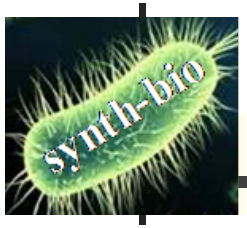
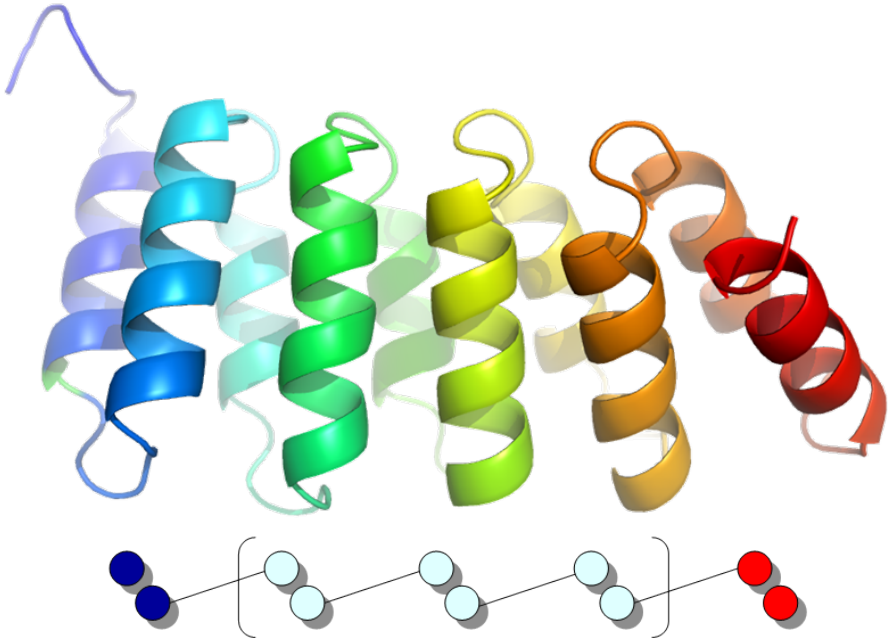


Computational Design of Biological Systems by Automatic Methods



Alfonso JARAMILLO

Synth-Bio group
Programme Epigenomique
CNRS-Genopole-UEVE
&
Ecole Polytechnique

http://synth-bio.org

Outline of Talk

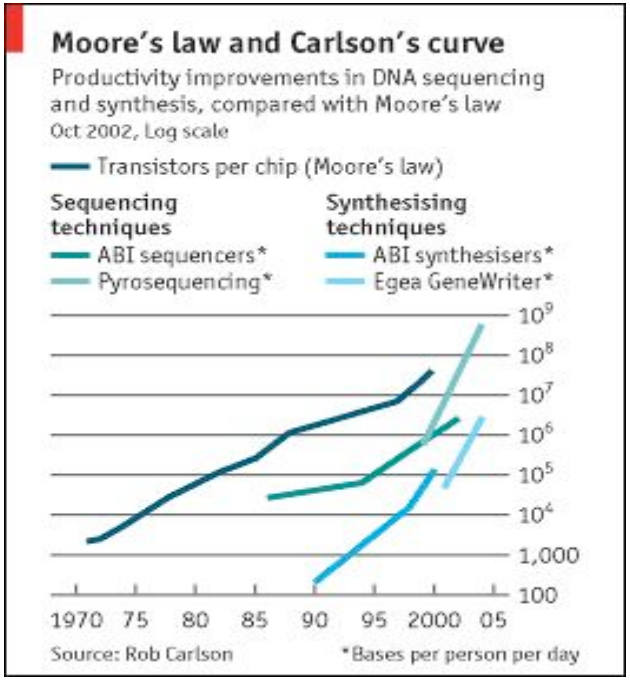
- Molecular Genetics in the post-genomic era
- Design of molecular parts (I): Macromolecules
- Design of molecular parts (II): Networks
- Design of molecular parts (III): Cells

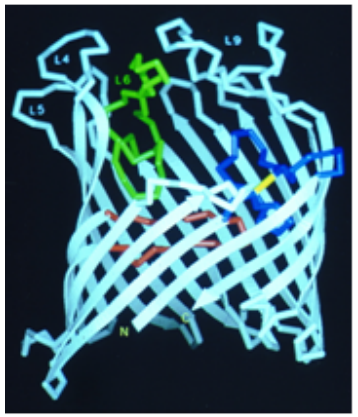
Molecular Genetics in the post-genomic era

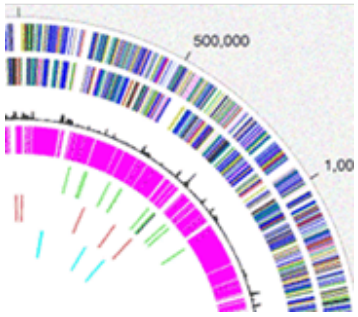
- Can we understand complex genetic systems as a combination of molecular parts?
  - Proteins, RNAs, Genetic circuits, Metabolic circuits, Genomes…
- Approach 1: build a list of parts, construct computational models and test their predictions against experimental data.
  - Systems Biology
- Approach 2: design, construct & validate synthetic systems from molecular parts
  - Synthetic Biology

Enabling breakthroughs in a postgenomic era

- Advances in computing power
- Internet
- Genomic sequencing
- Crystal structures of proteins
- High through-put technologies





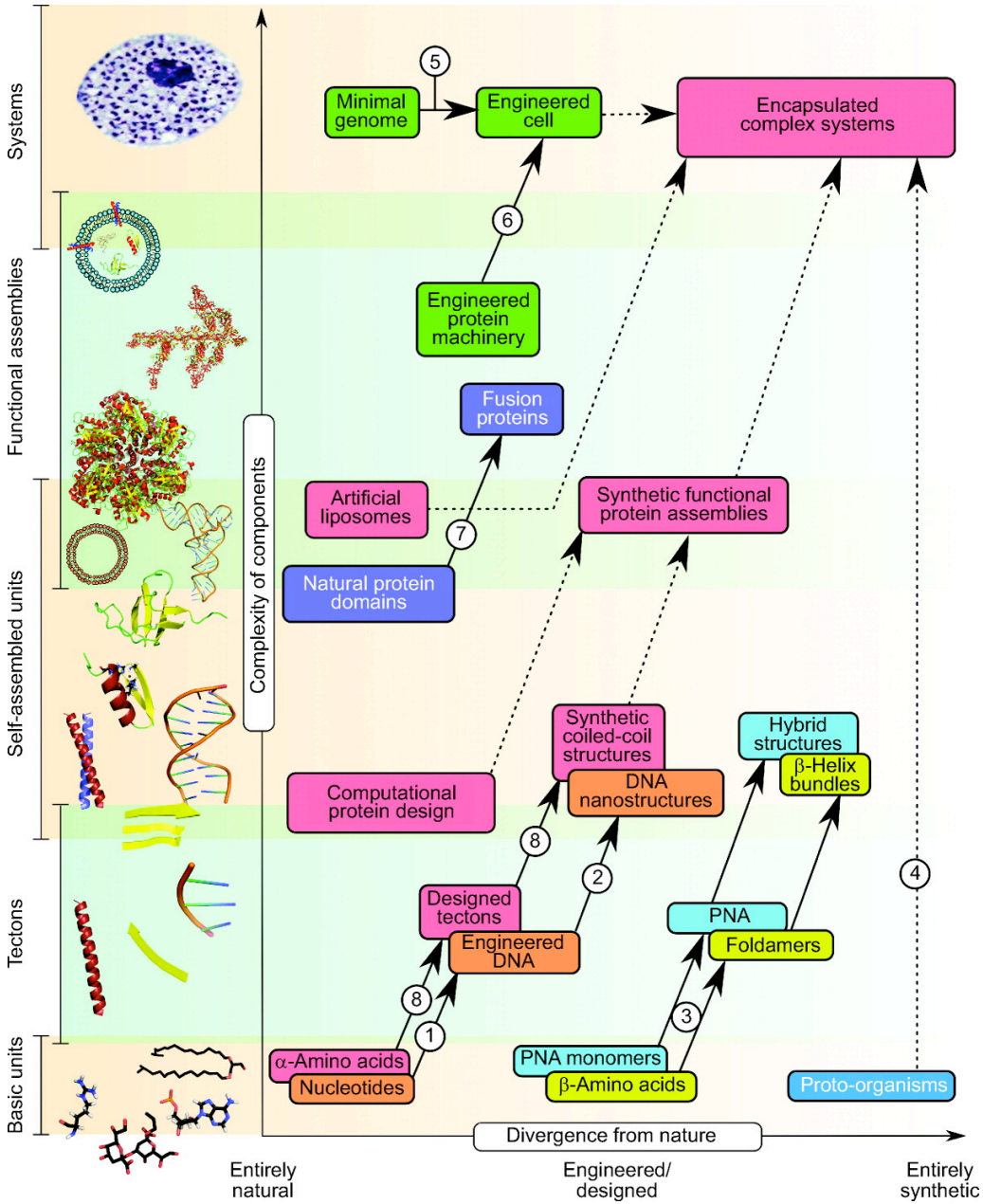


Synthetic Biology

Understanding complex genetic systems as a combination of molecular parts

- Approach: design, construct & validate synthetic systems from molecular parts
  - Problem: Genetic Engineering has been around for more than 30 years and its technology does not scale with current molecular part lists.
  - Solution: Make biology more engineerable
- How we could facilitate the engineering of genetic systems?
  - By embracing engineering paradigms
    - Abstraction, Modularization and standardization
  - By developing computational design methods that apply our knowledge

Engineering new biological systems



Path 1: the construction of engineered DNA, which allows manipulation at every level of the natural hierarchy.

Path 2: the use of engineered DNA to produce novel nanostructures.

Path 3: the development of nonstandard amino acids and base pairs, which can then be assembled into foldamers and DNA analogs.

Path 4: the creation of alternative genetic systems.

Path 5: producing minimal genomes (synthetic chromosomes) and transplanting them into prokaryotic hosts.

Path 6: adding new functions to living organisms by manipulating cell machinery.

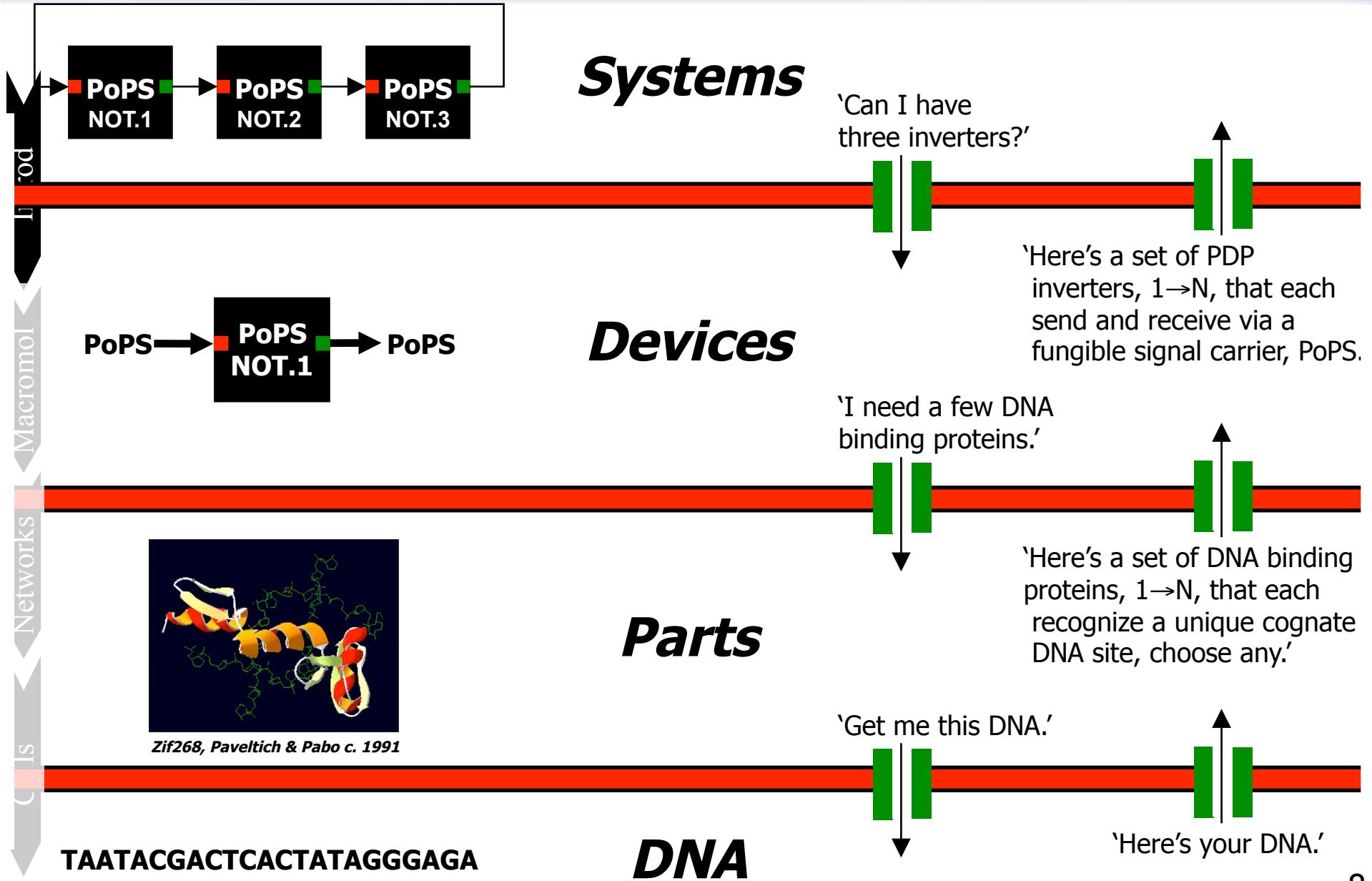
Path 7: the fusion of proteins to produce assemblies with novel functions.

Path 8: the use of peptide synthesis to create programmable building blocks that can assemble further into functional protein components.

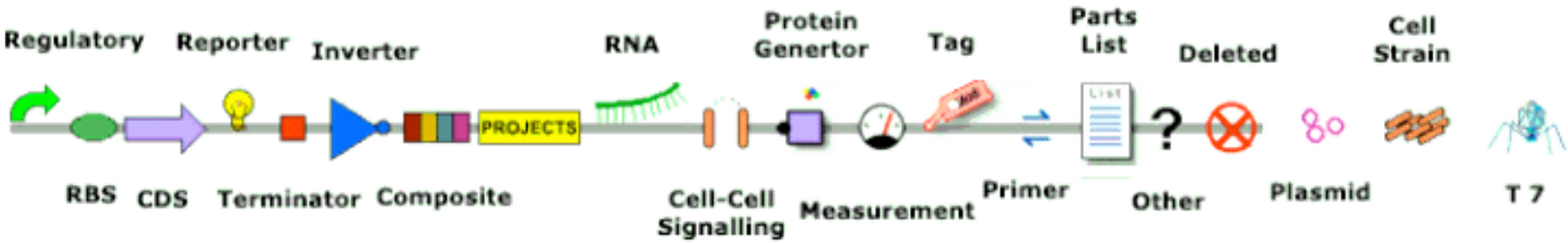
Design principles of SB

- Decoupling Design & Fabrication
  - Rules insulating design process from details of fabrication
  - Enable parts, device, and system designers to work together
  - VLSI electronics (1970s)
- Abstraction
  - Insulate relevant characteristics from overwhelming detail
  - Simple components that can be used in combination
  - From Physics to Electrical Engineering (1900s)
- Standardization of Components
  - Predictable performance
  - Off-the-shelf
  - Mechanical Engineering (1800s) & the manufacturing revolution (e.g. Henry Ford)

Abstraction levels



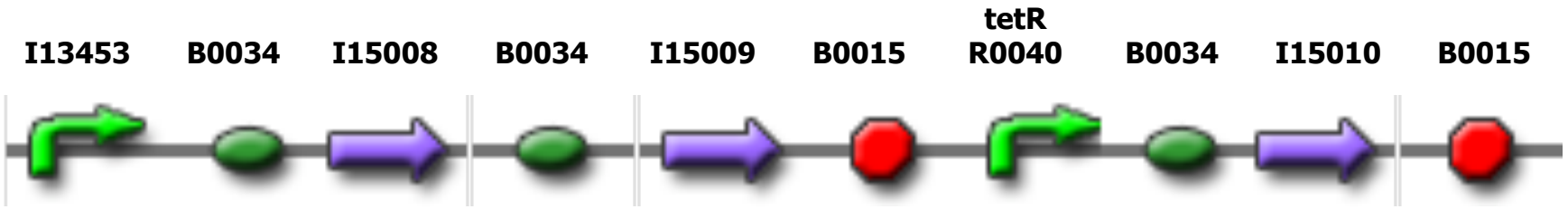
Off-the-shelf biological parts and devices



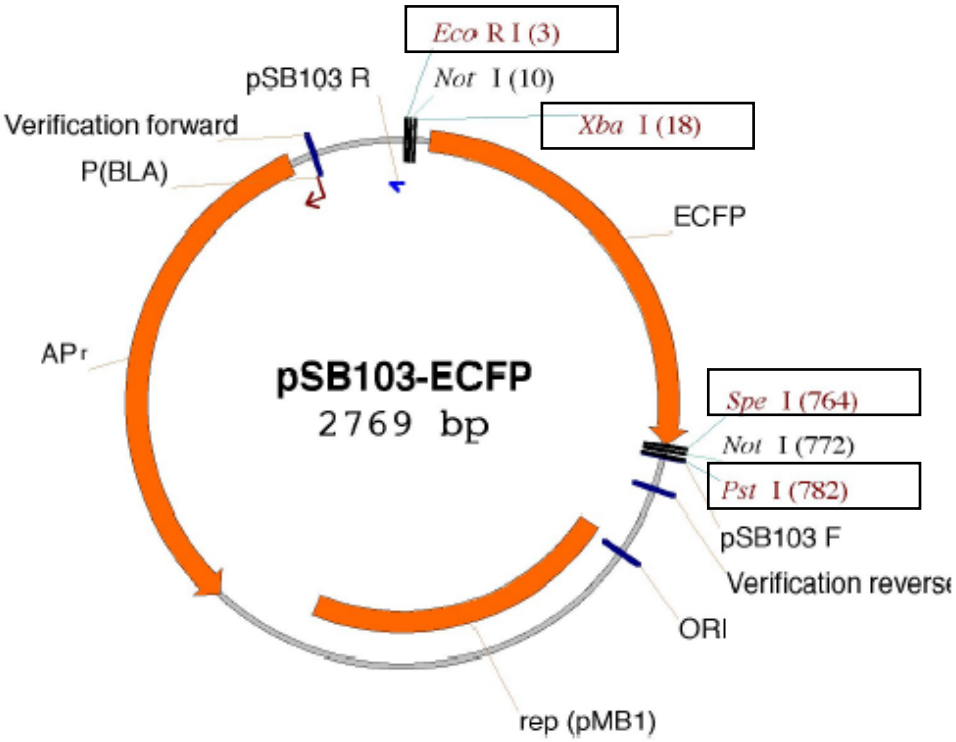
- Promoter
- RBS
- CDS
- Terminator
- Tag
- Primer
- *Operator*

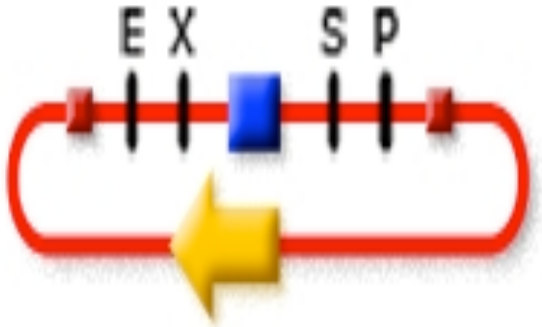
Notice that for the MIT registry, any combination of parts (e.g. devices and systems) is a part.

BBa_M30109 =



Biobricks






```
5' --gca GAATTC GCGGCCGC T TCTAGA GT ACTAGT A GCGGCCG CTGCAG gct---  3'
3' --cgt CTTAAG CGCCGGCG A ACATCT CA TGATCA T CGCCGGC GACGTC cga---  5'
         EcoRI   NotI       XbaI      SpeI     NotI    PstI
```

E X S P

Application: hydrogen production

- Hydrogen is considered the energy carrier of the future
- Use cyanobacteria for photoproduction of hydrogen:
  - Solar energy is inexpensive
  - Production is clean and sustainable

The problem:

- Photosynthesis can produce hydrogen *(hydrogenase)*
- Photosynthesis produces oxygen
- Oxygen inhibits hydrogen production by hydrogenase!

Options:

