September 9, 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 Search Committee,

I am writing to apply for the tenure-track faculty position in the Department of Biology at Indiana University.

I received my PhD in Mathematics and Physics in 1984 from the Institute of Physics of the Earth, Russian Academy of Sciences (Moscow, Russia) and my MS in Electrical Engineering in 1975 from Moscow Technical University. I also took additional computer classes and courses in the period 1994-1998.

After graduation from the University I worked as a computer engineer there and in parallel I taught the laboratory course of thermodynamics. From 1980 to 1993 I worked at the Russian Academy of Sciences as assistant research scientist, research scientist, and senior researcher involved in multiple scientific projects. During that time I participated in numerous domestic and international conferences, symposiums and scientific schools and taught several courses of Physics for graduate students. I moved to United States in 1993, since than I have been working as a researcher as well as a computer programmer.

I have diverse scientific experience, including bioinformatics, applied mathematics, physics and geophysics, with over 60+ journal articles, patents, and scientific reports. My current scientific interests include: regulation of gene expression, computational approaches to identifying *cis*-regulatory elements, promoter mapping, nonlinear waves in DNA, and software development for scientific and educational purposes.

I am a fluent computer programmer (C++ and Visual Studio are my favorite language and tool, respectively). I have worked on diverse software- related projects, including computation of cooling-down face of hydroelectric turbine fabrication (Moscow Technical University), computer tools such as "Full Authority", "Check Site" and "Site Right" (Lexis-Nexis), and virtual laboratory experiments for the general physics laboratory (Wright Sate University).

I think that my unique combination of scientific experience, computer skills and ability to generate and implement ideas, together with my dedication to scientific work, makes me an appropriate candidate for this position.

Enclosed are the required research statement, teaching statement, two reprints, and curriculum vitae. Recommendations can be obtained from:

1. Michael Ostrowski
   Professor
   Department of Molecular Virology, Immunology and Medical Genetics

834 Riffe Building
496 W 12th Avenue
Columbus OH 43210-1214
Office Phone: (614) 688-3824
email: Michael.Ostrowski@osumc.edu

2. Gary D. Stormo
   Professor
   Department of Genetics
   Director, Computational Biology Program
   Washington University Medical School
   660 S. Euclid, Box 8232
   St. Louis MO 63110
   Office Phone: (314) 747-5534
   FAX: (314) 362-7855
   email: stormo@genetics.wustl.edu

3. Gust Bambakidis
   Professor
   Physics Department
   Wright Sate University
   Dayton OH
   Office Phone: (937) 774-2760
   email: gust.bambakidis&wright.edu

4. Brian A. Lewis, Ph.D.
   Adjunct assistant professor
   Department of Biochemistry, UMDNJ-RWJMS
   683 Hoes Lane
   Piscataway, NJ 08854
   Office Phone: (732) 235-4194
   email: lewlisba@umdnj.edu

5. Michael B. Gokhberg
   Professor
   Academician, Russian Academy of Sciences
   Department of the Physical Fields of Earth.
   Institute of Physics of the Earth
   Moscow, Russia
   Office Phone: (7-095) 254-6134
   Email :gmb@ifz.ru

6. Leonid S. Alperovich
   Professor
   Geophysics & Planetary Sciences
   Tel Aviv University
   Tel Aviv, Israel
   Phone Number : 972-3-6407411
   Email : leonid@frodo.tau.ac.il

Sincerely,

Naum Gershenzon, PhD
Postdoctoral Researcher
Biomedical Informatics
The Ohio State University
3184 Graves Hall
333 W. 10th Avenue
Columbus, OH 43210
Office Phone: (614) 292-0187
email: naum.gershenzon@wright.edu
        or gershenzon-1@medctr.osu.edu

Research Statement

Naum Gershenzon

The major areas of my scientific interest include:
- Regulation of gene expression
- Computational approaches to identifying *cis*-regulatory elements
- Combinatorial analysis of *cis*-regulatory elements
- Transcriptional networks
- Eukaryotic promoter mapping
- Modeling of DNA dynamic processes
- Software development.

Regulation of gene expression

Gene expression is a fundamental process involving participation of many regulatory proteins. We human been are now at the preliminary stage of understanding of this multi-step multi-level process. Less than one fourth of all known promoters can be described on the level of basal transcription machinery by known transcription scenarios. We developed and successfully applied a computational approach, allowing classification of eukaryotic promoters based on known transcription mechanisms as well as prediction of new mechanisms. I'm going to further develop and apply this technique to investigate the transcriptional mechanisms for the remaining three fourths of the promoters.

Computational approaches to identifying *cis*-regulatory elements

The identification of thousands of transcription factor binding sites (TFBS) (so-called *cis*-regulatory elements) for each of thousands of transcription factors is a necessary part of the elucidation of the gene expression puzzle. Many experimental and computational approaches have been developed during the last three decades to solve this problem. One of the most widely applied methods for searching TFBS is Position-Weight Matrices (PWM), yet the majority of existing PWMs provide a low level of both sensitivity and specificity. Recently, we developed and applied a computational algorithm allowing improvement of PWM quality based on the putative sites revealed from the promoter database. I am planning to use this algorithm extensively to improve PWMs for dozens of important TFs. The described technique works for TFBSs having a well-defined bias positional distribution in the proximal promoter area. However, there are many TFs effective at distant promoter areas such as members of the p53 protein family. I'm going to further develop the algorithm for PWM improvement to include such TFs based on combinatorial analysis of *cis*-regulatory elements and using putative sites conserved between species.

