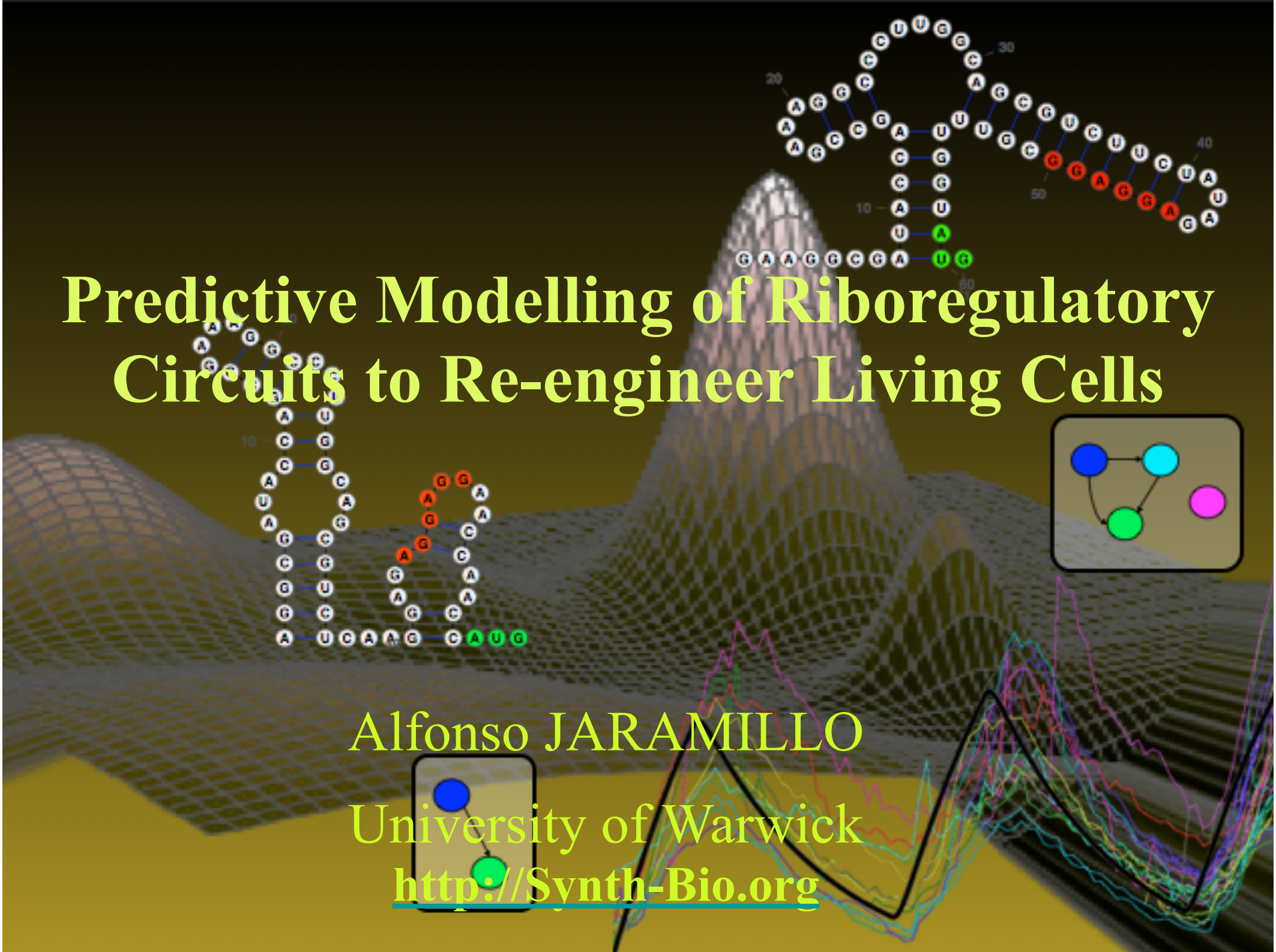
Predictive Modelling of Riboregulatory Circuits to Re-engineer Living Cells



Alfonso JARAMILLO

University of Warwick

http://Synth-Bio.org

Outline

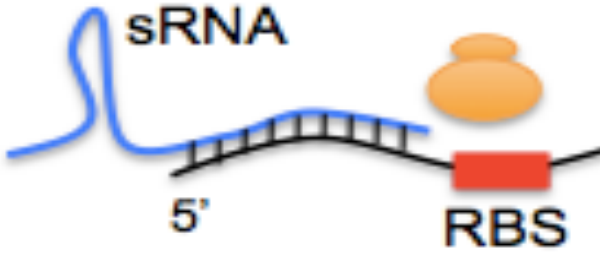
- Introduction to RNA, its folding and its design.
- Single-cell characterisation
- Engineering RNA circuits

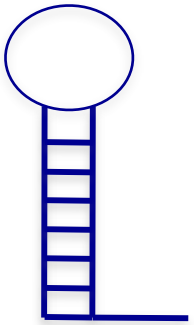
About RNA

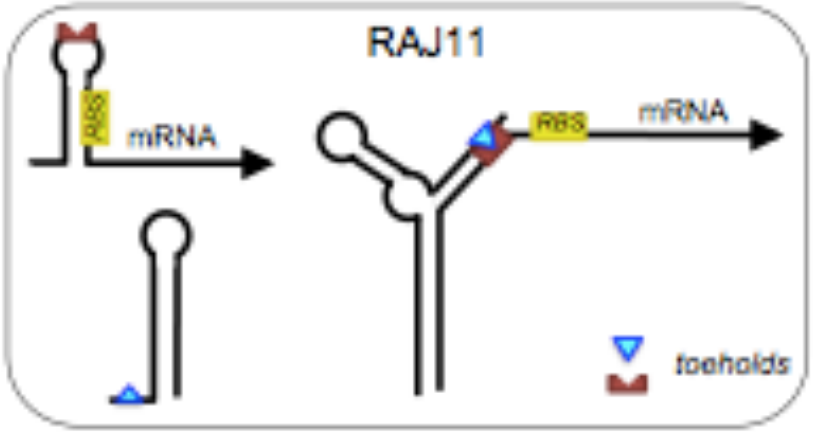
A brief introduction of RNA

RNA is a single-stranded molecule that is transcribed from a double-stranded DNA template.

As RNA is single-stranded it can pair with another RNA molecule, or with itself to form complex structures.



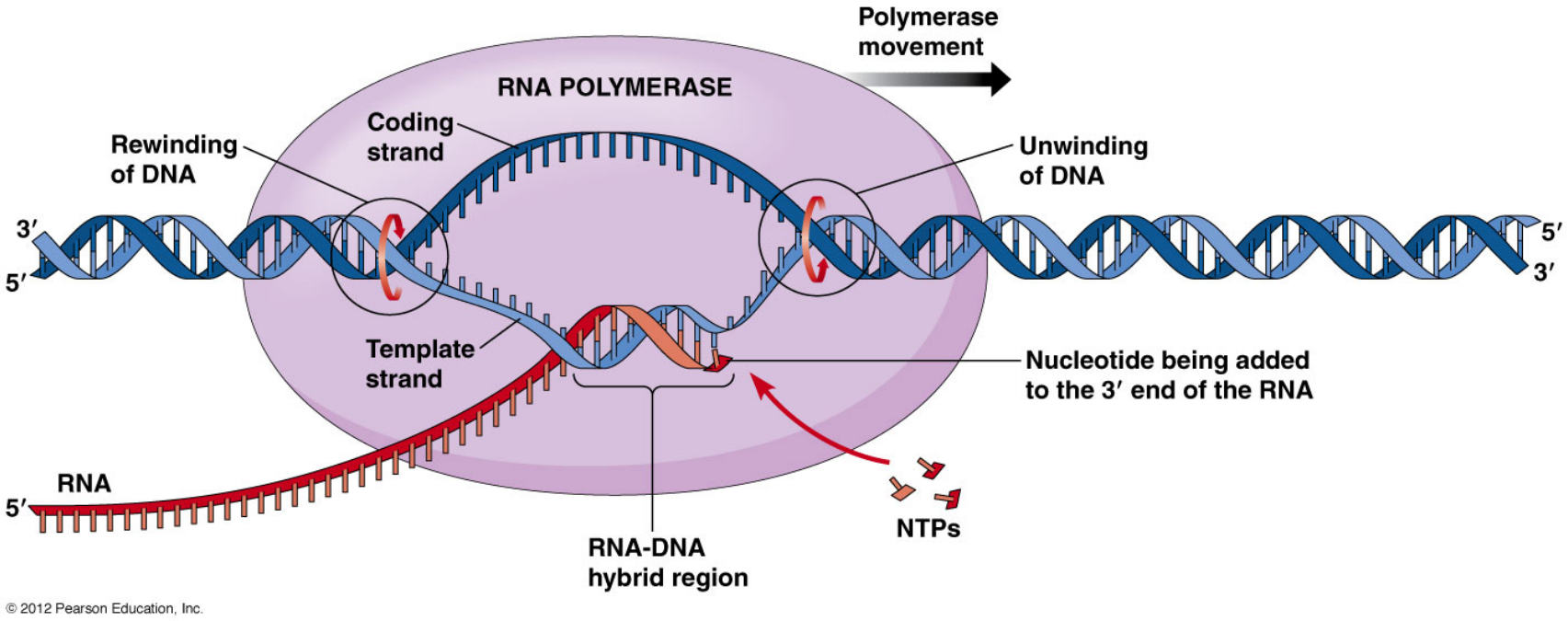




These structures can be used as the basis for producing RNA switches or RNA circuits.

Transcription

RNA is produced when an RNA polymerase moves along a dsDNA. The polymerase attaches to a region called the promoter and proceeds in the 5' -> 3' direction, this process is called transcription.

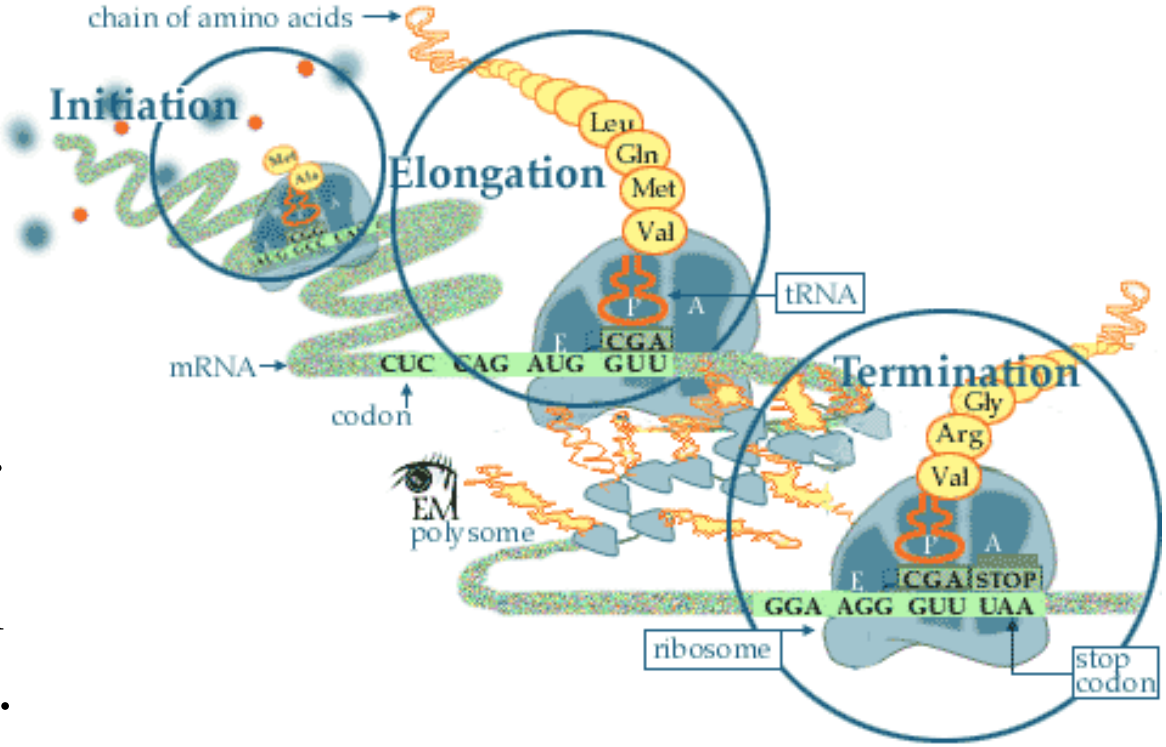


The RNA polymerase will continue to generate an RNA until it reaches a terminator. At this point the polymerase pauses, tracks in the reverse direction (3' -> 5'), and after a short journey detaches from the dsDNA. The RNA molecule is now decoupled from the polymerase and dsDNA.

Translation

- Proteins are composed of many amino acids joined together. Amino acids are encoded on the RNA as a triplet base (i.e. 3 nucleotides next to each other), e.g. ACG GCU UCC GAA encodes Threonine-Alanine-Serine-Glutamine.

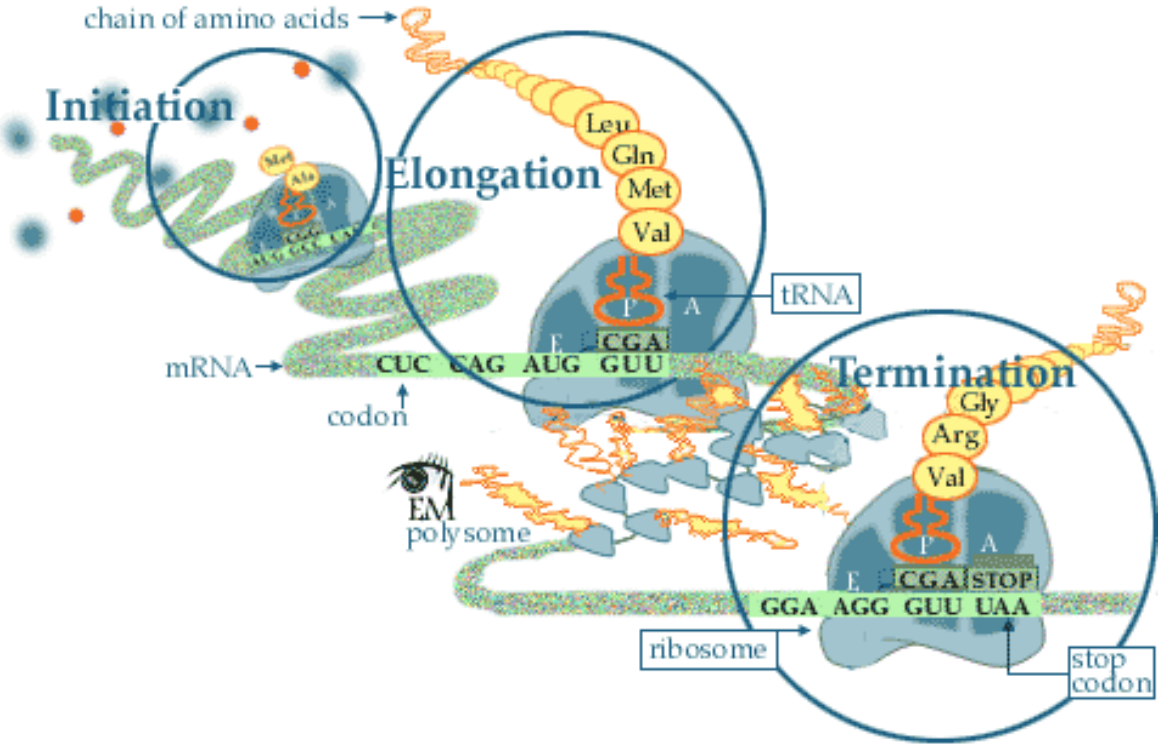
Translation is the process that generates proteins from an RNA template. The ribosome binds to a region of RNA near the 5' end that is called the Ribosome Binding Site (RBS). The ribosome then moves in the direction of the 3' end until it reaches a start codon (AUG).



Nobelprize.org

- Once the ribosome has attached to the start codon it proceeds along the RNA generating an amino acid chain until a stop codon is reached (UAA, UGA or UAG). When the stop codon is reached the extension of the amino acid chain stops and the peptide is released from the RNA.

In bacteria translation occurs at the same time that the RNA is being produced in transcription – they are coupled. The force of the ribosome moving along the emerging RNA can be enough to force the polymerase to keep transcribing the RNA.



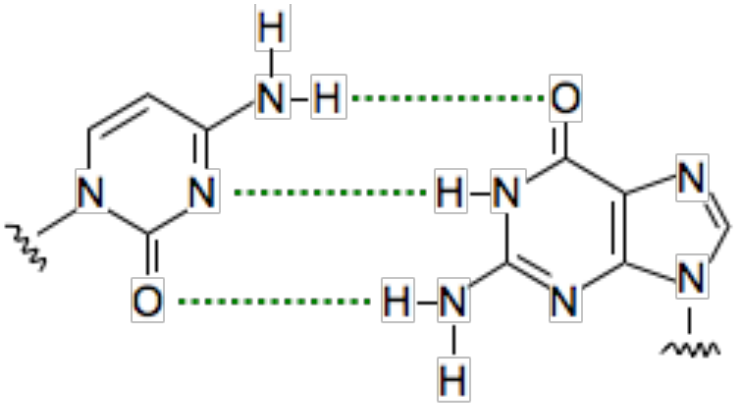
Nobelprize.org

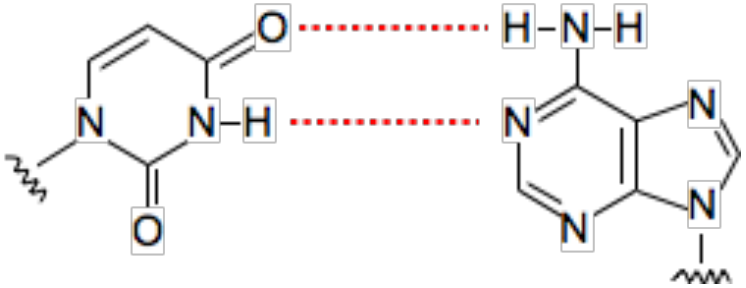
Energies

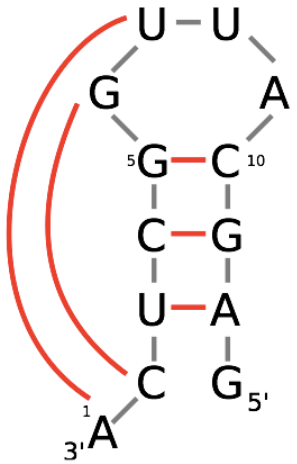
Base pairing energy

$$E(S) = \sum_{i.j \in S} \epsilon(r_i, r_j)$$

G-C	-3 kcal/M
A-U	-2 kcal/M
G-U	-1 kcal/M





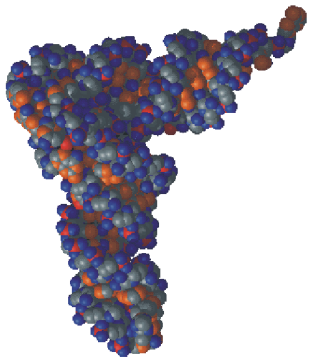



```
A C U C G G U U A C G A G
[ [ ( ( ( ] ] . . ) ) ) )
```

$$P = \{(1,7),(2,6),(3,12),(4,11),(5,10)\}$$

Folding

Query sequence



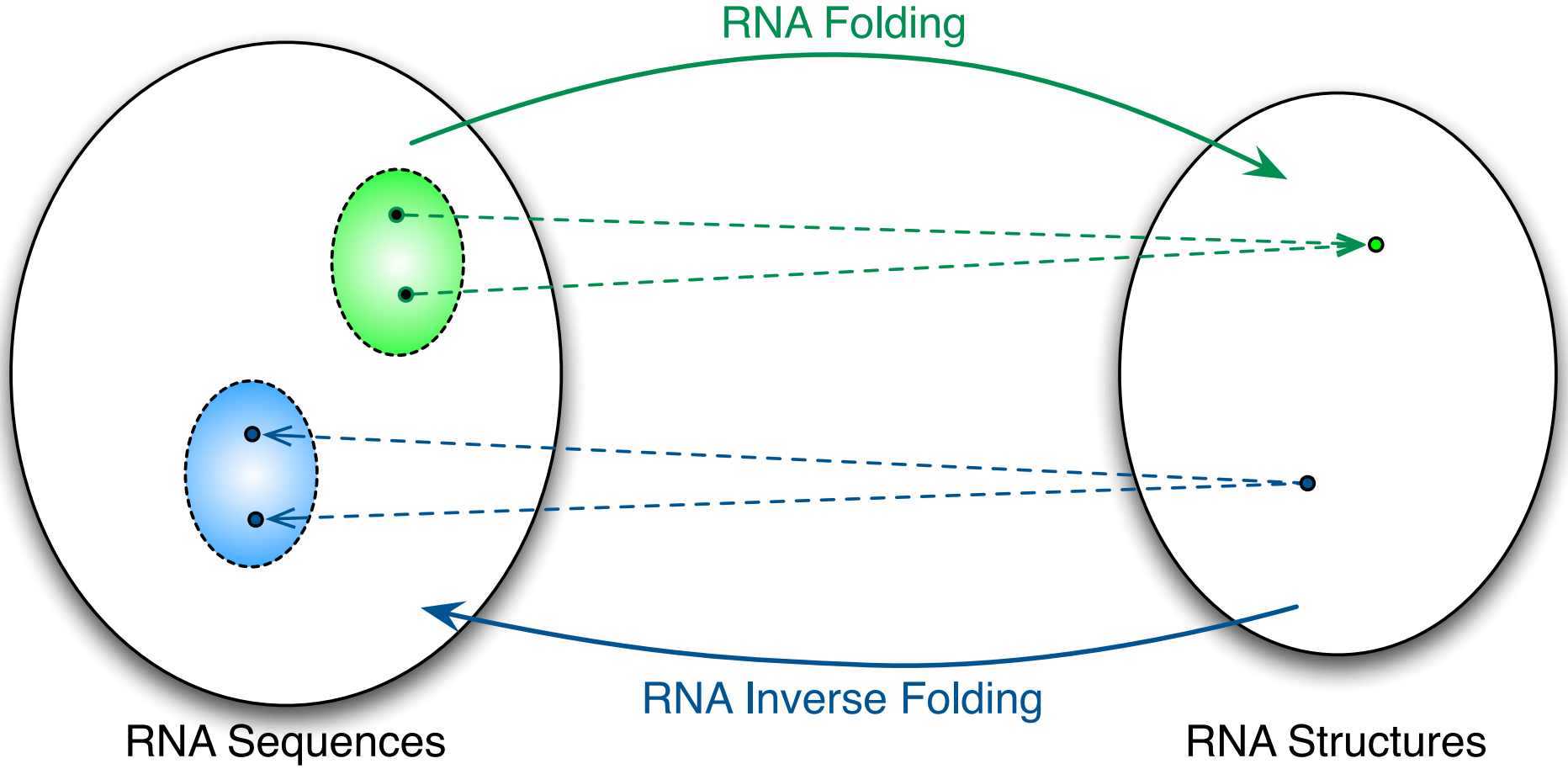
Fold 1

Fold 2

Fold 3

Fold N

RNA inverse folding



Folding can be solved in O(n3) time complexity and O(n2) space complexity (Nussinov algorithm)