1) Use resistent hydrogenase (NiFe), less efficient

2) Use efficient hydrogense (Fe), fight inhibition

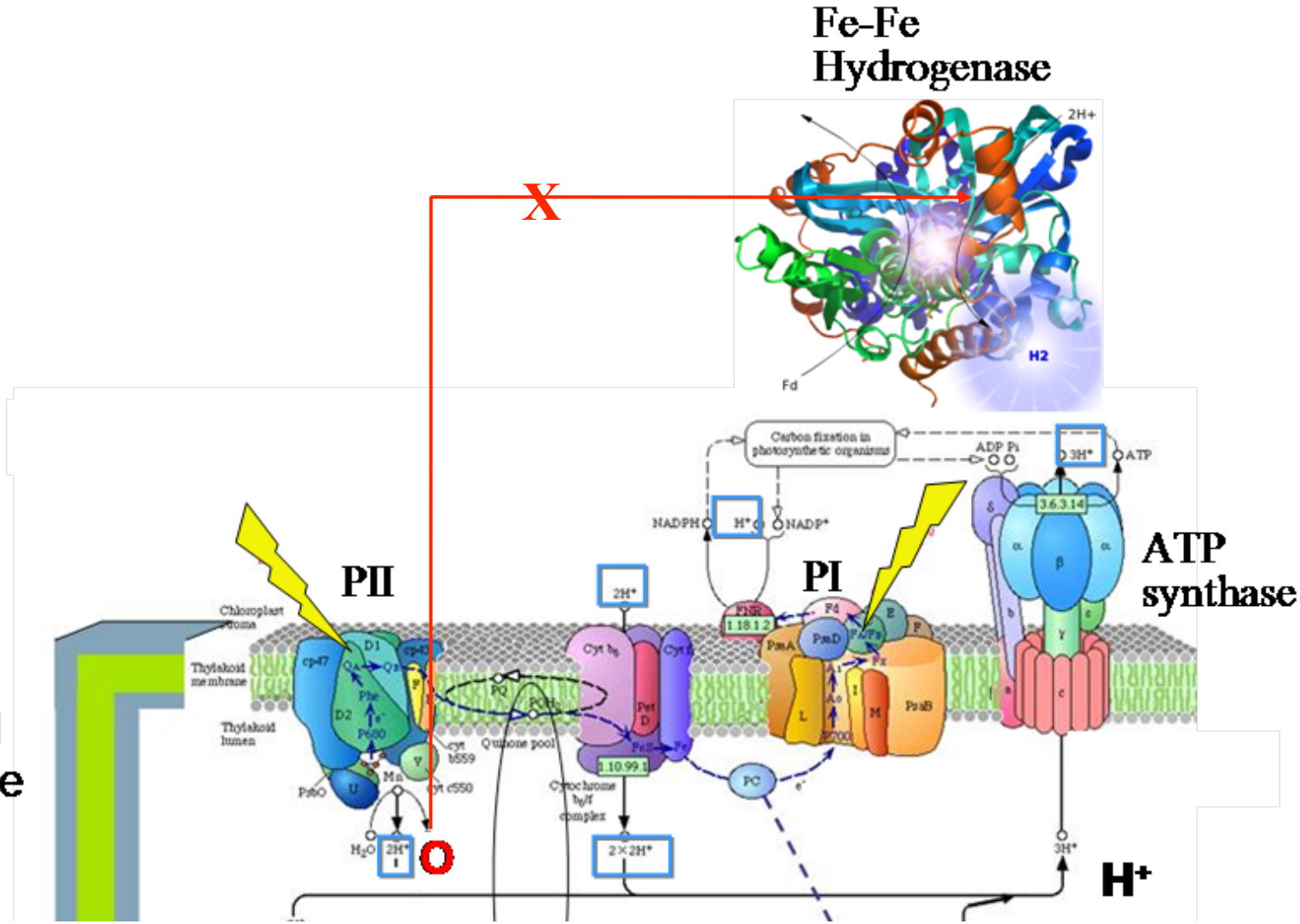
Inhibition

Introd

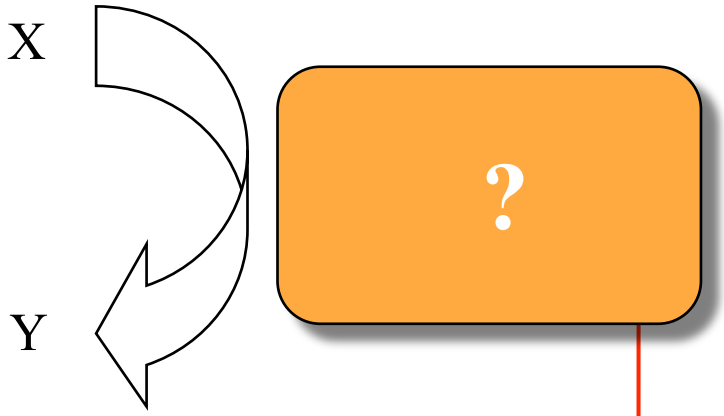
Macromol

Networks

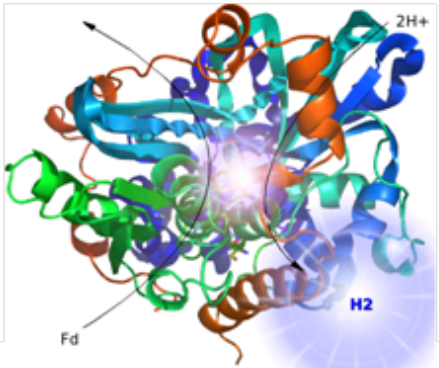
Cells

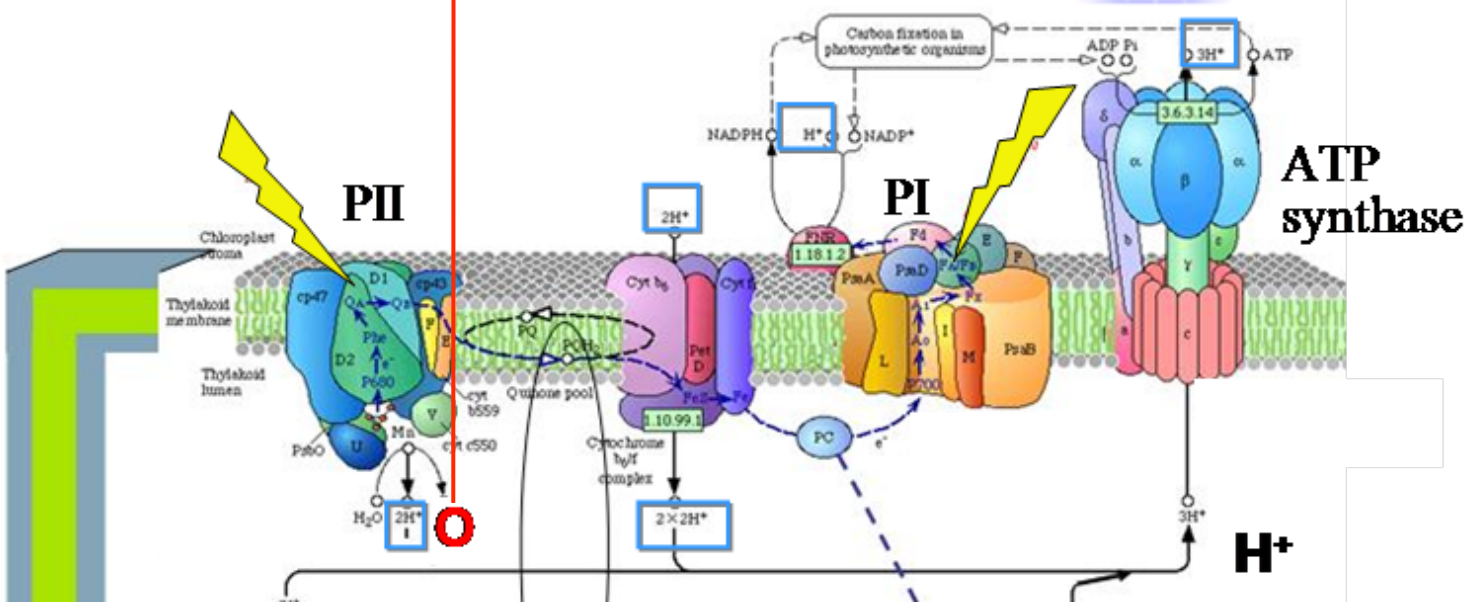


Oxygen consumption



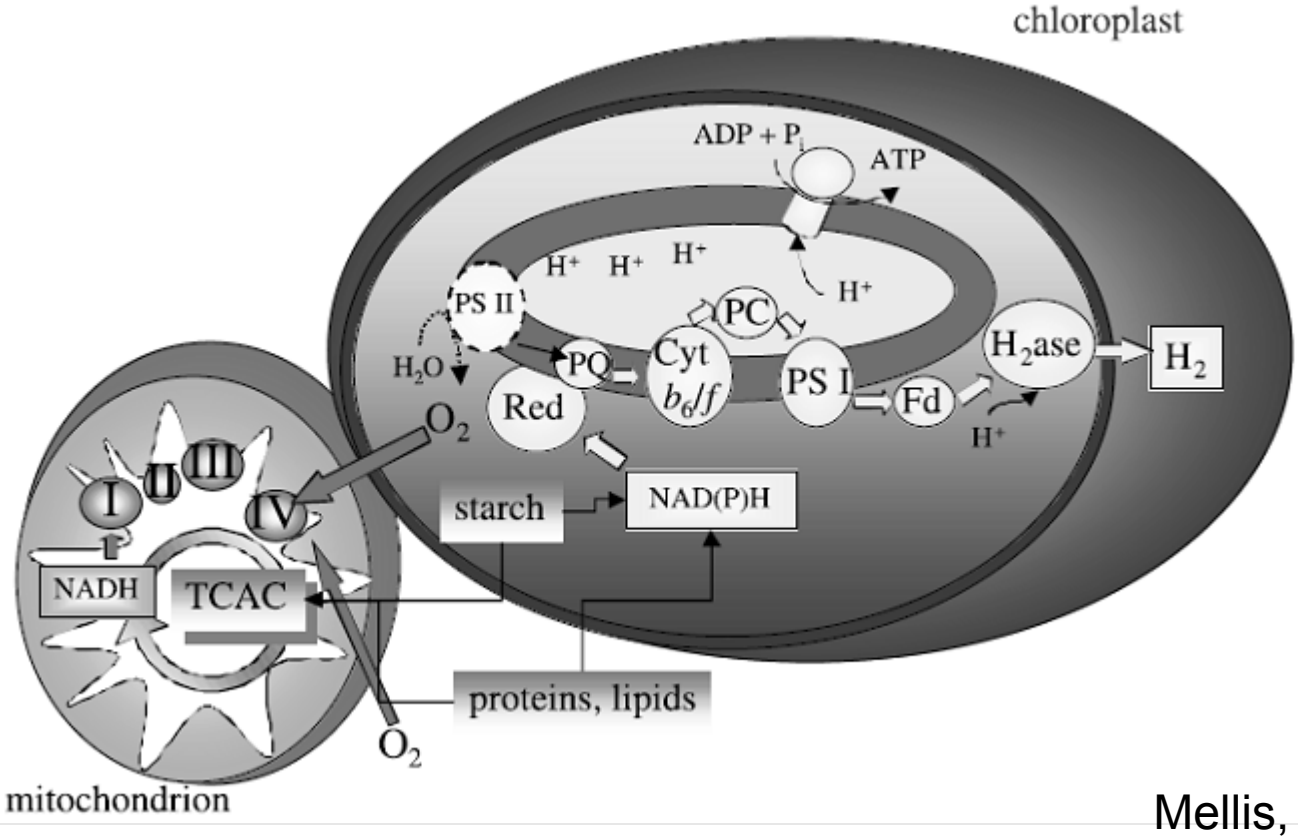
Fe-Fe Hydrogenase





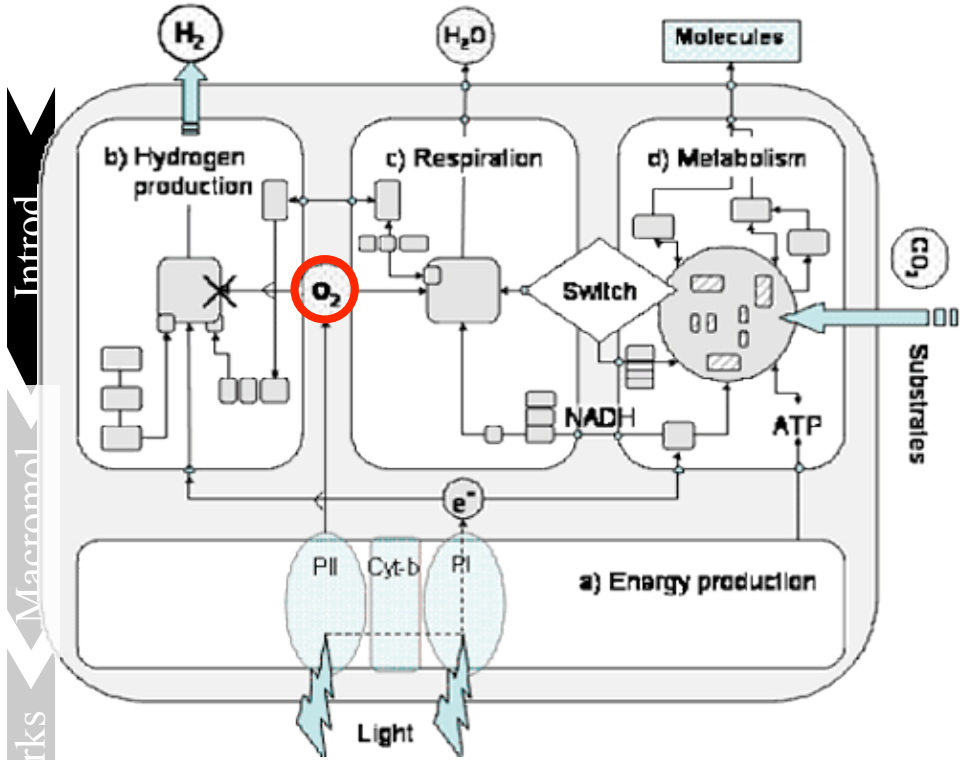
Fighting inhibition

- Use mitochondrion to consume oxygen
- Tune photosynthesis so production and consumption match



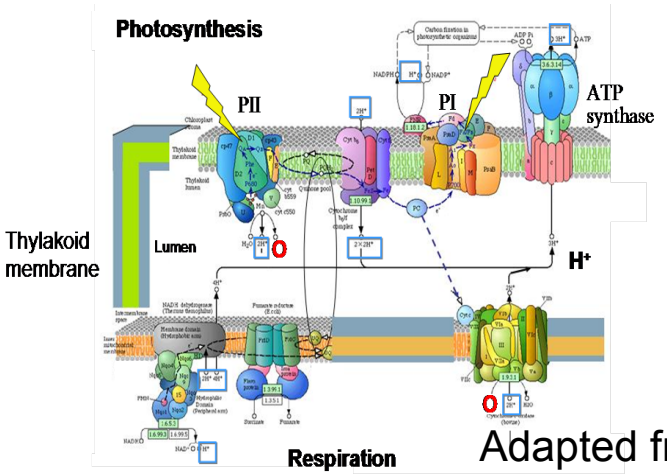
Mellis, 2004

BioModularH2 project



Intro
Macromol
Networks
Cells

- Abstraction
  - -Parts
  - -Devices
  - -Systems
    - H_2 production
    - Oxygen consumption and sensing
    - Regulation
- Specification
- Modularity
- Simulation
- Optimization

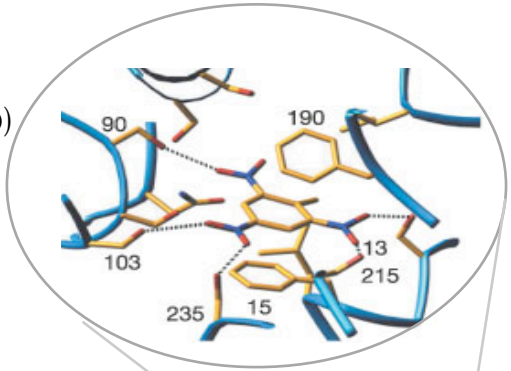


Adapted from KEGG

Multi-scale computational design

Macromolecules

$$E = \sum_{bonds} K_b(b-b_o)^2 + \sum_{angles} K_\theta(\theta-\theta_o)^2 + \sum_{torsions} K\phi(1+\cos(n\phi-\delta))$$
$$+ \sum_{impropers} K_\varphi(\varphi-\varphi_o)^2 + \sum_{Urey-Bradley} K_{UB}(r_{1,3}-r_{1,3,o})^2 +$$
$$+ \sum_{nonbonded} \frac{q_i q_j}{4\pi D r_{ij}} + \varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right]$$



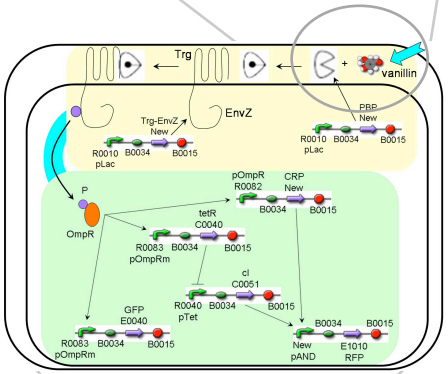
De novo design of proteins:

- DESIGNER
- PROTDES

Biological networks

$$\frac{d[Y]}{dt} = \alpha\frac{1}{1+(\frac{[U]}{K})^n} - \beta[Y] + \gamma$$

$$\frac{d[Y]}{dt} = \alpha\frac{(\frac{[U]}{K})^n}{1+(\frac{[U]}{K})^n} - \beta[Y] + \gamma$$



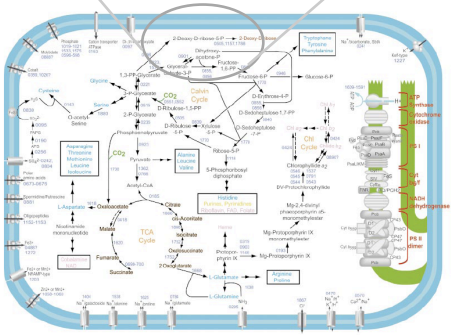
De novo design of:

- Transcriptional networks
  - GENETDES
  - ASMPARTS
- RNA networks
  - RNADES

Genomic background

$$\max \mu = cv$$
$$\text{st}: Sv = b$$
$$v_{\min} \le v \le v_{\max}$$

$$\frac{dA}{dt} = -v_1 - v_2 + v_3 + b_1$$
$$\frac{dB}{dt} = v_1 + v_4 - b_2$$
$$\frac{dC}{dt} = v_2 - v_3 - v_4 - b_3$$

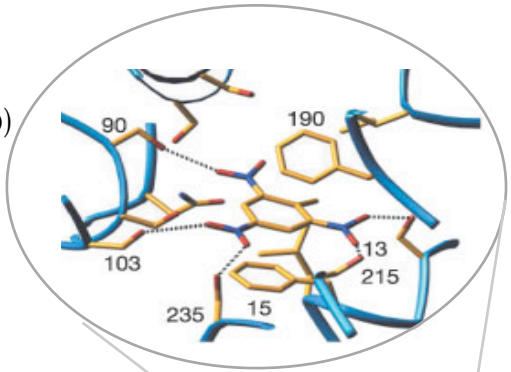


- *De novo* design of metabolic pathways by retro-biosynthesis
  - DESHARKY
- Network inference from microarray & proteomics resp.
  - INFERGENE

Multi-scale computational design

Macromolecules

$$E = \sum_{bonds} K_b(b-b_o)^2 + \sum_{angles} K_\theta(\theta-\theta_o)^2 + \sum_{torsions} K\phi(1+\cos(n\phi-\delta))$$
$$+ \sum_{impropers} K_\varphi(\varphi-\varphi_o)^2 + \sum_{Urey-Bradley} K_{UB}(r_{1,3}-r_{1,3,o})^2 +$$
$$+ \sum_{nonbonded} \frac{q_i q_j}{4\pi D r_{ij}} + \varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right]$$



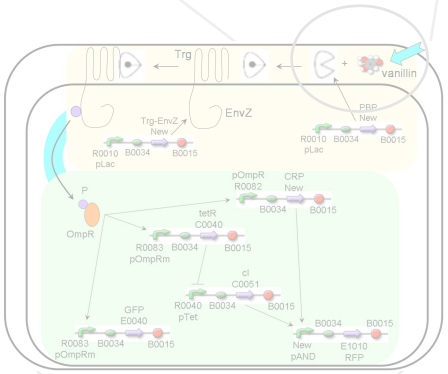
De novo design of proteins:

- DESIGNER
- PROTDES

Biological networks

$$\frac{d[Y]}{dt} = \alpha\frac{1}{1+(\frac{[U]}{K})^n} - \beta[Y] + \gamma$$

$$\frac{d[Y]}{dt} = \alpha\frac{(\frac{[U]}{K})^n}{1+(\frac{[U]}{K})^n} - \beta[Y] + \gamma$$



De novo design of:

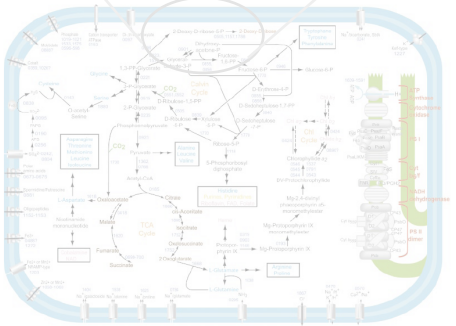
- Transcriptional networks
  - GENETDES
  - ASMPARTS
- RNA networks
  - RNADES

Genomic background

$$\max \mu = cv$$

$$\text{st} : Sv = b$$

$$v_{\min} \le v \le v_{\max}$$



- *De novo* design of metabolic pathways by retro-biosynthesis
  - DESHARKY
- Network inference from microarray & proteomics resp.
  - INFERGENE

How can we design protein structure and function?

Protein design: Inverse folding problem

Physical model at atomic scale

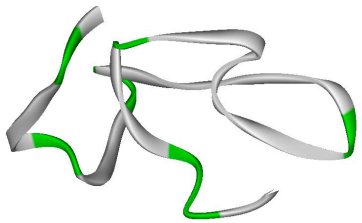
$$E = \sum_{bonds} K_b (b - b_o)^2 + \sum_{angles} K_\theta (\theta - \theta_o)^2 + \sum_{torsions} K\phi (1 + \cos(n\phi - \delta))$$
$$+ \sum_{impropers} K_\varphi (\varphi - \varphi_o)^2 + \sum_{Urey-Bradley} K_{UB} (r_{1,3} - r_{1,3,o})^2 + \sum_{nonbonded} \frac{q_i q_j}{4\pi D r_{ij}} + \varepsilon_{ij} \left[\left(\frac{R_{\min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{\min,ij}}{r_{ij}} \right)^{6} \right]$$

Protein design: Inverse folding problem

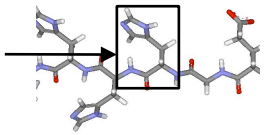
Physical model at atomic scale

$$E = \sum_{bonds} K_b (b - b_o)^2 + \sum_{angles} K_\theta (\theta - \theta_o)^2 + \sum_{torsions} K\phi (1 + \cos(n\phi - \delta))$$
$$+ \sum_{impropers} K_\varphi (\varphi - \varphi_o)^2 + \sum_{Urey-Bradley} K_{UB} (r_{1,3} - r_{1,3,o})^2 + \sum_{nonbonded} \frac{q_i q_j}{4\pi D r_{ij}} + \varepsilon_{ij} \left[\left(\frac{R_{\min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{\min,ij}}{r_{ij}} \right)^{6} \right]$$

Unfolded



$G_U(\mathrm{AA}) \approx \langle E \rangle$

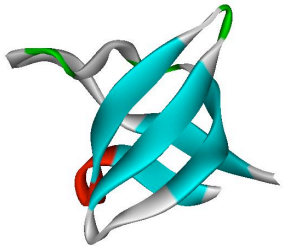


Dipeptide random model partition function

Folding

Rotamer library

Folded



$G_F \approx E$

Protein design: Inverse folding problem

Physical model at atomic scale

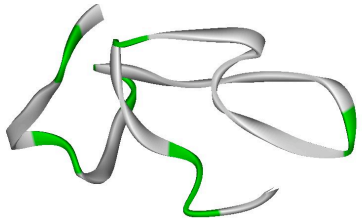
$$E = \sum_{bonds} K_b (b - b_o)^2 + \sum_{angles} K_\theta (\theta - \theta_o)^2 + \sum_{torsions} K\phi (1 + \cos(n\phi - \delta)) + \sum_{impropers} K_\varphi (\varphi - \varphi_o)^2 + \sum_{Urey-Bradley} K_{UB} (r_{1,3} - r_{1,3,o})^2 + \sum_{nonbonded} \frac{q_i q_j}{4\pi D r_{ij}} + \varepsilon_{ij} \left[\left(\frac{R_{\min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{\min,ij}}{r_{ij}} \right)^{6} \right]$$

Score sequences with:

$$\Delta G_{U \to F} = G_F - G_U$$

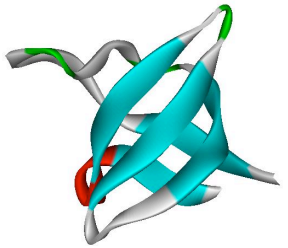
Combinatorial optimisation

Unfolded

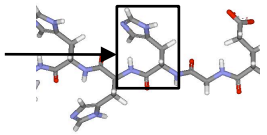


Folding

Folded



$G_U(\mathrm{AA}) \approx \langle E \rangle$



Dipeptide random model
partition function

Rotamer library

$G_F \approx E$

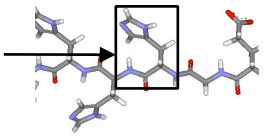
Main challenges in protein design

The main challenges in protein design require methodological advances.

- Model unfolded state

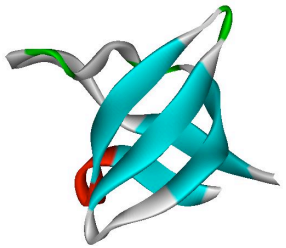
  Syst & Synth Biol 2009

$$G_U(\mathrm{AA}) \approx \langle E \rangle$$



- Model folded state

  Syst & Synth Biol 2009. PROTDES software



- Implicit solvation

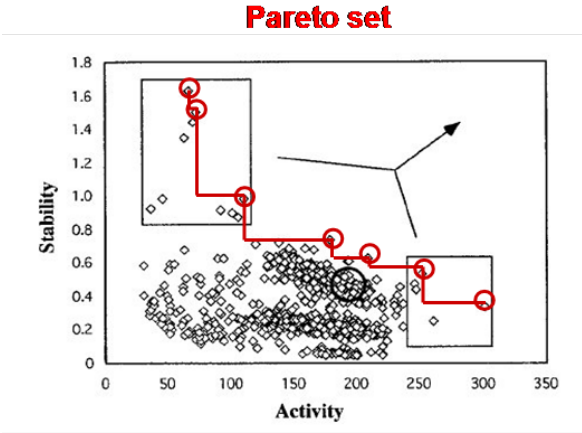
  Biophys J. 2005

- Side-chain and backbone flexibility

  Proteins 2009

- Combinatorial explosion

  J Comput Chem 2008

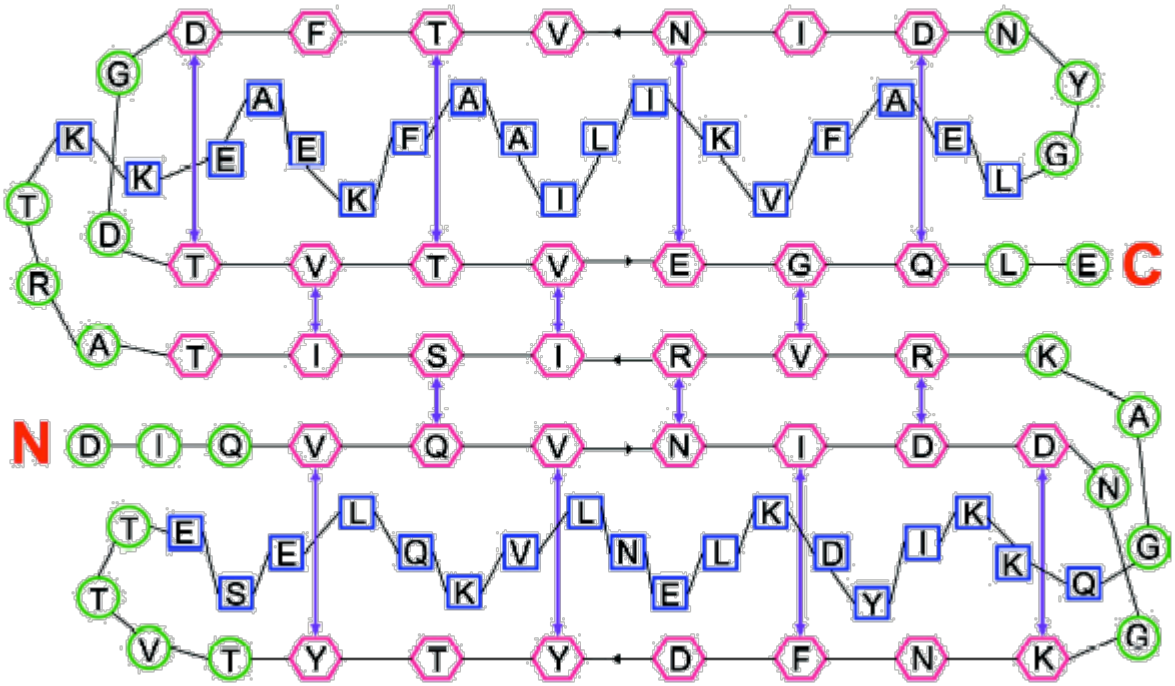


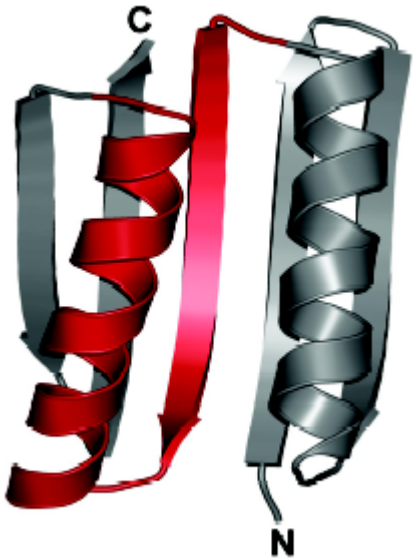
Data from F. Arnold Acc. Chem. Res 1998

Synthetic Protein Scaffolds

Design of a New Fold

A new topology, not in PDB, was chosen:





Baker's group, (Science 2003)

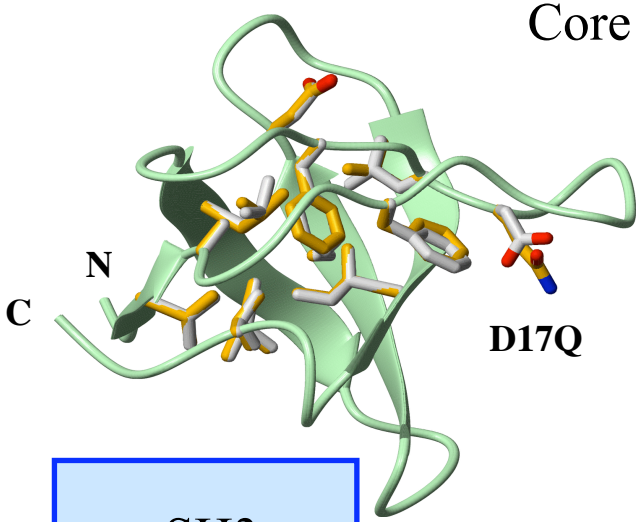
Designed protein

Blue computationally designed, red x-ray structure

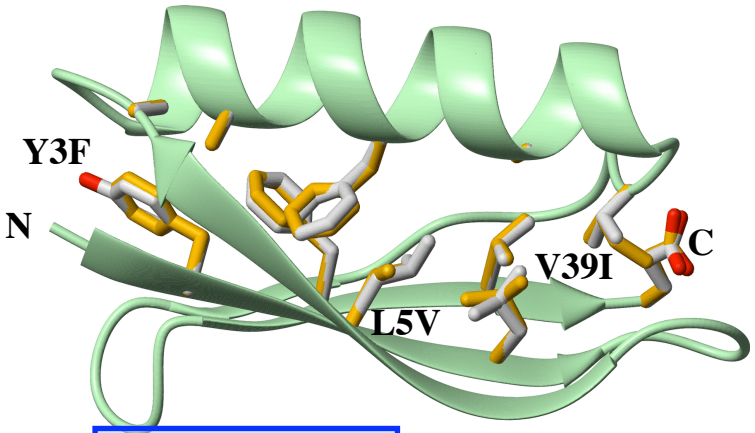
RMSD 1.17A

Redesign of natural protein domains

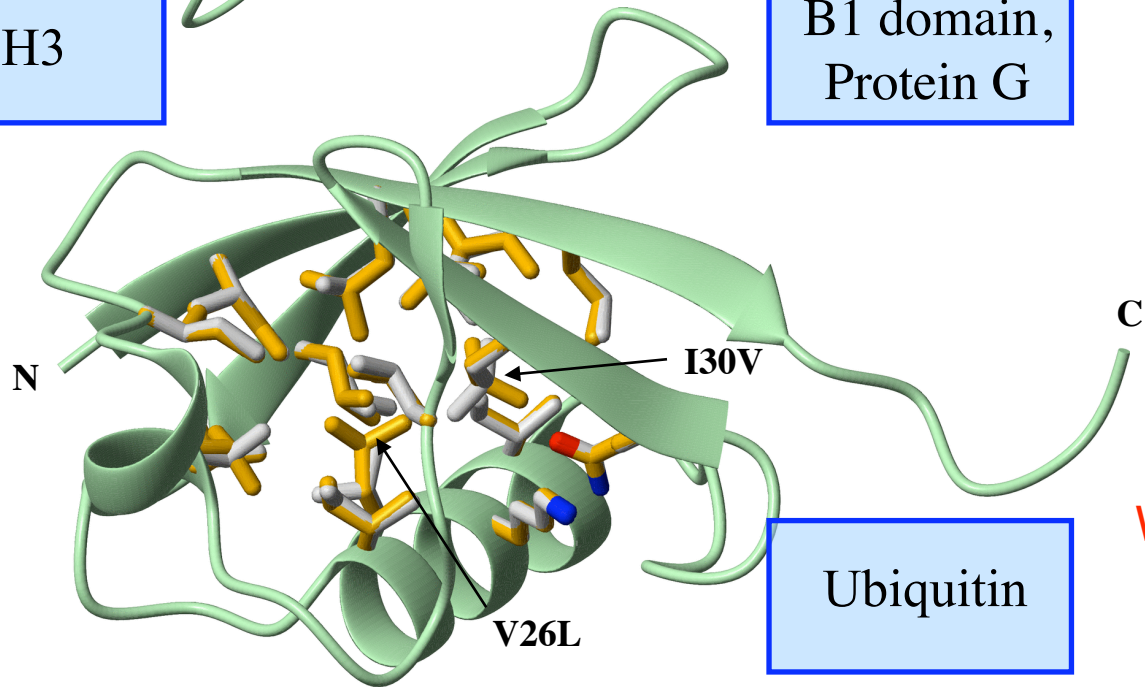
Core re-design



SH3



B1 domain,
Protein G



Ubiquitin

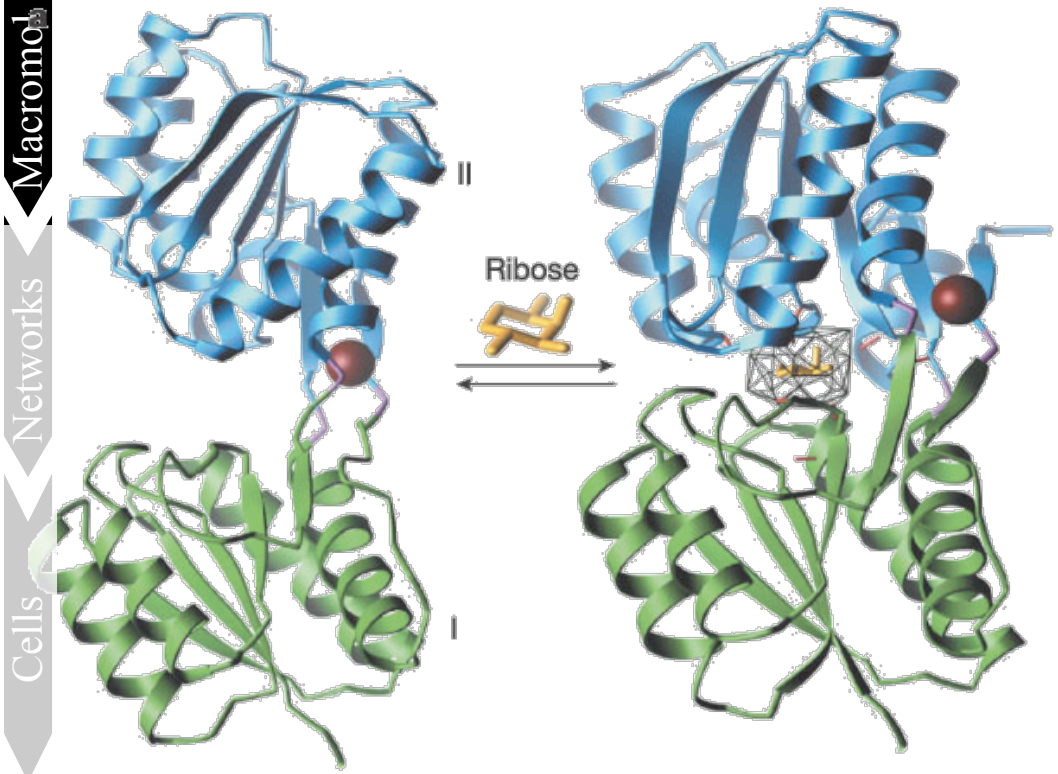
Wernisch et al.,
JMB 2000

New Molecular Recognition

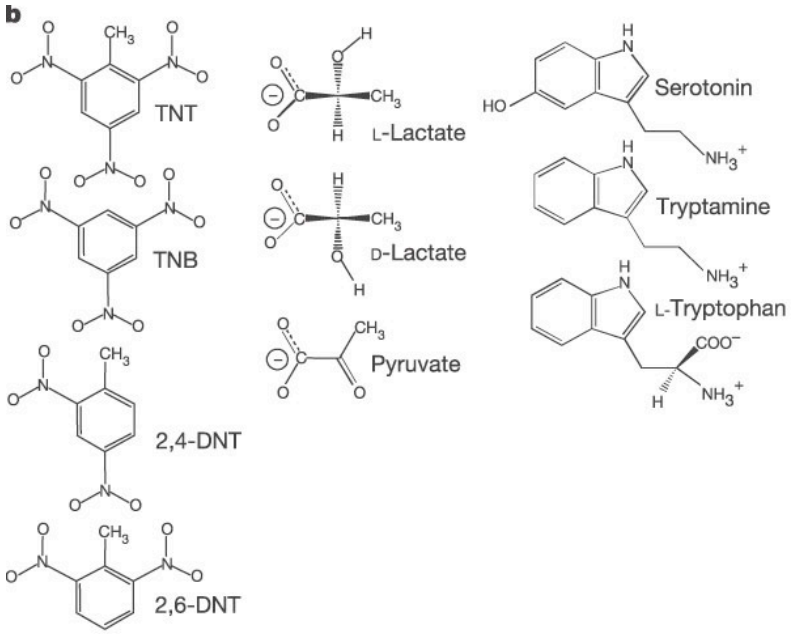
Design of new sensor proteins

Redesign 5 periplasmic binding proteins (PBP) to bind trinitrotoluene (TNT), L-lactate or serotonin in place of the wild-type sugar or amino-acid ligands

