I. Introduction – Project Objectives

Humans can now design and piece together DNA sequences in order to construct new biological systems and organisms. We can do this more quickly and less expensively than ever. Applications abound for our synthetic biological constructs, from sensors of environmental pollutants, to devices that capture sunlight and produce biofuels. Still, an important challenge facing the synthetic and systems biology communities, hampering their efforts, is to reduce the enormous volume and complexity of biological data and dynamic behaviors into concise theoretical formulations with predictive ability.

The proposed work focuses on the development and dissemination of mathematical methods and modeling tools for computer-aided design of synthetic gene regulatory networks. The proposed predictive mathematical models will connect synthetic DNA sequences to complex biological dynamics and facilitate forward engineering of gene networks. With the proposed computational tools, users will be able to generate testable hypothesis, design, perturb and optimize the sequence of synthetic DNA constructs, test predictions and rationally engineer a targeted dynamic behavior in living bacterial organisms.

We recently published version 1.0.1 of the Synthetic Biology Software Suite (SynBioSS), a software package that puts sophisticated modeling tools at the fingertips of practitioners [1]. SynBioSS is a community resource for the semantic description, distributed management, simulation, visualization, analysis and optimization of synthetic biological constructs. All the components of SynBioSS are publicly available with Open Licenses on synbioss.sourceforge.net. With SynBioSS as a platform, we propose to launch a suite of sophisticated, user-friendly modeling tools as components of the DOE Systems Biology Knowledgebase (Kbase).

We also propose to investigate gene regulatory networks with applications in microbial production of biofuels, such as logical AND gates and comparators. Here are our objectives:

1) Develop and disseminate software tools for synthetic biology applications. We propose to create a community resource with the following components: 1) a database of standard synthetic biological components; 2) user-friendly interfaces to flexibly access, retrieve and manipulate the database information; 3) a tool that automatically generates models of these synthetic constructs; 4) a repository of these models; 5) tools to modify and improve the models ; 6) a simulation engine to conduct computer simulations of these constructs.

2) Develop multiscale dynamic models of gene expression systems. At the core of the simulation tools are algorithms for the numerical simulation of gene networks. These algorithms are challenging to use in numerical simulations because models of gene networks are stiff, i.e., are manifested over multiple scales. Furthermore, biomolecular systems and not close to the thermodynamic limit. We propose to develop algorithms that tackle multiscale, stochastic behavior. We will also develop a suite of analysis and optimization tools, appropriate for synthetic biology applications.

3) Simulate and design novel synthetic networks. We propose to simulate synthetic gene networks in *E. coli*. We will work with logical AND-gates and comparators. These systems are finding application in the production of biofuels. We have already successfully modeled and experimentally built bio-logical AND gates in *E. coli* [2, 3]. The proposed systems are a natural extension of this previous work.

We should note that with available funding from the BioTechnology Institute at the University of Minnesota and from an NSF CAREER proposal to the PI of the current proposal, we will experimentally construct and validate these designs in a wet lab. Although this is an important component of our overall efforts, it is beyond the scope of the current proposal and we will not discuss the details further.

II. Background

Synthetic biology is a flourishing discipline that shares principles, tools, and objectives with biological and engineering sciences [4-32]. Its goal is the construction of new DNA or RNA sequences that give rise to new biological behaviors, usually in the form of specified temporal and spatial expression of proteins. This expression profile is what we call 'dynamic phenotype' in this proposal.

The building blocks used in synthetic biology applications are the components of molecular biology processes: promoter sequences, operator sequences, ribosome binding sites, termination sites, reporter proteins, and transcription factors, such as activator and repressor proteins. With inexpensive DNA synthesis and manipulation technologies it is becoming possible to engineer physical assemblies of biological building blocks and create arbitrary gene regulatory relations. Thus, numerous synthetic gene circuits have been created in the past decade, including bistable switches, oscillators, and logic gates [4-32]. Possible areas of application include biofuels production, environment cleanup, chemical sensing devices, disease diagnosis, and gene therapies.

A significant challenge is to determine how to construct a synthetic system given a desired dynamic phenotype. In other words, what are the molecular components and how are they to be assembled for a particular temporal and spatial protein expression profile to be realized? For example, if we want to engineer an oscillating gene regulatory network in *E. coli*, or a metabolic pathway to produce ethanol from sugars, how do we choose the DNA sequence that will give rise to these precise phenotypic behaviors? This is a typical forward engineering problem: assembling components to a functional whole.

Arguably, the task of connecting DNA sequences to dynamic phenotypes can become rational if mechanistic explanations of complex biological phenomena were available in terms of biomolecular interactions, which can, in turn, be connected to DNA sequences. Indeed, with knowledge of the molecular species involved and of their interactions, including the thermodynamic and kinetic rules of these interactions, a rational conceptual design of synthetic gene networks is conceivable.

Admittedly, because biological phenomena are very complex, mechanistic explanations will be fragmentary and approximate. With the Babel of sequences and structures that all living organisms are, it is difficult to anticipate that we will possess as detailed and accurate a picture of biomolecular systems as we do for physical or chemical systems. Nevertheless, the careful collection, management and dissemination of hard-earned, experimental information on biomolecular processes may advance the production of mechanistic explanations in a way that is fit for analysis and design.

Importantly, mathematical modeling, which has always been a significant component of engineering disciplines, can play an important role in synthetic biology. Modeling can assist synthetic biology the same way modeling helps in aircraft or architecture design: models and computer simulations can quickly provide a clear picture of how different components influence the behavior of the whole.

In this proposal, we propose to develop data repositories, knowledge management tools, and mathematical modeling and simulation tools that provide mechanistic explanations of synthetic gene networks.

Notably, the goal to provide mechanistic explanations of complex biological phenomena in terms of biomolecular interactions is commonly shared by both systems biology and synthetic biology. Indeed, systems biology and synthetic biology are the two sides of the same coin: synthetic biology is a forward engineering approach, whereas systems biology is a reverse engineering one. The former attempts to assemble components into a new whole. The latter attempts to capture the behavior of existing biological systems in a holistic way. Their paths are complementing: systems biology generates information on components and interactions that can be used in synthetic biology applications. Synthetic biology can be employed to probe mechanisms and provide mechanistic insight on how phenotypic complexity emerges from interacting molecules.

Consequently, although the theories, methods and tools we propose to develop have a synthetic biology focus, they can become integral components of Kbase. Indeed, work in computational synthetic biology can be considered as a distinct, complementary dimension to systems biology efforts.

II.1 The Synthetic Biology Software Suite

SynBioSS is a suite of software for modeling and simulation of arbitrary synthetic genetic constructs. SynBioSS assists in connecting DNA sequences to dynamic phenotypes [1].

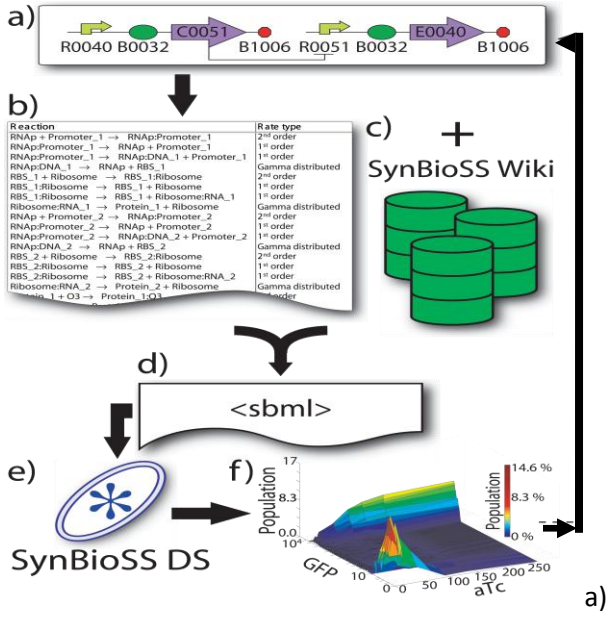
There are three components in SynBioSS: *Designer, Wiki* and *Desktop Simulator*. With **SynBioSS *Designer,*** users can rapidly construct models of gene networks. The manual construction of a model that can capture the dynamic behavior of complex biomolecular systems is a slow, arduous process. It is also prone to arbitrary choices that result in limited applicability, with a model hardly ever being used by a group other than then one that developed it.

With *Designer*, model building is automated and standardized. Users only need to enter in the *Designer* web interface the sequence of molecular components as they will be constructed in the synthetic DNA sequence (promoters, operators, ribosome binding sites, genes, terminators) and define the regulatory relations. *Designe*r then automatically generates a network of reactions that model all biomolecular interactions.

