UNIVERSITY OF CALIFORNIA, DAVIS

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

COLLEGE OF ENGINEERING
DEPARTMENT OF
BIOMEDICAL ENGINEERING

1 SHIELDS AVENUE
DAVIS, CALIFORNIA 95616-5294
FAX (530) 754-5739
(530) 752-1033

Dear Prof. Brun:

I am an Assistant Research Scientist in the Biomedical Engineering Department at UC Davis and I am in the process of applying for tenure-track faculty positions. I saw your add to hire new faculty in the general area of Computational Systems Biology for the Department of Biology and the Biocomplexity Institute at the University of Indiana, on the Nature Jobs webpage (Online Ref: 73583) and I would like to be considered for an Assistant Professor position.

My Ph. D. work was done at the Microbiology & Immunology Department of the University of Michigan Medical School, in Prof. Savageau's lab. I received the degree from the University of Lisbon in July 2000. My work focused on developing methods to analyze design principles in molecular biology systems. I applied those methods to metabolic and signal transduction networks. After, I went to London for post doctoral work in Prof. Michael J. E: Sternberg's Lab, at the Imperial Cancer Research Fund. I used genome and enzyme databases to study evolution of enzyme networks. I then continued my post doctoral work in the University of Lleida in Spain, at the Departament de Ciencies Mediques Basiques. There, my work with Prof. Albert Sorribas combines structural bioinformatics and dynamic modelling to reconstruct metabolic systems *in silico*. In the last two years I have also been working on large scale genome analysis to understand molecular biology design principles and evolution of microbial genomes and microbial interactions. My profile, experience and interests fit well with the Computational analysis of mechanisms of bacterial cell function and Biomolecular networks, including signalling, gene regulatory and metabolic networks, which are two of the areas that you target for your search.

In addition to my research activities I have experience in teaching Computational Biology, Bioinformatics and Analysis of Cellular Networks to a Graduate student audience and in supervising graduate students.

Together with this letter I submit my *CV*, Past Research Accomplishments statement and a Research Statement/Plan for your evaluation. I add five of my publications. I also asked for reference letters to be sent to you by Profs. Michael A. Savageau (UC Davis), Enric Herrero and Albert Sorribas (U. Lleida, Spain), Armindo Salvador (U. Coimbra, Portugal) and Michael J. E.

UNIVERSITY OF CALIFORNIA, DAVIS

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

COLLEGE OF ENGINEERING
DEPARTMENT OF
BIOMEDICAL ENGINEERING

1 SHIELDS AVENUE
DAVIS, CALIFORNIA 95616-5294
FAX (530) 754-5739
(530) 752-1033

Sternberg (Imperial College, London).I thank you for your consideration and look forward to hearing from you. Sincerely,

Rui Alves

Past Research Accomplishments

Rui Alves

October 2005

In the past I have been interested and have accomplishments in the theoretical analysis of different Molecular Biology problems, ranging from oxidative stress to the evolution of microbial evolutionary networks to the development of mathematical formalisms for the analysis of populations.

Early Research Career Accomplishments:

As a senior undergraduate student I was involved in the theoretical analysis and mathematical modeling of lipid peroxidation in membranes. In this research I have contributed with analysis and with models of the mitochondrial respiratory chains that focused on the production and usage of reactive oxygen species and were then integrated in a larger scale model of hepatocytes [1]. This model helped in understanding the contribution of different factors to the molecular effects of oxidative stress in membranes.

Ph. D. Research Accomplishments:

After finishing my undergraduate degree in Biochemistry I was accepted in an interdisciplinary Ph.D. program that ultimately allowed me to perform my research in Prof. Michael Savageau's Lab at the University of Michigan Medical School Microbiology and Immunology department. During this period I was interested in the analysis of biological design principles in Molecular Biology.

One of the more prevalent regulatory design principles observed in nature is that, in microorganisms, the final product of a biosynthetic pathway inhibits the rate of the first reaction of the pathways. using a theoretical method known as Mathematically Controlled Comparison I have investigate why this design is so prevalent and, based on objective efficiency criteria, like robustness of steady state concentrations and fluxes, responsiveness and stability. The work shows that a pathway in which this type of feedback exists, in general, responds faster and is more robust to changes in the environment than any other alternative type of inhibitory regulation [2]. This work provides a rationale for the natural selection of such a design.

Another accomplishment of my Ph. D. was the following. It is well known that the irreversible reactions of the biosynthetic pathways mentioned in the previous paragraph tend to be at the beginning of the pathways. Using the methods and objective criteria mentioned earlier I have investigated why this is so. It was found that, by keeping irreversibility close to the beginning of a biosynthetic pathway, on average, allows for faster physiological responses and larger robustness to changes in the environment [3].