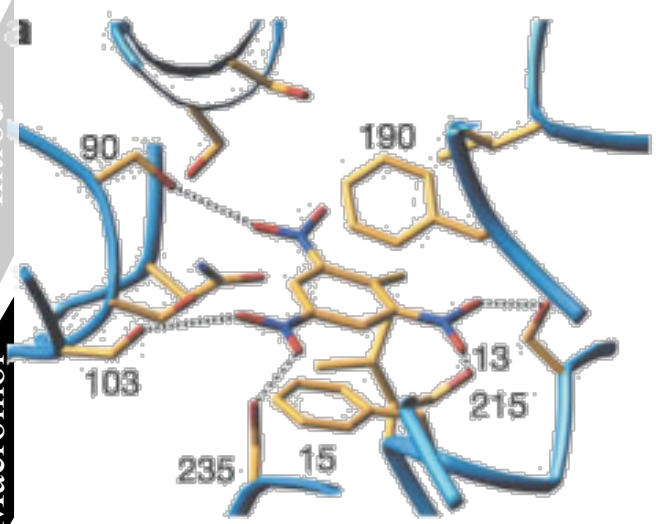
Hellinga's group, (Nature 2003)



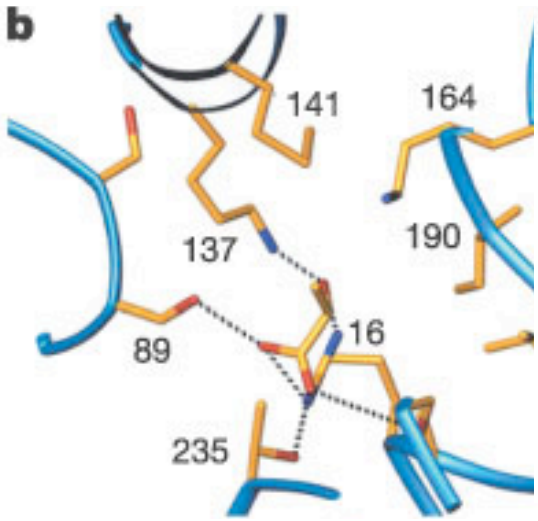
open closed



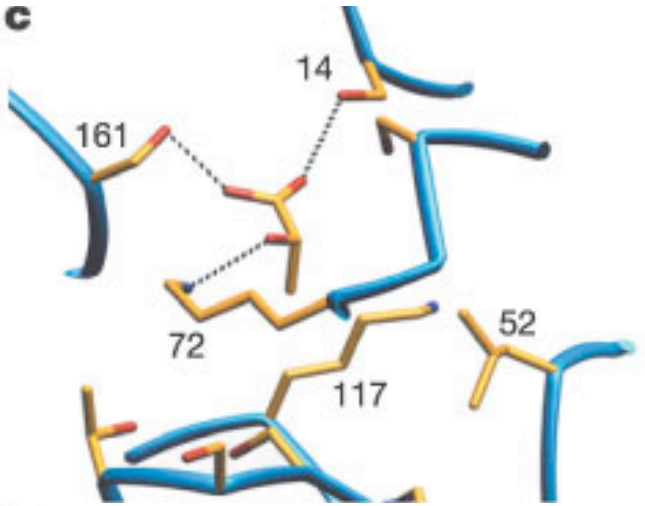
Design of new sensor proteins



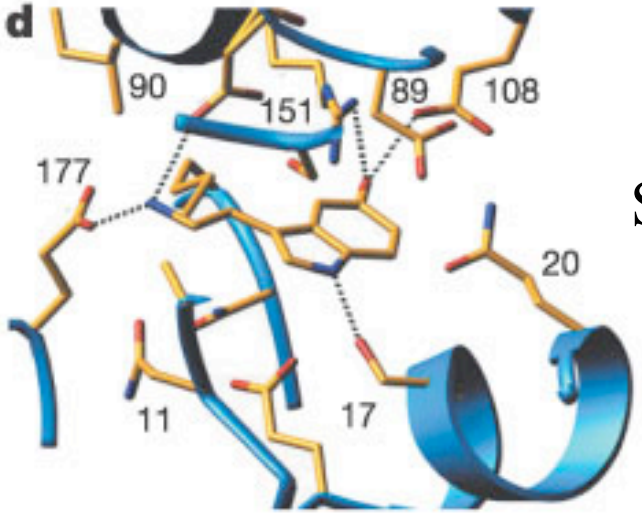
TNT.R3
(Affinity 2 nM)



Lac.R1
(Affinity 7.4 μM)



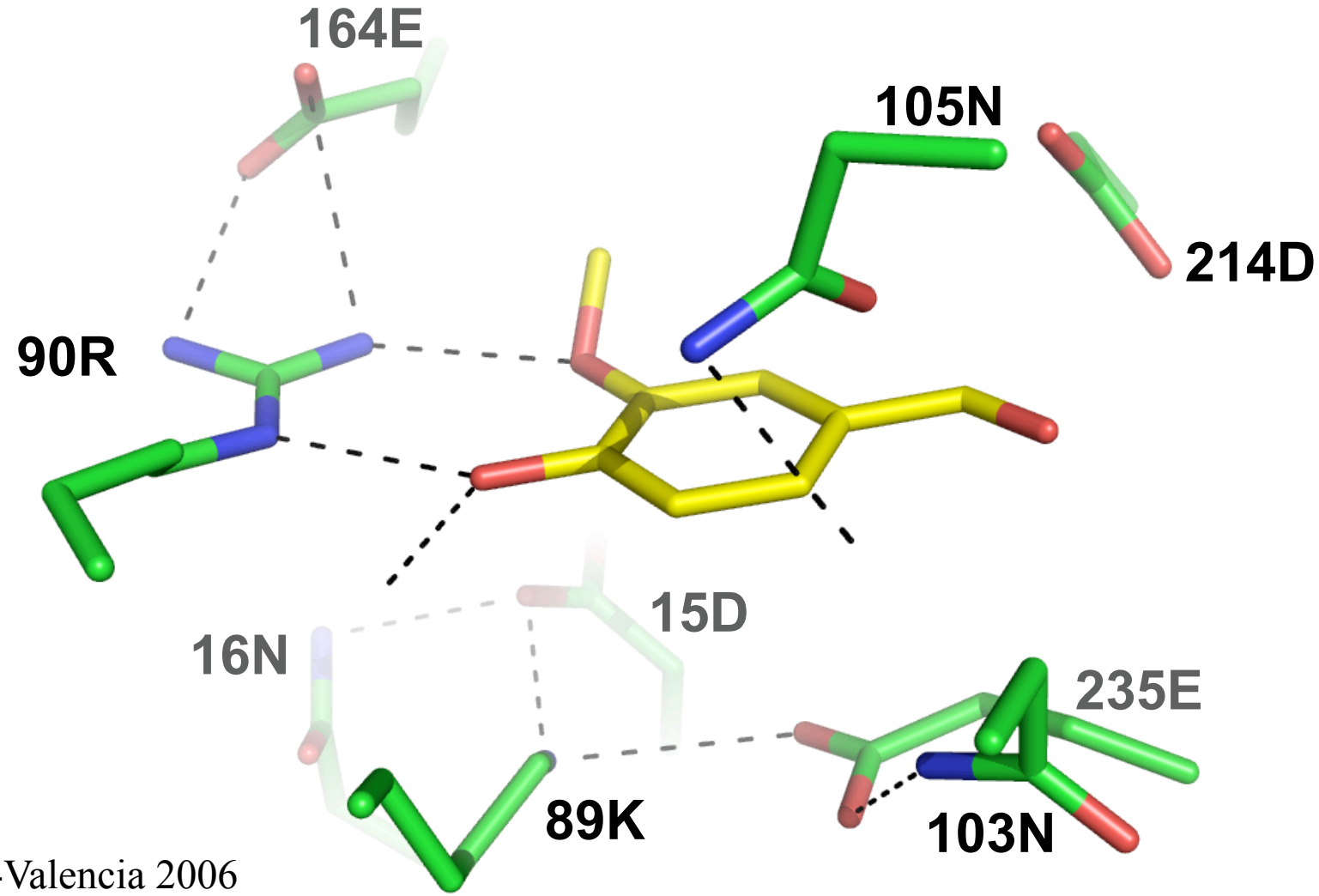
Lac.H1
(Affinity 1.8 μM)



Stn.A1
(Affinity 50 μM)

Hellinga's group,
(Nature 2003)

Designed Binding Site for Vanillin



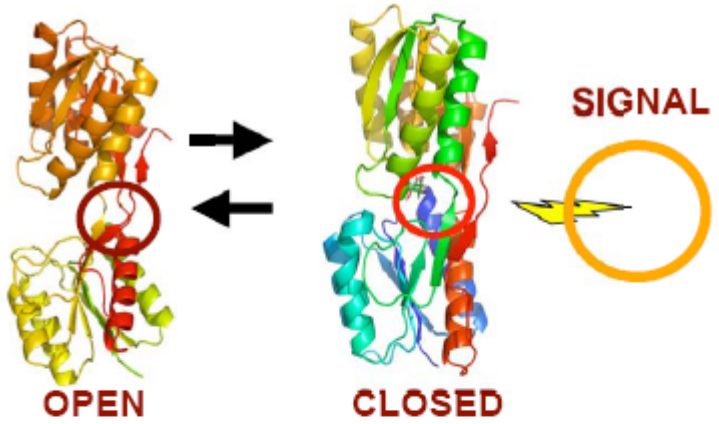
GEM-Valencia 2006

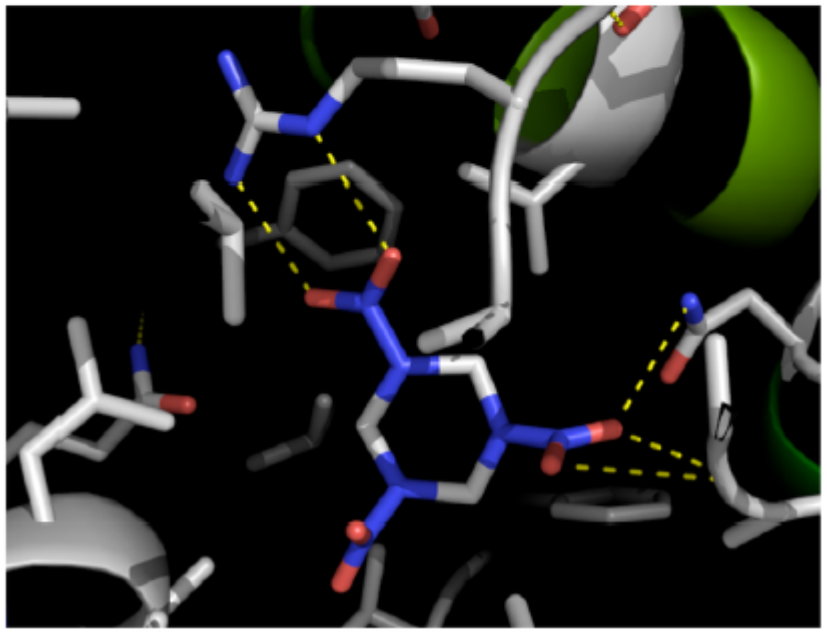
http://www.intertech.upv.es/wiki/

RDX biosensor

- RDX: a very common explosive used in landmines.
- Traditionall landmine detection mechanism (metal detection) useless in plastic made ones. Has more than 80% of false positives.
- Use *Pseudomonas putida* as a specific RDX detector.
- Use ribose binding protein and redesign binding site for specificity towards RDX.

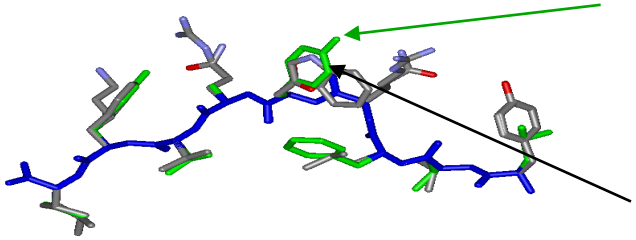
Collaboration with Prof. V. de Lorenzo (CNB-CSIC).

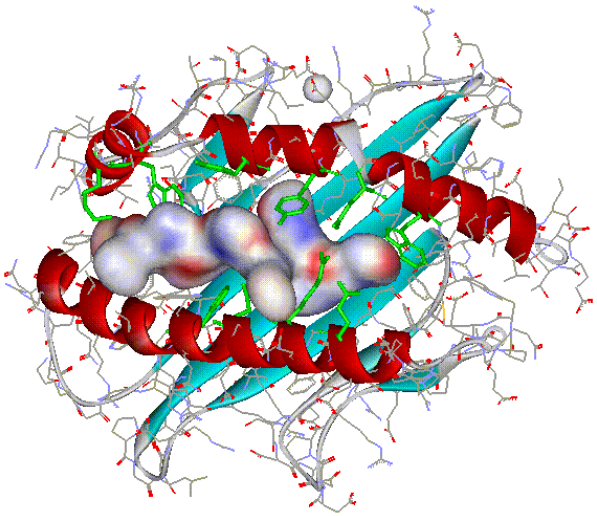




Design of MHC-I inhibitors

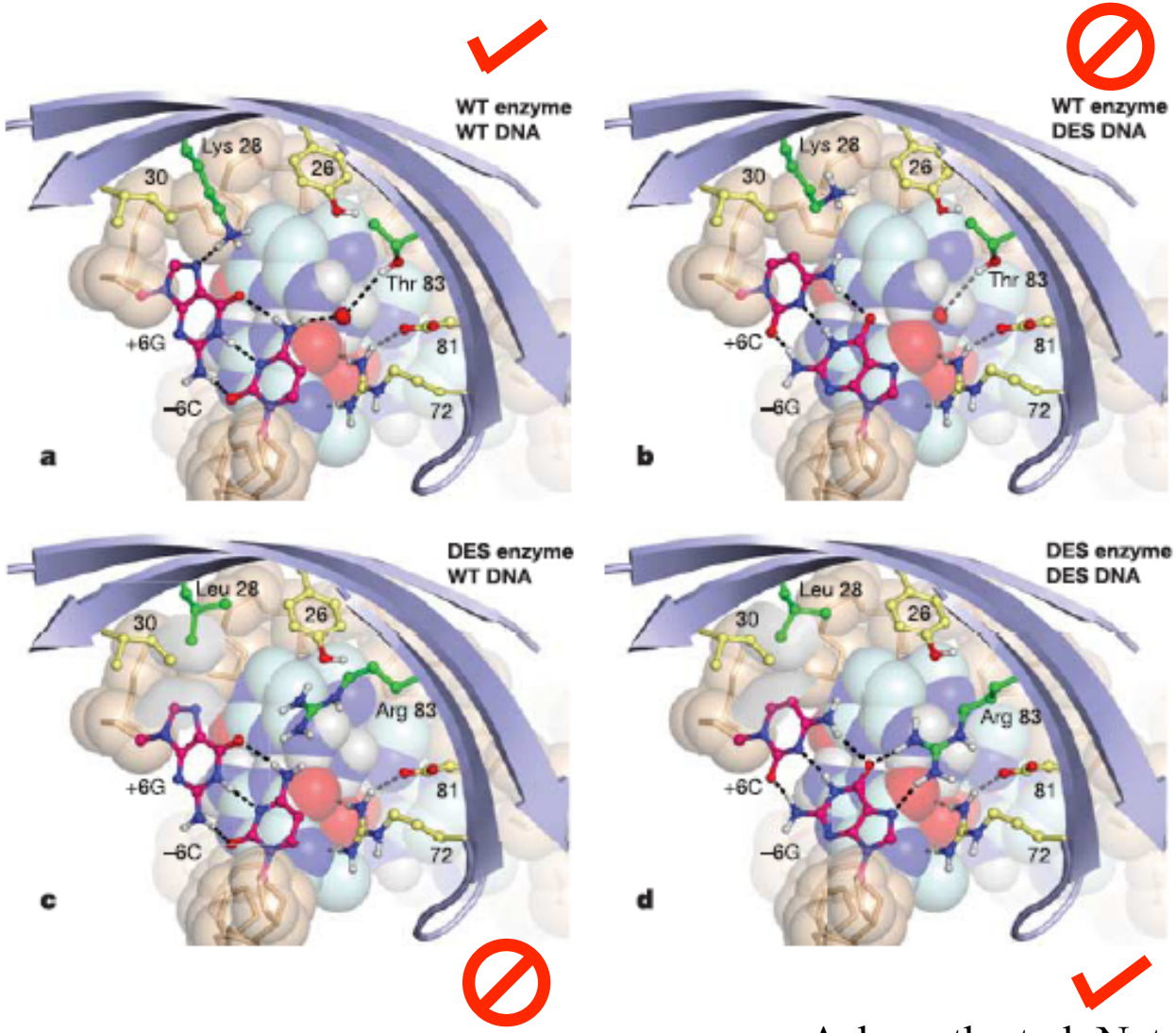
- Find sequences of 9 residues long binding to MHC- I
- Minimize the binding energy between the MHC-I and the peptide





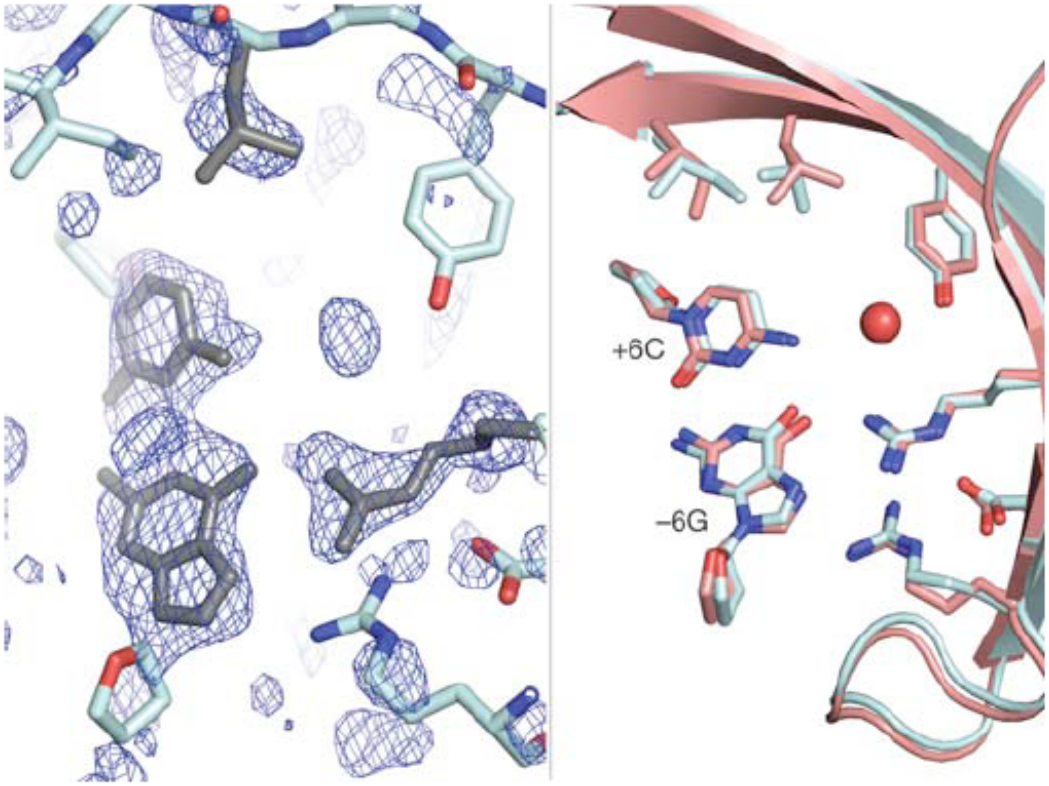
- W[illegible] 10 peptides:
  - All with binding
  - 131% of reference binding
  - Less than 55% identity with known peptides
  - 3 peptides recognized by the TCR

Computational Redesign of Endonuclease



Ashworth et al. Nature 2006

Crystal Structure of the DES Enzyme-DNA Complex



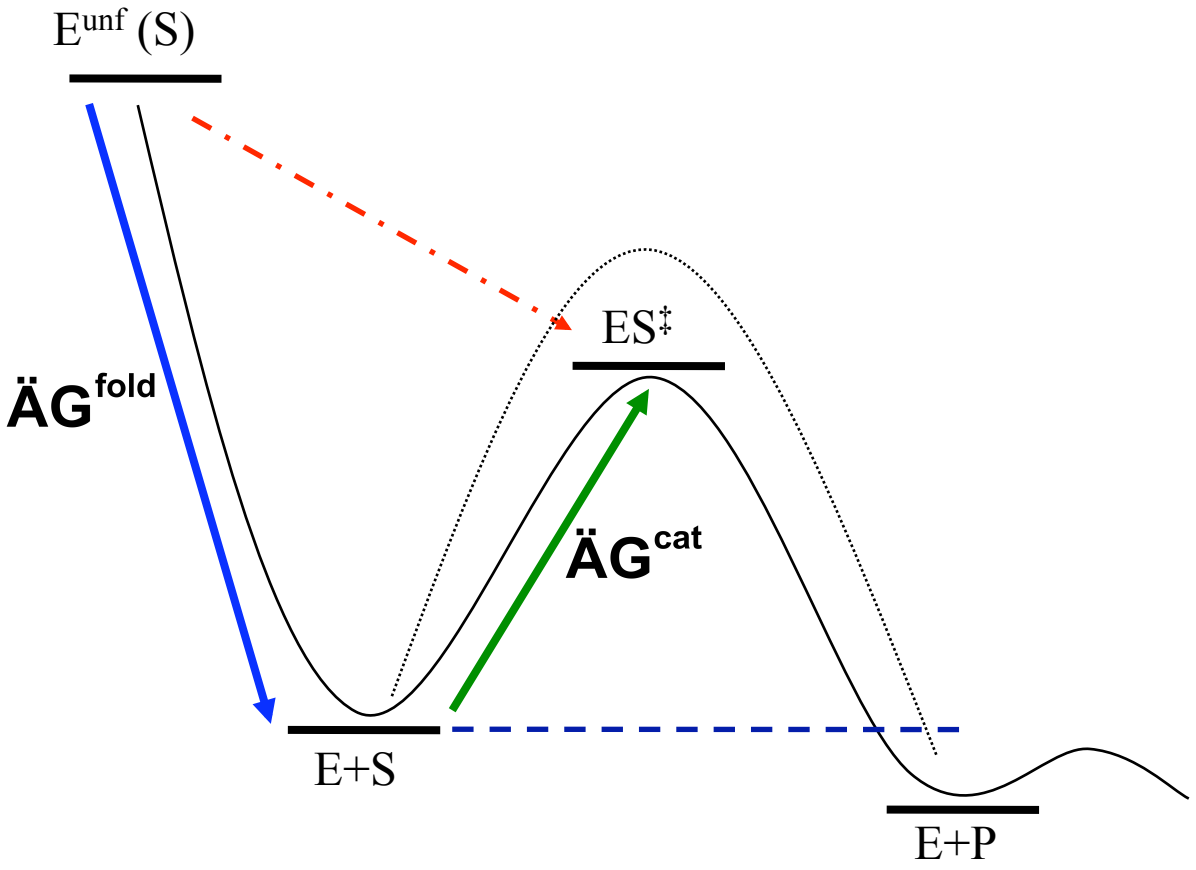
Electro-density map of the redesigned region:
gray: computational designed model

Superposition:
salmon: design model
cyan: crystal structure

De Novo Design of Novel Enzymes

Catalytic site design

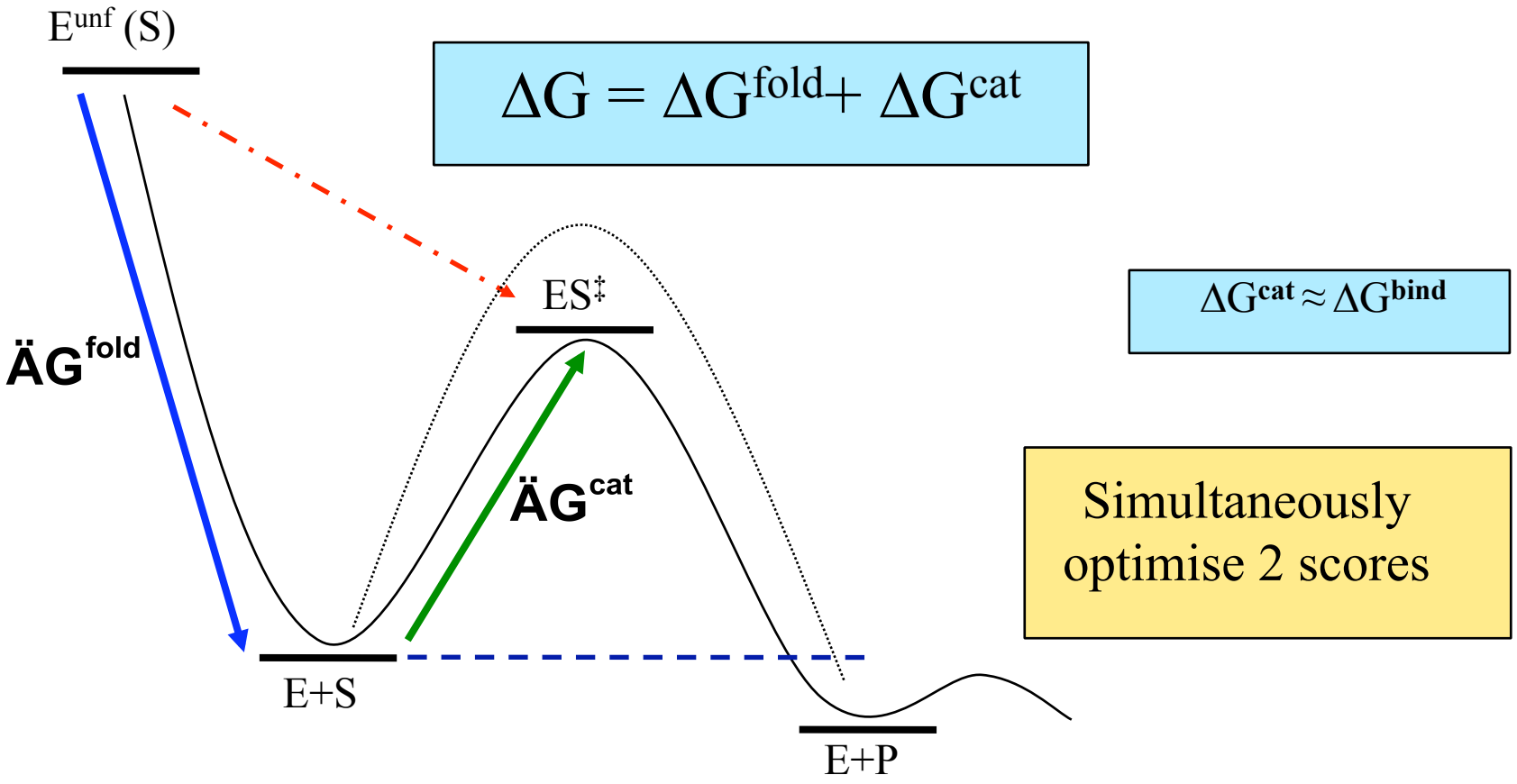
We want to lower ΔG^{cat} by minimising the binding energy to 3D model of ES‡



$\Delta G^{cat} \approx \Delta G^{bind}$

Catalytic site design

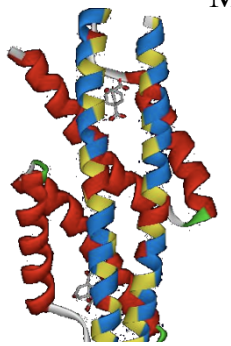
- Find sequence that both folds AND has activity: two-objective problem

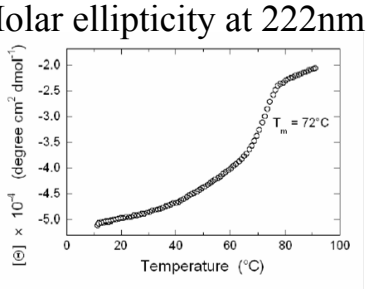


Experimental validation of min. energy designs

- Thermostable choristmate mutase with restricted AI/KE alphabet

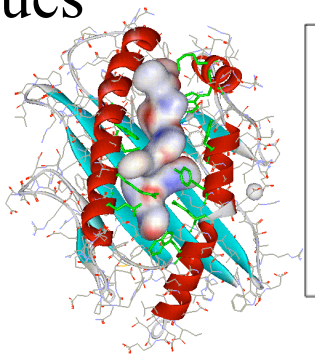
  Coll. Profs. Hilvert (ETH-Zurich), Wodak (Toronto) & Karplus (Harvard &ISIS)

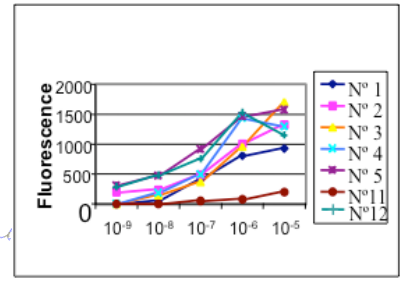




- Design of 10 peptide sequences of 9 residues long binding to MHC- I

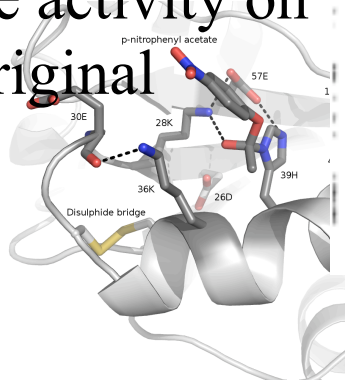
  J Biol Chem 2003

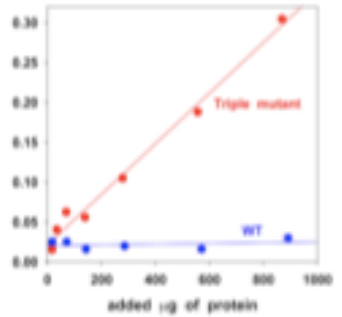




- Redesign thioredoxin by grafting esterase activity on p-nitrophenyl acetate while preserving original function. Promiscuous enzyme design.