In *Designer*, the expression of genes progresses in a systematic way, following the molecular biology dogma: RNAp binds the promoter site, forms an open complex, proceeds with transcriptional elongation, and synthesizes mRNA; then ribosome binds the RBS of mRNA and proceeds with translational elongation and polypeptide synthesis; the protein is formed, it folds and functions. In the process, mRNA and protein molecules are degraded, enriching the pools of RNA bases and amino acids. This network of reactions can be built for any sequence of DNA with defined genetic components and can be exported in a SBML format file [33].

For any arbitrary gene network, the model-generating algorithm of *Designer* remains the same. With a flexible web interface *Designer* guides users in building and annotating computational models of synthetic gene networks.

Every reaction in the model has a corresponding kinetic rate that describes the rate of association of its reactant molecules and the formation or destruction of any covalent bonds or stable non-covalent interactions. **SynBioSS *Wiki*** has been specifically created to store, manage and recall just this sort of kinetic data.



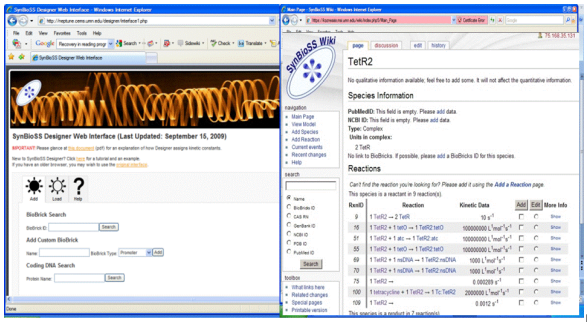


Figure1 (adapted from [1]). a) Synthetic Biology Software Suite Tools that can help address the synthetic biology challenge: with our algorithms synthetic biologists can quickly experiment with alternative synthetic designs, comparing simulated dynamic phenotypic behaviors to targeted ones. Without prior knowledge of computational mathematics, a user can enter molecular components (e.g. in BioBricks format, described later in the proposal), generate reaction networks representing the recombinant DNA sequence and simulate with sophisticated algorithms that generate concentration probability distributions. b) Screenshots of SynBioSS Designer (left) and SynBioSS Wiki.

SynBioSS *Wiki* has two components: i) a web interface based on the MediaWiki package and ii) a database for storing molecular components, their interactions, and pertinent biological information. SynBioSS *Wiki* goes beyond the MediaWiki software in storing kinetic information in a formatted, machine-searchable, format. The database of kinetic constants is easily searchable for participating species, reaction

type, etc. Users can search or browse the web site and select reactions to interactively build a model that can be exported in a SBML format. Each kinetic constant entered in the database is correlated with a reference field in the database as well as type-specific reference information (pdb ID for proteins, CAS ID for small molecules, PubMed ID for everything, etc).

Given the vast and varied nature of biochemical reaction data, no single person or research group is well suited for the task of curating such a database, thus necessitating this distributed approach – in spite of the accompanying challenges faced by any open wiki approach, such as Wikipedia. To avoid abuse and vandalism, SynBioSS Users are asked to register with a valid email address in order to make changes.

The third component**, SynBioSS *Desktop (DS),*** is a package that is currently available for Windows platforms (MacOS version is currently in beta). With SynBioSS DS, users can numerically simulate networks of chemical reactions in time and obtain the temporal expression profile of any molecule in the synthetic network. SynBioSS DS includes multiscale, stochastic algorithms (discussed in the following section), with a Windows GUI interface that can bring sophisticated algorithms at the fingertips of scientists without any UNIX/Linux/FORTRAN knowledge. SynBioSS DS is available on http://synbioss.sourceforge.net/.

With SynBioSS, users can start with a DNA sequence that encapsulates a synthetic gene network, automatically generate a set of chemical reactions that model the gene network, simulate this set of chemical reactions and obtain a dynamic phenotype. This dynamic phenotype can be connected directly to experimental measurements (e.g. flow cytometry), thereby guiding synthetic biology efforts.

II.2 Multiscale algorithms in SynBioSS

Biomolecular systems can be far from the thermodynamic limit with reactants/products numbering only very small numbers of molecules in the system. This hinders the application of traditional mathematical models for modeling kinetic and thermodynamic processes in living organisms. Using deterministic-continuous, ordinary differential equations for simulating the reaction kinetics of these systems can produce distinctly false results. The need arises for stochastic models that account for inherent thermal noise, which is manifested as phenotypic probability distributions. Indeed, it has been conclusively demonstrated that there are networks of biomolecular interactions (e.g. gene regulatory networks) whose stochastic kinetic behavior is key for emerging physiologic phenotypes [34, 35].

Since the work of early work of McQuarrie and Oppenheim [36-38], numerous groups have worked on methods to model stochastic chemical reacting systems [39-52]. In our group we have also developed theoretical methods and algorithms to simulate biomolecular systems not close to the thermodynamic limit. We developed a hybrid stochastic-discrete, stochastic continuous method combining Gillespie's discrete-stochastic kinetic Monte Carlo method [53] with a chemical Langevin equation representation of stochastic-continuous reaction events [38].

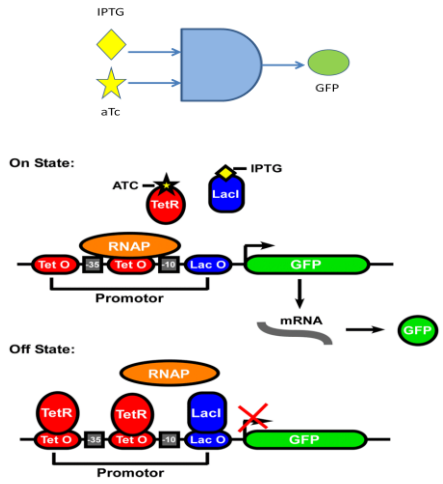


Figure 2. Schematic representation of a biological AND gate.

We then developed a pseudo-steady state approximation method to simulate stiff sets of reactions with small numbers of reactant molecules [54]. We also developed an adaptive time step algorithm for the numerical integration of stochastic differential equations, e.g. chemical Langevin equations [55]. More recently, we devised a reduction scheme for stiff stochastic differential equations based on singular perturbation methods [56]. All these algorithms are incorporated in Hy3S, another open software suite we have developed, and in SynBioSS DS [57]. These new methods result in significant computational gains over traditional kinetic Monte Carlo methods, without sacrificing the accuracy of the results.

With Hy3S and SynBioSS, we have simulated numerous synthetic biological systems. We modeled the repressilator, a three gene network with oscillating protein expression profile, and identified the biomolecular interactions that are important for oscillations [58]. We then developed a simulated annealing method to optimize the strength of key biomolecular interactions in order to obtain oscillations of certain frequency [59]. We also modeled two types bio-logical AND gates: ones that emerge from gene regulatory relations [2] and a second type that emerges from protein modular devices [60]. Moreover, we modeled inducible systems, such as bistable switches and tetracycline-inducible gene networks, studying the dynamic behavior and the influence of gene network topology on the dynamic phenotype [61].

II.3 Experimental validation: Forward engineering of synthetic bio-logical AND gates.

Because we believe that scientific insights and engineering certainty can best be ascertained by the contention of both theory and experiment, we combined our mechanistic-kinetic models with *in vivo* genetic engineering to build and test a synthetic, bio-logical AND gate [2]. This device is composed of elements of the tetracycline (Tet), lactose (Lac), and λ-phage promoters and is responsive to the commonly-used inducers IPTG and aTc, producing GFP as an output signal (Figure 2).

The quantitative behavior of the AND gate phenotype was studied both numerically and *in vivo* as a function of promoter topology. The model was constructed from kinetic data obtained from the literature, using *Designer* as described previously. It is a set of approximately 60 reactions that model transcription, translation, regulation, degradation, and induction. The model yields clearly-defined ON/OFF logical behavior at realistic inducer concentrations.

These behaviors are matched to observed *in vivo* data obtained through fluorescence-activated cell sorting. Experimentally, we constructed an *in vivo* synthetic, hybrid system consisting of multiple operators within a single promoter (Figure 2). The operator sequences employed are derived from three unrelated natural regulatory elements: Lac, Tet, and λ-phage operons arranged logically within a single transcriptional unit. Specifically, we built six single-promoter regulatory motifs by shuffling *tet* and *lac* operator sites (T and L, respectively) in and around the P_L (λ-phage): LLT, LTL, TLL, TLT, TTL and LTT).