Combinatorial analysis of *cis*-regulatory elements

The *trans*-regulatory elements usually work in complexes and some of the members of these complexes bind to DNA. So the *cis*-regulatory elements also form complexes. The necessity for TFs to work with partners explains how nature provides the required level of specificity for the regulation of transcription. Combinatorial analysis of such complexes could be used for studying and classifying transcription patterns as well as increasing the specificity of *cis*-regulatory element identification. I have recently

developed a program to search for potential pairs of TFBSs and I'm planning to continue this search.

Transcriptional network

The complex functions of a cell are carried out through the synchronized activity of many genes. This activity is often coordinated by the sets of co-regulated genes, regulatory modules. Identifying these modules is crucial for understanding cellular responses to external and internal signals. Several computational techniques have been developed to reveal regulatory modules using genome-wide expression profiles. I propose new approach. The main idea is to use temporal (dynamic) expression data. This allows connect the intensities of gene expression by the system of ordinary differential equations. Knowing the dynamic behavior of each gene expression one could determine the coefficients of these equations by solving inverse problem. From these coefficients one is able to identify the dependency between genes. Fully developed this method should allow revealing transcriptional networks responded on variety of external and internal signals.

Promoter mapping for eukaryotic genomes

Although nearly all protein coding genes in many of the eukaryotic genomes have been successfully identified, the exact promoter mapping is still an unsolved problem. The techniques developed by us for promoter classification, PWM improvement, and combinatorial analysis are the necessary elements to build a reliable state-of-the-art algorithm and computer program for promoter mapping. I'm planning to release the preliminary version of such a program in a few months.

Modeling of DNA dynamic processes

For over a century a paramount task of traditional mathematical physics has been the solution of three linear partial differential equations: the wave equation, the heat conduction equation and Laplace's equation. Their importance is explained by their exceptional universality. In the last few decades, this list of fundamental equations has been enriched by some essentially nonlinear equations, such as Korteweg-de Vries, nonlinear Schrodinger, and the sine-Gordon equations. These equations have much in common. In particular, they have special solutions (solitons) which are localized in time and space and are similar to classical particles. I have had previous experience applying the sine-Gordon equation to seemingly unrelated objects and processes such as dislocation motion in crystals, laser beam propagation through so-called two-level systems, and movement of tectonic plate. It is known that DNA can be also described by sine-Gordon equation, which means that all mathematical apparatus developed to find nontrivial solutions of this equation could be apply to describe dynamic processes in DNA. Recent experimental data such as "DNA breathing" is the appropriate candidate for such analysis. I'm planning to spend part of my time developing this promising area.

Software development

There is a countless number of publicly available software packages which could be utilized for various type of biological data analysis. Yet every new task seems to require more suitable and flexible software. Being a fluent C++ programmer I usually prefer to develop the necessary software myself. Thus, I created the package "Promoter Classifier", which includes seven programs for promoter database analysis, a program utilizing the algorithm for PWM improvement, a program for combinatorial analysis of *cis*-elements, and many other programs for scanning DNA sequences. For the future, I'm

also planning to develop additional software for my own research and for the research community.

Teaching Statement

Naum Gershenzon

1. Teaching Philosophy

Teaching is a necessary part of scientific work providing the possibility to constantly update one's knowledge of the subject, to exchange ideas and points of view, and to generate new ideas through student questions and discussions. I grew up in a scientific community influenced by the spirit of great Russian scientists such as Lev Landau and Peter Kapitza. They always dedicated time and energy to their pupils and considered on-going contact with them as an essential part of their scientific life.

2. Teaching Experience

From 1975 to 1980 I worked as computer engineer at Moscow Technical University and in parallel I was a teaching assistant for the course *Thermodynamics* (fall quarters from 1976 to 1979) .

During my work at the Russian Academy of Sciences (1980 - 1993) I participated in six summer schools where I taught a series of lectures including "Tectono-electromagnetic Phenomena", "Electrodynamics of Magnetosphere and Ionosphere Plasma", and "Mechano-electromagnetic Phenomena".

At Wright State University I was closely involved in the creation and teaching of the software package "Virtual Laboratory Experiments for the course *Physics 202: General Physics Laboratories: Electricity and Magnetism*". This software package includes complete simulation of eight experiments and allows students to prepare themselves for the real experiments, understand more deeply the underlying physical laws through simulation and visualization of processes which cannot be seen in the real experiments, and even perform virtual experiments instead of real experiments.

During the last two and a half decades I have participated in over 40 conferences, symposiums, and schools presenting review lectures and the results of my works.

3. Teaching Plans

Initially I plan to involve postgraduate students in the research process through computer analysis of various biological structures as well as through the creation of software.

For the future, using my interdisciplinary background in physics and applied mathematics, I am planning to create a bioinformatics course (or a portion of it) in which the physical properties of biological structures (such as stacking energy or bending stiffness of DNA) as well as sophisticated mathematics (such as non-linear equations) will be incorporated in modeling of biological processes and bioinformatics analysis.