The Mathematically Controlled Comparison technique mentioned earlier allows one to average out possible compensatory physiological adaptations along the evolutionary paths of the different network designs being compared retaining, in the end, only the differences in network behavior that are unique to each design. The standard way to apply this method is the following: i) write the differential equations that describe the dynamical behavior of the alternative networks in power law formalism; ii) obtain the analytical steady state solutions for the different networks being compared; iii)determine the important physiological properties that should be analyzed; iv) calculate the constraints that average out compensatory adaptations; iv)calculate the analytical ratios of corresponding properties in the comparable alternative designs; v) analyze the signs of the parameters (no values are known) to determine if the ratios are smaller than equal to or larger than 1; vi) based on these ratios decide which system performs best under different conditions. Before my work, to have a general interpretation, this method was fully analytic, and in many cases, one could not, simply by knowing the signs of the parameters, determine if the ratios are smaller than, equal to or larger than 1. Attributing parameter values is not of great assistance because it may answer the question for a given set of parameter values but will not provide a generally valid answer. I have accomplished the development of an extension of the Mathematically Controlled Comparison, introducing statistical and numerical concepts that allow useful general information to be drawn from cases that can not be decided analytically [4,5]. In parallel to this I have developed a method that allows the study of how admissible physiological behavior of a network limits and influences allowed parametric values [6].

Post Graduate Accomplishments

In my first post doc I changed fields and went to work in Prof. Sternberg's lab at the ICRF and latter at the Imperial College, in London. My goal was to learn bioinformatics and genome analysis techniques and, latter integrate these with what I had already learned during my Ph. D. in order to have a fuller set of tools that would allow me to tie together evolution and systems biology. While at Prof. Sterneberg's Lab I applied structural bioinformatics to analyze fully sequenced genomes of many different microbes. What I feel was the major accomplishment in this work was that for the first time, pathway evolution was studied from a full enzyme network perspective and not by considering the different biochemical pathways individually. This changed somewhat the picture one has about the evolution of metabolic pathways[7]. It was found that retro evolution of enzymes is more common than it was thought up until that point, although recruitment of enzymes with similar chemical function from other pathways was still the dominant form of enzyme evolution within the enzyme networks.

As I continued pursuing the analysis of genomic information to understand molecular evolution I focused on the evolution of amino acid biosynthetic pathways in microbes. Genes coding for enzymes that synthesize a given amino acid in bacteria are repressed when the cognate amino acid is abundant in the medium. The amino acid content of these enzymes should be biased towards low values of the cognate amino acid (**Low Cognate Bias**). Otherwise, when the cognate amino acid is depleted from the medium, the biosynthesis enzymes could not be effectively made *de novo,* due to lack of the amino acid, creating a "Catch 22" situation. I accomplished to find support for this hypothesis after analyzing parallel lines of evidence including whole genome sequence data, genetic regulation, estimations of enzyme activity, mathematical models and correlation between amino acid composition of proteins and environmental supply of individual amino acids. The analysis was done using the genomes and additional relevant information from *Escherichia coli, Salmonella typhimurium* and *Bacillus subtilis* [8].

At this time I had started to integrate the two broad areas of mathematical modeling and bioinformatics and apply them to the analysis of the microbial signal transduction networks known as Two Component systems. These are two steps signal transduction cascades in which a sensor protein recognizes a signal and autophosphorylates. After this, the sensor transfers its phosphate to a regulator protein that will regulate some physiological effect or gene expression. In some cases the sensor protein, when unphosphorylated also acts as a phosphatase to the response regulator (bifunctional sensor design), while in others, the sensor protein only transfer its phosphate an independent phosphatase exists (monofunctional sensor design). On one hand I used structural bioinformatics and whole genome analysis over a range of different microbes to analyze two component systems signal transduction networks. I was able to predict structural features for sensor bifunctionality in a two component system. The ATP lid of bifunctional sensors appears to be properly folded in a α-helix conformation in bifunctional sensors, while it appears to be disordered in monofunctional sensors. Using objective physiological criteria it was found that a bifunctional sensor design buffers against crosstalk or signal leakage to and from other signal transduction modules, while a monofunctional design favors the integration of signals from different sources [9].

At this point I was equipped with the theoretical tools to go back and analyze another aspect of the molecular biology of oxidative stress, which tied in to my pre-graduate school research interests. Combining structural bioinformatics, mathematical modeling and experimental two-hybrid system analysis, I was able to initiate the metabolic reconstruction of the iron sulfur cluster (FeSC) assembly pathway in yeast. I was able to theoretical test for alternative roles of the different proteins known to be involved in the process. Comparing the results of the *in silico* experiments with those from corresponding wet lab experiments I was able to predict the role of the different proteins involved in the process[10-12]. My pre-graduate school research interests in oxidative stress.

Reference List

1. Antunes, F., Salvador, A., Marinho, H. S., Alves, R. & Pinto, R. E. Lipid peroxidation in mitochondrial inner membranes. I. An integrative kinetic model. *Free Radic. Biol. Med.* **21**, 917-943 (1996).

2. Alves, R. & Savageau, M. A. Effect of overall feedback inhibition in unbranched biosynthetic pathways. *Biophys. J.* **79**, 2290-2304 (2000).

3. Alves, R. & Savageau M.A. Irreversibility in unbranched pathways: preferred positions based on regulatory considerations. *Biophys. J.* **80**, 1185 (2001).

4. Alves, R. & Savageau M.A. Comparing systemic properties of ensembles of biological networks by graphical and statistical methods. *Bioinformatics* **16**, 527-533 (2000).