  Coll. Prof. Sánchez-Ruiz (Granada, Spain)

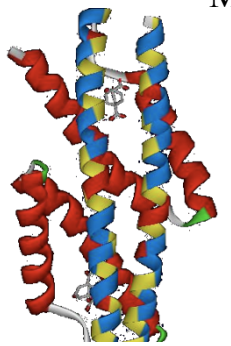




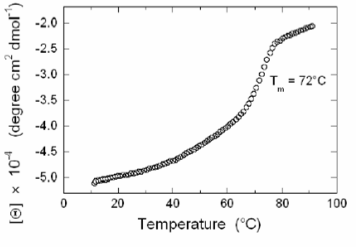
Experimental validation of min. energy designs

- Thermostable choristmate mutase with restricted AI/KE alphabet

  Coll. Profs. Hilvert (ETH-Zurich), Wodak (Toronto) & Karplus (Harvard &ISIS)

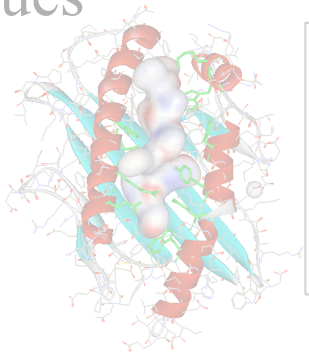


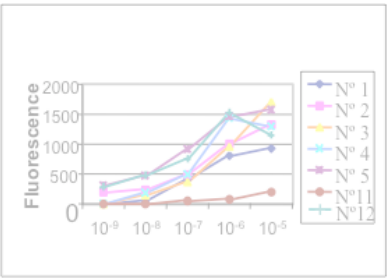
Molar ellipticity at 222nm



- Design of 10 peptide sequences of 9 residues long binding to MHC- I

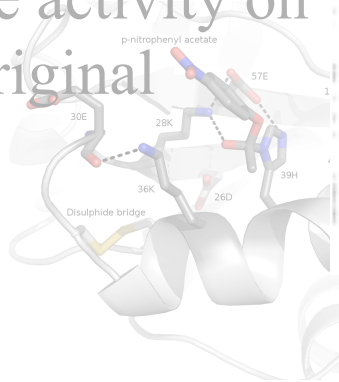
  J Biol Chem 2003

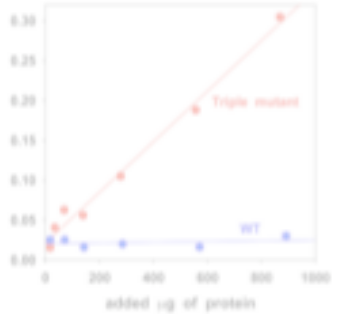




- Redesign thioredoxin by grafting esterase activity on p-nitrophenyl acetate while preserving original function. Promiscuous enzyme design.

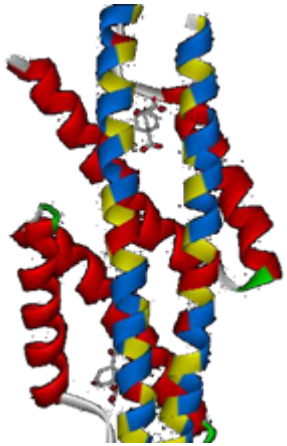
  Coll. Prof. Sánchez-Ruiz (Granada, Spain)

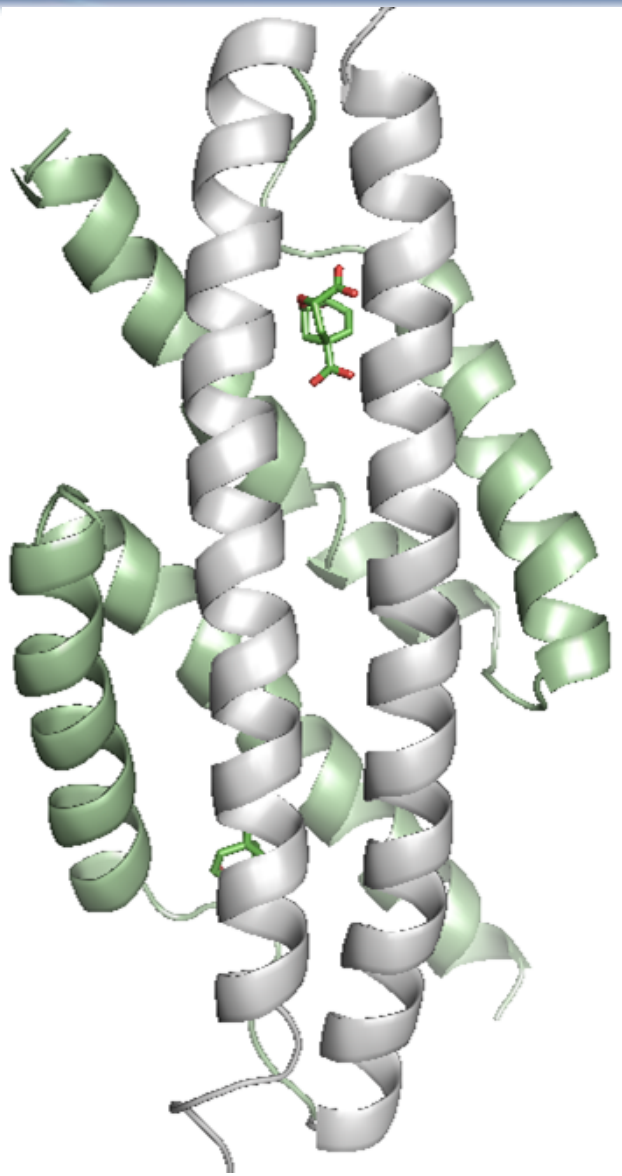




Enzyme with minimal aminoacid alphabet

Chorismate mutase with restricted AI/KE alphabet?

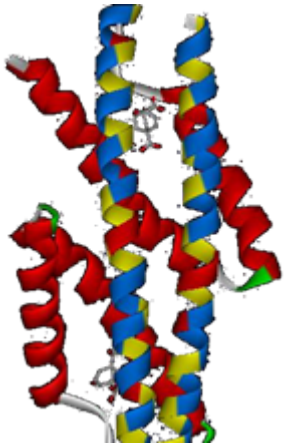


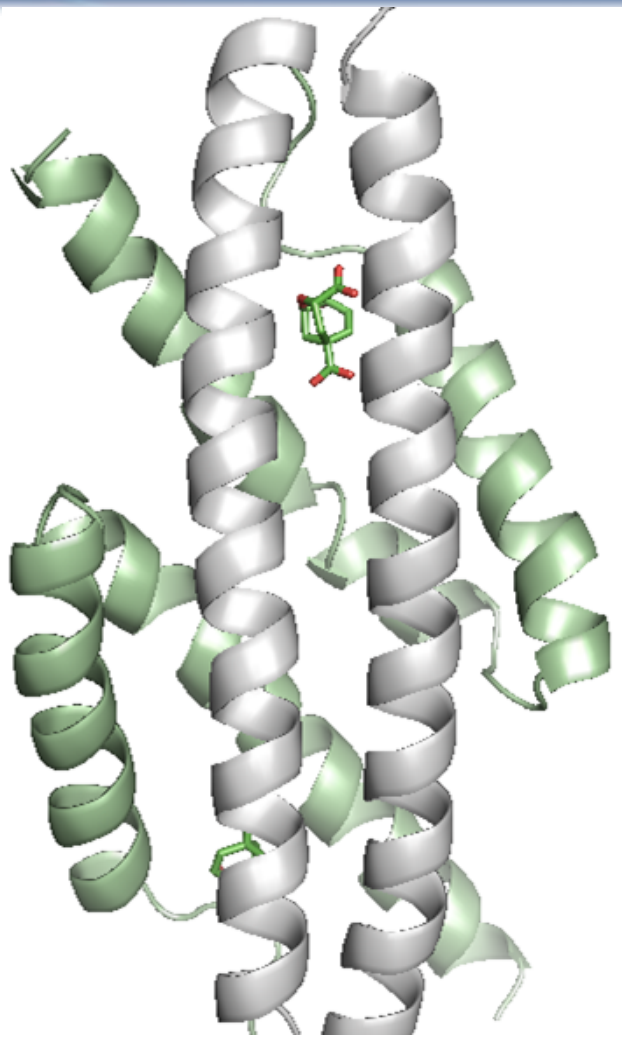


- Redesign of helix 1 using hydrophobic /hydrophilic patterning.

Enzyme with minimal aminoacid alphabet

Chorismate mutase with restricted AI/KE alphabet?



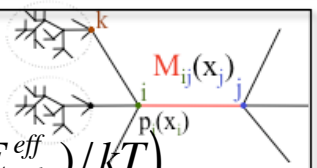


- Redesign of helix 1 using hydrophobic /hydrophilic patterning.
- We did two parallel experiments:
  - Computational protein design → **Coll. Wodak (Toronto) & Karplus (Harvard & ISIS)**
  - *In vivo* directed evolution → **Coll. Profs. Hilvert (ETH-Zurich)**

Methodology Computational Protein design

State space: $\{x_i\}$ = Aminoacid side-chain at residue i

Mathematical description: **Calculation of folding free energy**

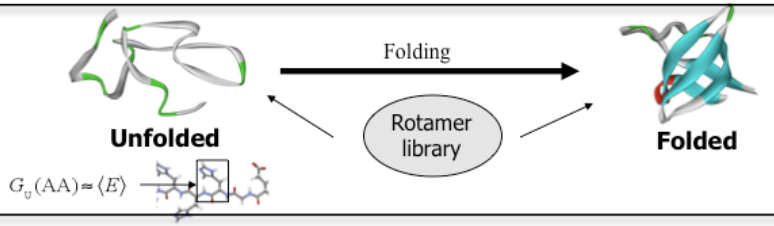


$$\exp(-G/kT) = \int d(solute)d(solvent)\exp(-E/kT) = \int d(solute)\exp\left(-(E + E^{eff}_{solv})/kT\right)$$

$M_{ij}(x_j) = \sum_{x_i} \exp(-E_{ij}^{pair}(x_i,x_j)/RT)\, \exp(-E_i^{sing}(x_i)/RT) \prod_{k\neq j} M_{ki}(x_i)$ **Belief propagation**

Objective function: $\Delta G_{U\to F} = G_F - G_U$

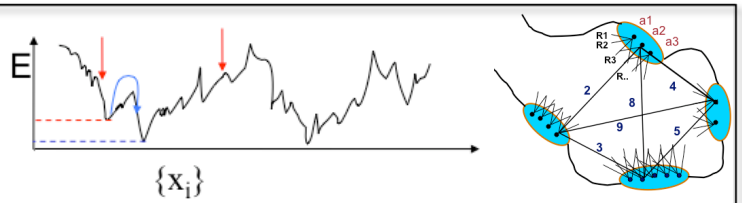
$G=\sum_i \sum_{x_i} G_i^{sing}(x_i) + \sum_{i<j} \sum_{x_i x_j} G_{ij}^{pair}(x_i x_j)$



Optimisation: **Inverse folding problem**

Heuristic: Monte Carlo Simulated Annealing

Exact: Branch & Bound



Challenges:

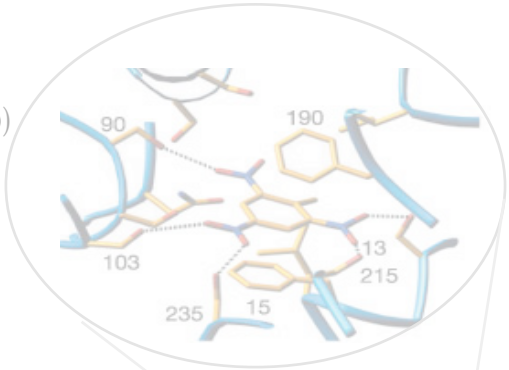
NP-Hard problem of large size: 10^{200}, new approaches are needed

We use a physical model with almost no fitted parameters

Multi-scale computational design

Macromolecules

$$E = \sum_{bonds} K_b(b-b_o)^2 + \sum_{angles} K_\theta(\theta-\theta_o)^2 + \sum_{torsions} K\phi(1+\cos(n\phi-\delta))$$
$$+ \sum_{impropers} K_\varphi(\varphi-\varphi_o)^2 + \sum_{Urey-Bradley} K_{UB}(r_{1,3}-r_{1,3,o})^2 +$$
$$+ \sum_{nonbonded} \frac{q_i q_j}{4\pi D r_{ij}} + \varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right]$$



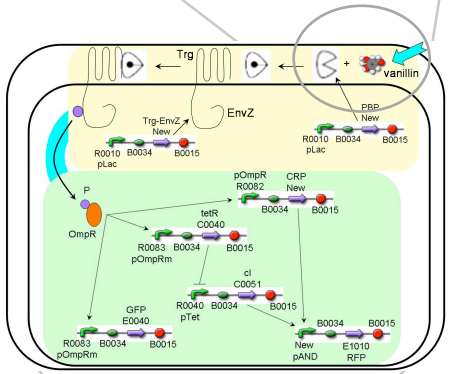
De novo design of proteins:

- DESIGNER
- PROTDES

Biological networks

$$\frac{d[Y]}{dt} = \alpha\frac{1}{1+(\frac{[U]}{K})^n} - \beta[Y] + \gamma$$

$$\frac{d[Y]}{dt} = \alpha\frac{(\frac{[U]}{K})^n}{1+(\frac{[U]}{K})^n} - \beta[Y] + \gamma$$



De novo design of:

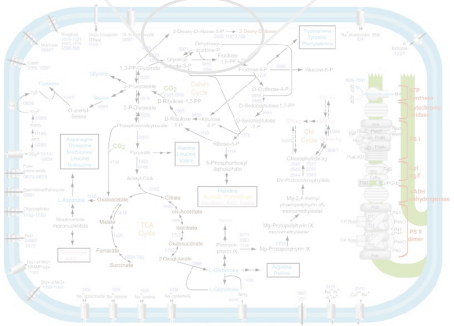
- Transcriptional networks
  - GENETDES
  - ASMPARTS
- RNA networks
  - RNADES

Genomic background

$$\max \mu = cv$$

$$\text{st} : Sv = b$$

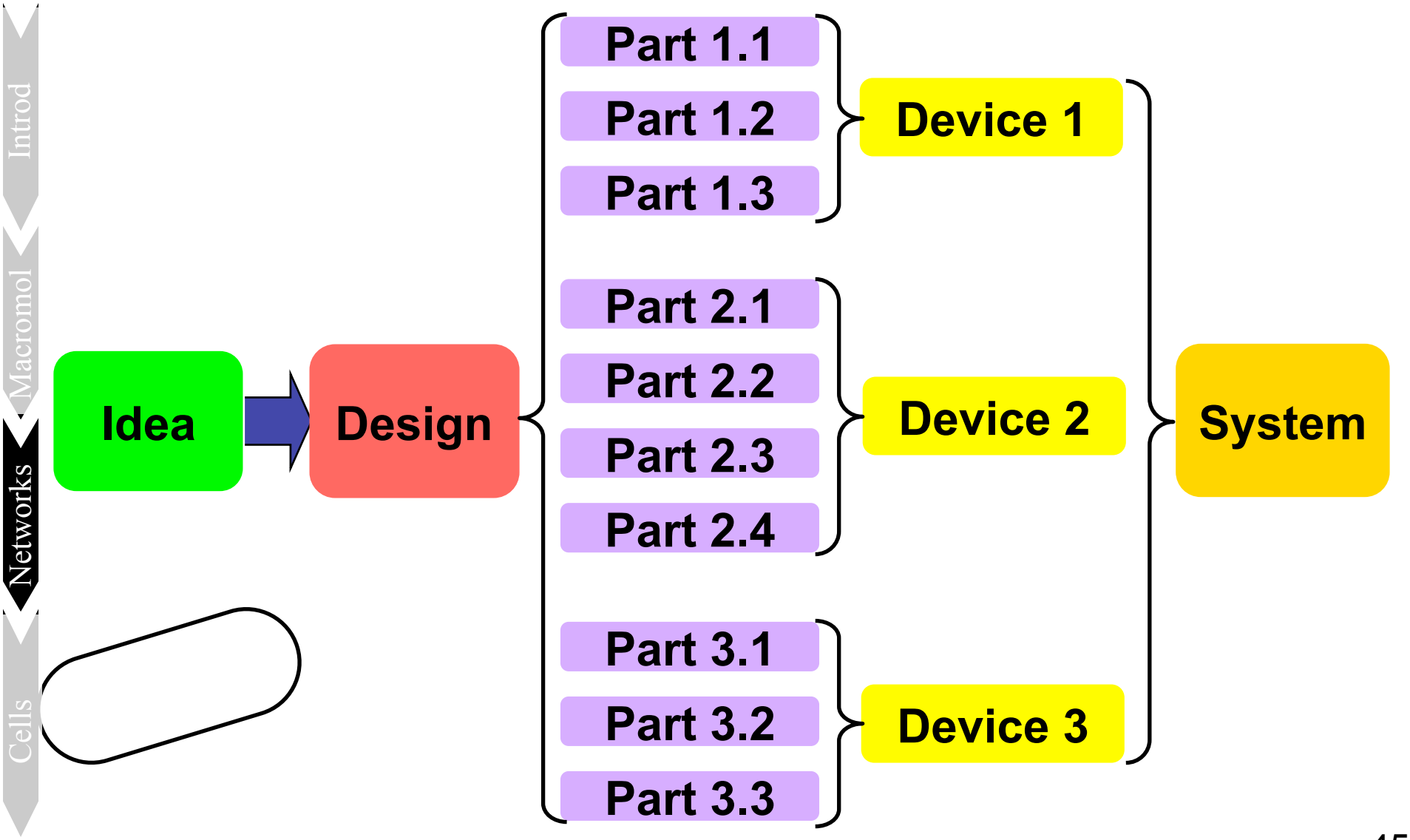
$$v_{\min} \le v \le v_{\max}$$



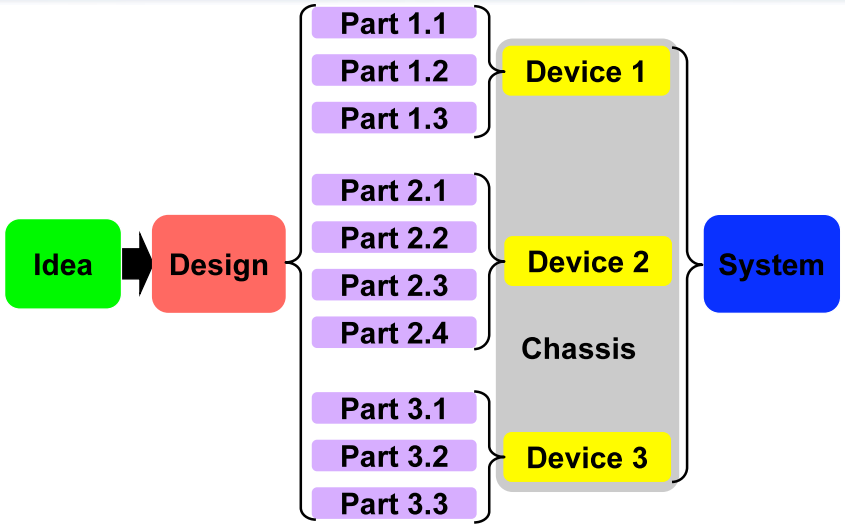
- *De novo* design of metabolic pathways by retro-biosynthesis
  - DESHARKY
- Network inference from microarray & proteomics resp.
  - INFERGENE

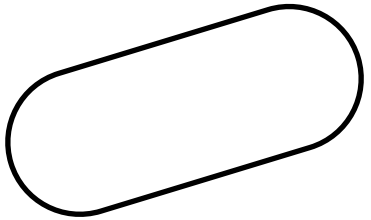
Can we design protein networks with targeted behaviour?

