International Genetically Engineered Machines Competition

Bio Building Basics:
A Conceptual Instruction Manual for Synthetic Biology

Prepared by: Noah Helman, Wendell Lim, Sergio Peisajovich, David Pincus, and Nili Sommovilla
University of California San Francisco, May 2007

SYNTHETIC BIOLOGY

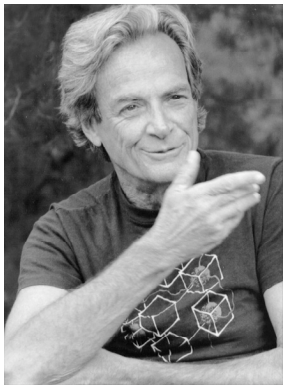
For iGEM, by Sergio Peisajovich (Lim Lab, May 2007)

What is synthetic biology?

Synthetic Biology is an emerging field of biology that aims at designing and building novel biological systems. Whereas "genetic engineering" has been around for many decades already, synthetic biology extends its spirit to focus on *whole systems* of genes and gene products, rather than on *individual* genes. Thus, synthetic biology aims, for instance, at adding or modifying biological functions to existing organisms or, in the future, creating novel organisms with tailored properties.

Why do we need synthetic biology?

Different scientific disciplines are interested in (and will benefit from) synthetic biology for a number of different reasons:



(1) For biologists: **it provides a direct and compelling method** for testing our current understanding of natural biological systems.
Citing the famous physicist Richard Feynman (pictured at right): "What I cannot create I do not understand." Disagreements between expected and observed behavior of the engineered biological system can serve to highlight areas of research that are worth exploring.
(2) For chemists: **biology is an extension of chemistry**, and as such, synthetic biology is simply the next logical step in synthetic chemistry (by extending the ability to create novel molecules and molecular systems, we allow continuing the development of useful diagnostic assays, medicines, or novel materials).
(3) For engineers: **biology is a technology**, but to fully develop as such, it requires a profound *conceptual* change in the way biology is done. Instead of the classical biotechnological approach, in which individual solutions are applied to individual problems, novel technologies and strategies need to be developed so that individual problems could benefit from standardized solutions (previously applied to other related or unrelated problems).

Is synthetic biology achievable?

Although clearly very complex, there are two aspects of biology that are inherently simple and likely to have arisen from evolution (the cause of biological complexity). They are:

(1) biology is hierarchical

and,

(2) biology re-uses a small set of simple parts to create complex behaviors.

It is true that multi-cellular organisms (animals, for instance) are extraordinarily complex. However, we can understand many aspects of animal biology if we divide complexity in hierarchies and then focus on understanding each individual level in a stepwise manner. In other words, understanding how, at the bottom of the hierarchy, DNA, proteins and metabolites function in individual cells, helps us to understand how different cell types are organized in tissues, how tissues make complex organs, and, finally, how many different organs are orchestrated in the top hierarchical level, the organism. Furthermore, different cell types and functions use re-current basic mechanisms of organization and communication; thus, understanding any single cell type provides already an incredibly rich amount of information to understand any other cell type (even from organisms as distant as yeasts and humans).

This hierarchical and re-current organization of biological systems is what allows: (1) biology to be understandable and (2) synthetic biology to be possible. We, as synthetic biologists, design new biological systems having in mind the top of hierarchy, but actually operating at the bottom of the hierarchy, by designing and testing novel gene and protein combinations, for how the smallest parts (genes and the proteins they encode) are wired is ultimately responsible for how all of the hierarchies above behave.

Discussion Question:

Could you think of any examples of "Modular Organization" in biology?

A possible hierarchy for synthetic biology

In analogy to electrical engineering, synthetic biology envisions a hierarchy for biology in which genes and proteins (individual parts) are assembled in increasingly complex devices and networks, in the same way in which individual transistors are assembled into integrated electronic circuits (see Fig. 1).

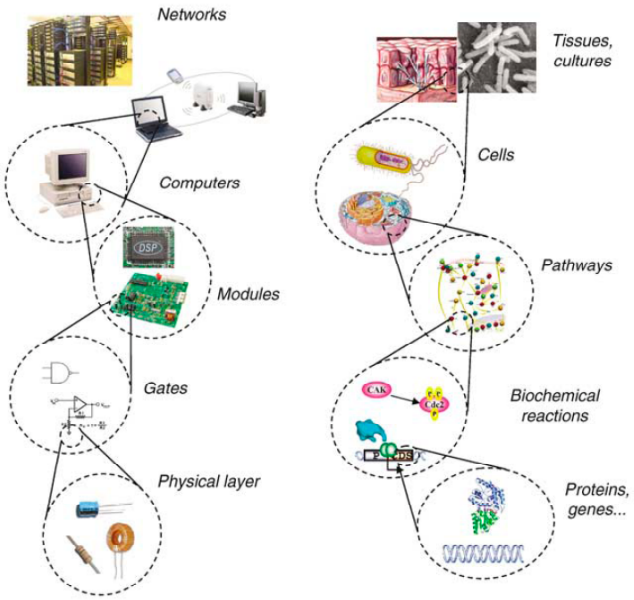


Fig.1

Biological Components

The simplest biological component that synthetic biologists use is the "part". **This is a defined sequence of DNA that encodes certain information and as a result performs a defined function.** This could be a gene regulatory function, such as a promoter of gene expression, a ribosome-binding site to direct protein translation, or an open reading frame coding for a particular protein. **Many different parts can be combined into a "device"** (see Fig. 2).

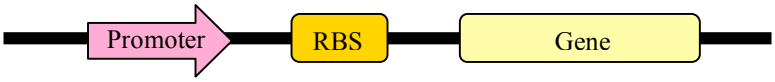


Fig.2: Three different parts (a promoter, a ribosome binding site and a gene) are assembled in a device.

Multiple devices can be hooked together thereby creating "systems" or "modules" capable of complex behaviors. In the example in Fig. 3, the device "A" responds to the presence of a chemical "a" by expressing the gene "A2", which in turn acts binding to a DNA promoter region (part "B1" in the device "B") thereby inducing transcription of the gene B2.

