Synthetic biology: basic concepts

Synthetic biology is in many aspects similar to **electric engineering**: cellular decision-making processes share basic operations with electronic control circuits. Intra- and extracellular information is collected by sensors that communicate the input signal states into a network, which processes the data according to logic and arithmetic operations. These operations result in decisions that are finally executed by output signals.

- Synthetic biology follows a hierarchicla structure, building up systems from smaller components.
- An important aspect of synthetic biology is the application of systemic design. This approach is based on the engineering principles of **modularity**, **characterization** and **standartization**.

Synthetic biology: abstraction hierarchy and modularity

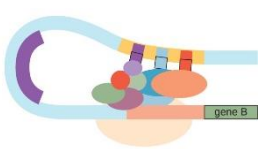
Engineering

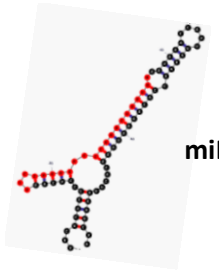
Synthetic Biology

Parts

protein

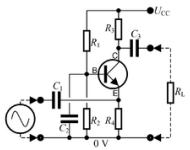
promoter

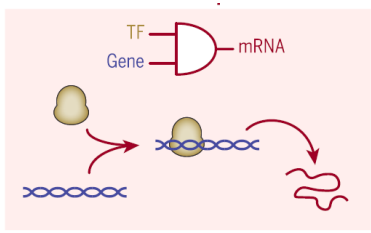




miRNA

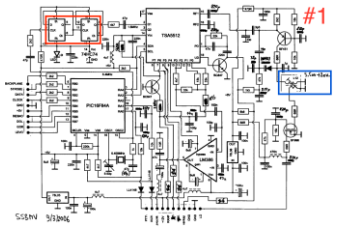
Devices

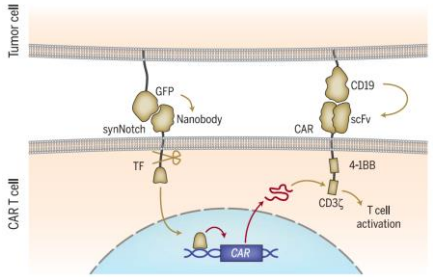




Inducible protein expression module

Systems





Synthetic therapeutic program to specifically kill a cancer cell

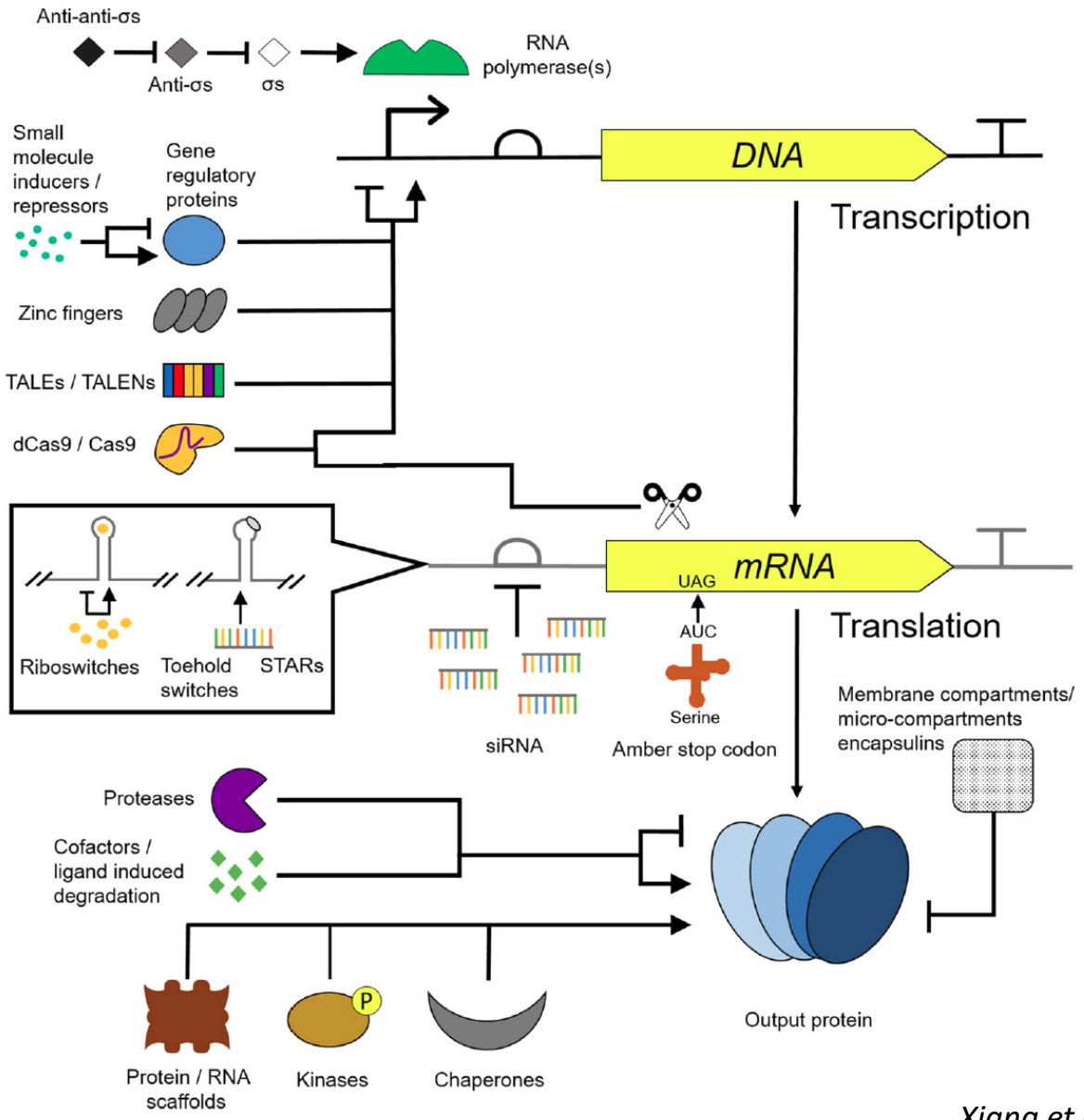
Basic biological functions

Combination of parts performing a human defined function

Combination of devices

Tools (parts) for regulating gene expression

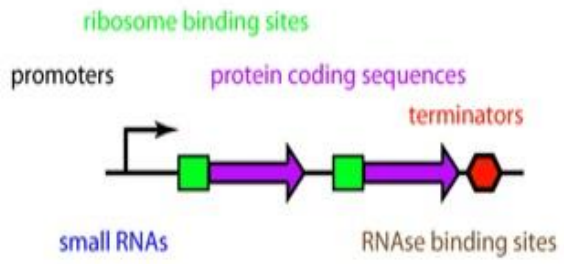
Fig. 3 Expanded toolbox for engineering complex gene regulation programs. These include using proteins that affect DNA transcription and RNA translation through protein-DNA and protein-RNA base pair binding. Also shown is the ability to use RNA secondary structure and base pair binding to control mRNA translation initiation. Protein activity can also be controlled by other proteins, through protein–protein interactions or enzymatic reactions that modulate activity. The activity of many regulators can be controlled by small molecule ligands/cofactors. (σ = sigma factors, STARs = small transcriptional activating RNAs, siRNA = small interfering RNA,
TALE(N)s = transcription activator-like effector (nuclease)



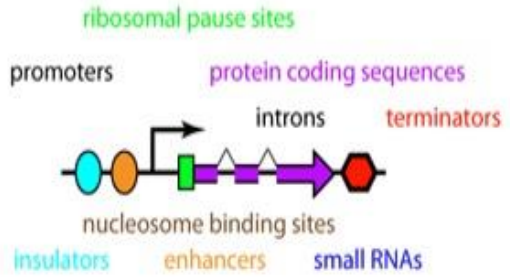
Xiang et al., Natural Comp 2018

Parts

- **Promoters**: recruits RNA polymerase and other accessory proteins to prime the transcription of mRNA. Promoters have different strength and can be constitutive or regulated (inducible, repressible).
- **Teminators**: signal termination of transcription, polyA signals.
- **Ribosome binding sites**: recruit ribosome for initiation of translation (Shine-Delgarno nad Kozak sequences). These sequences affect efficiency of transaltion and hence protein production.
- **Translational riboswitches**: a regulatory segment within a mRNA that can bind a small molecule, which in turn affects translation (eg lysisn riboswitch).
- **Protein coding sequences**: encode transcriptional activators and repressors, transcription factors, sensors (e.g. cryptochromes for light perception, receptors for chemical ligands), signaling molecules (kinases, proteases), protein scaffolds and „output“ proteins (antibiotic resistance, GFP, LUX, suicidal proteins).
- **Factors affecting RNA stability**: bacterial mRNA has very short half life (average 2 min in E.coli). In contrast, stability of eukaryotic mRNA can range form minutes to days. *Cis*-elements affecting RNA stability: length of 3´UTR, hairpins, RNA binding motives, introns.
- **siRNAs, miRNAs**: affect gene expression on postrancriptional or translational level.



Prokaryotic Genetic Parts

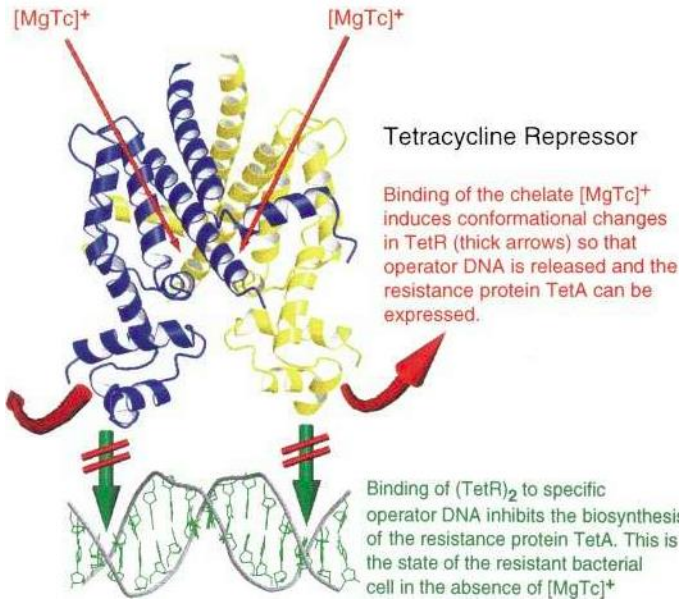


Eukaryotic Genetic Parts

Orthogonality in synthetic biology

Synthetic biology approaches commonly introduce heterologous gene networks into a host to predictably program cells, with the expectation of the synthetic network being **orthogonal** (non-interferig) to the host background and to other synthetic networks. It also implies context independent performance of a synthetic network.