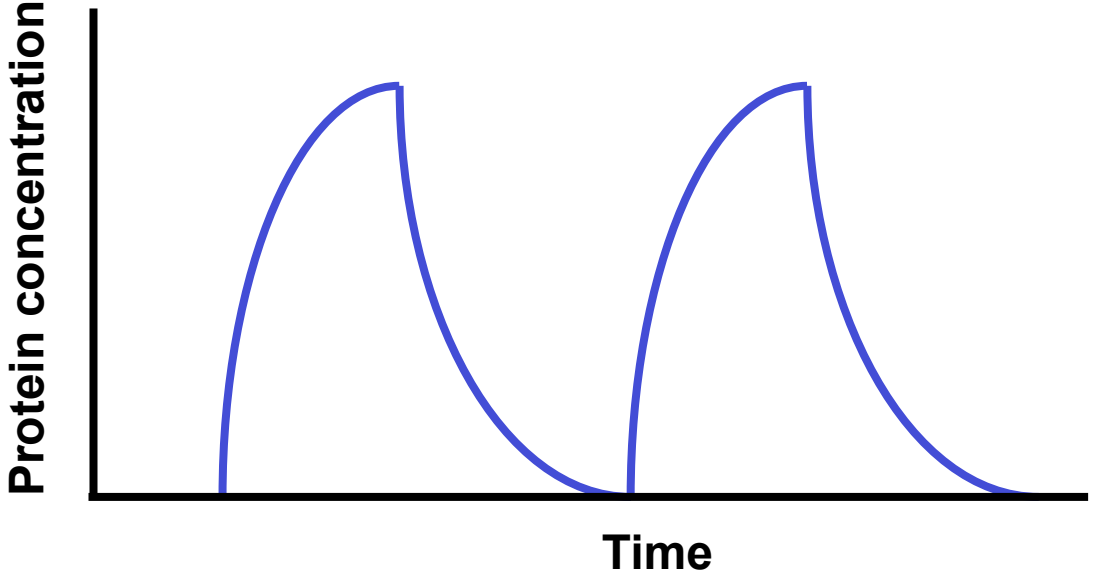
Building a bacterial blinker



Building a bacterial blinker

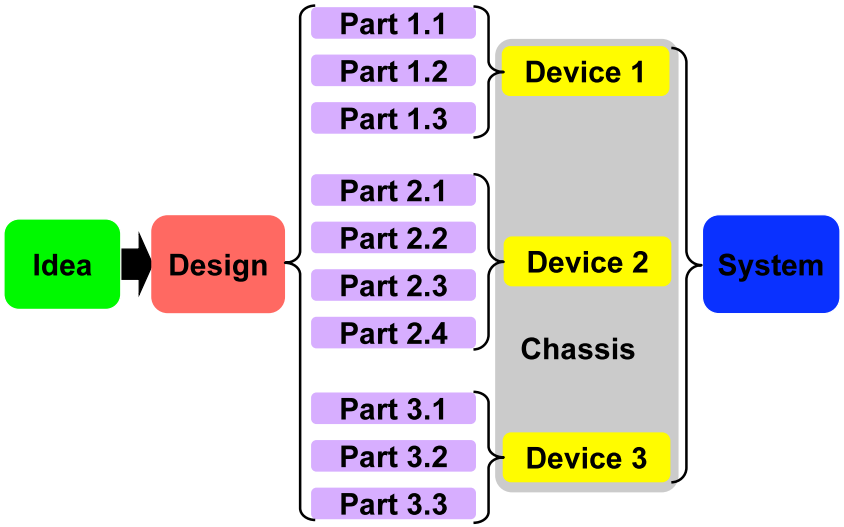


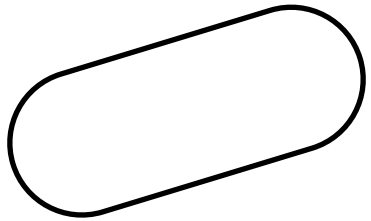


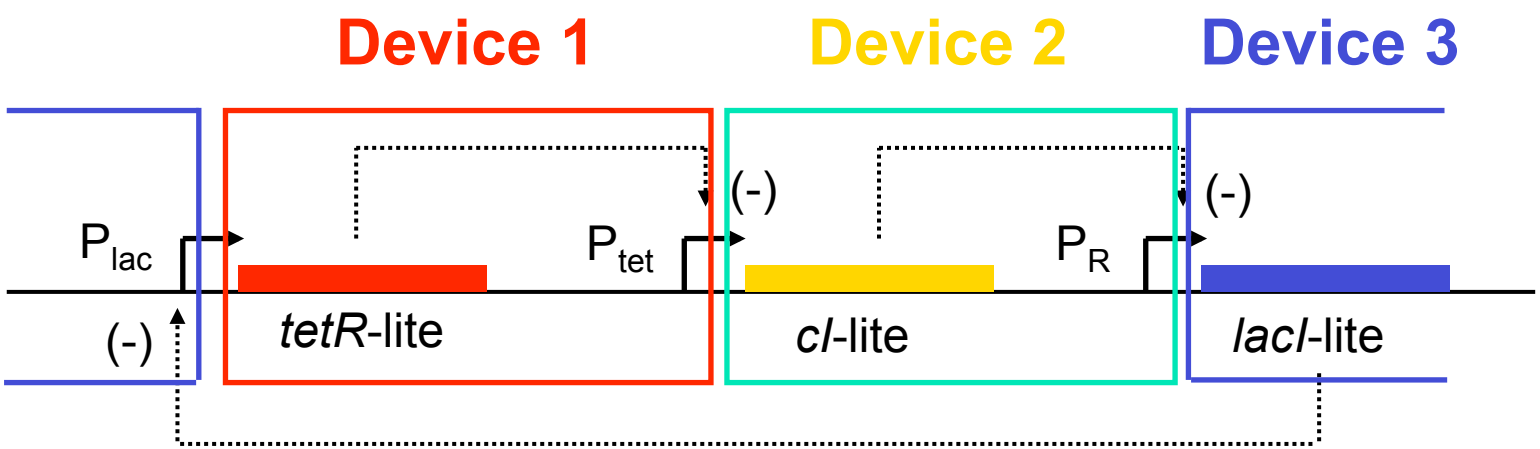


Elowitz & Leibler. 2000. *Nature* **403**:335-8

Building a bacterial blinker

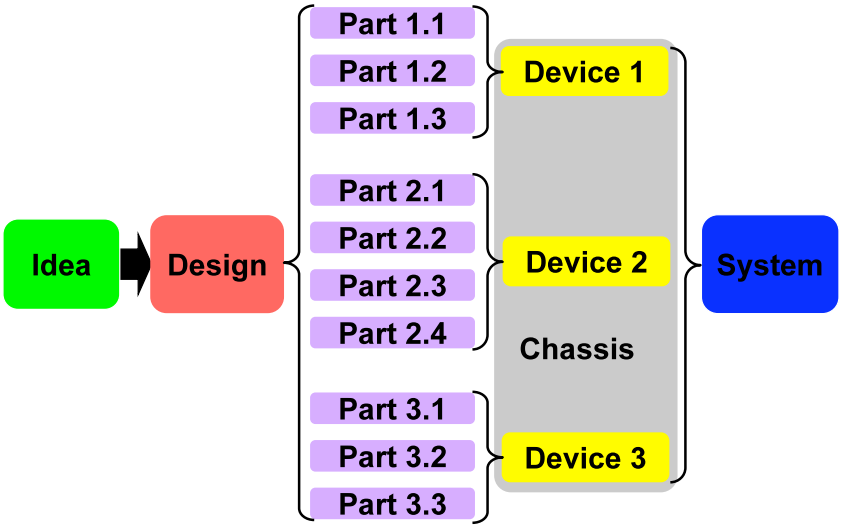


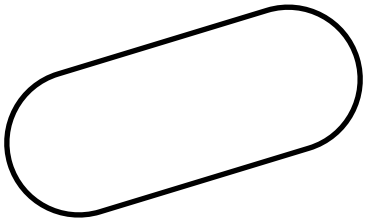


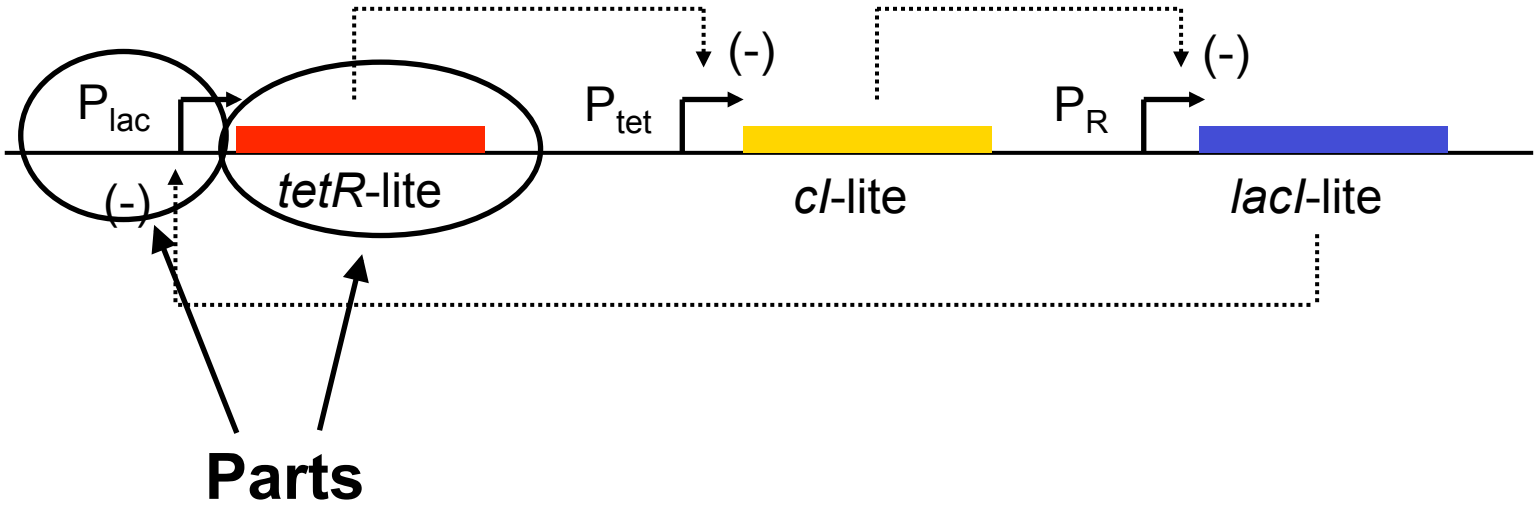


Elowitz & Leibler. 2000. *Nature* **403**:335-8

Building a bacterial blinker

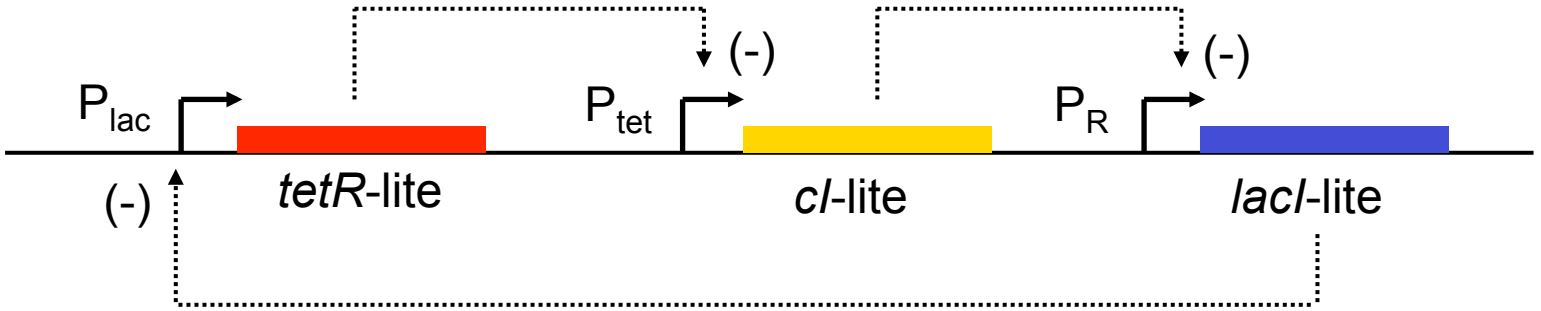


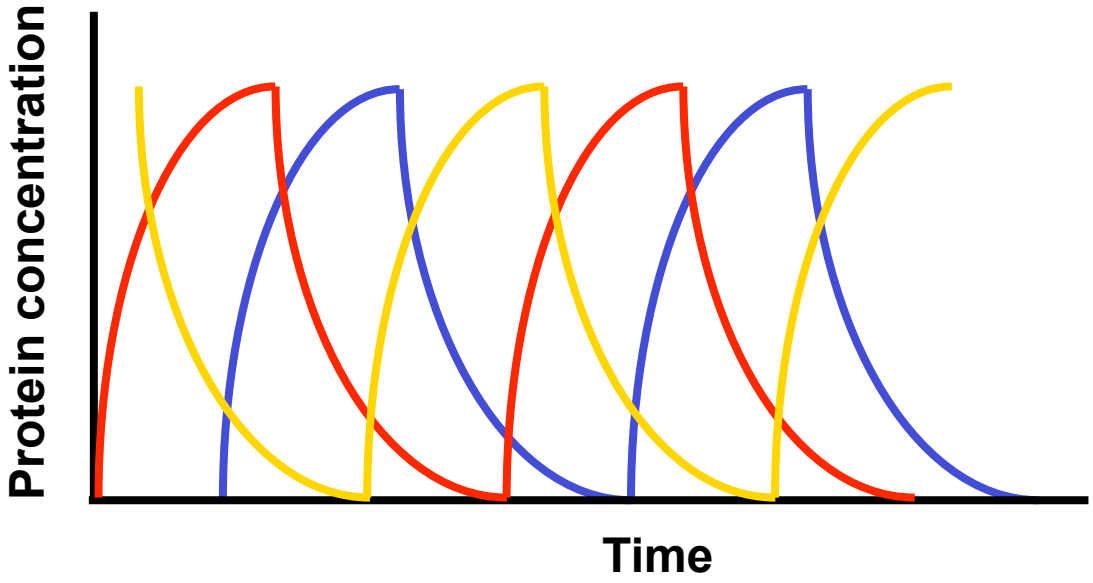




Elowitz & Leibler. 2000. *Nature* **403**:335-8

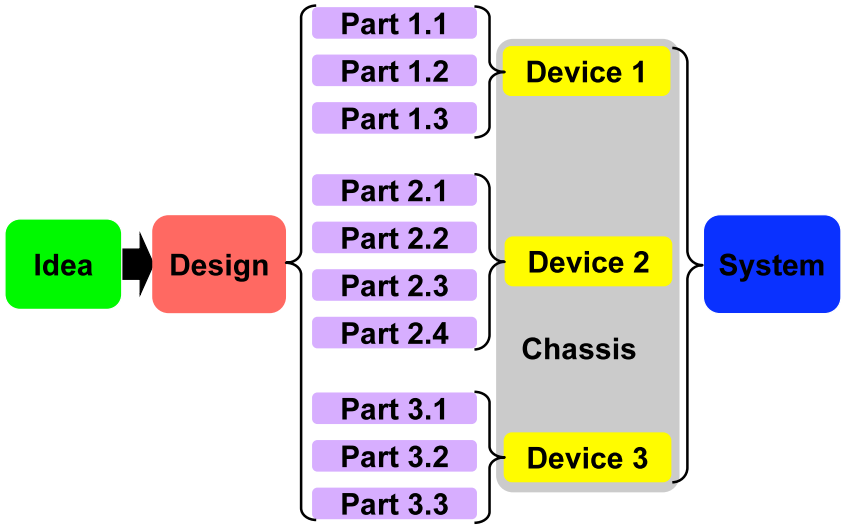
Building a bacterial blinker

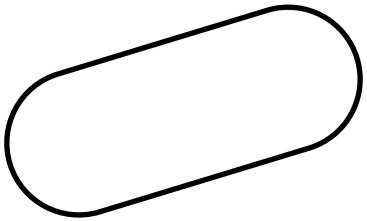


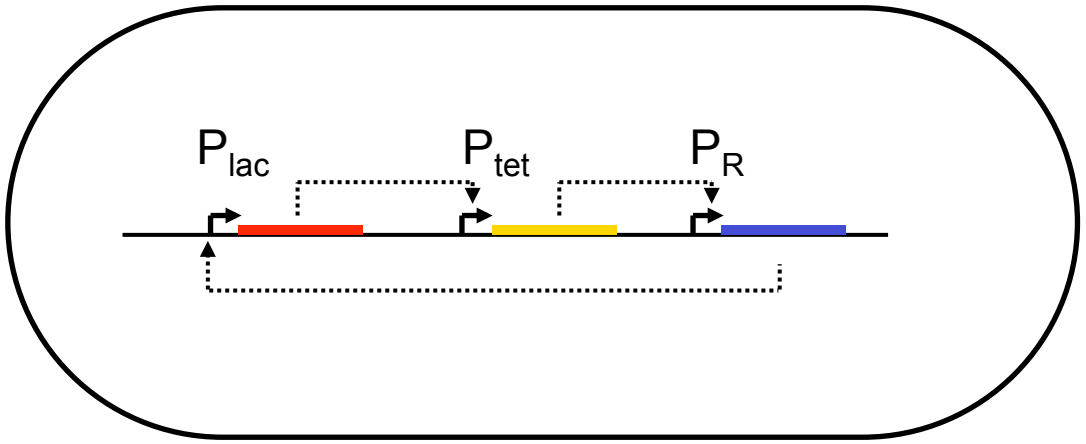


Elowitz & Leibler. 2000. *Nature* **403**:335-8

Building a bacterial blinker



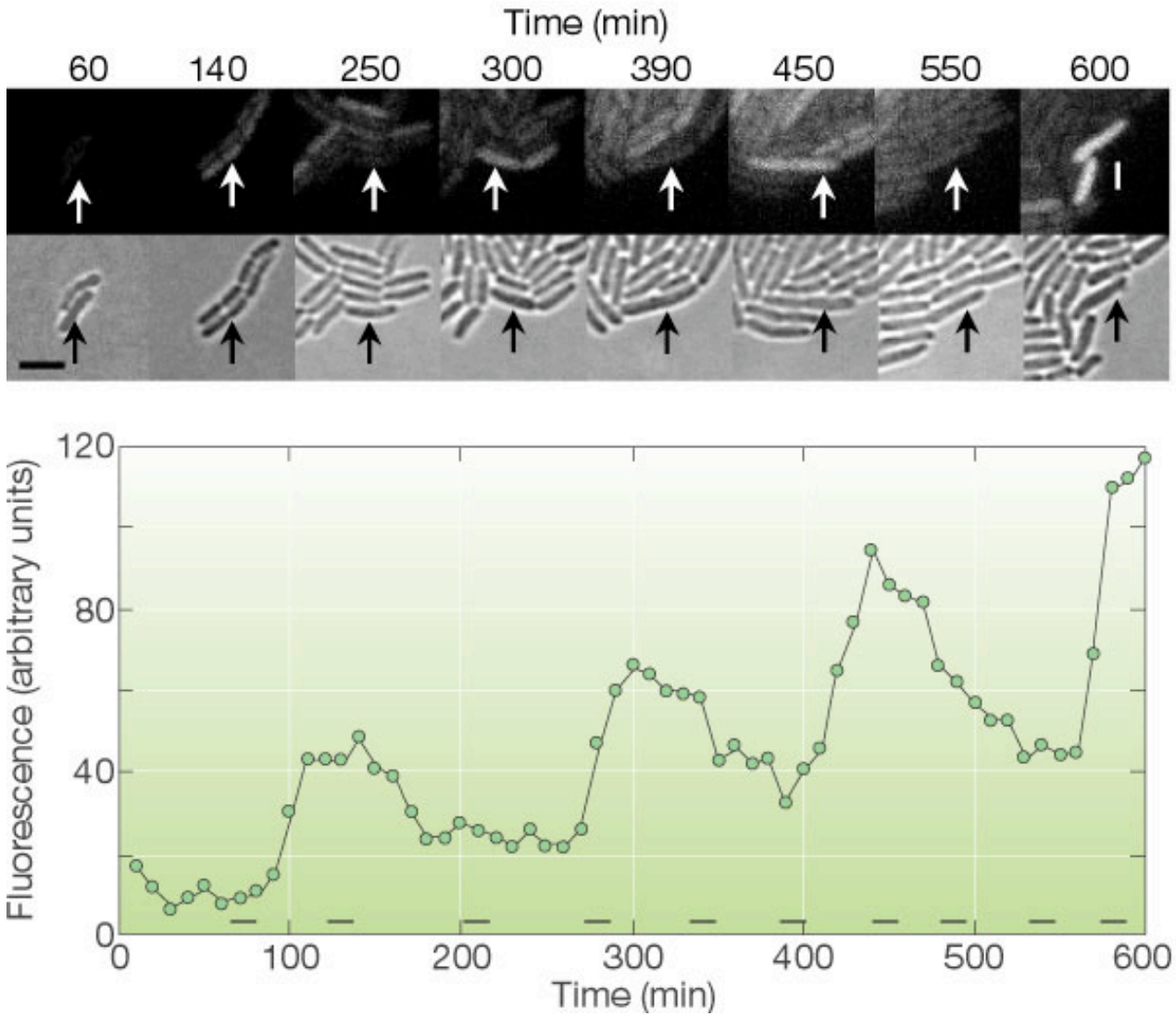




E. coli as the Chassis

Elowitz & Leibler. 2000. *Nature* **403**:335-8

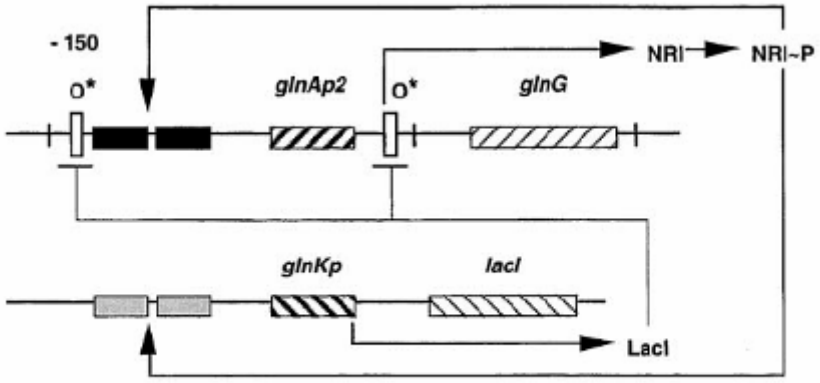
Building a bacterial blinker: repressilator

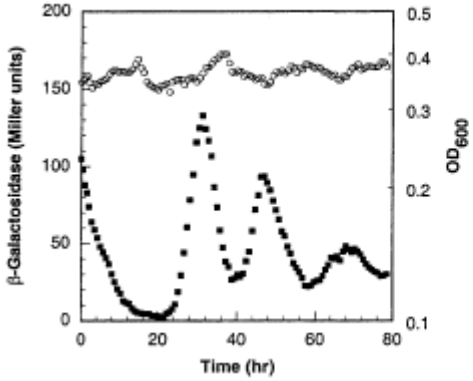


Elowitz & Leibler. 2000. *Nature* **403**:335-8

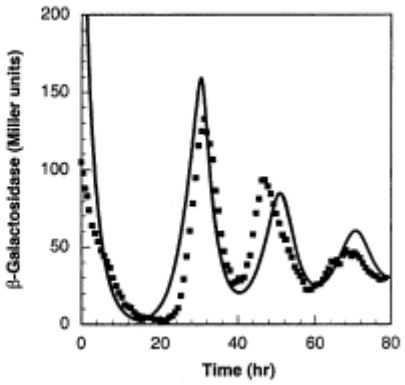
Atkinson oscillator

Development of Genetic Circuitry Exhibiting
Toggle Switch or Oscillatory Behavior
in Escherichia coli
Mariette R. Atkinson et al. Cell, Vol. 113, 597–607, May 30, 2003,

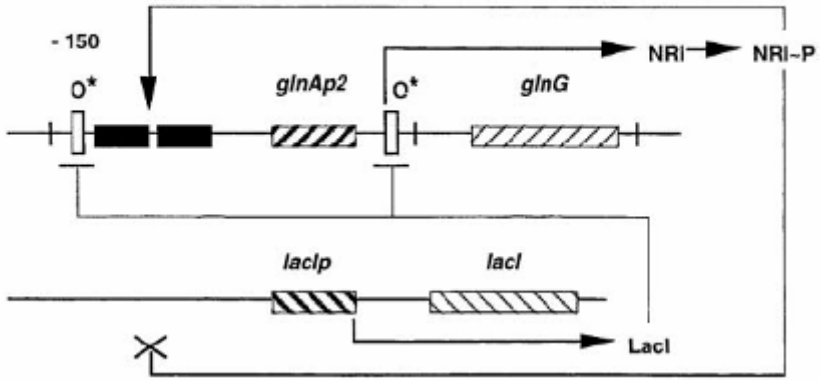


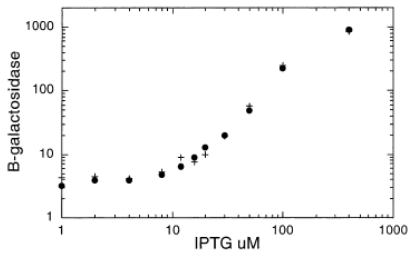


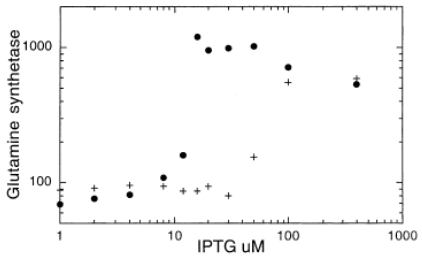
C



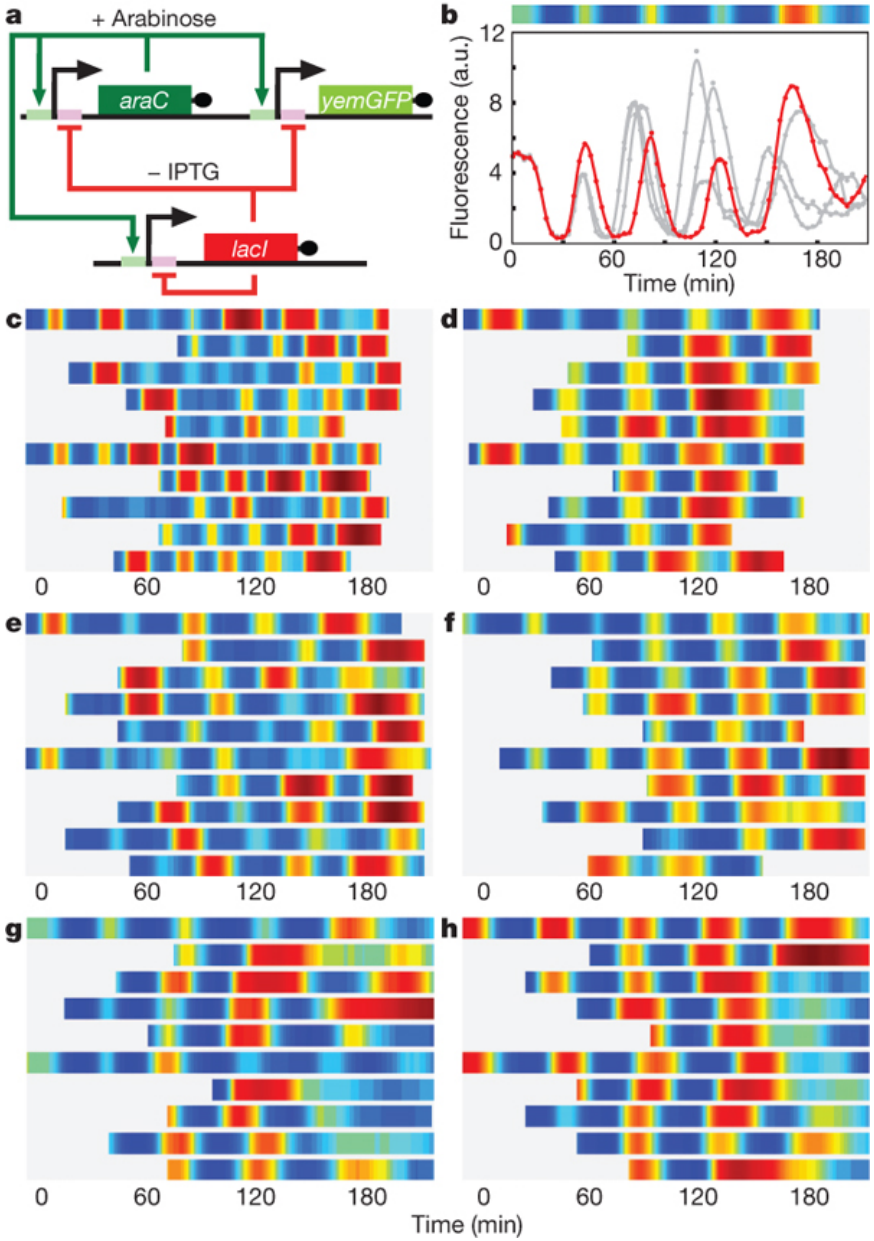
B







Hasty oscillator



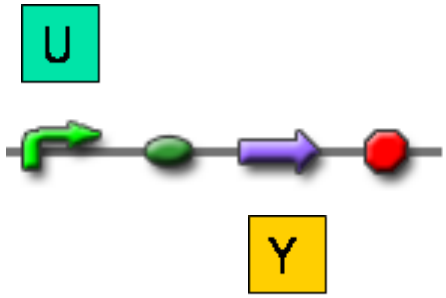
J Stricker *et al. Nature*, (2008) doi:10.1038/nature07389

nature

Can we design such protein networks in an automated way?

Automatic design gene networks

- We are going to use a coarse-grained description at the protein level
- Focus on transcription regulation
- Combinatorial optimisation

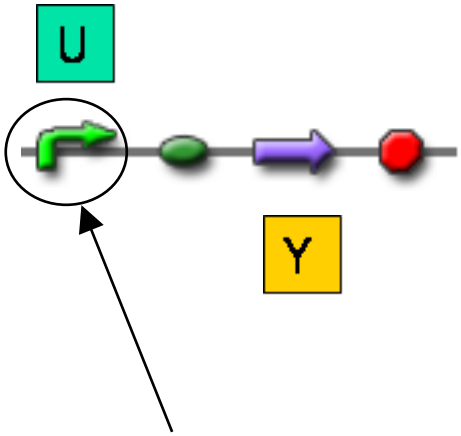


$$\frac{d[Y]}{dt} = \alpha \frac{1}{1 + (\frac{[U]}{K})^n} - \beta[Y] + \gamma$$ Repression

$$\frac{d[Y]}{dt} = \alpha \frac{(\frac{[U]}{K})^n}{1 + (\frac{[U]}{K})^n} - \beta[Y] + \gamma$$ Activation

Methodology to design gene networks

- We are going to use a coarse-grained description at the protein level
- Focus on transcription regulation
- Combinatorial optimisation



$$\frac{d[Y]}{dt} = \alpha \frac{1}{1 + (\frac{[U]}{K})^n} - \beta[Y] + \gamma$$ Repression

$$\frac{d[Y]}{dt} = \alpha \frac{(\frac{[U]}{K})^n}{1 + (\frac{[U]}{K})^n} - \beta[Y] + \gamma$$ Activation

TYPE	REGULATORY LOGIC	REGULATORY FUNCTION	PROMOTER WITH OPERATORS
0	-	1	CONSTITUTIVE
1 NOT	A → out	R^r_A	OR_A
2 YES	A → out	R^a_A	OA_A
3 NOR	A, B → out	$R^r_A R^r_B$	OR_B OR_A
4 XNOR	A, B → out	$1-R^r_A-R^r_B+2R^r_A R^r_B$	OR_B OR_A
5 NAND	A, B → out	$R^r_A+R^r_B-R^r_A R^r_B$	OR_B OR_A
6 AND	A, B → out	$R^a_A R^a_B$	OA_B OA_A

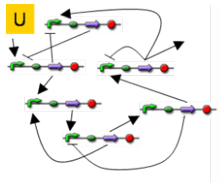
$$R_{r_X} = \frac{1}{1+([X]/K)^n}$$

$$R_{a_X} = \frac{([X]/K)^n}{1+([X]/K)^n}$$

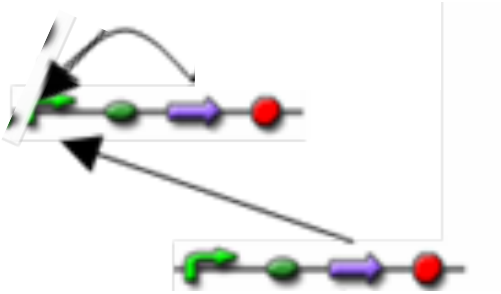
▭ = Induction operator site

▬ = Repression operator site

GENETDES software



$$\begin{cases} \dfrac{d[Y]}{dt} = \alpha \dfrac{1}{1+(\frac{[U]}{K})^n} - \beta[Y] + \gamma \\ \vdots \\ \dfrac{d[Y]}{dt} = \alpha \dfrac{(\frac{[U]}{K})^n}{1+(\frac{[U]}{K})^n} - \beta[Y] + \gamma \end{cases}$$

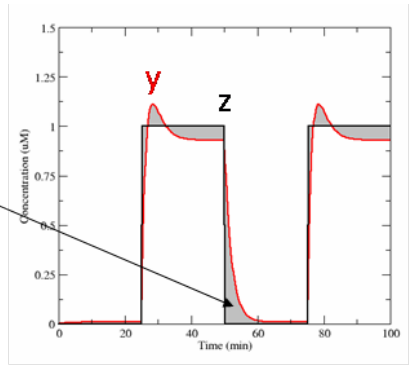


GENETDES software

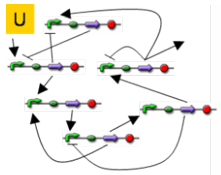
Rodrigo et al.
Bioinformatics 2007

Evolve and optimise network using a targeted time-course to construct a **score**

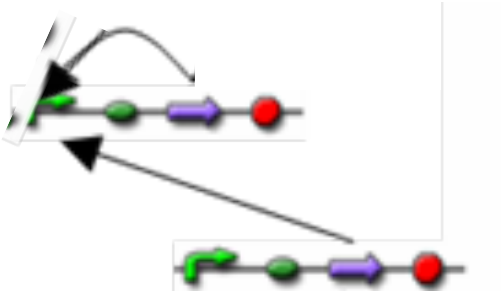
$$J = \frac{\int_{t_i}^{t_f} |y - z| \; dt}{\int_{t_i}^{t_f} z \; dt}$$