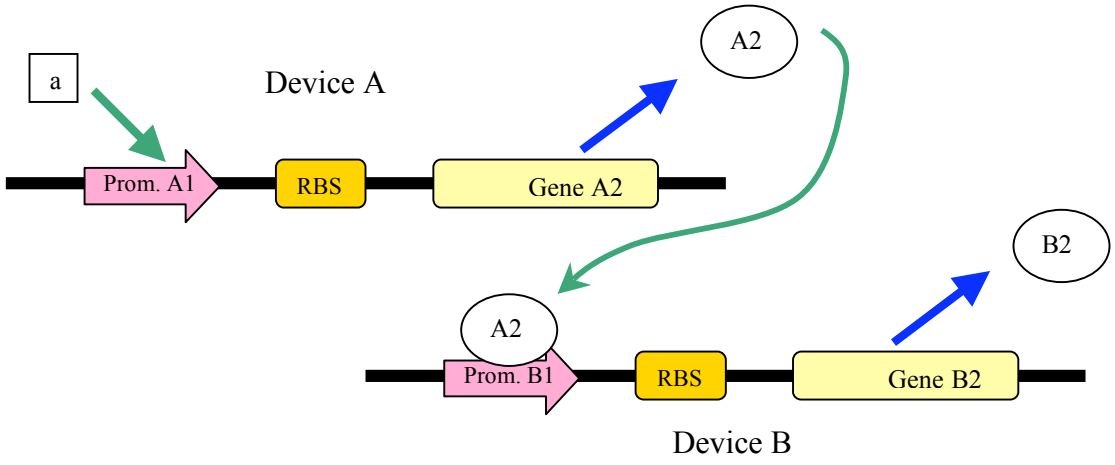


Fig. 3

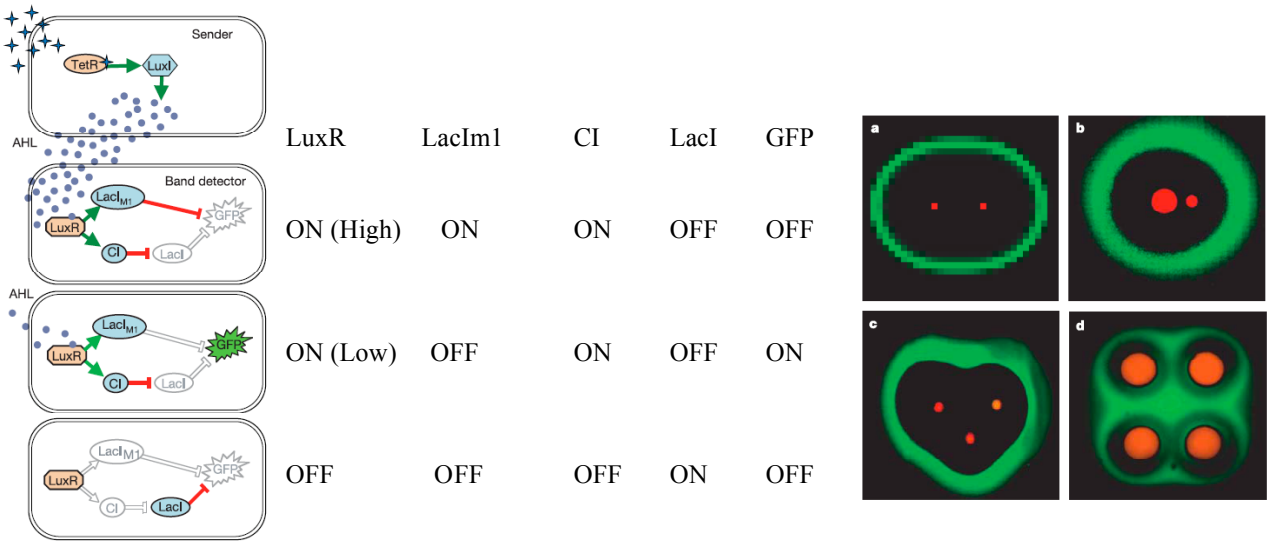
What parts are available for the synthetic biologist?

The list of possible functions encoded in individual parts is unlimited, simply reflecting the diversity of biology. Among those we can mention:

(i) Parts that regulate gene expression (promoters or enhancers, which are DNA sequences to which activators or repressors (proteins that regulate the activity of DNA polymerases in a positive or negative manner) bind
(ii) Different enzymes responsible for catalyzing a myriad of chemical reactions within organisms
(iii) Localization signals, which are generally short amino acid segments that when added to a given protein localize them to particular regions within the cell
(iv) Interaction modules, which are protein domains (or smaller motifs) that help proteins interact with specific partners
(v) Etc, etc, etc…

By combining many different systems (or modules), complex behaviors can be engineered. Moreover, an even higher level of complexity can be achieved by utilizing different cell types (each one with different embedded synthetic modules). In the example in Fig. 4 (Basu S. et al. (2005) Nature, 434: 1130-1134) the LuxI gene in the cell type "sender" synthesizes a small molecule called AHL that diffuses to the extracellular milieu. Once in the "receiver" cells, AHL binds to the transcriptional inducer LuxR, resulting in the expression of two different proteins: LacIm1 and CI, both repressors of gene expression. At high concentrations of AHL (that is in the proximity of "sender" cells), LacIm1 represses the expression of a green-fluorescent protein, while CI represses the expression of LacI (a more potent version of LacIm1).

Far away from sender cells, AHL concentrations are too low for LuxR to be active, thus LacIm1 and CI are poorly expressed. The absence of CI in turn allows the expression of LacI and again the expression of the green-fluorescent protein is repressed (remember both LacIm1 and LacI block its expression). On the contrary, at intermediate AHL concentrations (that is at intermediate distances from the "sender" cells) LuxR is only partially active. In this situation, the low expression levels of CI are still enough to block expression of LacI, whereas the low expression levels of LacIm1 fail to block expression of the green fluorescent protein (because CI is a much stronger transcriptional repressor than LacIm1). In this scenario, the green-fluorescent protein is expressed, conferring a typical green color to the cells (see Fig. 4 on the next page).



	LuxR	LacIm1	CI	LacI	GFP
ON (High)	ON	ON	OFF	OFF	
ON (Low)	OFF	ON	OFF	ON	
OFF	OFF	OFF	ON	OFF	

Fig. 4: The left shows modules operation for sender cells (top) and receiver cells at high, medium and low AHL concentrations (blue circles). The middle table indicates the ON/OFF state of the different devices at the different AHL concentrations. At the right, photographs of Petri dishes show different green patterns formed by placing sender cells (visible as red areas, due to the expression of a red-fluorescent protein not related to the modules described above) in different locations. Note that, as predicted by the modules design, cell that are too close or too far away from the senders fail to express green-fluorescent protein.

Synthetic biology as a truly engineering discipline

While the example showed above is a remarkable success in this emerging field, for synthetic biology to develop into a truly engineering discipline, we will need to be able to predict functions of even the simplest devices in engineered cells and to construct systems that perform complex tasks with **precision and reliability**. Our current difficulty to achieve these goals arises from several sources: (i) incomplete knowledge of biological systems; (ii) inherent functional overlap (that is: some parts perform many –some unknown- functions, some of which interfere or are detrimental to the goal in mind); (iii) inherent incompatibility between different parts (interactions between different parts within a device are not optimal); and (iv) different cellular (or extracellular) milieu affects parts behavior (parts or devices work well in a given cell type but not in a different one).

Overcoming these problems will likely require a gigantic effort by a large community of synthetic biologists (as well as biologists in other areas of research); however, applying some concepts that were fundamental to other, more established areas of engineering, could certainly be of great help. Some of these concepts are:

(i) **Standardization:** parts (and devices) should be designed and constructed in a way that they could be easily exchangeable between different devices (as well as between different laboratories). In this way, their use and study (in multiple conditions) would be optimized. Ideally, in the future it shall be possible to design and build parts the behavior of which will not be affected by changes in the cellular environment or by the presence of other parts (see Fig. 5).

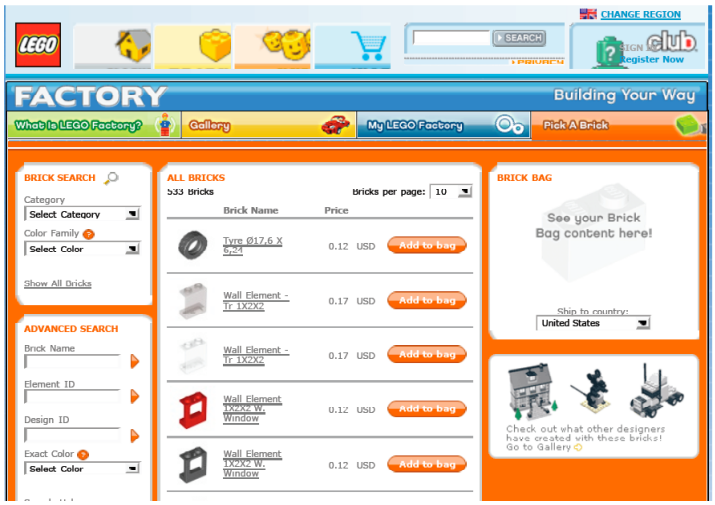


Fig.5: Biological parts should be modular and standardized, so that they can be easily combined and exchanged between different devices and/or laboratories. Ideally, one day we should be able to order and use biological parts as we now can order and use Lego parts.

(ii) **Abstraction:** information describing biological functions shall be organized across levels of complexity using hierarchies, in a way that individuals working at any one level do not need to know the details that are inherent to other levels, **yet they can exchange information across levels** in a useful way. For example, whereas some synthetic biologists will be interested in designing and building parts (i.e. a set of DNA binding proteins capable of acting as transcriptional repressors that are sensitive to varying concentrations of a given metabolite), others will prefer to focus on how parts are combined into devices, or how devices are linked to form systems (see Fig. 6).

