

# 3<sup>rd</sup> ANNUAL INSTITUTE FOR GENOMICS AND BIOINFORMATICS (IGB) BIOMEDICAL INFORMATICS TRAINING (BIT) PROGRAM SYMPOSIUM

Tamkin Lecture Hall - Room F114

Tuesday, May 3<sup>rd</sup>, 2005

## Opening

9:00-9:05 Opening Remarks. **G.W. Hatfield, P. Baldi**

## Morning Session I: The Genome

9:05-9:20 Population dynamics of transposable element insertions in *Arabidopsis thaliana* and *Arabidopsis lyrata*. **Jesse Hollister**, Brandon Gaut, Richard H. Lathrop

9:20-9:35 Naturally Occurring Cryptic Variation in Drosophila Notch Signaling Pathway. **Farida Nissan**, Anthony D. Long, and Pierre F. Baldi

9:35-9:50 Loss of diversity in combinatorial oligonucleotide-directed mutagenesis. **Phillip Tam, Christopher Wassman**, Greg Weiss, and Richard H. Lathrop

9:50-10:05 Bioinformatic identification and functional confirmation of  $\sigma^{28}$ -regulated genes in *Chlamydia*. **Hilda H.Y. Yu**, Sean Lee, Dennis Kibler, and Ming Tan

10:05-10:20 Global landscape of recent inferred Darwinian selection for *Homo sapiens*. **Eric T. Wang**, Greg Kodama, Pierre Baldi, and Robert K. Moyzis

10:20-10:35 Mitomaster: Mitochondrial Data Processor. **Marty C. Brandon**, Pierre F. Baldi, Marie T. Lott, Douglas C. Wallace

10:35-10:55 Coffee Break in the Lobby

## Morning Session II: The Transcriptome

11:00-11:15 The mouse hair growth cycle as a model for studying cyclic biological processes. **Kevin K. Lin**, Darya Chudova, Padhraic Smyth, Bogi Andersen

11:15-11:30 Targeting cis-regulatory variation through differential allelic expression in *Drosophila melanogaster*. **Jonathan D. Gruber**, Klemens J. Hertel, Anthony D. Long

11:30-11:45 Identification of novel transcriptional regulators in *Chlamydia*. **Johnny Akers**, Dennis Kibler, Ming Tan

11:45-12:00 Evolutionary changes in gene expression during temperature adaptation in *E. coli*. **Nancy Maria Aguilar-Roca**, Al Bennett, and Anthony D. Long

12:00-12:15 The global role of DNA supercoiling in the regulation of basal level gene expression in *Escherichia coli*. **Kimberly A. Aeling**, Padhraic J. Smyth, G. Wesley Hatfield

12:15-1:00 Lunch Break in F108

### **Afternoon Session I: The Proteome**

1:00-1:15 De Novo Design of a Folded Peptide. **R. Jeremy Woods**, **Arlo Randall**, James S. Nowick, and Pierre F. Baldi

1:15-1:30 Predicting the effects of Amino Acid Mutations on DNA Mismatch Repair Protein Function. **Bob Chan**, Steven Lipkin, and Dennis Kibler

1:30-1:45 TY3 virus-like particle assembly. **Liza Z. Larsen**, Suzanne Sandmeyer, Min Zhang, Alex McPherson, Yurii G. Kuznetsov, Richard H. Lathrop, G. Wesley Hatfield

1:45-2:00 Chemical informatics and computational docking. **S. Joshua Swamidass**, Jonathan Chen, Yimeng Duo, Jocelyne Bruand, Hartmut Luecke, Pierre Baldi

2:00-2:15 Coffee Break in F108

### **Afternoon Session II: The Molecular Modeling and Systems Biology**

2:15-2:30 Prediction of protein disordered regions using machine learning Techniques. **Michael J Sweredoski**, Jianlin Cheng, Pierre F Baldi

2:30-2:45 Predicting mutant protein function using computer models. **Samuel A. Danziger**, Richard Lathrop, Rainer Brachmann

2:45-3:00 Hidden Markov modeling of JNK docking sites. **David Ho**, **Ryan W. Benz**, Pierre Baldi and Lee Bardwell

3:00-3:15 The generalized Monod, Wyman, Changeux model for mathematical modeling of metabolic enzymes with allosteric regulation. **Tarek S. Najdi**, Chin-Rang, Yang, Bruce E. Shapiro, G. Wesley Hatfield and Eric D. Mjolsness