TetR repressor

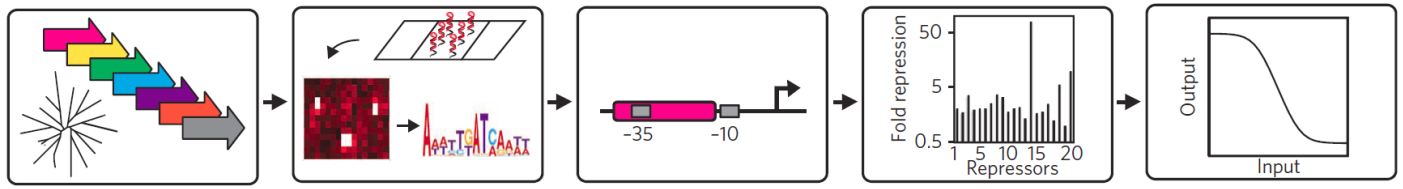


- Relatively short operator sequence
- Often used in synthetic genetic circuits
- Thousands of homologues that exhibit binding specificities to disparate operators are available

Mining of prokaryotic genomes for orthogonal TetR repressors

Gene Bank: 82,017 TetR reprssors

73 preselected for experimental characterization

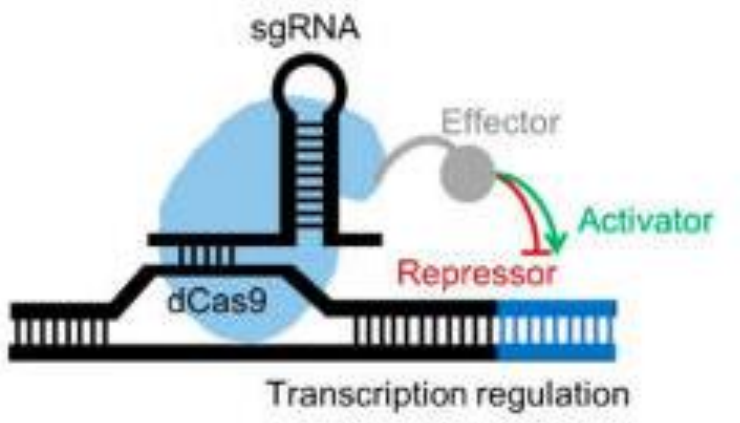


2.1 Mio 28 bp inverted repeats screened on microarrays using Cy5 labeled TetRs

16 operator-repressor pairs that exhibit strong repression and minimal corss-talk with other promoters/repressors

Stanton et al, Nat Chem Biol 2014

Synthetic transcription regulators based on TALE and CRISPR/Cas9



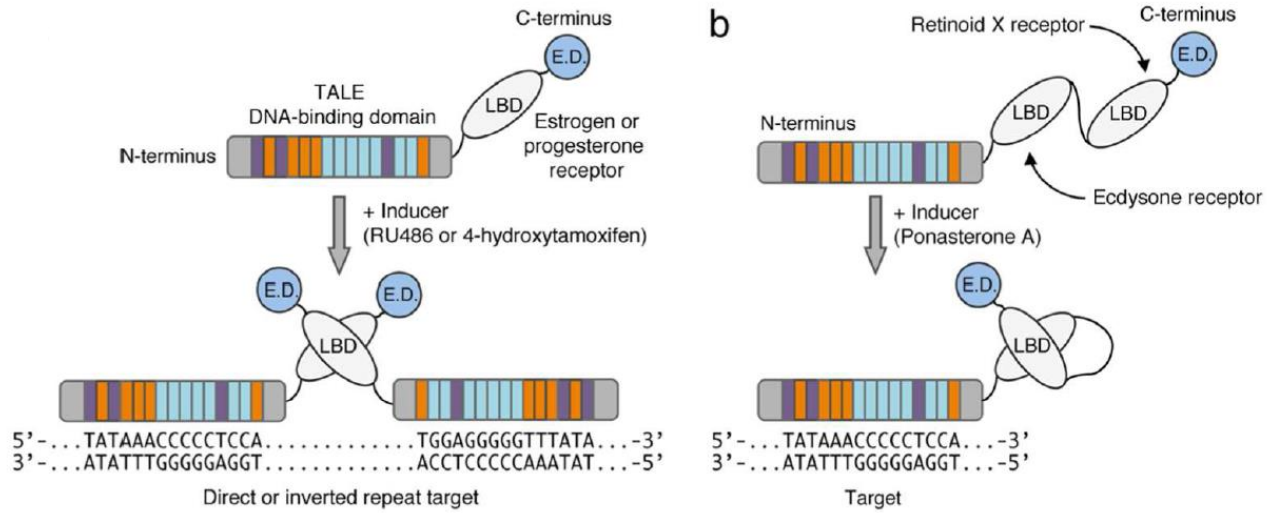
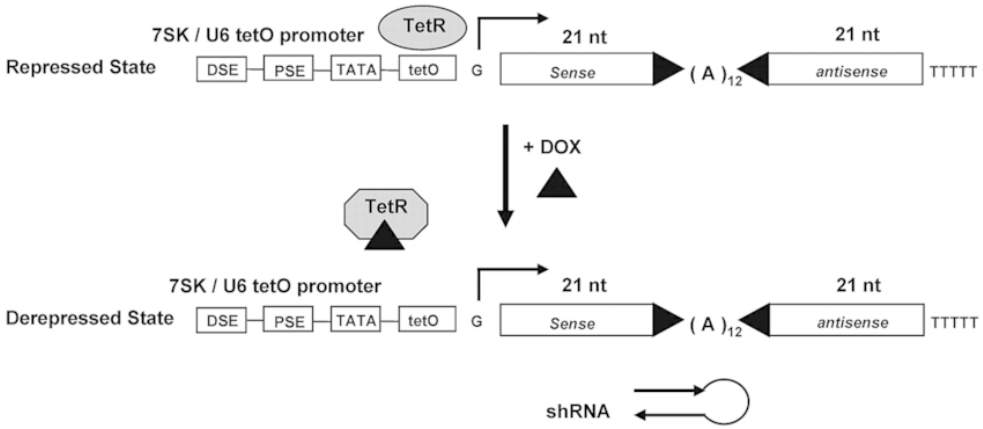


Figure 1. Ligand-inducible TALE transcription factors. (a) TALE-TF proteins fused to ligand-binding domains (LBDs) from the estrogen receptor (ER) or progesterone receptor (PR) undergo intermolecular dimerization in response to 4-hydroxytamoxifen (4-OHT) or RU486, respectively, and up-regulate gene activation from DNA sequences that contain two direct or inverted repeat TALE binding sites. (b) TALE-TF proteins fused to the chimeric single-chain retinoid X-α/ecdysone (RXE) LBD undergo intramolecular rearrangement in response to ponasterone A (PonA) and up-regulate gene activation from target DNA that contains only a single TALE binding site. E.D. indicates effector domain.

Mercer et al, ACS Synt Biol 2013

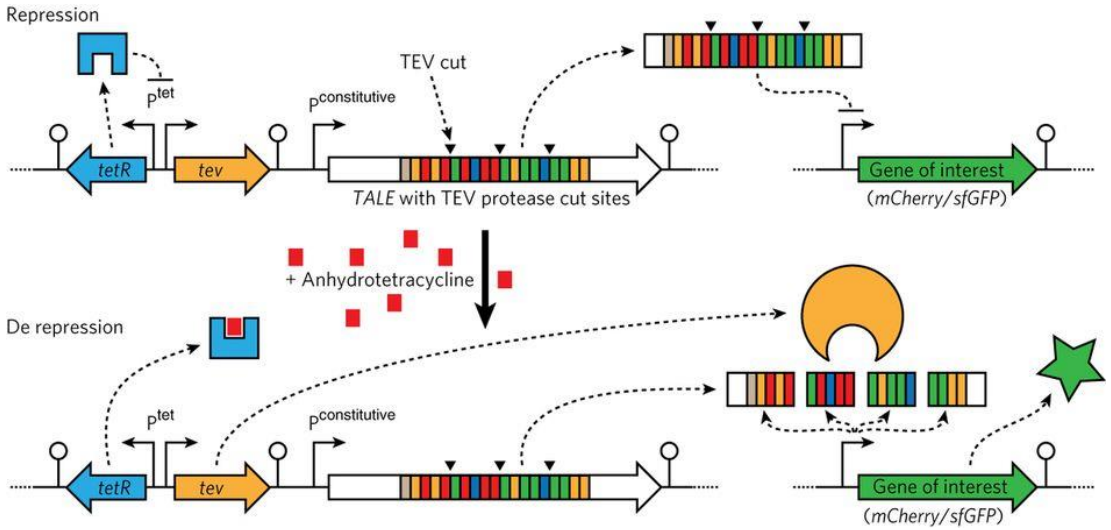
Devices

- Combination of parts performing a human defined function.



