Teaching Plan

I used to see teaching as a chore taking time away from research, which led me to apply and obtain a "charge de recherche" position in 1994. Since then, I have progressively revisited this opinion and today I see teaching and research as intimately intertwined. It now appears to me difficult to sustain a high level of scientific productivity without contributions from generations of graduate students and post-docs. The fast turnover typically found in academic labs ensures a very stimulating renewal of ideas. In environments like corporate R&D departments with low turnover rates, I witnessed good scientists losing their creativity for lack of opportunities to be challenged by younger scientists. When a research group insulates itself from the rest of the scientific community, a consensus progressively builds up preventing the emergence of new ideas that might compromise the stability of the group. I have come to question the assumption that a full time research scientist not engaged in a teaching activity is more productive than a scientist who is obligated to dedicate a significant fraction of his time to teaching.

Similarly, the teaching of someone disconnected from research would quickly be challenged by the fast developments in the life sciences. The best teachers I had were the ones who could share a unique scientific perspective structured around their passion for a particular research problem. Just like writing grants and other fund-raising activities are keys to securing the material resources necessary to conduct experiments, I would regard my teaching activities as a competition to attract the best human resources to my group.

What is the most important for me to transmit to future generations of scientists is an ability to embrace interdisciplinary research projects. It is my experience that multidisciplinary research teams composed of specialists of different disciplines are rarely very effective. Modelers who often have no idea of the complexity of what they model are not very well positioned to make relevant analyses of data. Experimentalists with limited understanding of experimental design tend to consider the modelers as computer support persons instead of integral parts of their scientific team. There is a strong need for scientists with a scientific personality strong enough that they are capable of outgrowing the territorial limits of their primary education.

I want to train students capable of approaching the molecular mechanisms of gene expression in quantitative ways by collecting data at the bench and relating these data to computer models. Some may be more gifted at the bench. Others might create incredible pieces of software. It is also possible that some of them will later be known for new theoretical developments. No matter what direction they go after graduation, they will all be required to spend some time in the lab, understand what a model is, be capable of writing basic scientific computing scripts in a high-level programming language, and have some appreciation of the computational cost of numerical simulation and optimization. Through the experience of being pushed out of their comfort zone while in my lab, they will gain the self-confidence necessary to explore new scientific territories and to adapt to the rapid changes of the life sciences after they graduate.

Part of my students scientific growth is the development of their leadership skills. No matter whether they join the industry or seek faculty positions after leaving my lab, they will soon need to demonstrate an ability to join or even build scientific teams. The most mature students will be asked to act as mentors/coach of younger members of my group. This coaching experience will be an opportunity for them to learn how to create a safe environment around them where more junior people can grow and shine. Ultimately, after their post-doc they should be capable of managing a project requiring the contribution of younger people with technical skills outside their personal field of expertise. They should be able to evaluate the contribution of all members of the team and adapt the level of supervision to their teammates evolving needs without loosing track of the project progress.

Research Plan

Background

Since 2000, engineers and physicists have revisited the qualitative methods of genetic engineering by assembling artificial gene networks in which two or more genes interact in a quantitative way to achieve sophisticated phenotypes. The founding papers of what is now known as synthetic biology were published in January 2000. Gardner described a bistable network with two genes repressing each other (Gardner et al., 2000). Elowitz described a three gene network exhibiting oscillations (Elowitz and Leibler, 2000).

While a lot of emphasis has been put by systems biologists on the discovery of the wiring diagrams of the cell regulatory mechanisms, synthetic gene networks demonstrated that wiring is not enough to determine a phenotype. Constructs with identical wiring diagrams could have different phenotypes simply by altering the strengths of regulatory interactions represented by this wiring diagram. Early on, Gardner experienced this limitation when one of the six toggle switches he assembled failed to exhibit bistability (Gardner et al., 2000). However, no one has demonstrated better than Guet that the phenotype of a gene network is not solely determined by the connectivity of network of interactions (Guet et al., 2002). By randomly inserting one of five promoters upstream of three repressors, he generated a library of 125 plasmids implementing as many regulatory networks. A subset of this library was phenotyped using a screen designed to characterize the Boolean function implemented by the network. Surprisingly, different logical gates could be implemented by networks with identical diagrams. In some cases, a single point mutation was sufficient to change the logical behavior of a network. This work illustrates that minor quantitative variations of the regulatory interactions in a regulatory network can result in dramatic qualitative changes in the phenotype.