3:15-3:30 Stochastic models for simulation of genetic regulatory mechanisms. **Chin-Rang Yang**, Bruce, E. Shapiro, Eric D. Mjolsness, Pierre F. Baldi and G. Wesley Hatfield

3:30-3:45 Biomechanical models of *Arabidopsis* meristems. P. Baldi, E. Mjolsness, **A. Sadovsky**, M. Heisler, T. Bacarian, E. Meyerowitz

### **Closing**

3:45-4:00 Closing Remarks. **G.W. Hatfield**, **P. Baldi**

# 3<sup>rd</sup> ANNUAL INSTITUTE FOR GENOMICS AND BIOINFORMATICS (IGB) BIOMEDICAL INFORMATICS TRAINING (BIT) PROGRAM SYMPOSIUM

Tamkin Lecture Hall - Room F114

Tuesday, May 3<sup>rd</sup>, 2005

Opening Remarks. **G.W. Hatfield, P. Baldi**

## Morning Session I: The Genome

### Population dynamics of transposable element insertions in *Arabidopsis thaliana* and *Arabidopsis lyrata*.

**Jesse Hollister**, Brandon Gaut, and Richard H. Lathrop

Transposable elements (TEs) are mobile DNA sequences found in virtually all prokaryotic and eukaryotic genomes. Population genetics theory predicts that mating system should be an important determinant of the dynamics of TE insertions in plant populations due to differences in selective forces on the elements and their hosts. In addition, selective forces on different families of TEs may vary, due to different site specificities and modes of transposition. In this study, Transposon display, an AFLP-based technique, was used to examine insertion polymorphism and abundances of several families of TEs in natural populations of the selfing plant *Arabidopsis thaliana* and its outcrossing relative *Arabidopsis lyrata*. These data were used to estimate the strength of selection against individual insertions in both species.

### Naturally occurring cryptic variation in *Drosophila* Notch signaling pathway

**Farida Nissan**, Anthony D. Long, and Pierre F. Baldi

Notch signaling pathway is involved in a developmental process termed lateral inhibition, the signaling out of one cell from a cell cluster for a given fate along with the inhibition of the remaining cells from assuming that fate. Lateral inhibition is one of a dozen or so major development processes used in multicellular animals. In humans many diseases result from aberrant Notch signaling (e.g. leukemia, and congenital syndromes associated with stroke and dementia). Lateral inhibition is a common theme in *Drosophila* development, being involved in cell fate decisions in several tissues. For instance, Notch signaling determines the dorsal-ventral boundary of the wing, as well as the patterning and spacing of bristles on the exoskeleton.

The objective of my work is the creation of a novel quantitative character sensitive to naturally occurring variation in Notch signaling. The character is displayed as variation in the level of "notching" of the *Drosophila* wing boundary. The idea central to the project is that by crossing a Notch loss of function mutant that produces a notched wing boundary to a set of wild inbred lines we can create a QTL-mapping population expressing phenotypic variation in the severity of the "notching" wing phenotype. The "notch" quantification will be carried out by processing the wing images and by developing a morphometric based computer program that not only provides shape descriptors but that is also able to calculate the "notched area". Genotyping SNPs in genes known to be part of the Notch signaling pathway (e.g., Delta, Suppressor of Hairless, Hairless, scabrous, groucho, etc.) will allow us to identify which of these genes harbor natural variation contributing to variation in Notch signaling. Identifying naturally occurring common variants that subtly affect Notch signaling is an important exercise, as these types of variants

may contribute to the severity of Mendelian diseases associated with loss-of-function mutations of large effect, as well as other more complex polygenic disorders whose genetic bases are poorly characterized.

### **Loss of diversity in combinatorial oligonucleotide-directed mutagenesis.**

**Phillip Tam, Christopher Wassman, Greg Weiss, and Richard H. Lathrop**

Protein engineering uses oligonucleotide-directed mutagenesis to modify DNA sequences through a two-step process of hybridization and enzymatic synthesis. Inefficient reactions confound attempts to introduce mutations, especially for the construction of vast combinatorial protein libraries. We applied computational approaches to the problem of inefficient mutagenesis. Several results implicated oligonucleotide annealing to non-target sites, termed “cross-hybridization”, as a significant contributor to mutagenesis reaction failures. Test oligonucleotides demonstrated control over reaction outcomes. A novel cross-hybridization score, quickly computable for any plasmid and oligonucleotide mixture, directly correlated with yields of deleterious mutagenesis side products. Cross-hybridization was confirmed conclusively by partial incorporation of an oligonucleotide at a predicted cross-hybridization site, and by modification of putative template secondary structure to control cross-hybridization. Even in low concentrations, cross-hybridizing species in mixtures poisoned reactions. Cross-hybridization in complex oligonucleotide mixtures could cause lost diversity in the cognate protein libraries. We have begun to investigate the nature of lost diversity using a novel set-partitioning algorithm in conjunction with microarray analysis.