Let us first focus on the experimental results (right column of Figure 3). A high-fidelity logic AND gate will have high GFP expression levels only at high concentrations of both aTc and IPTG. It is clear that none of the designed biological gates is of perfect digital fidelity. In the double-*tetO* systems (promoters containing two *tetO* sites and one *lacO* site: LTT, TLT, and TTL), there is always a GFP signal for non-zero aTc concentrations, even without IPTG present. This is clearly the result of leakiness of promoters containing the lactose operator. Despite the imperfect AND-gate phenotype, the double-*tetO*

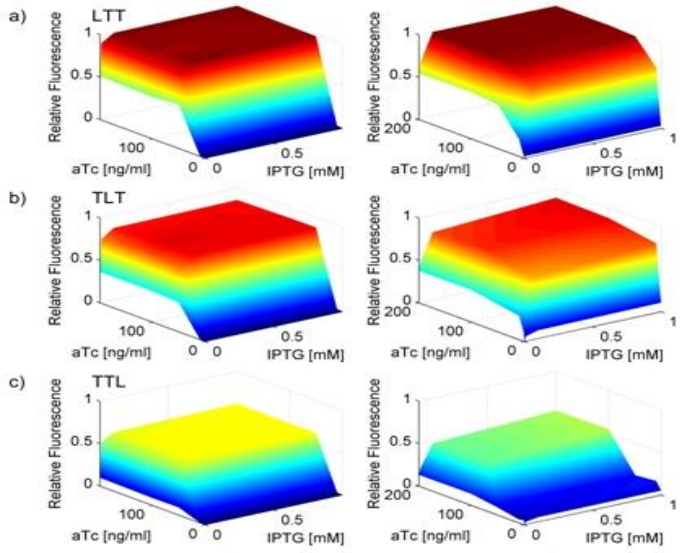


Figure 3 (from [2]). Comparison of model (left column) and experimental (right) results. The x and y axes form a grid of 36 used inducer concentrations: aTc (0-200ng/ml) and IPTG (0-1mM). The z axis is the average strength of fluorescence from the experiments or the average number of GFP molecules in the simulations, scaled by the maximum strength/number of GFP molecules. These three pairs of profiles depict the modeled (left) and observed (right) behavior of three promoter designs: (a) LTT, (b) TLT, and (c) TTL. In all cases, behavior is depicted 6 hours after induction. The plotted model values are the means of 1,000 independent stochastic kinetic simulations, whereas experimental values are the means of 100,000 FACS observations.

systems exhibit varying degrees of AND gate functionality with no GFP expressed in the absence of inducers and high GFP levels in response to high inducer concentrations.

Turning to the models, Figure 3 demonstrates that they capture the experimentally observed synthetic phenotypes. The fit of the simulated dynamic behavior of a complex network with more than 60 reactions modeled stochastically to experimental flow cytometry measurements is remarkable. This is achieved by only fitting the parameters in the leakiness of the promoters as a function of the position of the lactose operator(s) relative to the -35 and -10 sequences of the promoter (promoter topology). This emerges as a critical model parameter that was unavailable in the literature.

With this work we illustrated how mathematical modeling can provide mechanistic explanations of biomolecular phenomena and guide the design of synthetic biological constructs.

III. Research Plan

In this section, we discuss how we can make SynBioSS an expressway cyberinfrustructure for synthetic biologists to model and design gene regulatory networks.

III. 1 Develop and disseminate software tools for synthetic biology applications.

Even though SynBioSS is far from being perfect, it represents a significant first step toward the direction of an automated design process. A user-friendly web interface allows the fast and simple construction of gene networks. Information stored and retrieved by the community facilitates the use of models for quantitative characterization of synthetic constructs and generation of testable hypothesis. ***Following are four proposed subtasks for continued development and integration of SynBioSS with the Kbase.***

III.1.1 Integrate SynBioSS with Synthetic Biology Standards

A priority of ours is to adopt standards used by the synthetic biology community. In this vain, we are adapting SynBioSS *Designer* to automatically generate a kinetic model from a construct composed entirely of BioBricks. BioBricks are synthetic DNA sequences catalogued in the Registry of Standard Biological Parts, a repository of synthetic biological constructs [62]. A BioBrick standard biological part has "a nucleic acid-encoded biological function (e.g., turn on/off gene expression), along with associated information defining and describing the part" [http://bbf.openwetware.org/FAQ.html]. The sequential ordering of these BioBricks therefore describes a sequence of DNA by its intended function within a cell.

A large and ever growing community of synthetic biologists stores information of synthetic constructs in the Registry. This community is also developing standards for the definition, review, annotation, and design of standard biological parts.

SynBioSS *Designer* now has a connection to the official Registry of Standard Biological Parts (partsregistry.org/). This local database is populated using information extracted from the official Parts Registry, but organized in a way that is machine-readable, allowing for structured queries.

We have also begun the process of mirroring the Registry on computers at the Minnesota Supercomputing Institute. A letter by MSI Director Thomas Jones details the MSI commitment and the resources provided for the proposed work.

Currently, a beta version of *Designer* has a tabbed interface, making the complete sequence of a BioBrick visually accessible and easily manipulated. Clicking on a tab pulls up properties of that individual brick and allows the user to add, edit, and delete said properties. Properties are also easy to edit; clicking directly on an editable field causes a text input field or drop-down menu will appear, allowing the user to make appropriate changes.

A user will be able to enter biological components, including BioBricks, in SynBioSS *Designer*, and receive as an output a file with a reaction network that models the BioBricks. With the database of BioBricks, *Designer* can be used to streamline model construction. All information is now automatically, quickly, and

accurately retrieved, added to the current *Designer* construct, and displayed, ready to be edited if necessary.

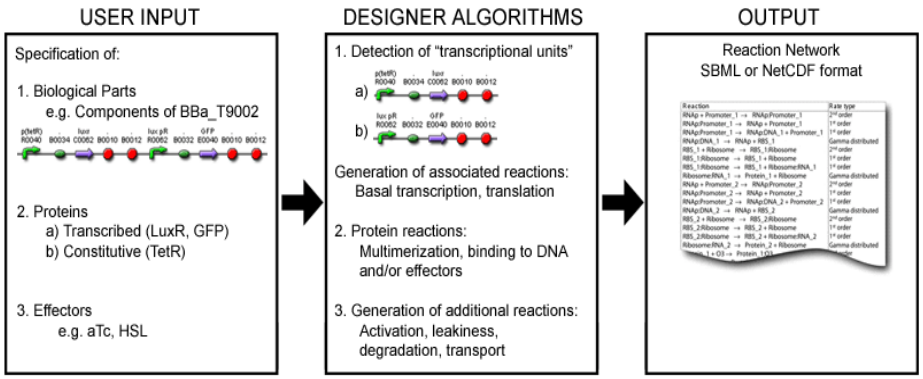


Figure 4: SynBioSS Designer steps. In the Designer webpage, a user can enter the BioBrick ID (e.g., BBa_T9002, a green fluorescent producer) and quickly obtain a chemical reaction network that models the BioBrick. Designer fetches the DNA sequence from the Registry, along with the functional unit annotation, i.e., the promoters, operators, terminators, etc. The user only needs to use drop-down menus in Designer to define regulatory relations. Designer then uses the principles of the molecular biology dogma to generate a reaction network of all the steps in transcription, translation, degradation, etc. A file is generated that can be uploaded in SynBioSS Wiki for editing and SynBioSS DS for simulation.

In Figure 4, a schematic representation is presented of the information a user enter in SynBioSS *Designer*, of the steps in *Designer*'s algorithm, and of the output reaction network. The process takes less than five minutes, from entering a BioBrick ID to saving an SBML file with a reaction network.

We envision our local databases as repositories of not only BioBricks, but of models of BioBricks and simulated expression profiles of BioBricks. Toward this goal, a second step will be to integrate the *Wiki* with BioBricks entries. There are two types of BioBrick entries: fundamental elements and composite devices. The fundamental elements will be described in detail in the *Wiki,* including all of their associated reactions, while composite devices can be represented in the *Wiki* as a collection of fundamental parts. Data exchange between the Parts Registry and SynBioss *Wiki* will automatically create the necessary entries in the *Wiki* for fundamental BioBricks, along with the "number of base-pairs" kinetic parameter for mRNA and protein production reactions. This automated importation of brick properties into the *Wiki* (e.g., how many base pairs in a coding region, which operator sites are present on a promoter) is possible with simple PHP querying scripts. Individual reactions stored in the *Wiki* and corresponding to a BioBrick would then be tagged with their appropriate BioBrick ID.

Importantly, we propose to generate reaction network representations of all the BioBricks that have been built for bacterial expression. We propose to build a dataset of reaction networks for all bacterial BioBricks and store it in SynBioSS *Wiki.* Each BioBrick can be passed through *Designer* and a reaction network be generated. A link will be created in *Wiki* to the SBML or NetCDF (binary file that can be indexed and searched quickly) file containing the reaction network. The wiki interface for the inventory of models will allow the community to modify and improve them.

