September 16, 2005

Prof. Dr. Yves Brun
Department of Biology
Indiana University
Bloomington IN 47405-7005
ybrun@bio.indiana.edu

Re: **Systems Biology/Microbiology Faculty Search**

Dear Dr. Brun:

This letter and attached *curriculum vita* are in application for the position of Faculty Position in Systems Biology/Microbiology at Indiana University advertised in the Science magazine. I am a principal investigator and senior research scientist II in Microbial Department and the Biomolecular Systems Initiative of the Pacific Northwest National Laboratory, a nationally recognized research institute on systems biology. I have a good training in molecular microbiology and more than 10 years' post-graduate research experiences on metabolic pathway engineering and signal transduction in various microbial systems. In the past five years my primary research focus has been on exploration of complex biological systems using systems biology approaches, such as transcriptomics, proteomic and metabolomics technologies. My other research efforts include comparative genomics, development of codon-based computational approach to predict gene expression level and development of statistical methods for integration and analysis of transcriptomic and proteomic data. I have authored or co-authored 30 scientific publications published in peer-reviewed scientific journals and 8 pending US patent applications. In addition, more than 10 manuscripts from past three years' research efforts were submitted or are currently in press. My research interests and expertise, therefore, would complement synergistically the systems biology effort at Indiana University.

A more detailed description of my recent research achievements and teaching interests is attached. Thank you for your consideration, and I look forward to discussing this opportunity with you further.

Very Sincerely yours,

Weiwen Zhang, Ph.D.

Principal Investigator
Senior Research Scientist II
Pacific Northwest National Laboratory
Battelle/US Department of Energy
P. O. Box 999, Mail Stop P7-50
Richland, Washington 99352, USA
Phone: (509)-376-2016; Fax: (509)-372-1632
Email: Weiwen.Zhang@pnl.gov
http://www.sysbio.org/resources/staff/zhang.stm

Summary of Past Research

My Ph.D. thesis work focused on metabolic regulation of methylmalonyl-CoA formation and aspartate pathways in rifamycin SV producing *Amycolatopsis mediterranei*. Rifamycin and its derive rifampin are the major drugs for clinical treatment of HIV-related tuberculosis. The aim of my project was to increase rifamycin SV production by modifying the rate-limiting precursor supply pathway using genetic engineering technology. Genes encoding methylmalonyl-CoA mutase, a major rate limiting step for precursor formation, and two key genes in the aspartate pathway, aspartokinase (*ask*) and aspartate semialdhyde dehydrogenase (*asd*) were cloned and characterized. The results showed that over-expression of *mut*AB gene can be used to increase the supply of the methylmalonyl-CoA pool. In addition, an eukaryotic-like serine/threonine protein kinase gene (*amk*) was cloned from *A. mediterranei*. By gene disruption analysis, we found that the *amk* gene is involved in the carbon source-dependent pigment formation. The study provided important evidence that this type of signal transduction systems was involved in secondary metabolism in *A. mediterranei*. The results of my research were summarized into ten research papers published in peer-reviewed journals, such as European Journal of Biochemistry, Archives of Microbiology, Gene, Applied Biochemistry and Biotechnology, Applied Biotechnology and Microbiology, and Biochemical and Biophysical Research Communication.

My first postdoctoral training was in Professor J.A.Robinson's group at the University of Zurich, Switzerland. My work in his laboratory focused on the cloning and analysis of vancomycin biosynthetic gene cluster, and its potential use to generate new antibiotics. A cosmid library for the vancomycin producing *A. orientalis* DSM 40040 strain was constructed and the vancomycin biosynthetic cluster was cloned and sequenced. In the biosynthetic cluster, three P450 oxygenase genes were identified. Gene-inactivation studies pointed to their involvement in catalyzing the oxidative phenol coupling reaction during glycopeptide antibiotic biosynthesis. To get more insight on the catalytic mechanism, one of the P450 genes, *oxyB*, was expressed and purified as hemoprotein from *Escherichia coli*. The protein was crystallized, and its structure determined to 1.7-A resolution. OxyB exhibits the typical P450-fold, with the high structural similarity to P450nor, P450terp, and P450eryF. This work was published in Journal of Biological Chemistry.