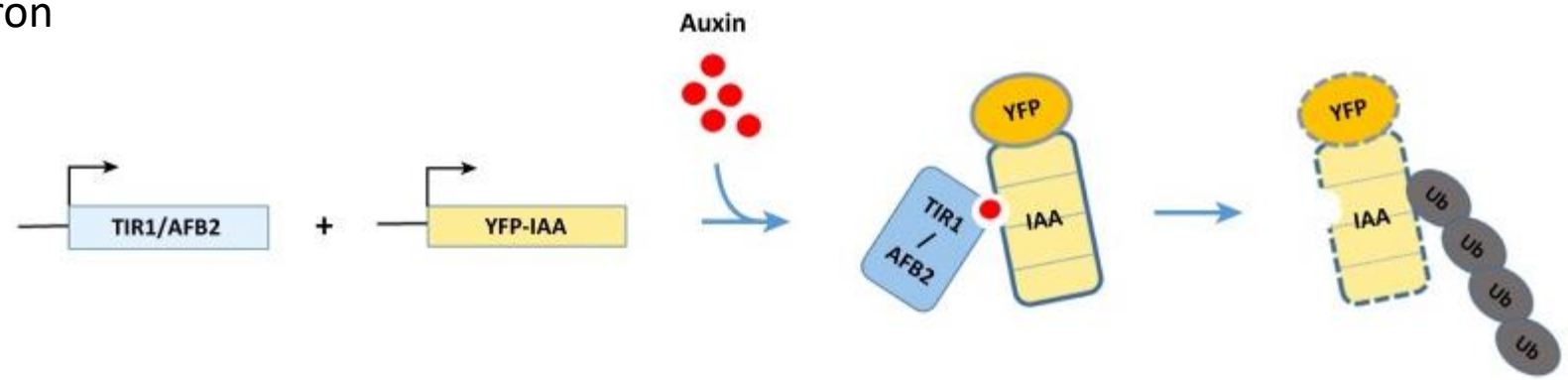
Tetracycline regulated siRNA production

Tetracycline regulated protein production via proteolytic cleavage

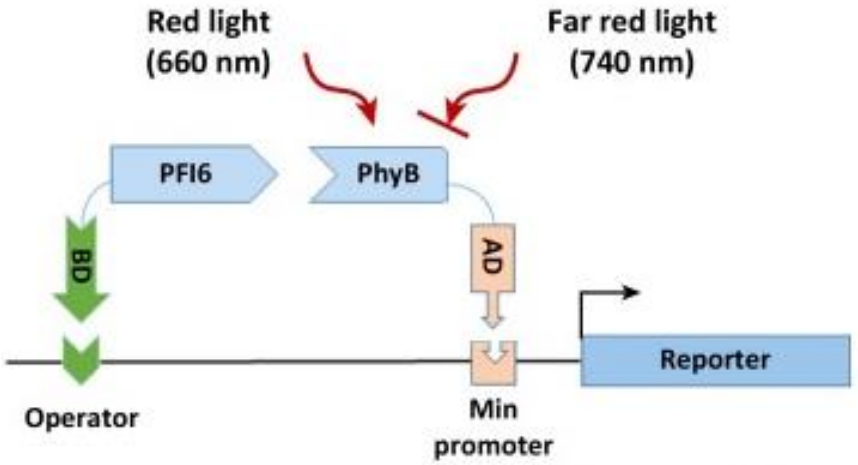


Devices

Auxin inducible degron



Red light regulated expression



Synthetic logic gates and cellular computation

A logic gate **in electronics**, is a physical device which is implemented with a Boolean function based on input and output signals (0 and 1). It executes a logical function on **one or more inputs** that produce a **single output.** Logic gates are used for storing the data that can be constructed by connecting several gates in a Flip-flops circuit which is a central building block of digital electronics systems in computers and communications.

In **biological systems**, logic gates are synthetic gene circuits programmed to permit the expression of an output protein only when a strictly defined signature of input signals is matched. Genetic elements interact with regulatory proteins to switch a **gene ON or OFF** while RNA or protein concentration can serve as input or output. One of the key approaches of synthetic biology is to **reprogram** the decision-making gene networks in order to implement them as **logic gates** in living systems.

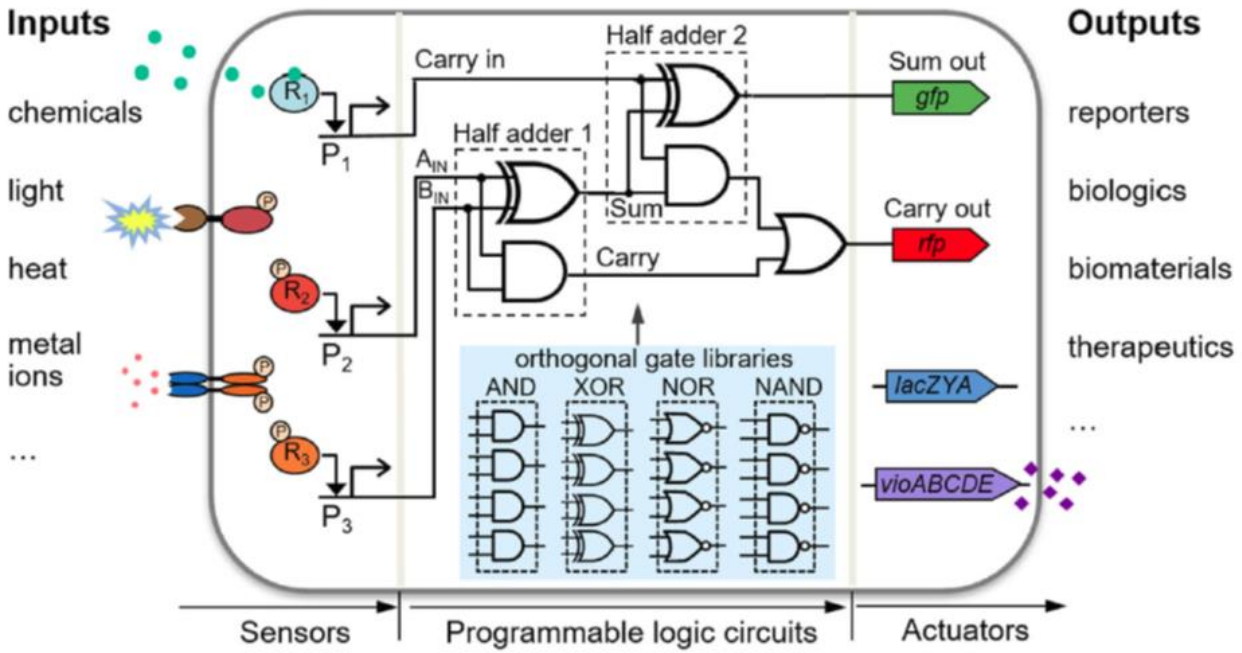


Fig. 1 Programmable cellular computation with scalable signal processing capacity. To achieve large-scale control of cellular behaviour, an expanded library of versatile orthogonal genetic regulatory blocks and associated wiring principles are needed. For example, a genetic 1-bit full adder program adds binary numbers, it has 3 inputs and 2 outputs, and can be constructed from 5 modular logic gates that are wired in 3 layers and selected from well-characterized orthogonal gate libraries. The genetic circuits can be coupled to modular input genetic sensors and output actuators to achieve complex decision making for a variety of human desired applications

Xiang et al., Natural Comp 2018

Logic gates: the AND gate

Input A	Input B	Output AB
0	0	0
1	0	0
0	1	0
1	1	1

The **AND** gate gives a high output (1) only if both the inputs are high.

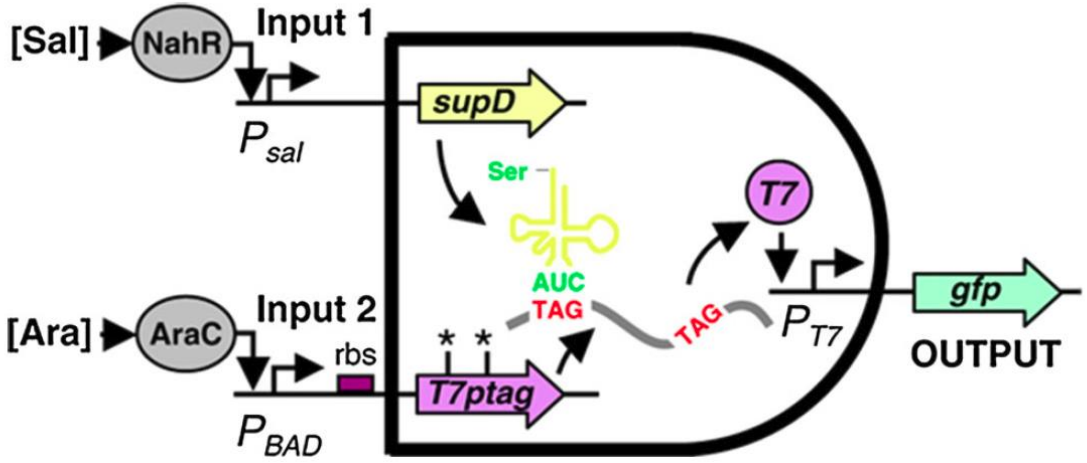
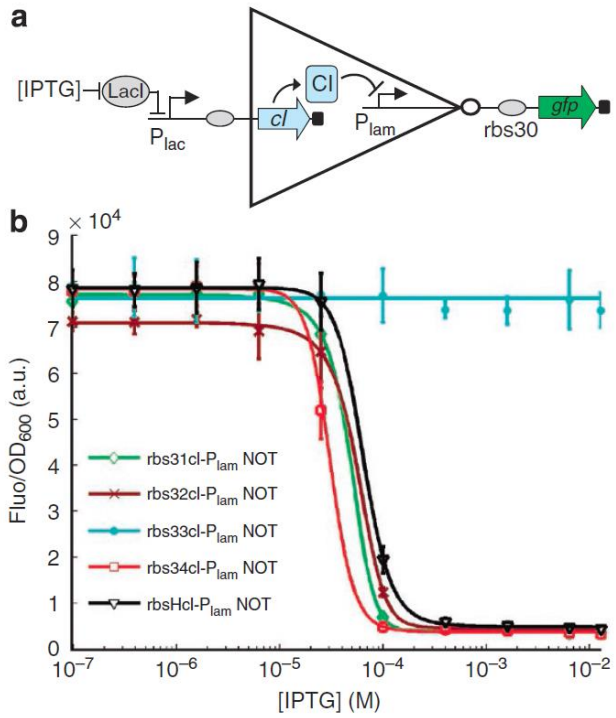


Fig. 2 Design of AND gate based on two promoters. The first promoter was linked with transcription of amber suppressor tRNA supD and second T7 RNA polymerase. Polymerase was modified to contain two amber stop codons and translated as serine when supD was transcribed and T7 pol was expressed when both SupD and T7ptag mRNA are present. Figure reproduced with permission from Anderson et al. (2007)

Singh, Syst Synth Biol 2014

The NOT gate module



- A **NOT** gate takes a single input and inverts it, so 0 becomes 1; 1 becomes 0.
- Even these simple gates can perform signal-processing functions, for example, converting a dark sensor into a light sensor.

