Michael L. Blinov

Theoretical Biology and Biophysics Group,
Los Alamos National Laboratory,
Los Alamos, NM 87544, USA
Phone: (505) 667-4111 (w), (505) 983-2060 (h)
Fax: (505) 665-3493
e-mail: mblinov@lanl.gov

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

Dear Members of the Search Committee,

I am writing to express my interest in a tenure-track faculty position in Computational Cell Biology at Indiana University. Attached are my curriculum vitae and research objectives.

I did my Ph.D. thesis in theoretical mathematics at the Weizmann Institute of Science (Rehovot, Israel) under the supervision of Professor Yosef Yomdin. My thesis led to five first-author publications in journals such as *Nonlinearity* and *Func. Differ. Equ*. I also taught "Mathematical Modeling in Biology", a unique graduate course designed by the late Professor Lee Segel for experimental biologists. It struck me how fascinating and complex biology is, and how mathematics can be applied to study biological systems. I took several bioinformatics courses and my enthusiasm for applying mathematics to problems in biology was heightened. By recommendation of Professor Lee Segel, I came to the Theoretical Biology and Biophysics Group (T-10) at Los Alamos National Laboratory in the summer of 2001 for an internship with Byron Goldstein and William Hlavacek. During this internship, I worked on computational modeling and analysis of cell signaling mediated by the high affinity receptor for the IgE antibody. The model developed with my assistance (*J. Immunology*, 2003) encompasses 354 different protein complexes derived automatically according to specified rules.

After graduation from the Weizmann Institute of Science, I accepted a Postdoctoral Research Associate position in T-10 (under supervision of Byron Goldstein). To generalize our rule-based approach to model a MAPK cascade in yeast and the initial signaling events mediated by the epidermal growth factor receptor (EGFR), I developed BioNetGen, general-purpose software for rule-based modeling of biomolecular interactions (*Bioinformatics*, 2004). Further, I worked on the development of enhanced software (*Lect. Notes Comp. Sci. Trans. Syst. Biol.*, in press) that allows the user to account for connectivity of proteins. Using the software, I have modeled the early events in signaling initiated by the epidermal growth factor receptor (EGFR), describing 356 protein complexes (*BioSystems*, 2005). The methods implemented in BioNetGen software serve as a basis for a patent application 20050042663 with a second patent application in preparation.

I believe I can make a strong contribution to the efforts of the Department of Biology at Indiana University in the area of computational and systems biology, focusing on analysis and simulation of biological processes at intracellular and cellular levels, and developing tools assisting in it. In particular, my initial plan would be to concentrate on the following projects:

- Mathematical modeling of Epidermal Growth Factor Receptor Signaling in collaboration with experimental groups. Collaborations with teams from Thomas Jefferson University (B. Kholodenko), University of Virginia (K. Ravichandran), and RIKEN (M. Hatakeyama) are already in place.

- Development of methods and software for rule-based modeling, enhancing and extending capabilities of BioNetGen software, in collaboration with cell signaling team at Los Alamos.

- Development of data structures required for rule-based modeling and electronic exchange of models.

I anticipate that new and interesting directions of research will present themselves while I am at Indiana University, and I would hope to establish new collaborations with research teams, such as

- Collaboration with Systems Biology Group of Professor S. Schnell to model biochemical pathways;
- Developing advanced algorithms for modeling, analysis and storage of protein-protein interactions together with Sun Kim's lab and Filippo Menczer's lab;
- Applying mechanistic modeling to protein-protein interactions in GroEL-mediated protein folding and microbial communications together with Lingling Chen's Lab.

I look forward to the opportunity to interview for this position. Thank you for your time and consideration.

