(Computational) Synthetic Biology

Natalio Krasnogor
Natalio.Krasnogor@Nottingham.ac.uk

ASAP - Interdisciplinary Optimisation Laboratory
School of Computer Science

Centre for Integrative Systems Biology
School of Biosciences

Centre for Healthcare Associated Infections
Institute of Infection, Immunity & Inflammation

University of Nottingham

Outline

❿Essential Systems Biology

•Synthetic Biology

•Computational Modeling for Systems/Synthetic Biology

•A Note on Ethical, Social and Legal Issues

•Conclusions

The Far Future

L. Cronin, N. Krasnogor, B. G. Davis, C. Alexander, N. Robertson, J.H.G. Steinke, S.L.M. Schroeder, A.N. Khlobystov, G. Cooper, P. Gardner, P. Siepmann, and B. Whitaker. The imitation game - a computational chemical approach to recognizing life. Nature Biotechnology, 24:1203-1206, 2006.

Outline

•Essential Systems Biology

•Synthetic Biology

•Computational Modeling for Synthetic Biology

•A Note on Ethical, Social and Legal Issues

•Conclusions

Essential Systems Biology

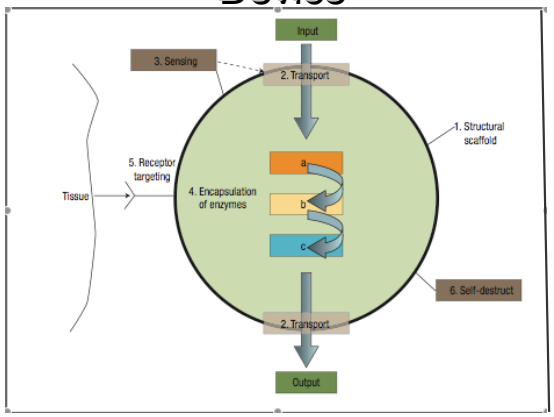
- The following slides are based on U. Alon's papers & excellent introductory text book:

"An Introduction to Systems Biology: Design Principles of Biological Circuits"

- Also, you may want to check his group's webpage for up-to-date papers/software:

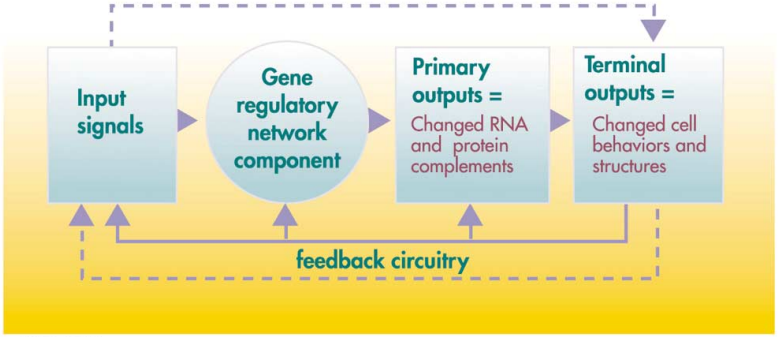
http://www.weizmann.ac.il/mcb/UriAlon/

The Cell as an Information Processing Device



LeDuc et al. Towards an in vivo biologically inspired nanofactory. Nature (2007)

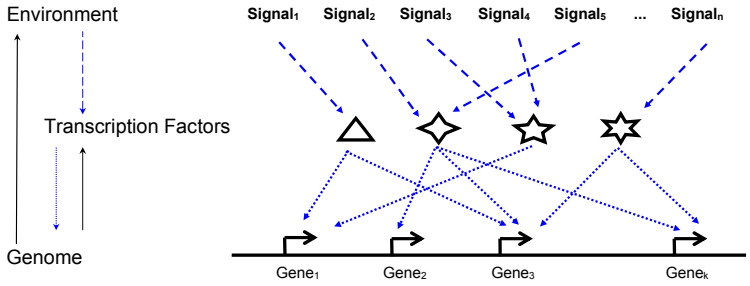
The Cell as an Intelligent (Evolved) Machine

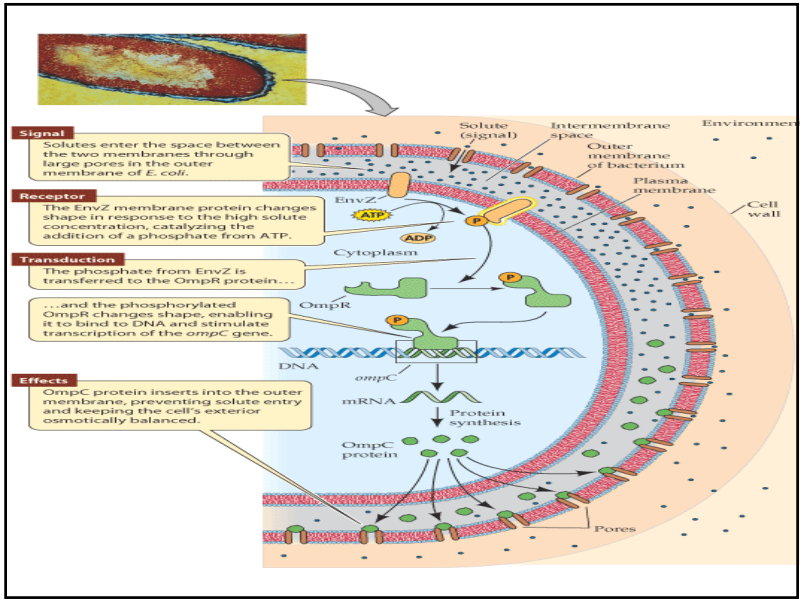


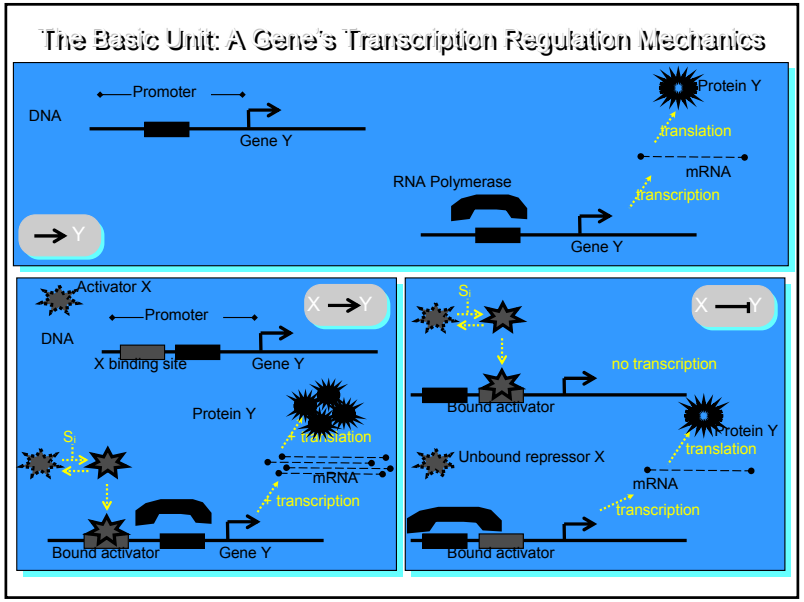
YGG 01-0086

Amir Mitchell, et al., Adaptive prediction of environmental changes by microorganisms. Nature June 2009.

Transcription Networks







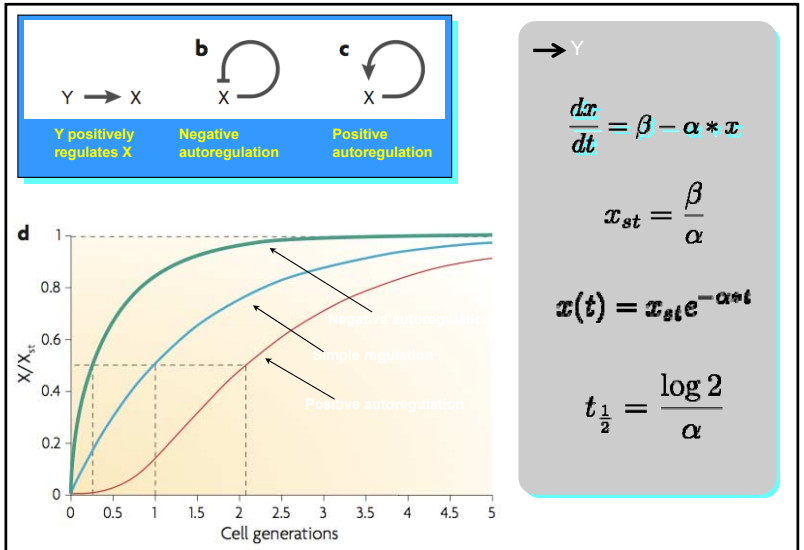
Network Motifs: Evolution's Preferred Circuits

- Biological networks are complex and vast
- To understand their functionality in a *scalable* way one must choose the correct abstraction

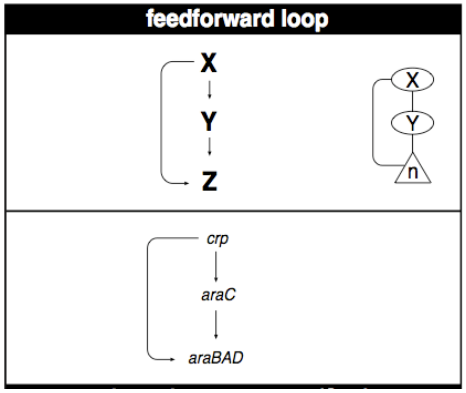
"Patterns that occur in the real network significantly more often than in randomized networks are called network motifs" Shai S. Shen-Orr et al., Network motifs in the transcriptional regulation network of Escherichia coli. Nature Genetics 31, 64 - 68 (2002)

- Moreover, these patterns are organised in non-trivial/non-random hierarchies Radu Dobrin et al., Aggregation of topological motifs in the Escherichia coli transcriptional regulatory network. BMC Bioinformatics. 2004; 5: 10.