GENETDES software



$$\begin{cases} \dfrac{d[Y]}{dt} = \alpha \dfrac{1}{1+(\frac{[U]}{K})^n} - \beta[Y] + \gamma \\ \vdots \\ \dfrac{d[Y]}{dt} = \alpha \dfrac{(\frac{[U]}{K})^n}{1+(\frac{[U]}{K})^n} - \beta[Y] + \gamma \end{cases}$$

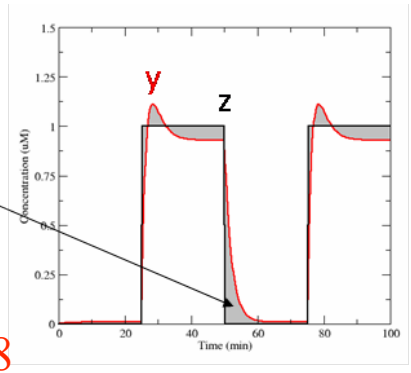


GENETDES software

Rodrigo et al.
Bioinformatics 2007

Evolve and optimise network using a targeted time-course to construct a **score**

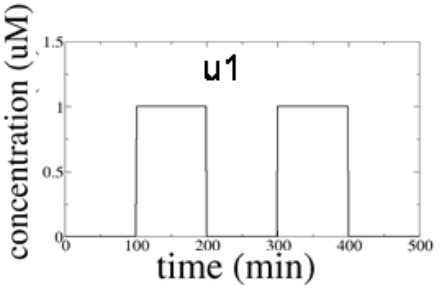
$$J = \frac{\int_{t_i}^{t_f} |y - z| \; dt}{\int_{t_i}^{t_f} z \; dt}$$

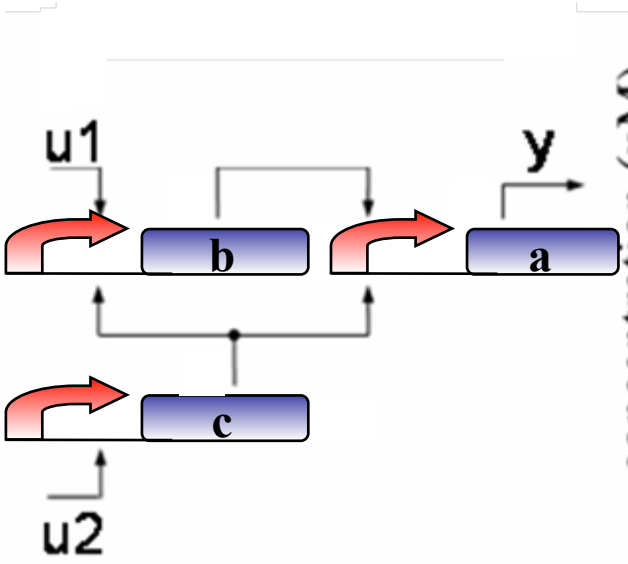


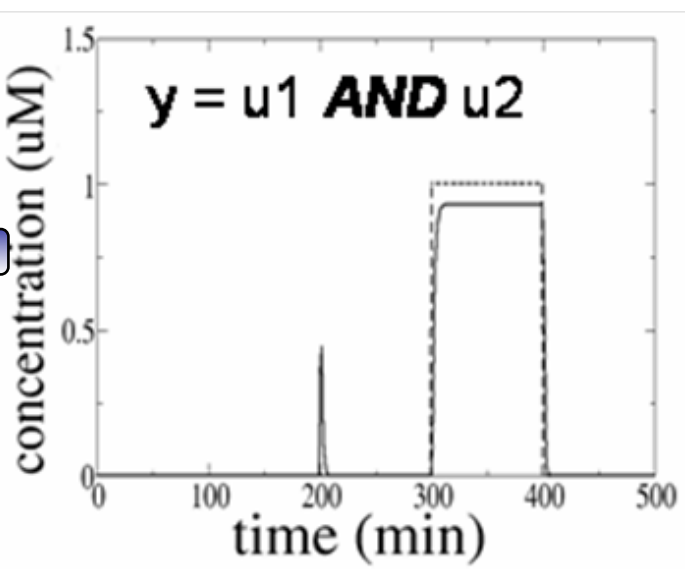
Extend to combinatorial assembly of SBML models

Rodrigo et al. J. Syst. & Synth. Biol. 2008

Transcriptional networks as logic gates

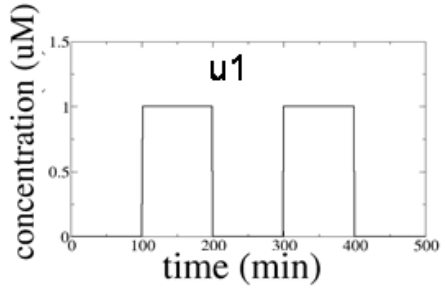


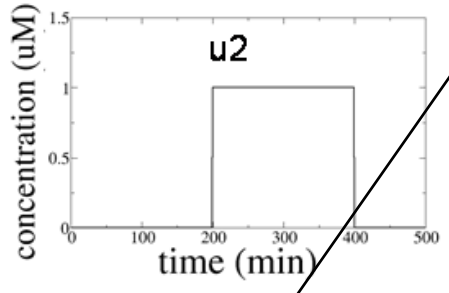


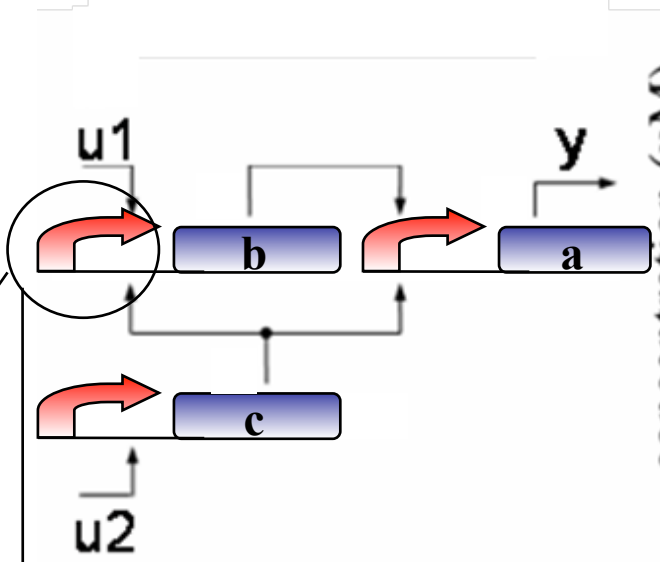


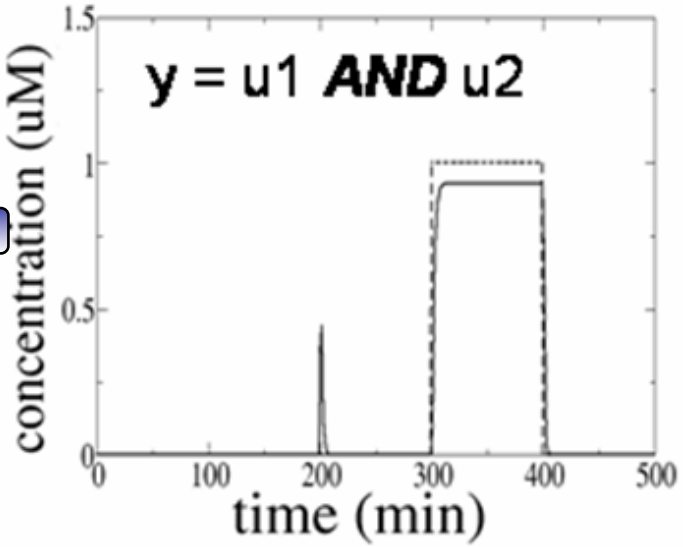
Input u1	Input u2	Output y
0	0	0
0	1	0
1	0	0
1	1	1

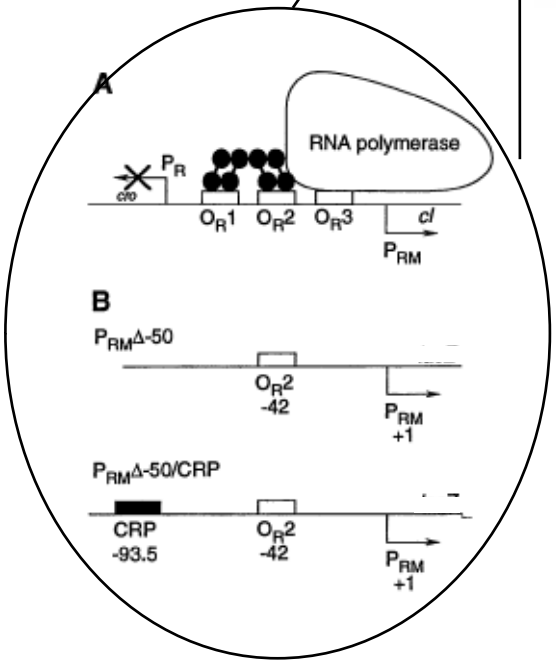
Transcriptional networks as logic gates

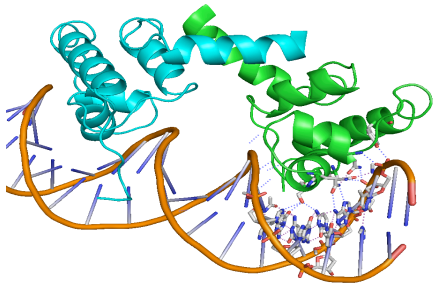




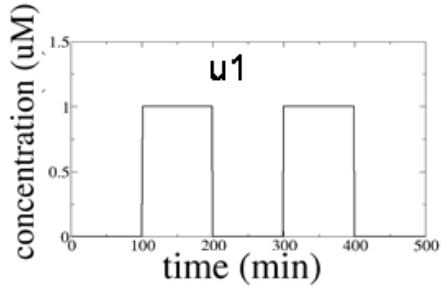


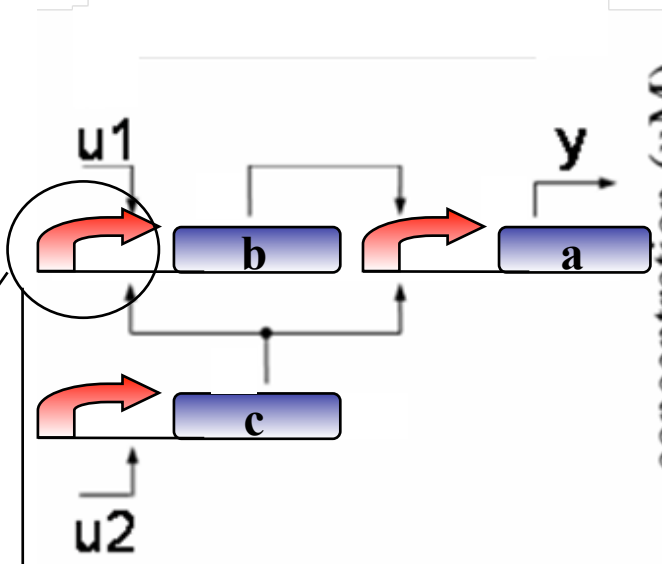


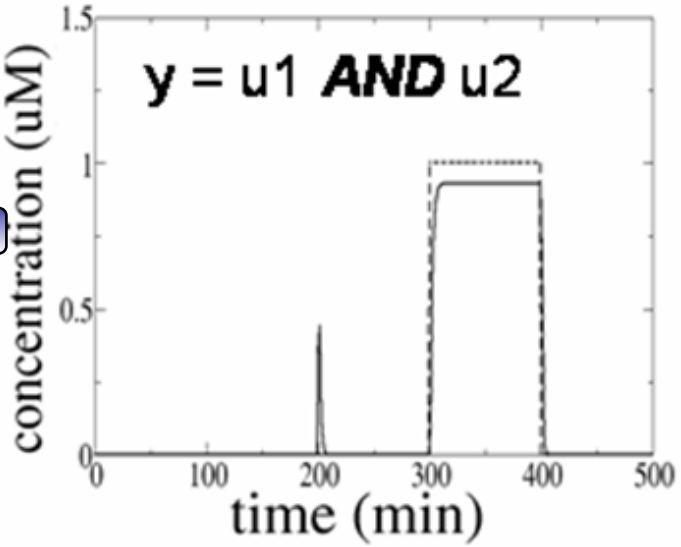


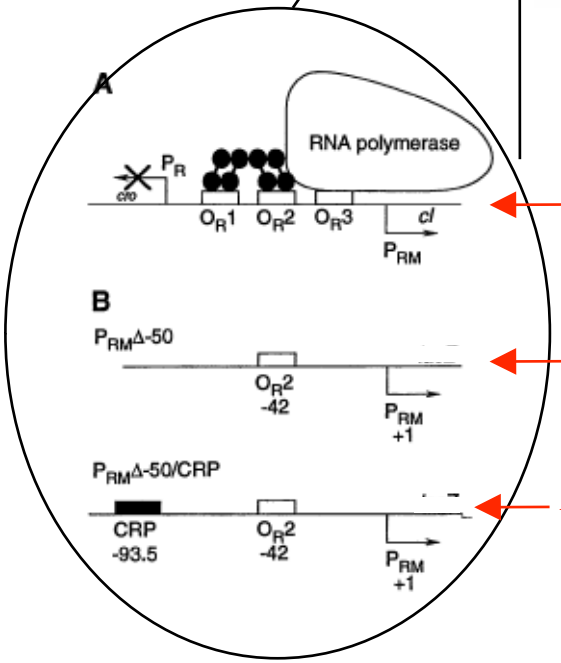


Transcriptional networks as logic gates





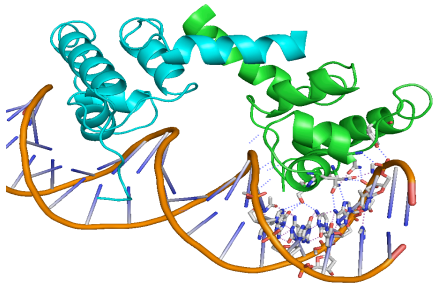




Lambda phage bidirectional promoter

Removed P_R (upstream -50) & O_R1 O_R3

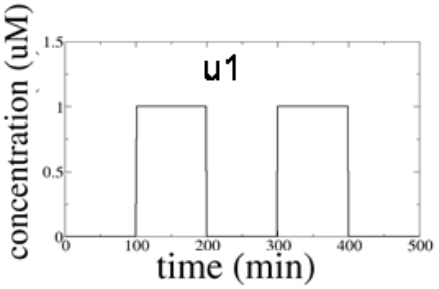
Added operator for CRP from consensus

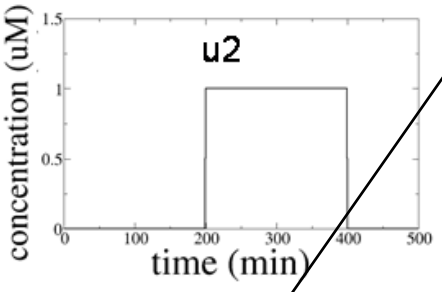


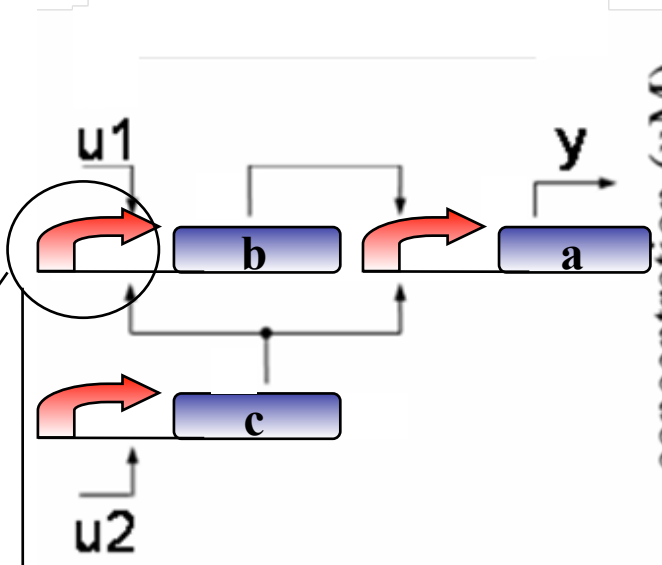
[Protocol by Joung et al. , Science 1994]

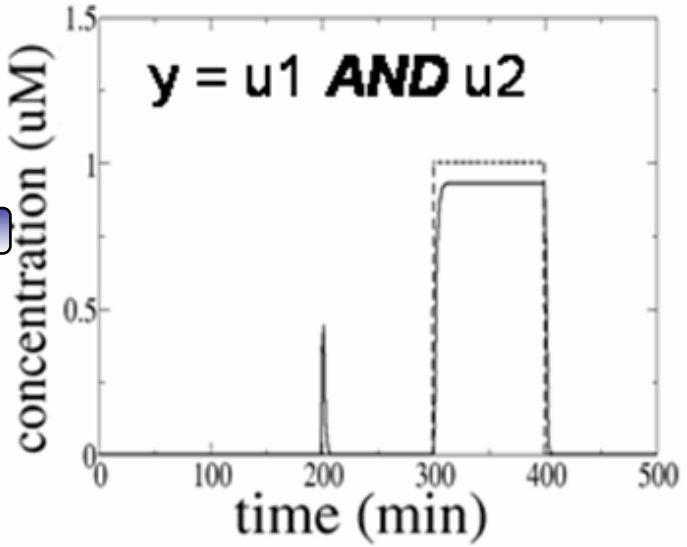
Transcriptional networks as logic gates

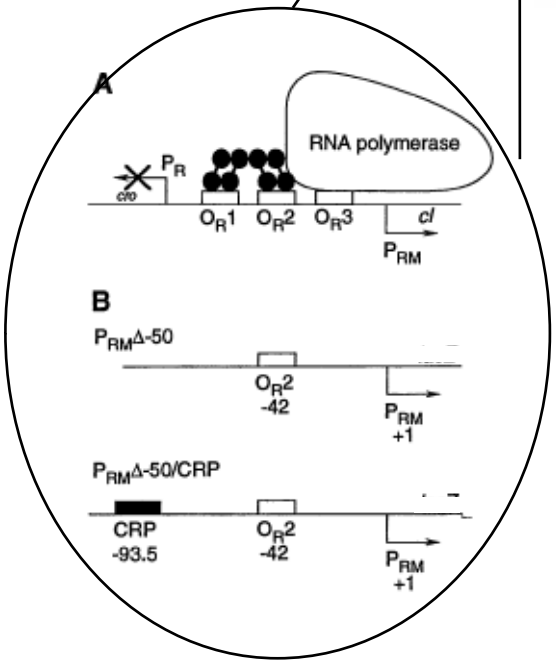
Introd
Macromol
Networks
Cells

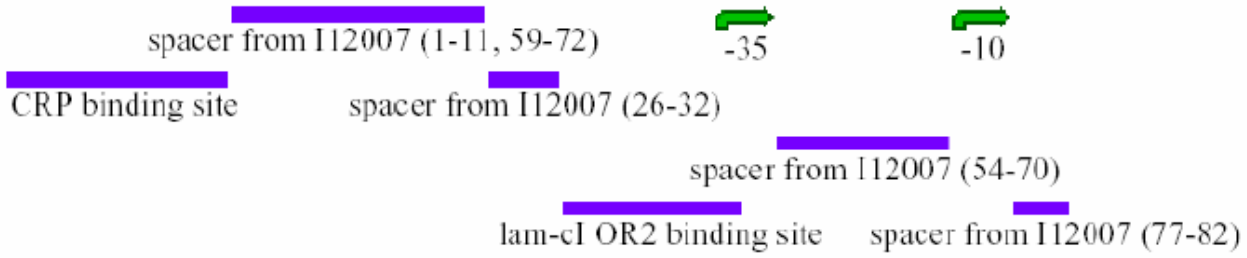


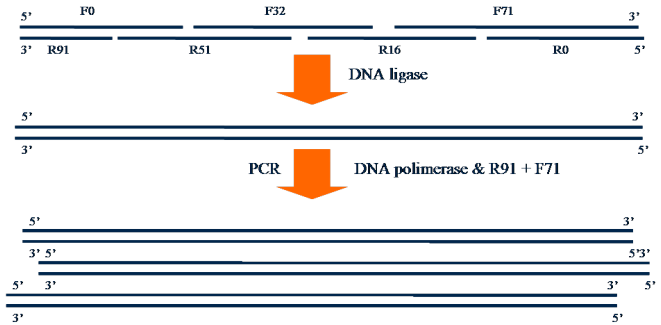








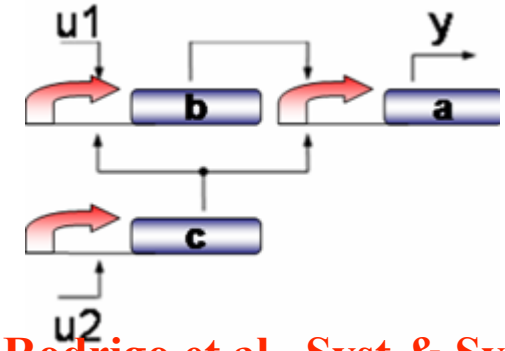




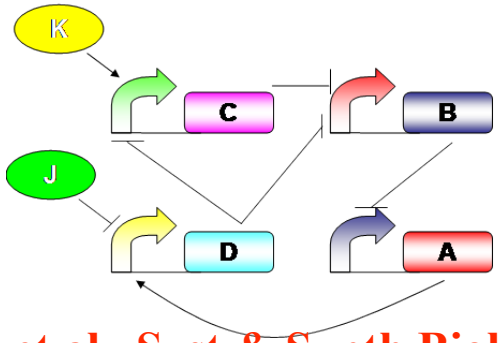
Rodrigo et al. IET Synth Biol 2007

Computationally designed gene networks

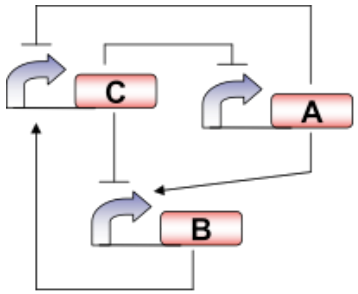
- Logic gates
  - AND
  - NAND
  - OR
  - NOR
- Memory devices
  - RS-Latch
  - JK-Latch
- Oscillatory circuits



Rodrigo et al., Syst & Synth Biol 2008b



Rodrigo et al., Syst & Synth Biol 2008a

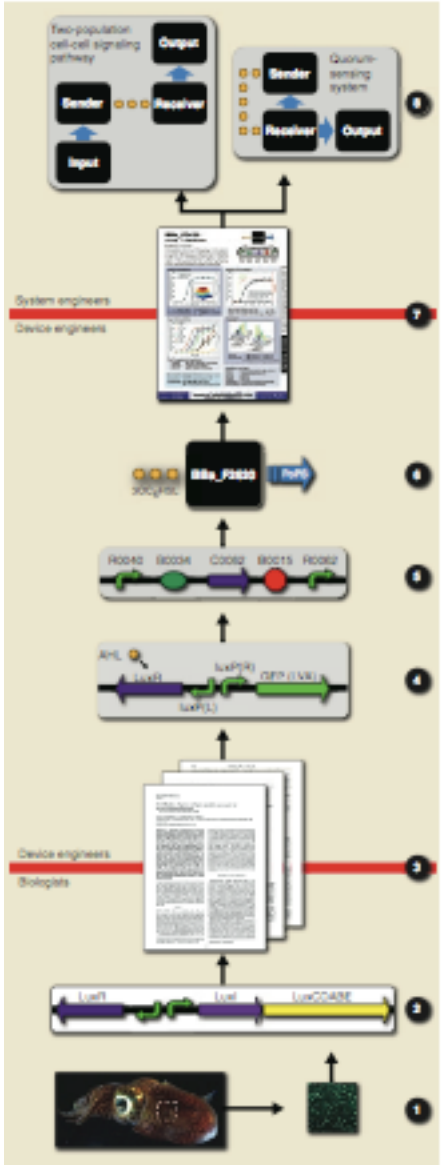


Rodrigo et al., CEJB 2007, Biochimie 2008

New developments: constructing the circuits by assembling

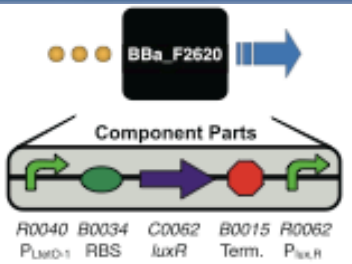
Workflow

Introd
Macromol
Networks
Cells



BBa_F2620

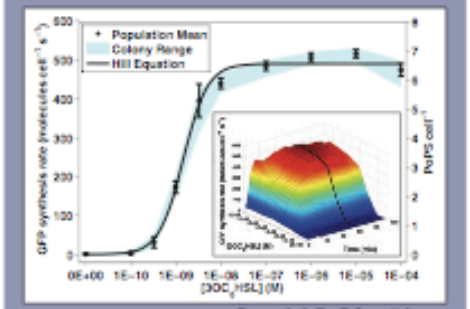
$3OC_6HSL \rightarrow$ PoPS Receiver



Mechanism & Function

A transcription factor (LuxR) that is active in the presence of a cell-cell signaling molecule ($3OC_6HSL$) is controlled by a regulated operator ($P_{LtetO-1}$). Device input is $3OC_6HSL$. Device output is PoPS from a LuxR-regulated operator. If used in a cell containing TetR then a second input such as aTc can be used to produce a Boolean AND function.