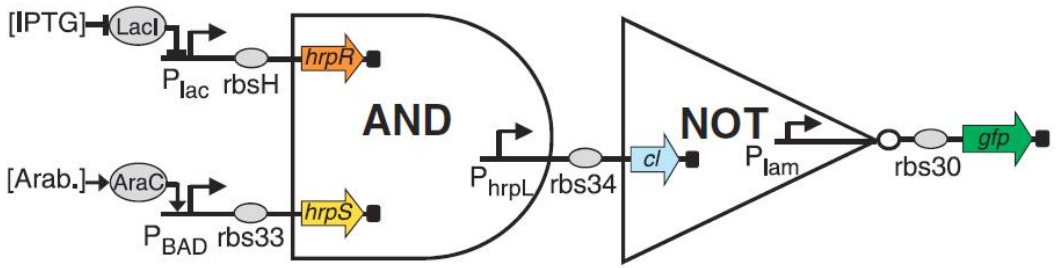
The modular NOT gate was designed on the basis of the cI/Plam repressor module consisting of lambda gene cI and its regulatory PR promoter.

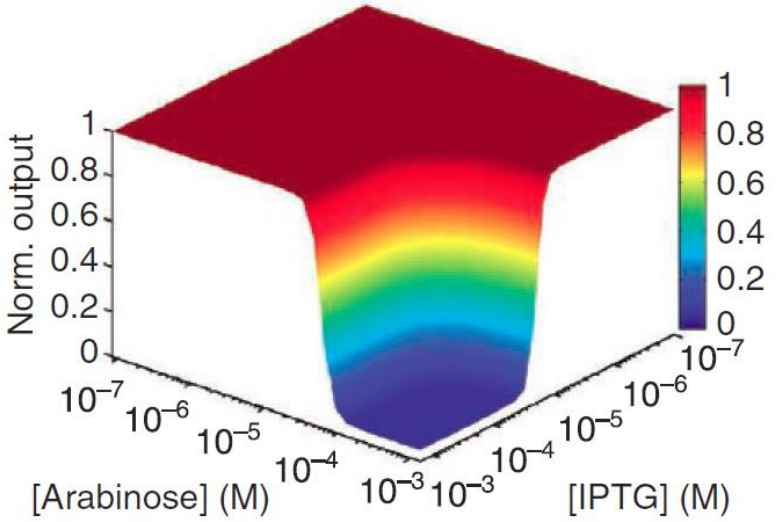
Wang et al., Nat Comm 2011

The NAND gate

Input A	Input B	Output AB
0	0	1
1	0	1
0	1	1
1	1	0

The output of **NAND** gate is high if any of inputs are low.





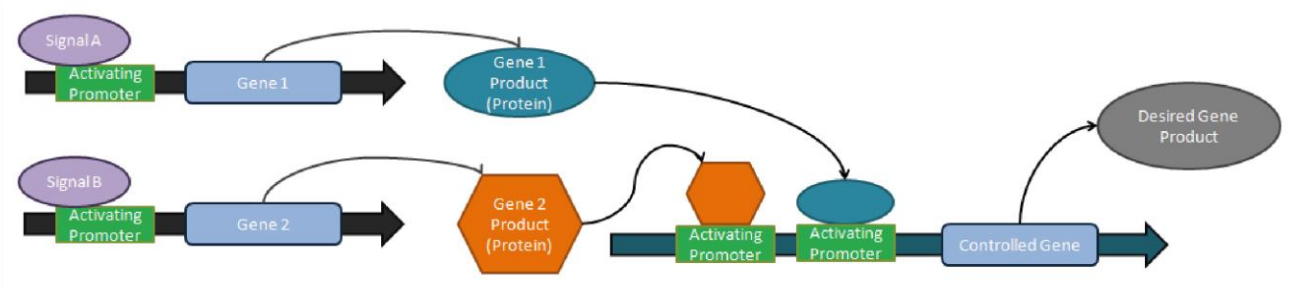
This NAND gate results from the combination of AND and NOT gates. The AND gate is derived from hetero-regulation module from *Pseudomonas syringae*. The device comprises two co-activating genes hrpR and hrpS controlled by separate promoter inputs, and a σ54-dependent hrpL promoter driving the output. The hrpL promoter is activated only when both genes are expressed.

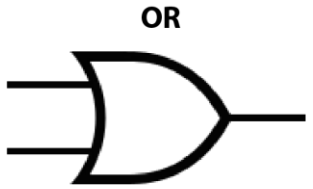
Wang et al., Nat Comm 2011

The OR gate

Input A	Input B	Output A + B
0	0	0
1	0	1
0	1	1
1	1	1

An **OR** gate outputs 1 as long as either (or both) of the inputs is 1.





The NOR gate

Input A	Input B	Output A + B
0	0	1
1	0	0
0	1	0
1	1	0

A **NOR** gate is equivalent to OR gate followed by a NOT gate. The outputs of NOR gates are low, if any of the inputs are high.

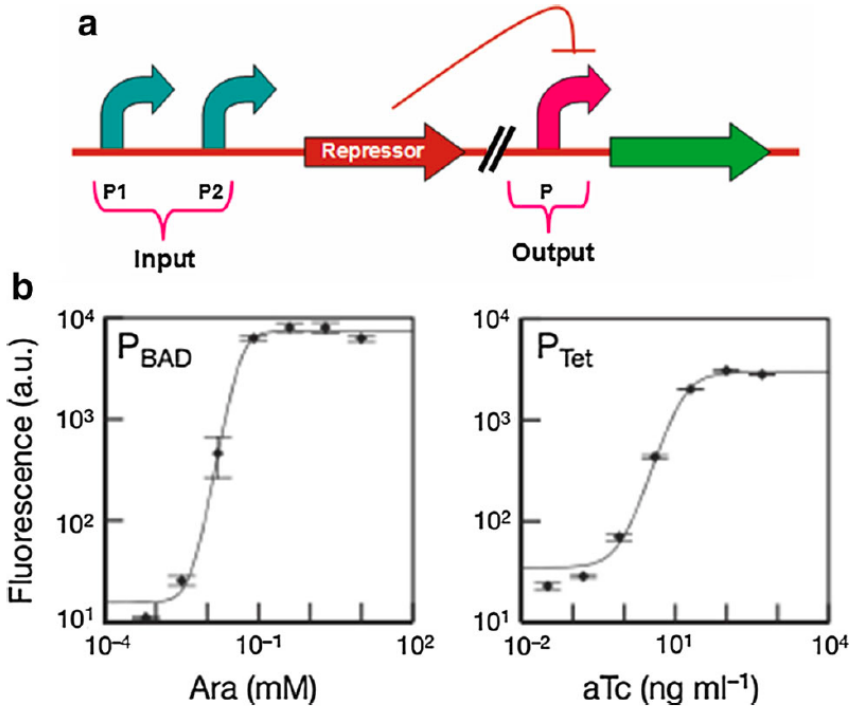
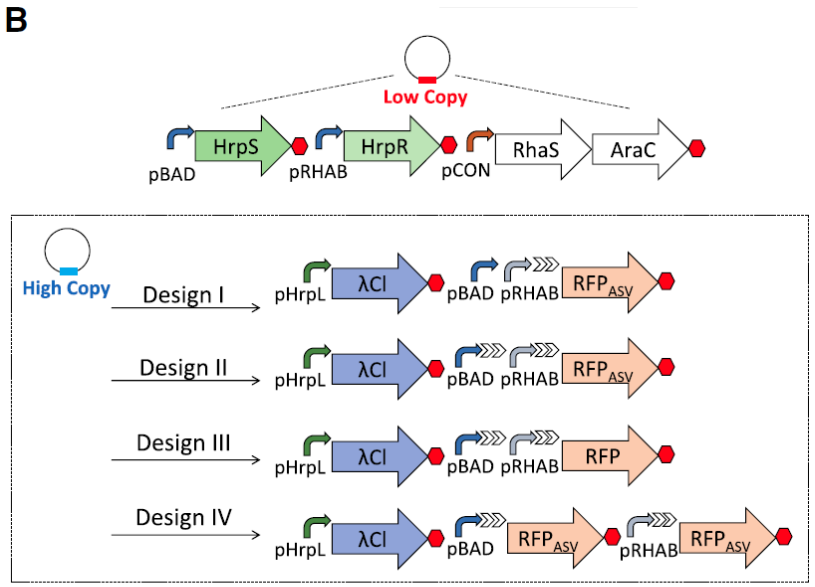


Fig. 4 NOR gate was engineered based on two tandem promoters (P1 and P2) express repressor and turn off downstream promoter (Pout) (**a**). **b** Different inducer concentrations were used for tandem promoter and NOR gate characterizations. Figure reproduced with permission from Nature (Tamsir et al. 2011) © 2011 Macmillan Publishers Ltd

The XOR gate

An **exclusive-OR (XOR)** gate gives a high output only if either input is present.