### **Bioinformatic identification and functional confirmation of $\sigma^{28}$ -regulated genes in *Chlamydia*.**

**Hilda H.Y. Yu, Sean Lee, Dennis Kibler, and Ming Tan**

We have previously demonstrated that an alternative RNA polymerase containing  $\sigma^{28}$  is active in *Chlamydia*, and transcribes a developmentally regulated gene, *hctB*. To identify additional  $\sigma^{28}$ -regulated genes, we have developed a computer algorithm that generates a weighted probability matrix based on known  $\sigma^{28}$  promoter sequences in other bacteria.

Using this training set, we have identified candidate  $\sigma^{28}$  promoters in *C. trachomatis* and tested them with our  $\sigma^{28}$  *in vitro* transcription assay. 3 out of 10 promoters tested were transcriptionally active, and  $\sigma^{28}$ -dependence was shown by specific inhibition with anti- $\sigma^{28}$  antibodies.

To improve the predictive value of the algorithm for *Chlamydia*-specific promoters, we will define the optimal promoter sequences recognized by chlamydial  $\sigma^{28}$  RNA polymerase. Using a mutational approach, we will assay the effect of each single base substitution in the *hctB* promoter on promoter activity *in vitro*. The ensuing results will provide us with a probability weighting specific to *Chlamydia* that comprises the relative contribution of each base at each position to promoter activity. This combined bioinformatics and functional approach has, thus far, allowed us to identify  $\sigma^{28}$ -regulated genes and provided us with insights into the role of  $\sigma^{28}$  RNA polymerase in regulating gene expression in *Chlamydia*.

### **Global landscape of recent inferred Darwinian selection for *Homo sapiens*.**

**Eric T. Wang, Greg Kodama, Pierre Baldi, and Robert K. Moyzis**

Using the International Human Haplotype Map (HapMap) dataset, a probabilistic search for the landscape exhibited by positive Darwinian selection was conducted. By computationally sorting each high frequency allele by genotype, the sigmoidal decay of adjacent SNP linkage disequilibrium (LD) was directly calculated, eliminating the need for inferring haplotype. Comparing all 1,073,663 HapMap SNPs (Phase I Freeze) to the sigmoidal pattern found at known selected alleles, such as human *G6PD* V202M,

1.7% of HapMap SNPs were found to exhibit the genetic architecture of selection. Simulation studies indicate that this novel approach effectively distinguishes selection from other causes of extensive LD, such as population bottlenecks and admixture. Each highly unusual site was mapped to its corresponding gene (4.9% of annotated genes) and clustered according to Gene-Ontology (GO) categories. Based on over-representation analysis, several predominant biological themes are common in these selected alleles, including host-pathogen interactions, reproduction, DNA metabolism/cell cycle, protein metabolism, and neuronal function.

### **Mitomaster: Mitochondrial data processor.**

**Marty C. Brandon**, Pierre F. Baldi, Marie T. Lott, Douglas C. Wallace

Mitochondrial DNA has desirable properties for use in both phylogenetics and forensics. Many clinical variations in the mitochondrial genome have also been discovered, making it important for diagnosticians. However, specialized knowledge of mitochondrial genomics is needed when performing these analyses. Mitomaster is a system that integrates information about the mitochondrial genome with computational algorithms, and provides this analytical functionality through a web interface. The goal is to provide a powerful tool that is also easy to use.