III.1.2 Develop models from conceptual design to reaction networks to DNA.

One of the great successes of integrated circuit design has been in abstracting and scaling of the design problem. The physical behavior of transistors is understood in terms of differential equations. However, the design of circuits proceeds at a more abstract level – in terms of switches, gates, and functional units. This modular approach makes design tractable; furthermore, it permits a systematic exploration of different configurations, leading to optimal designs. Although driven by experimental expertise, synthetic biology has reached a stage where it calls for a similar degree of automation.

The design flow that we are proposing will allow for virtual experimentation: one can vary the inputs and parameters of synthetic designs and observe the outputs – in a manner analogous to traditional in vitro and in vivo experimentation. In this task we will develop a modular and automated flow for synthesizing

computation, such as signal processing functions, with biochemistry. Synthesis first will be performed at a conceptual level, in terms of abstract biochemical reactions – a task analogous to technology-independent synthesis in integrated circuit design. Then the results will be mapped onto specific biochemical components, selected from libraries – a task analogous to technology mapping in integrated circuit design.

Our methodology produces a set of biochemical reactions that satisfies this I/O specification – the equivalent of transistor "netlist". Given such a netlist, established simulation methods and tools are used to characterize the chemical kinetics. The resulting waveforms, specifying quantities of proteins as a function of time, confirm the validity of the design.

With the set of abstract biochemical reactions at hand, the users should be able to draw actual biochemical reactions with appropriate kinetic strengths. Currently, default kinetic constants are hard-coded in *Designer* (cf. Table in *Designer* website for sources of kinetic constants). They are carefully chosen from the literature of the tetracycline, lactose, and arabinose operons, and used for any reaction of the same type: protein-protein, transcription factor-DNA, protein-effector, RNAp-promoter, ribosome-RBS, protein degradation, RNA degradation. Careful examination of kinetic constants is needed, as the default values are only a starting point that are not appropriate for all simulation purposes. Such an examination a modification of constants can be accomplished manually in SynBioSS *DS*.

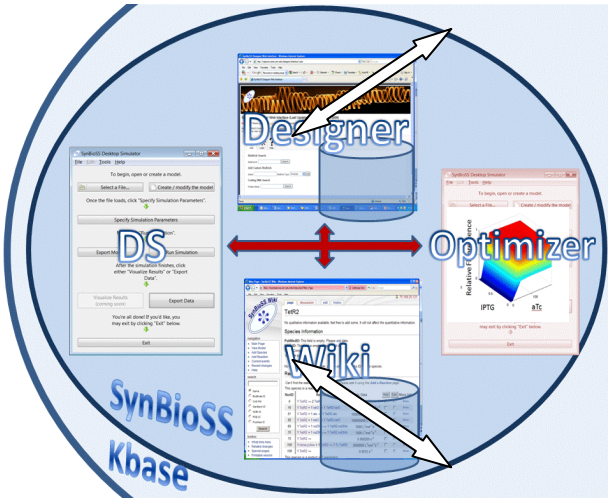


Figure 5. Synthetic Biology Software Suite Tools that can help address the synthetic biology challenge: with the proposed algorithms synthetic biologists can quickly experiment with alternative synthetic designs, comparing simulated dynamic phenotypic behaviors to targeted ones. SynBioSS Wiki and SynBioSS Designer will comprise of web-interfaces on top of publicly available data repositories. The databases are constructed with a wiki interface for the community to enter, change and extract information on standard synthetic biological parts and kinetic models of synthetic gene networks. Simulation, optimization and analysis tools will be available for Kbase users with Open Licenses.

In order to proceed from conceptual design to actual reaction, we will work on further integrating *Designer* and *Wiki.* Ideally, the model file built by *Designer* will be imported into the *Wiki.* We will develop the PHP scripts so that the user can quickly search the *Wiki* for component and reaction entries. The search can return a link or a pull-down menu that automatically shows a list of the different reactions of the parts that are presently included in the *Wiki.* In addition, to assist users to understand the *Designer* process, *Designer* can generate a special text file in addition to the model file that would contain the following: 1) a brief description of each reaction included in the model, 2) literature references and a description of where this type of kinetic expression has been used in previous biological models, and 3) additional relevant information including pertinent assumptions and kinetic approximations (e.g., assumption of pseudo steady state for Michaelis-Menten). Compiling this information in a text file will help the user understand the foundation of the model. In the absence of available quantitative information, the user will at least be given a clear sense of the specific interactions for which to search the literature and be able to specify ranges of kinetic constant values.

III.1.3 Augment available biochemical processes in Designer.

It is challenging to formalize interactions between biomolecules in an algorithm that is standardized and widely applicable. We will continue examining the first principles used in *Designer* and continue building it to accommodate possible synthetic biological systems. Following is a list of limitations we have identified and plan to work on.

- *Designer* does not yet understand polycistronic DNA, although there are applications where a single promoter expresses multiple proteins.
- *Designer* does not understand some of the more sophisticated regulation strategies, such as DNA looping, multiple types of sigma factors, or the expression of different RNA polymerases.
- *Designer* does not handle enzymatic reactions that often play regulatory roles (e.g., phosphorylation).
- The current modeling formalism is not set up to handle transcription from multiple copy plasmids.
- Transcription from different promoters, each residing on a separate plasmid, cannot be accommodated (e.g., ColE1 plasmid vs. a p15A plasmid).

Currently, a user needs to manually construct the model to include these mechanisms. This is possible in SynBioSS *Wiki*, using the web-based tools for constructing reactions. It is also possible in SynBioSS DS, with the GUI enabling the addition, deletion and edition of reactions in any reaction network. Our goal is to improve Designer to take into account these often encountered molecular biology processes.

III.1.4 Software Architecture and Sharing with Kbase.

In order to best connect with Kbase, we will continue making our codes available through synbioss.SourceForge.net. Sourceforge is the *de facto* forum for dissemination of open source code. All of the data will be deposited in SynBioSS *Wiki* and mirrored in potential Kbase resources. In making our software available, we use the BSD license format from OpenSource.org. We have moved Hy3S into SynBioSS, making available the executables for Windows using Intel compilers and proposing to make them available for MacOS machines. To enable rapid cross-platform development, we have leveraged several open-source technologies. The *DS* interface uses the GTK+ toolkit (http://www.gtk.org), and the interaction is scripted in Python (http://www.python.org). For all our models, we use SBML, the commonly used model representation format, enabling the integration of our tools with other multiscale modeling efforts. The SBML parser is used that uses libSBML (http://www.sbml.org/software/libsbml), SymPy (http://code.google.com/p/sympy) for evaluation of kinetic law expressions, and Unum (http://home.scarlet.be/be052320/Unum.html) for units validation and conversion. All of our models and the results will be published online in SBML and NetCDF formats. We have done this with all our previous simulations, having deposited all models on synbioss.sourceforge.net. We will explore the CellML format.

Help documentation is already available for SynBioSS on the website. We have prepared manuals with test cases and walkthrough exercises. We will continue updating this information.

We will conform to community recognized standard formats for synthetic biological parts and gene regulatory network models. Although there are currently no universally accepted standards in synthetic biology, a consortium of scientists, similar to the Genomic Standards Consortium (GSC), is working on the development of minimum information about synthetic constructs, in a process similar to the development of the genome sequence standards (MIGS). We are actively involved in this process.

III.2 Develop multiscale models of gene expression systems.

We have developed a general framework for modeling gene regulatory networks, focusing on the multiple time scales that emerge in these systems. We will quantify the limits of correspondence between the various multiple scales and develop hybrid algorithms that bridge these scales. Of particular importance is the manner in which information about molecular populations and reactions are shared between partitions. Specifically, since population information is shared between discrete and continuous modeling partitions [53], all the numerical integrators mush share a common time-step and synchronously proceed through time. Also, reactions may dynamically move from one partition to another (stochastic and deterministic reaction events as detailed in [53]) as time progresses, allowing such a partitioned algorithm to adapt to temporal stiffness and still integrate through simulation-time quickly. We propose a framework for identifying the limits between macroscopic-deterministic equations and stochastic simulation. We also propose to develop analysis and optimization tools for multiscale reaction networks, as well as tools for

conducting simulations on high-performance computing systems. ***Following are three proposed subtasks for multiscale simulations and optimization of gene networks.***

III.2.1 Develop stochastic-deterministic hybrid algorithms.