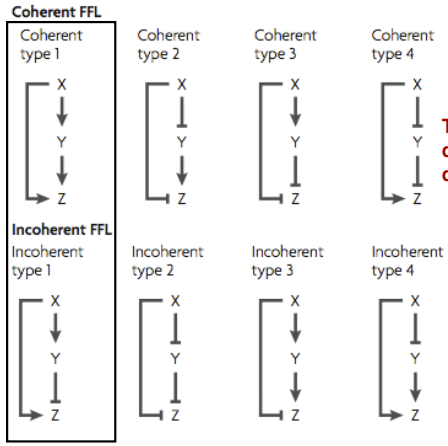
- ***Each network motif carries out a specific information-processing function***



U. Alon. Network motifs: theory and experimental approaches. Nature Reviews Genetics (2007) vol. 8 (6) pp. 450-461



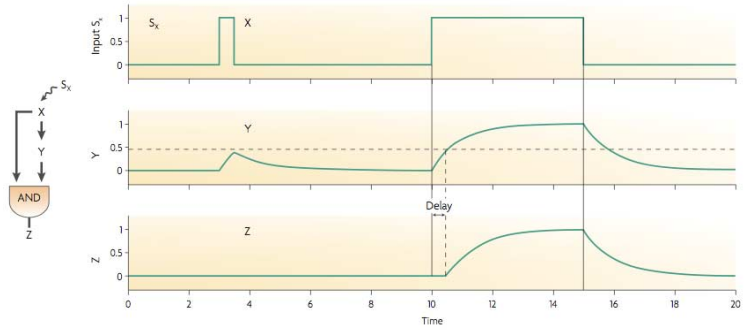
Shai S. Shen-Orr et al., Network motifs in the transcriptional regulation network of Escherichia coli. Nature Genetics 31, 64 - 68 (2002)



The C1-FFL is a 'sign-sensitive delay' element and a persistence detector.

The I1-FFL is a pulse generator and response accelerator

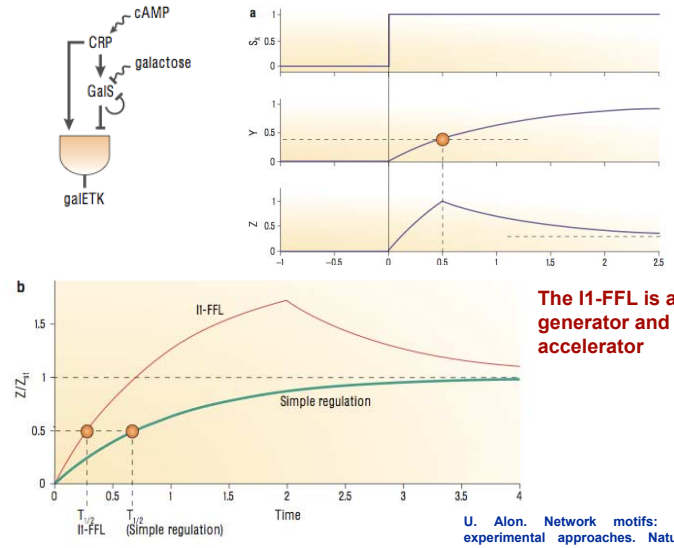
U. Alon. Network motifs: theory and experimental approaches. Nature Reviews Genetics (2007) vol. 8 (6) pp. 450-461



The C1-FFL is a 'sign-sensitive delay' element and a persistence detector.

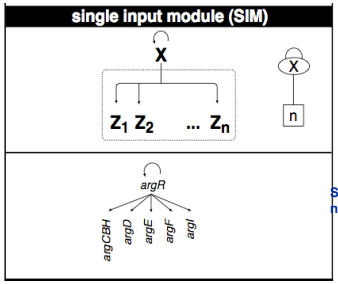
If the integration function is "OR" (rather than "AND"), C1-FFL has now delay after stimulation by Sx but, instead, manifests the delay when the stimulation stops.

U. Alon. Network motifs: theory and experimental approaches. Nature Reviews Genetics (2007) vol. 8 (6) pp. 450-461

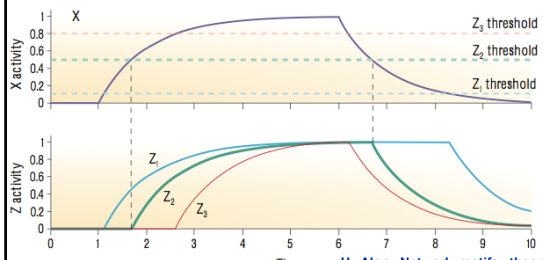


The I1-FFL is a pulse generator and response accelerator

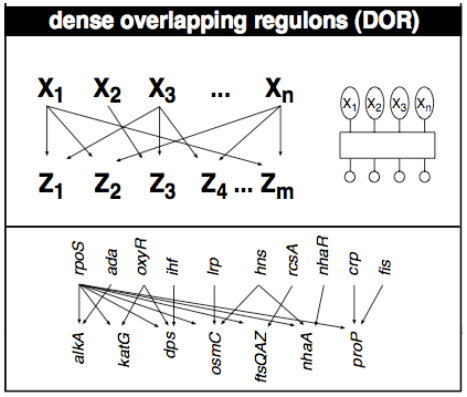
U. Alon. Network motifs: theory and experimental approaches. Nature Reviews Genetics (2007) vol. 8 (6) pp. 450-461



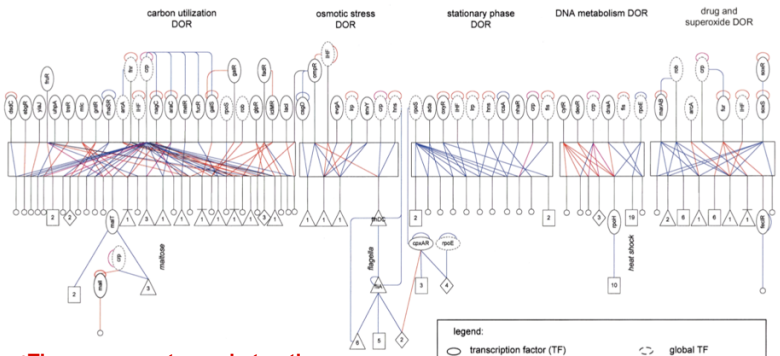
Shai S. Shen-Orr et al., Network motifs in the transcriptional regulation network of Escherichia coli. Nature Genetics 31, 64 - 68 (2002)



U. Alon. Network motifs: theory and experimental approaches. Nature Reviews Genetics (2007) vol. 8 (6) pp. 450-461



Shai S. Shen-Orr et al., Network motifs in the transcriptional regulation network of Escherichia coli. Nature Genetics 31, 64 - 68 (2002)



- **The correct abstractions facilitates understanding in complex systems.**
- **Provide a route to engineering & programming cells.**

Shai S. Shen-Orr et al., Network motifs in the transcriptional regulation network of Escherichia coli. Nature Genetics 31, 64 - 68 (2002)

Outline

- Essential Systems Biology
- Synthetic Biology
- Computational Modeling for Synthetic Biology
- A Note on Ethical, Social and Legal Issues
- Conclusions

What is Synthetic Biology

Synthetic Biology is http://syntheticbiology.org/

A) the design and construction of new biological parts, devices, and systems, and

B) the re-design of existing, natural biological systems for useful purposes.

C) Through rigorous mathematical, computational & engineering routes

Synthetic Biology

- Aims at designing, constructing and developing artificial biological systems
- Offers new routes to 'genetically modified' organisms, synthetic living entities, smart drugs and hybrid computational-biological devices.
- Potentially enormous societal impact, e.g., healthcare, environmental protection and remediation, etc
- **Synthetic Biology's basic assumption:**
  - **Methods readily used to build non-biological systems could also be use to specify, design, implement, verify, test and deploy novel synthetic biosystems.**
  - **These method come from computer science, engineering and maths.**
  - **Modeling and optimisation run through all of the above.**

Synthetic & Systems Biology: Sisters Disciplines

Synthetic Biology

Systems Biology

Basic goal: to ***clarify*** current understandings by formalising what the constitutive elements of a system are and how they interact

Intermediate goal: to ***test*** current understandings against experimental data

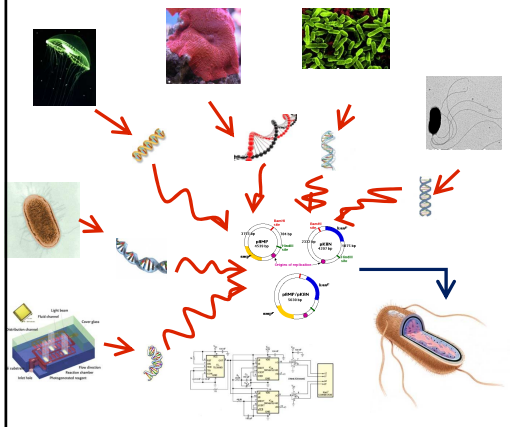
Advanced goal: to ***predict*** beyond current understanding and available data

Dream goal:

(1) to ***combinatorially combine*** *in silico* well-understood components/models for the design and generation of novel experiments and hypothesis and ultimately

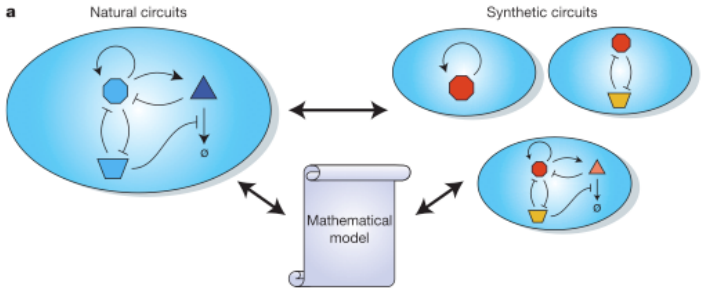
(2) to ***design, program, optimise & control*** (new) biological systems

Top-Down Synthetic Biology: An Approach to Engineering Biology



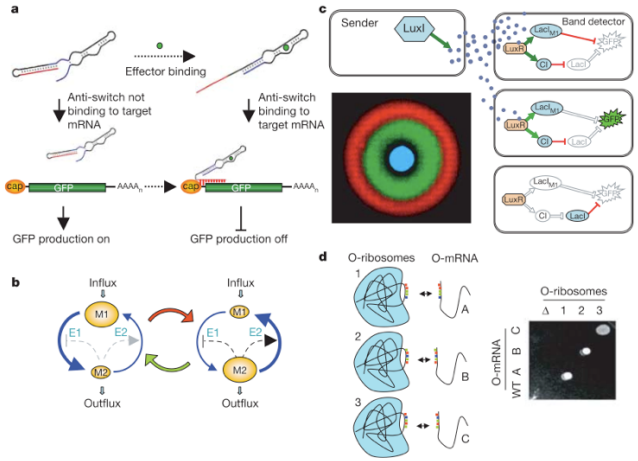
- Cells are **information processors**. DNA is their programming language.
- **DNA sequencing and PCR**: Identification and isolation of cellular parts.
- **Recombinant DNA and DNA synthesis** : Combination of DNA and construction of new systems.
- **Tools to make biology easier to engineer:** Standardisation, modularisation and abstraction (blueprints).