### **Coffee Break in the Lobby**

## **Morning Session II: The Transcriptome**

### **The mouse hair growth cycle as a model for studying cyclic biological processes.**

**Kevin K. Lin**, Darya Chudova, Padhraic Smyth, Bogi Andersen

Understanding the regulation of hair follicle cycling is of great interest because aberrant regulation of hair cycle control genes is responsible for several types of abnormal hair loss and skin cancers. We recently reported the identification of hair-cycle associated genes from time-course gene expression profile data of the complex skin tissue by using a computational approach. However, some of these genes may only be playing a key role in hair follicle morphogenesis and not necessarily involved in hair follicle cycling. Hence, we refined this list of hair-cycle associated genes by performing microarray analysis (Affymetrix Mouse Genome 430 2.0 arrays) at representative time points during the second hair growth cycle. A more accurate estimation of the actual expression profiles was achieved by increasing the sampling rate of the time-course process; selecting nine representative time points in profiling the second hair growth cycle instead of five time points when we profiled the first cycle. Using a mixture model clustering algorithm developed by our group, we clustered the gene expression profile data of the second hair growth cycle for those probe sets that correspond to genes identified to be hair cycle-associated in the first cycle. Comparison of the clusters for the first and second cycles revealed striking differences in gene expression patterns for a set of genes that are involved in the initial hair follicle morphogenesis but are not associated with the recurrent hair growth cycles. For those genes that belong to the same clusters for the first and second cycles, we used quantitative real-time PCR to validate the temporal expression pattern over the two cycles. Future investigations of this refined list of hair-cycle associated genes would lead to a better molecular understanding of the mechanisms regulating hair follicle cycling.

### **Targeting cis-regulatory variation through differential allelic expression in *Drosophila melanogaster*.**

**Jonathan D. Gruber**, Klemens J. Hertel, Anthony D. Long

Variation in gene expression has long been suspected as an important contributor to evolutionary differences between species, differences between disease and non-disease states, and other quantitative traits. Under a standard interpretation of gene expression, variation can be partitioned into differences in the net activity of *trans*-regulatory molecules or DNA sequence differences in *cis*-regulatory motifs. Our technique masks the former, allowing us to focus on functional polymorphisms in *cis*-regulatory motifs. Using heterozygotes at coding-sequence Single Nucleotide Polymorphisms (SNPs), the ratio of these two alleles serves as a proxy for differential allelic expression, in turn implying *cis*-regulatory genetic variation. Genes on the X chromosome have been sequenced in search of coding-sequence SNPs in the correct population orientation; more than 25% of these have an appropriate polymorphism. Preliminary data demonstrating the measurement of allelic expression in test-case genes is presented. When we infer that a gene has *cis*-regulatory variation, discovering the regulatory polymorphism will increase our understanding of the mechanisms of gene expression as well as the nature of genetic variation for this phenotype within the species.

### **Identification of novel transcriptional regulators in *Chlamydia*.**

**Johnny Akers**, Dennis Kibler, Ming Tan

The pathogenic bacterium *Chlamydia* is an obligate intracellular parasite whose genes are coordinately expressed during the developmental cycle. Although there is evidence of transcriptional regulation in *Chlamydia*, the mechanisms have not been well defined. In particular, very few transcription factors have been predicted, and only a handful have been studied for function. We have developed an approach to identify unrecognized transcription factors in *Chlamydia* by first identifying *cis*-acting DNA regulatory elements that serve as binding sites for the transcription factors. The algorithm searches the chlamydial genomes for sequences containing characteristics of transcription factor binding sites such as inverted or direct DNA repeats. We hypothesize that conserved DNA motifs located in intergenic regions are likely to serve as binding sites for transcription factors. We will use selected DNA motifs as the ‘bait’ for directly purifying candidate transcription factors on the basis of specific binding to each DNA sequence. These DNA binding proteins will be further tested *in vitro* with functional assays to determine if they are transcription factors that can regulate promoter activity.

### **Evolutionary changes in gene expression during temperature adaptation in *E. coli*.**

**Nancy Maria Aguilar-Roca**, Al Bennett, and Anthony D. Long

Changes in gene expression during acute temperature changes have been studied extensively in *Escherichia coli*, but evolutionary changes in gene expression are not well characterized. High-density genomic arrays are currently under construction to test hypotheses about the role of specific genes in experimentally evolved lines of *E. coli*. Six lines of *E. coli* evolved for 2000 generations in a fluctuating thermal environment, thus both cold and heat-induced genes have the potential to play a role in the physiological adaptations previously documented for these lines. A key component of this study will be the application of Bayesian statistics to the gene expression data to account for the high levels of variation between a relatively small number of replicates. This combination of physiological, molecular and computational methods will demonstrate whether genes for the acute acclimation response are important over longer evolutionary time periods, and if changes in gene expression evolve in parallel in separate lines of *E. coli*.