Input A	Input B	Output A + B
0	0	0
1	0	1
0	1	1
1	1	0



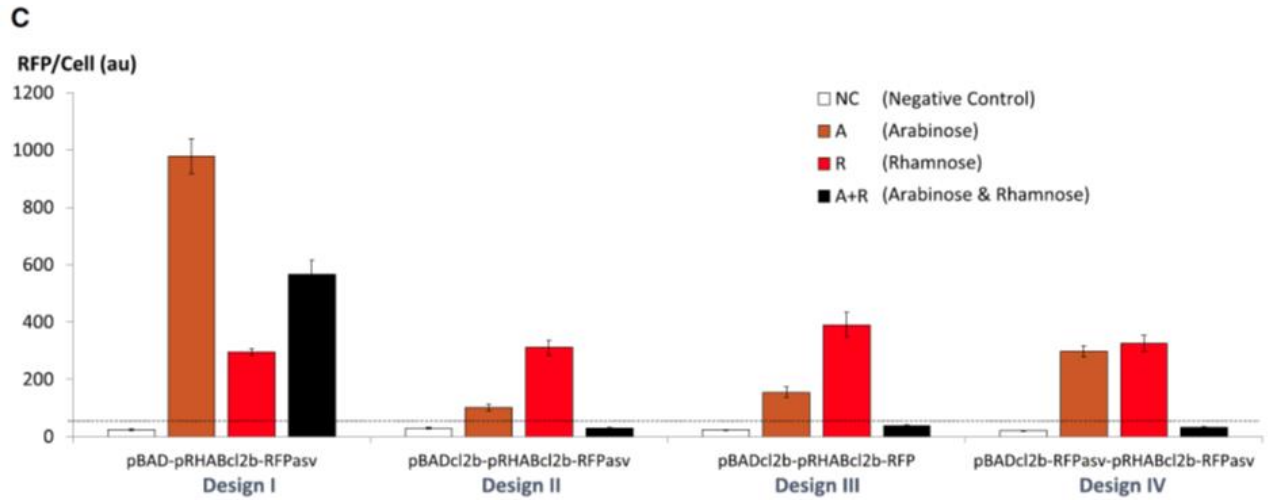
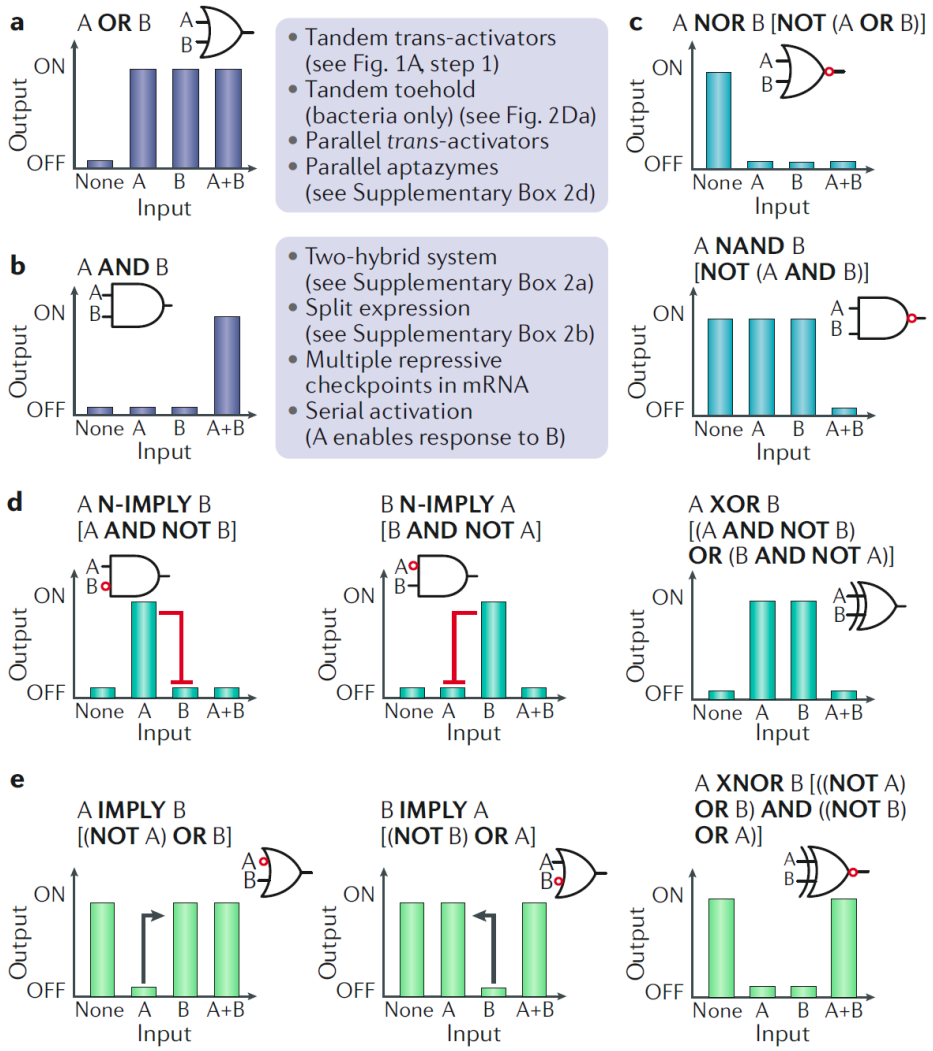
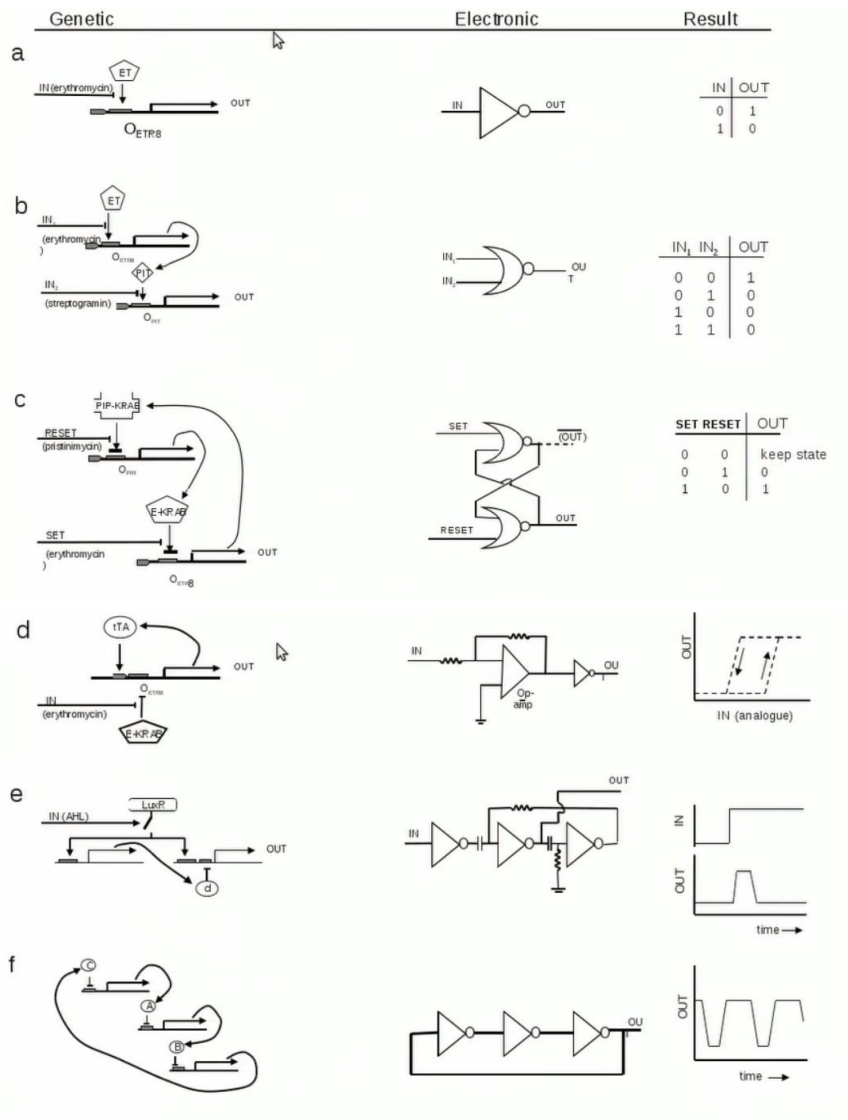


Fig. 5 Design and characterisation of biological XOR gates. **a** The logic output of XOR gate. **b** Genetic blueprint of four biological XOR gate designs. The XOR gate comprises serially layered AND, NOT and OR gates. HrpRS transcription factors are carried in a low copy plasmid, while pHrpL-λCI and distinct modules of OR gates with lambda repressor binding sites expressing RFP reporter are carried in high copy plasmids. Design I comprises tandem promoters with repressor binding sites downstream of pRHAB promoter and an RFP reporter engineered with the ASV protein degradation tag. Designs II and III comprise tandem promoters with repressor binding sites downstream of each promoter and RFP with and without the ASV degradation tag, respectively. Design IV is modified from design II with RFP expressed in two disparate transcripts. **c** Digital performance of various designs of biological XOR gates at steady state. **d** The steady state profile of XOR gate IV for various concentrations of arabinose (input A) and rhamnose (input B). Error bars represent the standard deviation of four independent experiments