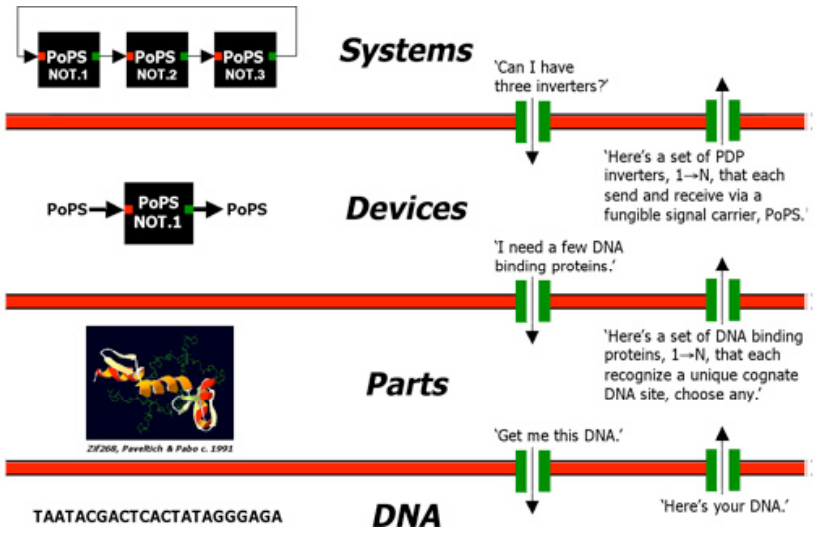


Fig.6: Abstraction hierarchies insulate different levels in a way that synthetic biologists working at any given level do not need to worry about other levels (i.e.: someone who designed a series of devices that require a set of DNA binding proteins, each one recognizing a unique DNA sequence, could simply ask someone else working at a lower level for the necessary parts).

Further reading:

Endy, D. (2005) Nature, 438: 449-453.

Elowitz, M. B. & Leibler, S. A. (2000) Nature, 403: 335-338.

Sprinzak D, Elowitz MB. (2005) Nature, 438: 443-448.

Guet CC, Elowitz MB, Hsing W, Leibler S. (2002) Science, 296:1466-1470.

Basu S, Gerchman Y, Collins CH, Arnold FH, Weiss R. (2005) Nature, 434:1130-1134.

Andrianantoandro, E. et al. (2006) Mol. Syst. Biol. Doi: 10.1038/msb4100073.

Park SH, Zarrinpar A, Lim WA. (2003) Science, 299:1061-1064.

Dueber JE, Yeh BJ, Chak K, Lim WA. (2003) Science, 301:1904-1908.

Dueber JE, Yeh BJ, Bhattacharyya RP, Lim WA. (2004) Curr. Opin. Struct. Biol. 14: 690-699.

ADAPTATION IN SIGNALING

SUMMARY. Biological signaling systems receive an input stimulus from the environment and generate a specific response. Such systems, however, are rarely simple on/off systems, but often show more sophisticated behavior. An important and common example is adaptation – many signaling systems turn on when presented with an input (or increase in input), but then turn back off when this input is persistent over a long period of time. Such adaptation allows the signaling system to respond to changes in input relative to background. Without adaptation, many sensory systems would only work within a small dynamic range, and any level of input above a certain threshold would saturate the response.

An example: Light Adaptation in Vision

Our own visual system contains photoreceptor cells within a photoreceptor organ (the eye) that turns ON in when stimulated by photons. Yet, the visual system is *not* a simple ON/OFF system that turns on when stimulated with a set number of photons. Instead, it shows adaptation, allowing the system to function in a wide range of light levels.

All of you have had the following experience: In dark room, like a movie theater, you can see very dim objects. But if you go outside, objects are hard to see – appearing washed out with low contrast (see Figure 1). Within minutes, however, your eyes adapt to this new level of ambient light and objects can be easily seen with high contrast.

Figure 1

before adaptation

after adaptation

This process, called **light adaptation**, involves several mechanisms that allow the visual system to reset its level of sensitivity—the amount of input (light) required to trigger response. Adaptation occurs at multiple levels. At the organ level, we see the constriction of the pupils, physically altering the amount of light that strikes the retina. At the cell level, we see calcium-mediated feedback that down-regulates photoreceptor sensitivity.

Definition of an Adaptive System:
A system that recognizes deviance of one of its states from a target set-point, and initiates counter-measures that return this state to this target level, even with the persistence of the perturbing factors ("perfect" adaptation returns exactly to the set-point).

Why is Visual Adaptation Important?
Adaptation allows vision to have a wide and dynamic range – it allows us to see differences over a wide range of ambient light. Otherwise, we would only be able to see within a very small range of light.

Other examples of Adaptation:

- Humans -- adaptation occurs to persistent stimuli, such as sound and touch

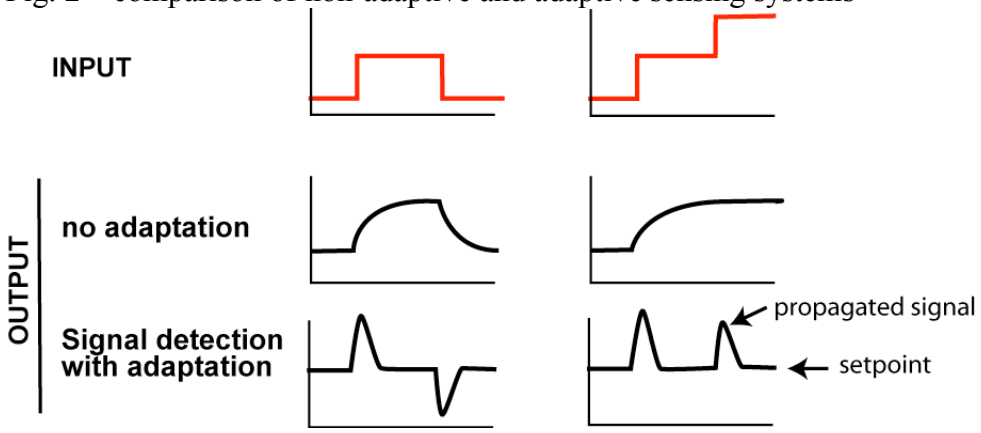
- Bacteria – chemotaxis. Here, bacteria respond to changes in the concentration of certain molecules in the surrounding environment by moving towards or away from such gradients. Adaptation allows bacteria to register whether they are doing better or worse (we will look at this in more detail shortly).

Discussion Question:

Can you think of other adaptive systems, either biological or man-made? Try to come up with a list of five examples.

A Quantitative Description of Adaptation
Figure 2 shows the quantitative behavior of a non-adaptive and adaptive sensory system:

Fig. 2 – comparison of non-adaptive and adaptive sensing systems



Discussion Question:

Why does the non-adaptive system fail to respond to the second increase in input on the right hand side?

Example: Bacterial Chemotaxis

Here we describe one of the best-characterized systems of adaptive sensing – **bacterial chemotaxis**.

Physiological behavior.

Bacteria have receptor proteins that can sense chemicals that are either nutrients or toxins. The information these receptors provide is used in a process known as chemotaxis ("chemical movement"). Bacteria can swim *up* a gradient of nutrients (chemicals referred to as "chemotactic attractants") and can swim *down* a gradient of toxins (chemicals referred to as "chemotactic repellents"). Chemotaxis thus allows the bacteria to **find sources of food** and **avoid harmful environments—**two behaviors crucial for survival.

One of the most remarkable aspects of bacterial chemotaxis is that the bacteria can find their way in very subtle gradients, and can perform this gradient sensing at a wide range of chemoattractant or repellent concentrations. How are bacteria able to do so? It should now come as no surprise that adaptation plays a critical role in this behavior.

Logic of the sensing system.

At first, there were two major hypotheses regarding how chemotaxis might work.

One was a **spatial model** – that the bacteria might detect a difference in attractant concentration at one end of the cell to the other, and then swim in the correct direction.

The second was a **temporal model** – that the bacteria might swim a step in a direction, then sample whether the attractant concentration is now higher (or lower, or the same) than before. If it is higher, it would continue swimming in that direction.

Which system is correct?

It turns out that bacteria use the second model, a temporal mechanism, whereby they constantly compare whether they are in a higher or lower chemoattractant concentration than before.

How does it work?