Sincerely,

Michael L. Blinov

Names of references:

- B. Goldstein, Laboratory Fellow, Theor. Biol., Los Alamos National Lab, bxg@lanl.gov
- W. S. Hlavacek, Staff Scientist, Theor. Biol., Los Alamos National Lab, wish@lanl.gov
- Y. Yomdin, Prof., Theor. Math, Weizmann Institute of Science, yomdin@wisdom.weizmann.ac.il
- J. R. Faeder, Staff Scientist, Theor. Biol., Los Alamos National Lab, faeder@lanl.gov
- K. Yusim, Research Scientist, Theor. Biol., Los Alamos National Lab, kysim@lanl.gov (teaching abilities)

Research objectives

Building and analysis of predictive mathematical models of signal transduction is a great challenge for cell biology and predictive science. This is the area I have worked in and plan to continue to work in.

My research is organized and will be organized around the following topics:

1. *Mechanistic modeling of epidermal growth factor receptor (EGFR) signaling*. The purpose of the work is to create the most detailed mechanistic model of EGF signaling. As a first step, I've created an extended model based on interactions described by Kholodenko et al. (1999) and used by Schoeberl et al. (2002) and several other research groups. The original model tracks the dynamics of 18 species that arise from interactions among five proteins: EGFR, the ligand EGF, the adapter proteins Grb2 and Shc, and the guanine nucleotide exchange factor Sos. However, many molecular species that are monitored experimentally are omitted from this model, e.g. different phosphoforms of EGFR, or complexes involving several adapter proteins bound to receptor-dimer. Indeed, the model is based on several implicit assumptions: 1) the only monomers of EGFR considered are those lacking cytoplasmic modifications; 2) dimers of EGFR may not be in direct contact with more than a single adapter protein; 3) phosphorylation and dephosphorylation is simultaneous on both receptors in a dimer and all domains within a receptor; 4) a single protein bound to a reeceptor protects all the sites in a dimer from dephosphorylation; and 5) phosphorylated dimers cannot break up. There are no data justifying these assumptions, and some data contradicts them. For example, Jiang and Sorkin (2002) have shown that the EGFR dimer may be in contact with more than one adapter protein. When we lift these assumptions, we obtain a model with 356 distinct chemical species and 3,749 reactions. This model generates a variety of new predictions, e.g. time courses of phosphorylation for different tyrosines, and suggests experiments that can be performed to test these predictions. This work was selected for an oral presentation at the international conference ICSB2004 in Heidelberg. While working on this project, I established collaboration with teams at Thomas Jefferson University (B. Kholodenko), University of Virginia (K. Ravichandran), and RIKEN in Japan (M. Hatakeyama). Letters of collaboration are available upon request. The goal of this project will be to implement a computationally-tractable comprehensive model for EGFR signaling, including all the details described in Oda et al. (2005) at the level of molecular domains. The next step would be to use network inference methods to obtain information about potential protein-protein interactions from publicly available proteomic data sets, to use for enhancing the mechanistic model of EGFR signaling and guide experimental research. *I plan to pursue this project as an independent investigator. My supervisors at LANL have indicated they are happy to let me assume management of this project. I'll attempt to obtain funding to proceed with this project in collaboration with experimental investigators (collaborations are already in place).*
2. *Development of advanced methodology, templates for different signaling systems, and enhancing capabilities of software for rule-based modeling of signal transduction based on the interactions of molecular domains*. Development of BioNetGen software has been my main task at Los Alamos National Laboratory. I have designed and implemented the first

version of BioNetGen that was publicly released (reported in first-authored publications in *Bioinformatics*, 2004 and *Complexity*, 2005; a patent application is filed by the Los Alamos Natl Lab) and used to model different signaling systems. In this approach, rules for protein-protein interactions are specified and then used as generators of species and reactions. I used this software to study initial steps in signaling by immune-receptors (coauthoring publications in *J. Immunology*, *Biotechn. Bioeng*. and *Syst. Biol*.), and developed test models for Toll-like receptor 4, interleukin 1, insulin, interferon gamma, and erythropoietin (Epo) receptor signaling. Currently, I'm developing advanced methods and software for rule-based modeling using graphs to specify biomolecular interactions (reported in *LNCS, 2005*), in order to explicitly and systematically represent the connectivity of components in multicomponent biomolecular complexes; and explicitly track and account for the internal structures of biomolecules. This progress allows me to model a large variety of different signaling systems, including systems where certain signaling complexes can form chains of different length. The goal of this project is to create a set of templates that allows the easy generation of a predictive model for almost any signaling system. Collaboration has been established with several teams, aimed at enhancing abilities and compatibility of BioNetGen with different software platforms (Virtual Cell, TeraNode, BioUML, and Gene Network Sciences). *Funding will be pursued for this project jointly with my team at LANL.*