Inverse folding is NP-hard

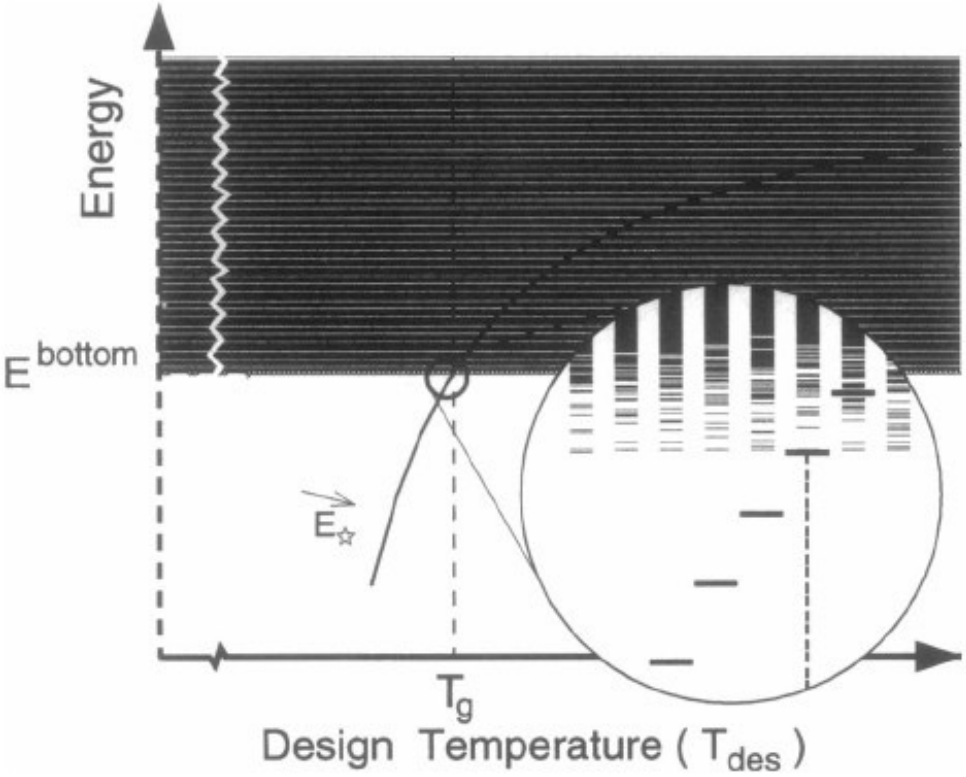
Conformational ensemble

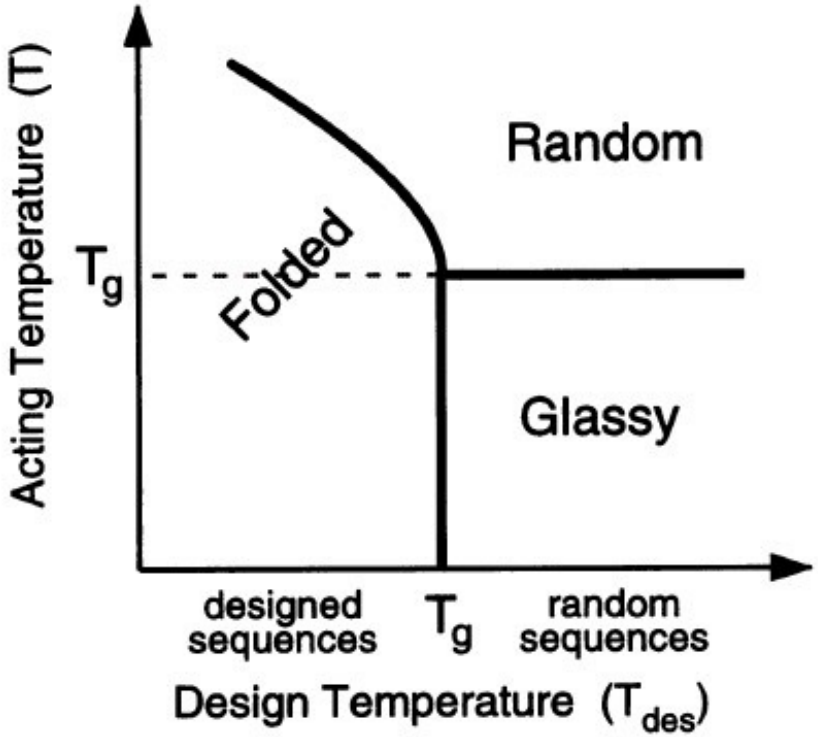
$$P(\text{Secondary Structure}) = \frac{e^{-\Delta G(\text{Secondary Structure})/RT}}{Q}$$

$$P_{i,j} = \sum_k \frac{e^{-\Delta G(k)/RT}}{Q} = \left(\frac{1}{Q}\right)\sum_k e^{-\Delta G(k)/RT} = \frac{Q_{i \text{ paired to } j}}{Q}$$

where Q is the partition function and k is the sum over all structures with the i-j base pair.

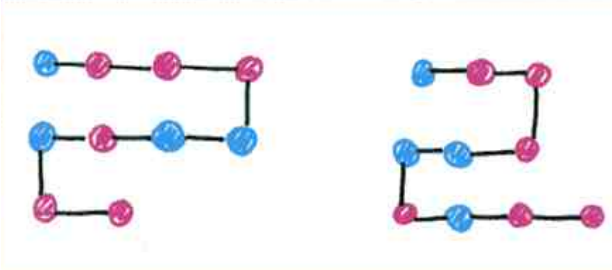
Sequence-structure space



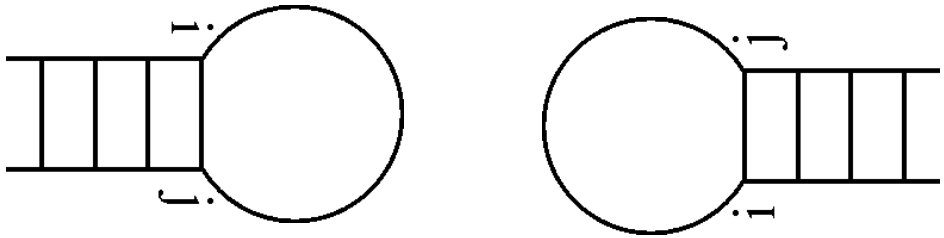


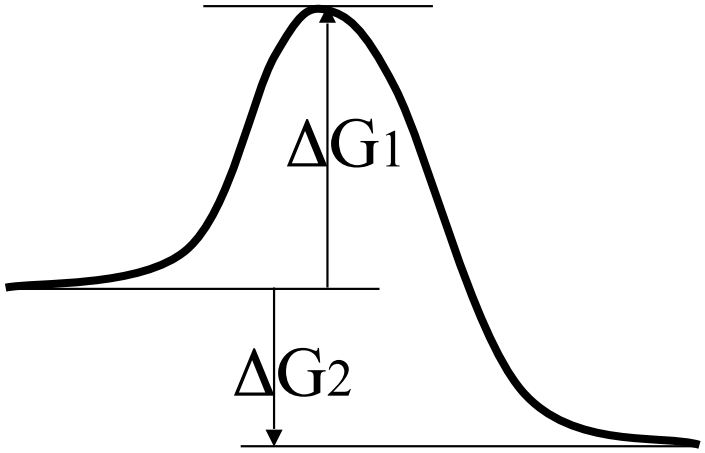
(Pande *et al.*, 2000)

Square heteropolymers as models of proteins gave a designability phase diagram.



RNA-RNA interactions





ΔG1 lowest

ΔG2 lowest

De novo design by evolutionary computation

- Generate sequence diversity, select with a fitness/objective function and iterate.
  - Start from a known/random sequence and suggest new sequences by single/multiple mutation/shuffling
  - Using a folding free energy as fitness: Inverse folding problem
- Improvement of fitness function by adding bio-molecular function
  - Adding interactions with other molecules/systems

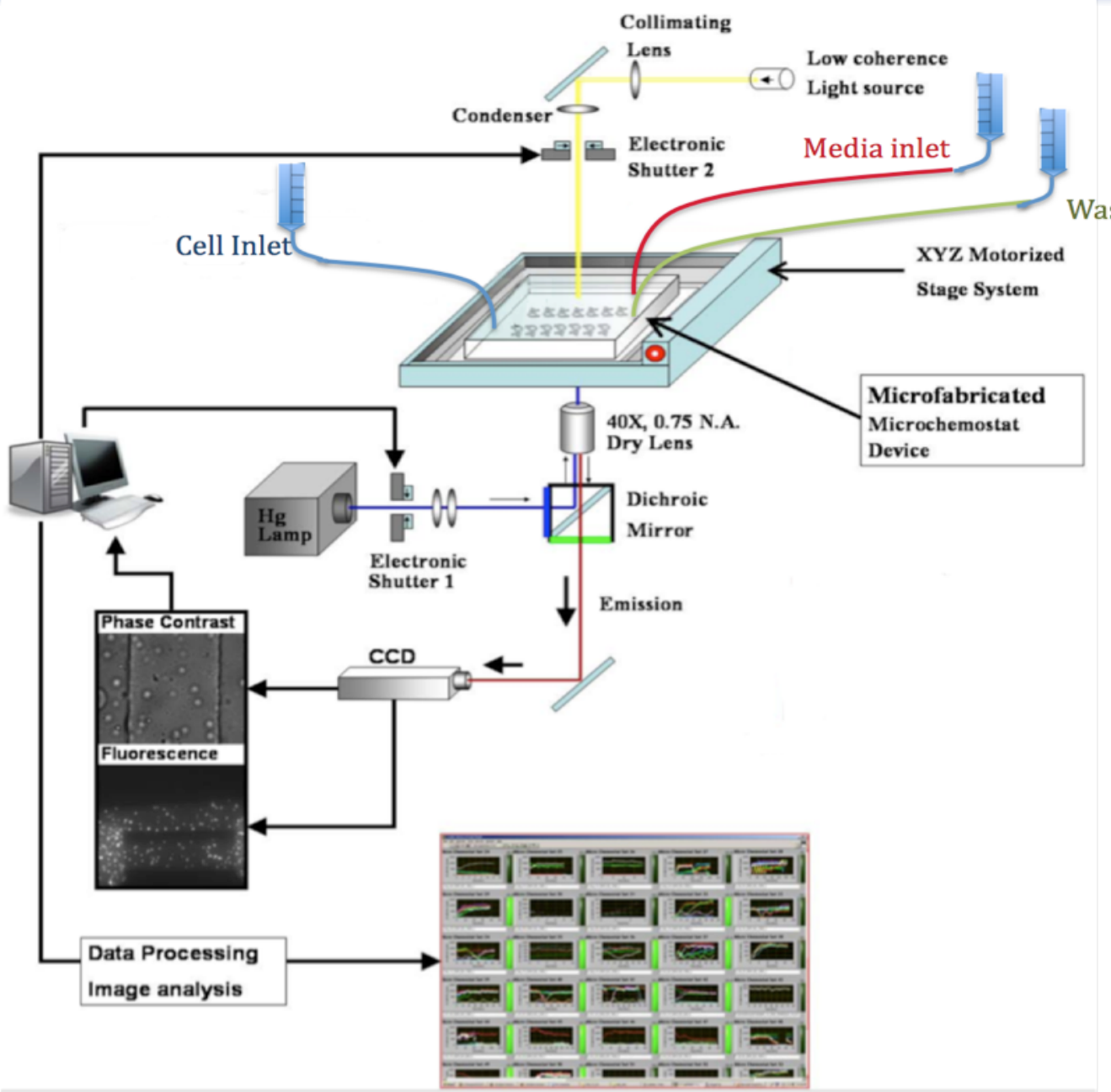
De novo design by evolutionary computation

- Examples of computational *de novo* design
  - Proteins
    - (see **Jaramillo et al. PNAS 2002**)
  - Non-coding RNA and regulatory circuits.
  - Transcription factor circuits
    - (see **Rodrigo et al. NAR 2011** & **Rodrigo et al. ACS Synth Biol 2012**)
  - Genome design
    - (see our **Carrera et al. PNAS 2012**)

About single-cell characterisation of gene dynamics

Overall microfluidics-microscope setup

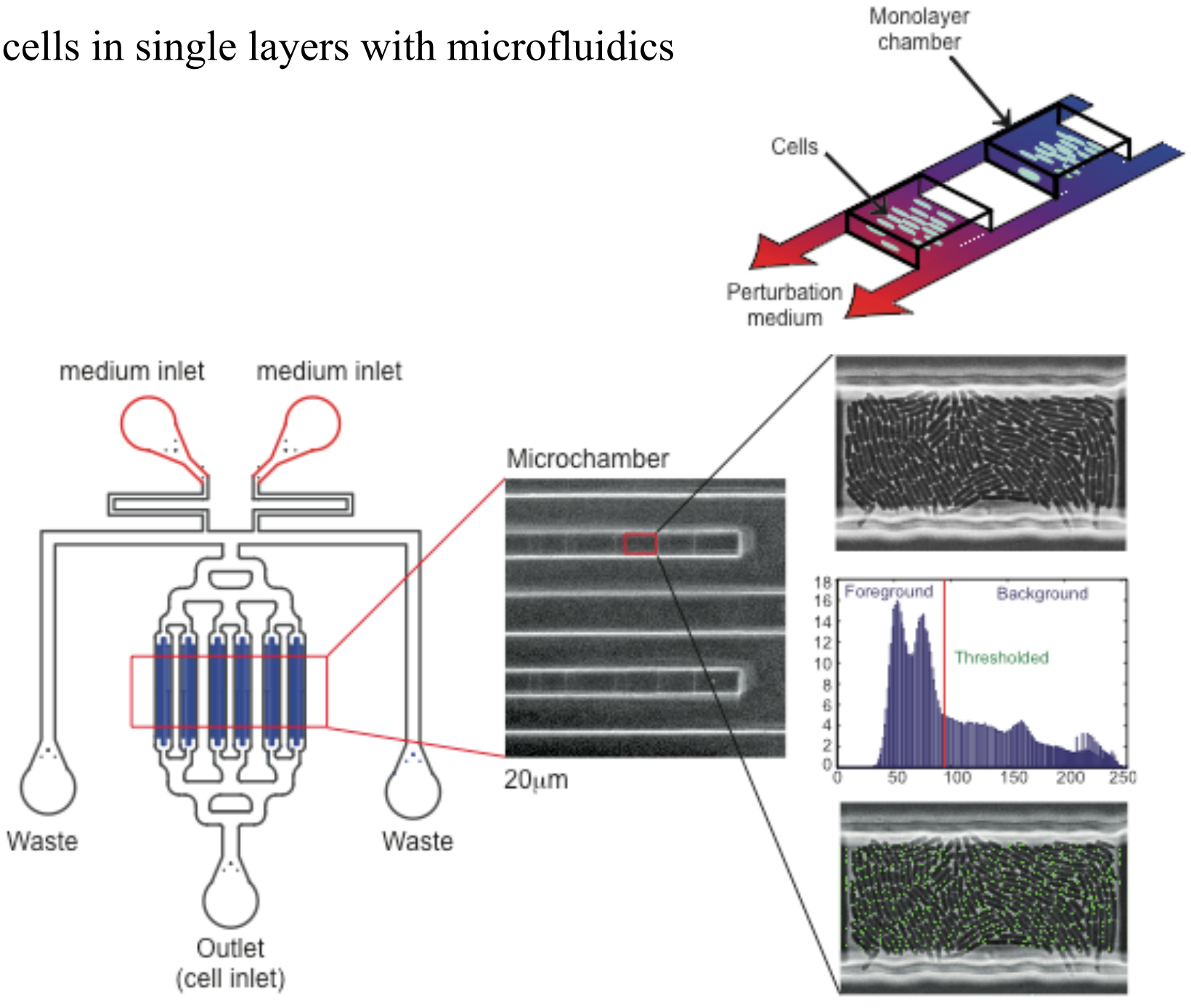
- Constant controlled cellular growth and environment
- Better control over the inducers' levels
- Fixed microscope focal plane



Adapted from Balagadde, 2007

Riboregulator *in vivo* dynamics

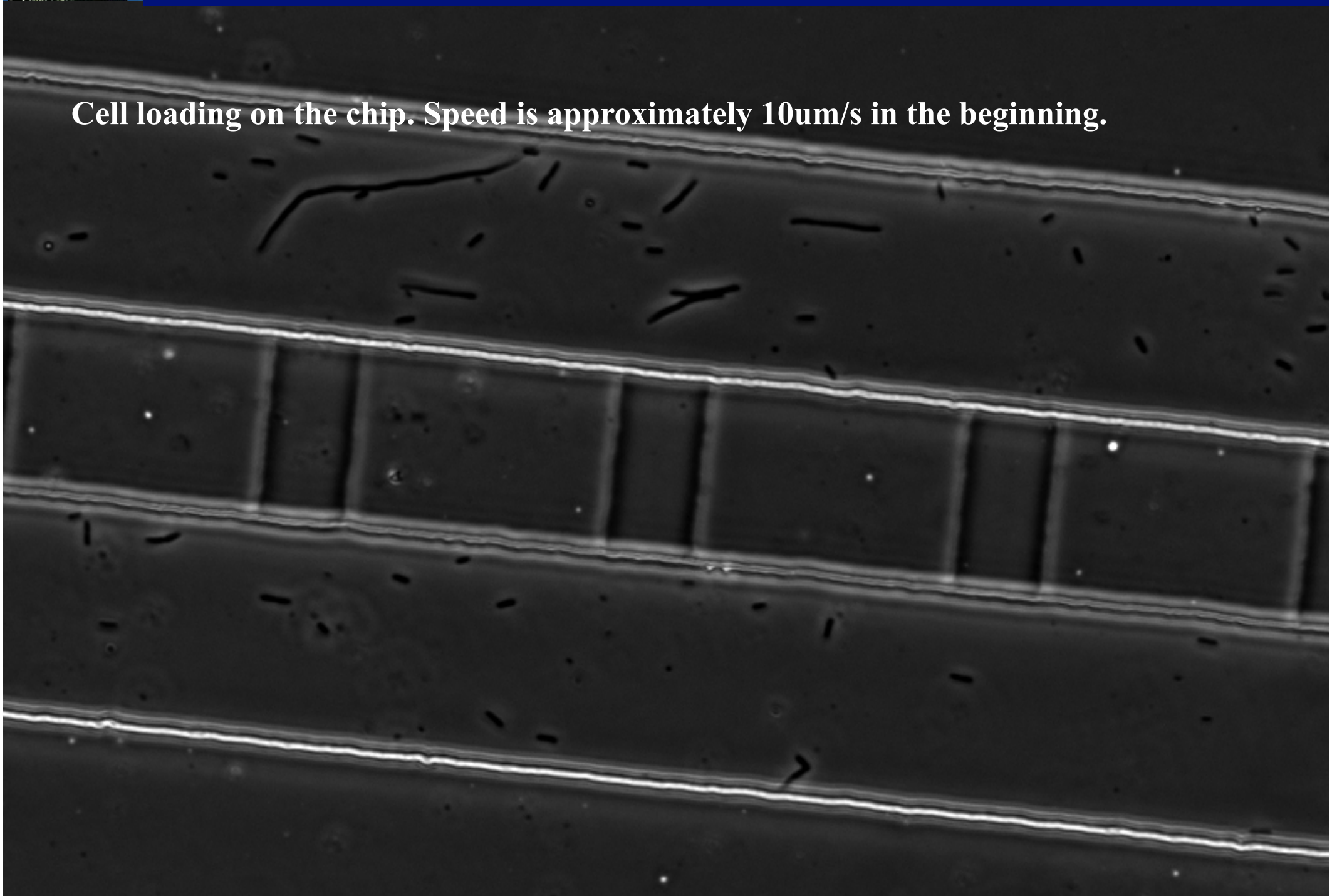
Growing cells in single layers with microfluidics



Filling the biochip

We have a biochip comprising 12 lines of 24 traps each. The traps are 50x40 um.

Cell loading



Cell loading on the chip. Speed is approximately 10um/s in the beginning.

