DEPARTMENT OF HEALTH & HUMAN SERVICES
Public Health Service
National Institutes of Health
Bethesda, Maryland 20892
www.nih.gov

October 26, 2005
Re: Assistant Professor

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

Dear Search Committee,

I am writing to be considered for an appointment as **Assistant Professor of Systems Biology** in the **Department of Biology**. My name is Vincent VanBuren and I am presently a postdoctoral fellow in the Laboratory of Genetics at the National Institute on Aging, NIH, in Baltimore. I earned my PhD in Molecular Biology from Lehigh University. Although my academic training is in Molecular Biology, I have almost exclusively used computer simulations and analysis to answer biological questions in my research.

To give you a quick idea of the impact of my research, I have enclosed a "New and Notable" about my upcoming work in Biophysical Journal, November 2005). That article, written by Henry Schek and Alan Hunt at the University of Michigan, describes my new work entitled, "A mechanochemical model of microtubule structure and self-assembly kinetics". Both articles are presently available online ahead of print:

New and Notable:
http://www.biophysj.org/cgi/rapidpdf/biophysj.105.067462v1
my article:
http://www.biophysj.org/cgi/rapidpdf/biophysj.105.06091.

My interests in computational genetics and computational molecular biophysics are diverse and forward thinking. I have made contributions to the field of Embryogenomics, and I am presently working on comparisons of microarray analysis performed by different groups using different platforms. I recently submitted a career transition grant proposal to the AHA in the area of gene regulation networks (July, 2005), and I am planning a future proposal to extend my microtubule simulation work in the area of cancer research.

You can reach me on my cellular phone at 410-746-0814 or by email at vincent_vanburen@nih.gov. I look forward to discussing this position with the Search Committee.

Kind regards,

Vincent VanBuren, PhD

Postdoctoral Fellow
Developmental Genomics and Aging Section
Laboratory of Genetics
National Institute on Aging, NIH
333 Cassell Drive, Suite 3000
Baltimore, MD 21224
Office: 410-558-8410
Cellular: 410-746-0814
Email: vincent_vanburen@nih.gov

DEPARTMENT OF HEALTH & HUMAN SERVICES
Public Health Service
National Institutes of Health
Bethesda, Maryland 20892
www.nih.gov

Statement of Research Interests

Vincent VanBuren
National Institutes of Health

My present position as a Postdoctoral Fellow at the National Institutes of Health has provided me with rich opportunities for collaboration with bench scientists, and I have seized those opportunities whenever it appeared that the results of collaboration would be productive. I have broad experiences in a leadership role in Bioinformatics and Computational Biology, and my work of late has focused on novel approaches towards an understanding of the relationship between measured fluorescence intensity from DNA microarray experiments and absolute transcript abundance, essential knowledge for the reconstruction and simulation of biochemical networks (see below). I would like to continue this work while organizing the infrastructure and team of individuals necessary for an internationally recognized research group in computational biology and bioinformatics. Below I will offer a description of some of my research experiences and plans, and discuss some of the general principles applied to my work.

The value of computational approaches to biology may be summarized in three categories: (1) *interpretation*, or the analysis of biological data for the purpose of uncovering hidden relationships using an accepted model, (2) *prediction*, or using models to formulate computational hypotheses that may be further explored with bench experiments, and (3) *model generation*, or producing a new model by fitting known constraints on an *in silico* system in order to generate known results and new hypotheses. Interpretation and prediction using computational approaches requires some computational knowledge, including programming skills for novel approaches, and requires some knowledge of the biological model employed. Attempts at model generation using computational approaches, however, require both computational skills and a thorough understanding of the biological system of interest.

My past and current research efforts in computational biology have incorporated all of these approaches. My graduate work at Lehigh University addressed the dynamic properties of microtubule assembly and disassembly. Microtubules are protein aggregates essential for cell division, and are the target of several anti-cancer treatments. I used Monte Carlo simulations to generate a new pseudo-mechanical model for the assembly, disassembly and the rates of switching between these two states. This was the first work that considered mechanical effects during microtubule assembly. This effort produced many insights, including predictions for the lateral and longitudinal bond strength between tubulin dimers, a prediction for the mechanical energy of strain within the microtubule dimers buried in the microtubule lattice, theoretical support for a model of XMAP215's action in binding to microtubules, as well as other predictions. This first part of my work was published in **PNAS**. The success of the pseudo-mechanical model motivated the development of a mechanochemical model that explicitly incorporated mechanical effects in a 3-D chemical kinetic model of microtubule assembly. This model is able to account for many of the forms microtubules are known to take in assembly and disassembly, including ram's horns, frayed ends, sheets, extensions, and blunt ends. This 3-D model of microtubule assembly and disassembly is the first of its kind, offering the first integration structural/mechanical effects and the chemical kinetics of microtubules. This latter work is in press at **Biophysical Journal** [Epub ahead of print June 10, 2005, in print November 2005], and Biophysical Journal will publish a companion "New and Notable" article about my manuscript (see enclosure). My collective work on microtubule simulations provides insights into rational drug design for the treatment of cancer by providing predictions for the effects of targeting either the disruption or strengthening of lateral and longitudinal dimer associations, thereby interfering with the dynamicity of microtubules required for cell division.