5. Alves, R. & Savageau, M. A. Extending the method of mathematically controlled comparison to include numerical comparisons. *Bioinformatics* **16**, 786-798 (2000).

6. Alves, R. & Savageau, M. A. Systemic properties of ensembles of metabolic networks, application of graphical and statistical methods to simple unbranched pathways. *Bioinformatics* **16**, 537-547 (2000).

7. Alves, R., Chaleil, R. A. & Sternberg, M. J. Evolution of enzymes in metabolism: a network perspective. *J Mol Biol* **320**, 751-70 (2002).

8. Alves, R. & Savageau, M. A. Evidence of selection for low cognate amino-acid bias in amino-acid biosynthetic enzymes. *Mol Microbiol* **56**, 1017-1034 (2005).

9. Alves, R. & Savageau, M. A. Comparative analysis of prototype two-component systems with either bifunctional or monofunctional sensors, differences in molecular structure and physiological function. *Mol. Microbiol.* **48**, 25-51 (2003).

10. Alves, R., Herrero, E. & Sorribas, A. Predictive reconstruction of the mitochondrial iron-sulfur cluster assembly metabolism: I. The role of the protein pair ferredoxin-ferredoxin reductase (Yah1-Arh1). *Proteins* **56**, 354-66 (2004).

11. Alves, R., Herrero, E. & Sorribas, A. Predictive reconstruction of the mitochondrial iron-sulfur cluster assembly metabolism. II. Role of glutaredoxin Grx5. *Proteins* **57**, 481-92 (2004).

12. Vilella.F. *et al.* Evolution and cellular function of monothiol glutaredoxins: involvement in iron-sulphur cluster assembly. *Comp Funct Genomics* **5**, 328-341 (2004).

Research Interests & Initial Research Program

Rui Alves

June 2005

It is the broad goal of my research program to use and develop different theoretical, bioinformatics and computational (TBC) techniques to address specific biological questions that can be framed within the fields of Molecular Evolution and Systems Biology in microbial species.

1 – Application of an Integrative Methodology for the Reconstruction of Metabolic Pathways to the Biogenesis of Iron Sulfur Clusters.

Different theoretical, bioinformatics and computational (TBC) techniques are used by molecular biologists to analyze their genes/proteins/phenomena of interest. Although each of these techniques has a well defined and limited scope, together they have a wide range of application that spans from the analysis of a full organism/cell to the analysis of a given signal transduction or metabolic pathway to the analysis of individual genes or proteins. However a structured, algorithm-like, methodology that allows researchers to combine the different techniques to perform vertical and horizontal analysis of their system of interest at different levels is lacking. Developing such a structured methodology to use the TBC set of techniques in tandem would provide a powerful integrative approach to the analysis of Molecular Biology phenomena from a systemic perspective. It is my goal to develop such a methodology and benchmark it by applying it to the reconstruction of the iron-sulfur cluster (FeSC) biogenesis pathways. Defects on some of the proteins in this pathway in humans lead to Reactive oxygen species related diseases like Friedriech's ataxia. The goal will be achieved through two consecutive research stages:

1 –Implementation and refinement of the algorithm presented in Figure 1 in a publicly accessible meta-server. Kinetic modeling, protein structure modeling/analysis, *in silico* protein docking, automated literature analysis, and large scale gene expression and protein expression data analysis will be combined in a structured manner. Benchmarking of different part of the methodology will be done using different target systems. The benchmarking of the network determination using bioinformatics methods will be done using the network of well characterized two component systems modules of E. coli. Once the benchmarking of this part of the methodology is done we will integrate it will be integrated with the remaining part of the methodology and benchmark of the assembled methodology will be done by applying it to the analysis of the network of proteins involved in iron sulfur cluster (FeSC) biogenesis. These clusters are fundamental in many different aspect of cellular function, with roles in catalysis, structure of proteins and signal transduction. The FeSC biogenesis pathway will be analyzed in different cells types, including bacteria, Archaea and Eukaryotes. This large scale approach will allow for a study of the evolution of the FeSC biogenesis pathways.

2 – Creation and maintenance of a searchable database and website with interactive capabilities where the data resulting from specific goal 1 will be stored and made available for validation and iterative refinement by the scientific community involved in Fe-S cluster (FeSC) biogenesis. The goal at this stage

is to provide an organizing center for the information, readily available and updatable, where data can be integrated at different levels.

a) Step 1: Implementation and refinement of the algorithm presented in Figure 1.

The procedure we propose in Figure 1 can be schematically described in the following way:

1) Determine the proteins and metabolites of interest that are thought to be involved in the process of interest. Use expert knowledge and literature analysis and phylogenetic profiling for this determination.

2) If possible, obtain structures for these proteins. If no structure is available, obtain structural models from the sequences.

3) Use all against all protein docking to derive the most likely interactions. Analyze those interactions. Additionally, use other methods to predict network interactions between the proteins (co-evolution, meta-text analysis)

4) Derive a degenerate/incomplete set of possible network structures based on the interactions obtained from 3).

5) Eliminate from/add to 4) any interactions that are eliminated/ suggested by other known data, from large throughput experiments or from primary literature meta-analysis.