Microchemostat

Exponential growth of *E. coli* cells in a single trap for about 6 hours

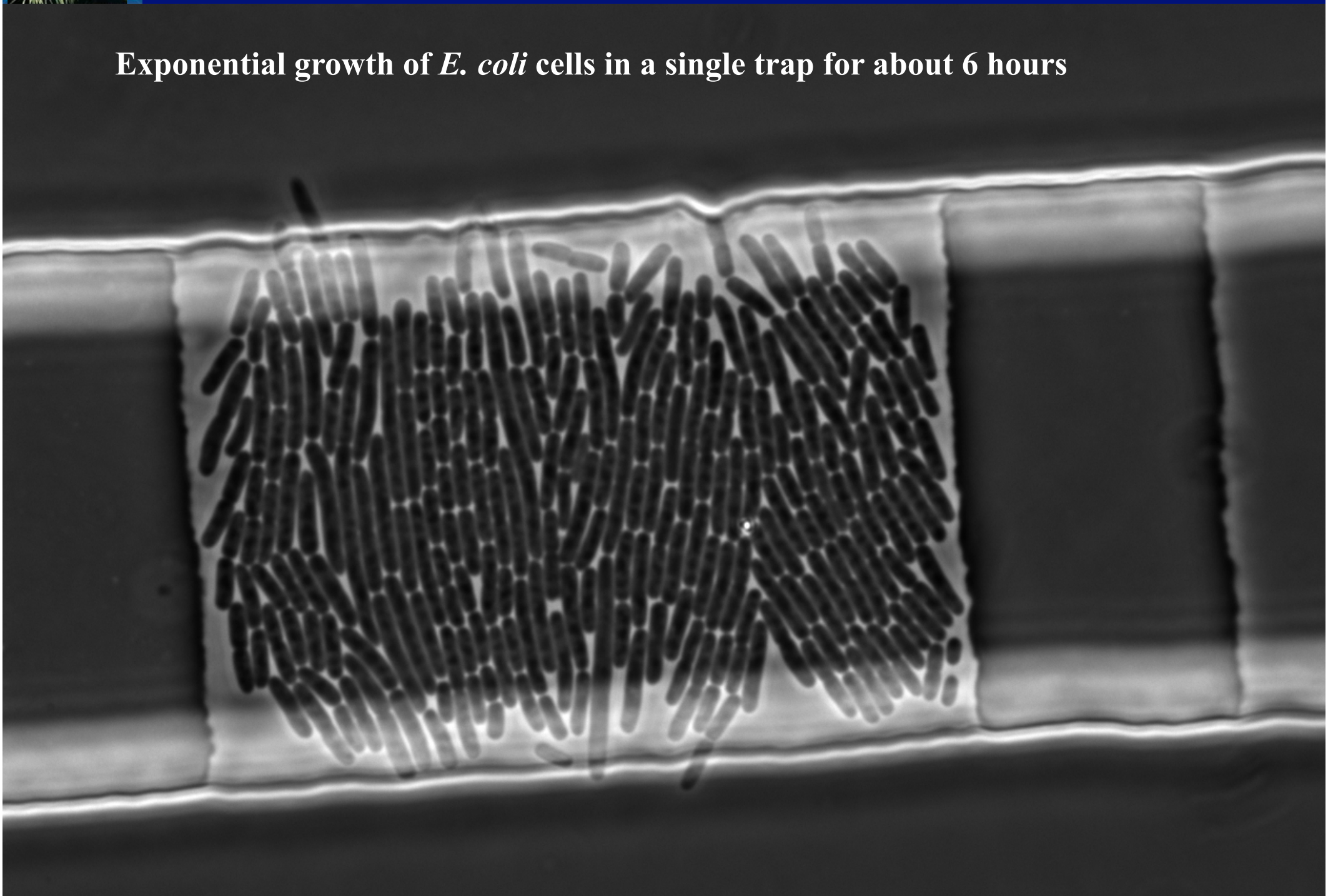


Image treatment

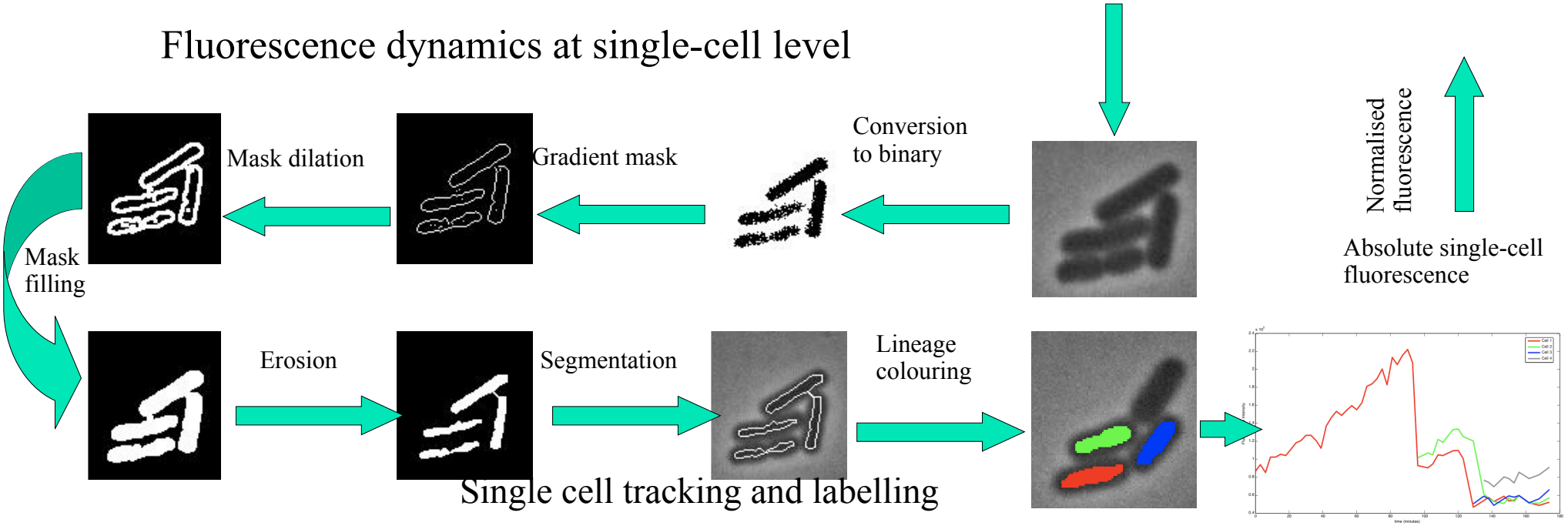
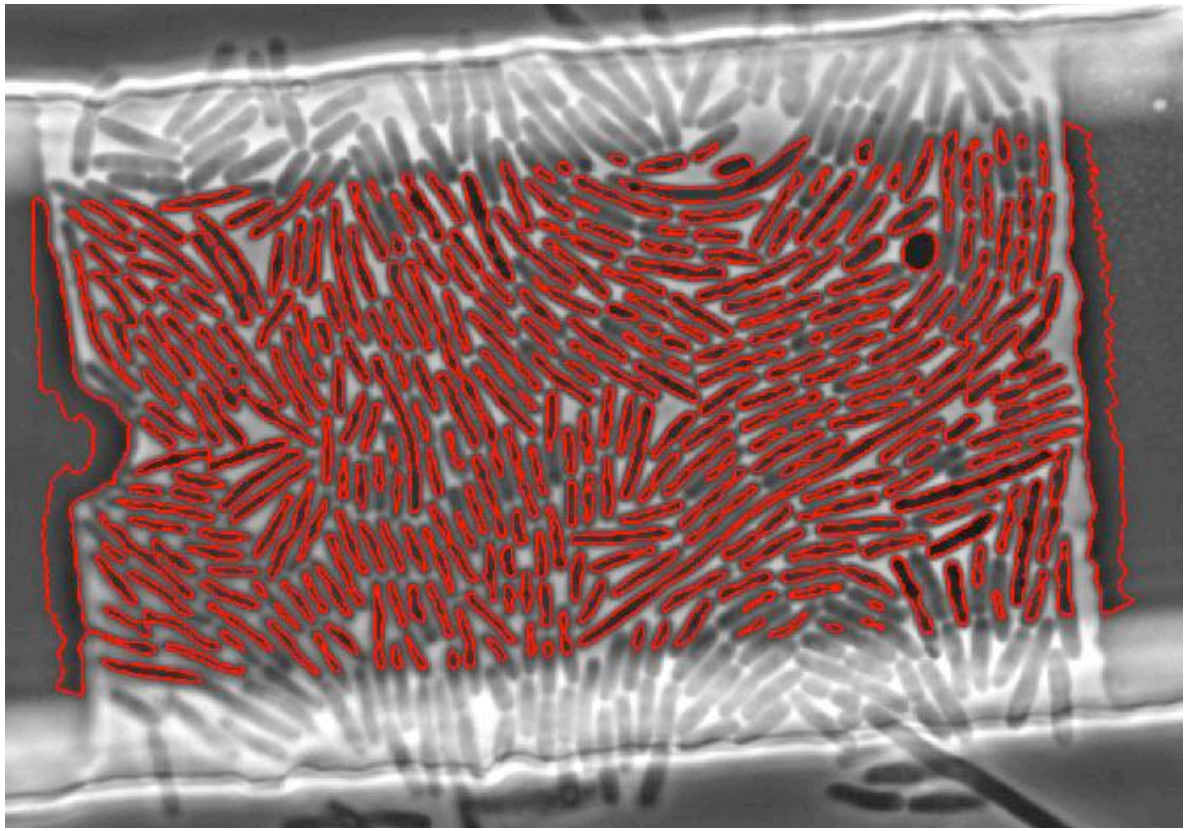


Image treatment

Contours of cells superimposed over the initial image. More than 80% of the cells are correctly indentified in spite of the inhomogeneity of the illumination and irregularities. A magnified image is shown on the right.



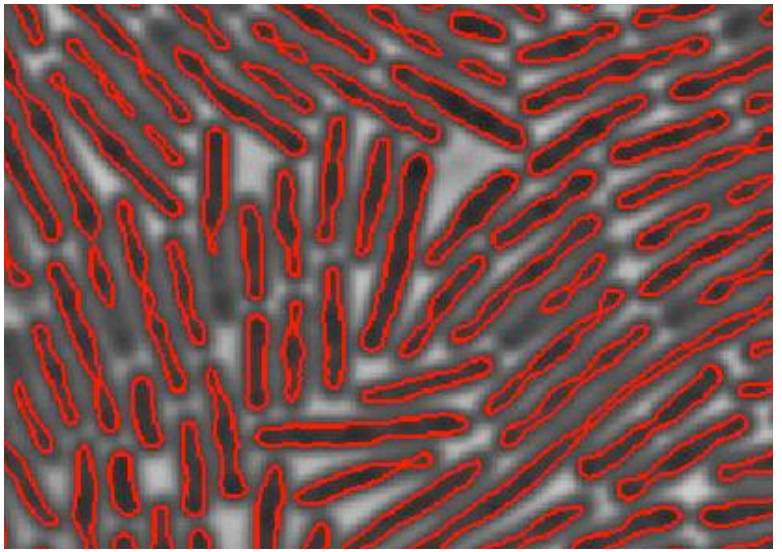
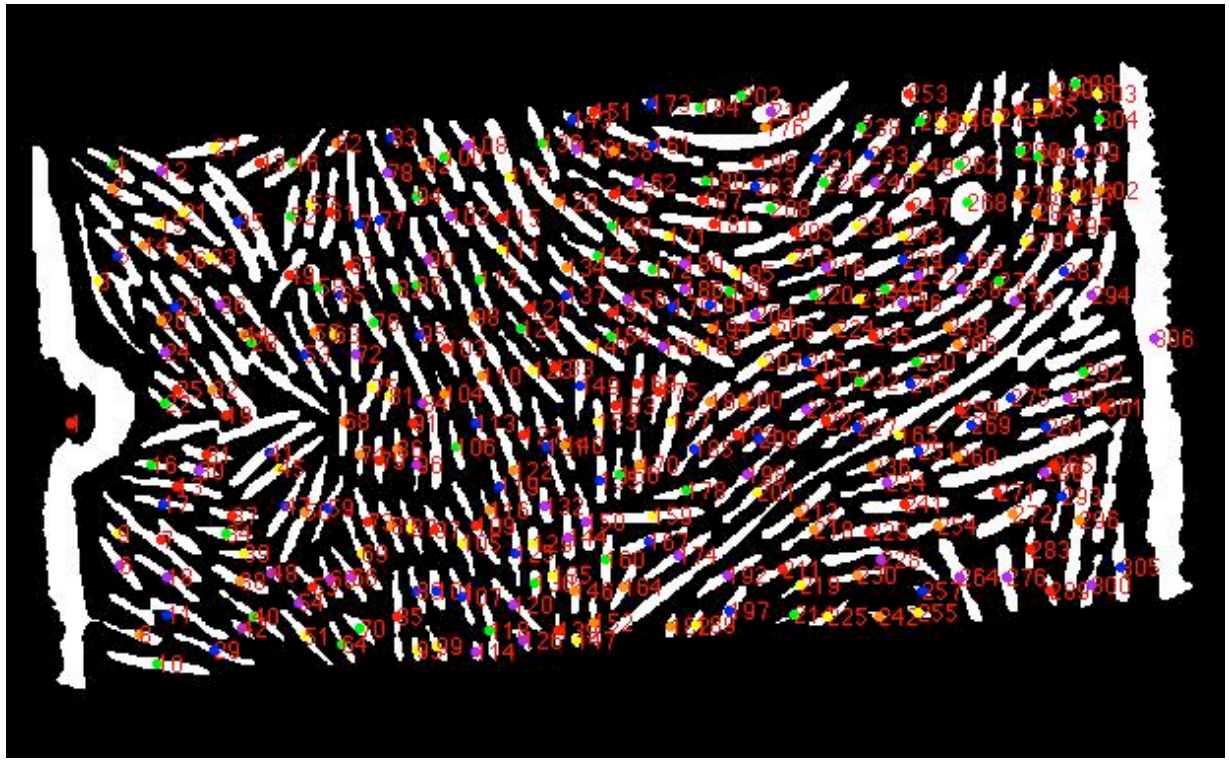
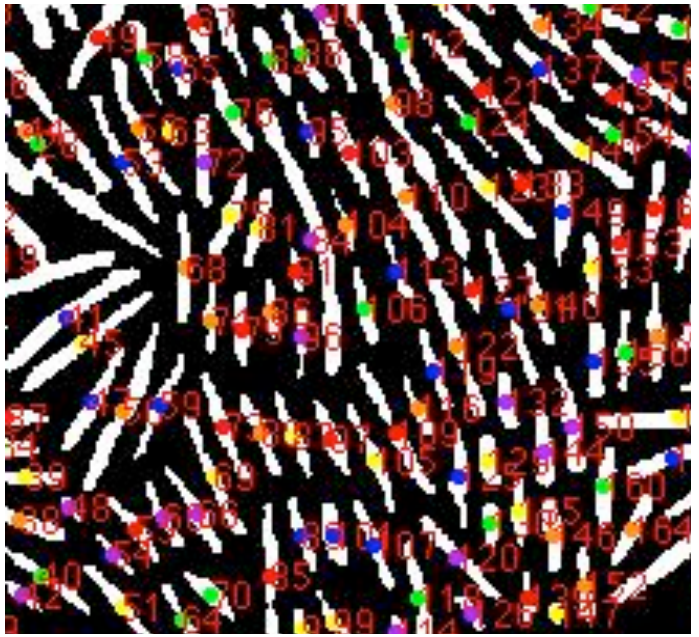


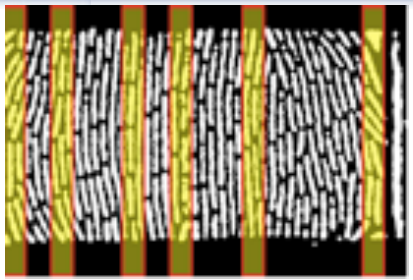
Image treatment

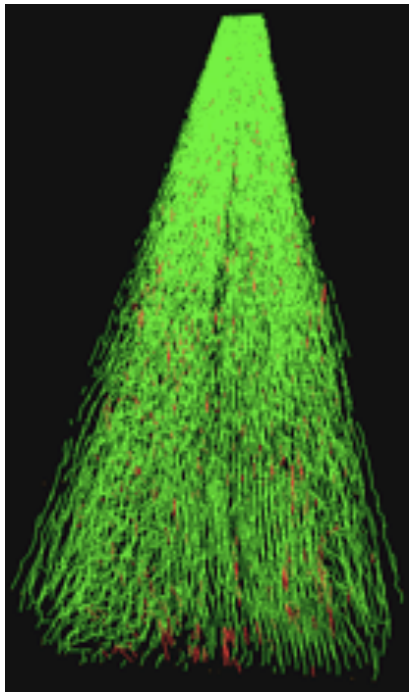
Segmented image with the centroid of each cell marked and numbered; the whole chamber on the left and a detail on the right.

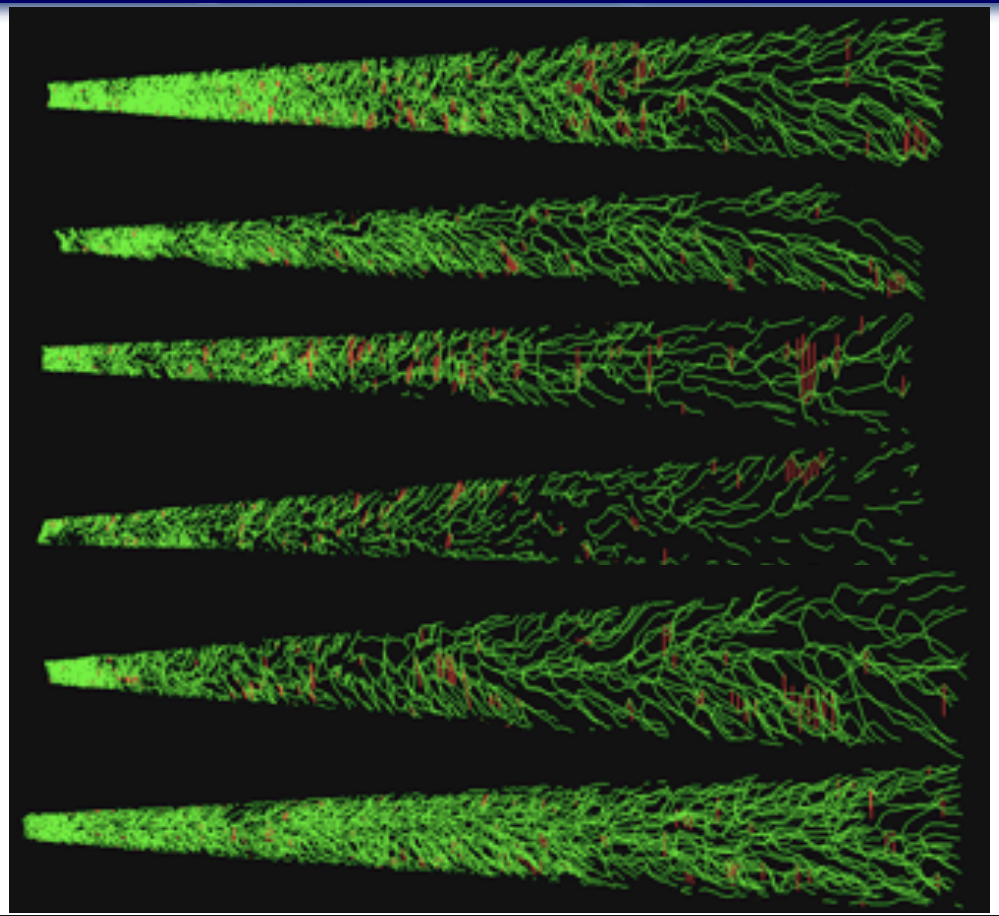


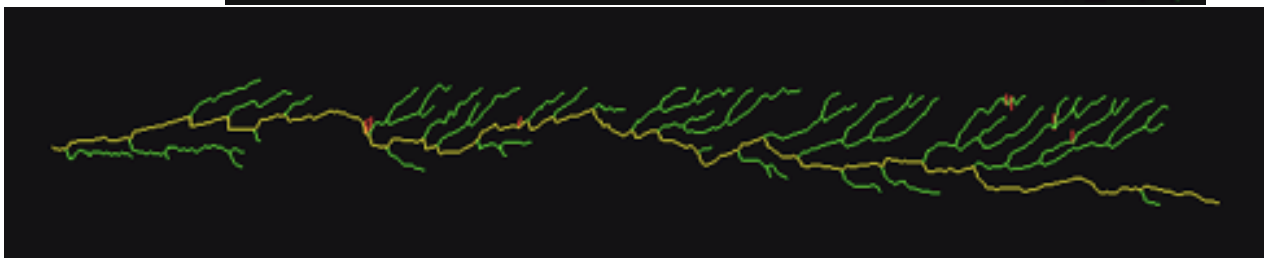


Automated segmentation and tracking of cells









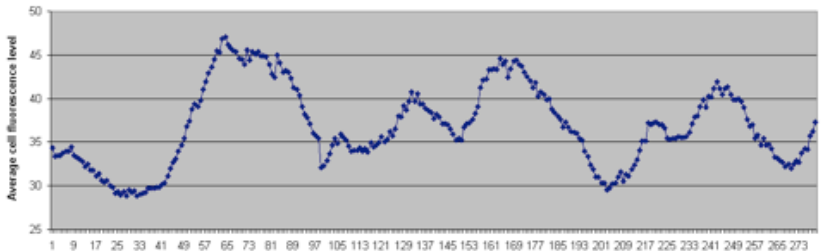
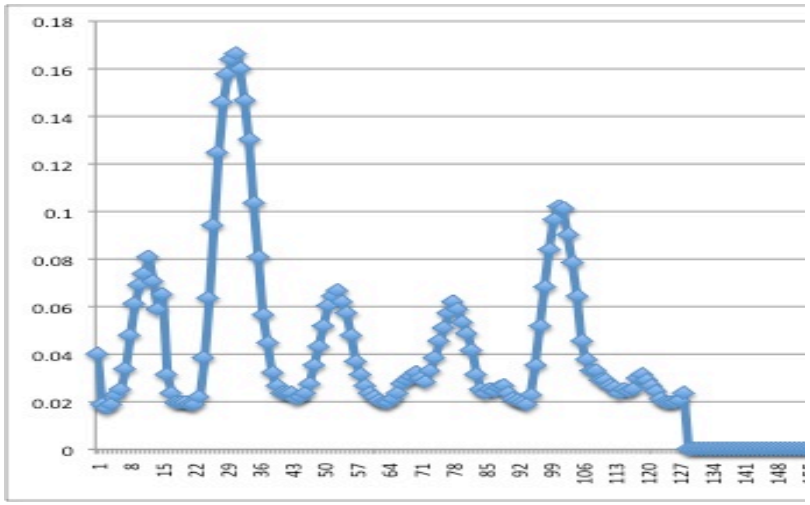
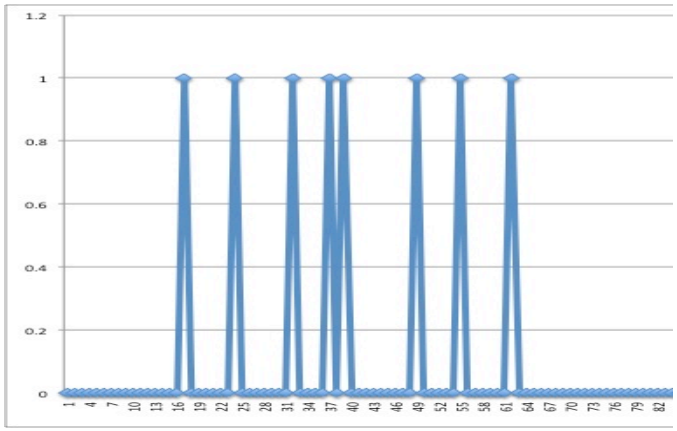


Image treatment

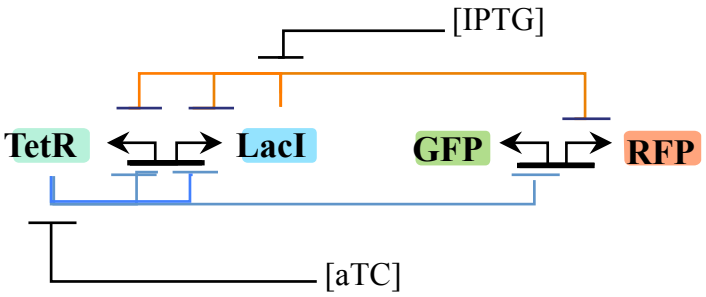
Fluorescence data from a single cell versus time (arbitrary units). The images were acquired every three minutes, and the cell was successfully tracked for more than 7 hours.

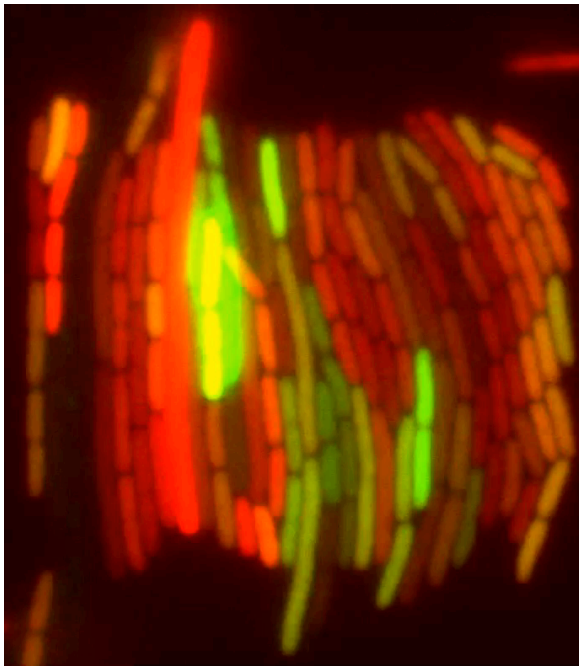


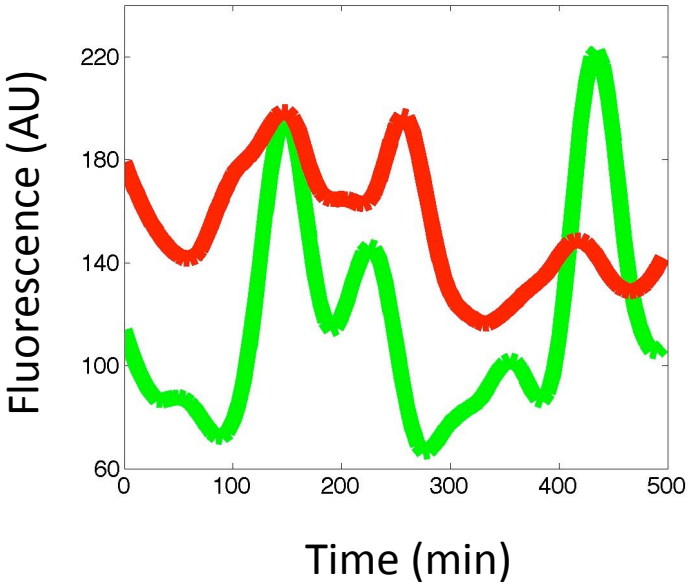
Cell-doubling events identified by the software. In spite of one false identification the software correctly tracks the division.