Wong et al., BMC Biology 2015

Assembly of logic gates allows building circuits with complex behaviour





Xie and Fussenegger, NatRevMCB 2018

A multilayer AND gate

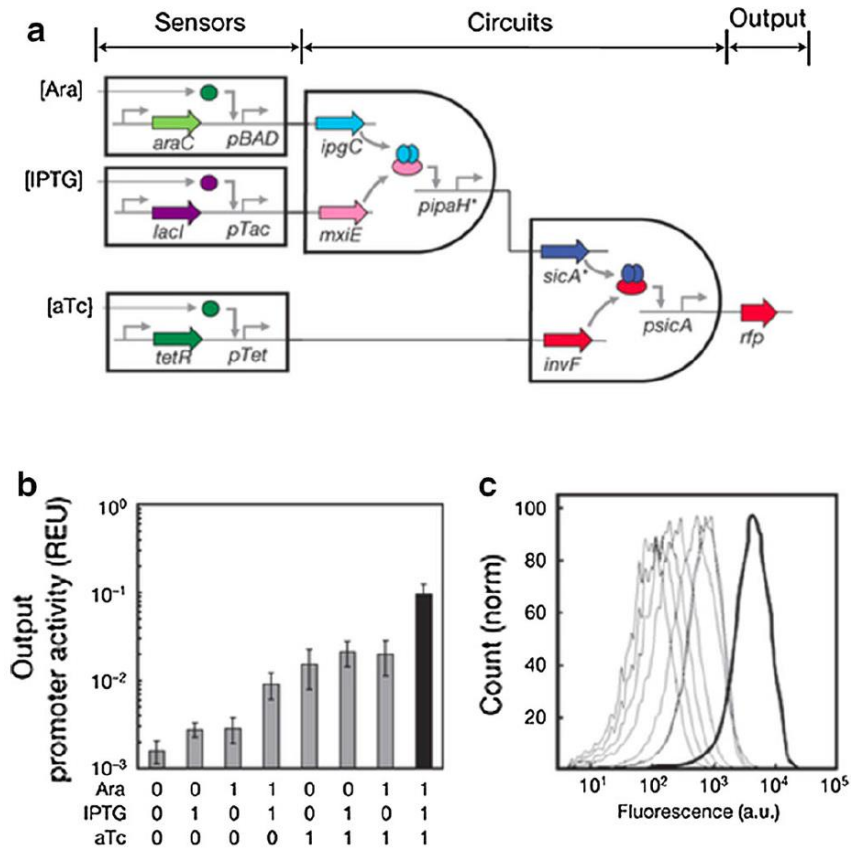
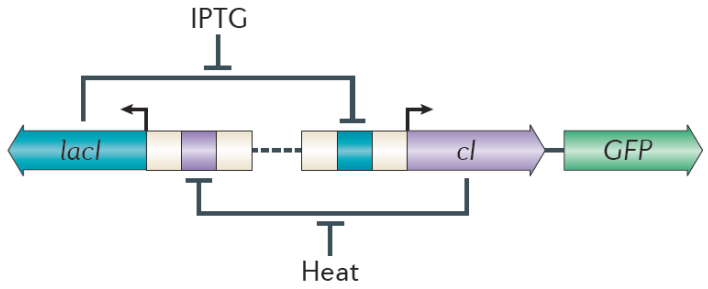


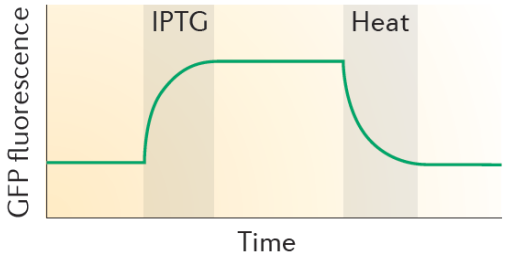
Fig. 3 Design of layering AND gate. **a** The 3-input AND gate. It consists of 3 sensors which is integrated circuit and a reporter gene. **b** The fluorescence was measured from cells containing the 3-input AND gate. The three inducers were used for on (1) input were Ara (5 mM), IPTG (0.1 mM) and aTc (10 ng/ml). **c** Cytometry data used for all sets of input states. **d** 4-input AND gate and **e** output

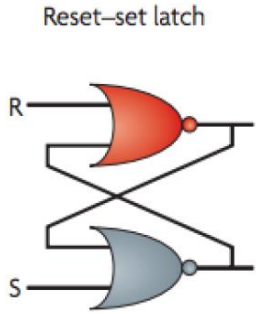
Singh, Syst Synth Biol 2014

Toggle switch and repressilator

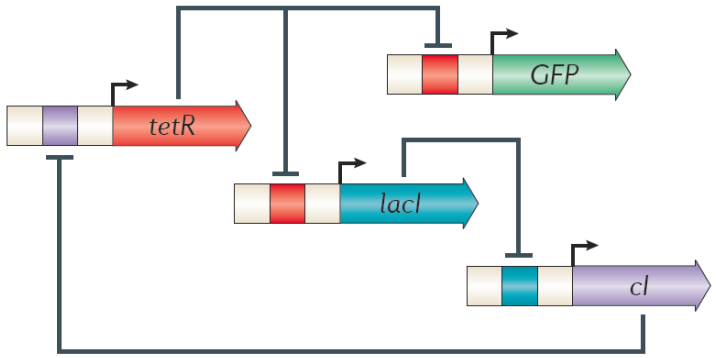
Toggle switch

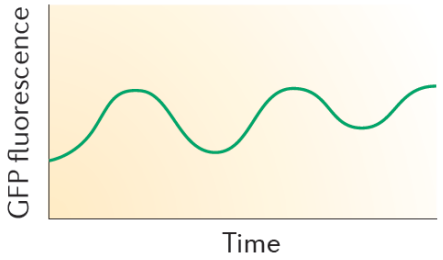


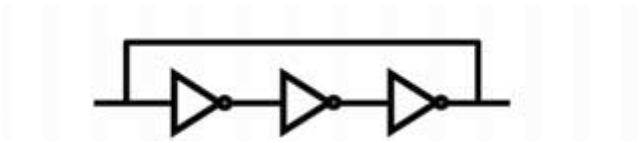




Repressilator

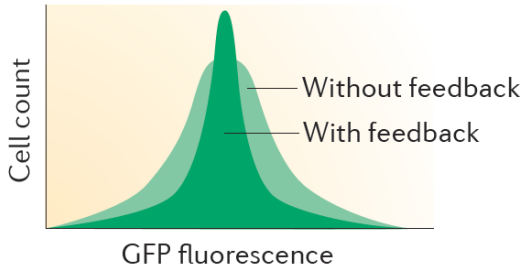


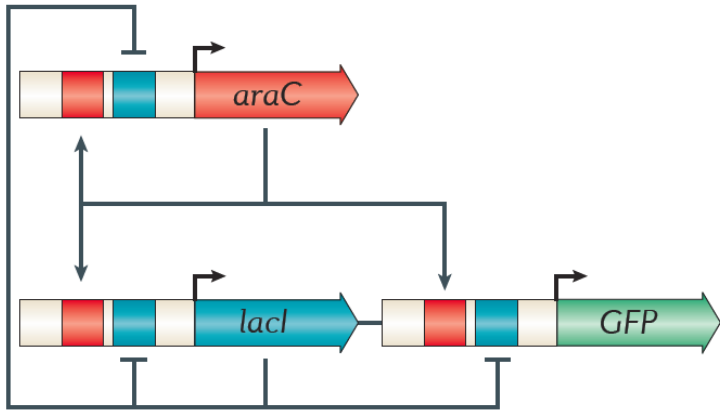


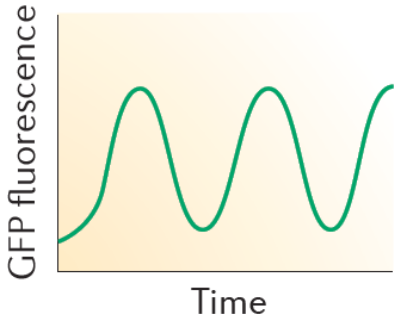


Relaxation oscillator uses autoregulatory feedback

Negative feedback loop rescues the noice and results in a narrow expression distribution







Registry of parts for synthetic biology

The basic concept in engineering is using a combination of standart parts to produce standard devices, which are then combined to produce standard systems.

MIT registry for the International Genetically Engineered Machine (iGEM)

http://parts.igem.org/Main_Page

- Standart parts must be thoroughly characterized and their performance well described
- Data on standartized parts are organized in registries of parts for synthetic biology
- iGEM Registry provides a resource of available biological parts that have been user-tested and characterized for users developing synthetic biology projects.
- iGEM Registry is an open community that runs on the "Get & Give & Share" philosophy. Users get parts, samples, data, and tools to work on their synthetic biology projects. They'll give back to the Registry the new parts they've made, as well as data and experience on new and existing parts.
- iGEM Registry contains about 20,000 parts
- The parts on the iGEM Registry adhere to the BioBrick standard allowing them to be assembled together creating new longer and more complex parts, while still maintaining the structural elements of the standard. This allows the engineer to focus on design instead of assembly.
- BioBrick Assembly Standart 10 is based on restriction cloning

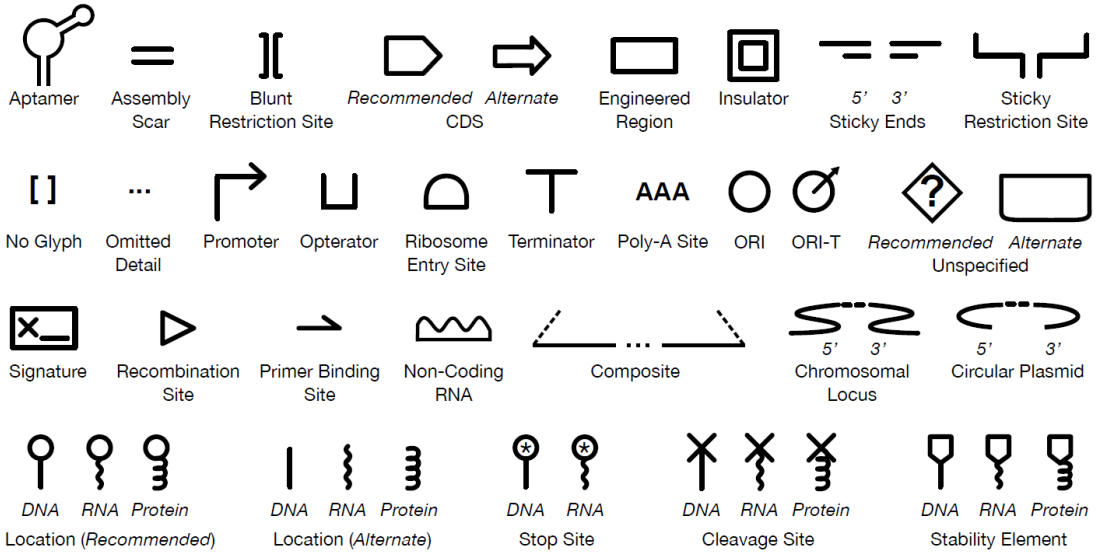
```
                Prefix                                          Suffix
5' - GAATTC GCGGCCGC T TCTAGA G ...part... T ACTAGT A GCGGCCG CTGCAG - 3'
     EcoRI    NotI      XbaI                   SpeI     NotI   PstI
```