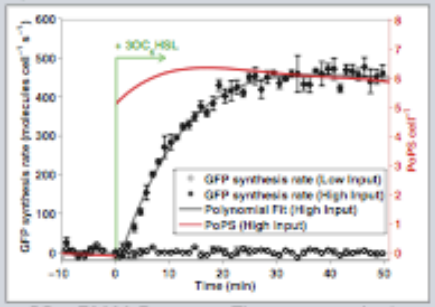
Static Performance*



$$P_{out} = \frac{P_{max}[3OC_6HSL]^n}{K^n + [3OC_6HSL]^n}$$

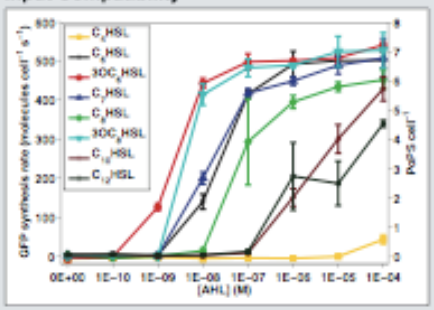
P_{max}: **6.6** PoPS cell^{-1}
K: **1.5E-09** M $3OC_6HSL$
n: **1.6**

Dynamic Performance*



BBa_F2620 Response Time: **<1** *min*
BBa_T9002 Response Time: **6±1** *min*
Inputs: 0 M (Low), 1E-07 M (High) $3OC_6HSL$

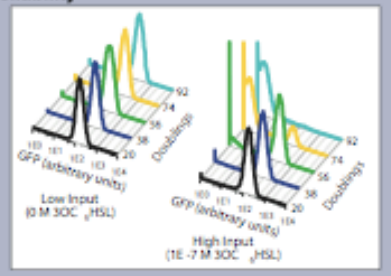
Input Compatibility*



Part Compatibility (qualitative)

Chassis: *MC4100, MG1655,* and *DH5α*
Plasmids: *pSB3K3* and *pSB1A2*
Devices: *E0240, E0430* and *E0434*

Reliability**

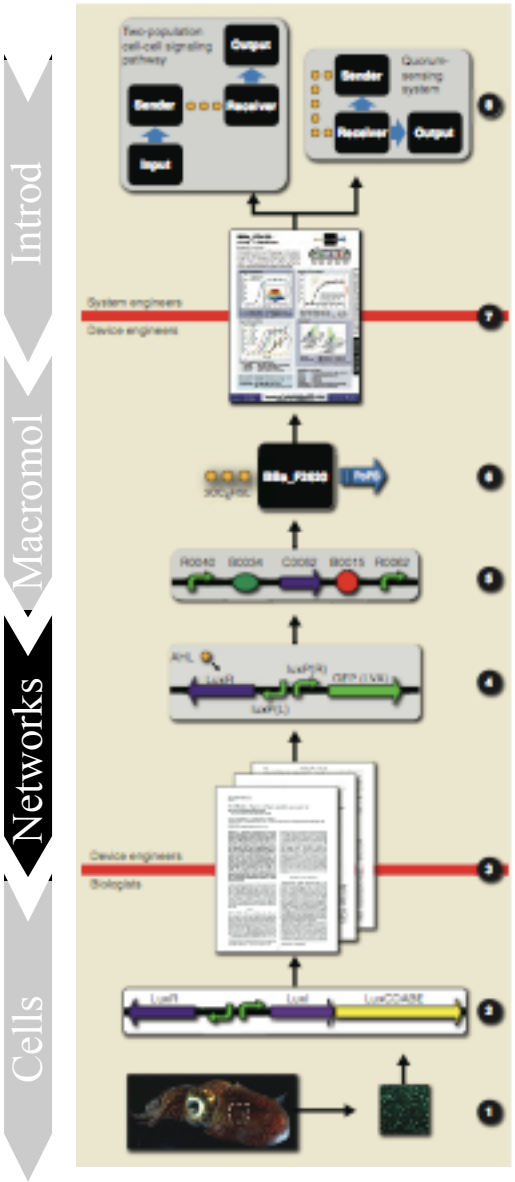


Genetic: **>92/>56** culture doublings
Performance: **>92/>56** culture doublings
(low/high input during propagation)

Transcriptional Output Demand (low/high input)
Nucleotides: **0 / 6.6xNt**ucleotides cell^{-1} s^{-1}
Polymerases: **0 / 1.5E-1xNt** RNAP cell^{-1}
(**Nt** = downstream transcript length)

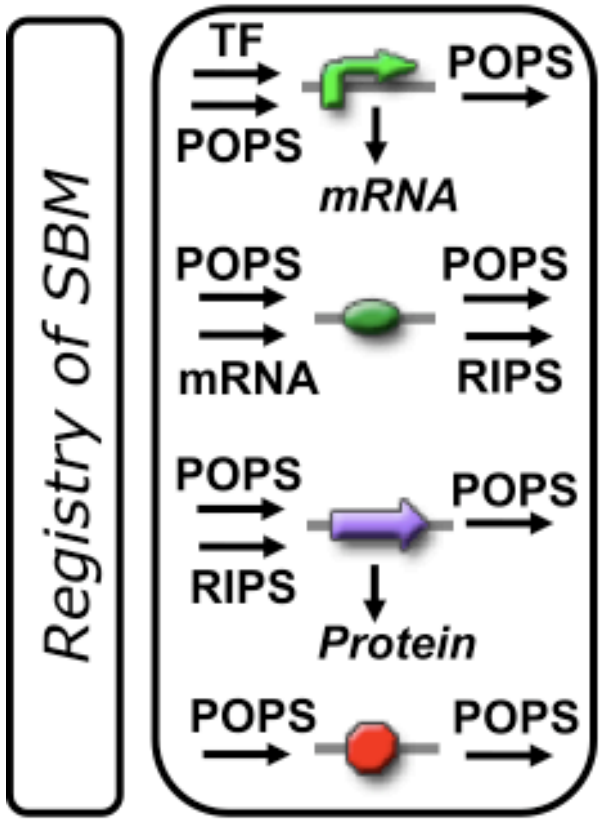
Conditions (abridged)
Output: PoPS measured via *BBa_E0240*
Culture: Supplemented M9, 37°C
Plasmid: pSB3K3
Chassis: MG1655
**Equipment:* PE Victor3 multi-well fluorimeter
***Equipment:* BD FACScan cytometer

From biological discovery to an engineered device



The device is re-engineered using standardised biological parts

Asmparts: *in silico* assembly of parts



$$\frac{d}{dt}[mRNA] = POPS - \delta[mRNA]$$

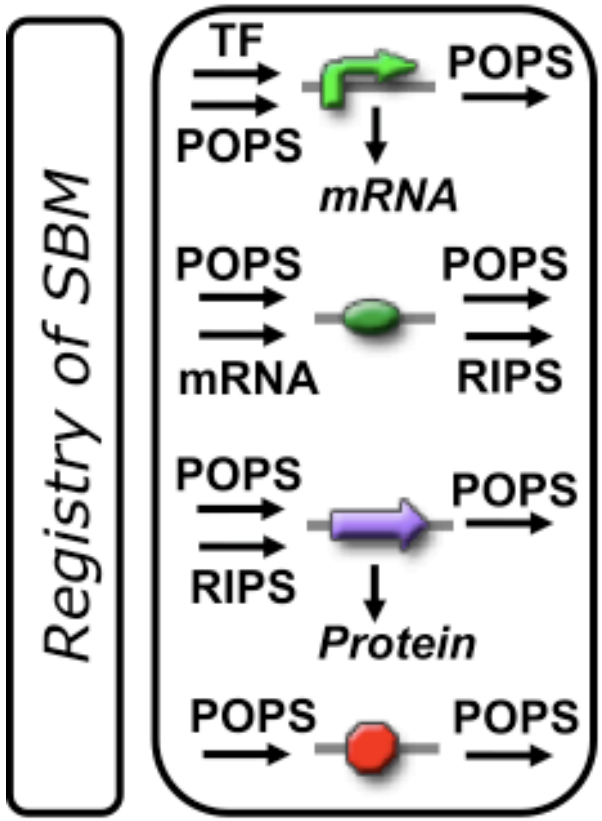
$$\frac{d}{dt}[Protein] = RIPS - \beta[Protein]$$

$$POPS = POPS_0 + \frac{\alpha_0 + \alpha(\frac{[TF]}{K})^n}{1 + (\frac{[TF]}{K})^n}$$

$$RIPS = \lambda[mRNA]$$

$$POPS_0 = \eta POPS$$

Asmparts: *in silico* assembly of parts



$$\frac{d}{dt}[mRNA] = POPS - \delta[mRNA]$$

mRNA degradation constant

$$\frac{d}{dt}[Protein] = RIPS - \beta[Protein]$$

protein degradation constant

transcription rate in presence of TF

basal transcription rate

$$POPS = POPS_0 + \frac{\alpha_0 + \alpha\left(\frac{[TF]}{K}\right)^n}{1 + \left(\frac{[TF]}{K}\right)^n}$$

Hill coefficient

regulatory coefficient

$$RIPS = \lambda[mRNA]$$

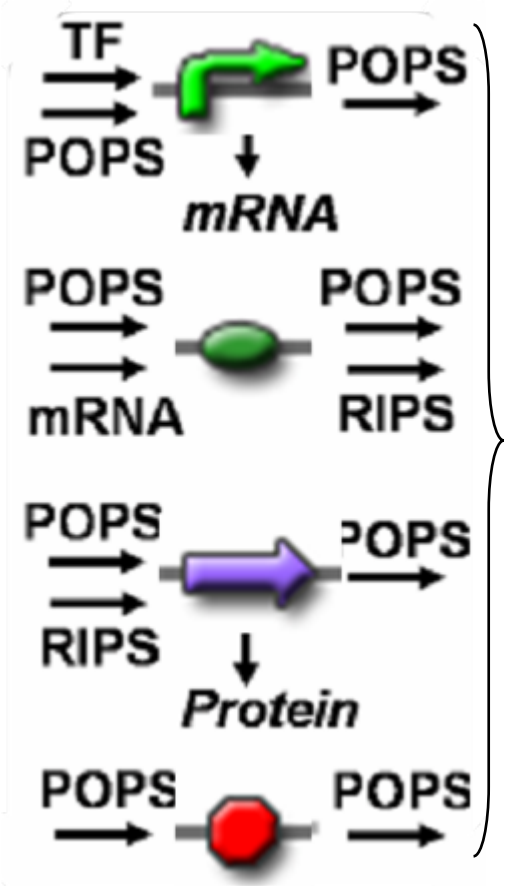
Ribosome binding constant

$$POPS_0 = \eta POPS$$

1-transcription termination efficiency

Asmparts: *in silico* assembly of parts

Biological part models
in SBML



Asmparts: *In silico* assembly of parts models

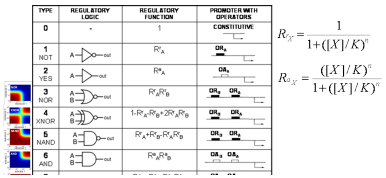
Rodrigo et al. Syst. & Synth. Biol. 2008

Genetdes 3.0: design of biological circuits using a combinatorial assembly of standard model parts

Methodology computational gene network design

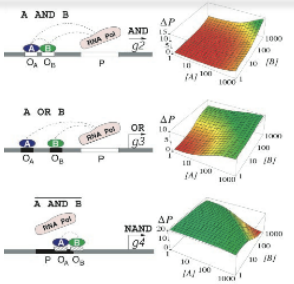
State space: $\{x_i\}$= Concentration/number of molecule i

Mathematical description:



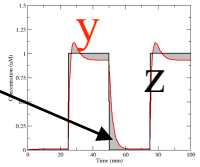
$$\frac{d[Y]}{dt} = \alpha \frac{1}{1+([U]/K)^n} - \beta[Y] + \gamma$$

$$\frac{d[Y]}{dt} = \alpha \frac{([U]/K)^n}{1+([U]/K)^n} - \beta[Y] + \gamma$$



Objective function:

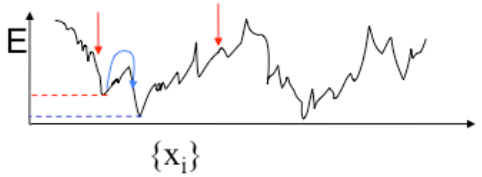
$$J = \frac{\int_{t_i}^{t_f} |y - z| \, dt}{\int_{t_i}^{t_f} z \, dt}$$



Solve the dynamics

Optimisation:

Heuristic: Monte Carlo Simulated Annealing



Challenges:

Avoid solving the dynamics by using analytical approximations for perturbations

Adapt it to stochastic processes and discrete events (*e.g.* signalling)

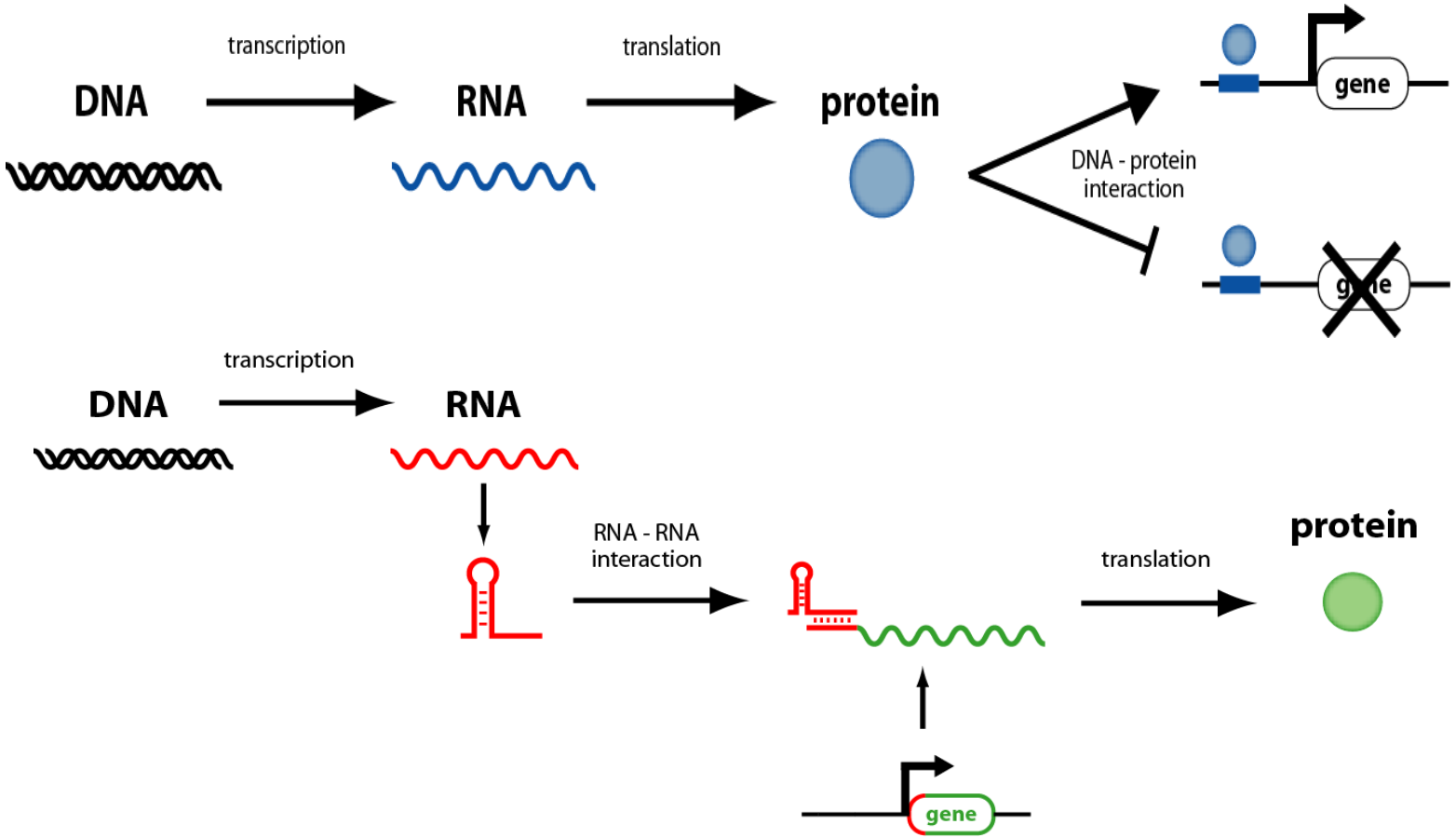
GENETDES software Rodrigo et al. Bioinformatics 2007

Naturally RNA-based gene regulation systems

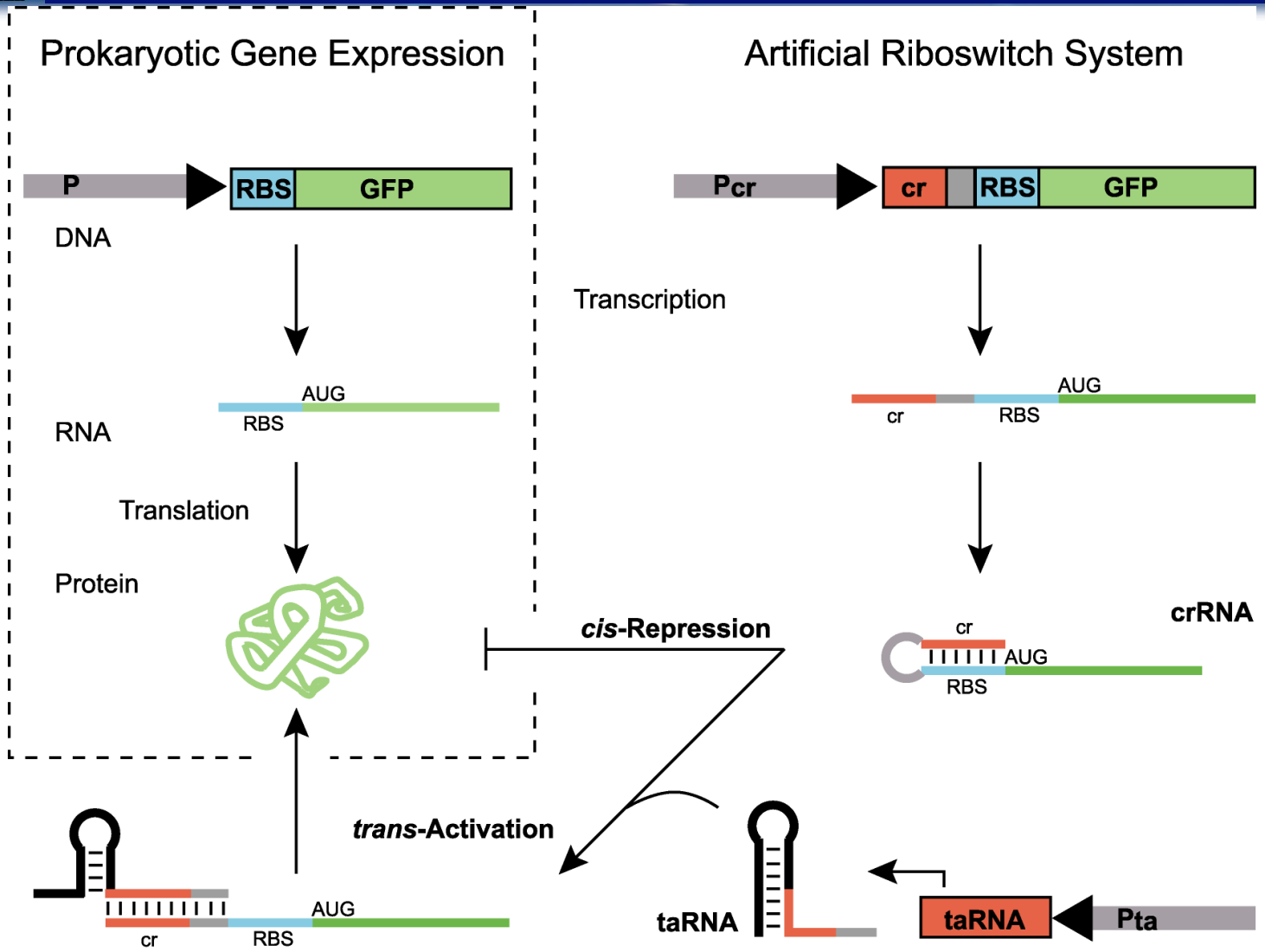
Introd
Macromol
Networks
Cells

Class	Mechanism	Activity
Antisense	Prokaryotic Eukaryotic AAAAA(A)n	Active in *trans* Binding represses translation
Riboregulators	+	Active in *trans* Binding may repress or activate translation
Ribozymes		Active in *cis* or *trans* Activity (cleavage) in *cis* will repress translation Activity (cleavage) in *trans* may repress or activate translation
Riboswitches	Transcriptional Translational Metabolite-binding ribozyme	Active in *cis* Ligand binding may repress or activate transcription Ligand binding may repress or activate translation
Small interfering RNA (siRNA)	AAAAA(A)n	Active in *trans* Binding represses translation
MicroRNA (miRNA)	AAAAA(A)n	Active in *trans* Binding represses translation

RNA-based Synthetic Biology



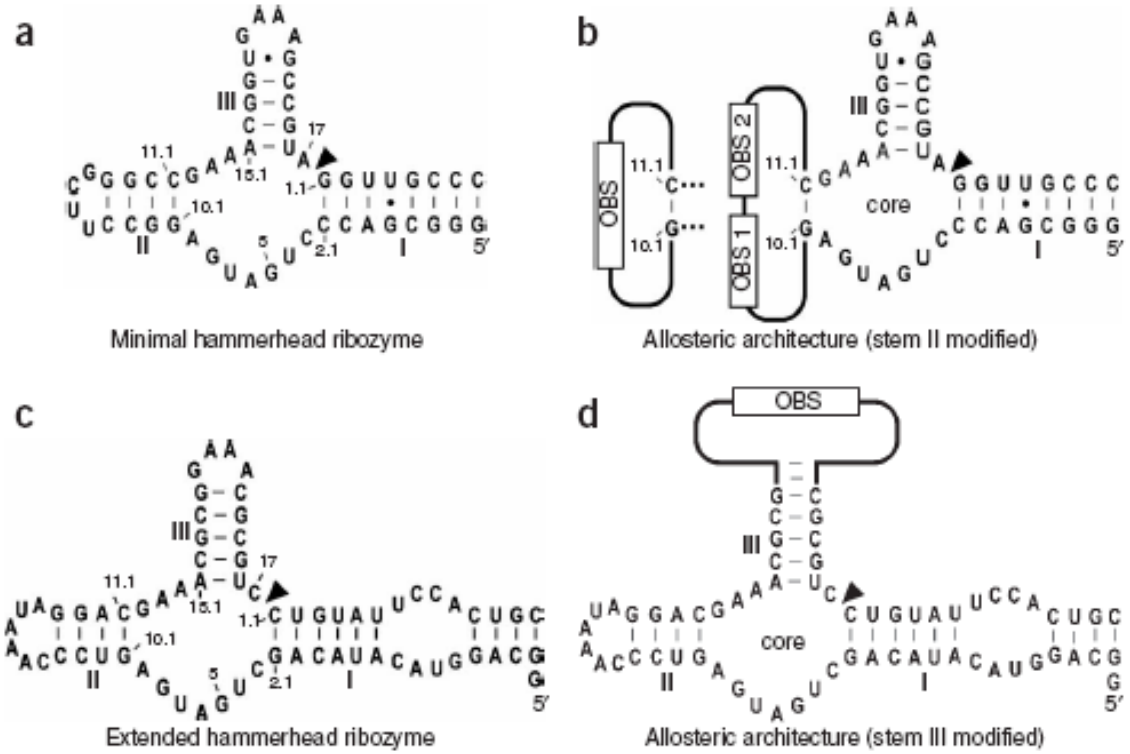
RNA Switches: Engineered Riboregulators



Introd
Macromol
Networks
Cells

FJ Isaacs et al., Nature Biotechnology, 2004

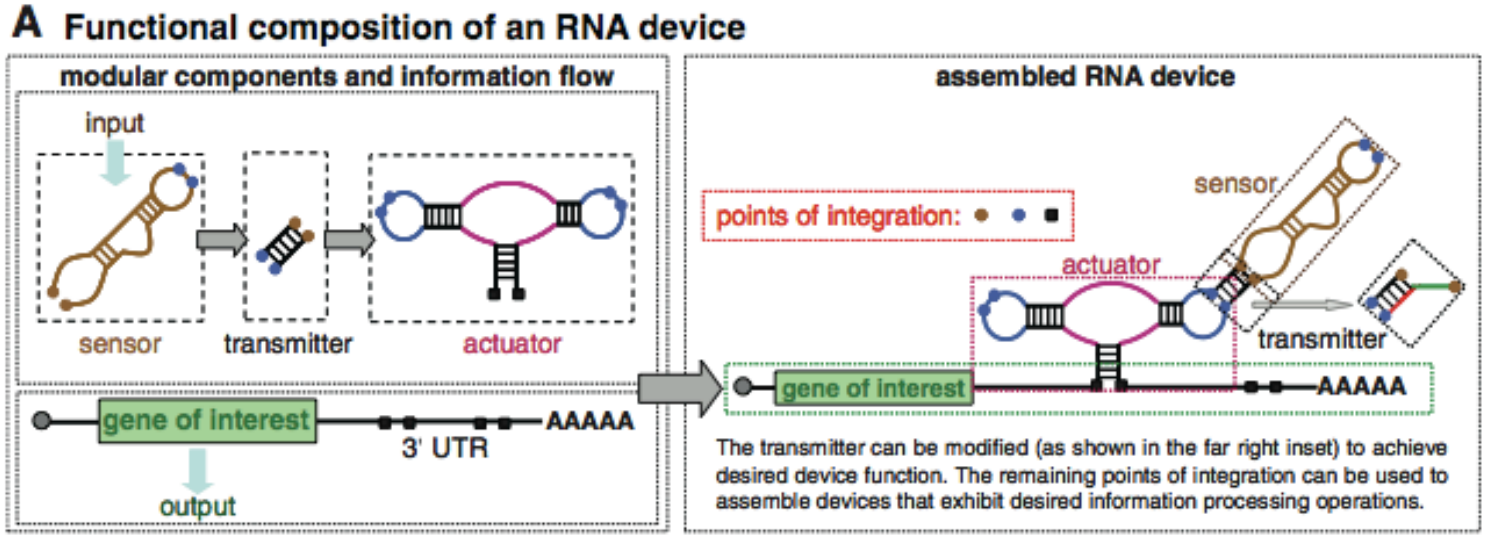
Computational Design of Riboswitches

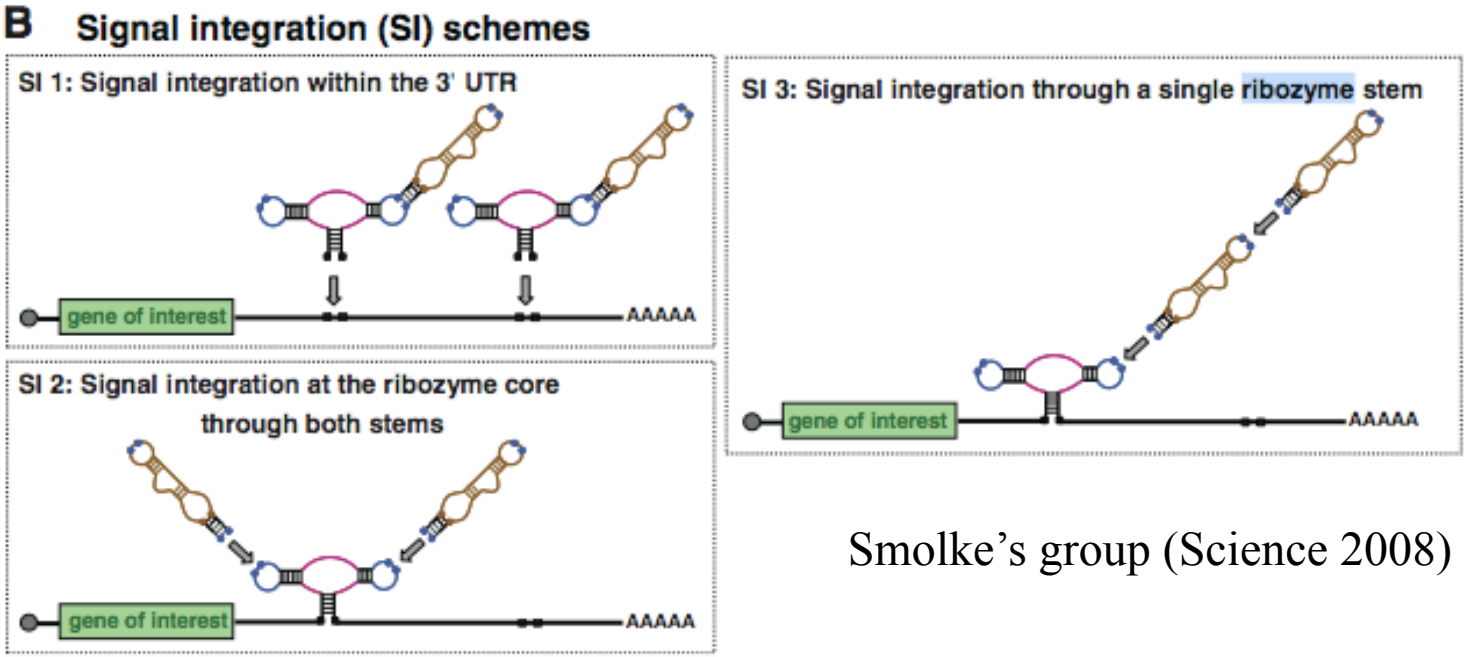


We can use combinatorial optimisation to stabilise an unbound active/inactive ribozyme and to destabilise a bound inactive/active conformation. Several logic gates can be created.