The problem is that parameters are unknown. In some cases, it is possible to perform a bifurcation analysis of the model representing the gene networks dynamics. This analysis can help identify large regions of the parameter space with qualitatively similar solutions. For instance, general conditions to achieve bistability (Gardner et al., 2000) or oscillations (Elowitz and Leibler, 2000) could be determined that way. However, the large uncertainty around the actual values of the physical constants associated with the model parameters prevented other groups from achieving the phenotype they intended to implement (Atkinson et al., 2003). Synthetic biology would benefit from estimates of the weight of regulatory interactions. Yet, a number of scientists dismiss the possibility of ever getting this information because any model simulating the dynamics of gene network would be too simplistic to fit a comprehensive set of experimental data (Kim and Tidor, 2003).

In order to deliver on its promise, in order to go beyond the proof of concept stage, synthetic biology needs to characterize the modularity and portability of genetic elements used in genetic constructs. Modularity refers to the idea that the quantitative effect of a part is somewhat independent of the environment in which is it is used. For instance, a strong promoter will lead to a high of expression of any gene it will control the transcription of. Numerous evidence show that genetic regulatory systems are modular at different levels of organization from pathways down to protein domains (Hartwell et al., 1999). However, numerous evidences also show that parts can interfere with one another in unexpected ways. There is therefore a strong need to develop methods to detect and quantify interferences between elements used in genetic constructs so that we can better understand them and use that knowledge in the design of new constructs. Apart from Fussenegger's group work (Kramer et al., 2004; Kramer and Fussenegger, 2005), synthetic biology has mostly prototyped gene networks in simple model organisms. This approach is practical if there is a clear path to adapting these prototypes to other more complex organisms where they could be used in various medical or commercial applications. Below are short outlines of the research programs I want to initiate to address these two important issues.

Modularity of artificial gene networks

- Using a limited number of genetic elements, I would build libraries of plasmids in E. Coli with increasing levels of complexity. A good example of this type of approach is the set of 21 calibration plasmids developed by Gardner prior to assembling the bistable switches. This library is described in the online supplement of (Gardner et al., 2000).

- After introducing environmental perturbations, I would observe the relaxation of the gene network toward a new steady-state using flow-cytometry. Since single cell expression data are more informative than population average, it is my expectation that these times series of gene expression distribution will give a better insight in the internal dynamics of the networks.

- Using mass-action models and the hybrid simulation algorithm developed in collaboration with Sanders' group (UIUC), I would fit the model to match its simulated distribution to the observed series of flow-cytometry data. Starting with models representing only a limited number of simple networks, I will progressively add more networks to the models using initial constraints corresponding the modularity hypothesis and relaxing these hypotheses as needed (Peccoud and Vander Velden, 2003) where interference is detected. Alternative model topologies could also be explored either manually or using an evolutionary approach (Peccoud and Vander Velden, 2003; Francois and Hakim, 2004).

- The goal is to build a model of a genetic space generated by a limited set of genetic elements used in a specific context. The model should be trainable using data collected on a subset of the genetic space and be capable of predicting the phenotype of constructs not observed in the training phase.

- The expected deliverables of this research direction are:

    - This progressive approach should help clarify aspects of artificial gene networks that can be modeled using simple models and the ones that required more complex models that could be indicative of novel biological mechanisms.

    - Variants of biological parts could be used to revisit on a larger scale and in a more quantitative way the traditional structure-function analyzes. This will help delimit sequences corresponding to parts in a way that minimizes interferences between parts.

    - This project will determine experimental protocols leading to informative data sets that can be used to fit models of genetic spaces.

