Herbert M Sauro
Keck Graduate Institute
535 Watson Drive
Claremont, CA 91711

October 10th, 2005

Yves Brun
Systems Biology/Microbiology Faculty Search
Department of Biology
Indiana University, Jordan Hall 142
1001 E 3rd St
Bloomington IN 47405-7005

Dear Sir

In response to the call for applicants to faculty positions in systems biology at Indiana University, Bloomington, I would like to submit my application. I enclose the relevant documentation, including research interests, teaching interests and relevant publications. I have requested that my letters of recommendation be sent to you as soon as possible.

I look forward to hearing from you in the future.

Yours sincerely

H. M. Sau

Herbert Sauro
Assistant Professor
Keck Graduate Institute

Statement of Research Interests

Herbert Sauro

My interest in how biological reaction networks function began in my undergraduate days. Unfortunately throughout the 1980s and 90s, systems biology was not popular; in fact, grant agencies were disinterested in the analysis of systems using computational approaches. It was only with the sequencing of the human genome that understanding biological systems using computational approaches became a more acceptable area of research. Mainstream researchers began to realize that the complexity of living organisms was far too great for Molecular Biology to remain a simple descriptive science.

In the past few years, there has been an incredible shift in attitude not only by mainstream scientists, but also by many grant agencies. This change in attitude allowed me to return to academia after some years of absence. As a result of my past experience, in 2000, I was fortunate to be invited to work at Caltech, Pasadena to begin the development of SBML (Systems Biology markup Language) which enables different software tools to exchange computational models. SBML proved to be the right idea at the right time; SBML is now used by over eighty software tools, both academic and commercial and in many ways has been a resounding success. Nature magazine recently (May 5th, 2005) wrote an editorial which endorsed our standard.

We are now entering a golden age for systems research in biology, encompassing experimentation, theory and computation as a means to understand biological function. My own specific areas of interest include:

1. Development of methods to understand network function.
2. Reverse engineering network function.
3. Synthetic Biology
4. Evolution of natural control systems.
5. Innovative software development.
6. Development of approaches to ensure that computational models are reliable.

I will briefly describe each topic as follows.

Innovative Software Development

I began developing software for simulating and analyzing cellular networks when I was an undergraduate. I believe I was the first person to write network simulation software (called SCAMP) for the PC. This software included a number of novel ideas which others authors have since followed. SCAMP proved to be very popular; over 80 citations record its use by the community. Since that time I have gone through a number of software iterations, and now focus attention on what is called the Systems Biology Workbench (SBW). One of the problems that plagues modern software development in systems biology is chronic reinvention. There is little true innovation in systems biology software today - most tools are simple wrappers around ODE solvers (or even simpler implementations of Gillespie). SBW is different in this regard, and represents a novel, extensible framework, that allows new functionality to be easily added in a variety of languages (including C/C++, Java, Delphi, C#, VB, Matlab, Python and Perl). Thus, rather than writing yet another ODE solver, one uses an existing ODE module on SBW to extend functionality. As a result of this body of work, we now have a wide variety of tools, ranging from optimizers, novel model editors to bifurcation tools. There are a number of groups in the world

which now actively contribute to this effort, including ISB, Virginia Tech, University of Tennessee, and groups in Europe and Japan (cellDesigner). Other groups have expressed an interest but have not yet made concrete contributions. Below is a screen shot of one of our model editors, a visual design tool that is tailored for users learning to model, or causal users who need a tool that is easy to use. We also develop more advanced tools for seasoned users; in particular Jarnac is a script based simulation tool that permits users to have precise control over simulations and analysis. We are currently achieving over 1000 downloads of SBW per month. All our software is hosted on sourceforge and is fully open source. This work is funded by DARPA and DOE/GTL .

We also conduct tutorials on who to use our software, for example, we will be giving a modeling tutorial at the ICSB 2005 conference in Boston, MA this October. We also recently gave a tutorial in Boston this August. The software is also used extensively in my teaching at KGI and we have a tutorial booklet available online for anyone to use.

Other related areas:

1. Continued development of SBML, particularly the addition of extensions to enable the graphical representation of biological networks (See figure below).

2. We are part of the team developing biomodels.net. We supply the curation expertise to the new biomodels.net database which is geared towards the storage of computational models of interest to systems biology.