Synthetic Biology's Brick & Mortar (I)



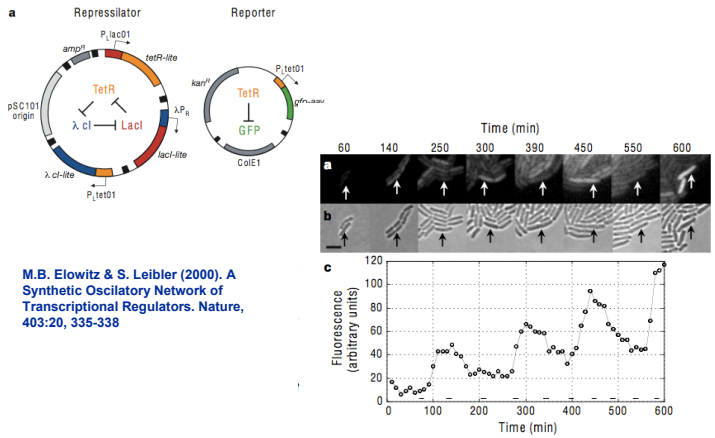
D. Sprinzak & M.B. Elowitz (2005). Reconstruction of genetic circuits, Nature 438:24, 443-448.

Synthetic Biology's Brick & Mortar (II)



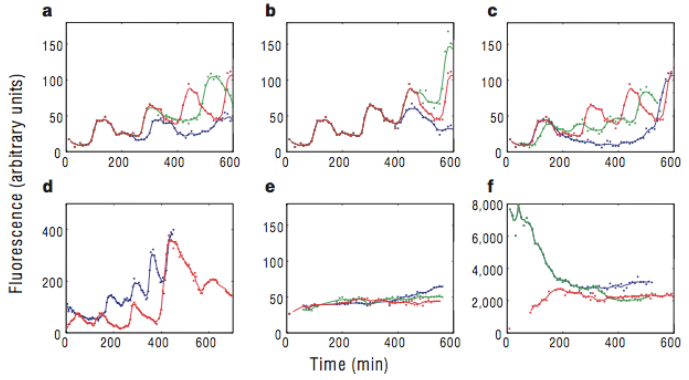
D. Sprinzak & M.B. Elowitz (2005). Reconstruction of genetic circuits, Nature 438:24, 443-448.

Example I: Elowitz & Leibler Represilator



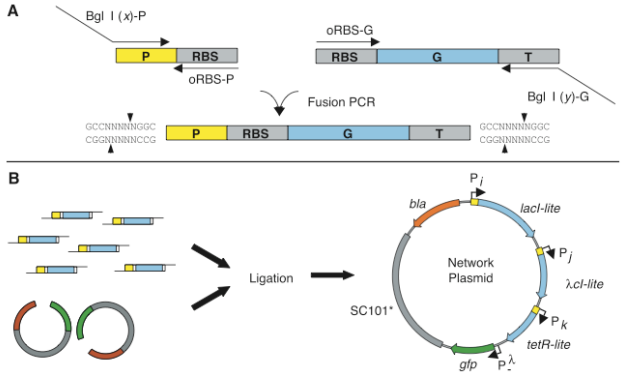
M.B. Elowitz & S. Leibler (2000). A Synthetic Oscilatory Network of Transcriptional Regulators. Nature, 403:20, 335-338

An Example: Elowitz & Leibler Represilator



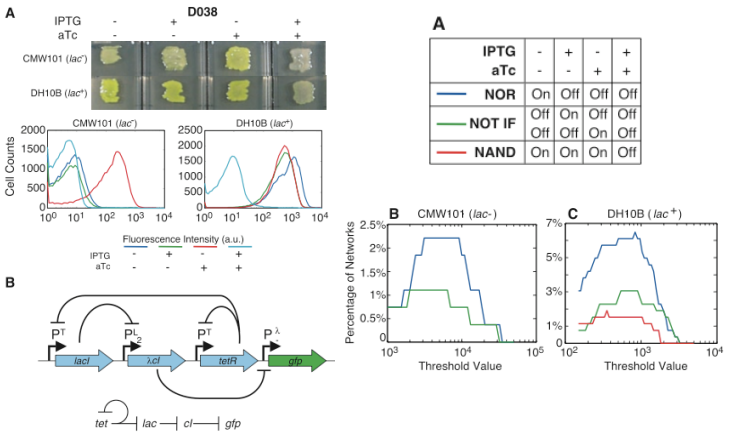
M.B. Elowitz & S. Leibler (2000). A Synthetic Oscillatory Network of Transcriptional Regulators. Nature, 403:20, 335-338

Example II: Combinatorial Synthetic Logic



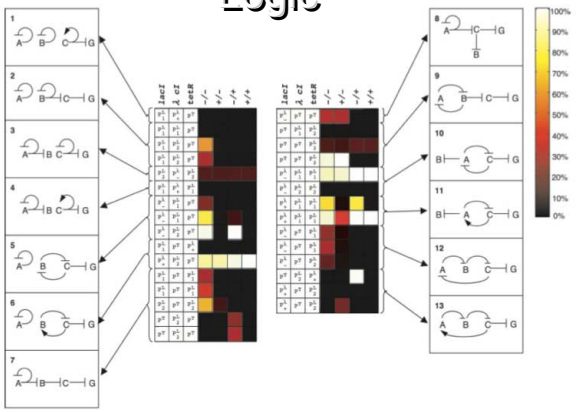
C.C. Guet et al., Combinatorial Synthesis of Genetic Networks, Science 296, 1466-1470, 2002

Example II: Combinatorial Synthetic Logic



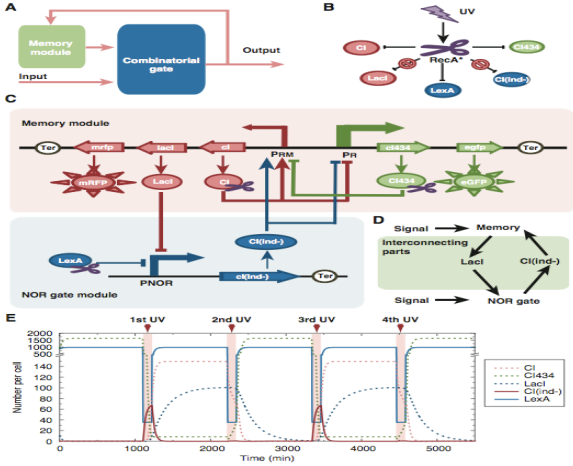
C.C. Guet et al., Combinatorial Synthesis of Genetic Networks, Science 296, 1466-1470, 2002

Example II: Combinatorial Synthetic Logic



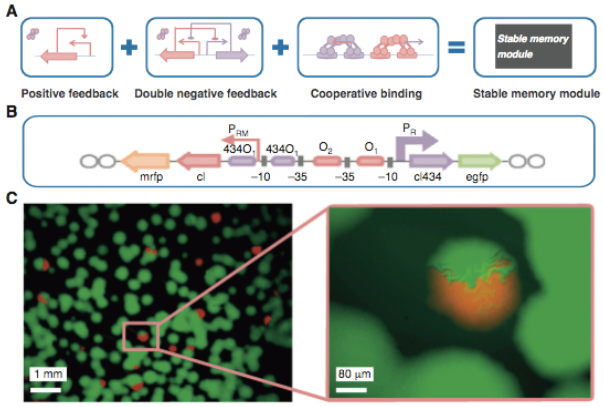
C.C. Guet et al., Combinatorial Synthesis of Genetic Networks, Science 296, 1466-1470, 2002

Example III: Push-on/Push-off circuit



C. Lou et al., Synthesizing a novel genetic sequential logic circuit: a push-on push-off switch, Molecular Systems Biology 6; Article number 350

Example III: Push-on/Push-off circuit

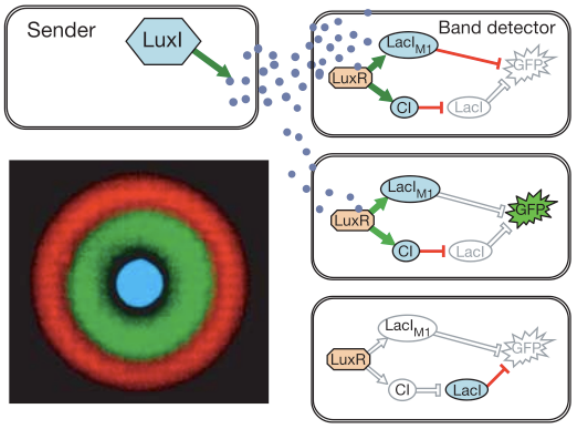


C. Lou et al., Synthesizing a novel genetic sequential logic circuit: a push-on push-off switch, Molecular Systems Biology 6; Article number 350

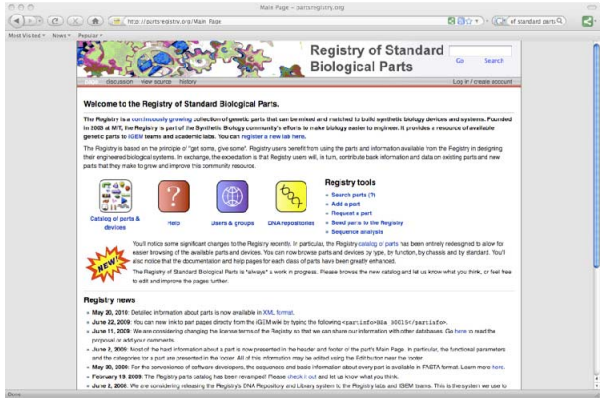
Example IV: Ron Weiss' Pulse Generator

- **Two different bacterial strains** carrying specific synthetic gene regulatory networks are used.
- The first strain produces a diffusible **signal AHL**.
- The second strain possesses a *synthetic gene regulatory* network which produces a **pulse of GFP** after AHL sensing within a range of values (Band Pass).

S. Basu, R. Mehreja, et al. (2004) Spatiotemporal control of gene expression with pulse generating networks, PNAS, 101, 6355-6360



The MIT Registry of Standard Parts



http://partsregistry.org/Main_Page

Outline

- Essential Systems Biology
- Synthetic Biology
- Computational Modeling for Synthetic Biology
- A Note on Ethical, Social and Legal Issues
- Conclusions

What is modeling?