In my postdoctoral training at the National Institute on Aging, I have worked on diverse bioinformatics and computational biology problems, including: leading an effort in annotation and quality control for the assembly of cDNA clones into a large collection representing libraries derived from early embryos, stem cells, and newborn

organs in the mouse (published in **Genome Research**), development of in-house Web-based applications to speed research efforts, and participation (co-author) in development of a mouse gene index that includes many genes unique to early mouse development (**PLoS Biology**). My current projects include studying the relationship between probe hybridization mean intensity on microarrays and the absolute abundance of the target transcript being measured, and comparison of microarray data across different labs and platforms. Knowledge of absolute transcript abundance is invaluable for any effort to reconstruct genetic networks. Presently, microarrays are used to measure the *relative* abundance of transcript given two collected samples, while the *absolute* abundance is obscured. At least two things contribute to clouding the relationship between probe intensity and absolute transcript abundance: probe quality and target amplification when the samples used are rare (such as preimplantation embryos). Probe quality is related to the physical properties of the probe, which include factors such as G+C content, probe length, position of the probe within the target, and the energy of hairpin formation. Target amplification may skew the relationship of mean intensity and absolute abundance in a way that is dependent on the number of rounds of amplification. To both provide a useful tool and better understand how abundance might be estimated from mean intensity, I created a Web-based application that automates probe design for large-scale applications, including DNA microarrays and large-scale *in situ* hybridization (ISH) (**IEEE CSB Proceedings, 2003** and a work in preparation). This program has different modes for each of these large-scale applications, and as there are presently much more microarray data than ISH data, the probe design tool is better optimized for microarrays. I recently developed a new probe scoring algorithm that uses a neural network to score the complex relationship between probe quality and average mean intensity, and thus selects the optimal probe (best scoring) for a target-of-interest. This should greatly facilitate the construction of custom arrays (providing an easier design process) and generally improve DNA microarray studies by providing more probes with measured intensities significantly different from background intensity (providing more data).

By incorporating further microarray data with multiple probes against different regions of the same gene or spike-ins of known quantities of target, it should be possible to adapt the tools I constructed for probe design to estimate absolute transcript abundance from measured mean intensity on DNA microarrays. A recent paper I co-authored in **Genome Biology** describes a method for estimating transcript copy number from measured fluorescence intensities. Neural network predictions should offer an improvement to the accuracy of those predictions. An estimate of absolute transcript abundance is a necessary ingredient in Monte Carlo simulations of metabolic networks. Monte Carlo methods applied to biochemical processes are simulations in which biochemical changes are treated as stochastic processes, where there is some probability that a given molecule will participate in a given reaction. Alternative approaches, including the use of differential equation models and mixed models, should also benefit from more accurate estimates of transcript abundances. The key idea here is that knowledge of absolute transcript abundance from microarray experiments will reveal a large amount of information about metabolic constituents, thus presenting the opportunity to build rich models that will more accurately describe the underlying networks of genetic and biochemical processes of the cell.

In the next 3-5 years, I am planning to (1) apply my efforts in nucleic acid probe design towards creating an application that will calculate absolute transcript abundance from measured intensities in microarray experiments, (2) use data mining techniques and Monte Carlo and differential methods to reconstruct metabolic pathways from microarray data and other high-yield data sets, including yeast two-hybrid data, (3) apply bioinformatics methods and data mining strategies towards a better understanding of the cellular mechanisms involved in RNAi, and (4) refine methods of simulating microtubules and interactions between microtubules and microtubule associated proteins (MAPs), and the interaction between microtubules and chemotherapeutic agents that interact with them. I recently submitted an application for a Scientist Development Grant that partly addresses (2) and (3) above. The involvement of graduate students, technicians, and postdoctoral fellows, as well as collaborative involvement will all contribute to the vibrancy of this research plan. Graduate students or technicians may contribute toward Web-based applications, including interface design, connectivity between the interface and the applications, and may participate in developing efficient algorithms for analyzing probe sequences and calculating abundances. Graduate students and technicians may also participate in any of the above research goals to the extent that they have an interest and aptitude for the work, but it seems likely that (1) and (3) above are the most likely candidates for graduate students or technicians, while (2) and (4) are more suitable for advanced graduate students or postdocs (or myself).