Consider *N* chemical species in a spatially uniform mixture of volume *V* involved in *M* reactions. Consider the N-dimensional vector $\underline{X}$ of the number of molecules of species *i,* X_i, and v_{ij} the stoichiometric coefficient of the ith species involved in the jth reaction (v_{ij}<0 for reactants, > for products, 0 for species not involved in j reaction). Traditionally, the chemical dynamics are solved using a continuous-deterministic formalism with ordinary differential equations:

$$\frac{d\left(X_i / N_A V\right)}{dt} = f_i(\underline{X}) \qquad (1)$$

where N_A is Avogadro's number and the functions *f* describe the reaction rate equations for each of the species. On the other hand, for systems away from the thermodynamic limit, the master equation formalism can be used in principle. The macroscopic concentration $\underline{C}(t)=\langle\underline{X}\rangle/N_A/V$ where $\left\langle \underline{X} \right\rangle = \int \underline{X} P(\underline{X}, t) d\,\underline{X}$, with $P(\underline{X}, t)$ being defined as the master probability that the system is a $\underline{X}$ at time t. We can write

$$\frac{d\left\langle \underline{X} \right\rangle}{dt} = \int \underline{X} \frac{\partial P(\underline{X}, t)}{\partial t} d\,\underline{X} \qquad (2)$$

or, using the master equation formalism [63]

$$\begin{aligned} \frac{d\left\langle \underline{X} \right\rangle}{dt} &= \int \underline{X} \left\{ T(\underline{X} \mid \underline{X}') P(\underline{X}', t) - T(\underline{X}' \mid \underline{X}) P(\underline{X}, t) \right\} d\,\underline{X}\, d\,\underline{X}' \\ &= \int (\underline{X}' - \underline{X}) T(\underline{X}' \mid \underline{X}) P(\underline{X}', t) d\,\underline{X}\, d\,\underline{X}' \end{aligned} \qquad (3)$$

where $T(\underline{X}|\underline{X}')$ is the transitional probability that the state will move to state $\underline{X}$ from $\underline{X}'$.

What is clear is that in the case of simple, linear kinetics, the solution of ODEs corresponds to the average of the master probability distribution. The variance of the distribution decreases with an increasing number of molecules, vanishing at the thermodynamic limit. What is not clear is the exact thresholds for the number of molecules and reaction propensities above which the higher moments of the probability distribution can be ignored. Especially in the case of nonlinear dynamics, it will be useful to theoretically explore the correspondence between macroscopic-deterministic and stochastic dynamics. We are intent on rigorously defining these limits and developing the algorithm for crossing the boundary accurately.

Van Kampen [63] and Moyal [64] provide an initial answer for the simplest of systems $A \rightarrow B$, but to our knowledge there is no systematic investigation for complex, nonlinear reacting systems. In what follows, we initiate the development of the framework for stochastic and deterministic simulations.
We define the derivative characteristic function of $\underline{X}$(t) as

$$L(\underline{\Theta} \mid \underline{X}, t) = \lim_{\tau \to 0} \frac{\Lambda(\underline{\Theta}, t+\tau \mid \underline{X}, t) - 1}{\tau} \qquad (4)$$

where $\Lambda(\underline{\Theta}, t+\tau \mid \underline{X}, t) = \int e^{-i\underline{\Theta}(\underline{X}'-\underline{X})} dP(\underline{X}', t+\tau \mid \underline{X}, t)$ (7) and $dP = T(\underline{X}' \mid \underline{X}) d\,\underline{X}'$ (8)

If L and Λ are analytical functions of Θ, we can write

$$L(\underline{\Theta} \mid \underline{\Xi}, t) = \lim_{\tau \to 0} \frac{1}{\tau} \sum_{k=1}^{\infty} \frac{(i\underline{\Theta})^k}{k!} \int (\underline{X} - \underline{\Xi})^k dP(\underline{X}, t+\tau \mid \underline{\Xi}, t) = \sum_{k=1}^{\infty} \frac{(i\underline{\Theta})^k}{k!} a_k(\underline{\Xi}, t) \qquad (5)$$

where the derivative moments are defined as

$$a_k(\underline{\Xi}, t) = \lim_{\tau \to 0} \frac{1}{\tau} \int (\underline{X} - \underline{\Xi})^k dP(\underline{X}, t+\tau \mid \underline{\Xi}, t) \qquad (6)$$

When they exist the derivative moments can be written as

$$a_k(\underline{X}) = \int (\underline{X}' - \underline{X})^k T(\underline{X}' \mid \underline{X}) d\,\underline{X}' \qquad (7)$$

Returning to the relationship between the macroscopic equation and the master equation we write

$$\frac{d\left\langle \underline{X} \right\rangle}{dt} = \int a_1(\underline{X}) P(\underline{X}, t) d\,\underline{X} = \left\langle a_1(\underline{X}) \right\rangle \qquad (8)$$

It is perhaps interesting to note that after Moyal developed this theory for single component systems in 1949, he started exploring the extension to a higher dimension state vector $\underline{X}$ but noted that "However, no applications have yet been found for such relations, and we shall not therefore pursue this line of though any further" [64]. We believe that these applications are ripe for investigating the correspondence between macroscopic deterministic and mesoscopic stochastic kinetics for multicomponent, nonlinear systems. Using the higher-order derivative moments, we can calculate the higher moments of the master probability, starting with the variance in $\underline{X}$(t) and progressively computing the third, fourth, etc. order derivative moments. With a calculation of these moments, we will determine the appropriate modeling formalism adaptively.

We will package all the methods in a single software suite, SynBioSS *DS*. Armed with increasingly fast supercomputers, it is now becoming practical to numerically simulate complex gene networks at the highest possible molecular resolution and connect to the cell population level. With multiscale simulations, we can gain invaluable insight into the molecular-level events responsible for tight inducible control.

We should note that a major shortcoming of the proposed work is that its focus is on spatially homogeneous systems. Diffusion-limited phenomena may be important for the emergence of biological phenotypes in bacterial systems, and are certainly of crucial importance in eukaryotic cells. There is certainly need for the development of multiple-length scale models of reacting/diffusing biomolecular system and we plan on tackling this interesting area in the future.

III.2.2 Develop and Build SynBioSS Optimizer.

We will develop a tool that wraps around SynBioSS DS and optimizes the strength of kinetic constants, comparing the simulated dynamic phenotype to the user-defined targeted one.

Previously, we presented an optimization method based on simulated annealing (SA) to locate combinations of kinetic parameters and determine the values that produce a desired dynamic behavior[59].

Simulated annealing is an optimization scheme for systems with many degrees of freedom. It is one of the most popular and widely-used optimization algorithms due to its versatility and wide-applicability[65, 66]. Simulated annealing draws an analogy between a multi-dimensional optimization problem and the minimization of energy that occurs within a metal as it cools and its atoms optimize their positions to minimize Gibbs free energy. In simulated annealing, perturbations to the model replace atomic vibration, a problem-specific quality metric takes the place of energy, and a virtual temperature is lowered to "anneal" the system towards the optimal value of that quality metric. In our work, a mechanistic, stochastically-integrated model of a gene network is used as the foundation for a Metropolis Monte Carlo / simulated annealing optimization scheme.

We recognize that locating the global optimum behavior of a gene network is of little value if the resulting set of optimum parameters do not correspond to the kinetic parameters of genetic components available to the experimentalist. Therefore, we will seek to use our optimization scheme in combination with a particular gene network model to locate many sets of parameters that correspond to many different optima. The experimentalist will then be presented with a larger "menu" of putative systems that yield a desired network dynamic behavior, within a certain tolerable error.

We should note here, that quantitative information is often available for biomolecular interactions in the form of equilibrium constants. In such cases, we can assume that the forward rate of binding of a large protein to its DNA or RNA binding site is diffusion-limited, use the size of the protein to calculate its forward binding kinetic constant, and then use the equilibrium data to calculate the unbinding kinetic constant. We will use a first principles description of two particles diffusing in a three-dimensional space. The

Smoluchowski rate for a diffusion-limited reaction is $k_f=4\pi Da$, with a diffusion coefficient, D, and target size, a. The diffusion coefficient of a free particle of diameter d in a homogeneous fluid undergoing a random three-dimensional walk may be calculated using Einstein's relation, $D=(k_bT)/(3\pi\eta d)$.

We can now obtain the kinetic rate of association in terms of temperature, T, the viscosity of the fluid, η, and the ratio between the diameter of the protein and its target site, a/d, so that $k_f =(T)(a/d)(4kb/3\ \eta)$.

If we assume that fluid is aqueous so that $\eta = 1\times10^{-3}$ Pa s and the temperature is T = 30°C, then the association rate is $k_f = 3.3590\times10^{9}$ (a/d) in units of $[M\ s]^{-1}$. The size of the protein will typically be much larger than the size of an effector molecule or a DNA binding site, and so we may assume that (a/d)= 0.1, leading to an approximate association constant of $k_f = 10^8\ [M\ s]^{-1}$. Consequently, if we lack kinetic data, we may approximate the backward kinetic constant as $k_b = 10^8\times Kd$.