6) Human curation of the resulting schemes to evaluate its appropriateness. Identify alternative models corresponding to different hypothesis for the component elements and processes.

7) Derive mathematical models for the selected schemes using a Generalized Mass action (GMA) approach.

8) Normalize the models and scan parameters over appropriate ranges to determine which alternative networks are able to reproduce known experimental behavior of the system.

9) If more than one of the alternative schemes reproduces known experimental results, devise new experiments to differentiate between the alternatives. Go back to step 5). If no network reproduces the experimental results go back to reanalyze the networks that have been discarded in step 6. If, after modeling none of those reproduces the experimental results, go back to step 1.

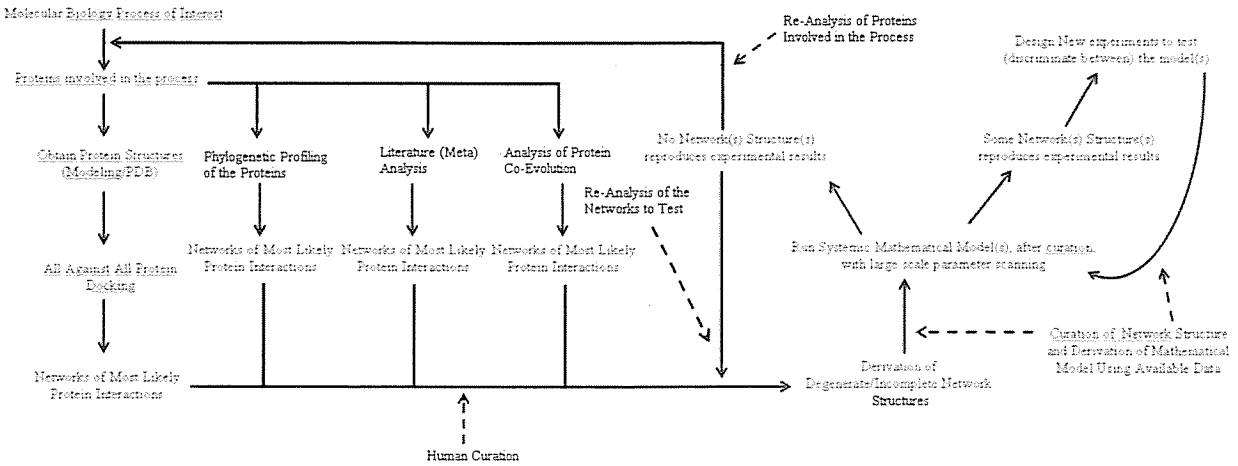

Figure 1 Schematic flow chart for the method proposed in this Research Plan

The integrative methodology proposed here will be implemented using several pre-existing computational tools, databases and servers. Structural models/crystal structures for the different proteins will be obtained using several different servers (Swissmodel[1], 3DJIGSAW[2], Robetta[3], etc.) or from the protein databank[4,5]. In house analysis will be used to improve the models. Existing docking techniques and algorithms are also freely available for academic use (Gramm[6-8], Hex[9], etc.). Using the PERL programming language, in house scripts that translate network information into differential equations in power law form have been developed. Simulation is done using well tested algorithms and programs (PLAS[10,11], GEPASI[12]). Software and servers for network extraction from literature meta-analysis are also available[13-15]. Phylogenetic analysis tables [16] can be created easily in-house by using BLAST [17] and the protein sequences available on-line. The limitations inherent to each of the individual techniques will be diminished or overcome by the integration of information derived from the application of the different TBC techniques. The software, servers and algorithms available to set up the methodology and perform the analysis is freely available either over the internet or at the UM library.

The biological data needed for this stage of the work is readily available. Published literature and public database are the primary source for the proteins involved in FeSC cluster biogenesis in the different organisms. Public access full genome and proteome data is also available for each the target organisms and will be used. The exception is *M. xanthus*. The proteins involved in FeSC biogenesis in *M. xanthus* will be identified by homology searches and comparison with the other organisms. Protein and genome sequence is available through collaboration with Dr. Mitchell Singer's Lab at UC Davis.

b) Step 2 - Creation of a searchable database available via an interactive website

The data resulting from this project will be made available to the scientific community via peer reviewed publications, meeting presentations and a public website. A database where the information regarding both how to apply the methodology and the analysis and reconstruction of FeSC biogenesis itself is to be created. This database will be interactive and fully searchable on-line. It will be freely available over the WWW and contain information regarding the analysis of the different organism, as well as the experiments that are proposed to test the predictions derived from the *in silico* analysis. The web site where this database will be made available will contain also means by which external researchers will be able to communicate both the results of proposed experiments and the requests for further *in silico* experiments

The database will be build using MySql technology. The website will be built using HTML and all cgi programming will be done using PERL.