### **The global role of DNA supercoiling in the regulation of basal level gene expression in *Escherichia coli*.**

**Kimberly A. Aeling**, Padhraic J. Smyth, G. Wesley Hatfield

*Escherichia coli* is adept for survival in a wide range of environments, varying in the availability of nutrients, temperature, pH, osmotic stress, and even oxygen levels. In order to accomplish this feat, it must detect and respond to environmental stimuli promptly and in an efficient manner specific to its current condition. Oftentimes, changes in environmental settings lead to temporary shifts in the cellular energy levels, which in turn affect the activities of enzymes that utilize ATP. These changes in cellular energy levels that accompany environmental shifts are correlated with shifts in DNA supercoiling levels. This indirect link between environmental stimuli and DNA supercoiling levels can act as a sensor of changing conditions. Most genes have an optimal supercoiling level for expression, and thus changes in DNA supercoiling may lead to global changes in basal levels of gene expression.

We hypothesize that supercoiling-dependent mechanisms serve to coordinate the expression levels of large numbers of genes during stress or during transitions from one growth condition to another, and therefore may be a global gene regulation mechanism. To test this hypothesis, our specific aims are: 1. To obtain gene expression profiles and measure the negative DNA supercoiling levels and energy charge of otherwise isogenic  $\pm$ IHF,  $\pm$ ArcA, and  $\pm$ Fnr *E. coli* K12 strains during aerobic to anaerobic growth transitions. 2. To identify new genes (operons) regulated by IHF, DNA supercoiling, and/or ArcA and/or Fnr. 3. To perform *in vitro* and *in vivo* studies to elucidate mechanistic interactions between a chromosome restructuring protein, IHF, and the global (regulon-specific) regulatory proteins, ArcA and Fnr.

## **Lunch Break in F108**

### **Afternoon Session I: The Proteome**

#### ***De novo* design of a folded peptide.**

**R. Jeremy Woods, Arlo Randall, James S. Nowick, and Pierre F. Baldi**

This paper describes the *de novo* design of a structured 39-residue peptide through a combination of bioinformatics and molecular modeling techniques. The peptide, which was designed with the eventual goal of making an intermolecular  $\beta$ -sheet dimerizer, includes three  $\beta$ -strands, an  $\alpha$ -helix, a  $\delta$ -ornithine  $\beta$ -turn unit, and a tryptophan spectroscopic probe. Structural optimizations, including Monte Carlo conformational searches and dynamics simulations, were used together with sequence optimizations to generate improvements in the stability of the structure. Sequence optimizations were made through the use of simulated annealing in a search for the sequence which is most likely to fold into the target structure. The move set consists of random mutations to the sequence under a set of sequence constraints. The objective function is  $P(\text{structure}|\text{sequence})$ , which is estimated by maximizing the probability that the predicted secondary structure (SS<sub>pro</sub>) and residue contacts (CMAP<sub>pro</sub>) from the sequence match those of the target structure. Synthesis and preliminary characterization of the peptide will be discussed.

#### **Predicting the effects of Amino Acid Mutations on DNA Mismatch Repair Protein Function.**

**Bob Chan, Steven Lipkin, and Dennis Kibler**

According to the Centers for Disease Control, colorectal cancer is the second leading cause of cancer-related deaths in the United States. Colorectal cancer (CRC) does have a significant genetic component, as first-degree relatives of CRC patients have a higher chance of being afflicted with CRC. However, the majority of cases do not exhibit a clear Mendelian inheritance pattern. This is most likely due to a large number of low-penetrance variants. Therefore, we have developed a method of predicting the effect of single nonsynonymous mutations on the function of two DNA mismatch repair proteins - MLH1 and MSH2 - that are important CRC susceptibility genes.