Ultimately, of course, only experimental information can provide accurate information, but we think that even a good starting estimate may be enough to create inroads in quantitative understanding.

III.2.3 Develop Tools for Simulations on Supercomputers.

Practically, an important shortcoming of our approach will be the computational cost of the proposed algorithms. Compared to the solution of a system of ordinary differential equations, our algorithms will be mathematically accurate but significantly slower. Thus we propose to leverage TeraGrid and DOE computational resources, the highest-end resource available based on commodity clusters, Linux/Unix, and Globus hardware. Currently the TeraGrid provides more than 50 Teraflops of computing power, distributed at multiple sites as well as facilities and personnel for storing and managing more than one Petabyte of data (www.teragrid.org). We were awarded 800,000 CPU hours on TeraGrid computers for 2009, half of which are for runs with SynBioSS. The newest version of our code on SourceForge.net is developed to run on thousands of Linux TeraGrid computers using OpenMP. With these resources, multiscale simulations of hundreds of biomolecular reactions are possible.

We propose to create a web interface for inexperienced users to upload files generated by SynBioSS *DS* for conducting simulations on the Teragrid or DOE supercomputers. The first step is for a user to request Teragrid resources by simply filling out a web-form on Teragrid.org. Without writing a proposal, any U.S. principal investigator can obtain up to 200,000 CPU hours, enough to conduct serious simulations of complex gene networks. With DOE, the process is more competitive. We will explore opportunities in programs like INCITE to access DOE resources and make them available for the community.

With the web interface, we will build the process of logging in to the Teragrid, processing the uploaded files, running the simulations, making the results available for download through a website, and sending an email message to the user alerting them of the completion of the simulation. We have already made supercomputing runs a simple task within our group by building UNIX/Linux scripts that launch jobs on hundreds of processors from the convenience of our desktop computers.

III.3 Simulate and design novel synthetic networks.

Atmospheric levels of CO_2 are projected to reach levels unprecedented in recent geologic history. This will likely happen during our lifetime. The resulting impact on earth's climate is unpredictable and possibly catastrophic. Carbon-neutral energy sources must be explored and investments in new technology development are required. Cellulosic biomass-derived biofuels, such as ethanol, are attractive alternative transportation fuels that have the potential of stemming the increasing rates of emitted CO_2. Cellulosic biomass can be hydrolyzed into mixtures of sugars, typically glucose/xylose mixtures. Fermenting sugar mixtures to ethanol does not lend itself to inexpensive continuous, as opposed to batch and semi-batch processes, because fermenting microorganisms prefer to metabolize glucose first and then metabolize xylose and other sugars. This is a technological bottleneck for developing processes that are inexpensive enough to be used at a small scale and scalable enough for ethanol mass production.

III.3.1 Design and simulate the comparator gene network.

We propose to engineer smart ethanologenic bacteria that sense and compare relative amounts of different sugars and optimize their own metabolism to exhaust all sugar sources concurrently. A strain of smart bacteria can lend itself to flexible, continuous fermentation processes, potentially lowering investment and operating costs.

The trait we want to engineer in bacteria is the capacity to sense relative amounts of sugars and adapt to changes by optimizing their own metabolism to metabolize mixtures of sugars concurrently. Specifically, we will construct a synthetic genetic controller/comparator to facilitate continuous fermentation of glucose/xylose mixtures. The comparator is a vital part of any control system. This component compares a set point value and a measured value and determines which is larger. The comparator then sends an appropriate signal to the rest of the system. As an example, consider the comparator in any thermostat, which compares the measured and set point temperatures (two inputs) and sends signals to turn on the heat or the air-conditioning (two outputs).

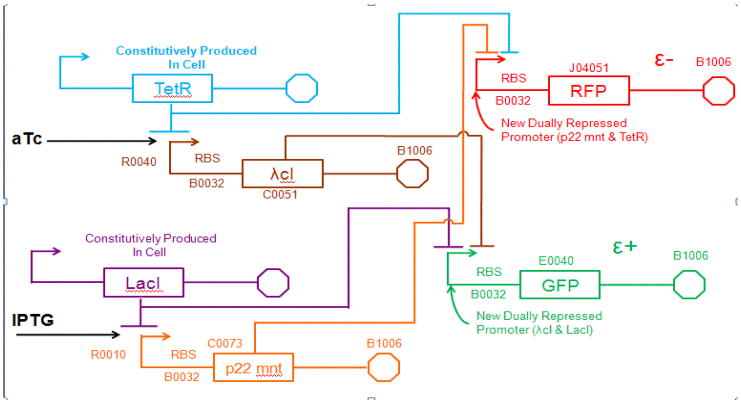


Figure 5. a) Schematic of our biological comparator with the six genes used to construct a synthetic controller. The device compares the levels of aTc and IPTG, two small organic molecules. It generates red fluorescence protein (RFP) is aTc>IPTG, and green FP (GFP) if aTc<IPTG and does nothing if aTc=IPTG.

In the genetic comparator, glucose and xylose amounts (two inputs) will be compared and the levels of proteins that transport these molecules inside the cell will be controlled correspondingly (two outputs). If glucose concentration is higher, the synthetic comparator/controller will repress expression of glucose transporter proteins and over-express xylose transporter proteins. If xylose concentration is higher, the synthetic comparator/controller will repress expression of xylose transporter proteins and overexpress glucose transporter proteins.

III.3.2 Simulate and optimize the dynamic behavior of the comparator.

A preliminary design of a genetic comparator is shown in Figure 5. As a proof-of-principle, we will model and simulate a synthetic biological device that compares the concentrations of two small molecules and controls expression of green fluorescence protein and red fluorescence protein. Specifically, a system involving six genes was designed using standard molecular biology techniques (Figure 5). For our proof-of-concept system, the two inputs (one representing the set point and one representing the measured value) are isopropyl β-D-1-thiogalactopyranoside (IPTG) and anhydrotetracycline (aTc), which are inducer molecules. These inputs interact with the LacI and TetR proteins which are constitutively produced by DH5alpha E. coli cells. Depending on the amounts of the two inducer molecules added to the system, either green fluorescent protein (GFP) or red fluorescent protein (RFP) are produced.

With SynBioSS *Designer* and SynBioSS *Wiki* we will generate a set of chemical reactions that model transcription, translation, degradation, induction steps in the comparator.

Starting with well-characterized components of the tetracycline, lactose, and p22 operons, affords the accurate definition and annotation of the biomolecular interactions that comprise the gene network.

With SynBioSS DS, we will simulate the dynamic behavior of the comparator. We will identify the key interactions. With simulated annealing we will focus on these interactions and determine the optimum values of kinetic constants to realize the comparison and the control mechanism.

III.3.3 Engineer the comparator for use in ethanologenic bacteria.

If successful, with proof-of-principle results at hand, we will construct a similar synthetic construct for bacteria to sense and utilize varying relative concentrations of xylose and glucose. After permeating

through the outer membrane of gram-negative bacteria, sugar molecules, such as glucose and xylose are recognized by specific protein receptors. These periplasmic receptors partner with cytoplasmic membrane components to transport the sugar molecules inside the cell. We will engineer the comparator to differentially express cytoplasmic transporter proteins to control the transport and metabolism of xylose and glucose in bacteria. For the outputs of the actual comparator we will express a xylose transporter (XylE or XylFGH) and a glucose transporter, II-BGl, which is part of the bacterial phosphoenolpyruvate:glycose phosphotransferase system (PTS) [67-75]. The PTS consists of many interacting cytoplasmic and membrane proteins that catalyze the phosphorylation and translocation of sugar substrates across the cell membrane [67-75].

When using *E. coli* we will also replace the native crp gene with a mutant (crp*), which is known to alleviate glucose repression of xylose transport [72]. The transformed cells will overproduce the xylose transporter and shut down expression of the glucose transporter, when the concentration of xylose is higher than the concentration of glucose, having calibrated the concentrations that are considered equal.

Coupling these synthetic gene and metabolic networks will represent an important advance in computational biology efforts to provide mechanistic explanations of biomolecular phenomena.

The strains used and more details on the molecular biology are beyond the scope of this proposal.

IV. Role of Participants

Prof. Kaznessis will be directing the project, coordinating with the co-PIs, supervising the graduate students of the project, and coordinating with Minnesota Supercomputing Institute personnel. With a Ph.D. in computational and statistical polymer physics (Notre Dame) and a post-doctoral fellowship in a computational biology and bioinformatics group (Pfizer), the PI is experienced in the development of multiscale models, the development of software tools and the numerical simulation of biomolecular interactions and gene regulatory networks.