To understand this system, one has to first understand the motor that drives bacterial motion. Bacterial have several flagella distributed on the cell surface. The flagella have a long helical structure attached to a molecular "motor" complex that can rotate either clockwise (CW) or counter-clockwise (CCW) (it is always rotating one way or another). Because of the helical handedness of the flagella, when the motor turns CCW, the flagella bundle behaves cooperatively as a single propeller, allowing the bacterium to move rapidly in one direction – a process referred to as "swimming" (Fig. 3). In contrast, when the motor turns CW, the flagella splay out in random directions, leading to a random rotation of the bacterium that is called "tumbling". While swimming propels bacteria in one direction, tumbling allows bacteria to reorient into new directions. The bacteria are constantly switching between either swimming or tumbling. As a bacterium swims up a gradient of attractant, one still observes this switching. However, the frequency of swimming behavior is greater than tumbling, leading to longer swimming runs as the cell moves into areas of higher concentration. Thus bacteria find their way up the gradient through a process that is technically called a **biased random walk** (Fig. 4).

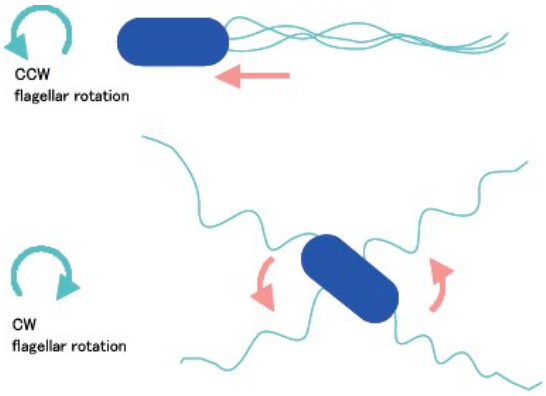


Fig. 3. Bacterial swimming (top) vs. tumbling (bot)

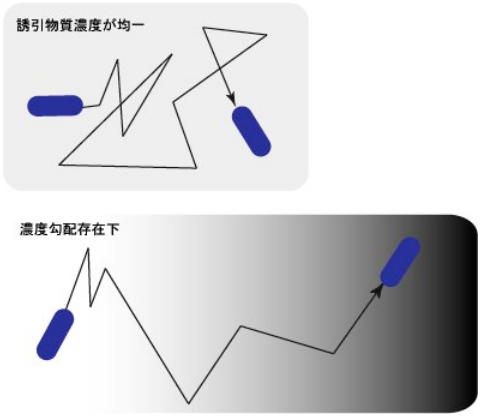


Fig. 4. Random walk vs. biased random walk up gradient

The basic signaling pathway: How chemostimuli modulate the direction of the flagellar motor.

When a chemostimulus binds to a receptor on the cell surface, it will activate a phosphorylation cascade (see Fig. 5 and Table 1). In the case of a chemorepellent, it will activate phosphorylation of the protein CheW, which in turn transfers its phosphate to the protein CheY (the opposite reactions are driven by activation of chemoattractant receptors). Phosphorylated CheY can bind to the flagella motor, switching its direction from CCW to CW (leading to more tumbling). Thus increasing chemorepellent leads to more tumbling (thereby allowing sampling of new directions), while increasing chemoattractant leads to more swimming (maintaining same direction). You might recognize, however, that the basic pathway, will **not** explain how the cell can tell if concentration is increasing of decreasing.

This behavior requires…adaptation!

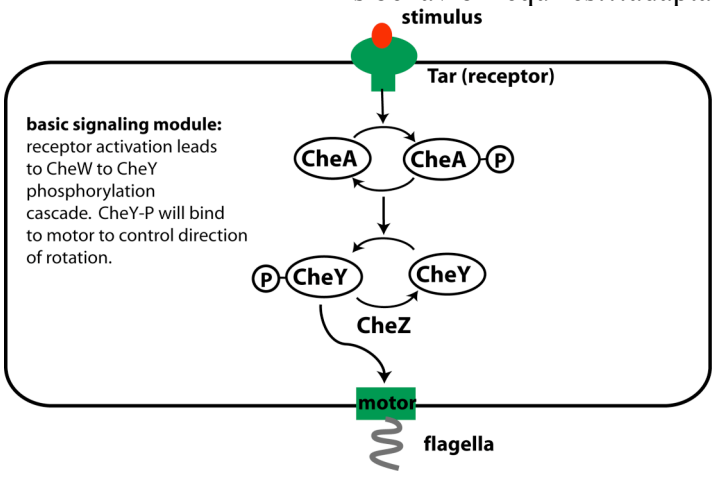


Fig. 5. Basic signaling pathway of bacterial chemotaxis.

Table 1. Chemotaxis Proteins

Protein	Function
MCP	Methyl accepting proteins (Tar, Trg, Tsr, Aer, Tap)
CheA	Autokinase; Phosphodonor for CheY/CheB
CheB	Methylesterase that removes methyl groups from the receptor
CheR	Methyltransferase that adds methyl groups to the receptor
CheW	Scaffolding protein required for stably coupling receptors and CheA
CheY	Cytosolic response regulator that carries the signal to flagellar motor
CheZ	Enzyme that facilitates dephosphorylation of CheY protein

The adaptation module.

When CheA is phosphorylated, it can activate CheY by transferring the phosphate to this protein. However, CheA can also transfer its phosphate group to the protein CheB. CheB is a methyltransferase that is activated upon phosphorylation. In the active state, it will transfer methyl groups to the receptor, a modification which reduces the sensitivity of the receptor. Thus, CheB forms a key part of a feedback adaptation module (Fig. 6). When the pathway is stimulated by chemorepellent (or attractant), not only is signal propagated to the flagella motor (which controls the final output – movement) but it is also fed back to the receptor. This leads to perfect adaptation, where after a persistent level of stimulation, the chemotactic signaling system returns to the basic 'off' state. This allows a further increase (or decrease) in signal to be detected (as shown in Fig. 2). What is important to recognize here is that the chemotactic receptors do not get saturated at a fixed concentration.

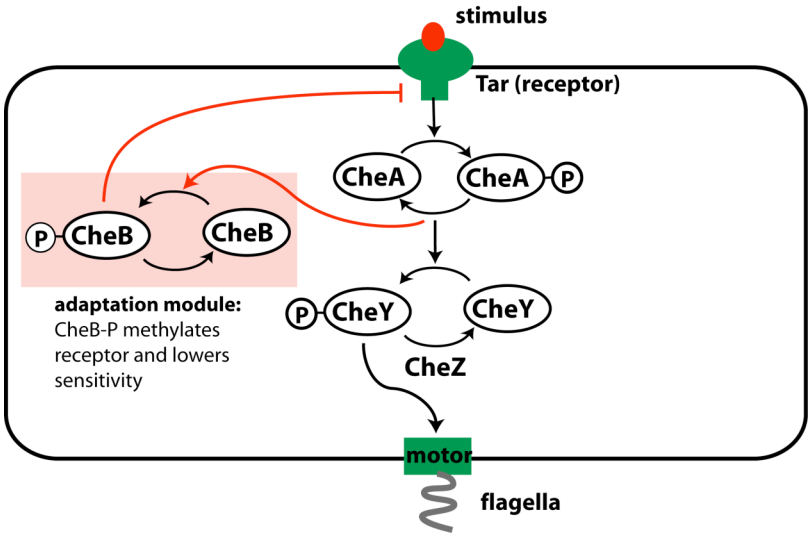


Figure 6. The feedback module that allows adaptation in the bacterial chemotaxis system.

PATHOGENS AS SYNTHETIC BIOLOGISTS

Bacterial proteins that can interface with host signaling pathways.

Many pathogenic organisms (Fig. 1) infect a host (like humans) and produce proteins that interface with the host signaling pathways. These proteins alter the behavior of the signaling pathway to the pathogen's advantage, essentially performing something similar to synthetic biology. For example, many pathogens produce proteins that can downregulate the immune system, facilitating the survival of the pathogen. Others produce proteins that can modulate cell movement and shape, allowing adhesion or spreading of the pathogen.

