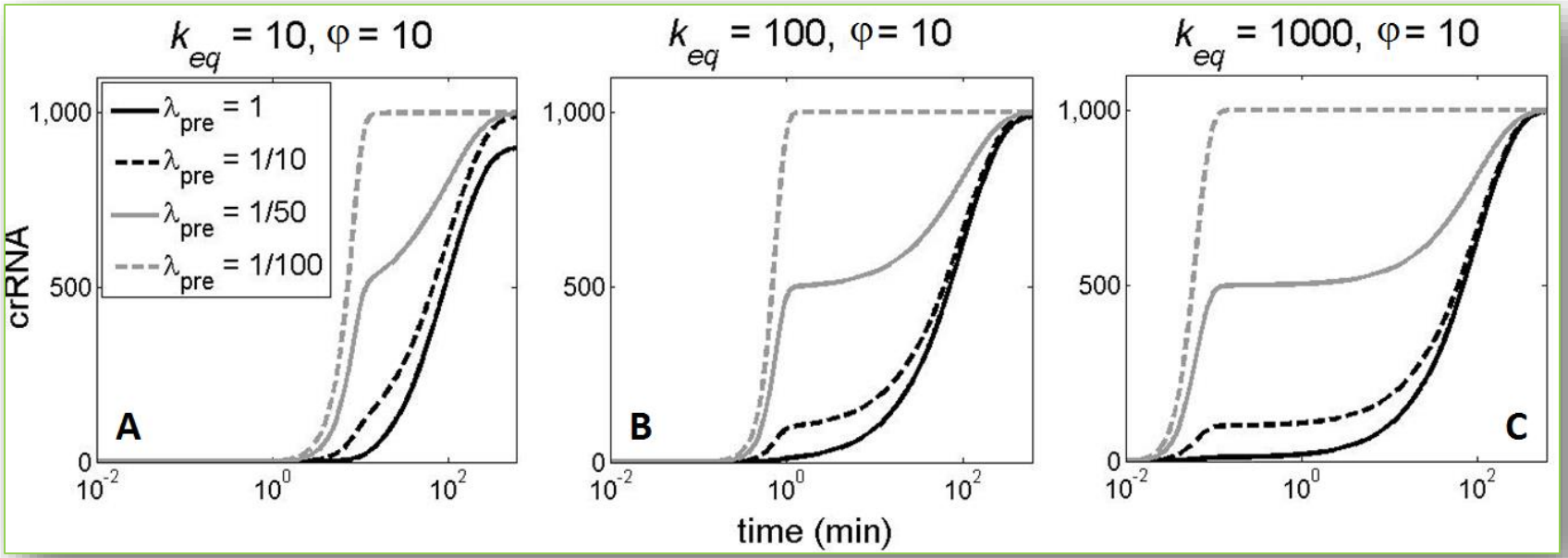
Transcripts concentrations at 20 min post-induction



crRNA expression can also be increased by eliminating repression of CRISPR promoter

Perturbing pre-crRNA processing rate

- **Key system feature – fast (non-specific) degradation of pre-crRNA**
- **If this degradation rate is perturbed (decreased) the time delay is decreased, and the sigmoidal (Hill-shape) response of the system is abolished.**



Conclusion

Both CRISPR-Cas and R-M dynamics is characterized by an initial time delay and a subsequent rapid transition from OFF to ON state.

These properties are ensured not only by the large binding cooperativity but also by a range of other mechanisms such as transcript processing (CRISPR), divergent overlapping promoters, dimerization constant and translation efficiency(R-M).

For Type I-E CRISPR-Cas in *E. coli* allow highly efficient expression of small RNAs in a narrow time interval, with a specified time-delay with respect to the signal onset – perhaps related with function other than immune.

Acknowledgements

Experimental measurements:

Konstantin Severinov's lab:
Rutgers U., Skoltech

Anton Sabanatsev (Esp 1396I)
Ekaterina Bogdanova (AhdI)
Ksenia Pougach (CRISPR/Cas)
Ekaterina Semenova

Modeling:

Computational Systems Biology group:
U. of Belgrade, Faculty of Biology

Andjela Rodic
Stefan Graovac
Bojana Blagojevic
Magdalena Djordjevic

Ministry of Science and Education of Serbia

FP7 Marie Curie International Reintegration grant

SNSF Scopes project