Prof. Reidel will be assisting with supervising the project personnel. He received his PhD and his MS. in Electrical Engineering at Caltech and his BEng in Electrical Engineering with a Minor in Mathematics at McGill University. His PhD dissertation, titled "Cyclic Combinational Circuits", received the Charles H. Wilts Prize for the best doctoral research in EE at Caltech. His research strives for new, transformative approaches to design automation. A broad theme is the application of expertise from an established field (digital circuit design) to new areas (nanotechnology and synthetic biology). A specific theme that cuts across these domains is constructing and deconstructing probabilistic behavior. Riedel and his group are developing techniques for designing biochemical pathways that produce different combinations of molecular types according to specified probability distributions.

Prof. Schmidt-Dannert will be assisting with supervising the project personnel as well as assist with direction to the project. After obtaining her Ph.D. at the National Research Center for Biotechnology (GBF, now Helmholtz Center for Infectious Research) in Braunschweig, she went to the University of Stuttgart to lead the Molecular Biotechnology Group in the Institute of Technical Biochemistry. She won a habilitation-fellowship from the German Science Foundation for research on in vitro pathway evolution and joined Frances Arnold Group at the California Institute of Technology. At the University of Minnesota, her research interests are on biosynthetic pathway engineering, natural products biosynthesis and protein engineering using rational and evolutionary design strategies.

It is worth noting that the PIs directed the Minnesota iGEM team. iGEM stands for International Genetically Engineered Machines and it is an international synthetic biology competition. It is worth noting that team Minnesota received a Gold Medal for its work on SynBioSS in the 2009 competition.

Funds to support three graduate students are requested.

Ben Swiniarski is Ph.D. candidate in the Department of Chemical Engineering and Materials Science. He will be the primary person responsible for the continued development of SynBioSS algorithms and tools. He is currently in his second year. He has worked extensively on SynBioSS *Designer* and *Wiki*.

Katherine Volzing is Ph.D. candidate in the Department of Chemical Engineering and Materials Science. She has a BS (magna cum laude) in Biochemistry, Molecular Biology and Genetics from the University of Minnesota. She will conduct numerical simulations of the comparator.

A graduate student (TBD) in the Department of Electrical and Computer Engineering will work on the development and maintenance of SynBioSS *Designer* and *Wiki*.

Personnel from the Minnesota Supercomputing Institute will assist in developing and maintaining the data repositories, and the web-interfaces. In his letter of support, the MSI Director Thomas Jones details the expertise of the MSI personnel.

V. Summary: SynBioSS as Kbase component

Imagine trying to engineer a bacterial strain that senses relative amounts of various sugars and turns on/off genes to produce biofuels with maximum yield. Now imagine having a software tool that can assist in engineering the synthetic strain quickly. The proposed activities will result in modeling tools for computer-aided construction of complex synthetic biological systems; inventories of data, open to modification and improvement by the community; multiscale modeling algorithms for numerical simulations of gene networks; tools for heuristic design and optimization of synthetic constructs. The activities will also produce novel synthetic gene regulatory networks with applications in biofuels production. We believe that SynBioSS as a Kbase component can become a tool-framework to provide valuable scientific insight and guide biological engineering.

V. Project Timeline

Task 1. Develop and disseminate software tools for synthetic biology applications.

Subtask 1.1. Integrate SynBioSS with Parts Registry and BioBricks format.

Subtask 1.2. Develop standardized tools from conceptual design to reaction networks to DNA.

Subtask 1.3. Augment available biochemical processes in *Designer*.

Subtask 1.4. Software architecture and sharing with Kbase.

Task 2. Develop multiscale dynamic models of gene expression systems.

Subtask 2.1. Develop stochastic-deterministic hybrid algorithms.

Subtask 2.2. Develop SynBioSS Optimizer.

Subtask 2.3. Develop Tools for Simulations on Supercomputers.

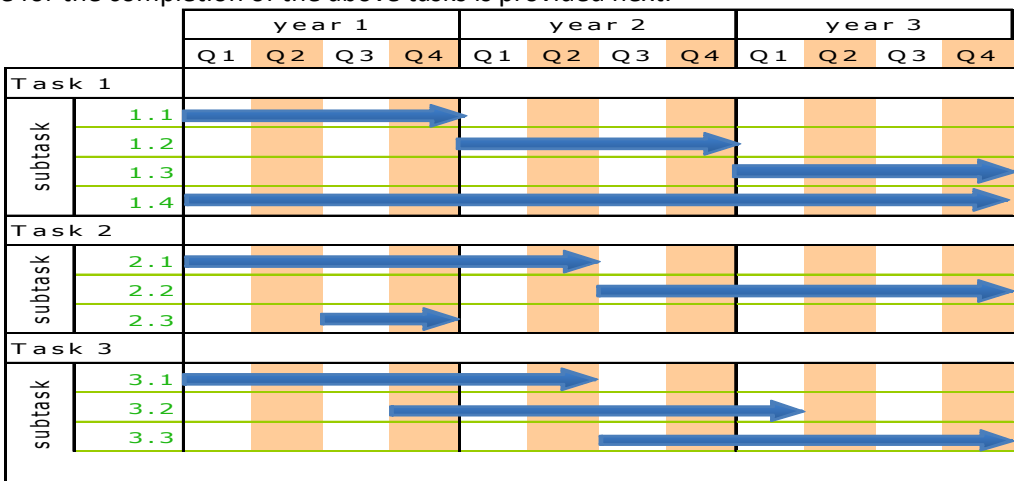
Task 3. Simulate and design novel synthetic networks.

Subtask 3.1. Design and simulate the comparator gene network.

Subtask 3.2. Simulate and optimize the dynamic behavior of the comparator.

Subtask 3.3 Engineer the comparator for use in ethanologenic bacteria.

The timeline for the completion of the above tasks is provided next.



Appendix 3

References Cited

1. Hill, A.D., et al., *SynBioSS: the synthetic biology modeling suite.* Bioinformatics, **2008**. 24(21): p. 2551-3.
2. Kaznessis, Y.N., *Models for synthetic biology.* BMC Syst Biol, **2007**. 1: p. 47.
3. Kaznessis, Y.N., *Computational methods in synthetic biology.* Biotechnol J, **2009**. 4(10): p. 1392-405.
4. Gardner, T.S., C.R. Cantor, and J.J. Collins, *Construction of a genetic toggle switch in Escherichia coli.* Nature, **2000**. 403(6767): p. 339-42.
5. Elowitz, M.B. and S. Leibler, *A synthetic oscillatory network of transcriptional regulators.* Nature, **2000**. 403(6767): p. 335-8.
6. Arkin, A.P., *Synthetic cell biology.* Curr Opin Biotechnol, **2001**. 12(6): p. 638-44.
7. Guet, C.C., et al., *Combinatorial synthesis of genetic networks.* Science, **2002**. 296(5572): p. 1466-70.
8. Alon, U., *Biological networks: the tinkerer as an engineer.* Science, **2003**. 301(5641): p. 1866-7.
9. Benner, S.A., D. Hutter, and A.M. Sismour, *Synthetic biology with artificially expanded genetic information systems. From personalized medicine to extraterrestrial life.* Nucleic Acids Res Suppl, **2003**(3): p. 125-6.
10. Benner, S.A., *Synthetic biology: Act natural.* Nature, **2003**. 421(6919): p. 118.
11. Ball, P., *Synthetic biology: starting from scratch.* Nature, **2004**. 431(7009): p. 624-6.
12. Blake, W.J. and F.J. Isaacs, *Synthetic biology evolves.* Trends Biotechnol, **2004**. 22(7): p. 321-4.
13. Ferber, D., *Synthetic biology. Microbes made to order.* Science, **2004**. 303(5655): p. 158-61.
14. Benner, S.A. and A.M. Sismour, *Synthetic biology.* Nat Rev Genet, **2005**. 6(7): p. 533-43.
15. Check, E., *Synthetic biology: designs on life.* Nature, **2005**. 438(7067): p. 417-8.
16. Church, G.M., *From systems biology to synthetic biology.* Mol Syst Biol, **2005**. 1: p. 2005 0032.
17. Levskaya, A., et al., *Synthetic biology: engineering Escherichia coli to see light.* Nature, **2005**. 438(7067): p. 441-2.
18. Pennisi, E., *Synthetic biology. Synthetic biology remakes small genomes.* Science, **2005**. 310(5749): p. 769-70.
19. Sprinzak, D. and M.B. Elowitz, *Reconstruction of genetic circuits.* Nature, **2005**. 438(7067): p. 443-8.
20. Andrianantoandro, E., et al., *Synthetic biology: new engineering rules for an emerging discipline.* Mol Syst Biol, **2006**. 2: p. 2006 0028.
21. Chopra, P. and A. Kamma, *Engineering life through Synthetic Biology.* In Silico Biol, **2006**. 6(5): p. 401-10.
22. de Lorenzo, V., L. Serrano, and A. Valencia, *Synthetic biology: challenges ahead.* Bioinformatics, **2006**. 22(2): p. 127-8.
23. Heinemann, M. and S. Panke, *Synthetic biology--putting engineering into biology.* Bioinformatics, **2006**. 22(22): p. 2790-9.
24. Isaacs, F.J., D.J. Dwyer, and J.J. Collins, *RNA synthetic biology.* Nat Biotechnol, **2006**. 24(5): p. 545-54.
25. Pleiss, J., *The promise of synthetic biology.* Appl Microbiol Biotechnol, **2006**. 73(4): p. 735-9.
26. Alon, U., *Network motifs: theory and experimental approaches.* Nat Rev Genet, **2007**. 8(6): p. 450-61.
27. Cox, R.S., 3rd, M.G. Surette, and M.B. Elowitz, *Programming gene expression with combinatorial promoters.* Mol Syst Biol, **2007**. 3: p. 145.
28. Ball, P., *Synthetic biology: Designs for life.* Nature, **2007**.
29. Forster, A.C. and G.M. Church, *Synthetic biology projects in vitro.* Genome Res, **2007**. 17(1): p. 1-6.
30. Serrano, L., *Synthetic biology: promises and challenges.* Mol Syst Biol, **2007**. 3: p. 158.

