


CS486C – Senior Capstone Design in Computer Science

Project Description

Project Title: GUI and web access design for parallelized BioNetFit software to model biomolecular interactions	
	Sponsor Information: Dr. Richard Posner, Department of Biological Sciences, NAU Email: Richard.Posner@nau.edu , Phone: +1 928 523 4209
	Dr. William Hlavacek, Theoretical Division, Los Alamos National Laboratory Email: wish@lanl.gov , Phone: +1 505 665-1355
	Dr. Abolfazl Razi, School of Informatics, Computing, and Cyber Systems, NAU Email: Abolfazl.razi@nau.edu , Phone: +1 928 523 8078

Project Overview:

As the field of big data informatics has matured, computers and software have become the driving force in nearly every scientific discipline. For instance, the **Illumina Company** recently began offering commercial whole-genome sequencing at the affordable cost of only \$1000 per person. Advancements like these present a unique opportunity for computer scientists to analyze data, evaluate biological hypotheses and generate new insights in ways that were not possible a decade ago.

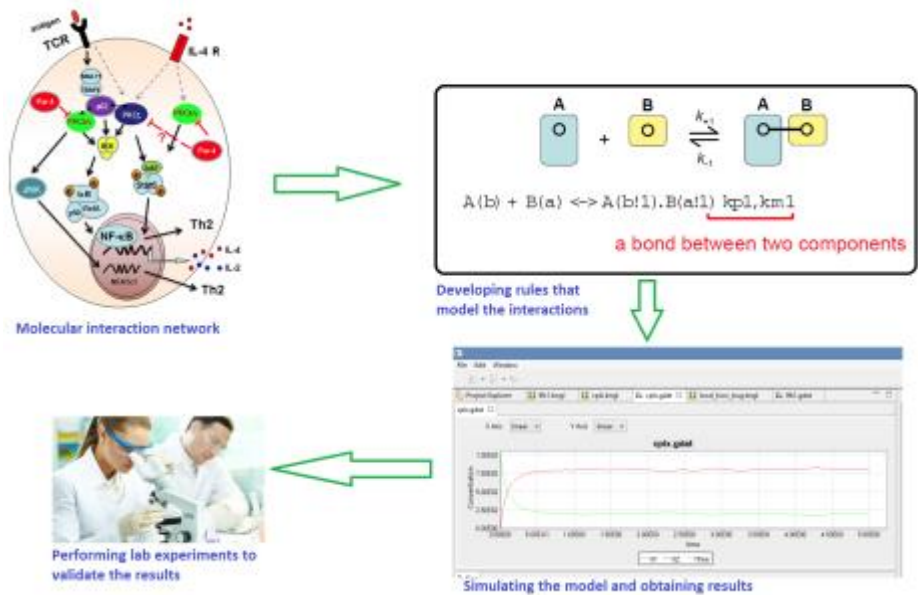


Figure 1 Work flow in biomolecular interaction hypothesis checking. Biological knowledge is used to develop a rule-based model. The model with initial conditions is simulated to predict the results of experiments. The results can be validated in the lab.

The informatics area that our group is working in is scientific simulation. One of the costliest and most time-consuming aspects of molecular level biology and chemistry is the “wet lab” experiments; hundreds or thousands of experiments with varying parameters might be required to arrive at some result. Our group of researchers at NAU, in collaboration with Los Alamos National Laboratory (LANL), has developed a set of computer-based automated tools to *digitally simulate* molecular chemical experiments; the reduction in time and cost has the potential to vastly accelerate the rate of discovery in chemical and biological sciences. Check out <http://bionetfit.nau.edu/> and <http://bionetgen.org/index.php/Tutorials> for more information.

We have already developed a powerful set of computation tools to simulate a wide range of experiments. The aim of this project is to develop a powerful and effective Web GUI that provides researchers anywhere with streamlined access to these tools. Researchers will access this secure web portal, will create and configure an experiment, run the simulation, and view and download the results.

Students will not need to have prior knowledge of molecular chemistry or biology; the client will provide the necessary expertise in these areas.

Further Project Details

The simulation core of the project is based around the BNGL language (<http://bionetgen.org/index.php/Tutorials>), which is commonly used by the systems biology community to write formal representations of biomolecular interactions. A BNGL model includes a set of molecule types, their properties, and their initial concentrations as well as a set of rules that govern these interactions. We then use NFsim, an open source C++ simulation engine (<https://github.com/msneddon/nfsim>), to analyze BNGL models and obtain simulation results. We also use RuleBender (<http://visualizlab.org/rulebender/>) as a graphical user interface to run NFsim; this basically converts the BNGL *models* into XML *files* which is the required input format for NFsim. Rulebender also visualizes the results provided by NFsim.

A very common use case is a scientist wanting to run essentially the same experiment, trying many different parameters. Unfortunately, to run a model with multiple parameter sets, we currently need to develop multiple models and manually execute each model using RuleBender and then validate the results. Obviously, this can be very tedious and begs for automation! We have currently developed a **command line** tool called BioNetFit to help with this process. The aim of this Capstone project is therefore to develop an elegant **graphical user interface** for BioNetFit to drive this entire exploratory process, designed specifically for non-technical scientist end-users. Some key features will include:

- A secure Web2.0 portal easily accessible from any desktop browser. The interface provides the ability to define and configure new experiments, as well as editing and re-running existing experiments within a registered user's "lab book".
- Able to configure the modeling run using a predefined pool of parameter sets; the system should provide some "intelligence" in helping to choose the best set of parameters that satisfy a predefined expected criterion.
- To provide a computational backend, the product should have the capability of running multiple instances of the simulation system within the *NAU Monsoon High performance computing cluster*.
- As a potential advanced feature, the system could allow adding additional computing resources; users could select which compute engine a particular model runs on.

Knowledge, skills, and expertise required for this project:

- a) **Programming Skills:** Solid general programming skills will be required, at a normal level of competency expected for graduating seniors. The programming language/frameworks for portal development are left open, and should be selected by the team as part of initial design.
- b) **Web development:** Competence in web development using Web2.0 techniques and frameworks will be required. The team will be expected to quickly fill missing gaps in information as the project progresses.

- c) **High-Performance Computing:** One core part of the program is to use parallel computing and high-performance computing (HPC) resources. Therefore, experience with NAU's HPC resources (viz., Monsoon and SLURM scripting) is a plus. However, the sponsor has significant technical resources and will provide any required training and information.
- d) **Biological Knowledge:** No prior knowledge about biomolecular interactions is required and the sponsors will provide enough information to develop the software.

Equipment Requirements:

- e) There is no specific equipment required for this project other than laptops and desktop computers. We encourage you to use your own laptops and computers to implement the software. However, a limited number of desktop computers can be provided by the sponsors and the department if required.

Software and other Deliverables:

- 1- **Web2.0 BioNetFit portal:** This is the main outcome of this project. The software should be bug-free and user-friendly. It is expected to have parallel computing capability through communication with clusters, such as NAU's Monsoon cluster.
- 2- **User manual:** Students will prepare a user manual, providing both editable (e.g. MSWord) and archival PDF versions.
- 3- **Final report:** A strong as-built report detailing the design and implementation of the product in a complete, clear and professional manner. This document should provide a strong basis for future development of the product.
- 4- **Source code:** Complete professionally-documented codebase, delivered both as a repository in GitHub, BitBucket, or some other version control repository; and as a physical archive on a USB drive.