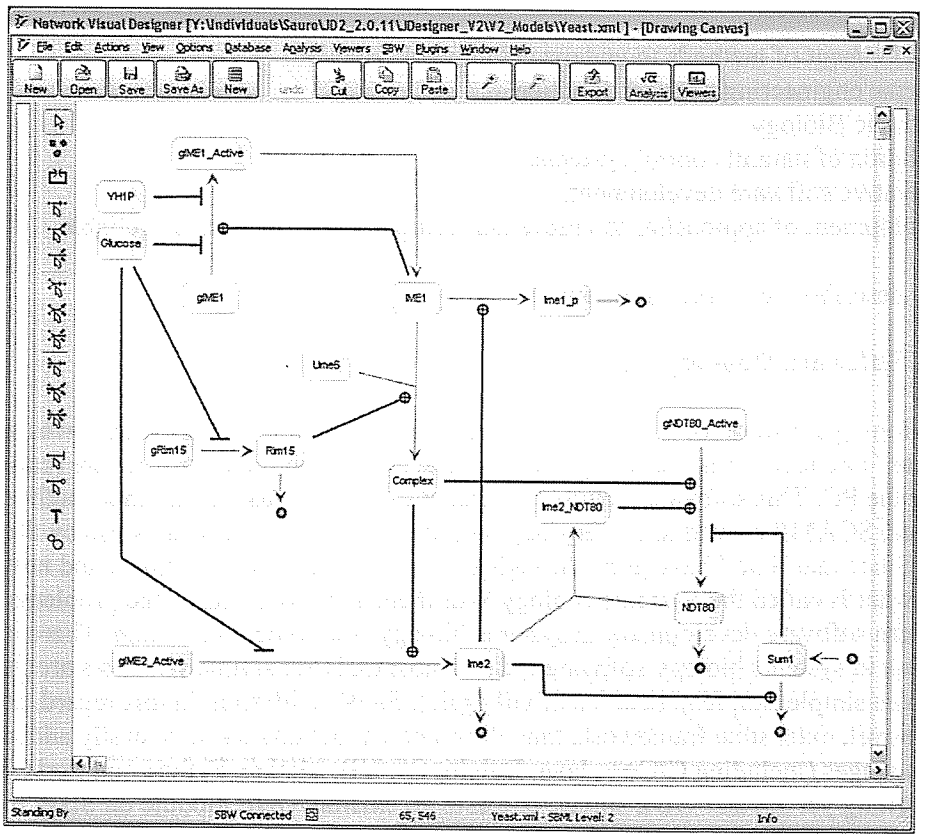

Figure: One module that comes with SBW is a visual design tool for building computational

network models. The above model illustrates one of our projects - modeling of yeast sporulation.

Development of Methods to Understand Network Function

When I started in this area, I naively thought that one could use simulation as a means to understand cellular networks. My first attempt at this was to model glycolysis, but I soon realized that simulation alone was not sufficient. Although the computer faithfully modeled glycolysis, understanding the dynamics of the simulation was very difficult because there was simply too much information. It was at this point that I realized the need for some kind of theory as a means to interrogate and distill the data into a more manageable form. In searching for such a theory I soon discovered the theory of Metabolic Control Analysis (MCA), developed by Kacser, Heinrich and Savageau. This theory permits one to express the properties of a pathway in terms of the properties of its individual components. MCA proved an ideal partner for simulation and most of my work prior to coming to the US was the development of this approach. In collaboration with other groups, including Oxford, Waterloo and Indiana, I have continued to explore MCA. For example in work with Ingalls (Waterloo) and Rao from Indiana, it was discovered that there was a deep connection between MCA and classical control theory. This means we can effectively combine the extensive theory and analysis techniques that one finds in control theory to understanding cellular networks. I think this is particularly welcome given the new and exciting developments in synthetic biology. Just as the design of a radio requires some understanding of control theory, I believe that developing non-trivial synthetic networks will require the input of some kind of design theory for biologists.