- Is an attempt at describing in a precise way an understanding of the elements of a system of interest, their states and interactions
- A model should be operational, i.e. it should be formal, detailed and "runnable" or "executable".

- "feature selection" is the first issue one must confront when building a model
- One starts from a system of interest and then a decision should be taken as to what will the model include/leave out
- That is, at what level the model will be built

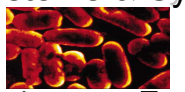
The goals of Modelling

- To capture the essential features of a biological entity/phenomenon
- To disambiguate the understanding behind those features and their interactions
- To move from qualitative knowledge towards quantitative knowledge

Modeling in Systems & Synthetic Biology

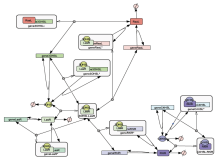
Systems Biology

- *Understanding*
- *Integration*
- *Prediction*
- *Life as it is*

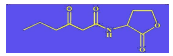


Colonies

Cells



Networks



Synthetic Biology

- *Control*
- *Design*
- *Engineering*
- *Life as it could be*

Computational modelling **to <u>elucidate and characterise</u> *modular patterns* exhibiting robustness, signal filtering, amplification, adaption, error correction, etc.**

Computational modelling **to <u>engineer and evaluate</u> possible *cellular designs* exhibiting a desired behaviour by combining well studied and characterised cellular modules**

There is potentially a distinction between modeling for Synthetic Biology vs Systems Biology:

- Systems Biology is concerned with Biology as it is
- Synthetic Biology is concerned with Biology as it could be

"*Our view of engineering biology focuses on the abstraction and standardization of biological components*" by R. Rettberg @ MIT newsbite August 2006.

"*Well-characterized components help lower the barriers to modeling. The use of control elements (such as temperature for a temperature-sensitive protein, or an exogenous small molecule affecting a reaction) helps model validation*" by Di Ventura et al, Nature, 2006

Co-design of parts and their models hence improving and making both more reliable

Model Development

From [E. Klipp et al, Systems Biology in Practice, 2005]:

1) Formulation of the problem
2) Verification of available information
3) Selection of model structure
4) Establishing a simple model
5) Sensitivity analysis
6) Experimental tests of the model predictions
7) Stating the agreements and divergences between experimental and modelling results
8) Iterative refinement of model

The Challenge of Scales

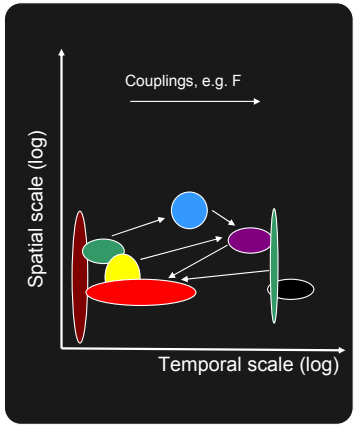
Property	E. Coli	Yeast	Mammalian Cell
Cell Volume	$\sim 1 \mu m^3$	$\sim 1000 \mu m^3$	$\sim 10000 \mu m^3$
Proteins per cell	$\sim 4 \times 10^6$	$\sim 4 \times 10^9$	$\sim 4 \times 10^{10}$
Diffusion time of proteins across the cell volume	~ 0.1 sec	~ 10 sec	~ 100 sec
Diffusion time of small molecules across the cell volume	~ 1 msec	~ 10 msec	~ 0.1 sec
Size of genome	4.6×10^6 bp	1.3×10^7 bp	3×10^9 bp
Size of: Regulator binding site	~ 10 bp	~ 10 bp	~ 10 bp
Promotor	~ 100 bp	~ 1000 bp	$\sim 10^4$ to 10^5 bp
Gene	~ 1000 bp	~ 1000 bp	$\sim 10^4$ to 10^6 bp
Time to transcribe a gene	~ 2 min	~ 2 min	~ 30 min
Time to translate a protein	~ 2 min	~ 2 min	~ 30 min
Timescale for equilibrium biding of small molecule to protein (diffusion limited)	$\sim 1 \mu sec$ ($1 \mu M$ affinity)	~ 1 sec ($1 nM$ affinity)	~ 1 sec ($1 nM$ affinity)
Typical mRNA lifetime	~ 2 - 5 min	~ 10 min to over 1 h	~ 10 min to over 10 h
Ribosomes per cell	$\sim 10^4$	$\sim 10^7$	$\sim 10^8$
Cell generation time	~ 30 min to several hours	~ 2 h to several hours	~ 20 h to nondividing

Within a cell the dissociation constants of DNA/ transcription factor binding to specific/non-specific sites differ by 4-6 orders of magnitude

DNA protein binding occurs at 1-10s time scale very fast in comparison to a cell's life cycle.

R. Milo, et al., BioNumbers—the database of key numbers in molecular and cell biology. Nucleic Acids Research, 1–4 (2009) http://bionumbers.hms.harvard.edu/default.aspx

The Scale Separation Map

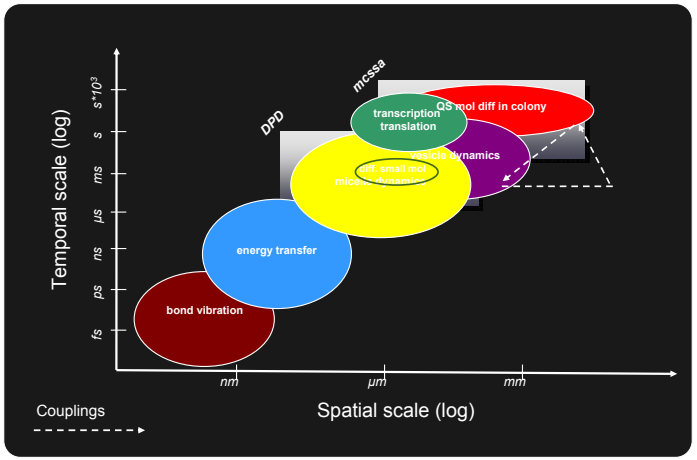


With sufficient data each process can be assigned its space-time region unambiguously

A given process may well have its Δx (respectively Δt) > than another's ξ_A (respectively τ_A)

Hence different processes in the SSM might require different modelling techniques

The Scale Separation Map



The Challenge of Small Numbers in Cellular Systems

- Most commonly recognised **sources of noise** in cellular system are *low number of molecules* and *slow molecular interactions*.

$$\text{Noise} \sim \frac{1}{\sqrt{\sharp\ molecules}}$$

- **Over 80% of genes** in *E. coli* express fewer than a hundred proteins per cell.
- **Mesoscopic, discrete and stochastic** approaches are more **suitable**:
  - Only relevant molecules are taken into account.
  - Focus on the statistics of the molecular interactions and how often they take place.

Mads Karn et al. Stochasticity in Gene Expression: From Theories to Phenotypes. Nature Reviews, 6, 451-464 (2005)

Purnananda Guptasarma. Does replication-induced transcription regulate synthesis of the myriad low copy number poteins of E. Coli. BioEssays, 17, 11, 987-997

Modelling Approaches

There exist many modeling approaches, each with its advantages and disadvantages.

Macroscopic, Microscopic and Mesoscopic

Quantitative and qualitative

Discrete and Continuous

Deterministic and Stochastic

Top-down or Bottom-up

Modelling Frameworks

•Denotational Semantics Models:

Set of equations showing relationships between molecular quantities and how they change over time.
They are approximated numerically.
(I.e. Ordinary Differential Equations, PDEs, etc)

•Operational Semantics Models:

Algorithm (list of instructions) executable by an abstract machine whose computation resembles the behaviour of the system under study. (i.e. Finite State Machine)

D. Harel, "A Grand Challenge for Computing: Full Reactive Modeling of a Multi-Cellular Animal", *Bulletin of the EATCS , European Association for Theoretical Computer Science*, no. 81, 2003, pp. 226-235

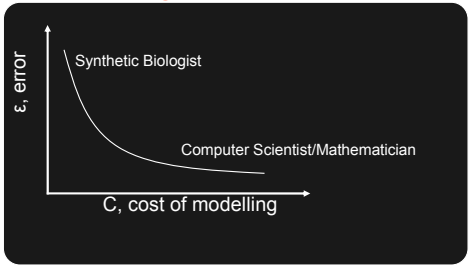
A. Regev, E. Shapiro. The π-calculus as an abstraction for biomolecular systems. Modelling in Molecular Biology., pages 1–50. Springer Berlin., 2004.

asmin Fisher and Thomas Henzinger. Executable cell biology. *Nature Biotechnology*, 25, 11, 1239-1249 (2008)

Tools Suitability and Cost

•From [D.E Goldberg, 2002] (adapted):

"Since science and math are in the description business, the model is the thing…The engineer or inventor has much different motives. The engineered object is the thing"



ODEs (Ordinary Differential Equations) Model

- mainstream model
- a deterministic and continuous model
- limitation: two key assumptions(continuity and determinism)

S-systems Model

- a particular case of ODEs model
- drawback: a large number of model parameters

Linear Weighted Matrices Model

- a deterministic model
- drawback: assuming only linear interactions between the components of cellular systems

Model Design in Systems/Synthetic Biology

• It is a hard process to design suitable models in systems/synthetic biology where one has to consider the choice of the model structure and model parameters at different points repeatedly.

• Some use of computer simulation has been mainly focused on the computation of the corresponding dynamics for a given model structure and model parameters.

• Ultimate goal: for a new biological system (spec) one would like to estimate the model structure and model parameters (that match reality/constructible) simultaneously and automatically.