2 – Molecular Evolution of Enzyme Network Composition and Connectivity

With the increase in available sequence data and annotated genomes, the evolution of enzyme networks can be extensively studied from a molecular evolution perspective. To investigate which mechanisms may be involved in evolving the reaction network of metabolic enzymes at a complete genome scale I have created and analyzed databases of completed genomes from publicly available information. These databases were used to determine the enzyme content and, upon identification of the different enzymes, created a reaction network for the metabolism of each of the studied genomes. The analysis of the homology patterns between enzymes of the different networks was then used to measure the prevalence of different mechanisms of enzyme evolution. The different mechanisms of enzyme network evolution that are more widely accepted are retro evolution (enzymes for the later steps of a pathway are used as templates to, upon gene duplication, evolve new enzymes to catalyze the early reactions of a pathway) or recruitment (enzymes from different pathways catalyzing similar reactions are, upon gene duplication, used as templates to evolve enzymes for another pathway). If one uses the traditional pathway definitions, it has been observed that recruitment is prevalent. However, if one considers that there is much branching in metabolism and drops traditional metabolic pathway definitions, using distance in the network as a determinant if two enzymes are in the same pathway or not retro, evolution emerges as a much more prevalent evolutionary design than previously thought [18,19].

Genomic data also contain data about how evolutionary pressures regarding the regulation of gene expression and protein activity lead to specific changes in the composition of the cellular proteins. For example, genes coding for enzymes that synthesize a given amino acid in bacteria are repressed when the cognate amino acid is abundant in the medium. The amino acid content of these enzymes should be biased towards low values of the cognate amino acid (**Low Cognate Bias**). Otherwise, when the cognate amino acid is depleted from the medium, the biosynthesis enzymes could not be effectively made *de novo,* due to lack of the amino acid, creating a "Catch 22" situation. This hypothesis is supported by parallel lines of evidence including whole genome sequence data, genetic regulation,

estimations of enzyme activity, mathematical models and correlation between amino acid composition of proteins and environmental supply of individual amino acids. The analysis was done using the genomes and additional relevant information from *Escherichia coli, Salmonella typhimurium* and *Bacillus subtilis* [20].

i) **Validation of the Generality of the Cognate Bias Hypothesis**

The cognate bias hypothesis has been tested *in silico* for three microbes. It is a specific goal of this research program to extend and test the hypothesis to other systems and to a larger number of organisms. To do so we intend to create a database of amino acid bias for the proteins of selected fully sequenced genomes. We will then analyze the proteins involved in amino acid biosynthesis and determine the extent of validity for the hypothesis. There will be some experimental cross validation for the results. Preliminary results suggest that, in some organisms, the enzyme that contributes with the largest flux for alanine production is in some case the same cysteine desulfurase that is involved in the FeSC biogenesis. Thus, at least for some organisms, it may be that this subproject will be relevant to the subproject described in the previous section.

ii) Investigating cognate bias in general amino acid metabolism

Amino acid metabolism will be analyzed to test the two following hypothesis.

1 – Low cognate bias should exist in amino acid transport proteins. In many cases, the first response to amino acid deprivation is the synthesis of membrane transporters. Thus, the composition of these transporters should be biased towards low values of the amino acid(s) they transport.

2 –High cognate bias should exist in amino acid utilization pathways. Catabolic pathways for the different amino acids, if induced or derepressed by an increase in amino acid levels should have a high bias for the amino acids that they utilize. Furthermore, the priority of usage of the amino acid by the different pathways should be inversely correlated with the cognate bias.

iii) Generalization of the bias hypothesis to other systems

Other systems with a similar form of positive feedback should have a similar bias. This hypothesis will be tested for the following cases:

1 – Proteins with regulatable gene expression involved in pathways that fix atoms in metabolism (e.g. nitrogen, carbon, sulfur) are likely to be biased towards low content of their cognate atoms.

2 – Proteins that are exclusively anaerobic are less prone to oxidative damage and thus, their amino acid content should be biased towards higher amounts of highly oxidizable amino acids.

iv) Use of amino acid bias profiles to derive a classification for the different organisms

The profiles of amino acid bias in amino acid biosynthetic pathways provide a finger print for an organism and its environment. This finger print can be used to classify the different organisms. Metrics that, based on the profiles, will allow the creation of such a classification will be derived and tested

3 – Analysis of Evolutionary Design Principles in Metabolic Networks

It has been observed that biological networks, both reaction and regulatory, have a limited number of designs. Furthermore, networks that appear to perform very similar function may have a core motif with slight variations in the connectivity. This is a consequence of life history and, in many cases, of selective pressures that are active on the networks. Can we identify the selection conditions that, constrained by life history, select (alternative) design(s) for a given network?

One of the more prevalent regulatory design principles observed in nature is that, in microorganisms, the final product of a biosynthetic pathway inhibits the rate of the first reaction of the pathways. using a theoretical method known as Mathematically Controlled Comparison I have investigate why this design is so prevalent and, based on objective efficiency criteria, like robustness of steady state concentrations and fluxes, responsiveness and stability. The work shows that a pathway in which this type of feedback exists, in general, responds faster and is more robust to changes in the environment than any other alternative type of inhibitory regulation [21]. This work provides a rationale for the natural selection of such a design.

It is well known that the irreversible reactions of the biosynthetic pathways mentioned in the previous paragraph tend to be at the beginning of the pathways. Using the methods and objective criteria mentioned earlier I have investigated why this is so. It was found that, by keeping irreversibility close to the beginning of a biosynthetic pathway, on average, allows for faster physiological responses and larger robustness to changes in the environment [22].