In January 1999, I started my second postdoc in Professor K.A.Reynolds's lab in Medical College of Virginia. My research in Virginia focused on the precursor supply pathways for monensin A antibiotic biosynthesis in *Streptomyces cinnamonensis*. By gene disruption, gene over-expression and C^{13} labeling studies, I demonstrated the presence of a previously unknown pathway for methylmalonyl-CoA formation in *S. cinnamonensis*. A novel B_{12} coenzyme–dependent mutase, *mea*A, is the central part of this novel pathway and critical for the monensin production under the tested conditions. This work was published in Journal of Bacteriology.

In January 2001, I started to work as a staff scientist in Microbial Research Department of Paradigm Genetics, Inc.. During the period of 2001-2002 I was involved in several projects of investigating various complex biological systems (*Corynebacterium*, yeast and filamentous fungi) using metabolomic and genomic (microarray) approach. One of my major efforts was to help develop novel software system and database for validating target gene discovery and metabolic pathway analyses based on metabolic profiles (metabolomics) and gene expression profiling data. The software, named as MetaTrace™, was built around Lion bioscience AG's pathSCOUT™, which contains a reference database of biochemical pathways, enzymes, and links to other relevant data, together with facilities for efficiently querying and displaying the data together with dense, validated empirical data around specific metabolite networks and a predictive statistical model of metabolism. The yeast lipid biosynthetic pathway was chosen as a experimental model for this software. As the principal biologist in a interdiscipline team, I was responsible for all the experimental planning for metabolomic and microarray analysis, data analyses and presentation. The achievements from this project were summarized into eight filed US patents.

Summary of Selective Topics of My Recent Research

1. Exploration of complex microbial systems and microbial interaction using systems biology approach

The multitude of genome projects in the past decade has generated vast amount of DNA sequence information and, at the same time, highlighted our lack of understanding of how cells function. My research interests are on the elucidation of the function, interaction and regulation of metabolic pathways using modern genetic and chemical tools. Because of the complexity of metabolic network as a living system, the global metabolism is not predictable from the complete description of each individual pathway in the network, approaches need to be implemented that focus on the systemic properties of the network. Microorganisms in their ecological niches in nature rarely exist alone. Instead, they participate in a web of chemical reactions that involve many other microbes. One of the most important types of interactions is termed syntrophy, which translates as "feeding together". In the past three years, we have been studying the *Desulfovibrio vulgaris-Methanosarcina barkeri* syntrophic model system using whole-genome DNA microarrays and proteomics approach. The goal of this project is to identify genome-wide metabolic networks in a two-microorganism syntrophic relationship and to understand how the syntrophic partnership modulates the partners' expression and biochemical machinery compared to those of the pure cultures. Metabolic networks have been determined for each organism and are being compared to the networks present when they are grown in a syntrophic co-culture. Batch and continuous cultures were studied. By analyzing and modeling this syntrophic relationship at the molecular genetic level will enable future understanding of other microbial interactions in complex microbial communities. Eight manuscripts have been so far finished on the response of pure cultures and bioreactor community to various environmental stress treatments, including the environmental conditions expected to be encountered during syntrophic growth using microarray and proteomic analysis.

2. Development of statistical model to integrate different types of genomic data

2.1. Integration of microarray and proteomic data: Parallel profiling of mRNA and protein on a global scale and integrative analysis of these two data types could provide additional insight into the metabolic mechanisms underlying complex biological systems. However, because protein abundance and mRNA expression are affected by many cellular and physical processes, there have been conflicting results on the correlation of these two measurements. In this study, we describe a novel data-driven statistical model to integrate whole-genome microarray and proteomic data collected from *Desulfovibrio vulgaris* grown under three different conditions. Based on the Poisson distribution pattern of proteomic data and the fact that a large number of proteins were undetected (excess zeros), Zero-inflated Poisson models were used to define the correlation pattern of mRNA and protein abundance. The models assumed that there is a probability mass at zero representing some of the undetected proteins because of technical limitations. The models thus use abundance measurements of transcripts and proteins experimentally detected as input to generate predictions of protein abundances as output for all genes in the genome. We demonstrated the statistical models by comparatively analyzing *D. vulgaris* grown on lactate-based versus formate-based media. The increased expressions of Ech hydrogenase and alcohol dehydrogenase (Adh)-periplasmic Fe-only hydrogenase (Hyd) pathway for ATP synthesis were predicted for *D. vulgaris* grown on formate. (Nie L, Wu G, Brockman F, & Zhang W, 2005, Bioinformatics, Submitted)