• Models should be clear & understandable to the biologist

There are good reasons to think that *information processing* is a key viewpoint to take when modeling

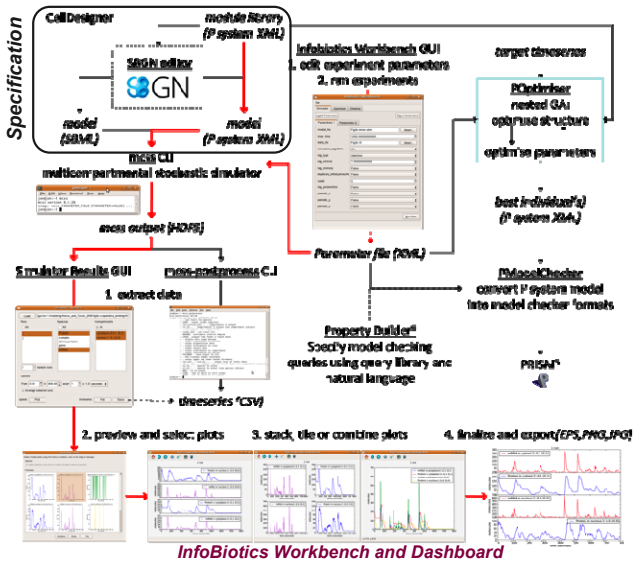
Life as we know is:
- coded in *discrete units* (DNA, RNA, Proteins)
- *combinatorially assembles interactions* (DNA-RNA, DNA-Proteins,RNA-Proteins , etc) through evolution and self-organisation
- Life emerges from these interacting parts
- Information is:
- transported in time (heredity, memory e.g. neural, immune system, etc)
- transported in space (molecular transport processes, channels, pumps, etc)
- Transport in time = storage/memory → a computational process
- Transport in space = communication → a computational process
- Signal Transduction = processing → a computational process

Computer Science Contributions

Methodologies designed to cope with:

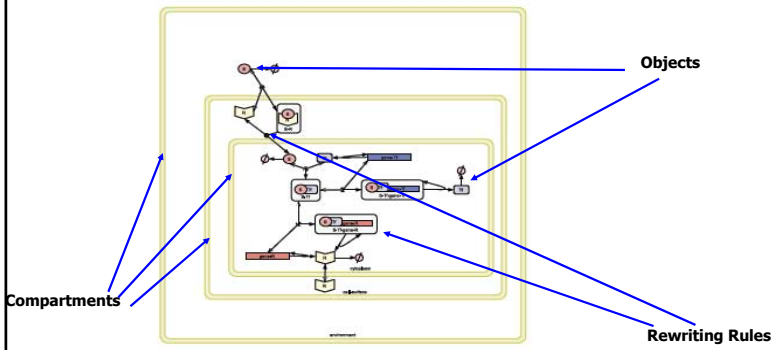
- Languages to cope with complex, concurrent, systems of parts:
  - ∏-calculus
  - Process Calculi
  - P Systems
- Tools to analyse and optimise:
  - EA, ML
  - Model Checking

J.Twycross, L.R. Band, M. J. Bennett, J.R. King, and N. Krasnogor. Stochastic and deterministic multiscale models for systems biology: an auxin-transport case study. BMC Systems Biology, 4(:34), March 2010



InfoBiotics Workbench and Dashboard

Distributed and parallel rewritting systems in *compartmentalised hierarchical structures*.



Computational universality and efficiency.

Modelling Framework

P-Systems: Modelling Principles

Molecules Structured Molecules	Objects Strings
Molecular Species	Multisets of objects/strings
Membranes/organelles	Membrane
Biochemical activity	rules
Biochemical transport	Communication rules

Stochastic P Systems

$$\Pi = (O, L, \mu, M_1, M_2, \ldots, M_n, R_{l_1}, \ldots, R_{l_m})$$

- O is a finite alphabet representing objects;
- $L = \{l_1, \ldots, l_m\}$ is a finite alphabet of labels identifying compartments types.
- μ is a membrane structure containing $n \geq 1$ membranes labelled with elements from L.
- $M_i = (l_i, w_i, s_i)$ is the initial configuration of membrane i with $l_i \in L$, $w_i \in O^*$ a finite multiset of objects and s_i a finite set of strings over O.
- $R_{l_t} = \{r_1^{l_t}, \ldots, r_{k_{l_t}}^{l_t}\}$ is a finite set of rewriting rules associated with compartments of type $l_t \in L$.

Rewriting Rules

Rewriting rules on Multiset of OBJECTS

$$r_j^{l_t} : \; obj_1 \, [\, obj_2 \,]_l \xrightarrow{c_j^{l_t}} obj_1' \, [\, obj_2' \,]_l$$

with $obj_1, obj_2, obj_1', obj_2' \in O^*$ some finite multisets of objects and l a label from L. A multiset of objects, obj is represented as $obj = o_1 + o_2 + \cdots + o_m$ with $o_1, \ldots, o_m \in O$.

Rewriting rules on Multiset of STRINGS

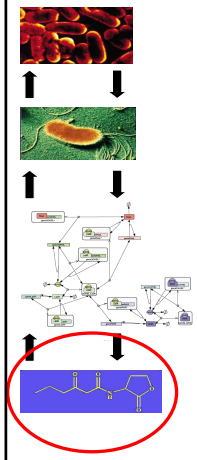
$$r_j^{l_t} : \; [\, obj_1 + str_1 \,]_l \xrightarrow{c_j^{l_t}} [\, obj_1' + str_1' + \cdots + str_p' \,]_l$$

A string str is represented as $str = \langle s_1 . s_2 . \cdots . s_i \rangle$ where $s_1, \ldots, s_i \in O$.

A stochastic constant $c_j^{l_t}$ is associated specifically with each rule.

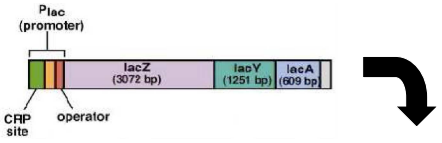
used by Multi-volume Gillespie's algorithm

Molecular Species



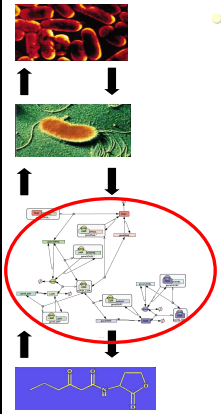
- A molecular species can be represented using **individual objects**.

- A molecular species with relevant internal structure can be represented **using a string**.



$$\langle cap \times prom \times op \times lacZ_1 \cdots lacZ_{30} \times lacY_1 \cdots lacY_{12} \times lacA_1 \cdots lacA_6 \rangle$$

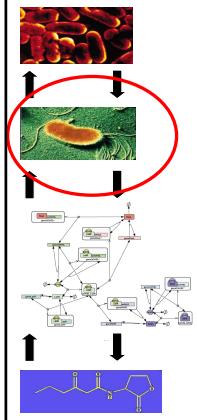
Molecular Interactions



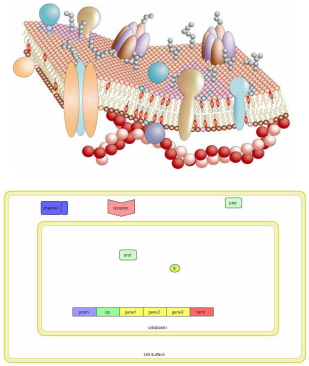
- **Comprehensive and relevant rule-based schema** for the most common molecular interactions taking place in living cells.

  - Transformation/Degradation
  - Complex Formation and Dissociation
  - Diffusion in / out
  - Binding and Debinding
  - Recruitment and Releasing
  - Transcription Factor Binding/Debinding
  - Transcription/Translation

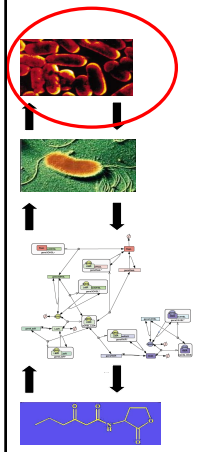
Compartments / Cells



- **Compartments and regions are explicitly** specified using membrane structures.

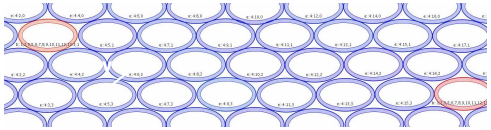


Colonies / Tissues



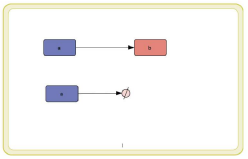
- Colonies and tissues are representing as **collection** of P systems distributed over a lattice.
- Objects can travel around the lattice through **translocation** rules.

$$[\,obj\,]_l \overset{\mathbf{v}}{\leftrightarrows} [\ ] \xrightarrow{c_i^j} [\,]_l \overset{\mathbf{v}}{\leftrightarrows} [\,obj\,]$$



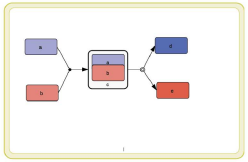
Molecular Interactions Inside Compartments

Transformation and degradation:



$$r_1: [\,a\,]_l \xrightarrow{c} [\,b\,]_l$$
$$r_2: [\,a\,]_l \xrightarrow{c} [\ ]_l$$
$$p(r_i) = c \cdot |a| \quad i = 1,2$$

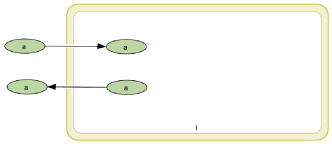
Complex formation and dissociation:



$$r_{cf}: [\,a+b\,]_l \xrightarrow{c_{cf}} [\,c\,]_l$$
$$r_{cd}: [\,c\,]_l \xrightarrow{c_{cd}} [\,d+e\,]_l$$
$$p(r_{cf}) = \begin{cases} c_{cf} \cdot |a||b| & a \neq b \\ c_{cf} \cdot \dfrac{|a|(|a|-1)}{2} & a = b \end{cases}$$
$$p(r_{cd}) = c_{cd} \cdot |c|$$

Passive Diffusion of Molecules

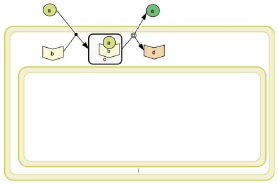
Diffusion in and out:



$$r_1: \quad a\,[\ ]_l \xrightarrow{c_{in}} [\ a\ ]_l \qquad p(r_1) = c_{in}|a|$$
$$r_2: \quad [\ a\ ]_l \xrightarrow{c_{out}} a\,[\ ]_l \qquad p(r_2) = c_{out}|a|$$

Signal Sensing and Active Transport

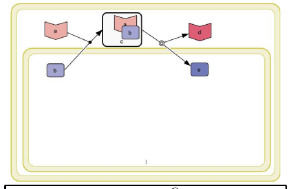
Binding and debinding:



$$r_1: \quad a\,[\ b\ ]_l \xrightarrow{c_{lb}} [\ c\ ]_l$$
$$r_2: \quad [\ c\ ]_l \xrightarrow{c_{ld}} d\,[\ e\ ]_l$$

$$p(r_1) = c_{lb}|a||b|$$
$$p(r_2) = c_{ld}|c|$$

Recruitment and releasing:

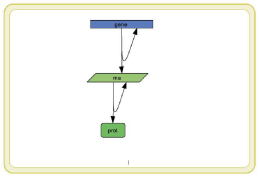


$$r_{cf}: a\,[\ b\ ]_l \xrightarrow{c_{cf}} c\,[\ ]_l$$
$$r_{cd}: c\,[\ ]_l \xrightarrow{c_{cd}} d\,[\ e\ ]_l$$

$$p(r_{rt}) = c_{rt}|a||b|$$
$$p(r_{rl}) = c_{rl}|c|$$

Specification of Transcriptional Regulatory Networks

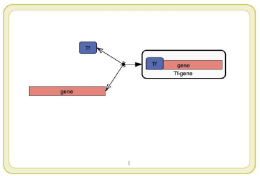
Transcription and Translation:



$$r_{tc}: [\ gene\ ]_l \xrightarrow{c_{tc}} [\ gene + rna\ ]_l$$
$$r_{tl}: [\ rna\ ]_l \xrightarrow{c_{tl}} [\ rna + prot\ ]_l$$

$$p(r_{tc}) = c_{tc}|gene|$$
$$p(r_{tl}) = c_{tl}|rna|$$

Transcription factor binding and debinding:

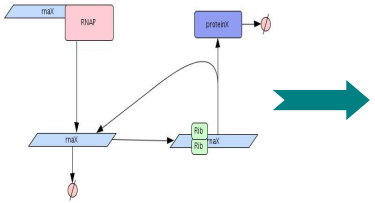


$$r_{gon}: [\ Tf + gene\ ]_l \xrightarrow{c_{gon}} [\ Tfgene\ ]_l$$
$$r_{goff}: [\ Tfgene\ ]_l \xrightarrow{c_{goff}} [\ Tf + gene\ ]_l$$

$$p(r_{gon}) = c_{gon}|Tf||gene|$$
$$p(r_{goff}) = c_{goff}|Tfgene|$$

Post-Transcriptional Processes

- For each protein in the system, post-transcriptional processes like translational initiation, messenger and protein degradation, protein dimerisation, signal sensing, signal diffusion etc are represented using modules of rules.
- Modules can have also as parameters the stochastic kinetic constants associated with the corresponding rules in order to allow us to explore possible mutations in the promoters and ribosome binding sites in order to optimise the behaviour of the system.



$$PostTransc(\{X\},\{c_1,\ldots,c_5\}) = \begin{cases} [\ rnaX{\cdot}RNAP\ ]_b \xrightarrow{c_1} [\ rnaX\ ]_b \\ [\ rnaX\ ]_b \xrightarrow{c_2} [\ ]_b \\ [\ rnaX\ ]_b \xrightarrow{c_3} [\ rnaX + proteinX{\cdot}Rib\ ]_b \\ [\ proteinX{\cdot}Rib\ ]_b \xrightarrow{c_4} [\ proteinX\ ]_b \\ [\ proteinX\ ]_b \xrightarrow{c_5} [\ ]_b \end{cases}$$

Scalability through Modularity

Cellular functions arise from **orchestrated interactions between motifs** consisting of many molecular interacting species.

A ***P System model*** is a **set of rules** representing molecular interactions motifs that appear in many cellular systems.

Basic P System Modules Used

Module Name	Type No	Module Size	Module	Biological Function
Com	3	1	$Com(\{X,Y,Z\},\{c\},\{l\}) = \{[X+Y]_l \xrightarrow{c} [Z]_l\}$	complex formation
Diss	4	1	$Diss(\{X,Y,Z\},\{c\},\{l\}) = \{[X]_l \xrightarrow{c} [Y+Z]_l\}$	complex dissociation
UnReg	0	4	$UnReg(\{G,R,P\},\{c_1,c_2,c_3,c_4\},\{l\}) = \left\{\begin{array}{l}[G]_l \xrightarrow{c_1} [G+R]_l \\ [R]_l \xrightarrow{c_2} [R+P]_l \\ [R]_l \xrightarrow{c_3} [] \\ [P]_l \xrightarrow{c_4} []_l\end{array}\right\}$	unregulated expression
Pos	1	6	$Pos(\{Act,G,R,P\},\{c_1,c_2,c_3,c_4,c_5,c_6\},\{l\}) = \left\{\begin{array}{l}[Act+G]_l \xrightarrow{c_1} [Act.G]_l \\ [Act.G]_l \xrightarrow{c_2} [Act+G]_l \\ [Act.G]_l \xrightarrow{c_3} [Act.G+R]_l \\ [R]_l \xrightarrow{c_4} [R+P]_l \\ [R]_l \xrightarrow{c_5} []_l \\ [P]_l \xrightarrow{c_6} []_l\end{array}\right\}$	positive regulated expression
Neg	2	2	$Neg(\{Rep,G\},\{c_1,c_2\},\{l\}) = \left\{\begin{array}{l}[Rep+G]_l \xrightarrow{c_1} [Rep.G]_l \\ [Rep.G]_l \xrightarrow{c_2} [Rep+G]_l\end{array}\right\}$	negative regulated expression

Characterisation/Encapsulation of Cellular Parts: Gene Promoters

- A **modeling language** for the design of synthetic bacterial colonies.
- A ***module***, set of rules describing the molecular interactions involving a cellular part, provides **encapsulation and abstraction**.
- Collection or **libraries of reusable cellular parts and reusable models.**

E. Davidson (2006) The Regulatory Genome, Gene Regulation Networks in Development and Evolution, Elsevier

Module Variables: Recombinant DNA, Directed Evolution, Chassis selection

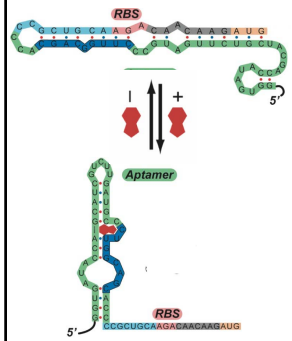
- 

- 

-

Characterisation/Encapsulation of Cellular Parts: Riboswitches

❑ A **riboswitch** is a piece of RNA that *folds* in different ways depending on the presence of absence of specific molecules regulating translation.

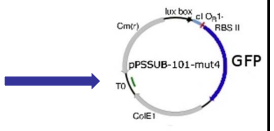


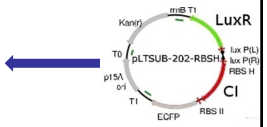
Characterisation/Encapsulation of Cellular Parts: Degradation Tags

❑ **Degradation tags** are amino acid sequences recognised by proteases. Once the corresponding DNA sequence is fused to a gene the half life of the protein is reduced considerably.

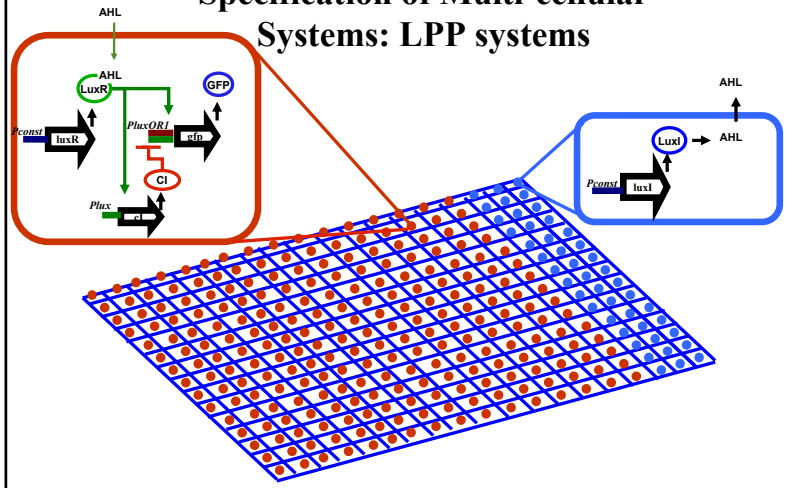
Higher Order Modules: Building Synthetic Gene Circuits

PluxOR1	ToppRibo	geneX	degLVA



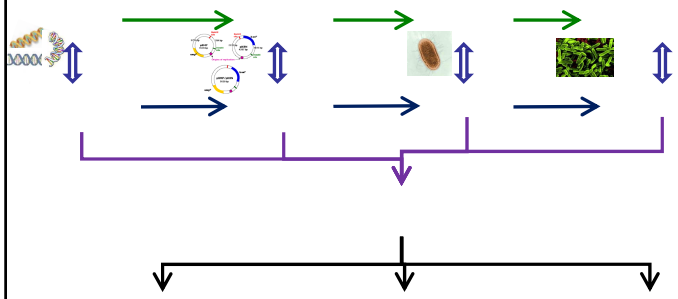


Specification of Multi-cellular Systems: LPP systems



Infobiotics: An Integrated Framework

http://www.infobiotics.org/infobiotics-workbench/



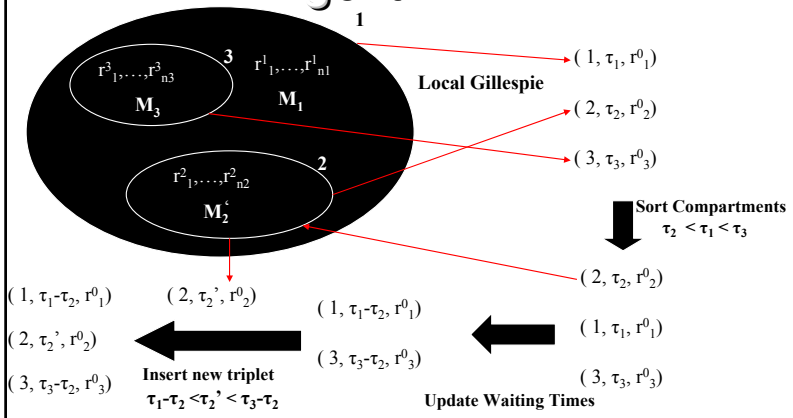
Stochastic P Systems Are *Executable* Programs

The virtual machine running these programs is a "Gillespie Algorithm (SSA)". It generates trajectories of a stochastic syste:

- A stochastic constant is associated with each rule.
- A propensity is computed for each rule by multiplying the stochastic constant by the number of distinct possible combinations of the elements on the left hand side of the rule.