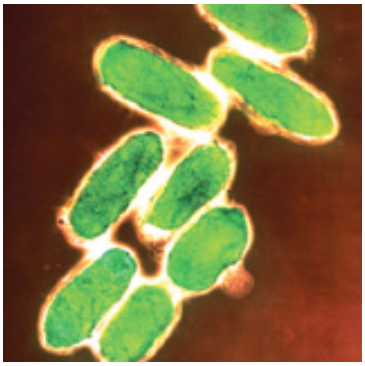


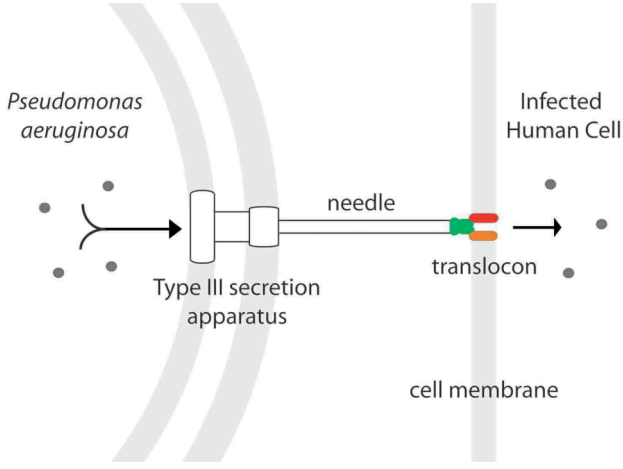
Fig. 1. *Yersinia Pestis*, the cause of black plague, is one of many bacteria that perform synthetic biology, systematically rewiring host regulatory pathways.

Although these pathogens are harmful, these individual proteins are not. In fact, they could be very useful. For example, proteins that can suppress immune response in a controlled way, could be useful for treating diseases involving an overreaction of the immune system, including autoimmune disorders or organ transplant rejection.

Bacteria use a molecular "syringe" to inject proteins into host cells.

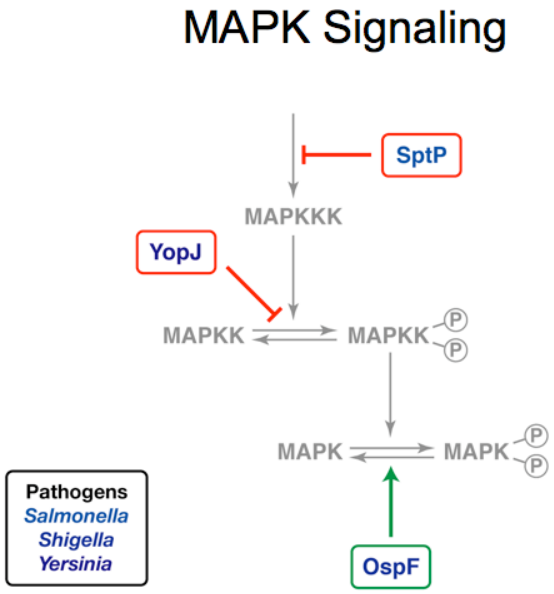
Many pathogenic bacteria have a molecular "syringe" structure, called a type III secretion system. This type III secretion system (Fig. 2) can be used to inject a cocktail of proteins into host cells.

Fig. 2. Cartoon of type III secretion apparatus. The grey dots represent proteins that the bacteria (right) injects into the host cell (left)



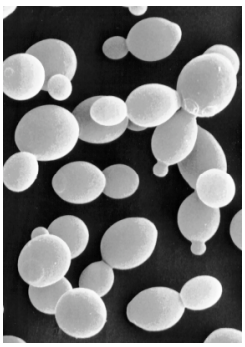
MAPK pathways are a target of many bacterial pathogens.
In many cases, bacterial proteins target MAP kinase pathways, because these are essential for immune response and cell growth. Figure 3 shows a variety of bacterial proteins that target different steps in MAPK signaling. YopJ is a protein that acetylates the two amino acids in MAPKKs that are normally phosphorylated. This blocks the pathway function. OspF cleaves the phosphorylated Thr residue in a MAPK. Both YopJ and OspF can downregulated MAPK pathways in diverse organisms, ranging from humans to yeast.

Fig. 3. Several pathogen proteins downregulate MAPK signaling



YEAST AS A MODEL SYSTEM FOR HIGHER ORGANISMS & CLONING STRATEGIES

For iGEM, by Sergio Peisajovich (Lim Lab, May 2007)



The yeast *Saccharomyces cerevisiae* (also called "baker's yeast") is probably the most ideal eukaryotic microorganism for biological studies. The "awesome power of yeast genetics" (i.e.: the ability to "manipulate" the yeast genome by adding or removing genes of interest) is ideal for synthetic biology. In addition, the complete sequence of its genome allows us to make rational and, in principle, precise genetic engineering. Furthermore, the basic mechanisms of yeast cell biology (such as DNA replication, recombination, cell division and metabolism) are highly similar to that of higher organisms (including humans); thus, many of the lessons we learn studying fungi (yes, yeasts are fungi!) are relevant to human health.

1- Yeast life cycle

Most yeasts are unicellular organisms with complex internal cell structures similar to those of plants and animals. There are two forms in which yeast cells can survive and grow, haploid and diploid, differing, among other things, in the number of chromosomes they carry. Haploid cells have a single copy of each chromosome, whereas diploid cells have two copies. The haploid cells undergo a simple lifecycle of growth by mitosis (cell division). The diploid cells (the preferential 'form' of yeast) similarly undergo a simple lifecycle of mitotic growth. However, under conditions of stress, they can undergo sporulation, entering meiosis and producing a variety of haploid spores, which can go on to mate, reforming the diploid (see Fig. 1).

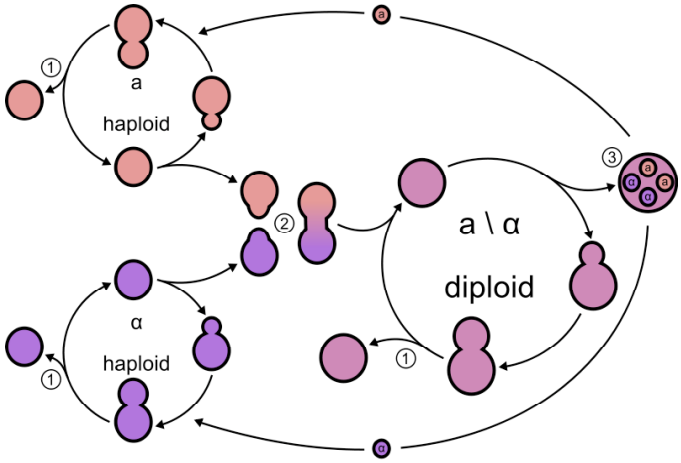


Fig.1: Yeast cells of type "a", "α" or "a/α" can grow by "budding" (mitotic division). Two haploid cells of opposite type ("a" and "α") can mate, leading to a diploid "a/α". Diploids can sporulate, producing 4 daughter haploid spores, two type "a" and two type "α" that could subsequently reproduce by budding, re-entering the cycle.

Yeast cells carry their genetic information in sixteen different nuclear chromosomes and a mitochondrial chromosome. However, most yeast strains carry also a so-called 2-µm circle plasmid, coding genes that apparently function solely for their own replication.

2- Yeasts are an ideal "platform" for synthetic biology

The ease at which the yeast genome could be manipulated was central to yeast becoming the ideal Eukaryotic model microorganism. Indeed, it is this very same attribute that could make yeast an ideal platform (or "chassis") for synthetic biology. What are in fact the "genetic operations" that we would like to be able to make for synthetic biology applications?

(a) We would like to add devices or even modules (see the section on synthetic biology for brief explanations about these terms) to yeast cells, so that the biological output expected (or not!) from these modules could be measured. In fact, we would like to do so in two ways: either by adding parts/devices/modules in an "extra-genomic" format (plasmid-based); or "integrating" these parts/devices/modules within the yeast genome.

(b) We would like to delete specific yeast genes, so that we could study the function of any given synthetic part/device/module we add without the "background" that could be caused by some genes (which are indeed "natural" parts/devices/modules) being already present in the cell.

(c) We would like to add "reporter genes" to more easily monitor, in real time, the function of the synthetic parts/devices/modules under study.