The Mathematically Controlled Comparison technique mentioned earlier allows one to average out possible compensatory physiological adaptations along the evolutionary paths of the different network designs being compared retaining, in the end, only the differences in network behavior that are unique to each design. The standard way to apply this method is the following: i) write the differential equations that describe the dynamical behavior of the alternative networks in power law formalism; ii) obtain the analytical steady state solutions for the different networks being compared; iii)determine the important physiological properties that should be analyzed; iv) calculate the constraints that average out compensatory adaptations; iv)calculate the analytical ratios of corresponding properties in the comparable alternative designs; v) analyze the signs of the parameters (no values are known) to determine if the ratios are smaller than equal to or larger than 1; vi) based on these ratios decide which system performs best under different conditions. Before my work, to have a general interpretation, this method was fully analytic, and in many cases, one could not, simply by knowing the signs of the parameters, determine if the ratios are smaller than, equal to or larger than 1. Attributing parameter values is not of great assistance because it may answer the question for a given set of parameter values

but will not provide a generally valid answer. I have developed an extension of the Mathematically Controlled Comparison, introducing statistical and numerical concepts that allow useful general information to be drawn from cases that can not be decided analytically [23,24]. In parallel to this I have developed a method that allows the study of how admissible physiological behavior of a network limits and influences allowed parametric values [25].

There are several questions that I would like to address, both biological and methodological. On one hand I would like to analyze the predictions that the first enzyme in an amino acid biosynthetic pathway should inhibit the tRNA synthetase for that amino acid [26,27]. By using the methodology proposed above and applying it to the different amino acid biosynthetic pathways of *E. coli*, I want to study how generalized this principles is likely to be. I would also like to analyze difference between linear and cyclic pathways and find a rationale for the different types of physiological situations where each design is favored.

Methodologically, I would to further develop the mathematically controlled comparisons, by providing stronger theoretical ground for some of the steps in the statistical mathematically controlled comparison. One of the steps that needs this clarification is, for example, the determination of the window size in the moving averaging part of the technique.

4 – Analysis of Evolutionary Design Principles and Metabolic Reconstruction of Two Component Systems

It is also well known that homologous signal transduction networks that accomplish similar goals in different organisms may have different regulatory designs. It is of interest to understand which objective advantages are associated with each of the design, if any. I have addressed this question in bacterial traditional two component signal transduction systems. These are two steps signal transduction cascades modules in which a sensor protein recognizes a signal and autophosphorylates. After this, the sensor transfers its phosphate to a regulator protein that will regulate some physiological effect or gene expression. In some cases the sensor protein, when unphosphorylated also acts as a phosphatase to the response regulator (bifunctional sensor design), while in others, the sensor protein only transfer its phosphate an independent phosphatase exists (monofunctional sensor design). Using objective physiological criteria it was found that a bifunctional sensor design buffers against crosstalk or signal leakage to and from other signal transduction modules, while a monofunctional design favors the integration of signals from different sources [28]. In addition to this we have found a possible structural marker for bifunctionality in a two component system sensor protein. The ATP lid of bifunctional sensors appears to be properly folded in a -helix conformation in bifunctional sensors, while it appears to be disordered in monofunctional sensors.

i) Regulatory Alternatives for Phosphatases in Two Component Systems

I want to start a follow up on this research, investigating the physiological difference caused by alternative regulatory designs for the independent phosphatases in two component systems. Specifically, why is the phosphatase, in some cases, regulated by the sensor protein of the TCS they regulate while in others it is not. In parallel I want to investigate, from a theoretical perspective, which objective differences exist between traditional two component systems and traditional eukaryotic signal transduction cascades. Early results point to two component systems being more effective in blocking crosstalk.

ii) Metabolic Reconstruction of the Two Component Systems Network in *M. xanthus.*

Additionally to the evolutionary study of the alternative designs in TCS, I plan to apply the Metabolic Reconstruction methodology described earlier to the reconstruction of the Two Component System signal transduction network in *Myxobacterium xanthus*, with a focus upon the developmental pathway regulated by SdeK and other histidine kinases[29-31]. I plan to create a database/website for this two component system signal transduction network. Possible extension to other two component systems. The fully annotated genome of *Myxobacterium xanthus* has been completed and is available through collaboration with Dr. Singer's Lab. This organism is especially important in several aspects. I) it is perhaps the simplest known organism that goes through developmental stages, thus providing a good model organism to theoretically understand basic principles of development. II) It has the largest number of TCS signal transduction proteins in any bacteria known until this day (291, counting response regulator proteins and sensor proteins). A large number of these are orphan, and it is unknown which response regulators are cognate to which sensor proteins. Our methodology, when applied to these proteins will assist in unraveling the network of interaction for these proteins. Previous work has identified the Histidine Sensor Kinase (HSK) SdeK as essential for the control of early developmental gene expression. SdeK acts in concert with the C-signaling pathway to activate developmental gene expression at the 6-hr stage of development[29,32]. Several experimental approaches to identify SdeR are under way, including genetic, biochemical and genomic approaches to identify the response regulator for SdeK. Other two component systems are also involved in regulating other stages of development. Understanding how the signal transduction network works will provide valuable insights to the fields of signal transduction and development. The use of the methodology and tools described in this project, when used together with the present experimental approaches, can be helpful in guiding the reconstruction of the two component systems network that regulates development. The theoretical/experimental interactions are likely to produce fruitful synergisms in our understanding of the system. III) *M. xanthus* is not a very well studied organism (827 papers in Medline and 1152 in ISI, with large overlap). Thus applying the methodology proposed here will contribute significantly to the understanding of the signal transduction network in this organism.