F. J. Romero-Campero, J. Twycross, M. Camara, M. Bennett, M. Gheorghe, and N. Krasnogor. Modular assembly of cell systems biology models using p systems. International Journal of Foundations of Computer Science, 2009

Multicompartmental Gillespie Algorithm



An Important Difference with "Normal" Programs

- Executable Stochastic P systems are *not* programs with stochastic behavior

```
function f1(p1,p2,p3,p4)
{
if (p1<p2) and (rand<0.5)
        print p3
else
        print p4
}
```

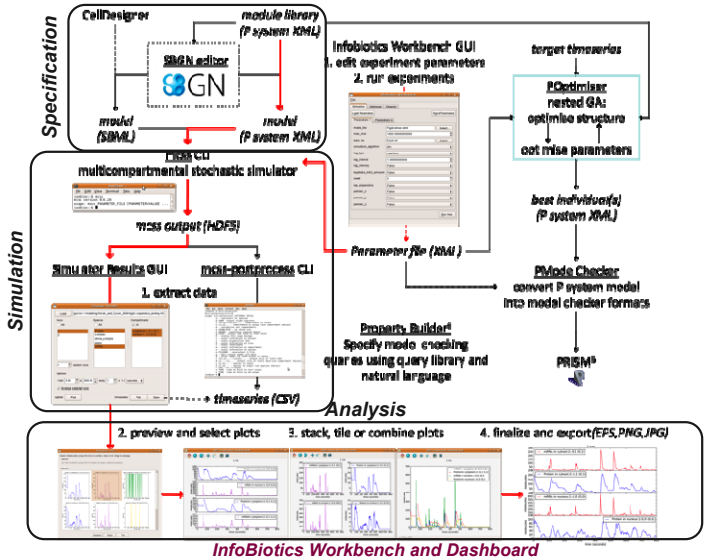
```
function f1(p1,p2,p3,p4)
{
if (p1<p2)
                RND
        print p3
                RND
else
                        RND
        print p4
                RND
}
```

- A cell is a living example of distributed stochastic computing.

Rapid Model Prototyping

- Using P systems modules one can model a large variety of commonly occurring BRN:
  - Gene Regulatory Networks
  - Signaling Networks
  - Metabolic Networks
- This can be done in an incremental way.

F. J. Romero-Campero, J. Twycross, M. Camara, M. Bennett, M. Gheorghe, and N. Krasnogor. Modular assembly of cell systems biology models using p systems. International Journal of Foundations of Computer Science, 2009

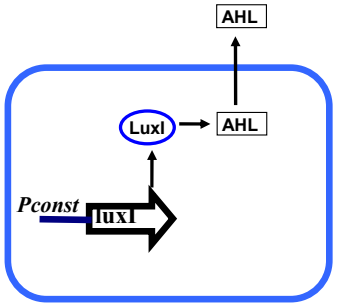


An example: A Pulse Generator

- **Two different bacterial strains** carrying specific synthetic gene regulatory networks are used.
- The first strain produces a diffusible **signal AHL**.
- The second strain possesses a *synthetic gene regulatory* network which produces a **pulse of GFP** after AHL sensing within a range of values (Band Pass).

S. Basu, R. Mehreja, et al. (2004) Spatiotemporal control of gene expression with pulse generating networks, PNAS, 101, 6355-6360

Sender Cells



SenderCell()=

{

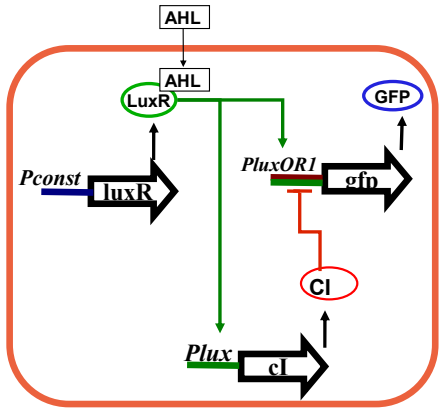
Pconst({X = luxI },…)

PostTransc({X=LuxI},{c_1=3.2,…}

Diff({X=AHL},{c=0.1})

}

Pulse Generating Cells



PulseGenerator()=

{

Pconst({X=luxR},…)

PluxOR1({X=gfp},…)

Plux({X=cI},…)

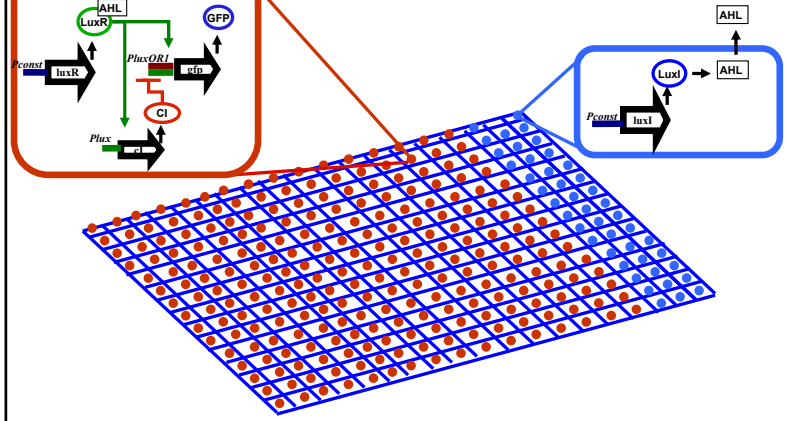
…

…

Diff({X=AHL},…)

}

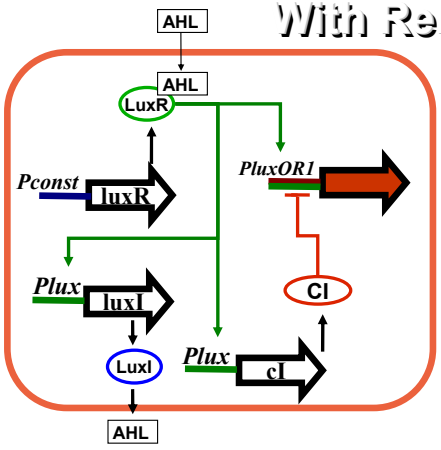
Spatial Distribution of Senders and Pulse Generators



Pulse propagation - simulation I

Simulation I

Pulse Generating Cells With Relay



PulseGenerator(X) =

{

Pconst({X=luxR},…)

PluxOR1({X},…) ,

Plux({X=cI},…) ,

…

Diff({X=AHL},…) ,

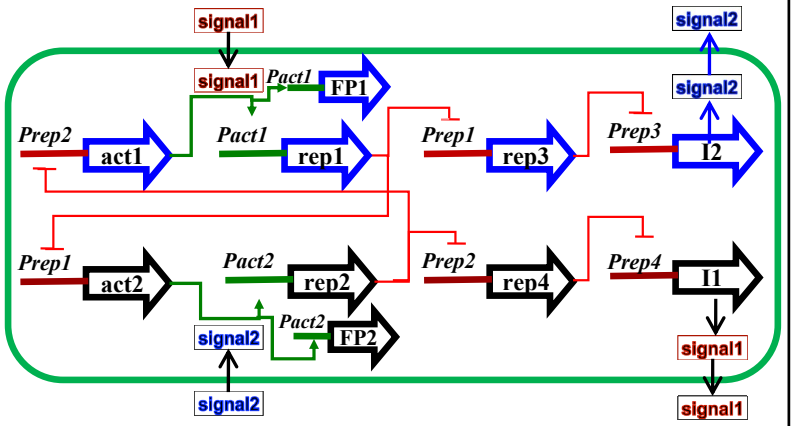
Plux({X=luxI},…)

}

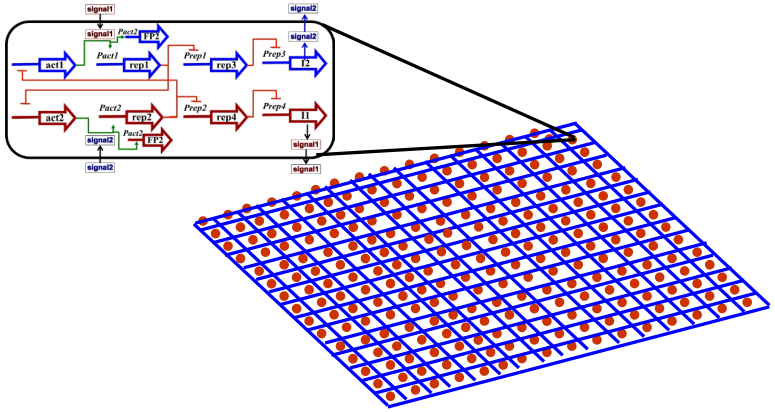
Pulse propagation & Rely-simulation II

Simulation II

A Signal Translator for Pattern Formation



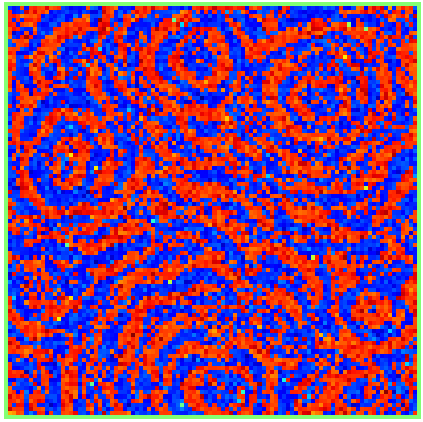
Uniform Spatial Distribution of Signal Translators for Pattern Formation



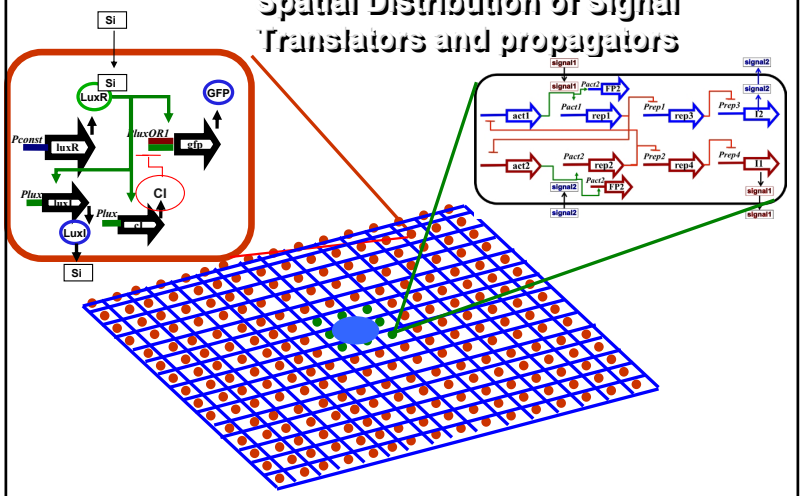
Pattern Formation in synthetic bacterial colonies

Simulation III

Pattern Formation in synthetic bacterial colonies



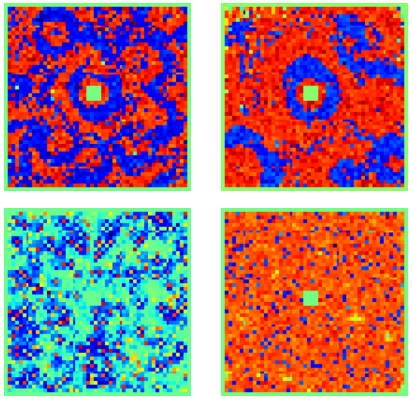
Spatial Distribution of Signal Translators and propagators



Alternating signal pulses in synthetic bacterial colonies

Simulation IV

"What if?" Scenarios



Outline

- Essential Systems Biology
- Synthetic Biology
- Computational Modeling for Synthetic Biology
- A Note on Ethical, Social and Legal Issues
- Conclusions

Synthetic Biology

The new science of synthetic biology aims to **re-engineer life** at the molecular level and even **create completely new forms of life**. It has the **potential** to **create new medicines, biofuels, assist climate change through carbon capture, and develop solutions to help clean up the environment**.