(d) Finally, we would like a life cycle fast enough so that we could do all these genetic manipulations in a reasonable amount of time (iGEM runs only for the summer…).

As you should have already imagined, yeasts meet all these requirements and, fortunately, there are available methodologies that shall allow us to perform all the mentioned procedures with relative ease and in a relatively short amount of time.

3- Cloning strategies

<u>Plasmid-based yeast transformation</u>

The simplest way to add any desired synthetic part/device/module to yeast is by "transforming" yeast cells with an appropriate plasmid. To refresh our minds: plasmids are circular DNA molecules that can be replicated autonomously (although in some special cases this will not be true in yeast, we will discuss that below) and that can be manipulated easily, so that DNA fragments could be added or removed from them. In

practice, we do all DNA manipulations in plasmid that are replicated in bacteria (*Escherichia coli*) mostly because bacteria divide much faster than yeast, and therefore we can get our "constructs" ready in less time. Once the plasmid contains all the parts/devices/modules required, we transform them into yeast for the real experiments to be done (see Fig. 2).

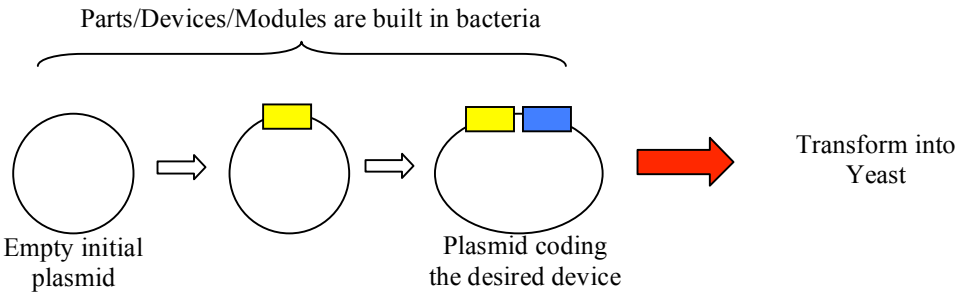


Fig. 2

In bacteria, plasmids are "selected" and "maintained" usually because they encode certain genes (mostly a single gene) that are required for the cell to grow in a particular "selective" environment. Most typically, the plasmid encodes an enzyme able to destroy (hydrolyze) a certain antibiotic. In this way, the cells that carry the plasmid are the only ones that are "protected" from the antibiotic toxicity, and are therefore able to grow. Similarly, there are antibiotics that are used to select and maintain plasmids (carrying genes that confer "resistance") within yeast cells. An alternative very frequently used to select and maintain plasmids in yeast is the use of "auxothrofic" strains. These are yeast strains that have "lethal mutations" in genes responsible for producing some nutrients that are required to grow, so that they need to take these nutrients from the growth medium. If one transforms a strain of these with a plasmid that carries a non-mutated version of the gene, then it is possible to select for the cells carrying the plasmid by growing the culture in a medium lacking the essential nutrient (see Fig. 3).

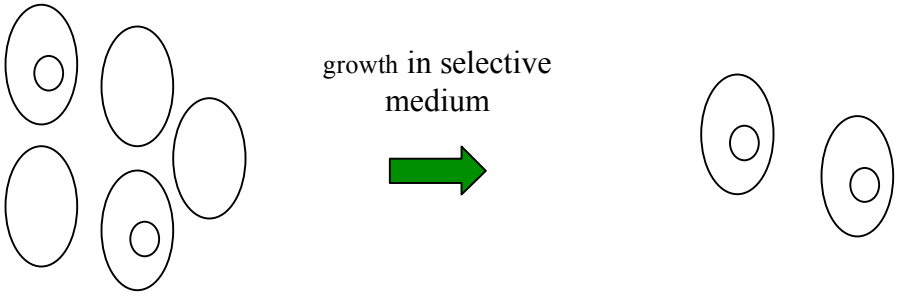


Fig. 3: by growing cells in a "selective" medium (with an antibiotic or without an essential nutrient) one could select cells carrying the adequate plasmid (shown as a small circle).

Note that the plasmids used to build parts/devices/modules in bacteria and be transferred into yeast, need to carry both selective genes to allow growth of bacteria in medium containing antibiotic and yeasts in medium lacking an essential nutrient.

Homologous recombination for gene knock out, replacement and integration in yeast

In some occasions one would like to "integrate" constructs (including parts/devices/modules) into the yeast chromosome instead of keeping them in plasmids. While the reasons for this could vary, generally speaking integrated constructs are more stable (plasmids many times "lose" some of the genes they carry, though could you guess which genes could never get lost?) and, one might say, they are also more likely to reflect a "natural" scenario, as most yeast genes are indeed encoded in chromosomes rather than in plasmids.

How is chromosomal integration of parts/devices/modules achieved? Here again, the basic yeast cell machinery comes to the rescue of the synthetic biologist: yeast cells have enzymes capable of "recombining" homologous DNA molecules (that is two DNA molecules that have the same sequence, see Fig. 4) in a way that any region of its chromosomes could be "replaced" by a foreign piece of DNA, as long as the incoming DNA possesses regions of homology on both ends.

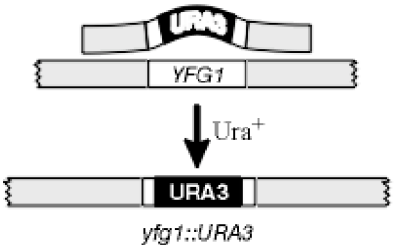


Fig. 4

In the example shown if Fig. 4, homologous recombination leads to a gene knock out. "Your favorite yeast gene 1" (yfg1) is targeted for knock out by transforming the yeast cells with a linear fragment of DNA that, on the sides, has regions of homology (the exact same DNA sequence) to the DNA flanking yfg1. Note that the incoming DNA contains also an internal gene (in this case named "URA3") that serves as a selection marker. URA3 encodes an enzyme required for the synthesis of uracile (a precursor in DNA synthesis). The yeast strain whose yfg1 gene is being targeted has a defective variant of its endogenous URA3, therefore the culture medium needs to be supplemented with uracile to allow growth. Cells in which the linear DNA coding for a functional copy of the URA3 gene is properly recombined (and thus the new URA3 gene is "integrated" into the yeast chromosome) are the only ones able to grow in medium lacking uracile. Note that, if the incoming DNA has a novel copy of yfg1 (that we could called yfg1*) in addition to a selective marker (URA3 or any other marker available), then integration as shown above would not lead to a knock out of the gene yfg1, but rather to its replacement by the gene yfg1* (or by a construct containing any desired combination of part/devices/modules!).

A similar method (based also on homologous recombination) could be easily used to integrate parts/devices/modules, but without disrupting any yeast genes within the chromosome (see Fig. 5). In this case, the linear, incoming DNA is flanked by regions homologous to a marker (URA3 for instance; note that the version of URA3 present in the yeast chromosome is defective, whereas the version encoded in the incoming DNA is functional, that is to say: they are homologous but still they differ slightly in some critical "mutations"). In addition to the flanking marker, the incoming DNA encodes any desired part/device/module.