Other Research Interests

Design Principles in Signal Transduction and Genetic Networks

I am also at this moment starting to collaborate with Mike Savageau at UC Davies in developing his Quantitative Demand Theory for Gene Expression that studies principles of design in Genetic Networks.

Cell Cycle Research

I have also been involved in the theoretical study of cell cycle in *Saccharomyces cerevisiae*, a research line that has been ongoing for the last two years. I have been developing a mathematical model of the cell cycle in this yeast. Preliminary unpublished results indicate that this model may be more accurate and biologically relevant than previously developed models. This increase in the accuracy and predictive is likely to be due to a more detailed modeling approach and to the realization that earlier deterministic models are unlikely to be able to capture many of the features of a process where, in some cases the amount of a given protein involved in the process is of less than 20 molecules at any given time. Therefore, we have implemented already existing stochastic algorithms and used them to solve the ordinary differential equations of the model. Additionally, we are in the latter stages of development of a method that, through some mathematical simplifications will allow us to model the cell cycle inside a cell in three dimensions, accounting for diffusion and budding. A paper is in the early stages of preparation for this work (Alves & Sorribas, in preparation). I am also involved in collaborating with Marti Aldea's experimental cell cycle group at the University of Lleida. This collaboration will allow the testing of future results of our model. The preliminary work to develop this line of research has already been funded by the Spanish and Portuguese governments.

Reference List

1. Guex, N. & Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer, An environment for comparative protein modeling. *Electrophoresis* **18**, 2714-2723 (1997).

2. Bates, P. A., Kelley, L. A., MacCallum, R. M. & Sternberg, M. J. E. Enhancement of protein modelling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins, Struct. Funct. Genetics.* **45**, 39-46 (2001).

3. Kim, D. E., Chivian, D. & Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* **32**, 526-531 (2004).

4. Bourne, P. E. *et al.* The distribution and query systems of the RCSB Protein Data Bank. *Nucleic Acids Res.* **32**, 223-225 (2004).

5. Westbrook, J., Feng, Z., Chen, L., Yang, H. & Berman, H. M. The protein data bank and structural genomics. *Nucleic Acids Res.* **31**, 489-491 (2003).

6. Tovchigrechko, A., Wells, C. A. & Vakser, I. A. Docking of protein models. *Protein Sci.* **11**, 1888-1896 (2002).

7. Vakser, I. A. Low-resolution docking, prediction of complexes for undetermined structures. *Biopolymers* **39**, 455-464 (1996).

8. Vakser, I. A. Evaluation of GRAMM low-resolution docking methodology on hemaglutinn-antibody complex. *Proteins, Struct. Func. Genetics.* **29**, 226-230 (1997).

9. Ritchie, D. W. Evaluation of protein docking predictions using Hex 3.1 in CAPRI rounds 1 and 2. *Proteins Struct Function Genetics* **52**, 98-106 (2003).

10. Voit, E. O. & Ferreira, A. E. N. *Computational analysis of biochemical systems. A practical guide for biochemists and molecular biologists.* Cambridge University Press, Cambridge, UK (2000).

11. Irvine, D. H. & Savageau, M. A. Efficient Solution of Nonlinear Ordinary Differential Equations Expressed in S-System Canonical Form. *SIAM J. Numerical Anal.* **27**, 704-735 (1990).

12. Mendes, P. Biochemistry by numbers, simulation of biochemical pathways with Gepasi 3. *Trends Biochem. Sci.* **22**, 361-363 (1998).

13. Hoffmann, R. & Valencia, A. Protein interaction: same network, different hubs. *Trends Genet.* **19**, 681-683 (2003).

14. Hoffmann, R. & Valencia, A. A gene network for navigating the literature. *Nature Genetics* **36**, 664 (2004).

15. Stapley, B. J. & Benoit, G. Altman, R. S. D. A. K. H. L. L. K. a. K. T. E. (ed.), pp. 529-540 (World Scientific Pub Co, Hawaii,2000).

16. Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D. & Yeates, T. O. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl. Acad. Sci USA.* **98**, 4285-4288 (1999).

17. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402 (1997).

18. Alves, R., Chaleil, R. A. & Sternberg, M. J. E. Evolution of enzymes in metabolism: a network perspective. *J. Mol. Biol.* **320**, 751-770 (2002).

19. Alves, R., Chaleil, R. A. & Sternberg, M. J. E. Evolution of enzymes in metabolism: a network perspective: Erratum. *J. Mol. Biol.* **324**, 387 (2002).

20. Alves, R. & Savageau, M. A. Evidence of selection for low cognate amino-acid bias in amino-acid biosynthetic enzymes. *Mol Microbiol* **56**, 1017-1034 (2005).