2.2. Relation between mRNA expression and sequence information in *Desulfovibrio vulgaris*: The context-dependent expression of genes is the core for biological activities, and significant attention has been given to identification of various factors contributing to gene expression at genomic scale. However, so far this type of analysis has been focused either on relation between mRNA expression and non-coding sequence features such as upstream regulatory motifs or on correlation between mRNA abundance and non-random features in coding sequences (*e.g.* codon usage and amino acid usage). In this study multiple regression analyses of the mRNA abundance and all sequence information in *Desulfovibrio vulgaris* were performed, with the goal to investigate how much coding and non-coding sequence features contribute to the variations in mRNA expression, and in what manner they act together. Using the AlignACE program, 442 over-represented motifs were identified from the upstream 100 bp region of 293 genes located in the known regulons. Regression of mRNA expression data against the measures of coding and non-coding

sequence features indicated that 54.1% of the variations in mRNA abundance can be explained by the presence of upstream motifs, while coding sequences alone contribute to 29.7% of the variations in mRNA abundance. Interestingly, most of contribution from coding sequences is overlapping with that from upstream motifs; thereby a total of 60.3% of the variations in mRNA abundance can be explained if both coding and non-coding information was included. This result demonstrates that upstream regulatory motifs and coding sequence information contribute to the overall mRNA expression in a *combinatorial* rather than an *additive* manner. In addition, the results suggest that by including coding sequence information the quantitative prediction of mRNA expression can be improved.
(Wu G, Nie L, & Zhang W, 2005, BMC Genomics, In press)

3. Regulation of microbial metabolism: protein phosphatases

Originally identified from eukaryotes, the Mg2+- or Mn2+-dependent protein phosphatases (PPMs) are a diverse group of enzymes whose members include eukaryotic PP2C and some prokaryotic serine/threonine phosphatases. We have identified unexpectedly large numbers of PPMs present in two *Streptomyces* genomes. Phylogenetic analysis was performed with all the PPMs available from a wide variety of microbial sources to determine the evolutionary origin of the *Streptomyces* PPM proteins. Consistent with earlier hypotheses, the results suggested that the microbial PPMs were relatively recent additions from eukaryotic sources. Results also indicated that the *Streptomyces* PPMs were divided into two major subfamilies at an early stage of their emergence in Streptomyces genomes. The first subfamily, which contains only six *Streptomyces* PPMs, possesses a catalytic domain whose sequence and architecture are similar to that of eukaryotic PPMs; the second subfamily contains 89 *Streptomyces* PPMs that lack the 5a and 5b catalytic domain motifs, similar to the PPMs SpoIIE and RsbU of *Bacillus subtilis*. Significant gene duplication was observed for the PPMs in the second subfamily. In addition, more than half (54 %) of the *Streptomyces* PPMs from the second subfamily were found to have at least one additional sensory domain, most commonly the PAS or the GAF domain. Phylogenetic analysis showed that these domains tended to be clustered according to the putative physiological functions rather than taxonomic relationship, implying that they might have arisen as a result of domain recruitment in a late evolutionary stage. This study provides an insight into how *Streptomyces* spp. may have expanded their PPM-based signal transduction networks to enable them to respond to a greater range of environmental changes.
(Zhang W & Shi L, 2004, Microbiology, 150, 4189-4197;
Shi L & Zhang W, 2004, Microbiology, 150, 2247-2256)