## **TY3 virus-like particle assembly.**

**Liza Z. Larsen**, Min Zhang, Yurii G. Kuznetsov, Alex McPherson, Richard H. Lathrop, G. Wesley Hatfield, Suzanne Sandmeyer

Ty3, a retrotransposon in *Saccharomyces cerevisiae* encodes homologs of retroviral Gag and Gag-Pol proteins, which, together with genomic RNA, assemble into virus like particles (VLPs) that undergo processing and reverse transcription. Similar to retroviruses, Ty3 CA contains a major homology region and a late domain. Ty3 particle maturation has advantages for study as a model system for retroviral assembly because Ty3 immature VLP structure and the Gag polyprotein has only CA, p3, and NC domains. In addition, the Ty3 lifecycle is intracellular making it simpler to study its entire morphogenesis process. Results from alanine scanning mutagenesis of *GAG3* will be reported. Using atomic force microscopy, VLPs were imaged from cells producing wild type, and protease and reverse transcriptase mutant Ty3. VLPs exhibited arrangements of capsomers on their surfaces which were consistent with icosahedral symmetry. The greatest number of VLPs were in the range of 42 to 50 nm diameter and consistent with T=7 symmetry. To further study Ty3 assembly, we have used *E. coli* system to characterize the basic components required for VLP production. With the use of a computationally optimized DNA assembly technology, we have achieved efficient expression of Gag3 in *E. coli* and shown that Gag3 is sufficient for formation of VLPs in *E. coli*. We have also purified Gag3 protein after expression in *Escherichia coli*. Currently we are testing conditions using this purified Gag3 protein for the production of VLPs *in vitro*.

## **Chemical informatics and computational docking.**

**S. Joshua Swamidass**, Jonathan Chen, Yimeng Duo, Jocelyne Bruand, Hartmut Luecke, Pierre Baldi

In conjunction with the computational docking work here at UCI, our lab has been developing a large curated database of purchasable chemicals, the Chemical Databank (CDB). The CDB is about 6 million compounds large, includes chemical structures for docking and will be made available for searching and download over the web. This database was docked against the well known cancer protein, p53, our hits are currently being co-crystallized with p53 in Hartmut Luecke's lab.

We have been developing new methods for searching and analyzing large chemical repositories such as the CDB. Kernel based machine learning methods, similarity measures, analysis tools are being used to understand chemical space in new ways that could help with important tasks such as predicting toxicity, solubility, bioactivity, and more.

## **Coffee Break in F108**

## **Afternoon Session II: The Molecular Modeling and Systems Biology**

### **Prediction of protein disordered regions using machine learning Techniques.**

**Michael J Sweredoski**, Jianlin Cheng, Pierre F Baldi

Intrinsically disordered regions in proteins are relatively frequent and important for our understanding of molecular recognition and assembly, and protein structure and function. From an algorithmic standpoint, flagging large disordered regions is also important for ab initio protein structure prediction methods. Here we first extract a curated, non-redundant, data set of protein disordered regions from the Protein Data Bank and compute relevant statistics on the length and location of these regions. We then develop an ab initio predictor of disordered regions called DISpro which uses evolutionary information in the form of profiles, predicted secondary structure and relative solvent accessibility, and ensembles of 1D-recursive

neural networks. DISpro is trained and cross validated using the curated data set. The experimental results show that DISpro achieves an accuracy of 92.8% with a false positive rate of 5%.

### **Predicting mutant protein function using computer models.**

**Samuel A. Danziger**, Richard Lathrop, Rainer Brachmann

Many biomedical problems relate to mutant functional properties across a sequence space of interest, e.g., flu, cancer, and HIV. Detailed knowledge of mutant properties and function improves medical treatment and prevention. A functional census of p53 cancer rescue mutants would aid the search for cancer treatments from p53 rescue. We devised a general methodology for conducting a functional census of a mutation sequence space, and conducted a blind predictive test on the functional rescue property of 71 novel putative p53 cancer rescue mutants iteratively predicted in sets of 3. Blind predictive accuracy (15-point moving window) rose from 47% to 86% over the trial ( $r = 0.74$ ).

### **Hidden Markov modeling of JNK docking sites.**

**David Ho**, **Ryan W. Benz**, Pierre Baldi and Lee Bardwell

Mitogen activated protein kinases (MAPKs) are a well-conserved family of proteins crucial for transducing a diverse array of extracellular signals. Of the four major MAPK pathways identified in mammals, the JNK pathway is of particular interest because of the critical role it plays in apoptosis, tumorigenesis, muscular dystrophy and Parkinson's disease, just to name a few examples. Though the JNK pathway is involved in the regulation of many important diseases and conditions, relatively few JNK substrates have been identified. In the current work, Hidden Markov Models (HMMs) are used for the purpose of searching through protein databases to find novel JNK docking sites, and ultimately novel JNK substrates. By training an HMM on experimentally known docking site regions of JNK substrates, the HMM is able to correctly find docking-sites from full-length protein sequences using a sliding window, Viterbi path scoring algorithm. From the training/validation sequences used, all of the known JNK docking site regions show a characteristic spike in the Viterbi path probability around the docking site regions, a fingerprint that may be useful in finding novel JNK substrates from protein databases. Finally, to test the computational methods, experimental techniques will be used verify the predicted JNK substrates.