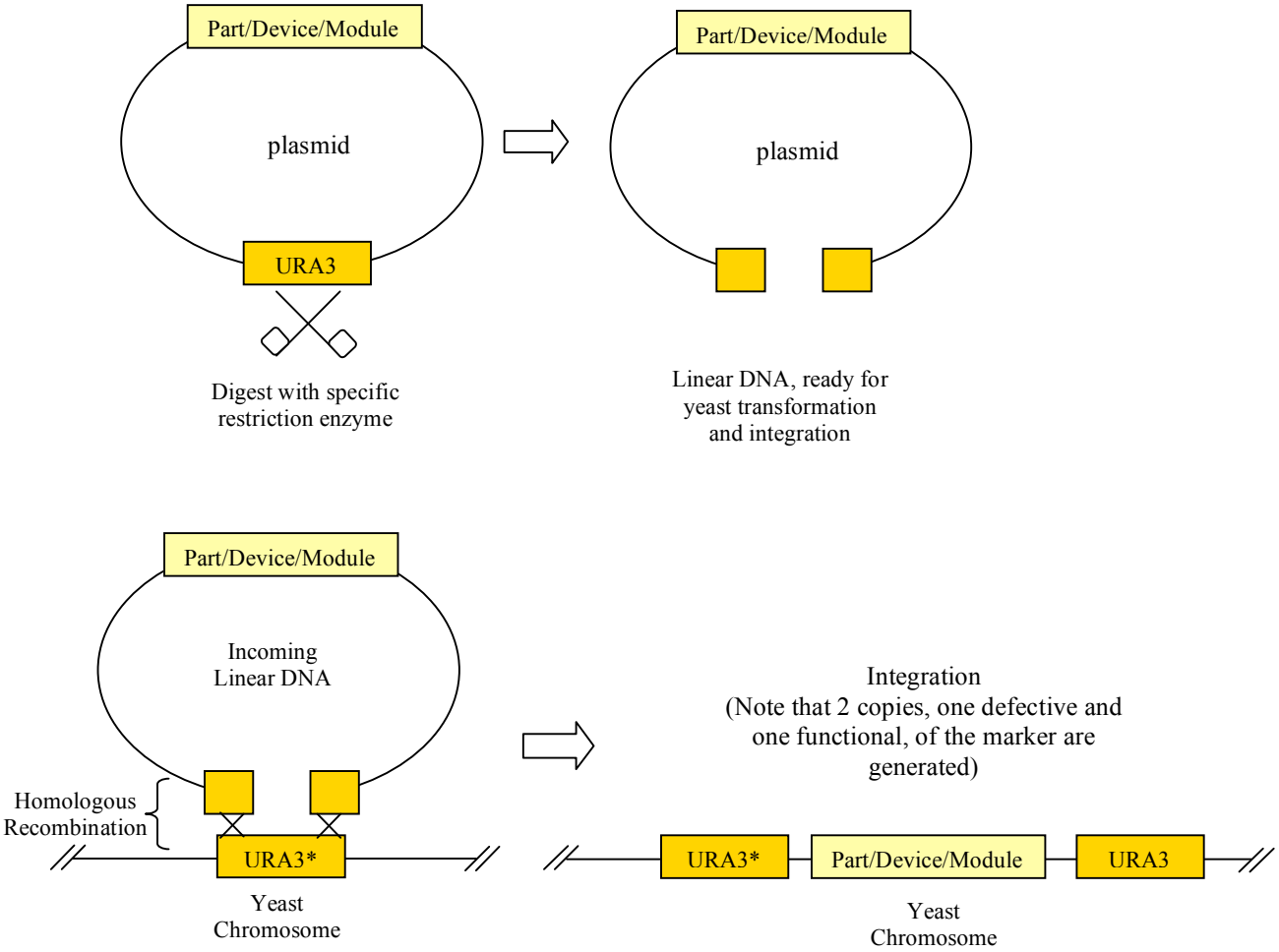


Fig. 5: The incoming DNA containing the desired part/device/module is first "linearized" by digestion with the appropriate restriction enzyme, so that both ends now contain sequences homologous to the target region in the yeast chromosome. After transformation into yeast, homologous recombination leads to the integration of the desired construct and generates two copies of the marker, one still defective but one functional that allows selection.

Combinatorial cloning strategy

The basic element in synthetic biology is the "part" (yes, go back to the section on "Synthetic Biology" and refresh those concepts!). We have said that this could be any DNA sequence coding for a particular function (e.g. a promoter for transcription, any gene, etc.). However, most synthetic biology applications require many different parts working together (that is "a device") or even many different devices combined (that is a "module" or "system"). Simple cloning strategies could be used to "assemble" all the required parts within a plasmid, before we transform them into yeast (either to stay as plasmids or to be integrated into one of the yeast chromosomes). In Fig. 6 we show how an iterative cycle of digestions and ligations could serve to "add" parts (one at a time) so that, after several cycles, a full device (or module) is ready for yeast transformation.

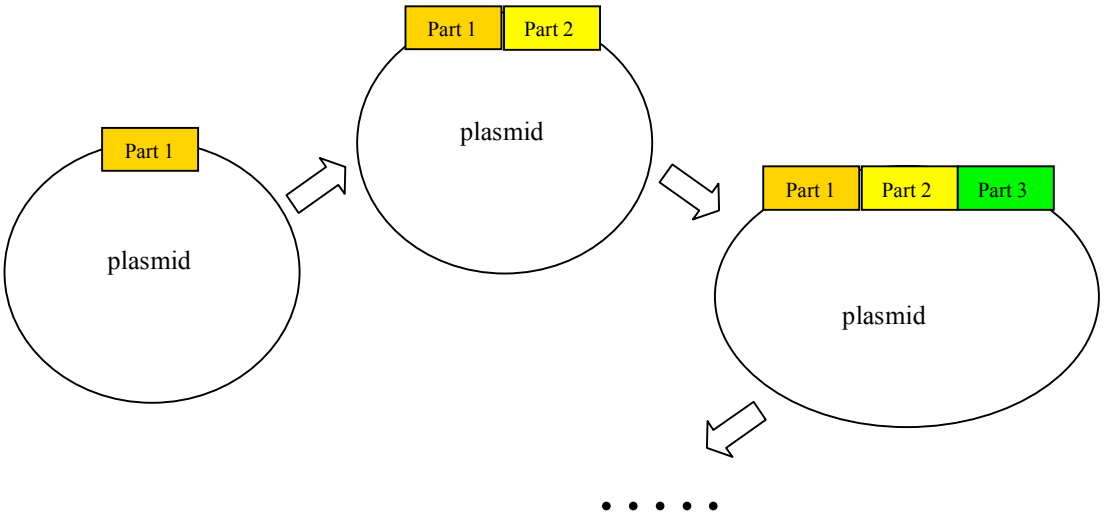


Fig.6

Even though this sequential building strategy could be applied to synthetic biology applications (indeed, it is still the most common building strategy in use), it is very time consuming (remember that each step requires the transformation of the ligated plasmid into bacteria before the next part can be added). Thus, synthetic biology would clearly benefit from a building strategy that would reduce the amount of time needed to build complex devices. For this purpose, we are working to adapt a novel **combinatorial building strategy** first proposed by Guet and co-workers (see references at the end of this section) to the specific needs of the synthetic biology applications of our laboratory. Theoretically, other laboratories could also easily adopt this technique as well. The novel idea of this combinatorial building strategy is that, instead of adding (by ligation) one part at a time, we could add two, three, or even more, parts in any single step, so that complete devices could be assembled in a **single ligation/transformation event** (see Fig. 7).