Reverse Engineering Network Function

One of the key areas in modern biology is functional analysis. Images of complex networks, such as those developed by Kurt Kohn at the NCI, can be overwhelming to the human eye. One of the areas I am particularly interested in is methods to breakdown such complexity into simpler functional units. Just as an electronic circuit can be modularized into different sections, I believe we can do the same for cellular networks, particularly signal and gene networks. Our approach (funding from NSF) to this problem is to develop a library of functional motifs and to then map these onto real networks. The problem is the initial development of a comprehensive motif library. To do this, we us a novel evolutionary method performed *in silico*. We start with a functional requirement, such as an oscillator or a frequency filter, and then apply an evolutionary approach to evolve gene networks, signaling networks and simple mass-action networks that can mimic the desired function. We have been very successful with the approach, and now have a suite of computer-evolved networks for most of the functional elements found in engineering. These include motifs that act as switches, oscillators, filters, mathematical operations, integration, differentiation, comparing, mixing, and more. Note that most of these motifs are analog in nature.

Synthetic Biology

My group has recently started a synthetic biology research project. Our first goal is to try an implement an amplitude filter, that is, a network that will only transmit signals of specific strengths. Once this is complete, the synthetic network will form the basis for additional modules. The network will use a hybrid system comprising of both protein and RNA control systems.

I am also collaborating with Brian Ingalls and Chris Rao on building additional synthetic biology network. Chris Rao in particular has a lab dedicated to synthetic biology.

Evolution of Natural Control Systems

Closely allied with the reserve engineering project is an interest in how natural control systems evolve, and whether the emergence of modularity is important and what role sex plays in the process. We are developing *in-silico* models which incorporate a linear genome, which in turn codes for a network. We can apply various mutational operators including point mutation, gene duplication, gene swapping, and crossover of homologous networks. We have already observed the evolution of modularity in one instance which has encouraged us to continue this work. At the moment this work is at an early stage. However we have recently been awarded a 4.9 million grant from the NSF (FIBR) to peruse this topic in much more detail.

Development of Approaches to ensure that Computational Models are Reliable

One of the most significant problems in model construction today is model validation. How can we know whether a model is reliable or not? The traditional approach is to fit models to experimental data; the simplest model with the best fit is usually considered the best model. This approach has draw back, for example it only tests a very limited number of models, and there is no way to be certain whether the model is indeed the best one. Consequently, we decided (in collaboration with Adam Arkin at Berkeley) to develop a completely new approach: instead of validating a model, which is essentially impossible, we attempt to falsify it instead. In addition, instead of testing just a few models, we will allow the computer to generate 10,000's of model variants (variants in the sense of network configuration), and the computer will attempt to falsify each model in turn. Very briefly, the method of falsification depends on carrying out a Monte Carlo fit under two experimental conditions, for example a partial deletion of a gene can be considered one condition. Under the two conditions, the kinetic constants are unchanged and therefore both fits should yield the same kinetic constants, if they don't then the model is falsified. This method is however very computationally intensive and we have developed special parallelized optimizers to deal with the large computational load. Both Microsoft and IBM have expressed an interested in the approach and IBM in particular have encouraged us to apply for support. In addition, several experimental groups have also expressed interest in this approach, including the Institute for Systems Biology and the Receptor Tyrosine Kinase consortium.

Industrial Collaboration

- Genomatica – Development of a flux balance model for the Yeast *Pichia pastoris.*
- Close links with others companies includes Gene Network Sciences (internships) and Entelos (employment opportunities for our students).
- Industry Panel, Boston October 2005

Currently Active Academic Collaboration

- Brian Ingalls (University of Waterloo) – Understanding cellular networks from a control engineering perspective.
- Steven Wiley (PNNL) – Development of data standards for systems biology.
- Adam Arkin (Berkeley) – Development of a new radical approach to model falsification.

- Animesh Ray (KGI) – Development of automated model development tools through literature mining.
- Chris Rao (Illinois at Urbana-Champaign) – Investigating the evolution of cellular networks using *in silico* methods.
- John Tyson (Virginia Tech) – Developing a new user friendly bifurcation analysis tool.
- Kurt Kohn and Mirit I. Aladjem (NCI Betheasda) – Developing graphical user interfaces for modeling signaling pathways.