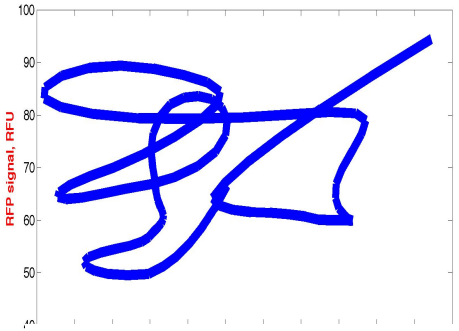
Coupling non-linear oscillators inside *E. coli*



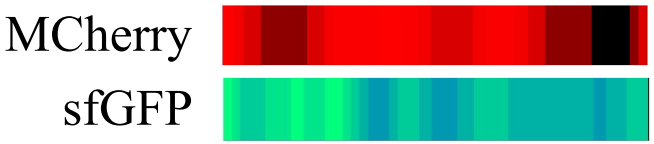




mCherry,

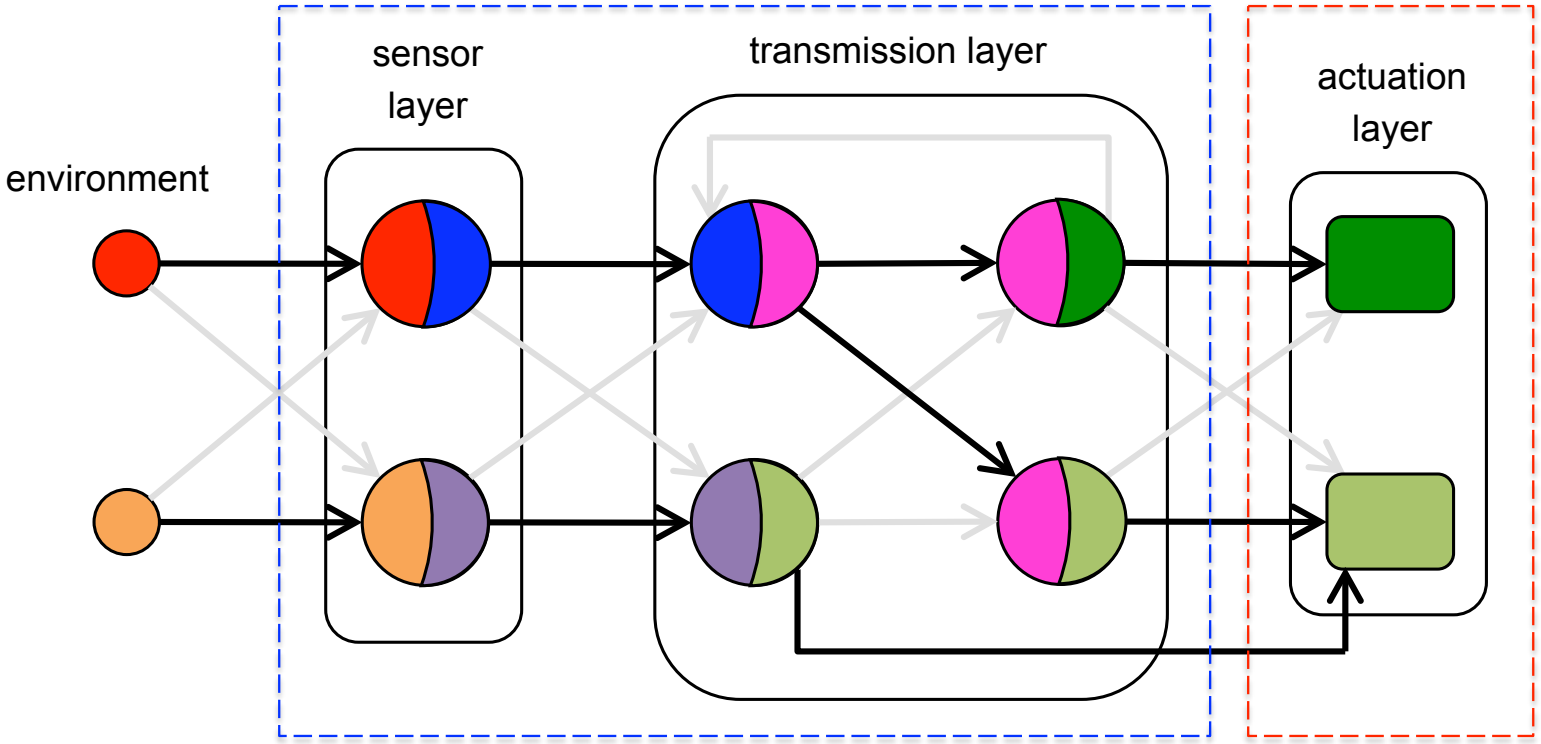


sfGFP,



Engineering RNA circuits

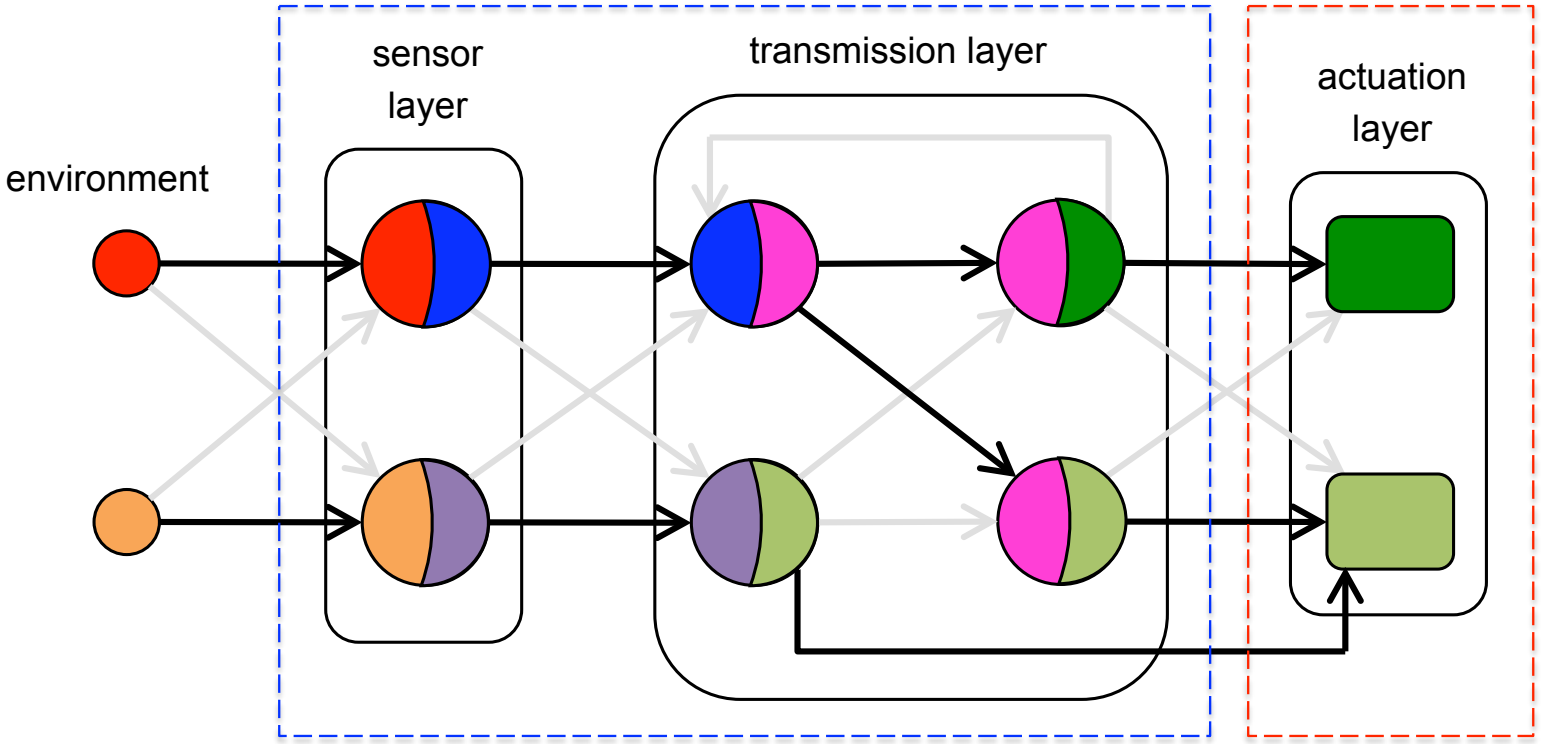
Synthetic RNA circuits



We need novel RNA functions:

Sensing, non-linearity, cascades and regulation.

Synthetic RNA circuits

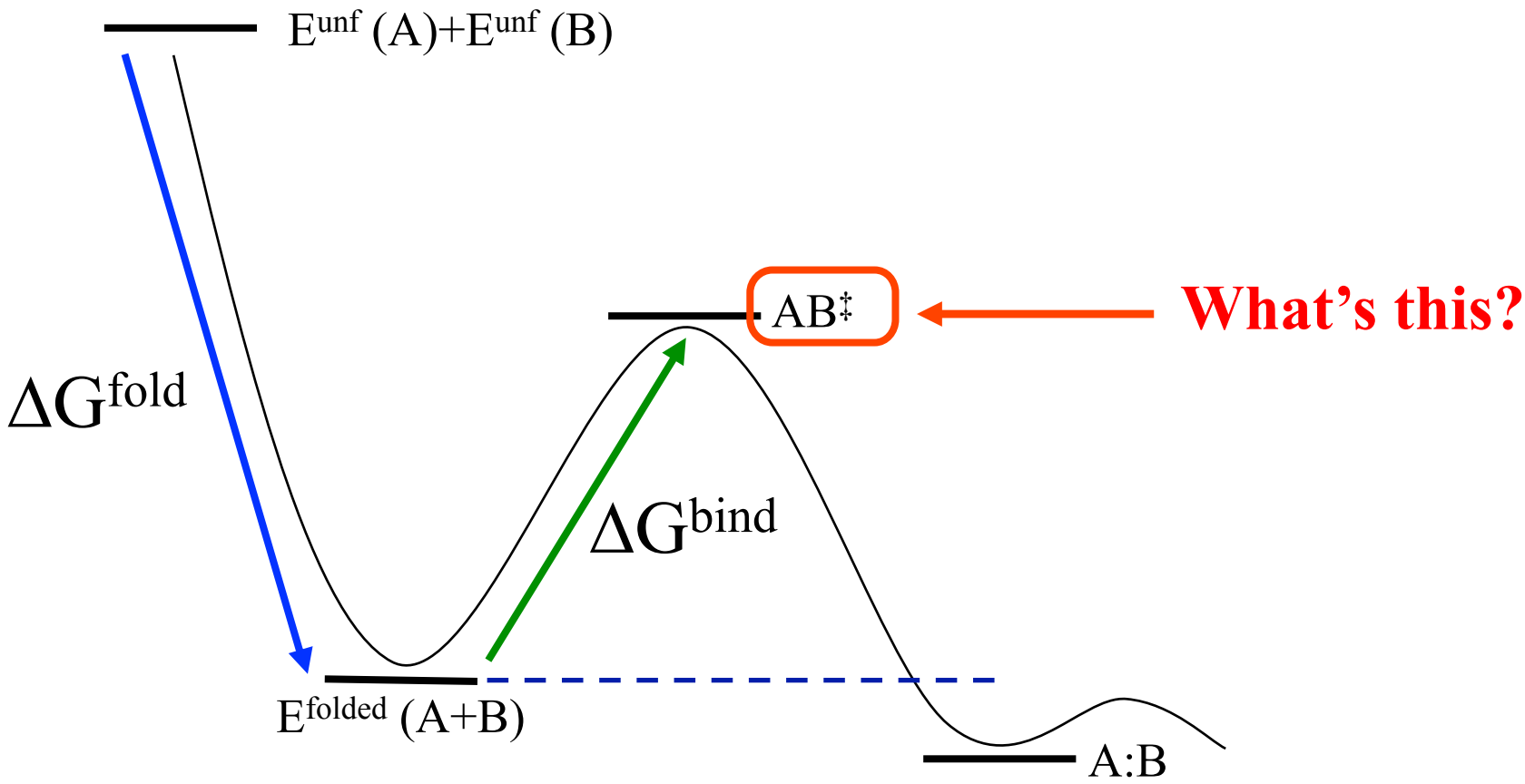


We need novel RNA functions:

Sensing, non-linearity, cascades and regulation.

De novo design by evolutionary computation

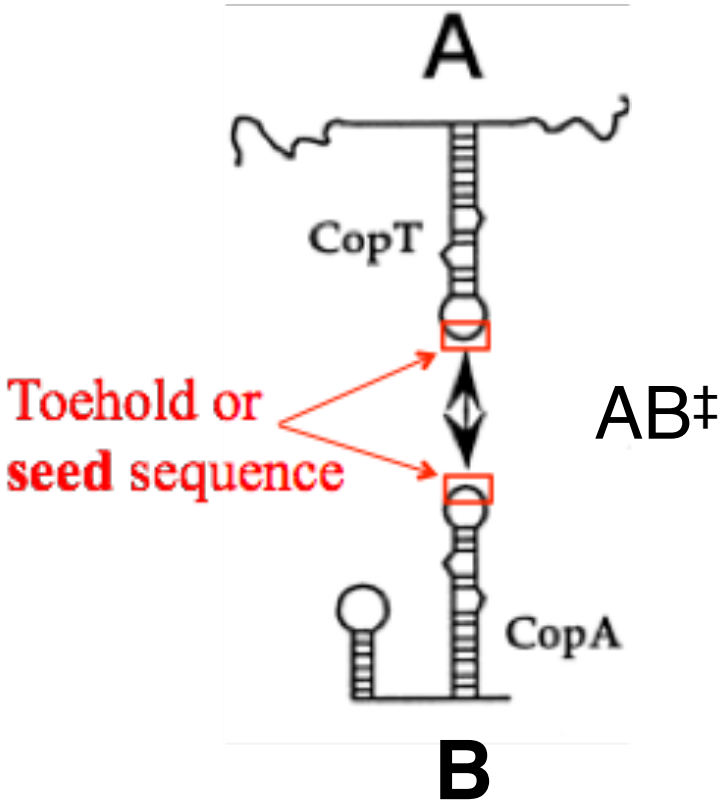
- Generate sequence diversity, select with a fitness/objective function and iterate.
- Improvement of fitness function by adding interactions with other molecules/systems

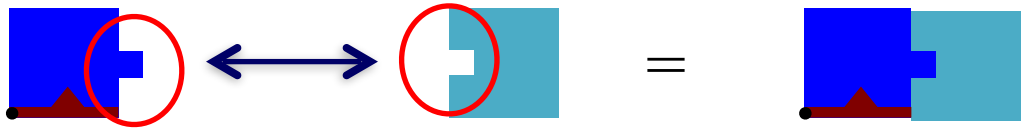


In vivo RNA-RNA interactions

- The active conformation for RNA-RNA interaction is assumed to require a ***kissing loop***

Minimisation of ΔG_{act}



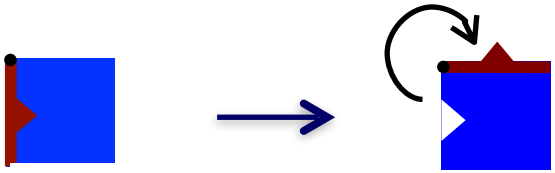


In vivo RNA-RNA interactions

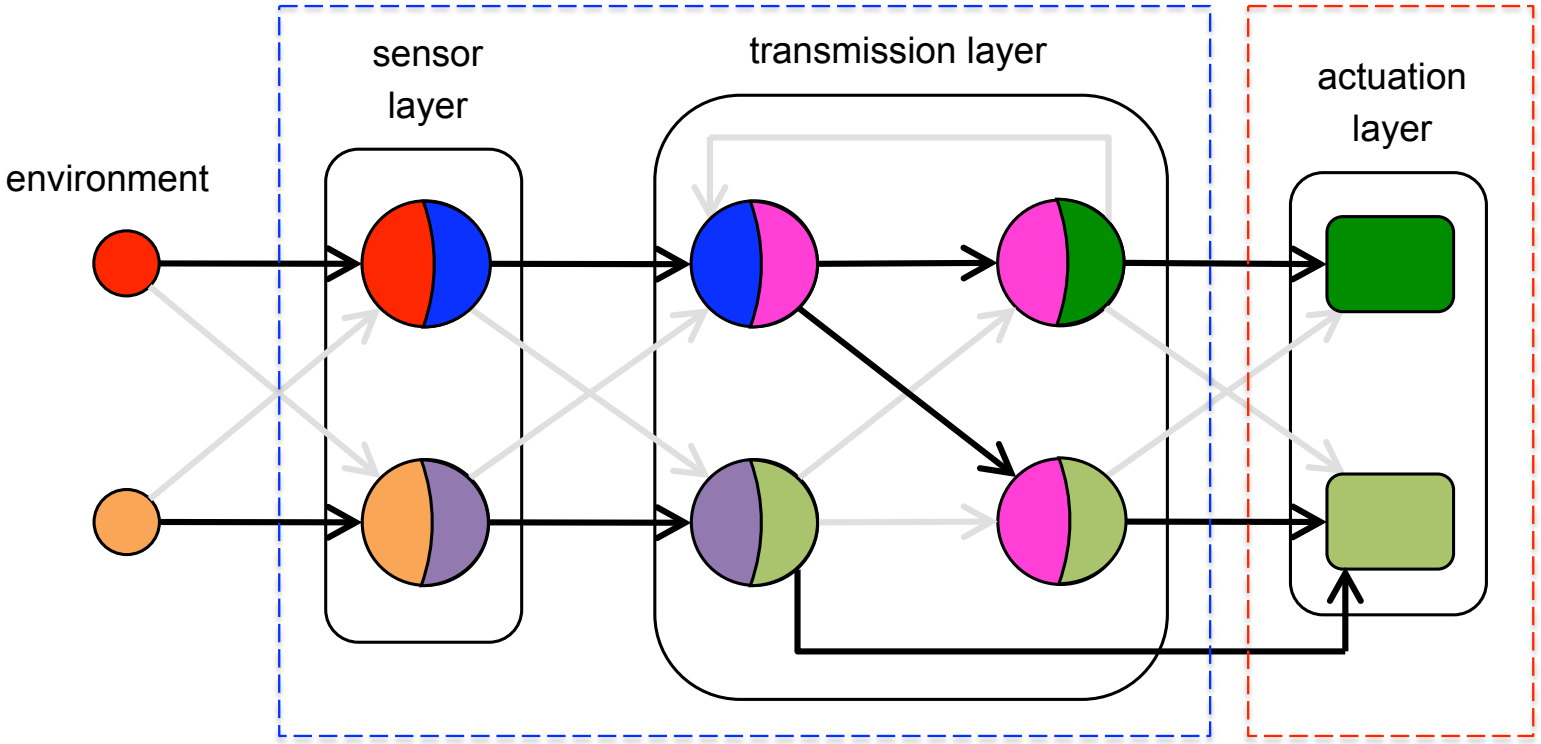
We will in/activate an RNA module by de/stabilising conformations

We assume that activity occurs in a precise conformation

Minimisation of ΔG_{form}



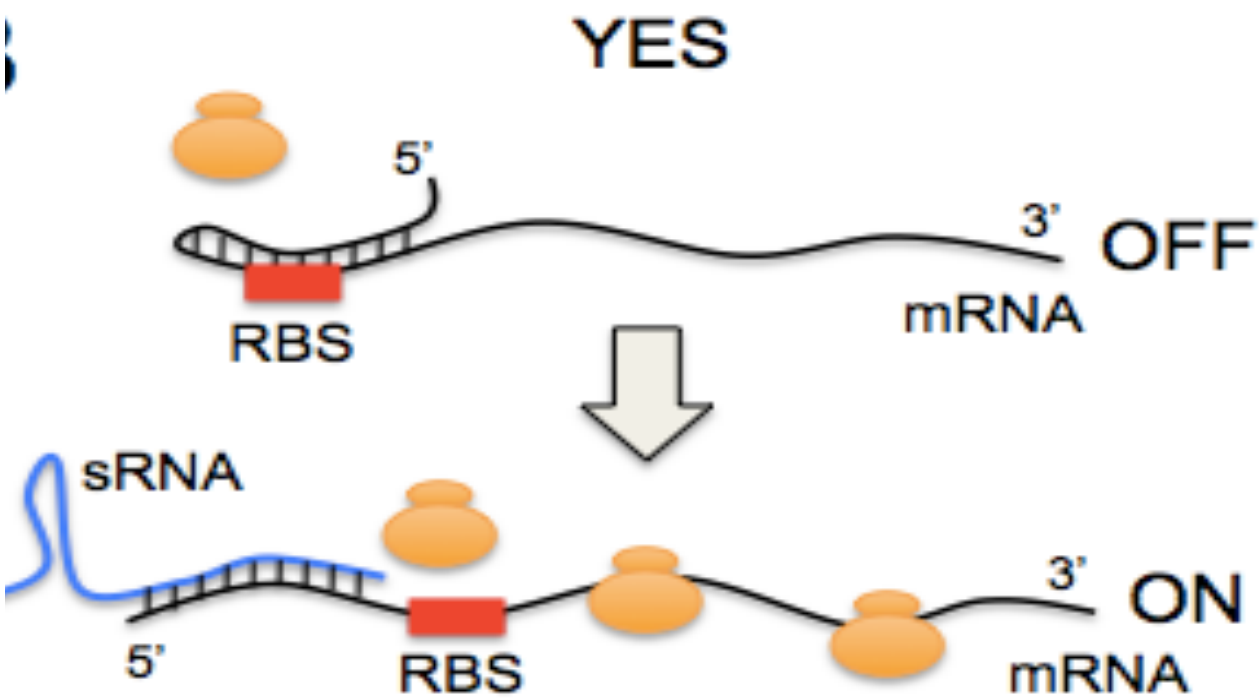
Synthetic RNA circuits



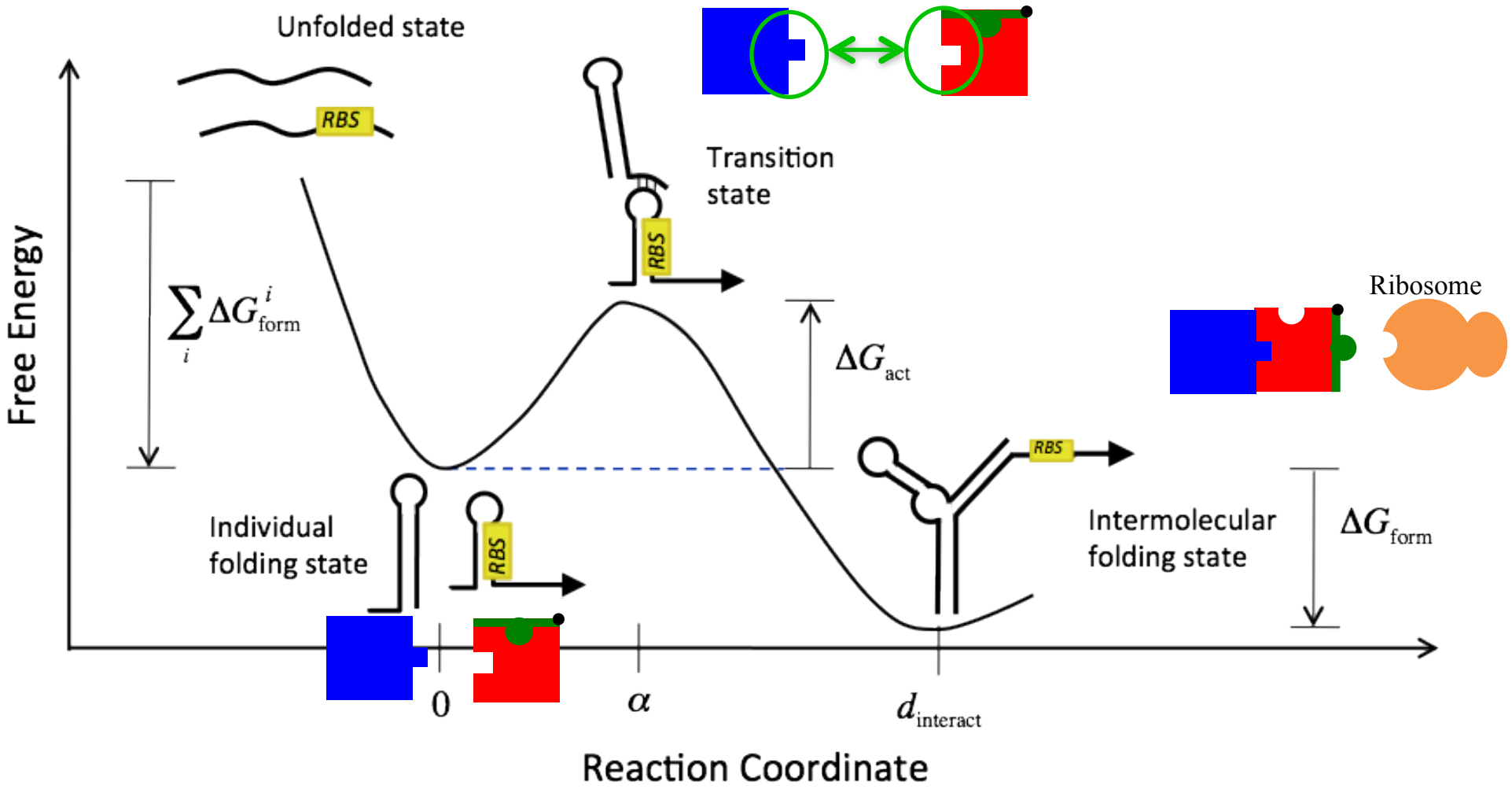
We need novel RNA functions:

Sensing, non-linearity, cascades and regulation.