The research described above demonstrates an aptitude and appetite for diversity. As it is research from the laboratory bench that has inspired all of my work, I am committed to building relationships and collaborations with bench scientists whenever possible (as with my graduate and postdoctoral work). In addition to cultivating collaborations with traditional molecular biologists, it has been important to my work to build relationships with specialists outside the field of molecular biology. In my previous work, I have formed close collaborations with chemical and mechanical engineers (graduate work) and a statistician (postdoctoral work), all of which yielded fruitful results. Looking towards the future, possible collaborations with clinical researchers offer the tantalizing potential to develop new prognostic and treatment strategies from high-throughput clinical data. My greatest strengths are to conduct productive research in a leadership role with diverse approaches and to synthesize diverse and complex ideas while effectively communicating, collaborating and building relationships with experts from diverse fields of study.

Vincent VanBuren, PhD
National Institute on Aging, NIH

DEPARTMENT OF HEALTH & HUMAN SERVICES
Public Health Service
National Institutes of Health
Bethesda, Maryland 20892
www.nih.gov

Statement of Teaching Philosophy and Interests

Vincent VanBuren
National Institutes of Health

Mark Twain has famously said, "I never let my schooling interfere with my education." I think the best teachers understand this notion in a very personal way. For myself, I played two roles in my undergraduate education, where one role was to follow where my curiosity would take me, reading about and learning subjects that excited my interest. The other role was that of an 'actor', acting like a good student, not always inspired about what I was learning, but performing well because I knew it would help provide me with foundation skills and a strong academic record. Of course I selected a field of study and elective courses that were in line with my own interests, but my expectations for particular classes were not always met, and some instructors failed to be inspiring or to convey ideas in palatable way. We all know of students that turned away from a subject that first interested them, after being discouraged by a poor educational experience. Sadly, schooling sometimes does interfere with education.

What are the ingredients of a good teacher? The most obvious aspects of teaching are (1) to convey some set of "facts" to students for their mastery, and (2) to facilitate an understanding of how to think critically about those facts. Many of the worst teachers know this much. Good teachers, however, also know that there are many other facets to teaching than to simply arm the students with some information and some intellectual tools for processing that information, including (3) effectively communicating those ideas, (4) appropriately challenging the students, (5) treating students fairly, (6) engaging the students with questions, (7) showing rather than telling, (8) applied learning, (9) adaptation to the needs of students (flexibility), (10) passion, (11) fun and humor, (12) providing take home messages, and what I think is one of the most important keys to learning (13) providing a motivation to learn other than getting a good grade. The very best teachers are inspiring, igniting the curiosity and passion of their students to learn more.

For personal development, teaching is a great way to keep an active understanding of both relatively broad subjects (undergraduate) and advanced topics (graduate), and shapes future science by sharing insights and guiding new scientists. I am experienced in teaching several laboratories as a Teaching Assistant, and was chosen to teach a faculty lecture course in Human Genetics and Reproduction when the professor for that course required a long-term absence. The laboratories I instructed included three semesters of two sections of Genetics Laboratory, two semesters of two sections of Histology Laboratory, and one semester of one section of Advanced Cell Biology Laboratory. I received strong evaluations for each course I instructed.

Presently, my strongest interests for undergraduate teaching would be courses in Introductory Bioinformatics, Introductory Biophysics, Introductory Genomics, Advanced Cell Biology, or Genetics. At the graduate level, my hope is that there is some flexibility in designing special topics classes. Some of my interests in this regard would be subjects like: advanced sequence analysis, computational discovery of *cis* regulatory elements, modeling biochemical systems, modeling complex biochemical systems with compartmentalization, modeling the biophysical/biochemical properties of macromolecules (computational molecular biophysics), reconstruction of cellular biochemical pathways from transcriptome and proteome data, large-scale biology, and the cytoskeleton. I am also capable of teaching any of the core courses of a graduate program in Molecular Biology.