4. Predicted highly expressed genes in the *Streptomyces* genomes and the implications for their metabolism.

Highly expressed genes in bacteria often have a stronger codon bias than genes expressed at lower levels, due to translational selection. In this study, a comparative analysis of predicted highly expressed (PHX) genes in the *Streptomyces coelicolor* and *Streptomyces avermitilis* genomes was performed using the codon adaptation index (CAI) as a numerical estimator of gene expression level. Although it has been suggested that there is little heterogeneity in codon usage in G+C-rich bacteria, considerable heterogeneity was found among genes in these two G+C-rich Streptomyces genomes. Using ribosomal protein genes as references, approximately 10% of the genes were predicted to be PHX genes using a CAI cutoff value of greater than 0.78 and 0.75 in *S. coelicolor* and *S. avermitilis*, respectively. The PHX genes showed good agreement with the experimental data on expression levels obtained from proteomic analysis by previous workers. Among 724 and 730 PHX genes identified from *S. coelicolor* and *S. avermitilis*, 368 are orthologue genes present in both genomes, which were mostly 'housekeeping' genes involved in cell growth. In addition, 61 orthologous gene pairs with unknown functions were identified as PHX. Only one polyketide synthase gene from each *Streptomyces* genome was predicted as PHX. Nevertheless, several key genes responsible for producing precursors for secondary metabolites, such as crotonyl-CoA reductase and propionyl-CoA carboxylase, and genes necessary for initiation of secondary metabolism, such as adenosylmethionine synthetase, were among the PHX genes in the two *Streptomyces* species. The PHX genes exclusive to each genome, and what they imply regarding cellular metabolism, are also discussed.
(Wu G, Nie L, & Zhang W, 2005, Microbiology, 151:2175-2187)

Future Research Plans

In the future, I would like to continue persuade active research in the following areas:

1. Understand the metabolic and regulatory networks of selective microbial systems at the genome level by systems biology approach

Exploration of complex biological systems using systems biology approaches will be my major focus in the future. In addition to various high through-put microarray, proteomic and metabolomic technologies which have been used in several of my current projects for studying metabolic and regulatory networks, I also have interests in studying protein-protein interaction network and cellular imaging work in microbes of interests. I have two projects recently funded by DOE LDRD program working in these areas. See below for details of my ongoing research in these areas.

A) Type III Secretion Systems and Their Roles in Bacterial Virulence Process: Secretion of proteins into the extracellular environment is important to most bacteria, and in particular mediates interactions between pathogenic bacteria and their eukaryotic hosts. Early studies of *Salmonella* entry into host cells showed that two type-III secretion systems (TTSS), SPI-1 (*Salmonella* pathogenicity island 1) and SPI-2 (*Salmonella* pathogenicity island 2) systems, are involved in mediation of enterocyte invasion and survival within macrophages, respectively. Although many candidate effector proteins have been associated with possible delivery through the Salmonella TTSS, information regarding interaction of these proteins with the eukaryotic hosts is still scarce. In this study, we are using the lambda general recombination system (?-Red) to tag the effector proteins with a tetracysteine motif, to provide specific binding sites for FlAsH-derivatized resin and tetracysteine-tagged protein. The immobilized FlAsH-tagged proteins and the interacting host cell proteins can then be purified for mass spectrometric analysis. Using this high through-put approach, we will develop a method to study the protein-protein interactions between mouse macrophages and *S. typhimurium*. The study will provide exciting insights into the mechanisms of *Salmonella* virulence.