System of 2 RNAs: riboregulation of translation

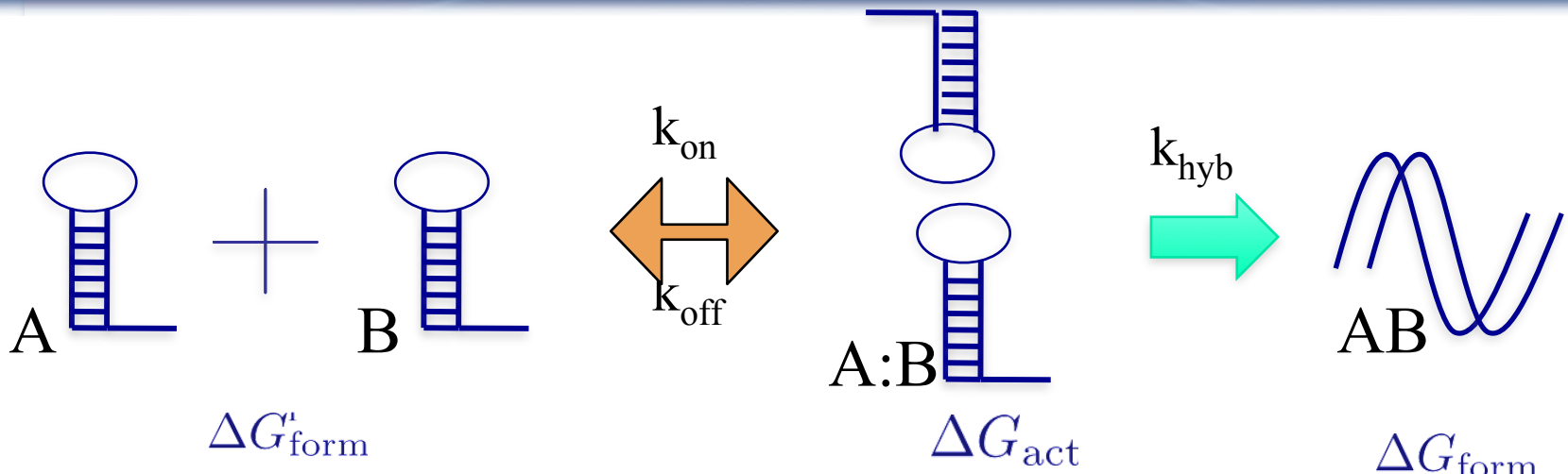


Engineering an sRNA heterodimer

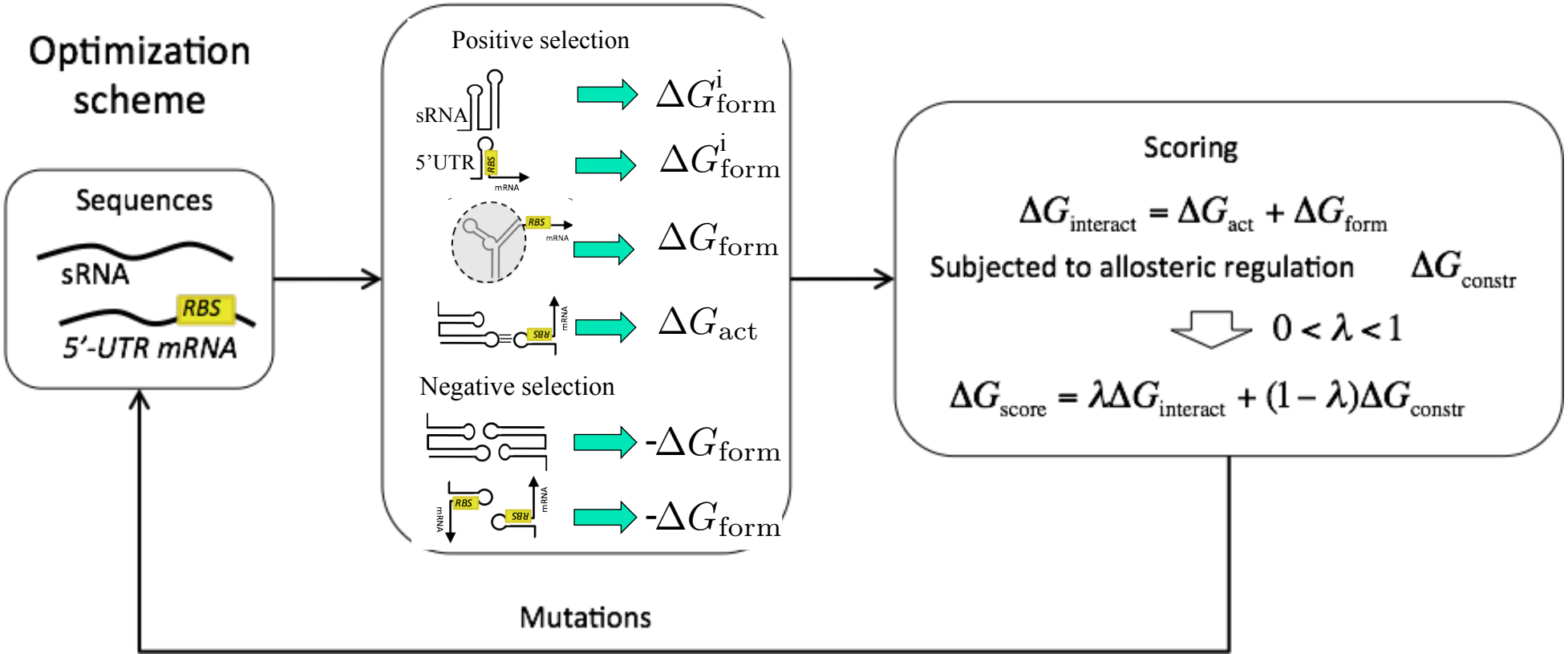


We use the number of inter-strand paired nucleotides as reaction coordinate

Engineering allostery with computational design



Fully automated design of nucleotide sequences



Scoring

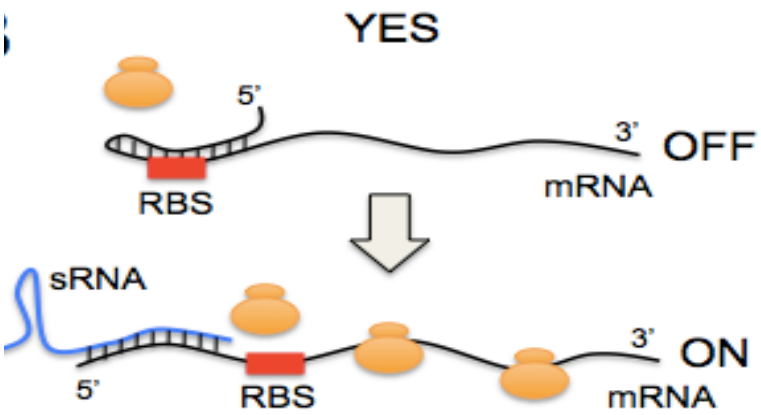
$$\Delta G_{score} = \lambda \sum_{i,j} \Delta G_{\text{interaction}\,ij} + (1-\lambda)\Delta G_{\text{constraints}}.$$

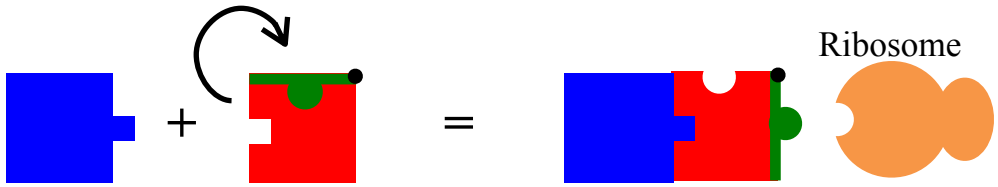
$$\Delta G_{\text{constraints}} = d_{\text{target, complex}} G_p . \qquad d = \text{Hamming distance to target structures}$$

$$\Delta G_{\text{interaction}\,ij} = \begin{cases} \Delta G_{ij} + \alpha_{ij} G_p, \text{ if no interaction is targeted (OFF)} \\ \min(0,\ \Delta G_{\text{sat}} - \Delta G_{ij}) + G_p \max(0, \alpha_{\text{sat}} - \alpha_{ij}), \text{ if interaction is targeted (ON)} \end{cases}$$

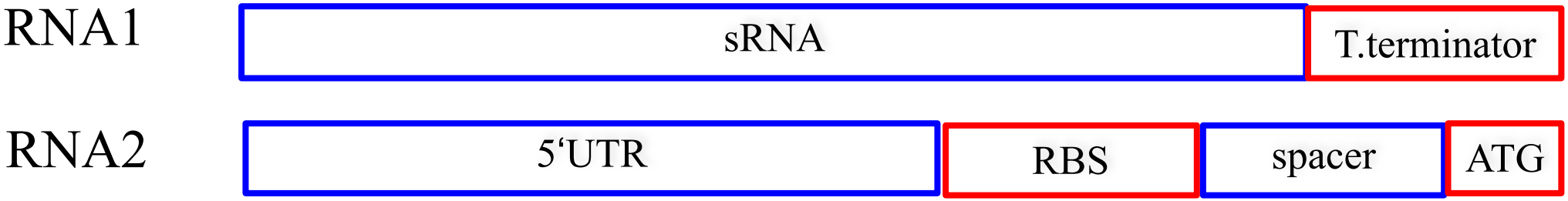
where ΔG_{sat}=-15 Kcal/mol and α_{sat}=6 (saturation levels). G_p=-1.28 Kcal/mol

System of 2 RNAs: riboregulation of translation



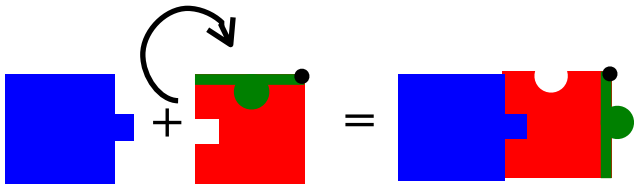


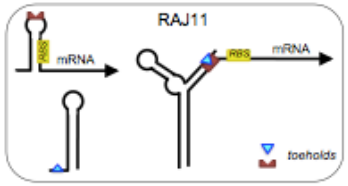
Sequence space search for a riboregulator:



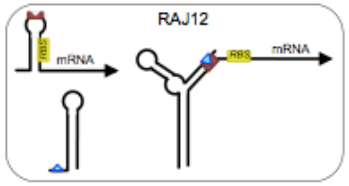
Fixed sequences: RBS, transcription terminator

Variable sequences: Nucleotide changes preserving secondary structure

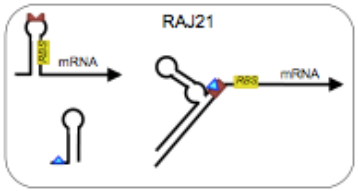




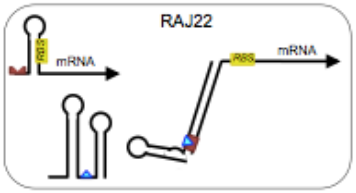
RAJ11



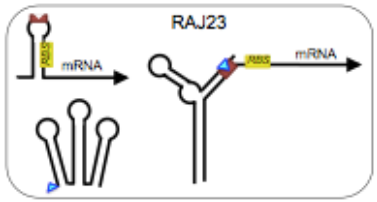
RAJ12



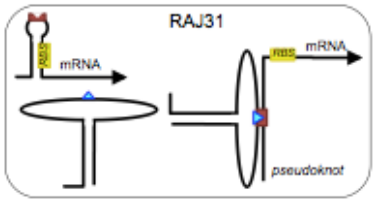
RAJ21



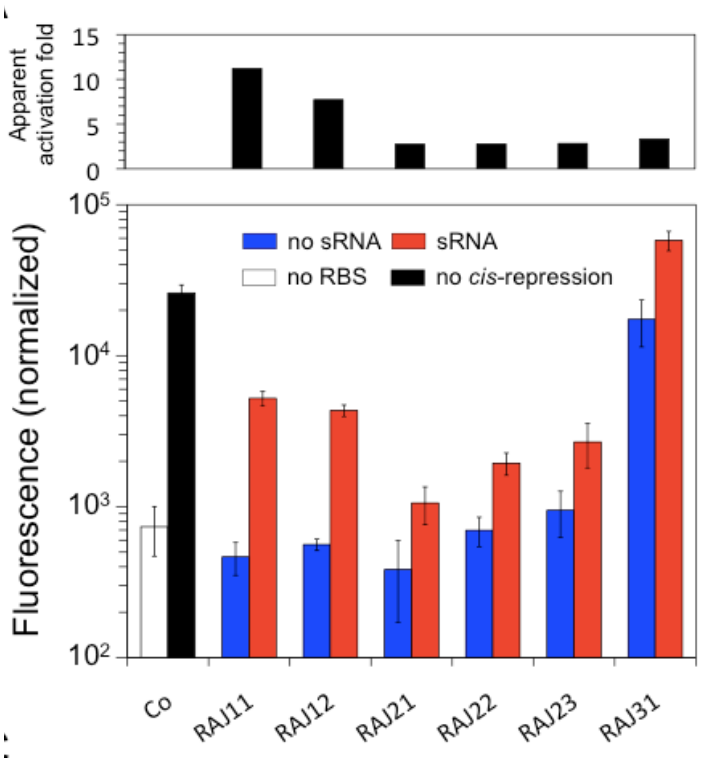
RAJ22

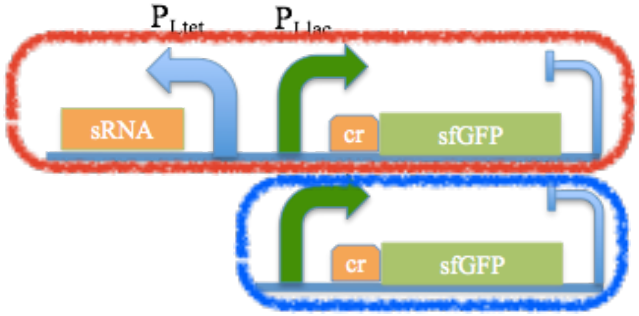


RAJ23



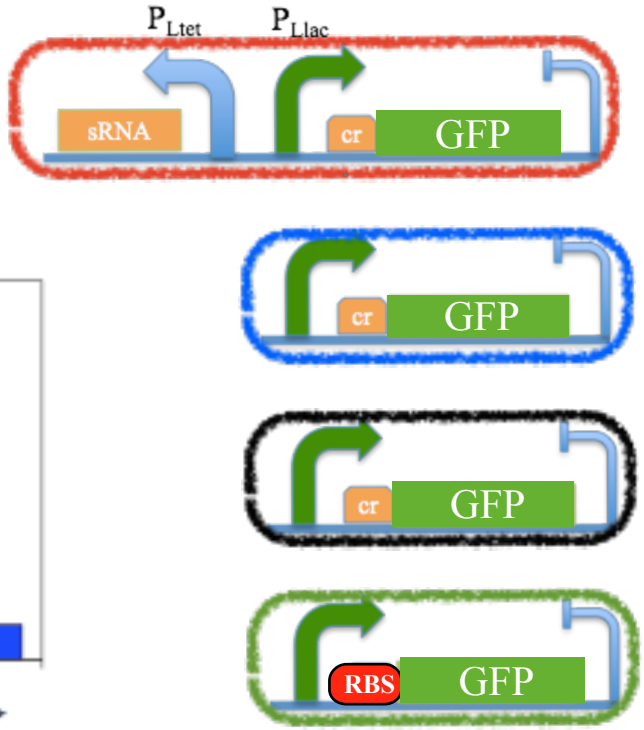
RAJ31

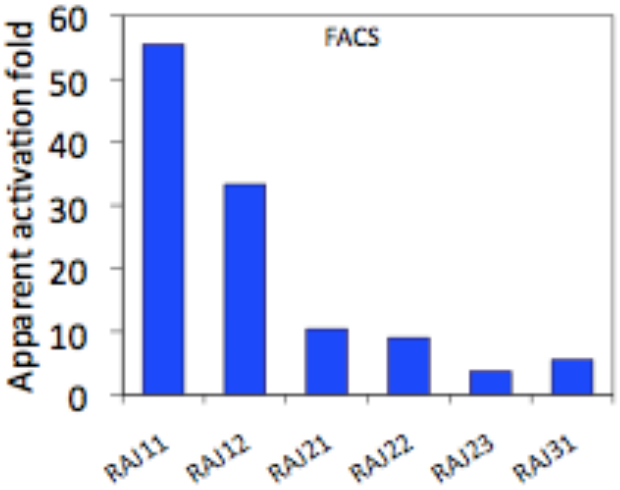


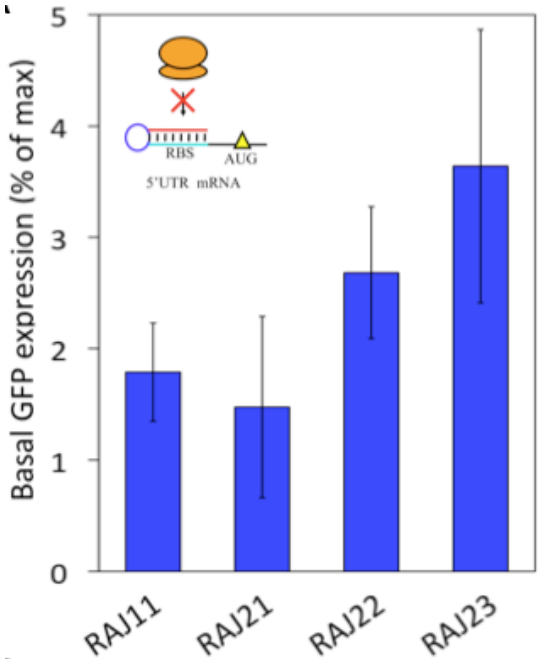


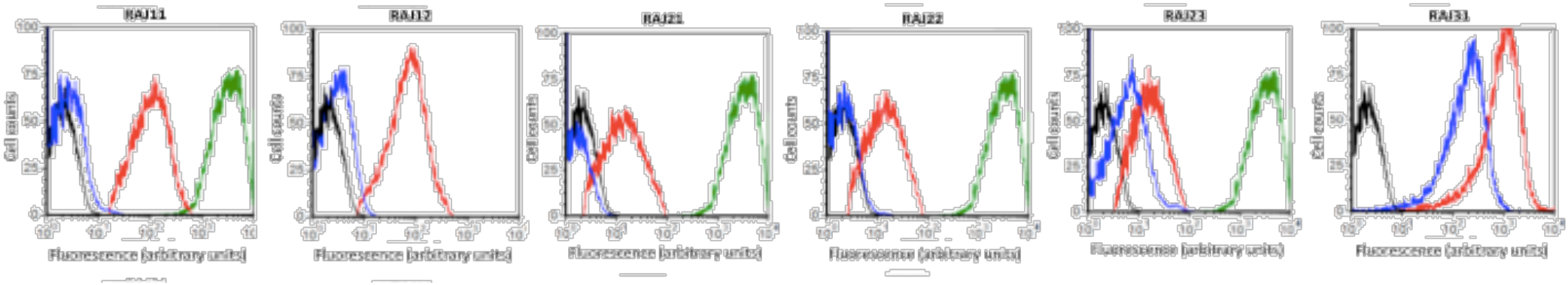
Rodrigo et al. PNAS 2012

Single-cell characterization of riboregulators



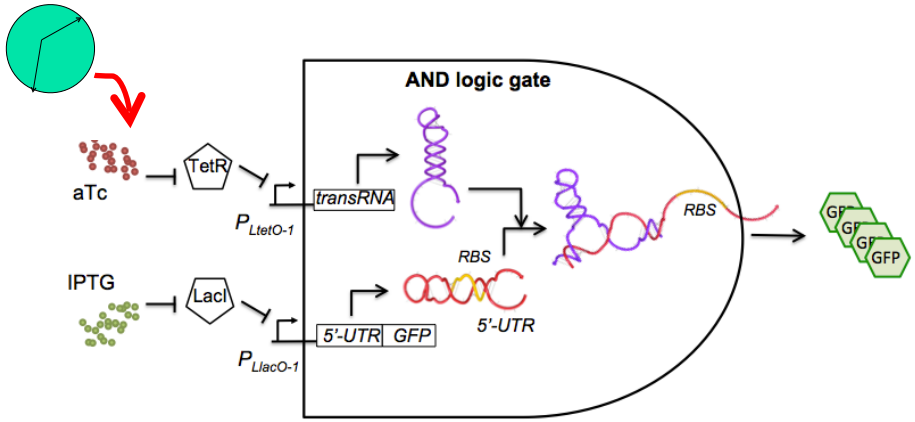


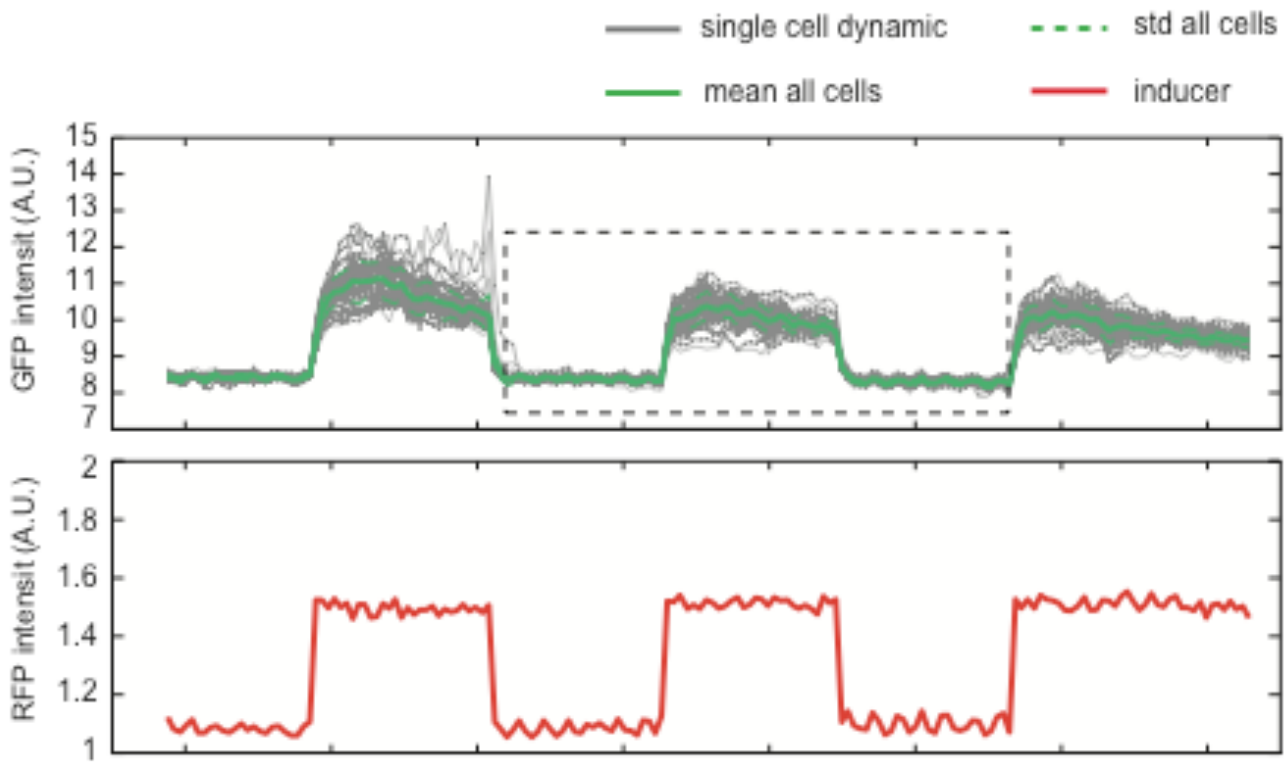




Riboregulator *in vivo* dynamics

Growing cells in single layers with microfluidics





Model riboregulator kinetics

$$P_m(LacI,IPTG) = P_m{}^0 \frac{1 + 1/f_{lac}\left(\frac{LacI}{K_{lac}(1 + IPTG/K_{IPTG})}\right)^{n_{lac}}}{1 + \left(\frac{LacI}{K_{lac}(1 + IPTG/K_{IPTG})}\right)^{n_{lac}}}$$

$$P_s(TetR,aTc) = P_s{}^0 \frac{1 + 1/f_{tet}\left(\frac{TetR}{K_{tet}(1 + aTc/K_{aTc})}\right)^{n_{tet}}}{1 + \left(\frac{TetR}{K_{tet}(1 + aTc/K_{aTc})}\right)^{n_{tet}}}$$

$$\frac{d}{dt} mRNA = CP_m(LacI,IPTG) - (\mu + \delta_m) mRNA$$

$$\frac{d}{dt} sRNA = CP_s(TetR,aTc) - (\mu + \delta_s) sRNA$$

$$\frac{d}{dt} GFP = \left[r_0 mRNA_{free} + r_1 sRNA :: mRNA\right] \frac{m}{m + \mu + \delta_g} - (\mu + \delta_g) GFP$$