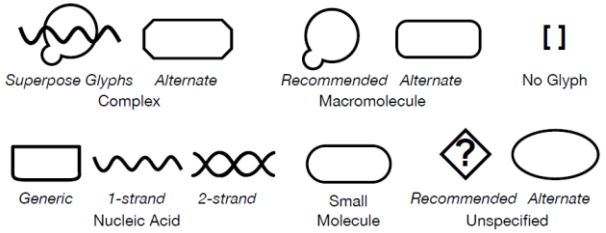
- Other registries: SYNBis Database http://synbis.bg.ic.ac.uk/synbis2/Welcome_Page.html

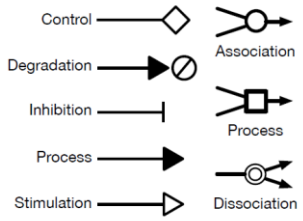
Standartization: The Synthetic Biology Open Language (SBOL)

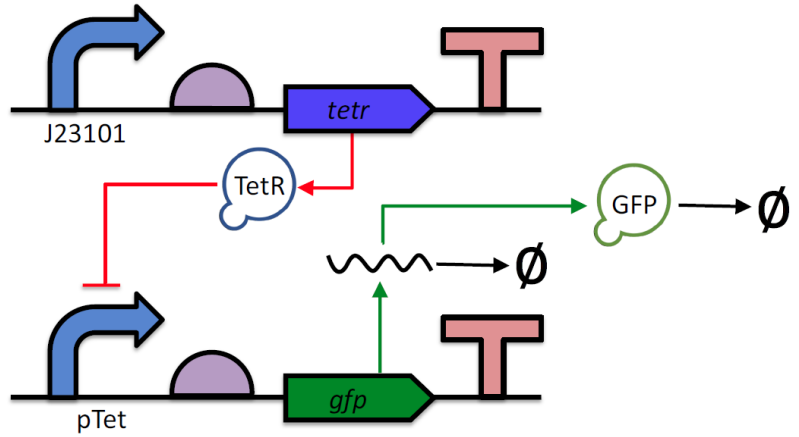
https://sbolstandard.org/

- SBOL is an open standard for the representation of in silico biological designs.
- SBOL also provides schematic glyphs to graphically depict genetic designs called SBOL Visual.



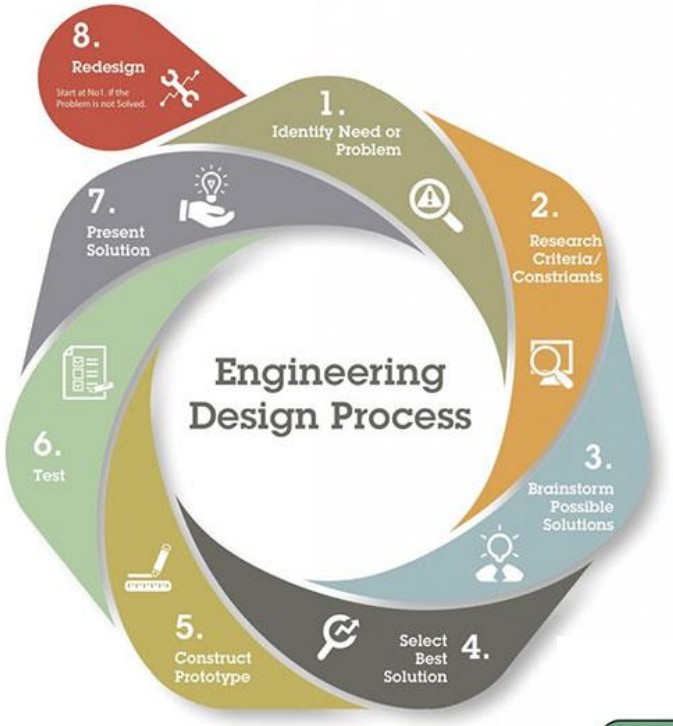






The top functional unit produces the TetR protein constitutively, under control of promoter J23101. TetR represses the pTet promoter, which is regulating production of GFP. The diagram of GFP production explicitly includes the intermediate mRNA and the degradation of both the mRNA and protein products.

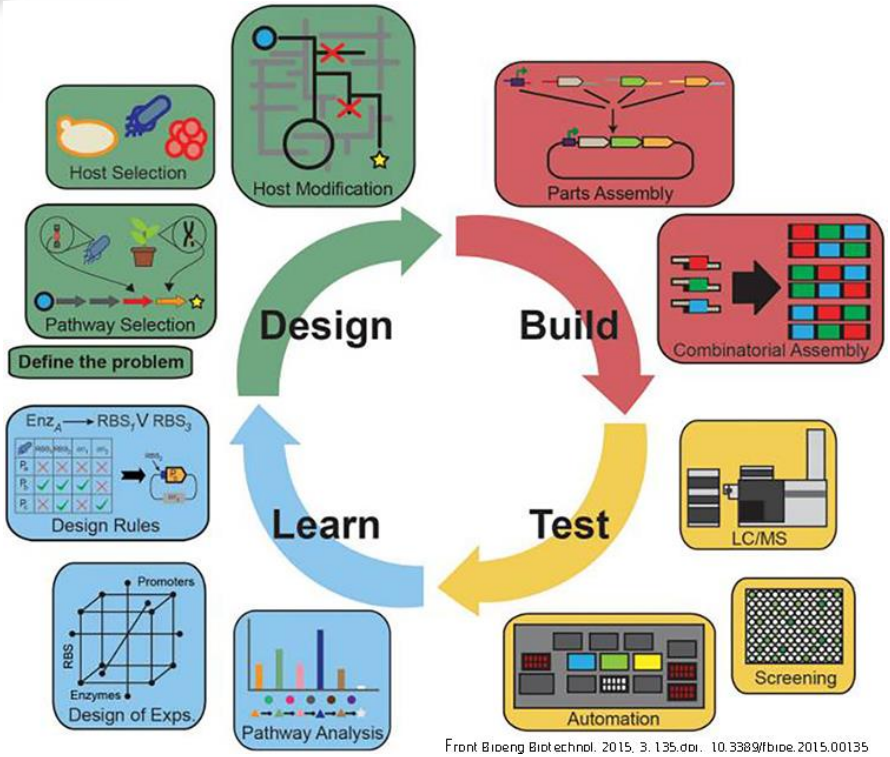
The synthetic biology design cycle



Specification: formal definition of the desired function and design of a target genetic system.

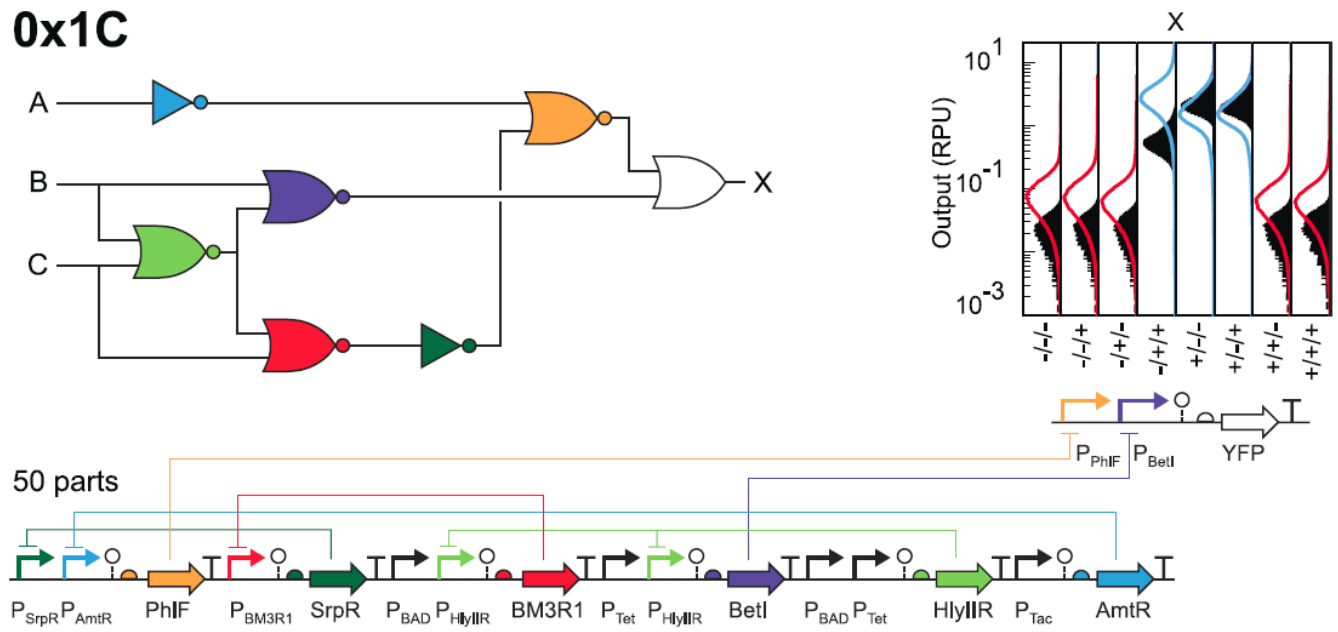
Design: the set of decisions needed to determine the constructs and hosts, to be used and/or modifications to the host to be made. Involves also creating a plan for composing the DNA constructs from their elements.

Build: Implementing DNA assembly plan and construction of the biological system.



Test: design and implementaiton of experiments for characterizing engineered systems and accompanying analysis and data interpretation.

Learn: include approaches to allow for revision of designs based on experimental outcomes.


0x1C
A
B
C
X
X
Output (RPU)
10^{1}
10^{-1}
10^{-3}
-/-/-
-/-/+
-/+/-
-/+/+
+/-/-
+/-/+
+/+/-
+/+/+
P_{PhlF}
P_{BetI}
YFP
50 parts
P_{SrpR} P_{AmtR}
PhlF
P_{BM3R1}
SrpR
P_{BAD} P_{HlyIIR}
BM3R1
P_{Tet}
P_{HlyIIR}
BetI
P_{BAD} P_{Tet}
HlyIIR
P_{Tac}
AmtR

Biological Computer Aided Design (BioCAD)

- BioCAD assists the *de novo* design and selection of existing genetic components to achieve a desired biological activity, as part of an integrated design-build-test cycle.
- BioCAD tools facilitate the design of larger systems from smaller genetic parts by providing users with visual, textual or programming-language-like interfaces, or automatically generating designs from intended function.
- As the field moves towards real-world applications, tools that can adequately predict functionality from design will be indispensable.
- Predictability of part behavior, particularly in different contexts, is still a major issue in synthetic biology design.