Breaker's group 2005

RNA-based digital devices





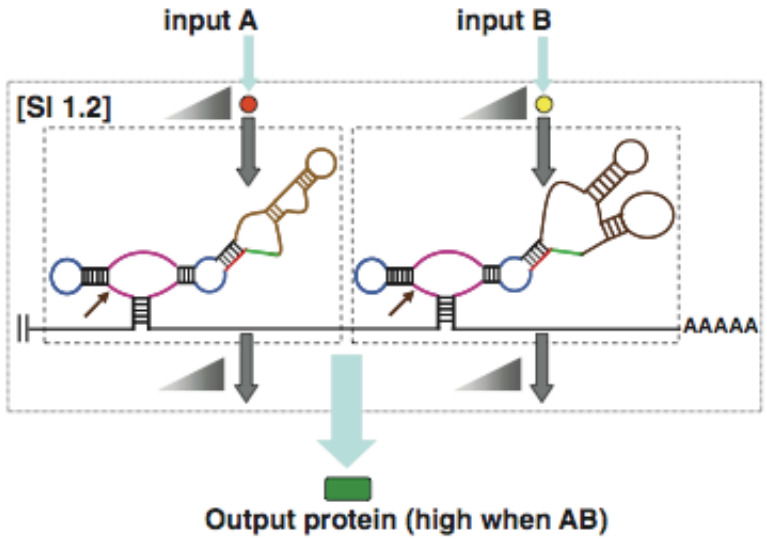
Smolke's group (Science 2008)

RNA-based digital devices

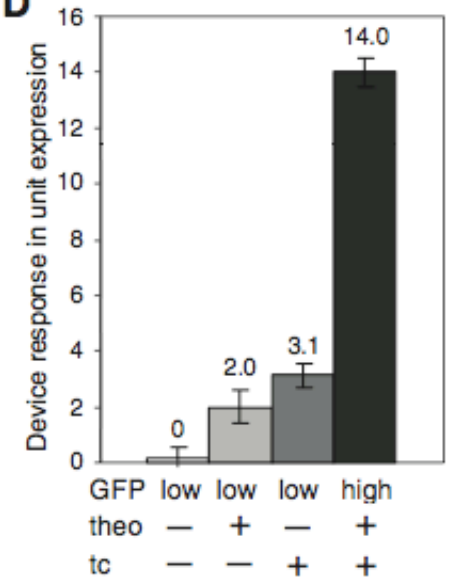
C

AND gate

A theo	B tc	output GFP
0	0	0
0	1	0
1	0	0
1	1	1

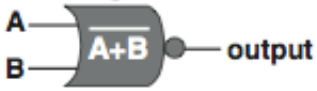


D

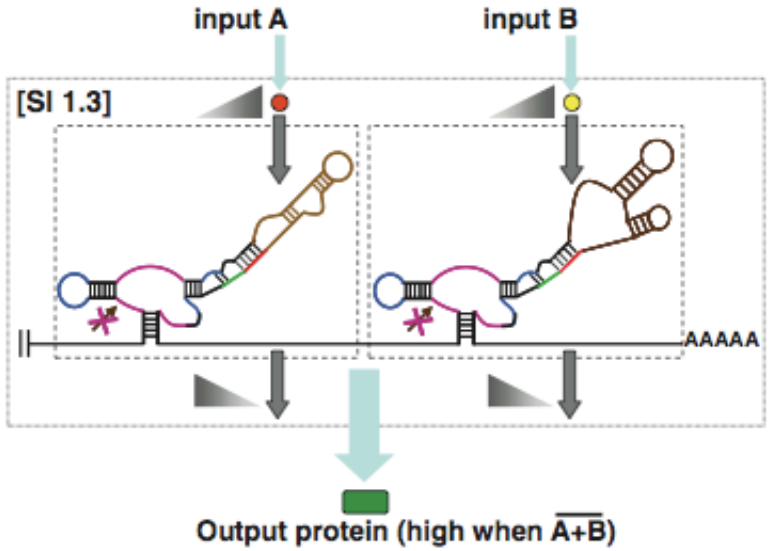


E

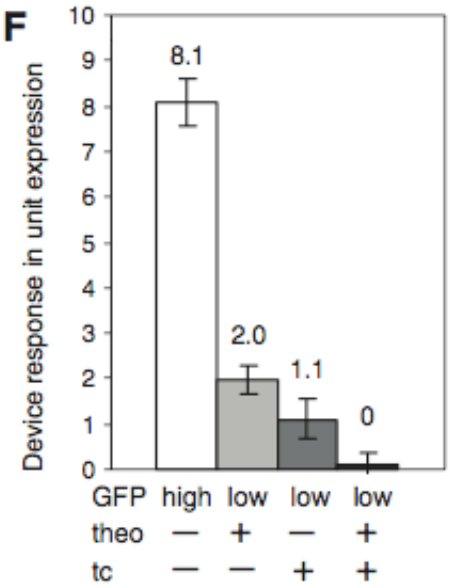
NOR gate



A theo	B tc	output GFP
0	0	1
0	1	0
1	0	0
1	1	0

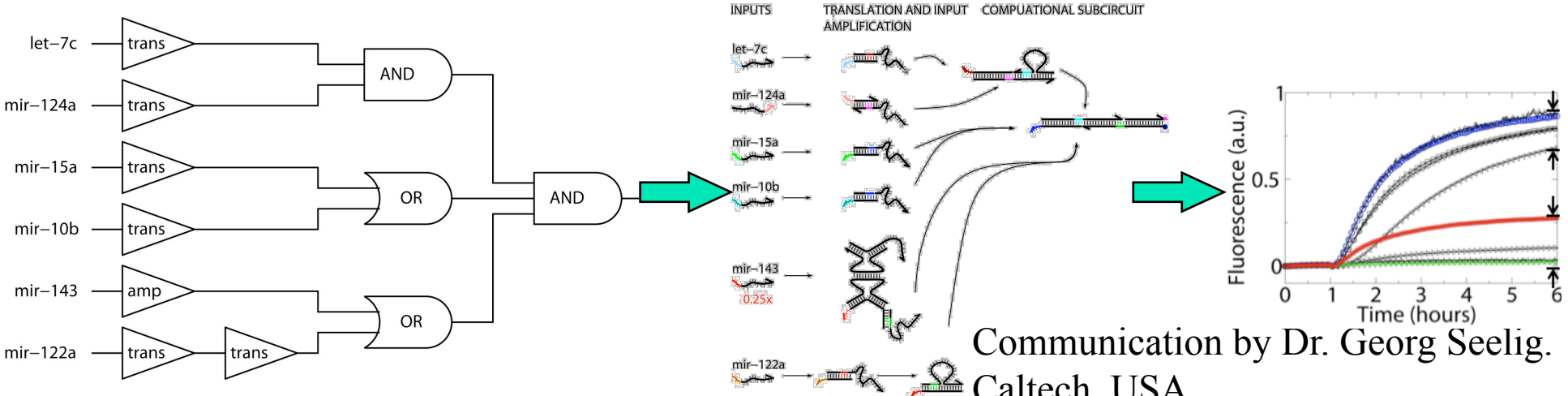


F



Automatic design with nucleic acids

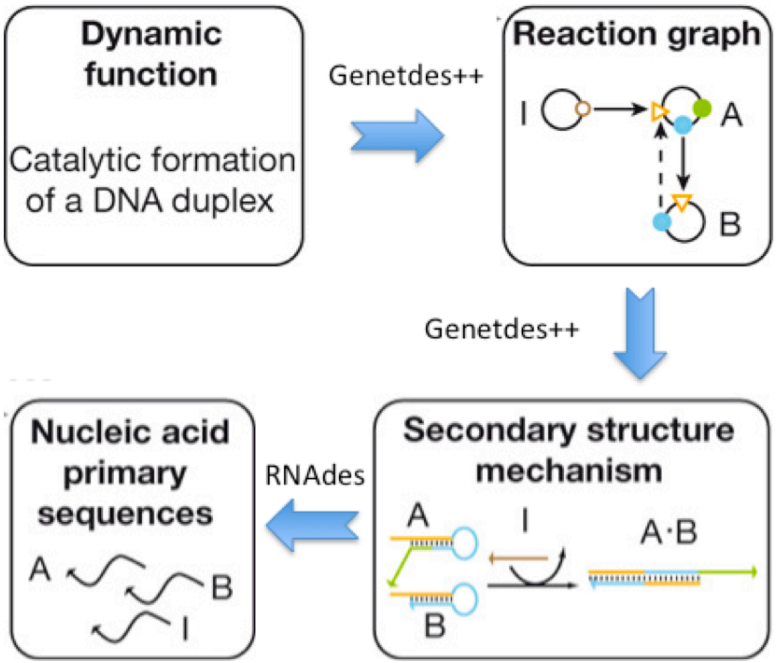
Example of biological integrated circuits by using RNA



Communication by Dr. Georg Seelig. Caltech, USA.

Multi-scale problem:

- Scale 1: Extend Genetdes to use generalised reactions in a modular way (Genetdes++).
- Macroscopic scale, governed by chemical reactions among several species.
- Scale 2: Obtain the nucleotide sequence that will produce a given reaction (RNAdes): **Inverse folding problem**
- Microscopic, controlled by statistical physics.



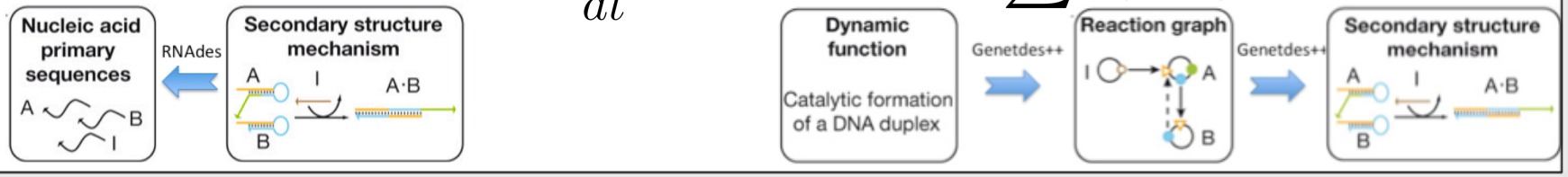
Methodology computational RNA network design

State space: Scale 1: $\{x_i\}$= Concentration of molecule i

Scale 2: $\{x_i\}$= Nucleotide at residue position i

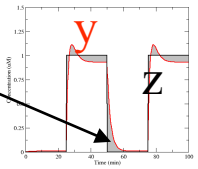
Mathematical description:

$$\frac{d}{dt}x_i = \alpha_i - \beta_i x_i - \sum k_{ij} x_i x_j$$

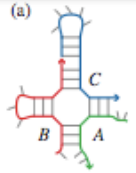


Objective function:

$$J = \frac{\int_{t_i}^{t_f} |y - z| \, dt}{\int_{t_i}^{t_f} z \, dt}$$



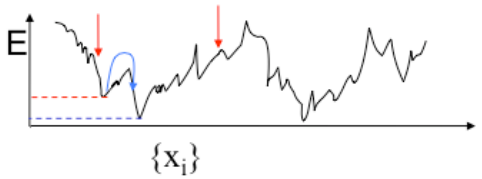
$$\Delta G_{U \circledR F} = G_F - G_U$$



Solve the dynamics

Optimisation:

Heuristic: Monte Carlo Simulated Annealing



Challenges:

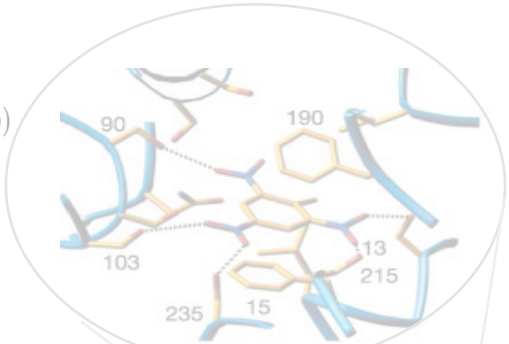
Kinetic modelling considering secondary structure

Generalise inverse folding to “inverse kinetics” problem

Multi-scale computational design

Macromolecules

$$E = \sum_{bonds} K_b(b-b_o)^2 + \sum_{angles} K_\theta(\theta-\theta_o)^2 + \sum_{torsions} K\phi(1+\cos(n\phi-\delta))$$
$$+ \sum_{impropers} K_\varphi(\varphi-\varphi_o)^2 + \sum_{Urey-Bradley} K_{UB}(r_{1,3}-r_{1,3,o})^2 +$$
$$+ \sum_{nonbonded} \frac{q_i q_j}{4\pi D r_{ij}} + \varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right]$$



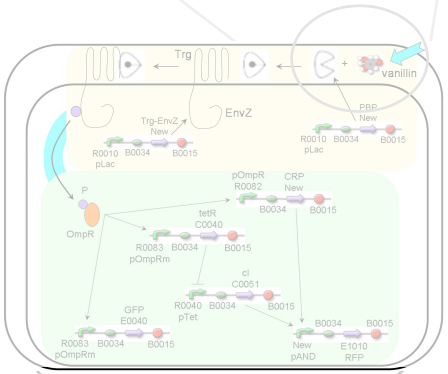
De novo design of proteins:

- DESIGNER
- PROTDES

Biological networks

$$\frac{d[Y]}{dt} = \alpha\frac{1}{1+(\frac{[U]}{K})^n} - \beta[Y] + \gamma$$

$$\frac{d[Y]}{dt} = \alpha\frac{(\frac{[U]}{K})^n}{1+(\frac{[U]}{K})^n} - \beta[Y] + \gamma$$



De novo design of:

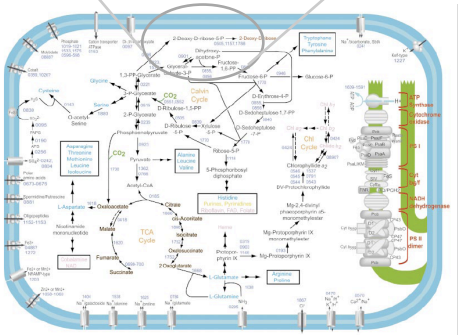
- Transcriptional networks
  - GENETDES
  - ASMPARTS
- RNA networks
  - RNADES

Cells: Genomic background

$$\max \mu = cv$$

$$\text{st}: Sv = b$$

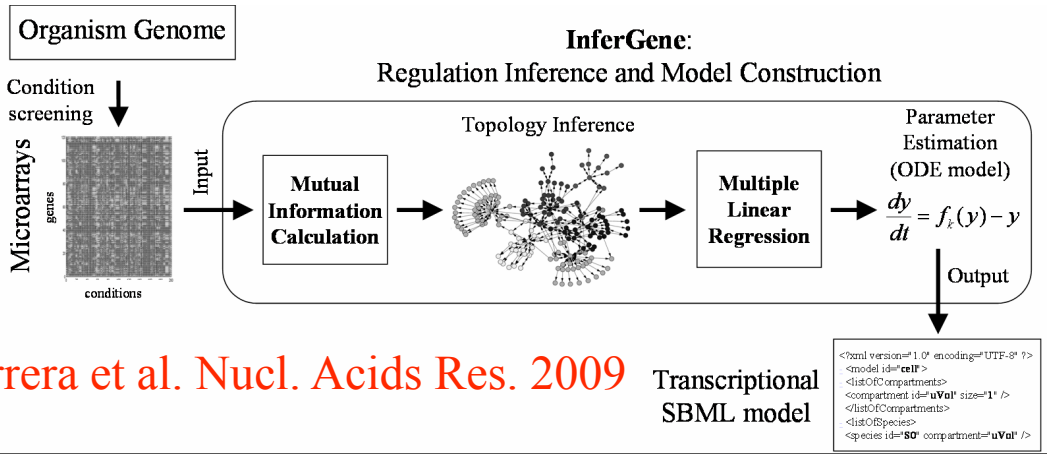
$$v_{\min} \le v \le v_{\max}$$



- *De novo* design of metabolic pathways by retro-biosynthesis
  - DESHARKY
- Network inference from microarray & proteomics resp.
  - INFERGENE

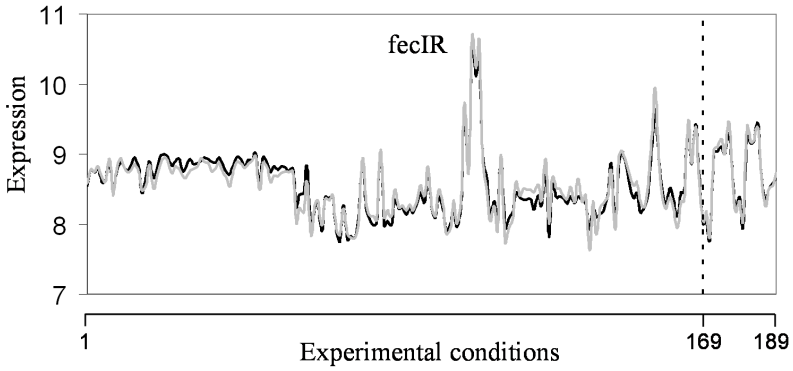
Computational designs in a genomic background

We obtained a ODE model for the global transcription network of *E. coli*:



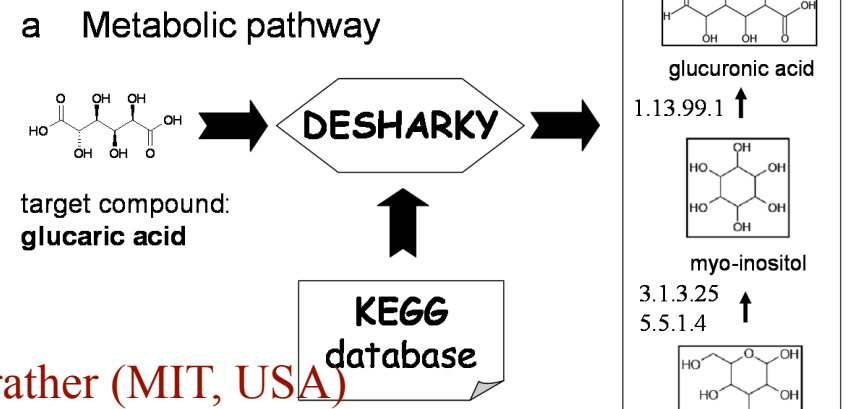
Carrera et al. Nucl. Acids Res. 2009

Predicted expression versus experimental



We developed a methodology for the automatic design of metabolic pathways:

We use a retro-biosynthesis algorithm



Improved sampling

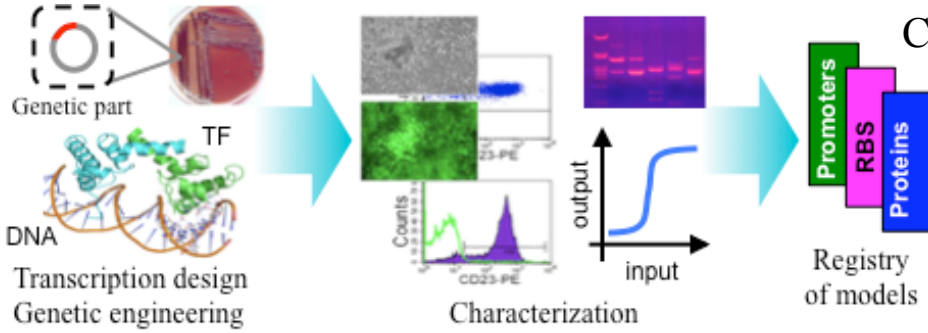
b List of enzymes

EC number	Organism	Gene
1.1.1.203	-	-
1.13.99.1	*S. usitatus*	Q01Z78
3.1.3.25	*E. coli*	SUHB
5.5.1.4	*N. oceani*	Q3J6X3

Coll. Prof. Prather (MIT, USA)

Rodrigo et al. Bioinformatics 2008

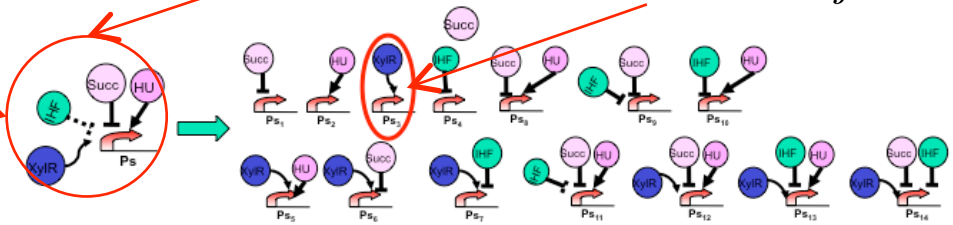
Combinatorial computational genome design



Characterization → biological part models (Asmparts)

$$\frac{d}{dt}X_i = \sum\prod \alpha_{ij}/(1+(X_j/K_{ij})^n_{ij}) - \beta_i X_i + \gamma_i$$

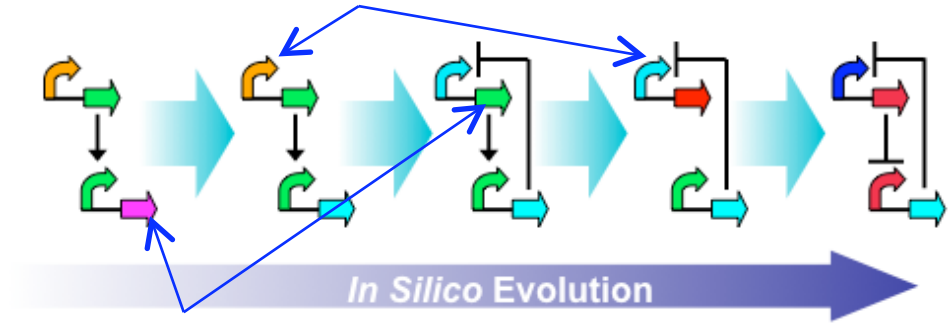
$$\alpha_i/(1+(X_j/K_{ij})^n_{ij})$$



Combinatorial promoter

Construction of a computational promoter library

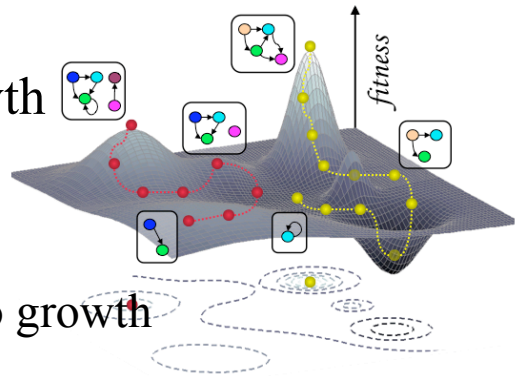
Replace promoter



Add/remove ORF

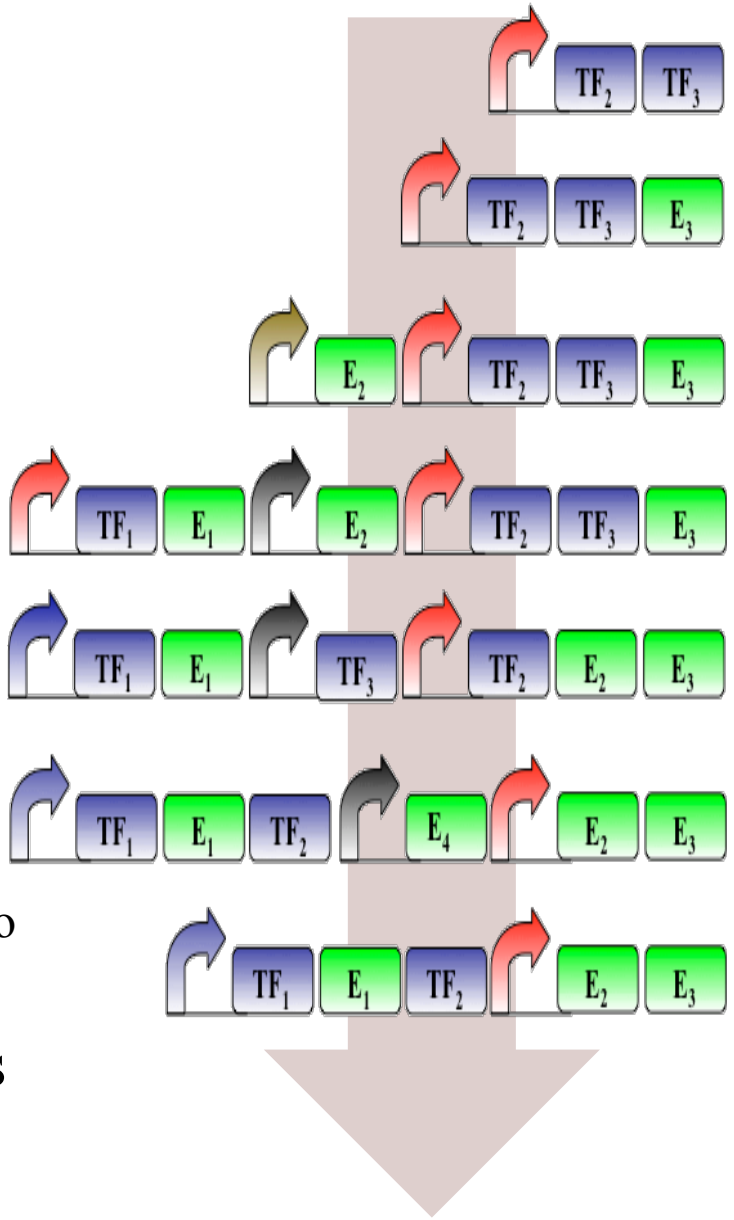
Fitness/scoring function:

- Use chassis model to estimate cell growth
- Cost/benefit model:
- Expressing genes is detrimental to growth
- Expressing "useful" pathways contributes to growth



In silico genome evolution and design

- Evolution moves:
  - Add/remove TF or enzyme
  - Replace promoter
  - Add/remove operon
  - Modify kinetic parameters
- Biological part models (Asmparts)
- Desharky to move in metabolic space
- Fitness/scoring function:
  - Use chassis model to estimate cell growth
  - Cost/benefit model:
    - Expressing genes is decremental to growth
    - Expressing “useful” pathways contributes to growth
- FBA for fast metabolic reactions, ODEs for slow transcriptional ones.



Methodology genome-scale modelling

State space: $\{x_i\}$= Concentration of metabolite i

$\{y_i\}$= Concentration of transcription factor i

Mathematical description:

Where v_i are the cell metabolic fluxes, c their contributions to the growth rate, S the stoichiometry matrix, and b the uptake fluxes

$$\frac{\mathrm{d}}{\mathrm{dt}} y_i = a_i + \sum_{j \in TF} b_{ij} y_j + \sum_{j \in TF} \sum_{k \in TF} b_{ijk} y_j y_k - \delta_i y_i,$$

$$\frac{d}{dt} x_i = \sum S_{ij} v_j - b_i$$

Objective function:

$$\max \mu = cv$$

Subject to $Sv = b$

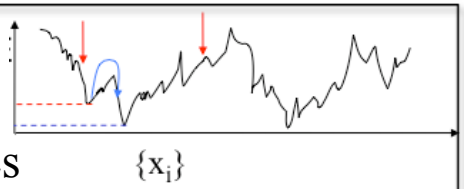
Steady state assumption

$$v_{\min} \leq v \leq v_{\max}$$

Optimisation:

Exact: Linear Programming $\{v_i\}$

Heuristic: Monte Carlo Simulated Annealing to evolve the ODEs

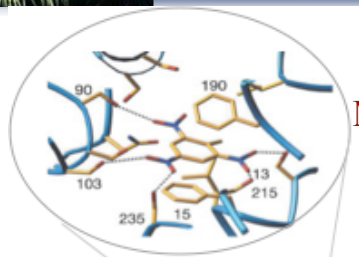


Challenges:

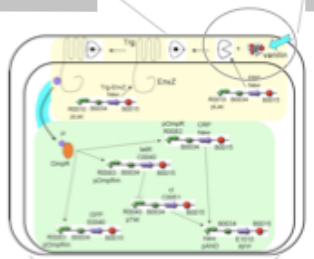
Couple transcription regulation to metabolic reactions

Integrate discrete events (*e.g.* signalling).

Conclusion

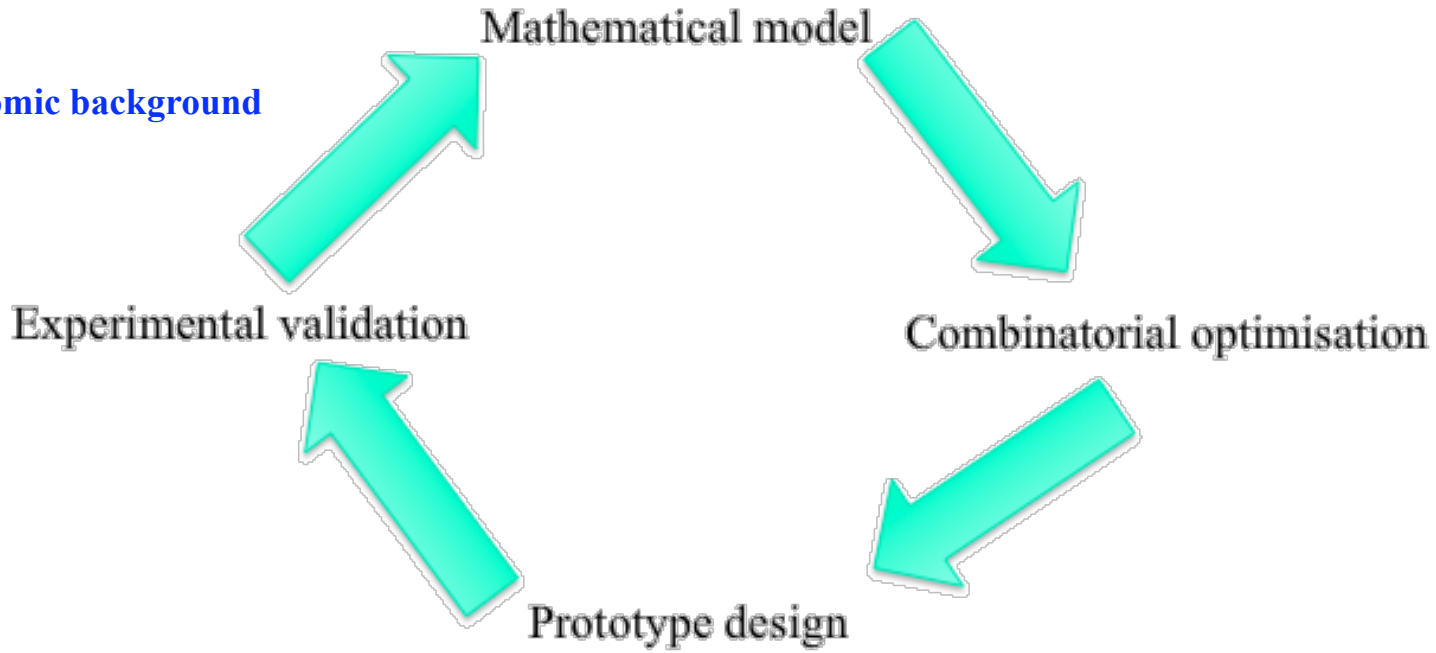


Macromolecules



Biological networks

Genomic background



Synth-Bio group

- Guillermo Rodrigo PhD Student (co-supervised IBMCP, Spain) (2006)
- Javier Carrera PhD Student (co-supervised IBMCP, Spain) (2007)
- Filipe Pinto PhD Student (co-supervised IBMC, Portugal) (2009)
- Daniel Camsund PhD Student (co-supervised Uppsala, Sweeden) (2009)
- Boris Kirov PhD Student (2008)
- Thomas Landrain PhD Student (2008)
- Bogdana Barlacu Technician (2008)
- Vijai Singh Postdoc (2009)
- Mariel Montesinos Administrative assist. & project management (2007)

Recent members:

- Pablo Tortosa EMBO postdoc (2004-2007)
- Maria Suarez Postdoc (2006-2009)

http://synth-bio.org

Postdoc openings!!!

Funding

ATIGE	Genopole/UEVE	2008-20
SynthBioClock	CNRS IPCB	2008
TARPOL	FP7	2008-20
BioModularH2	FP6 NEST	2007-20
Solar ethanol	IFCPAR/CEFIPRA	2008-20
Aide projets EU	Ile-de-France	2008-20
Emergence	FP6 NEST	2006-20
Laccase design	Alliance (Columbia)	2006-20
Glucaric acid p.	MIT-France	