Riboregulator *in vivo* dynamics by forcing

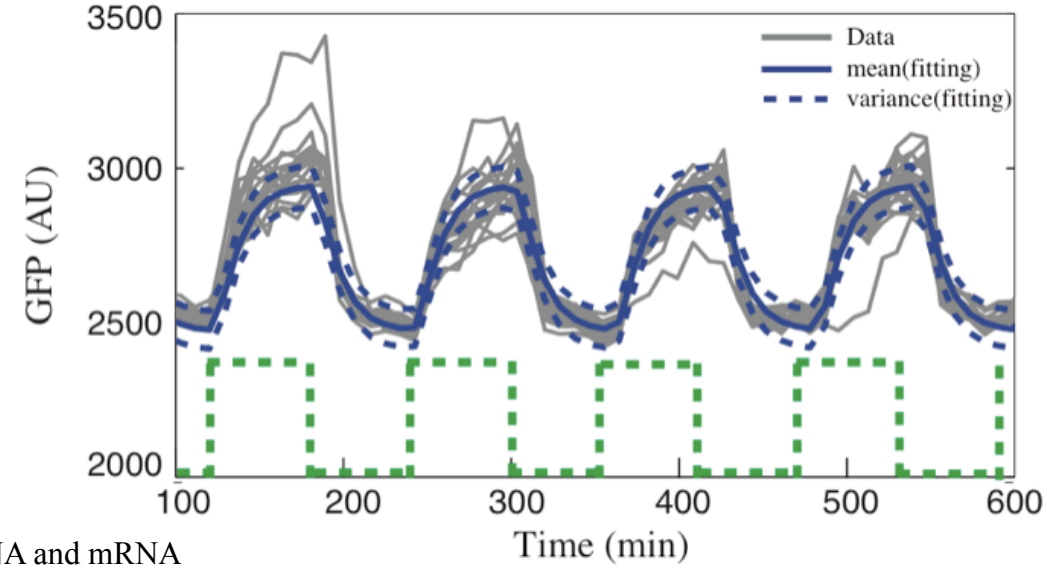
Mass action kinetics model trained with data from aTc forcing at 60/60min cycles

$$\phi \xrightarrow{\alpha_{tsx}} mRNA$$
$$mRNA \xrightarrow{\delta_m} \phi$$
$$\phi \xrightarrow{\alpha_{Tet}} sRNA$$
$$sRNA \xrightarrow{\delta_s} \phi$$

mRNA birth and death

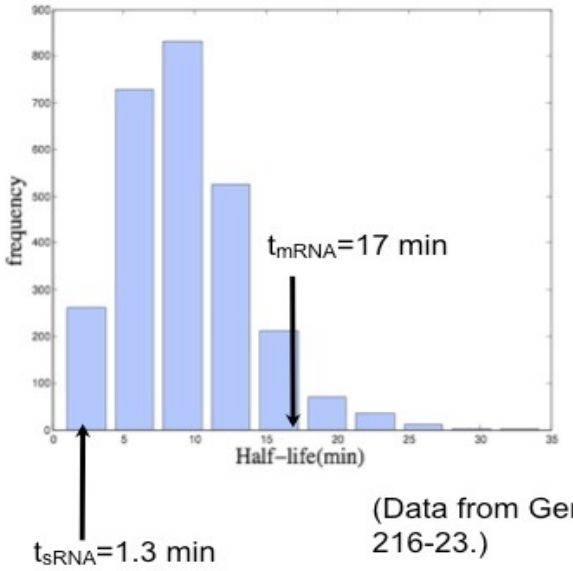
$$sRNA + mRNA \underset{k_{off}}{\overset{k_{on}}{\leftrightarrow}} sRNA:mRNA_{int}$$
$$sRNA:mRNA_{int} \xrightarrow{\delta_{im}} \phi$$
$$sRNA:mRNA_{int} \xrightarrow{k_{bsb}} sRNA:mRNA_{stable}$$
$$sRNA:mRNA_{stable} \xrightarrow{\delta_C} \phi$$
$$sRNA:mRNA_{int} \xrightarrow{\beta} sRNA:mRNA_{int} + GFP_{unmature}$$

sRNA and mRNA Interaction



$$sRNA:mRNA_{stable} \xrightarrow{f_S \cdot \beta} sRNA:mRNA_{stable} + GFP_{unmature}$$
$$GFP_{unmature} \xrightarrow{\delta g} \phi$$
$$GFP_{unmature} + ClpX \underset{K_Z}{\leftrightarrow} GFP:ClpX \xrightarrow{\nu_Z} \phi$$
$$GFP_{unmature} \xrightarrow{\gamma} GFP_{mature}$$
$$GFP_{mature} \xrightarrow{\delta g} \phi$$
$$GFP_{mature} + ClpX \underset{K_Z}{\leftrightarrow} GFP:ClpX \xrightarrow{\nu_Z} \phi$$

GFP translation



Comparison with In vitro globa analysis of mRNA decay in E. Coli

(Data from Genome Res. 2003 Feb;13(2): 216-23.)

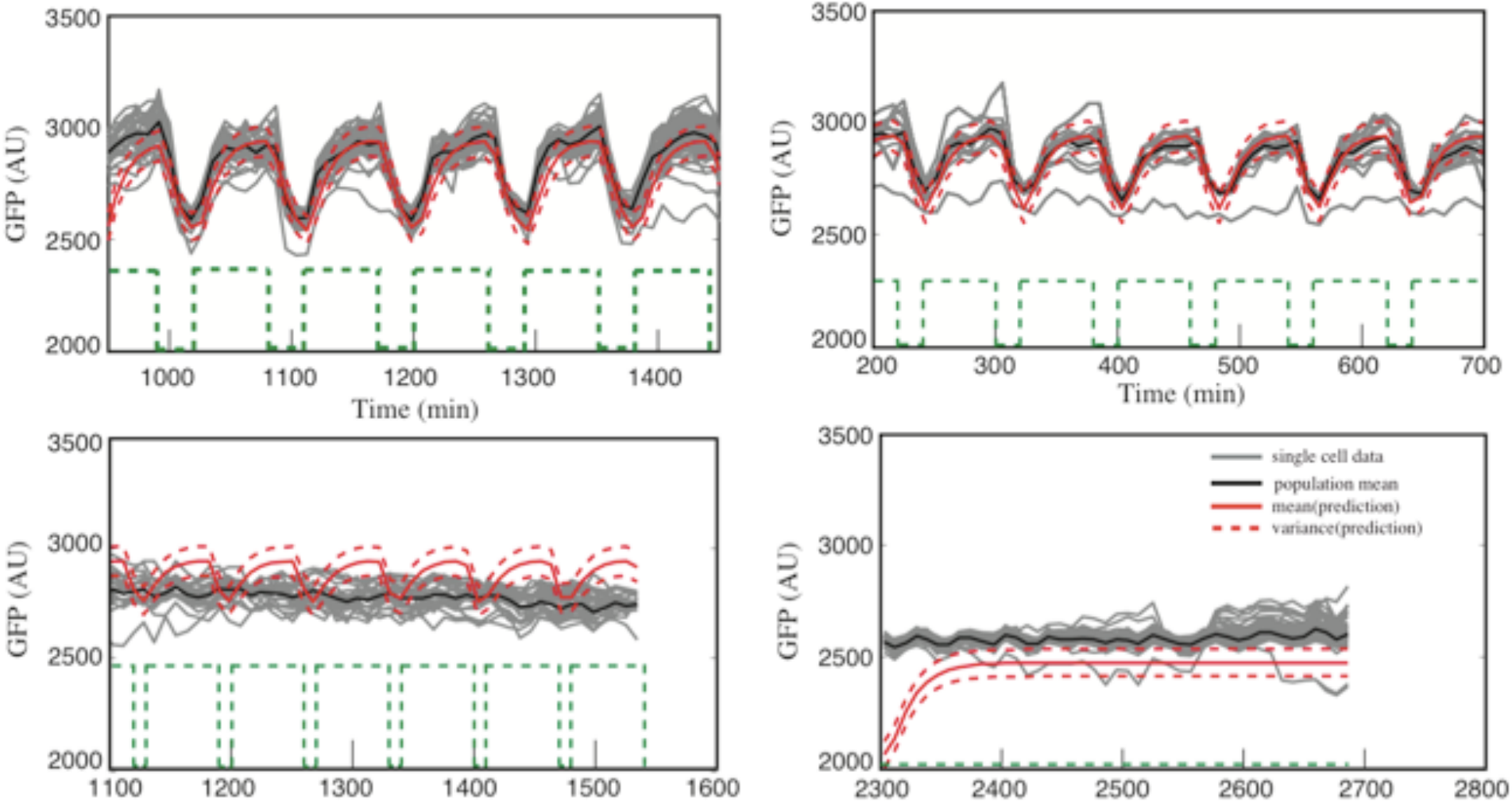
Characterization of AND gate by forcing

$$\text{sRNA} + \text{mRNA} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \text{sRNA:mRNA}_{intermediate} \xrightarrow{k_{hyb}} \text{sRNA:mRNA}_{stable}$$

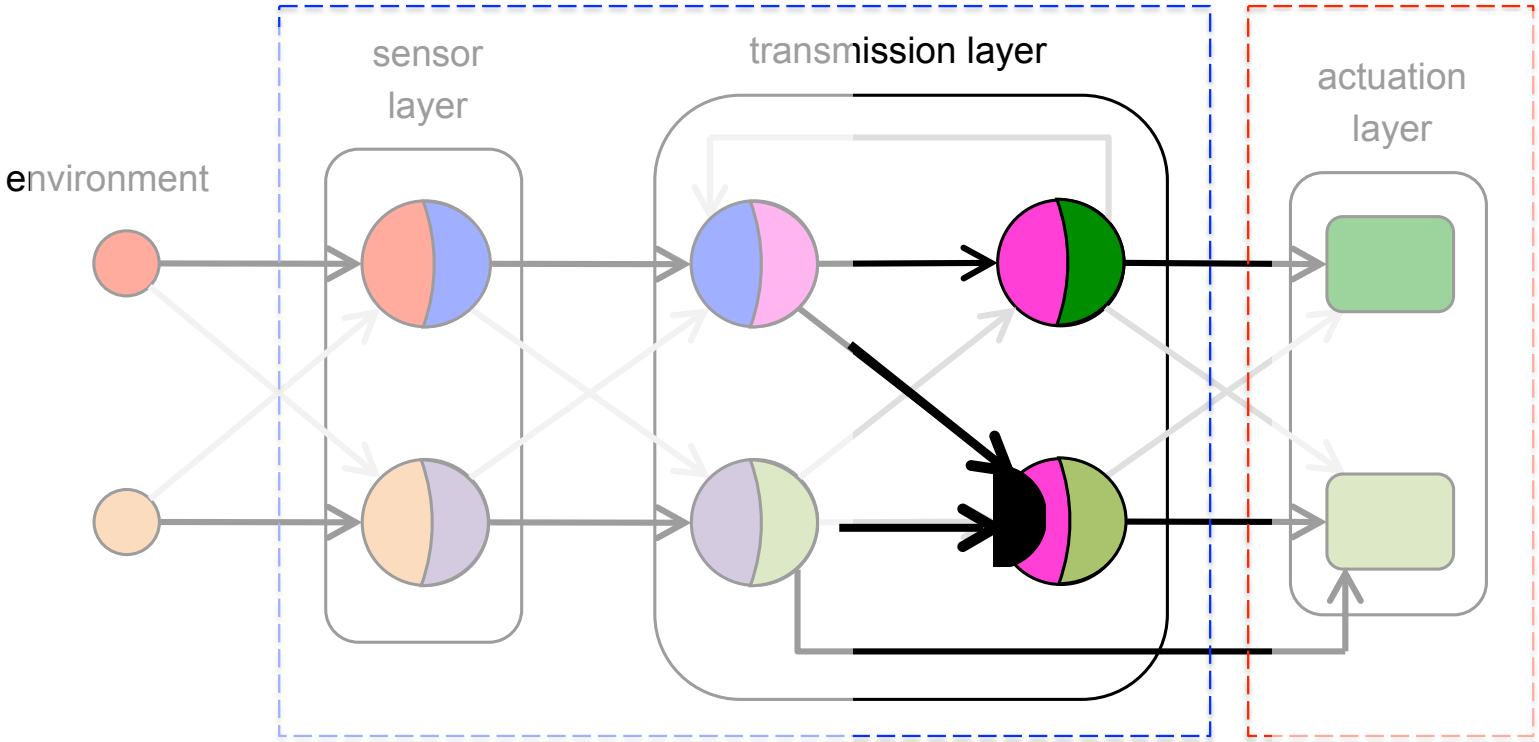
$$\text{sRNA:mRNA}_{stable} \xrightarrow{\text{Translation}} \text{GFP}$$

Activation fold of the sRNA in translation rate: $\frac{R_{11}}{R_{10}} = 7.4$

Prediction of the dynamics upon forcing with aTc



Synthetic RNA circuits

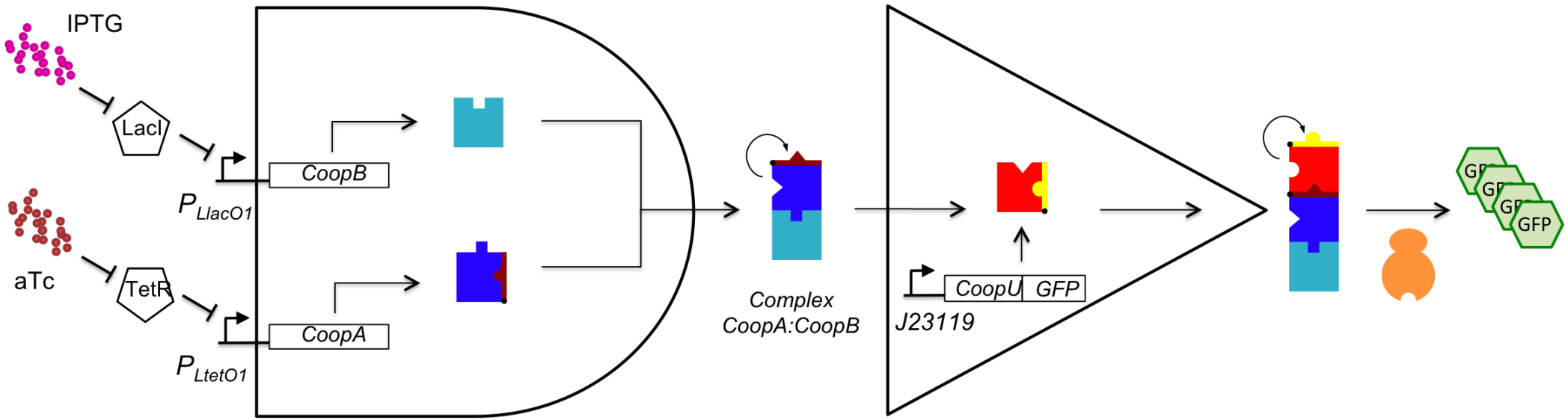


We need novel RNA functions:

Sensing, non-linearity, cascades and regulation.

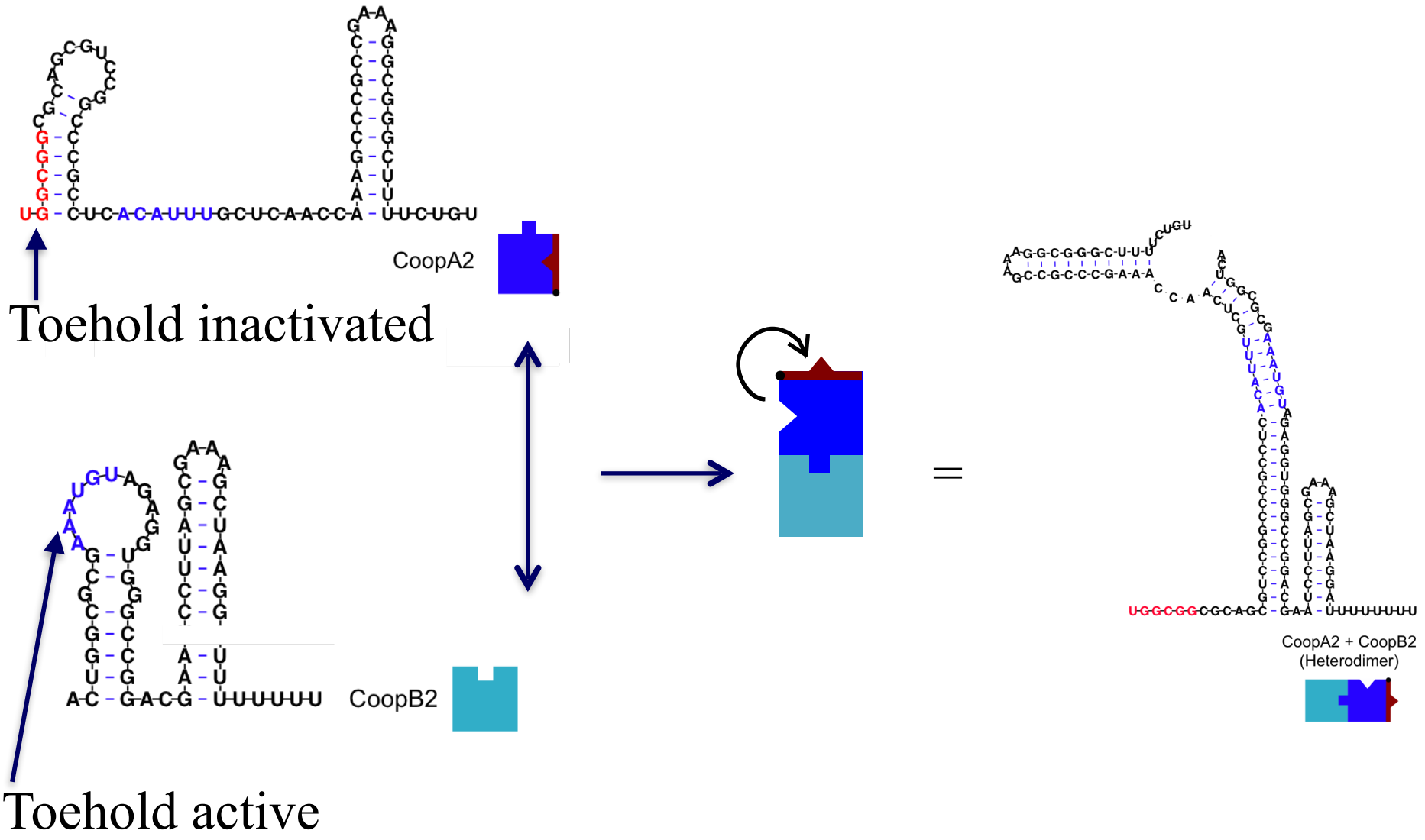
Aim

- To *de novo* engineer circuits with synergistic interactions

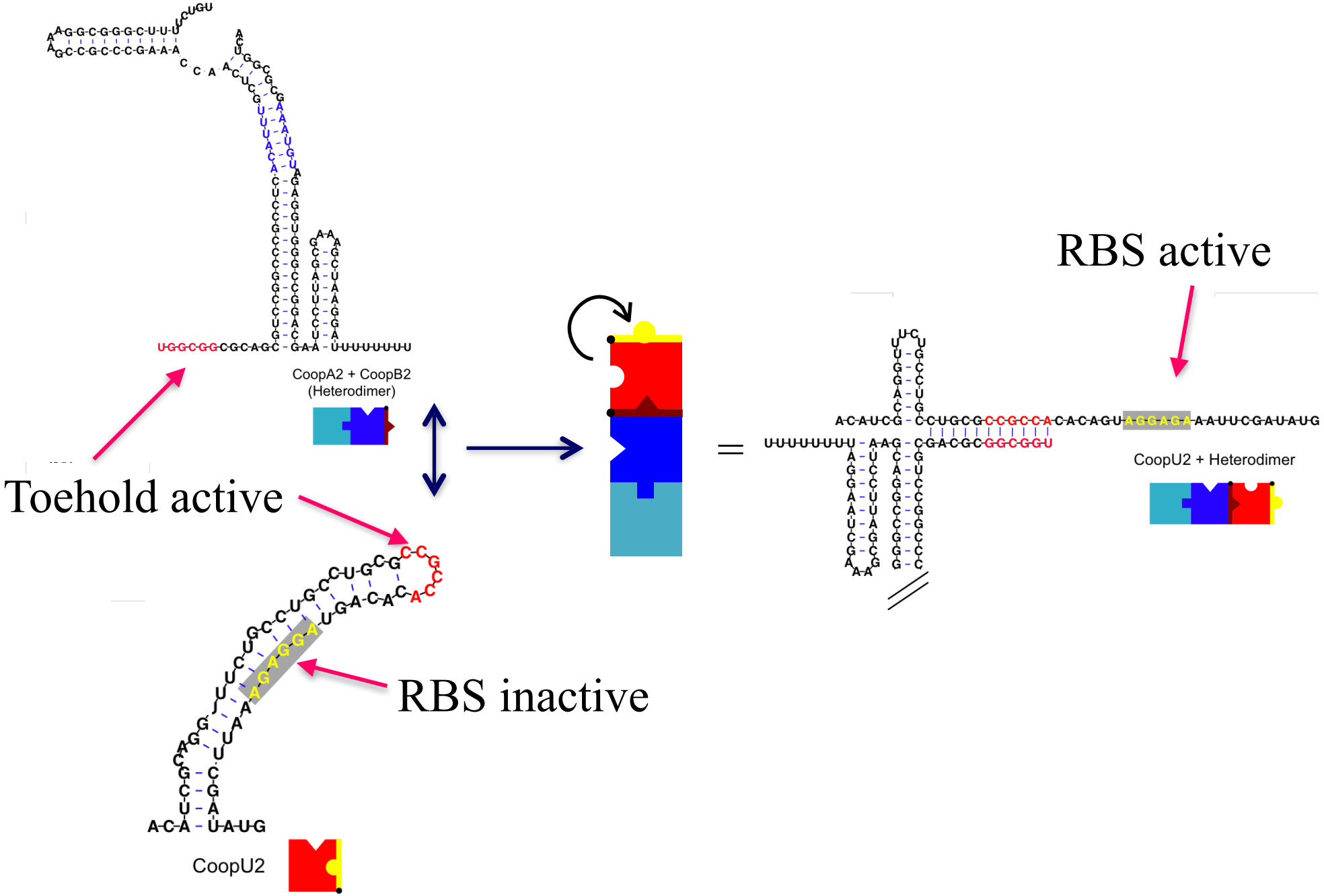


Coll. Dr. Rodrigo and Dr. Daros (IBMCP, Valencia)