21. Alves, R. & Savageau, M. A. Effect of overall feedback inhibition in unbranched biosynthetic pathways. *Biophys. J.* **79**, 2290-2304 (2000).

22. Alves, R. & Savageau M.A. Irreversibility in unbranched pathways: preferred positions based on regulatory considerations. *Biophys. J.* **80**, 1185 (2001).

23. Alves, R. & Savageau M.A. Comparing systemic properties of ensembles of biological networks by graphical and statistical methods. *Bioinformatics* **16**, 527-533 (2000).

24. Alves, R. & Savageau, M. A. Extending the method of mathematically controlled comparison to include numerical comparisons. *Bioinformatics* **16**, 786-798 (2000).

25. Alves, R. & Savageau, M. A. Systemic properties of ensembles of metabolic networks, application of graphical and statistical methods to simple unbranched pathways. *Bioinformatics* **16**, 537-547 (2000).

26. Savageau, M. A. & Jacknow, G. Feedfoward inhibition in biosynthetic pathways: inhibition of the aminoacyl-tRNA synthetase by intermediates of the pathway. *J. Theor. Biol.* **77**, 405-425 (1979).

27. Singer, P. A., Levinthal, M. & Williams, L. S. Synthesis of the Isoleucyl- and Valyl-tRNA Synthetases and the Isoleucine-Valine Biosynthetic Enzymes in a Threonine Deaminase Regulatory Mutant of *Escherichia coli* K-12. *J. Mol. Biol.* **175**, 39-55 (1985).

28. Alves, R. & Savageau, M. A. Comparative analysis of prototype two-component systems with either bifunctional or monofunctional sensors, differences in molecular structure and physiological function. *Mol. Microbiol.* **48**, 25-51 (2003).

29. Garza, A. G. *et al.* SdeK is required for early fruiting body development in Myxococcus xanthus. *J. Bacteriol.* **180**, 4628-4637 (1998).

30. Pham, V. D., Shebelut, C. W., Hodgson, D. A., Singer, M. & Whitworth, D. E. PhoP4 is Required for Proper Sporulation and Development-Specific Phosphatase Activities in *Myxococcus xanthus*. *To Be Submited.* (2004).

31. Pham, V. D., Shebelut, C. W., Zumstein, E. J. & Singer, M. BrgE is a Regulator of *Myxococcus xanthus* Development. *To Be Submited.* (2004).

32. Pollack, J. & Singer, M. SdeK, a Histidine Kinase Required for *Myxococcus xanthus* Development. *J. Bacteriol.* **183**, 3589-3596 (2001).

Teaching Interests & Statement

Rui Alves

June 2005

Ultimately, students will learn whatever one demands of them. If a teacher asks for fact learning they will learn facts. If a teacher asks for problem solving skills they will develop those skills. Because I feel that both types of learning are important and need to be balanced, my main goal as a teacher is to create this balance in the students' learning process. To achieve this I aim to create in the audience a conscience of what are the general goals, questions and approaches driving the progress of the field I am teaching, both in the past and in the present. I then present them with problems and direct them to sources where they can learn more, look for answers to the problems they are presented with, and critically access and question what they are reading and learning in the context of the problems they have to solve.

Between 2002 and 2005 I taught Bioinformatics and Computational Biology module to graduate students of experimental molecular biology at the University of Lleida, Spain. This year I also taught a similar, shorter course to first year Ph. D. students in Lisbon, Portugal. I approached the teaching of the course in two stages. First I presented them with the general methods, tools and problems that are current in bioinformatics, during a limited number of classes. At the second stage of the course, the students choose the area that each thought s/he would be more interested in and prepared a class about what the state of the art in that area is. The third stage was to guide them through the use of the tools and methods they had been learning about, in an integrated and critical way, making them aware of how to solve problems using what they had learned. I gave each of the students a set of cDNA sequences and asked them to identify those sequences, model their structure, discuss their evolution in various organisms and discover the physiological process that the proteins were involved in. Once they had accomplished this, I presented them with a few questions regarding the physiological consequences of different types of mutations to the proteins they had identified and helped them with the mathematical and kinetic modeling they had to do to answer those questions. During the entire course I always make myself available to them in order to help with any possible difficulties that they might have with the work. Ideally, I would like to do a final stage where each of the students would bring in a problem of their own choosing/work and have him/her solve it for the entire class, but there has not been enough time to do that. This approach has been developed with the collaboration of the students. At the beginning of the course I tell them what my goals are for them and

perform an anonymous pool to learn what their goals in taking the course are. Then, at the end of the course, I take another anonymous pool to learn how close the students felt they came to achieve the common goals of the course and to learn any suggestions for future course improvement.

I would also be interested and capable of teaching courses on the more quantitative aspects of Molecular Biology and Molecular Evolution, Mathematical Biology and Networks in Cellular and Molecular Biology or Integrative Bioinformatics. I could also teach basic Biochemistry or other similar courses at the undergraduate level. I feel that, due to my work, I am well qualified to show the students how to build bridges between the different areas and thus maximize what they can learn. I fell that this is one of the problems in today's education, where a multidisciplinary perspective is becoming more and more important.