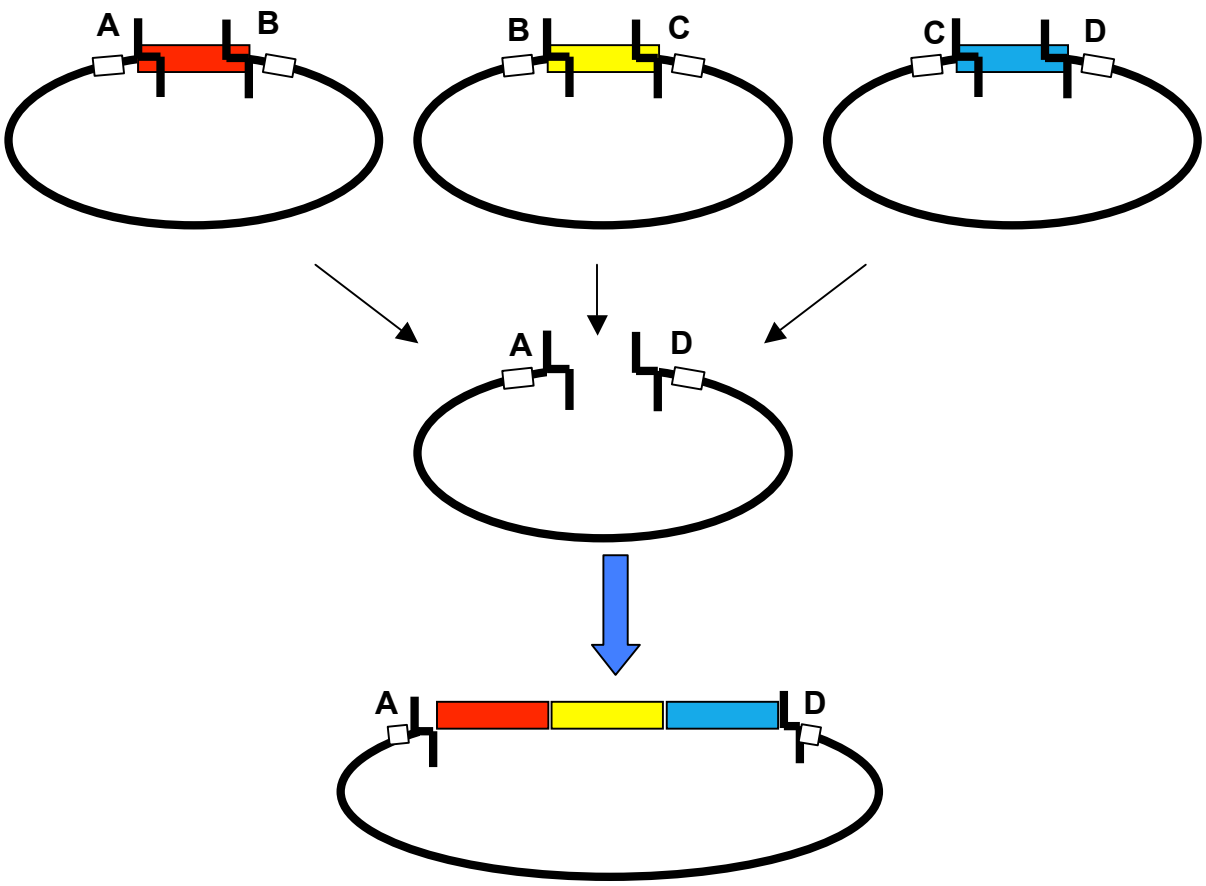


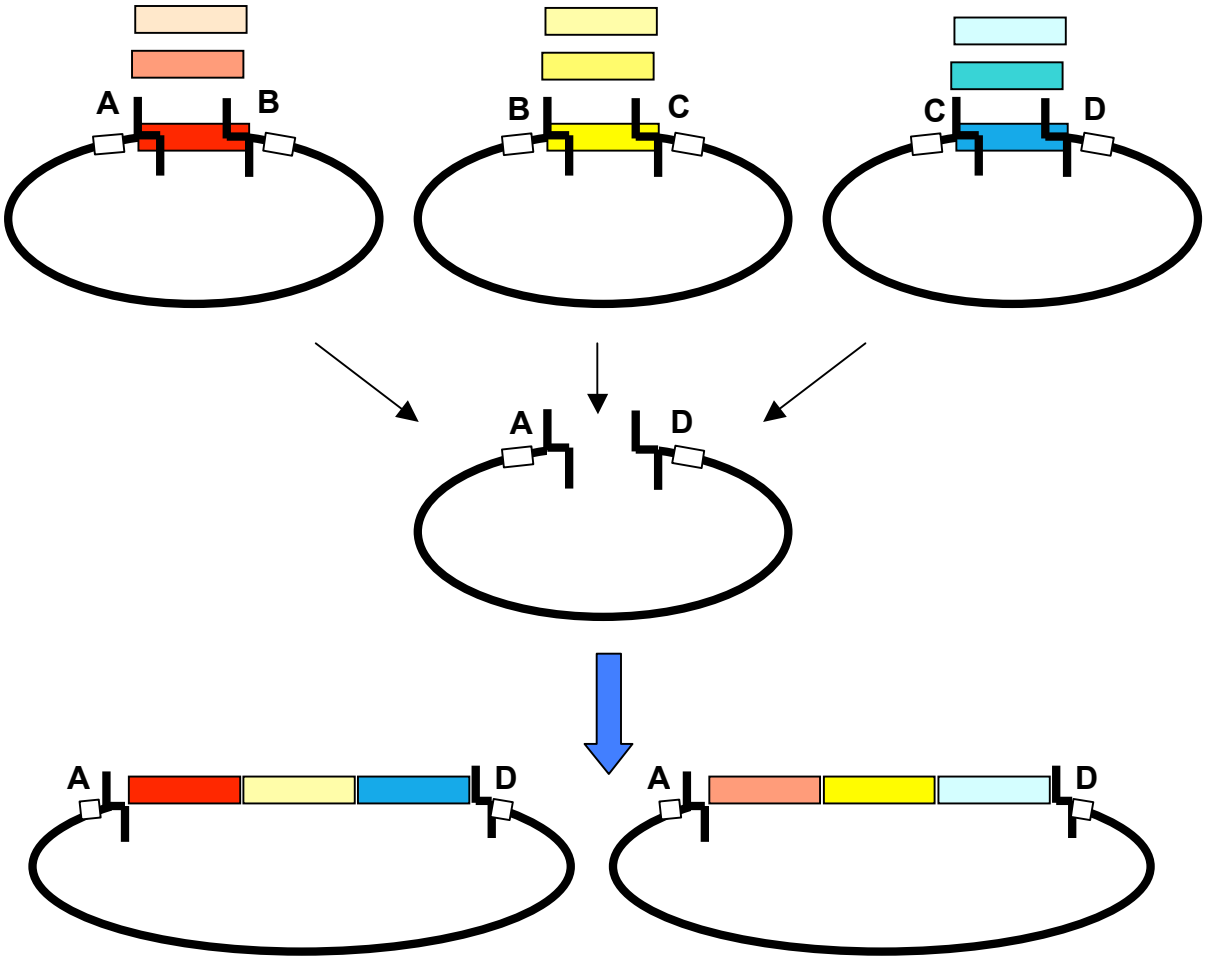
Fig. 7: A novel combinatorial building strategy allows the simultaneous ligation of multiple parts, creating a full device in a single ligation/trasnformation step.

This strategy is based on the use of a particular group of **restriction enzymes** called **Type IIs**. As any other restriction enzymes, type IIs enzymes recognize a specific sequence in the DNA. However, unlike other more common enzymes, type IIs enzymes do not cut the DNA ***within*** the recognition sequence, but rather a certain number of bases ***away*** from the recognition site. Thus, the "sticky ends" that are generated after DNA cleavage can be ***rationally designed*** by the synthetic biologist. Of course, we use this to our advantage; by choosing the apropriate sequence for the sticky ends, we can create any series of consecutive matching patterns. In this way, parts that are required to be in a particular order in the final device, are actually generated with the adequate sticky ends.

In the example shown in Fig.7, three parts are needed in the final device, in a particular order: red, yellow, blue. For that, each part is first cloned into a "donor plasmid" that is flanked by two restriction sites for a type IIs enzyme (in the lab, we used an enzyme named "AarI") but that on each side has different "sticky ends". For convenience, we called them "A, B, C or D." Remember, though, that each one is 4-bases long and contains a unique combination of DNA bases. In particular, "A" is GGAG, "B"

is CCCT, “C” is GCGA, and “D” is TGCG. Although there is no particular reason why we have chosen these sequences, note that they should be (and indeed are) non-palindromic (paindromic sequences look the same when they are read in the same direction in the two complementary DNA strands, e.g. 5’-ATGCAT-3’ is palindromic, because the complementary strand would read 3’-TACGTA-5’). By being non-palindromic, they cannot anneal (and get ligated) to themselves (you might want to check if and why palindromic sequences could do so). This is the reason why multiple parts could be ligated at once, still with a relatively good yield.

A second advantage of this strategy is that all parts are cloned in “donor vectors” in a standard way. This allows us to assemble different devices by simply combining parts in any possible way (in a particular order or in any possible order by having the same part cloned in all possible donor vectors, see Fig. 8).



… and any other possible combination!

Fig. 8

References

For general information about yeast techniques:

1- Virtual library of the Yeast Genome Database
http://www.yeastgenome.org/VL-yeast.html

2- Fred Sherman's yeast site
http://dbb.urmc.rochester.edu/labs/sherman_f/yeast/Cont.html

Combinatorial cloning:

Guet, C.C., Elowitz, M.B., Hsing, W. & Leibler, S. (2002) Combinatorial synthesis of genetic networks. Science, 296:1466-1470.