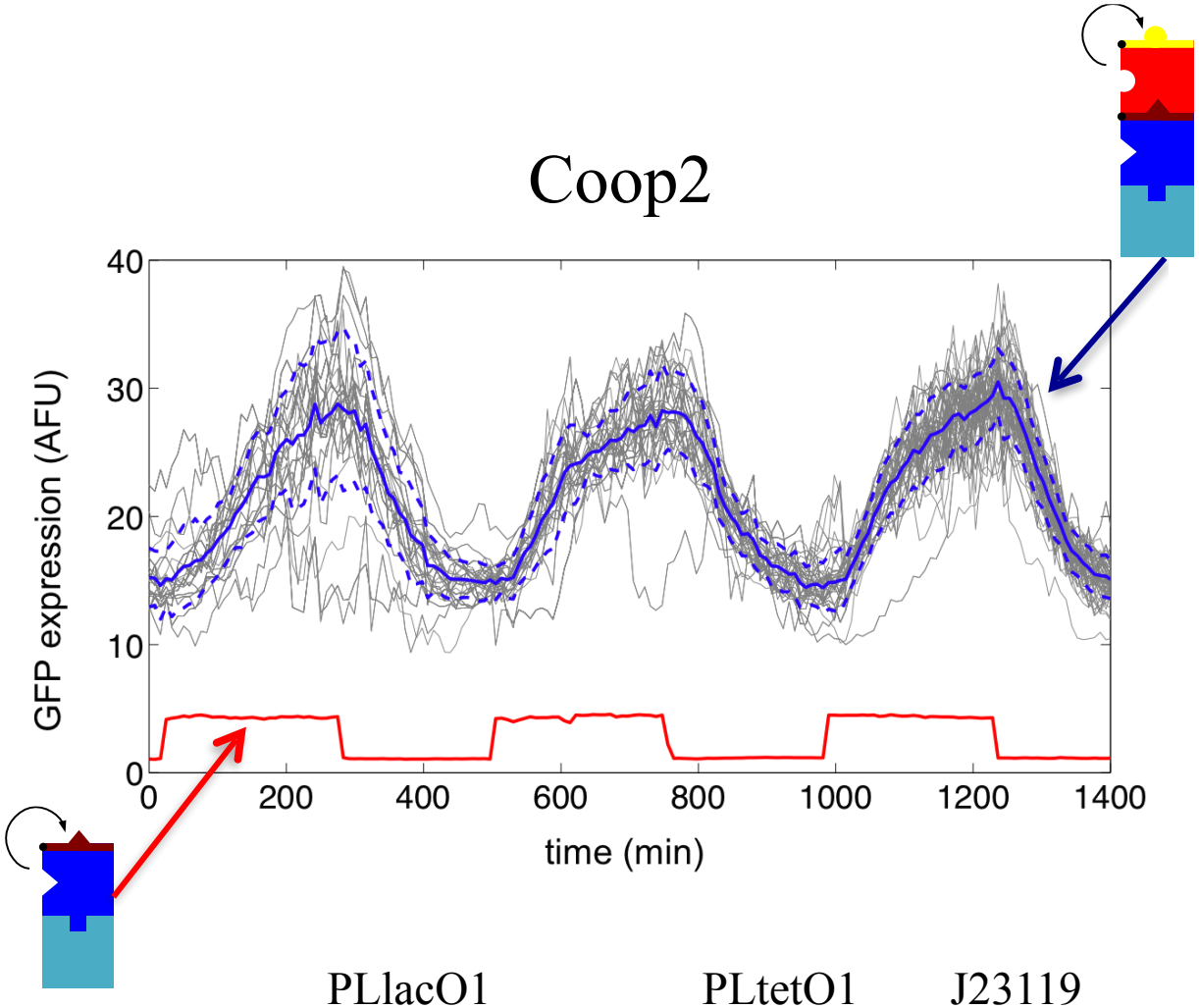
Engineering an sRNA heterodimer

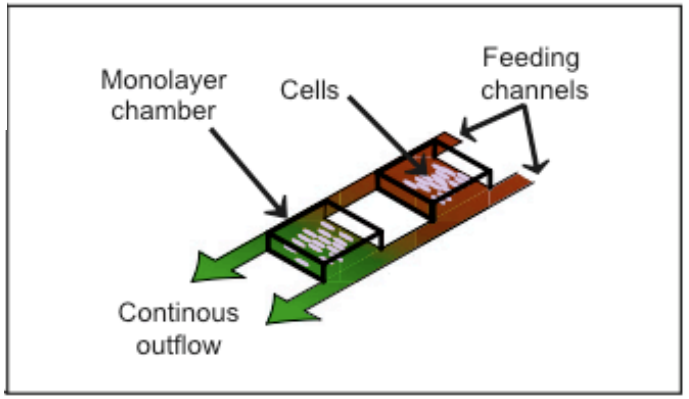


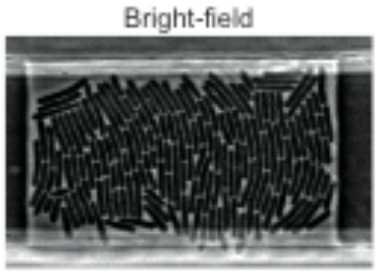
Engineering an sRNA heterodimer

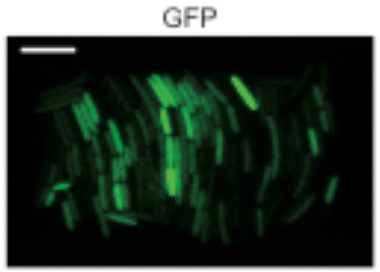


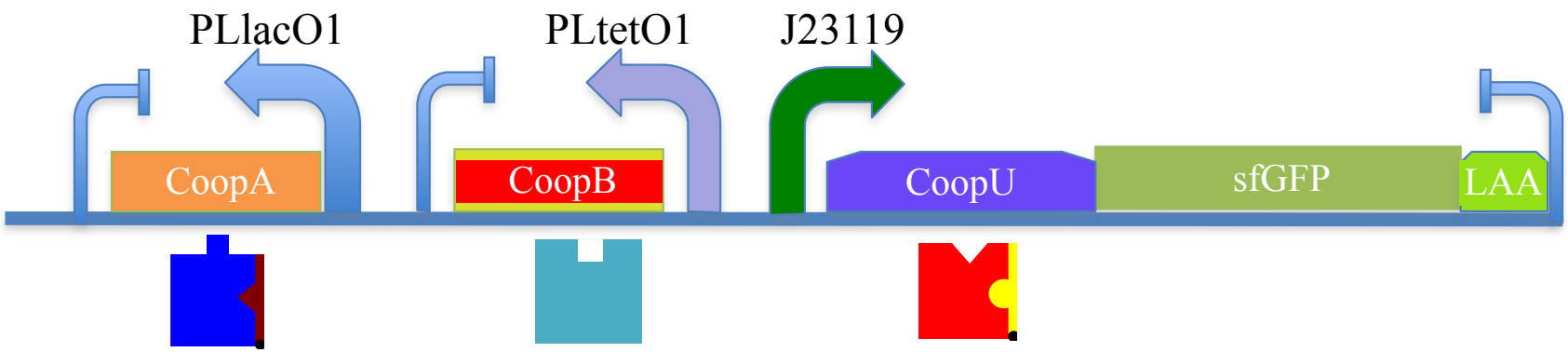
Characterization of Coop2 system



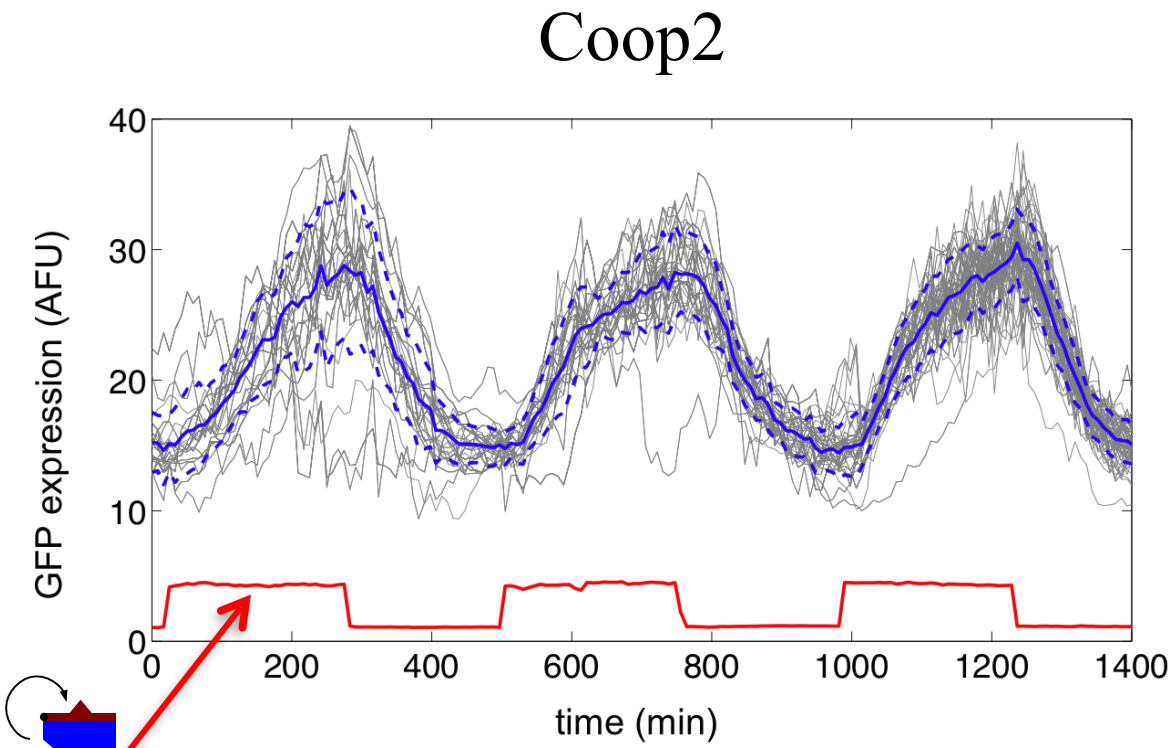


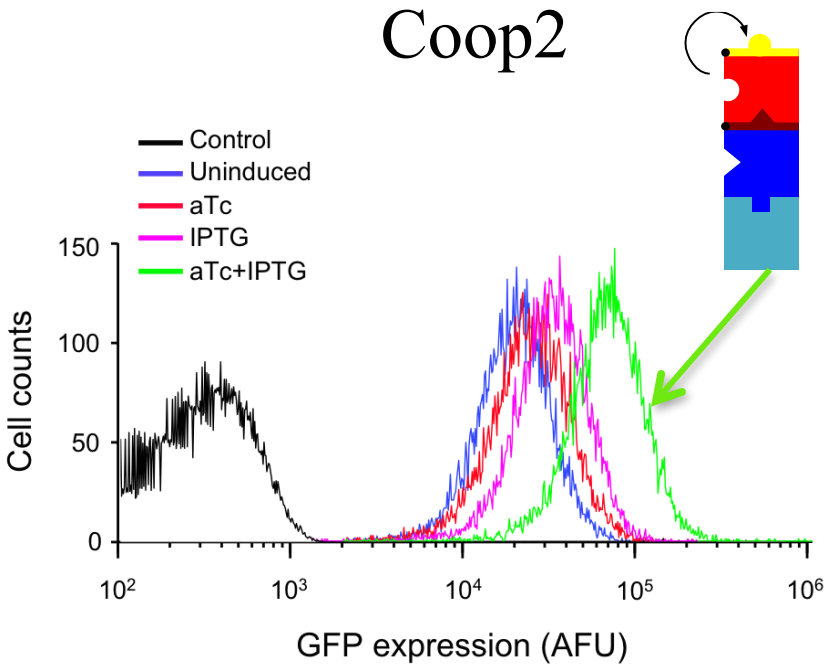


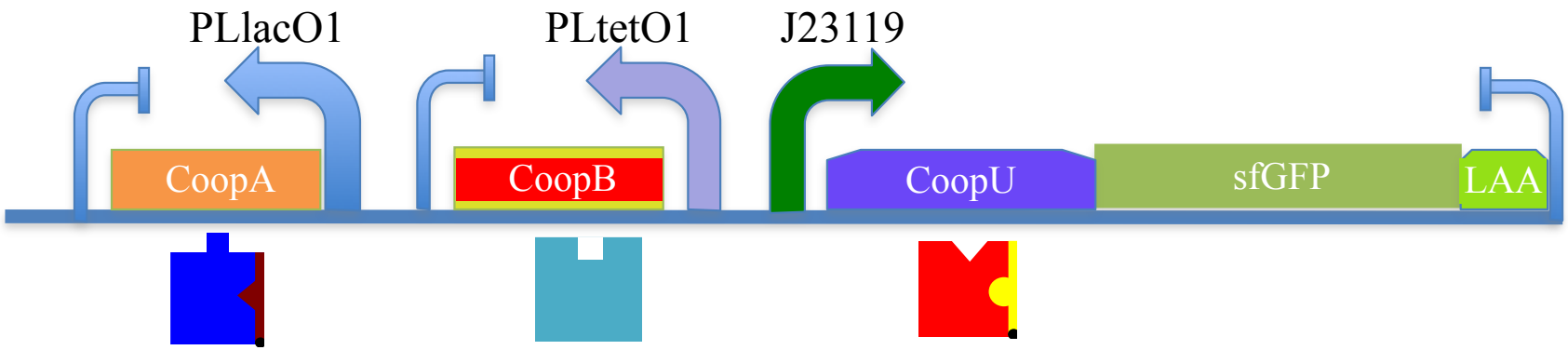




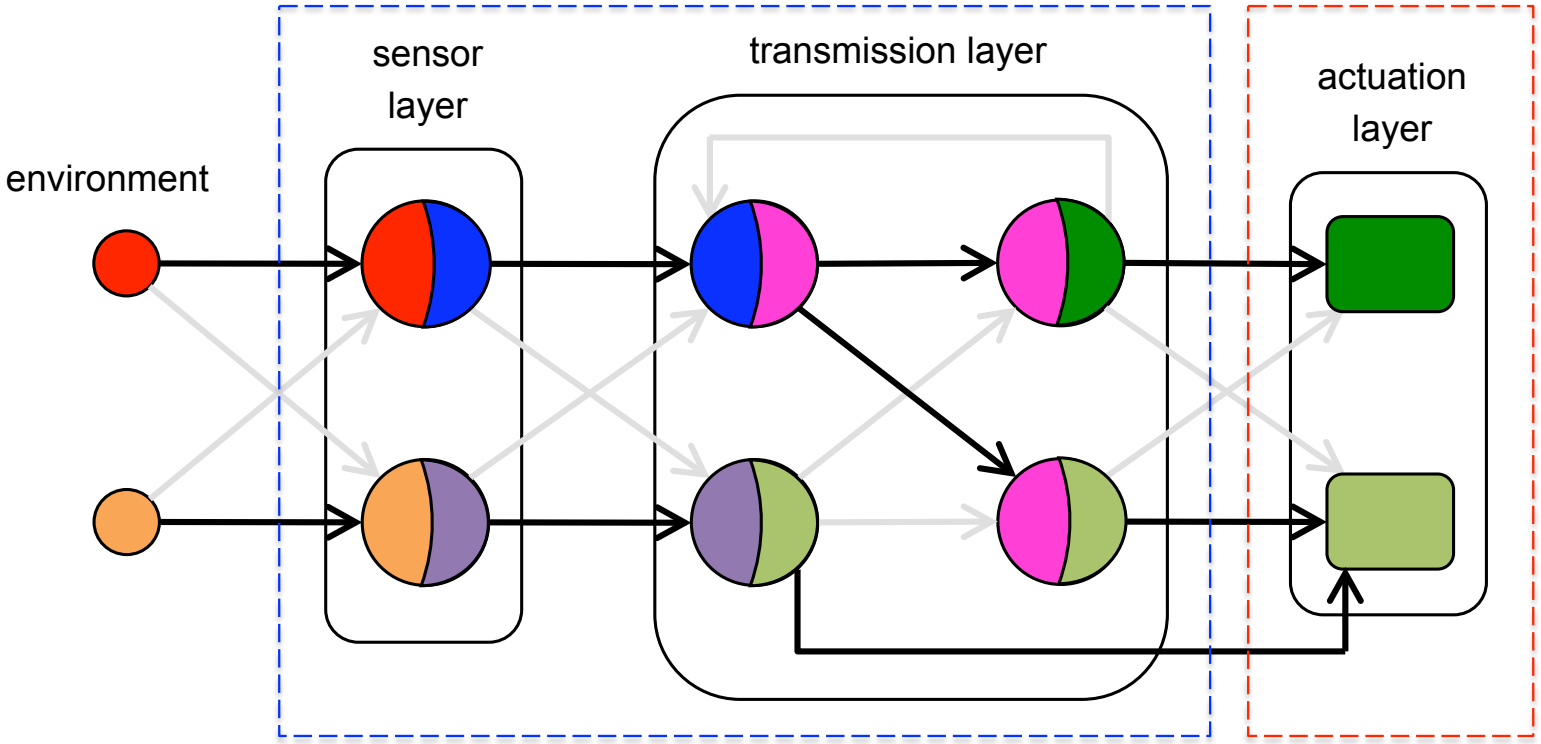
Characterization of Coop2 system







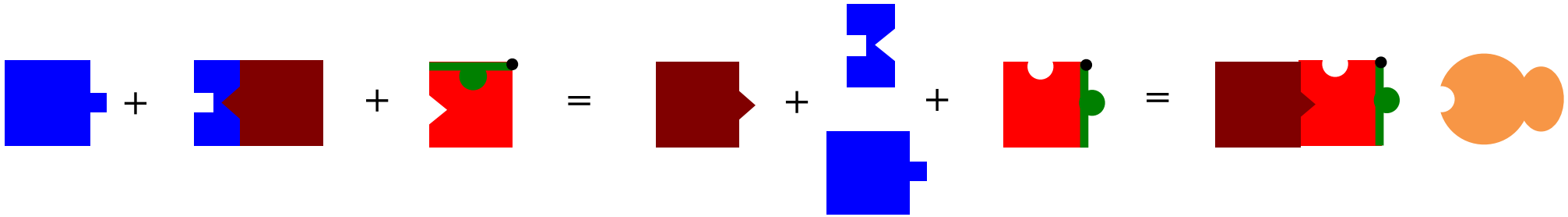
Synthetic RNA circuits

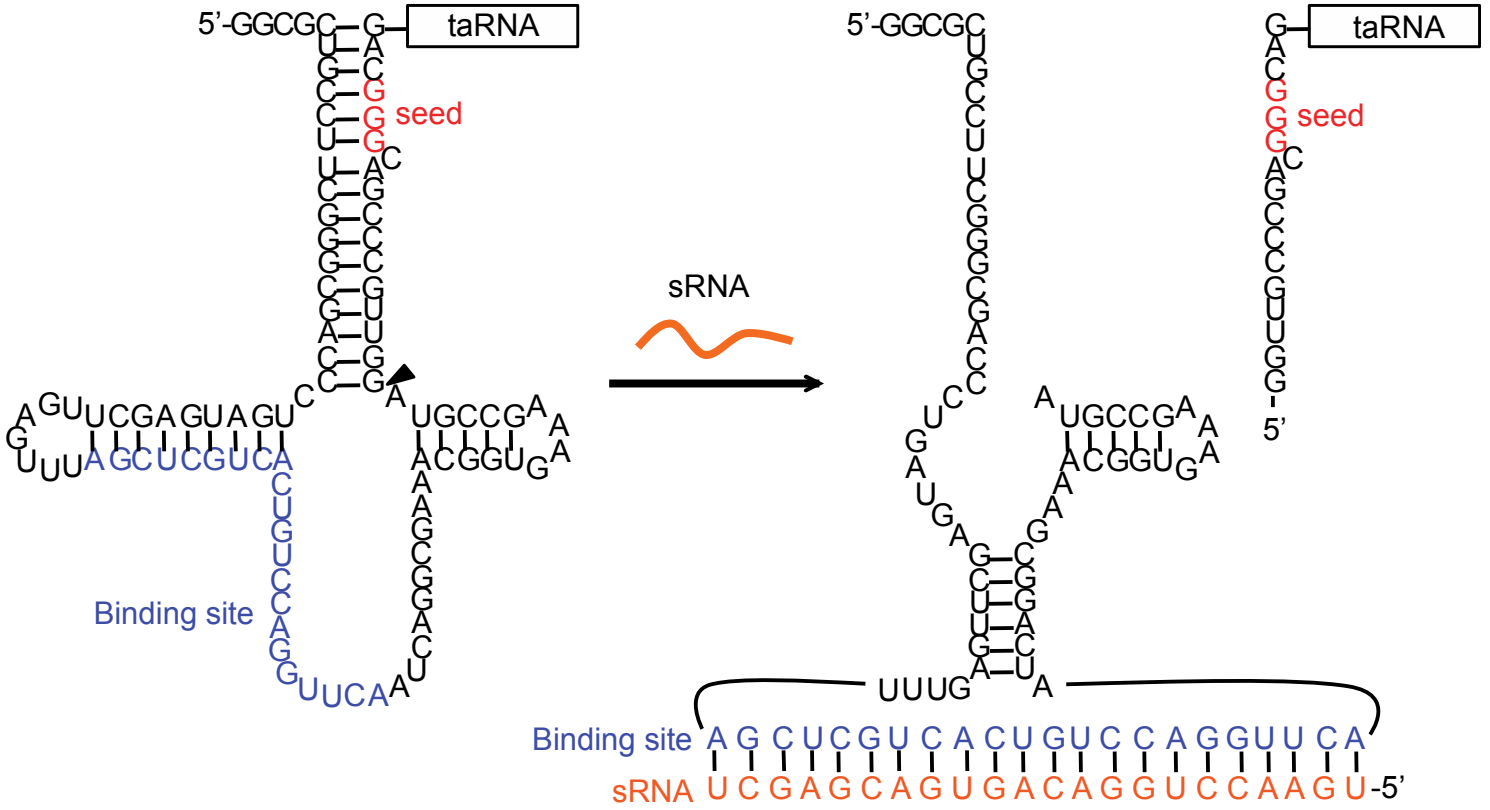


We need novel RNA functions:

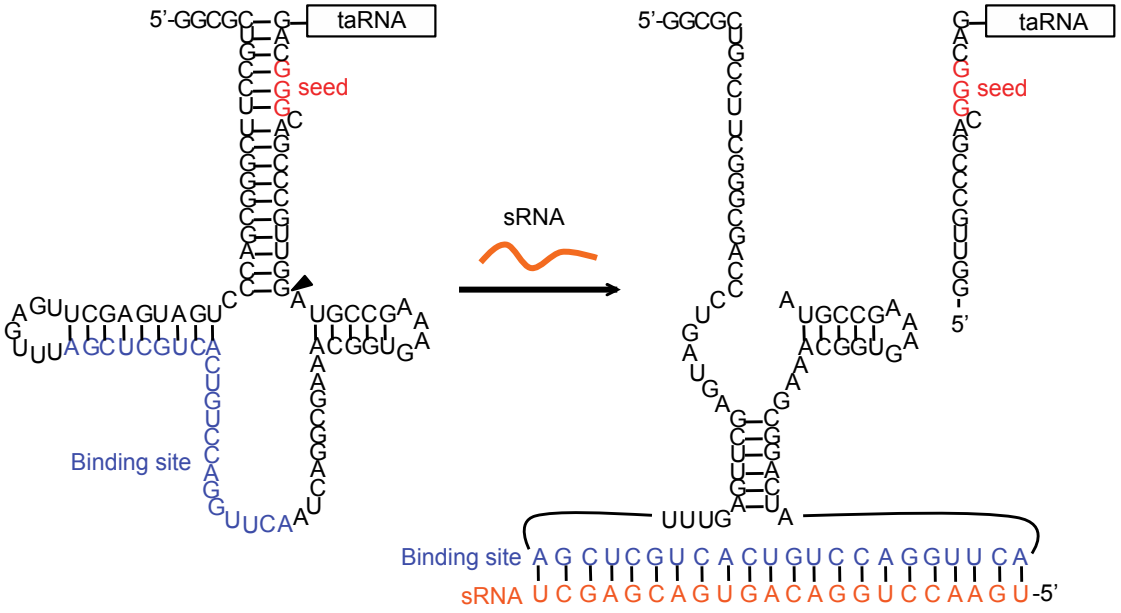
Sensing, non-linearity, cascades and regulation.