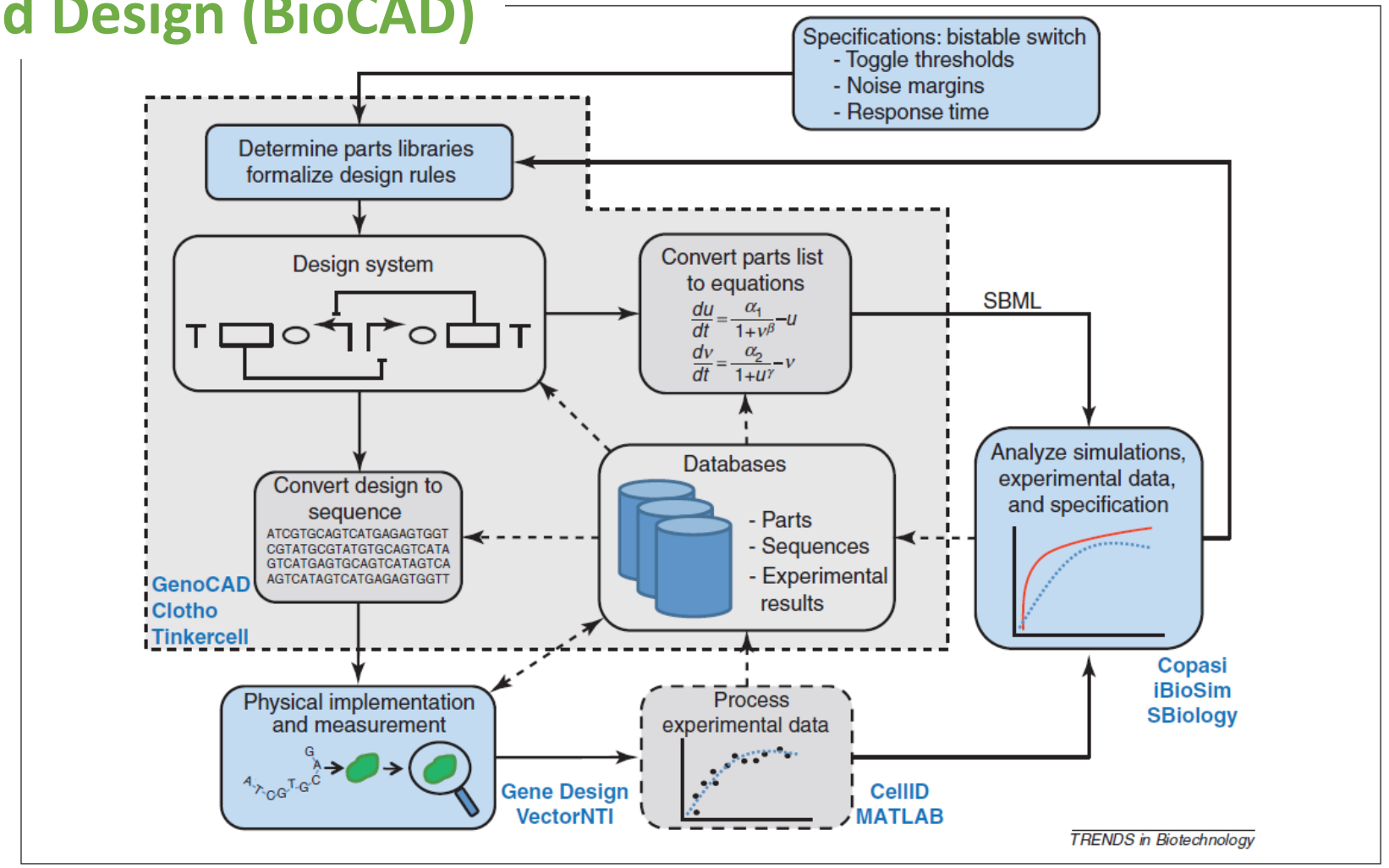


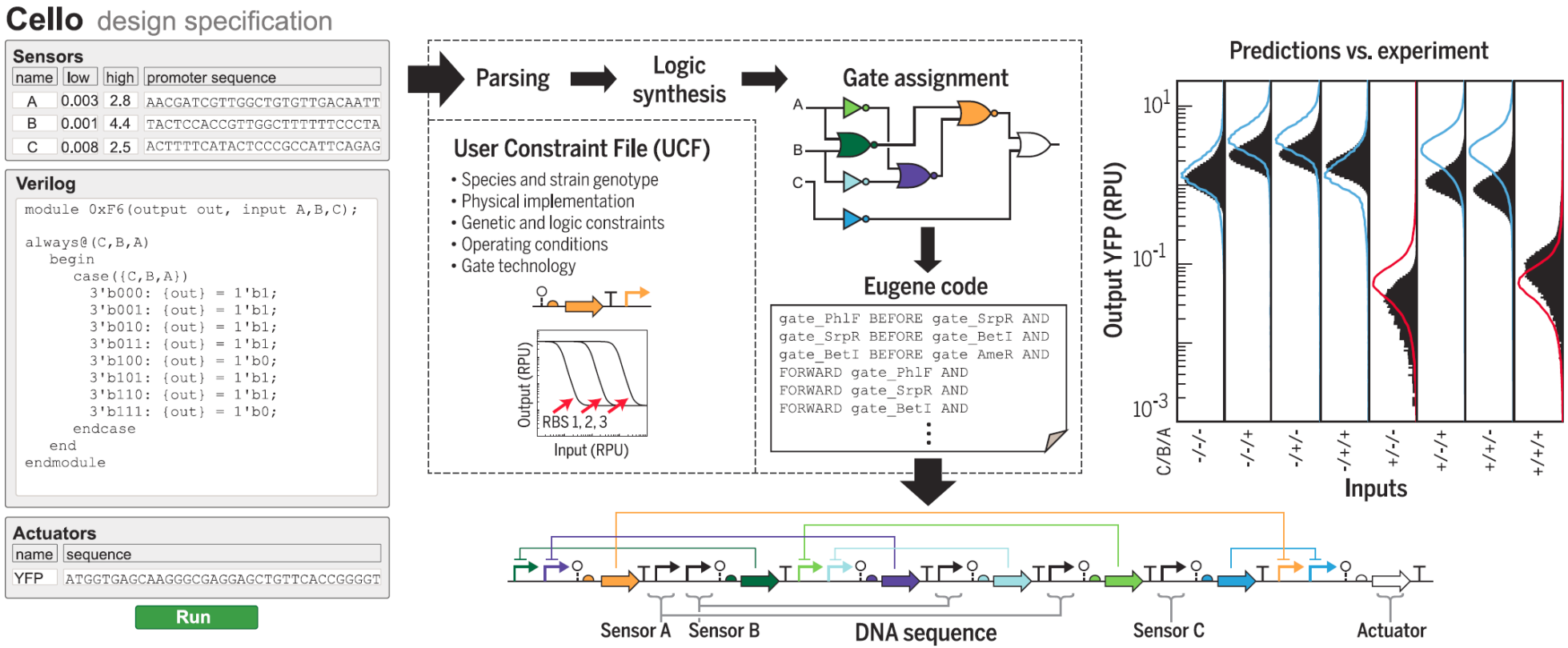
Figure 1. GDA design flow. Synthetic biology projects typically rely on iterative workflows composed of different tasks. Emerging GDA tool chains rely on numerous software applications that support different phases of the project workflow. The development of a genetic switch [72] will first involve expression of the design objective as a list of quantitative requirements: input toggle thresholds, noise margins, switching response time, etc. Once the objective is specified, it is possible to develop a list of genetic parts useable for the project. The choice of biological parts will involve factors such as use of the parts in prior projects, quality of the data characterizing the parts function, or intellectual property considerations. Formalization of design rules often takes place in parallel with parts library development. Design rules may cover issues such as whether it is acceptable to have polycistronic expression cassettes or if the design should be split between different plasmids. Only after parts have been selected and a strategy has been agreed on is it possible to start designing constructs. In the fabrication phase, the construct is assembled, usually by combining *de novo* gene synthesis and cloning of existing DNA sequences. Operators use molecular biology software suites to facilitate assembly or to order the sequence from a gene synthesis company. Experimentalists insert the synthetic DNA molecule into the host of choice and collect phenotypic data. Experimental data are then processed, for example by reducing microscopy images to time series of quantitative data. Performance is evaluated by considering simulations, experimental data and the original specifications. At nearly every stage, software interacts with databases to reuse past work or to store current work for future use. The shaded area delimited by dashes denotes stages facilitated by synthetic biology CAD software, whereas other stages are handled by more general purpose software. Text in blue indicates examples of software that provide assistance in each stage.

Lux et al., TiBiotech 2012

SYNTHETIC BIOLOGY

Genetic circuit design automation

Alec A. K. Nielsen,[1] Bryan S. Der,[1,2] Jonghyeon Shin,[1] Prashant Vaidyanathan,[2]
Vanya Paralanov,[3] Elizabeth A. Strychalski,[3] David Ross,[3]
Douglas Densmore,[2] Christopher A. Voigt[1]*



Genetic programming using Cello. A user specifies the desired circuit function in Verilog code, and this is transformed into a DNA sequence. An example circuit is shown (0xF6); red and blue curves are predicted output states for populations of cells, and solid black distributions are experimental flow cytometry data. The outputs are shown for all combinations of sensor states; plus and minus signs indicate the presence or absence of input signal. RBS, ribosome binding site; RPU, relative promoter unit; YFP, yellow fluorescent protein.

Nielsen et al., Science 2016

Recommended reading:

Wang et al. (2011) Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. Nat. Comm 1:508

Xiang et al. (2018) Scaling up genetic circuit design for cellular computing: advances and prospects. Natural Comp. 4:833