3. *Development of data structures required for rule-based modeling and electronic exchange of models.* This is an extension of the previous project proposed to allow electronic exchange of models generated with BioNetGen software. Rules can not be encoded in Systems Biology Markup Language (SBML) due to several limitations of SBML in its current form, for example it carries no information about domains of proteins and composition of multi-protein species. I'm developing a way to incorporate BioNetGen data structures into SBML. I have established collaboration with the Caltech team working on SBML (M. Hucka). This work was reported at the 9th SBML Forum in Heidelberg and available online at http://www.sbml.org/wiki/SBML_Level3_Efforts. A prototypical database for storage of BioNetGen models is available at http://cellsignaling.lanl.gov/database . Based on this work, I plan to develop a software workbench that would incorporate software for rule-based modeling and a database for storing of biological information in the form of rules, as outlined in *LNCS* publication. This workbench would allow mining of public databases, incorporating the information about domains of biomolecules, their activities and interactions, conditions for reactions, and numerical information like rate constants and cellular concentrations where determined. The workbench will stimulate development of rule-based models for a wide range of biological signaling systems. The developed workbench will be compatible with software tools and standards developed by other teams. *The first application of this workbench would be the modeling of the EGFR system. I'll try to get funding to proceed with this project independently or in collaboration with LANL team.*

Teaching interests

I have significant teaching experience. For two years, 1999-2000, I was a lecturer at the Weizmann Institute of Science (one lecture per week) teaching "Mathematical Modeling in Biology", a unique graduate course designed by Professor Lee Segel for experimental biologists. I advanced to lecturing this course after serving for one year as a teaching assistant. Professor Lee Segel has recently passed away, but references can be obtained from Dr. K. Yusim (kyusim@lanl.gov), a former graduate student of Professor Segel, who taught this course before me. Appreciation of my teaching abilities can be judged by the fact that Professor Segel invited me to teach a course for a Summer School at the Santa Fe Institute in 2003. In 1998 I worked as a teaching assistant for a Computer Algebra Course at the Weizmann Institute of Science taught by Professor Nurit Zehavi (Nurit.Zehavi@weizmann.ac.il). In addition, I worked as a mentor for a Summer Science Institute at the Weizmann Institute of Science, and was promoted to the position of Coordinator of Mathematical Projects. In this position I interviewed, accessed and graded students. I also trained mentors, and tutored students on preparing scientific manuscripts.

As a well trained mathematician, I am fully capable of teaching all the standard mathematical courses to a broad range of students. In addition, as a scientist involved in biological research for the last five years, I can bring a unique combination of mathematical skills and biological knowledge to the teaching of interdisciplinary courses. Below is a sample list of courses that I can design and teach:

- Basic: Mathematical Modeling in Biology (biochemical kinetics, discrete networks of genes and cells, parameters evaluation, affordable ranges of parameters, fitting and sensitivity, etc)
- Basic: Mathematics for Biologists (understanding deterministic and stochastic simulations, probability, graph theory, Bayesian networks etc)
- Basic: Tutorial on Biological Software Packages
- Basic: Systems Biology (different approaches to modeling and analysis of biological systems, understanding predictions and their limitations, etc)
- Advanced: Modeling Receptor Signaling
- Advanced: Biochemical Network Analysis (basic reaction networks, metabolic and gene networks, mass and flux analysis, parameter sensitivity, network inference etc).