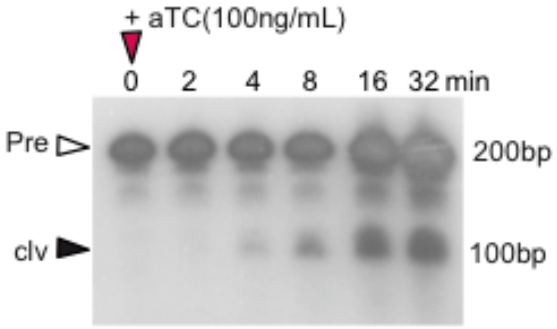
Single-cell analysis the novel RNA sensing Regazyme

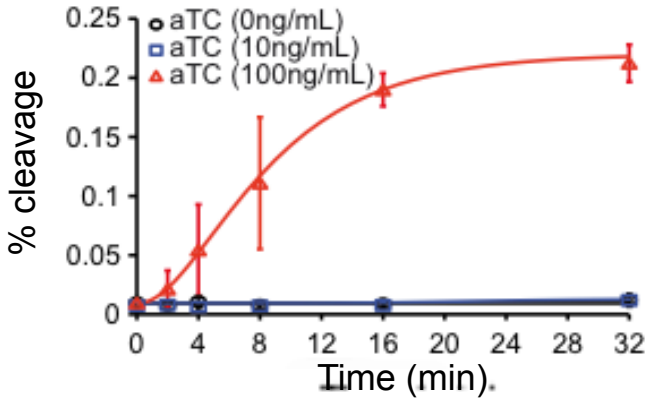


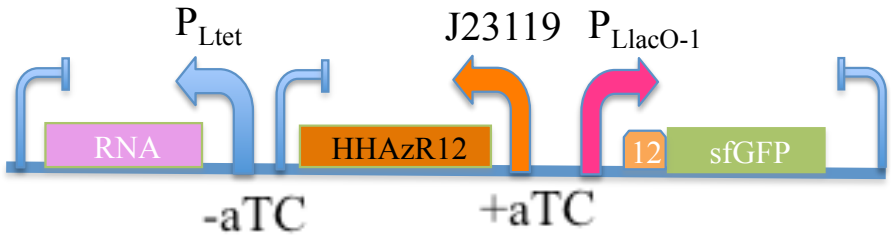


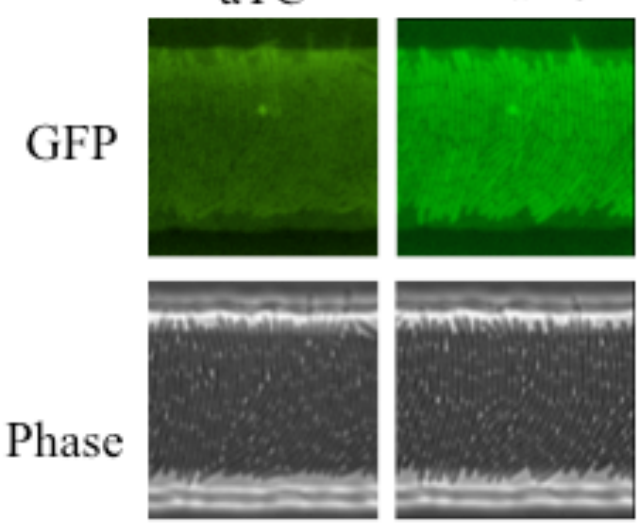
Single-cell analysis the novel RNA sensing Regazyme

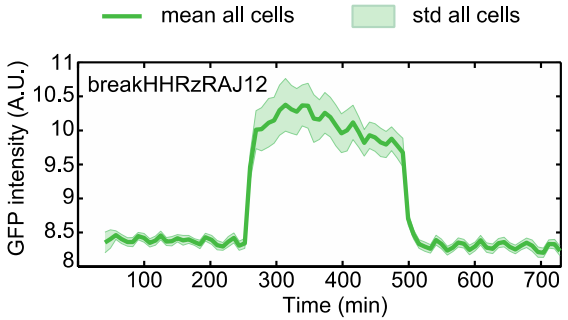


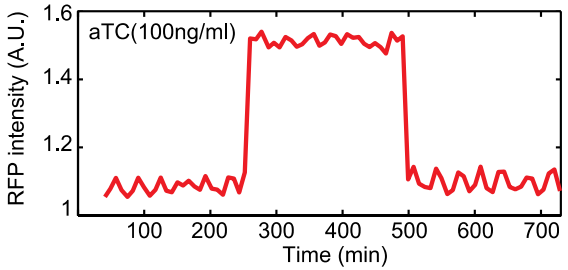




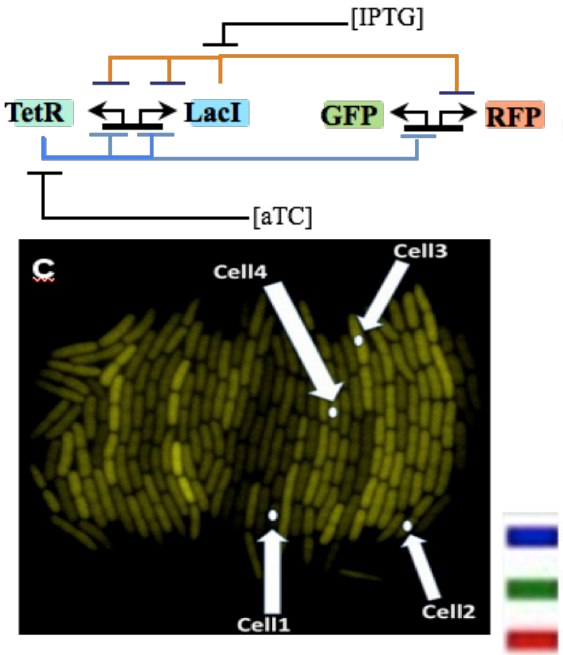


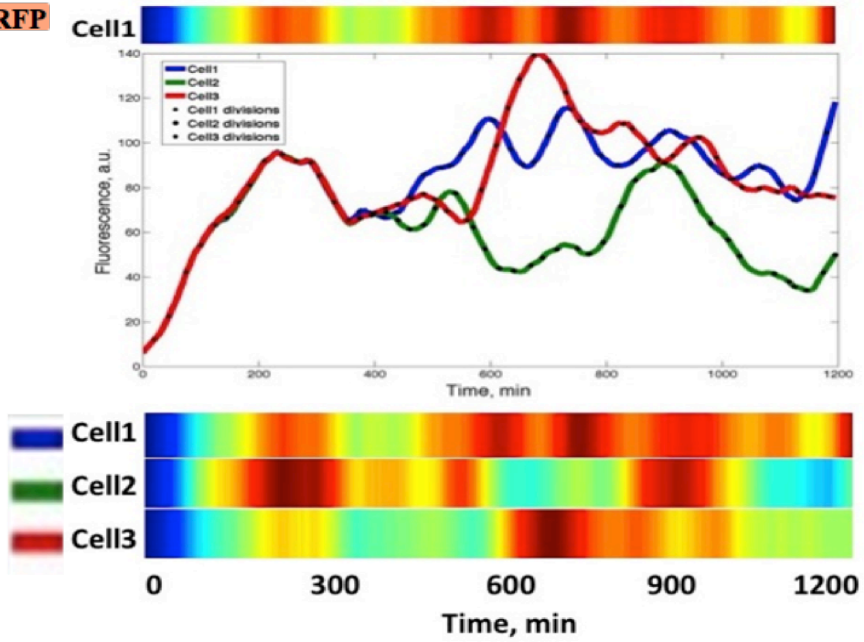


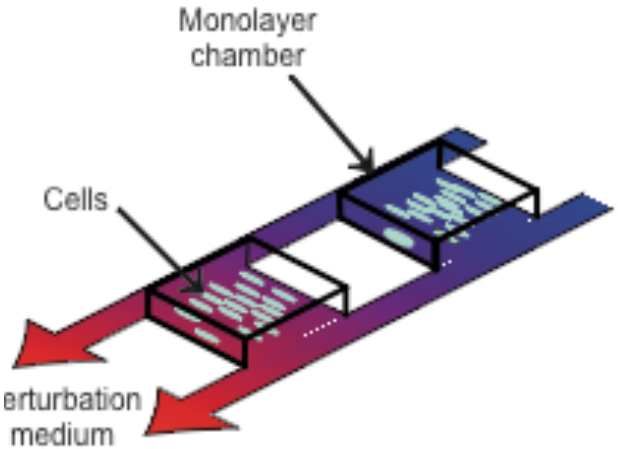




Discussion







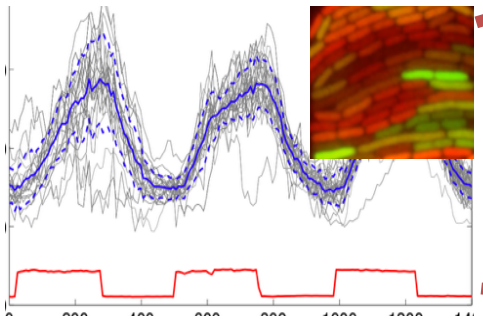
Can we engineer gene/RNA networks with decreased uncertainty in the gene response under variable environments?

Microfluidics-enabled data-rich measurements

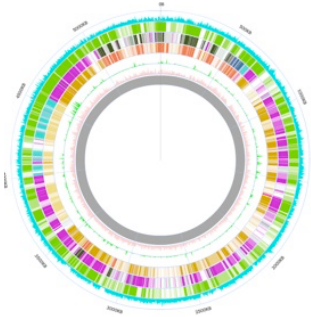
Discussion

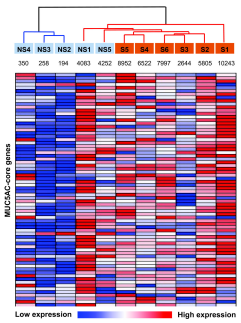
Producing better models accounting for cellular environment uncertainty

Mesurement of single-cell time-course gene expression

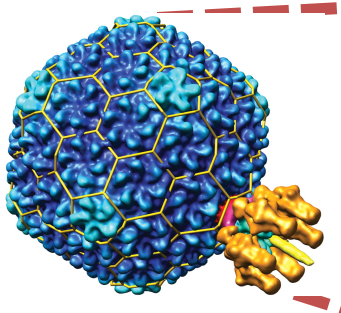


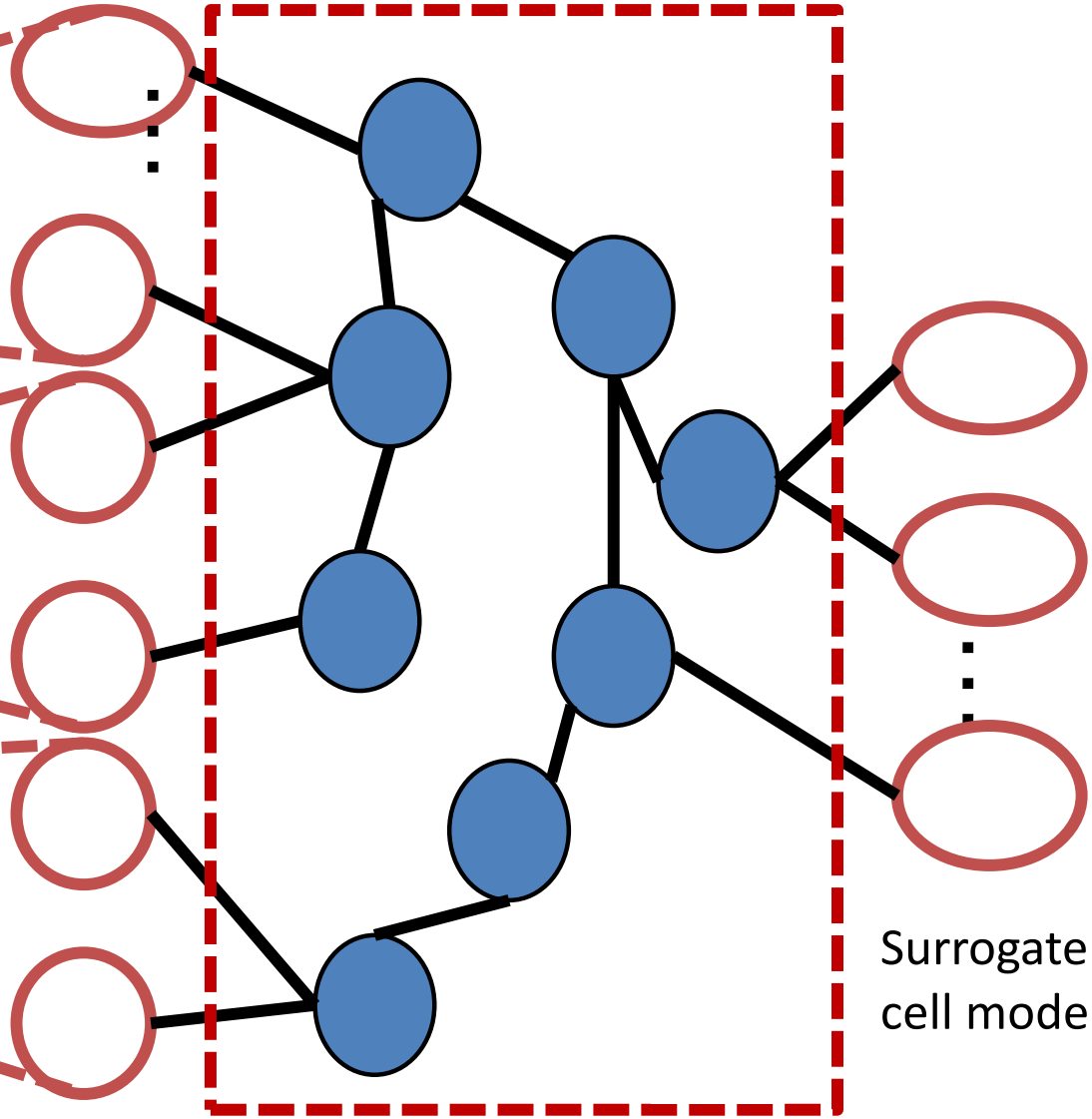
Genomic data





Macromolecular assembly data

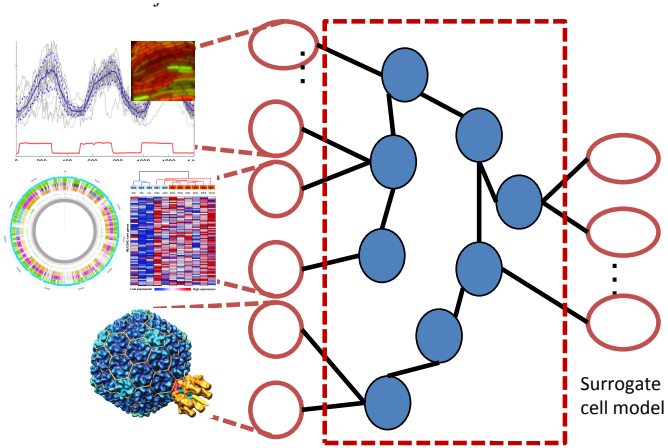


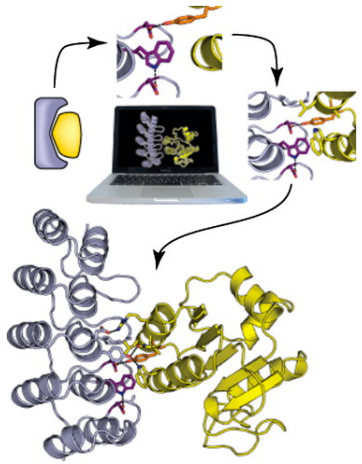


Surrogate cell model

Discussion

Re-engineer existing molecular biological systems to avoid molecular interactions producing unpredictable responses





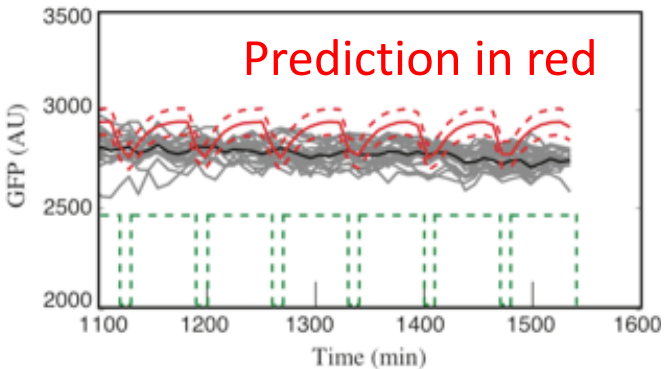
Computational design

Simplified design

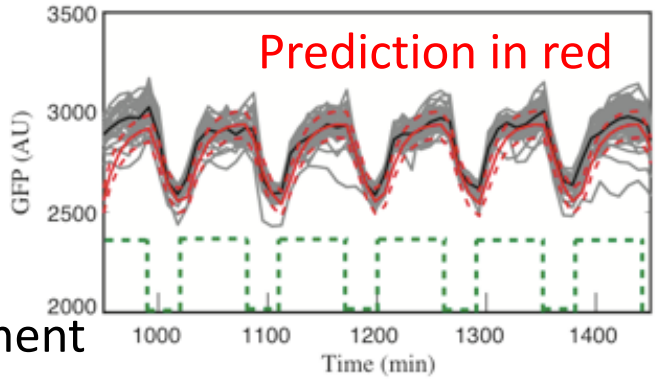
Measurement

Surrogate model inference

Bad prediction

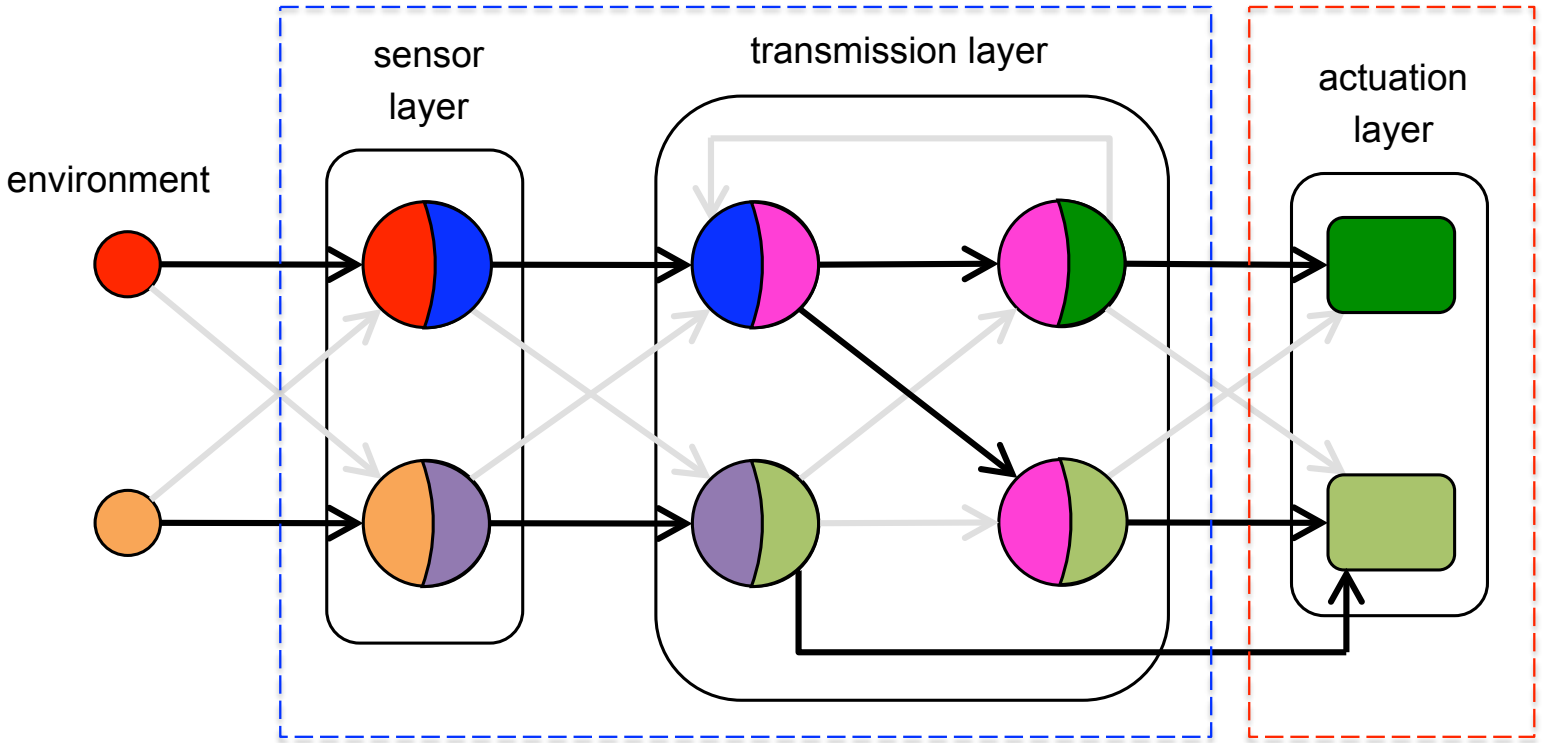


Good prediction



Conclusions

- Working with folded RNA provides many possibilities to engineer circuits in cells by utilising **regulation by allostery**.
  - activation/inactivation of RNA modules
- We had to rely on **computational design to obtain sequences but we need a more predictable model**



Acknowledgements: Jaramillo group

- Michal Legiewicz – Postdoctoral fellow
- John Duncan – Postdoctoral fellow
- Rui Rodrigues – Postdoctoral fellow
- Vijai Singh – Postdoctoral fellow
- Satya Prakhash – PhD student
- William Rostain – PhD student
- Paul MacDonald – PhD student
- Fabio Polesel – Researcher
- Jack Hassal – Researcher
- Nan Papili – Researcher
- George Kimberley – Researcher
- Matthew Tridgett – Researcher
- Peter Johnson – Undergraduate student
- Mariel Montesinos – EU project manager

http://jaramillolab.org
http://synth-bio.org

Collaborators contributing to this talk:

- Jose Antonio Daros (IBMCP, Spain)
  - Regazyme quantification gels
- Catalin Fetita (Telecom SudParis, France)
  - Image treatment
- Brian Munsky (LANL, USA)
  - *In vivo* model inference
- Jeff Hasty (UCSD, USA)
  - Microfluidics

Special thanks to former postdocs:
Dr. Guillermo Rodrigo (group leader; IBMCP, Valencia)
Dr. Shensi Shen (group leader; Ecole Normale Superieure, Lyon)