What is evoGEM?

evoGEM is a system for automating the design of synthetic biology devices. This system was created to make the design of bio-circuits and the manipulation of iGEM (the international Genetically Engineered Machine Competition) bio-bricks more efficient and rubout.

Bio-bricks are small DNA fragments that deliver one, and only one, function when inserted into living cells. By concatenating different permutations of these bio-bricks into bio-circuits, one can introduce novel and exciting functions into Escherichia coli (E.coli) bacteria, turning the bacteria into small programmable machines.

The evoGEM system uses a unique Evolutionary Strategies (ES) approach in order to test designs engineered by the machine itself. Through the basic principles of adaptation and selection, the system iteratively increases the fitness of the population to perform the required task. Successful candidates are propagated through the generations and are altered to better and better fit the required task. The highest fit individuals are available as designs for the expert biologist to inspect and attempt in the lab.

By combining data mining techniques, agent based simulation for the evaluation and a user-oriented graphical interface our software presents a cutting edge in systems biology. Using several hundreds of designs the system has shown great potential for success and has been able to engineer simple bio-circuits as well as analogues of human made systems.

Agent Based Simulation

A powerful agent based system supplies the main pillar for this system. The individual biological agents in this system operate only according to local, simple, rules and the resulting behavior observed emerges by itself.

The agent based model has been inspired by social creatures such as ants and bees. These display fairy simple behavior as individuals, but act like a “super-organism” when a critical number of them comes together. Our system applies the same principle to organic and biological molecules since these have a limited scope by their nature, but when they come together the amazing environment known as the intracellular space comes to being.

The different molecules the system accounts for are the DNA bio-circuits, the RNA polymerases, ribosomes, messenger RNA transcripts and any other proteins that are encoded by the bio-bricks in the simulation. These bio-agents interact in order to simulate the intracellular environment and the resulting reactions of the bio-circuits inside the E.coli bacterium. This allows biologists to quickly and efficiently inspect designs in silico before implementing them in vivo (in a living organism). This powerful simulation engine also supports graphical representation of the events to allow for easy and quick observation of the circuits speculated behavior in the cells.

The Team

Boris Shabash, Taras Karpachevskyy, Neven Dimic, Josh Leung, Kent Thachek

Christian Jacob, Sonja Georgijevic, Anders Nygren and Rolando Feyt

University of Calgary, Calgary, Canada

igem@ucalgary.ca

Data Gathering

The system uses Perl (www.cpan.org) in order to retrieve information from several different databases, in order to gather information to characterize the different bio-agents. The iGEM registry is the initial source of information where information about the different bio-bricks’ information such as type, efficiency, and their DNA sequence is retrieved. For protein coding regions, the system then inspects the product proteins through UniProt (www.uniprot.org) and retrieves all the information required to simulate the proteins’ behavior in silico (virtually). Finally, the system inspects each enzyme that is retrieved and the catalytic activity associated with it. The system then obtains information about every chemical species that is involved in such a reaction through the entries in ChemSpider (www.chemspider.com). This integrated set of data is then written to a local file on the machine for quick and easy retrieval of the information during the system’s simulation.

User Interface

The evoGEM system presents a convenient, intuitive and friendly interface for practitioners of the biological sciences to interact with this software. The basic interface allows users to specify wanted products from the system and their relative quantities, and supply these requests in several different formats (common names, amino acid sequences and InChI formats are all acceptable).

More advanced users can manipulate the system’s evolutionary parameters such as mutation frequency, or any length constraints on the circuit. Users can also specify a set of parts they believe are part of a possible solution to further hasten the search and design process.

Users may even use this interface to inspect the parts and their associated information that has been retrieved from the different databases. Thus this system can also display information that is important for the researcher for any manipulation of the bio-bricks in the lab.