BLUE
GREEN
RED

What is Synthetic Biology?

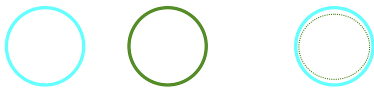
Synthetic Biology is http://syntheticbiology.org/

A) the design and construction of new biological parts, devices, and systems, and

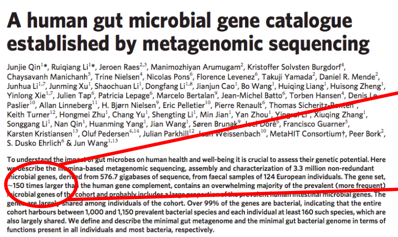
B) the re-design of existing, natural biological systems for useful purposes.

Been There, Done That

The Unjustified Fear of Chimera and Foreign Genes Transplantation



Synechosystis
Bacillus subtilis



A human gut microbial gene catalogue established by metagenomic sequencing

Junjie Qin[1]*, Ruiqiang Li[1]*, Jeroen Raes[2,3], Manimozhiyan Arumugam[2], Kristoffer Solvsten Burgdorf[4], Chaysavanh Manichanh[5], Trine Nielsen[4], Nicolas Pons[6], Florence Levenez[6], Takuji Yamada[2], Daniel R. Mende[2], Junhua Li[1,7], Junming Xu[1], Shaochuan Li[1], Dongfang Li[1,8], Jianjun Cao[1], Bo Wang[1], Huiqing Liang[1], Huisong Zheng[1], Yinlong Xie[1,7], Julien Tap[6], Patricia Lepage[6], Marcelo Bertalan[9], Jean-Michel Batto[6], Torben Hansen[4], Denis Le Paslier[10], Allan Linneberg[11], H. Bjørn Nielsen[9], Eric Pelletier[10], Pierre Renault[6], Thomas Sicheritz-Ponten, Keith Turner[12], Hongmei Zhu[1], Chang Yu[1], Shengting Li[1], Min Jian[1], Yan Zhou[1], Yingrui Li, Xiuqing Zhang[1], Songgang Li[1], Nan Qin[1], Huanming Yang[1], Jian Wang[1], Søren Brunak[9], Joel Doré[6], Francisco Guarner[5], Karsten Kristiansen[13], Oluf Pedersen[4,14], Julian Parkhill[12], Jean Weissenbach[10], MetaHIT Consortium†, Peer Bork[2], S. Dusko Ehrlich[6] & Jun Wang[1,13]

To understand the impact of gut microbes on human health and well-being it is crucial to assess their genetic potential. Here we describe the Illumina-based metagenomic sequencing, assembly and characterization of 3.3 million non-redundant microbial genes, derived from 576.7 gigabases of sequence, from faecal samples of 124 European individuals. The gene set, ~150 times larger than the human gene complement, contains an overwhelming majority of the prevalent (more frequent) microbial genes of the cohort and probably includes a large proportion of the prevalent human intestinal microbial genes. The genes are largely shared among individuals of the cohort. Over 99% of the genes are bacterial, indicating that the entire cohort harbours between 1,000 and 1,150 prevalent bacterial species and each individual at least 160 such species, which are also largely shared. We define and describe the minimal gut metagenome and the minimal gut bacterial genome in terms of functions present in all individuals and most bacteria, respectively.

150 times larger than the human genome

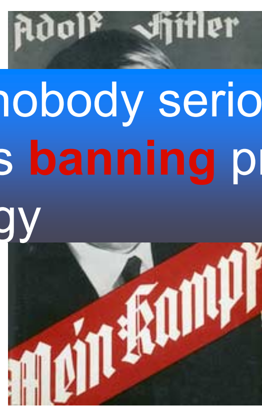
A Technology Not a Product

" The problem of deaths and injury as a result of road accidents is now acknowledge to be a global phenomenon.... publications show that in 1990 road accidents as a cause of death or dissability were in ninth place out of a total of over 100 identified causes.... by 2020 forecasts suggest... road accidents will

And yet nobody seriously considers **banning** mechanical engineering

Asia-Pacific	228,405	1,15	262,666	1,25	328,332	1,50	393,999
LAC	64,699	1,15	74,404	1,25	93,005	1,50	111,606
MENA	28,864	1,15	33,194	1,25	41,492	1,50	49,790
Global	542,661		616,056		745,769		876,539

A Technology Not a Product



And yet nobody seriously considers **banning** printing technology

But ....

- Technologies ***are*** regulated:
  - Cars **have** seat belts and laws establish speed limits
  - Mein Kampf **is banned** in Germany and by, restricting google, China (e.g.) bans uncountable written material of all sorts!
- Societies **must** establish *an **informed dialogue*** involving:
  - tax payers
  - Scientists
  - Lobbies of al sorts
  - Government

What *IS* Synthetic Biology?

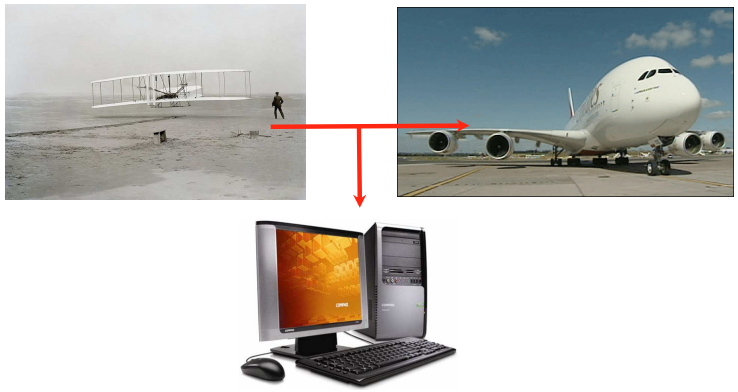
Synthetic Biology is http://syntheticbiology.org/

A) the design and construction of new biological parts, devices, and systems, and

B) the re-design of existing, natural biological systems for useful purposes.

C) Through rigorous mathematical, computational engineering routes

Biology only smarter, safer and clearer



Outline

- Essential Systems Biology
- Synthetic Biology
- Computational Modeling for Synthetic Biology
- A Note on Ethical, Social and Legal Issues
- Conclusions

Summary & Conclusions

- These lectures have focused on an integrative methodology for Systems & Synthetic Biology
  - Executable Biology
  - Parameter and Model Structure Discovery
- Computational models (or executable in Fisher & Henzinger's jargon) adhere to (a degree) to an operational semantics.
- Refer to the excellent review [Fisher & Henzinger, Nature Biotechnology, 2007]

108

Summary & Conclusions

- The gap present in mathematical models between the model and its algorithmic implementation disappears in computational models as all of them are algorithms.
- A new gap appears between the biology and the modeling technique and this can be solved by a judicious "feature selection", i.e. the selection of the correct abstractions
- Good computational models are more intuitive and analysable

109

Summary & Conclusions

- Computational models can thus be executed (quite a few tools out there, lots still missing)
- Quantitative VS qualitative modelling: computational models can be very useful even when not every detail about a system is known.
- Missing Parameters/model structures can sometimes be fitted with of-the-shelf optimisation strategies (e.g. COPASI, GAs, etc)
- Computational models can be analysed by model checking: thus they can be used for testing hypothesis and expanding experimental data in a principled way

110

Summary & Conclusions

- Some really nice tutorials and other sources:
  - Luca Caderlli's BraneCalculus & BioAmbients
  - Simulating Biological Systems in the Stochastic π−calculus by Phillips and Cardelli
  - From Pathway Databases to Network Models by Aguda and Goryachev
  - Modeling and analysis of biological processes by Brane Calculi and Membrane Systems by Busi and Zandron
  - D. Gilbert's website contain several nice papers with related methods and tutorials

111

Other Sources

F. J. Romero-Campero, J. Twycross, M. Camara, M. Bennett, M. Gheorghe, and N. Krasnogor. Modular assembly of cell systems biology models using p systems. International Journal of Foundations of Computer Science, (to appear), 2009.

F.J. Romero-Camero and N. Krasnogor. An approach to biomodel engineering based on p systems. In Proceedings of Computation In Europe (CIE 2009), 2009.

J. Smaldon, N. Krasnogor, M. Gheorghe, and A. Cameron. Liposome logic. In Proceedings of the 2009 Genetic and Evolutionary Computation Conference (GECCO 2009), 2009

F. Romero-Campero, H.Cao, M. Camara, and N. Krasnogor. Structure and parameter estimation for cell systems biology models. In Maarten Keijzer et.al, editor, Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2008), pages 331-338. ACM Publisher, 2008. This paper won the Best Paper award at the Bioinformatics track.

J. Smaldon, J. Blake, D. Lancet, and N. Krasnogor. A multi-scaled approach to artificial life simulation with p systems and dissipative particle dynamics. In Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2008). ACM Publisher, 2008.

112

Acknowledgements

Members of my team working on SB[2]

- Jonathan Blake  Integrated Environment
- Daven Sanassy  Modeling & Model Checking
- Karima Righetti  Molecular Micro-Biology
- Jamie Twycross  Stochastic Simulations

BB/F01855X/1
BB/D019613/1

EP/E017215/1
EP/H024905/1

University of Nottingham

Prof. M. Camara, Dr. S. Heeb, Prof. P. Williams

GECCO 2011 organisers for inviting this tutorial.

You for listening!

Any Questions?

www.synbiont.org

Become a member and have access to a large international community of Synthetic Biologists