### **A generalized Monod, Wyman, and Changeux (MWC) model for mathematical modeling of metabolic enzymes with allosteric regulation.**

**Tarek S. Najdi**, Chin-Rang, Yang, Bruce E. Shapiro, G. Wesley Hatfield and Eric D. Mjolsness

We have previously described a mathematical model based on the concerted transition model of Monod, Wyman, and Changeux (the MWC model) that simulates the allosteric regulation of single substrate, product, and effector ligand enzymes. To achieve this goal, we used kMech, a Cellerator language extension that describes enzyme mechanisms for the mathematical modeling of metabolic pathways. These mechanisms are converted by Cellerator into ordinary differential equations (ODEs) solvable by Mathematica™.

Here, we describe a flexible model in Cellerator, which generalizes the MWC model for allosteric regulation that allows for multiple substrate, product, activator and inhibitor binding sites. We have used this generalized model to simulate the kinetic behavior of a bifunctional allosteric enzyme, aspartate Kinase I-Homoserine Dehydrogenase I (AKI-HDHI).

### **Stochastic models for the simulation of genetic regulatory mechanisms.**

**Chin-Rang Yang**, Bruce, E. Shapiro, Eric D. Mjolsness, Pierre F. Baldi and G. Wesley Hatfield

We recently published, kMech, a high-level computer language for the deterministic modeling enzyme kinetic mechanisms [Yang, C.-R., Shapiro, B. E., Mjolsness, E. D., and Hatfield, G. W. (2005) *Bioinformatics*, 21(6):774-80]. However, unlike most reactions of central metabolism, many crucial events of living cells depend on the interaction of small numbers of molecules that are sensitive to the underlying stochasticity of reaction processes, such as genetic regulatory mechanisms that involve the interaction of a small number of regulatory proteins with a single high-affinity DNA binding site. Therefore, to accurately simulate these types of reactions, we have developed gMech. This is another high-level computer language extension, based on the Gillespie algorithm, that describes a suite of stochastic genetic regulatory mechanisms.

We have used gMech to model the dynamics of transcriptional attenuation of the *ilvGMEDA* operon in the model organism *Escherichia. coli*. Attenuation is the regulation of transcription termination at a site preceding a gene. For amino acid biosynthetic genes, attenuation is effected by the formation of mutually exclusive secondary mRNA structures that form in response to *in vivo* levels of amino acyl-tRNAs during the coupled transcription and translation process. A gMech model that discretizes this continuous dynamic attenuation process into different states according to the common features of the attenuation mechanism was constructed. To identify each state uniquely, we applied an index notation in which: the first number in the bracket is the DNA base-pair location of a transcribing RNA polymerase in the leader region of the gene; the second number is the location of a ribosome in the leader polypeptide coding region the nascent leader RNA; and the third number indicates the secondary RNA structure given the positions of RNA polymerase and the ribosome. With this notation, the attenuation mechanism of the *ilvGMEDA* operon can be modeled into a tree-like structure of discrete states for stochastic modeling. We show that gMech accurately simulates attenuation of isoleucine, valine, and leucine, biosynthetic gene expression in response to *in vivo* levels of isoleucyl-tRNA, valyl-tRNA, and leucyl-tRNA.

### **Biomechanical models of *Arabidopsis* meristems.**

**A. Sadovsky**, P. Baldi, E. Mjolsness, M. Heisler, T. Bacarian, E. Meyerowitz

We study the relation between the mechanical and genetic regulatory properties of the shoot apical meristem (SAM) of *Arabidopsis Thaliana*. The content of the talk has three parts. First, we formulate a continuum-mechanical model of SAM regarded as an elastic continuum, and provide some numerical solutions. In the second part, we analyze the hypothesis made by Hejnowicz regarding the existence of curvilinear coordinate systems that give a "natural" geometric description of the \*growth tensor\*. In the third, final, part, we discuss the SAM tissue at the level of individual cells. An attempt is made to characterize the growth of a cell as an interaction between expansin enzyme activity, cell wall synthesis, and the mechanical behavior of the cell.

## **Closing**

Closing Remarks. **G.W. Hatfield, P. Baldi**