    - The size of the genetic spaces that can be described in a single model remains an open question. Is it possible to build a model of the 21 plasmids generated by Gardner? Of the 125 plasmids of the Guet library? Since the two libraries use the same elements, is it possible to build a model encompassing both libraries?

Portability of artificial gene networks

I am not aware of any artificial gene network that was implemented in more than one organism using the same set of genetic elements. Yet, the idea of developing a switch or a quorum sensing system in E. Coli and adapting it to mammalians cells to use it in gene therapy applications or other applications is very attractive if there is a predictable path from E. Coli to mammals.

This work would require developing a process that would streamline the customization of the genetic construct sequence to different organisms. A couple of issues need to be addressed. First it is necessary to develop experimental methods to quickly assemble large transformation vectors by combining molecular cloning and *de novo* gene synthesis approaches. The second problem has to do with the "editing" of the DNA sequence based on a number of well-known rules such as codon usage, etc. As our ability to quickly synthesize DNA molecules increases, I would be interested in evaluating the possibility of developing a grammatical approach to convert a logical description of artificial gene networks into multiple DNA implementations suitable for transformation into different species (Searls, 2002). I would see this project developing in collaboration with a group active in synthetic genomics such as George Church's group, Codon Devices, Inc., or the company recently created by Craig Venter, Synthetic Genomics, Inc..

Once the same constructs are implemented in different organisms it would be very interesting to use comparable phenotyping and modeling approaches on data collected on different organisms to quantitatively characterize the portability of all the components used in the constructs. This problem can be studied at two levels. The portability of the construct will be ensured if its dynamics is conserved across species. Another important issue, especially in the perspective of using the constructs in medical applications is to evaluate how invasive is the artificial gene network on the host cell physiology. In order to achieve a predictable function, it is desirable to minimize the risk of

interferences between the synthetic network and the host cell endogenous regulatory mechanisms. Perturbations caused by the artificial gene networks could be assessed using large-scale profiling techniques such micro-arrays or metabolic profiling. These differential expression experiments could be related to the prediction power of models of genetic space. They could also help in designing minimally invasive artificial gene networks.

After having developed a process to adapt constructs to complex organisms and having gained evidence of the predictable phenotypes of these constructs across species, collaborations will be established to test the constructs in transgenic animals to support research projects.

Resources needed

- Wet lab to conduct molecular biology work, E. Coli, Yeast, and mammalian cell cultures.
- Easy access to a flow-cytometry facility.
- Access to gene expression profiling services (microarrays)
- Access to a high-performance computing environment such as a Linux cluster.

Bibliography

Atkinson,M.R., Savageau,M.A., Myers,J.T., and Ninfa,A.J. (2003). Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in Escherichia coli. Cell *113*, 597-607.

Elowitz,M.B. and Leibler,S. (2000). A synthetic oscillatory network of transcriptional regulators. Nature *403*, 335-338.

Francois,P. and Hakim,V. (2004). Design of genetic networks with specified functions by evolution in silico. PNAS *101*, 580-585.

Gardner,T.S., Cantor,C.R., and Collins,J.J. (2000). Construction of a genetic toggle switch in Escherichia coli. Nature *403*, 339-342.

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

Hartwell,L.H., Hopfield,J.J., Leibler,S., and Murray,A.W. (1999). From molecular to modular cell biology. Nature *402*, C47-C52.

Kim,P.M. and Tidor,B. (2003). Limitations of quantitative gene regulation models: a case study. Genome Res. *13*, 2391-2395.

Kramer,B.P. and Fussenegger,M. (2005). Hysteresis in a synthetic mammalian gene network. Proceedings of the National Academy of Sciences of the United States of America *102*, 9517-9522.

Kramer,B.P., Viretta,A.U., Baba,M.D., Aubel,D., Weber,W., and Fussenegger,M. (2004). An engineered epigenetic transgene switch in mammalian cells. Nat. Biotechnol.

Peccoud, J. and Vander Velden, K. Computer System for Genotype to Phenotype Mapping Using Molecular Network Models. [US2005/0086035 A1]. 9-1-2003.

Searls,D.B. (2002). The language of genes. Nature. *420*, 211-217.