B) Dynamics and Spatial Expression of Signal Proteins in the *Desulfovibrio vulgaris* Biofilm and Its Implication to Iron Corrosion: Corrosion of iron presents a serious economic problem. Whereas aerobic corrosion is a chemical process, anaerobic corrosion is frequently linked to the activity of sulphate-reducing bacteria (SRB). These bacteria often form biofilm on surface of steel. However, the underlying mechanisms of biofilm formation are complex and insufficiently understood. Recent studies showed that biofilm formation is controlled by two-component signal transduction systems in many bacteria. In our previous study, a large number of two-component signal transduction systems (TCSTS) including 59 putative sensory histidine kinases (HK) and 55 response regulators (RR) were identified from the *D. vulgaris* genome, indicating their important roles in regulation of cellular metabolism. In this study, we will construct reporter proteins that consist of the fusion of the promoter of selective two-component signal genes to *gfp* gene encoding a green fluorescent protein, and study their expressions following the formation of *D. vulgaris* biofilm grown on an iron metal surface. The experiments will illustrate for the first time the dynamic and spatial expression of two-component signal proteins in *D. vulgaris* biofilms growing in on metal surface. The efforts will facilitate investigations of the complex regulatory networks controlling biofilm formation and iron corrosion in natural environments.

2. Regulation of microbial secondary metabolism by protein phosphatases

Compared to two-component signal transduction systems, the signal transduction mediated by protein kinases and protein phosphatases and their roles in various secondary metabolites-producing Actinomycetes are not well-studied. Recently we have identified unexpectedly large numbers of the Mg2+- or Mn2+-dependent protein phosphatases (PPMs) present in two *Streptomyces* genomes (55 each genome). Phylogenetic analysis showed that they might have arisen as a result of gene duplication and domain recruitment. The existence of such large numbers of PPMs in streptomycetes genome clearly implies that they might play very important physiological roles. Thus the elucidation of their functions will certainly benefit our understanding of the complex regulatory networks of secondary metabolites

production and present intelligent guidance to our efforts of increasing polyketide titers and producing novel activities, through metabolic engineering technologies. Currently there is almost no research going on protein phosphatases of *Streptomyces*, I would like to spend some efforts in this area.

3. Develop computational methods for integration of high through-put genomic and proteomic data

DNA microarrays are now a popular platform for measuring changes in messenger RNA transcript levels on a genome-wide scale, while gel-free shotgun profiling methods (LC-MS/MS) based on tandem mass spectrometry are increasingly being used to determine the identity, modification states, and relative abundance of large numbers of proteins. While comparison of the complementary information contained in proteomic and mRNA data sets poses considerable analytical challenges, these efforts should provide added insight into the fundamental mechanisms underlying physiology, development, and the emergence of disease. I have established close collaboration with Prof. Lei Nie of Department of Biostatistics, Biomathematics, and Bioinformatics of Georgetown University. The goal of our collaboration is to develop statistical methods to compare heterogeneous databases and identify correlation patterns between proteomic and microarray data. In the past year, we have made some encouraging progresses, and one manuscript was submitted and another is in preparation. Currently Prof. Nie and I are working on a NIH proposal on this topic.

Statement of Teaching Interests

Teaching has been an integral part of my graduate training and scientific career. I have a strong commitment to share my knowledge and understanding of science to young people with aspirations to embark upon a scientific career. My teaching philosophy is simple: students are always the first priority.

I believe teaching and training graduate students is one of the critical responsibilities as a faculty member. Students need to be mentored so that they can develop into professionally successful people. It is important to spend time with them, help them to acquire professional skills, and truly care about them and their future. The advance of our discipline of science depends largely on the availability of motivated and well-trained young individuals. I have been involved in training and supervising junior scientists to work on specific projects. I am always excited when I see them learn under guidance and then achieve results in their own development.

I am also eager to interact with students in classrooms. I am interested in teaching fundamental classes in microbiology, microbial physiology, microbial genomics or metabolic engineering. This is an important step to stimulate the interests of young people in the disciplines of microbiology and to provide them the background needed to pursue further training in more specialized areas. Students will be encouraged to participate in an interactive learning environment. In addition, I will explore the possibilities of teaching an advanced course in my area of expertise, such as the biosynthesis and regulation of microbial secondary metabolites, and microbial metabolic engineering. My experience as a teaching assistant was both challenging and enjoyable. I was responsible for giving introductory lectures and providing technical guidance to undergraduate students in laboratory classes.

I look forward to the opportunities of teaching at different levels in an academic environment. I am confident that I have the communication skills, knowledge, and enthusiasm necessary to be an effective teacher.