31. Arkin, A., *Setting the standard in synthetic biology.* Nat Biotechnol, **2008**. 26(7): p. 771-4.
32. Tan, C., et al., *A synthetic biology challenge: making cells compute.* Mol Biosyst, **2007**. 3(5): p. 343-53.
33. Hucka, M., et al., *The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models.* Bioinformatics, **2003**. 19(4): p. 524-31.
34. Blake, W.J., et al., *Noise in eukaryotic gene expression.* Nature, **2003**. 422(6932): p. 633-7.
35. Elowitz, M.B., et al., *Stochastic gene expression in a single cell.* Science, **2002**. 297(5584): p. 1183-6.
36. Oppenheim, I., K.E. Shuler, and G.H. Weiss, *Stochastic Theory of Nonlinear Rate Processes with Multiple Stationary States.* Physica A, **1977**. 88(2): p. 191-214.
37. Hanggi, P., K.E. Shuler, and I. Oppenheim, *On the Relations between Markovian Master-Equations and Stochastic Differential-Equations.* Physica a-Statistical Mechanics and Its Applications, **1981**. 107(1): p. 143-157.
38. Manninen, T., M.L. Linne, and K. Ruohonen, *Developing Ito stochastic differential equation models for neuronal signal transduction pathways.* Comput Biol Chem, **2006**. 30(4): p. 280-91.
39. Cao, Y., D.T. Gillespie, and L.R. Petzold, *The slow-scale stochastic simulation algorithm.* J Chem Phys, **2005**. 122(1): p. 14116.
40. Cao, Y., D.T. Gillespie, and L.R. Petzold, *Adaptive explicit-implicit tau-leaping method with automatic tau selection.* J Chem Phys, **2007**. 126(22): p. 224101.
41. Chatterjee, A., et al., *Time accelerated Monte Carlo simulations of biological networks using the binomial tau-leap method.* Bioinformatics, **2005**. 21(9): p. 2136-7.
42. Gillespie, D.T., S. Lampoudi, and L.R. Petzold, *Effect of reactant size on discrete stochastic chemical kinetics.* J Chem Phys, **2007**. 126(3): p. 034302.
43. Gillespie, D.T., M. Roh, and L.R. Petzold, *Refining the weighted stochastic simulation algorithm.* J Chem Phys, **2009**. 130(17): p. 174103.
44. Lampoudi, S., D.T. Gillespie, and L.R. Petzold, *The multinomial simulation algorithm for discrete stochastic simulation of reaction-diffusion systems.* J Chem Phys, **2009**. 130(9): p. 094104.
45. Peles, S., B. Munsky, and M. Khammash, *Reduction and solution of the chemical master equation using time scale separation and finite state projection.* J Chem Phys, **2006**. 125(20): p. 204104.
46. Munsky, B. and M. Khammash, *Transient analysis of stochastic switches and trajectories with applications to gene regulatory networks.* IET Syst Biol, **2008**. 2(5): p. 323-33.
47. Khammash, M., *Reverse engineering: the architecture of biological networks.* Biotechniques, **2008**. 44(3): p. 323-9.
48. Vlachos, D.G., *Temporal coarse-graining of microscopic-lattice kinetic Monte Carlo simulations via tau leaping.* Phys Rev E Stat Nonlin Soft Matter Phys, **2008**. 78(4 Pt 2): p. 046713.
49. Chatterjee, A. and D.G. Vlachos, *Multiscale spatial Monte Carlo simulations: multigriding, computational singular perturbation, and hierarchical stochastic closures.* J Chem Phys, **2006**. 124(6): p. 64110.
50. Chatterjee, A., M.A. Katsoulakis, and D.G. Vlachos, *Spatially adaptive grand canonical ensemble Monte Carlo simulations.* Phys Rev E Stat Nonlin Soft Matter Phys, **2005**. 71(2 Pt 2): p. 026702.
51. Gillespie, D.T., L.R. Petzold, and Y. Cao, *Comment on "nested stochastic simulation algorithm for chemical kinetic systems with disparate rates" [J. Chem. Phys. 123, 194107 (2005)].* J Chem Phys, **2007**. 126(13): p. 137101; discussion 137102.
52. Gillespie, D.T., et al., *The subtle business of model reduction for stochastic chemical kinetics.* J Chem Phys, **2009**. 130(6): p. 064103.
53. Salis, H. and Y. Kaznessis, *Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions.* J Chem Phys, **2005**. 122(5): p. 54103.
54. Salis, H. and Y.N. Kaznessis, *An equation-free probabilistic steady-state approximation: dynamic application to the stochastic simulation of biochemical reaction networks.* J Chem Phys, **2005**. 123(21): p. 214106.

55. Sotiropoulos, V. and Y.N. Kaznessis, *An adaptive time step scheme for a system of stochastic differential equations with multiple multiplicative noise: chemical Langevin equation, a proof of concept.* J Chem Phys, **2008**. 128(1): p. 014103.
56. Sotiropoulos, V., et al., *Model reduction of multiscale chemical langevin equations: a numerical case study.* IEEE/ACM Trans Comput Biol Bioinform, **2009**. 6(3): p. 470-82.
57. Salis, H., V. Sotiropoulos, and Y.N. Kaznessis, *Multiscale Hy3S: hybrid stochastic simulation for supercomputers.* BMC Bioinformatics, **2006**. 7: p. 93.
58. Tuttle, L.M., et al., *Model-driven designs of an oscillating gene network.* Biophys J, **2005**. 89(6): p. 3873-83.
59. Tomshine, J. and Y.N. Kaznessis, *Optimization of a stochastically simulated gene network model via simulated annealing.* Biophys J, **2006**. 91(9): p. 3196-205.
60. Salis, H. and Y.N. Kaznessis, *Computer-aided design of modular protein devices: Boolean AND gene activation.* Phys Biol, **2006**. 3(4): p. 295-310.
61. Sotiropoulos, V. and Y.N. Kaznessis, *Synthetic tetracycline-inducible regulatory networks: computer-aided design of dynamic phenotypes.* BMC Syst Biol, **2007**. 1: p. 7.
62. Smolke, C.D., *Building outside of the box: iGEM and the BioBricks Foundation.* Nat Biotechnol, **2009**. 27(12): p. 1099-102.
63. Li, H., et al., *Algorithms and software for stochastic simulation of biochemical reacting systems.* Biotechnol Prog, **2008**. 24(1): p. 56-61.
64. Lampoudi, S., D.T. Gillespie, and L.R. Petzold, *Effect of excluded volume on 2D discrete stochastic chemical kinetics.* J Comput Phys, **2009**. 228(10): p. 3656-3668.
65. Kirkpatrick, S., C.D. Gelatt, and M.P. Vecchi, *Optimization by Simulated Annealing.* Science, 1983. 220(4598): p. 671-680.
66. Kirkpatrick, S., Optimization by Simulated Annealing - Quantitative Studies. Journal of Statistical Physics, 1984. 34(5-6): p. 975-986.