Currently Held Federal Grants:

NSF (2004-2007)	Understanding the computational role of cellular networks.
NSF FIBR (2005-2010)	Structure and Origins of Functional Modules (CoPI).
DARPA (2002-2004)	Development of model fitting software for BioSPICE and SBW.
DARPA (2005-2006)	Curation of Models, support for SBW development, including parallelization of data fitting for large scale models.
DOE (2004-2007)	Development of SBW, graphical user interfaces.

Conclusion

In summary, I have an active research program with a good balance of theory, computation and collaboration with experimentalists. In the past several years, I've been an active community member, and have been successful in procuring financial support from a number of different grant agencies. I have also successfully designed and developed undergraduate and graduate-level courseware, and have integrated my research interests into a successful REU program.

Statement of Teaching Interests

Herbert Sauro

Qualifications and Experience:

- I have a formal qualification (PGCE) in teaching obtained from a UK University. This is a one year degree course to train graduates for teaching positions in high schools and colleges in the UK.

- I spent two years teaching science and IT to 11 to 18 year olds, with specialization in chemistry and biology for senior students to prepare them for UK University entrance.

- I have developed two full length semester courses at the Keck Graduate Institute:

  1. Introductory computer programming for biologists.

  2. Computational Systems Biology to second year graduates.

- I teach an intense three week introductory course on computer modeling to the first year graduates.

- I have currently redesigning the 1st year graduate courses at KGI to emphasize systems and molecular biology (see next page).

- Graduate Students: three, with a forth on rotation.

- REU Students: Two each summer funded by a NSF REU grant (Currently seven students to date, one resulted in the publication of a peer review paper).

- I have been invited to speak and organize a number off site courses on systems biology and modeling (One in August and one in October this year).

- When time permits, I help out at the local Claremont high school by judging a science fair.

- My group will contribute a four to five hour tutorial on modeling at this year's ICSB 2005 conference in Boston.

Design of a New Systems Biology Curriculum

Although KGI is a graduate school we accept students from many disciplines, including non-scientific areas such as finance. Therefore the first year 'graduate' courses do not tend to contain graduate level material. Instead the curriculum is a mix of early and late undergraduate courses.

We have split the new curriculum into four broad areas, Genomic, Proteomic, Systems and

Applications. The Genomic and Proteomic areas are standard introductions with the addition of computational elements and will not be discussed further here. The areas of primary interest to me in terms of novelty are the Cellular Organization and Systems Level Analysis. The course is quite intense, we have one semester (12 weeks) to cover a lot of material. I will be teaching all of Cellular Organization and some of Systems Level Analysis. The courses will be mixed in with practical classes which will largely be computational and mathematical in nature, for example building models, carrying out a flux balance analysis or studying the behavior of specific network motifs. Because of my interest in the application of engineering principles to systems biology, the course will have an engineering bias. The course could easily be expanded to cover a number of undergraduate level courses.

Cellular Organization

Topic	Duration (weeks)
Introductory Computational Systems Biology	½
Continuous and Stochastic Modeling	1
Signaling Mechanisms and Signaling Pathways	1
Building Computational Models	½
Flux Balance Analysis	½
Network Motifs and Control	1
Synthetic Biology	½
Programmed Cell Death	½
DNA Damage Repair and Cell Cycle Control	½
Computational Cell Cycle Control	½
Total Duration	6.5

Systems Level Analysis

Topic	Duration
Integration of Genomic and Proteomics	1 (3 hrs)
Impact of Systems Biology on Studying Complex Diseases	1½ Paper Review (4.5) hrs
Cancer, Neurodegenerative disorders and Autoimmune Diseases	2 Project (~6 hours)
Total Duration	4.5 Weeks

Long Term Plans

The Keck Graduate Institute is quite small and other than a reworking of the second year systems biology curriculum (which is also used by our PhD students) I have probably reached the limit for course development. I am therefore interested in extending my experience in curriculum development in systems biology to the undergraduate level, particularly to try and run a theme of computation and systems throughout the undergraduate curriculum. In addition to the undergraduate level there is also graduate teaching. I currently teach courses for our small PhD group, bringing them up to speed in computational methods and theory in systems biology. Once again, our PhD students are from a wide background, and therefore the course needs a broad curriculum, encompassing a number of seemingly separate topics.

I am also interested in developing further software tools for assisting education in systems biology and ultimately to develop a text book suitable for undergraduate teaching.