Programming Life - Seminar report

Group 5/E: Felix Akkermans, Niels Doekemeijer,
Thomas van Helden, Albert ten Napel, Jan Pieter Waagmeester

March 16, 2012

Abstract

For the development of a modeling and simulation package for BioBricks we attended a seminar in synthetic biology, in which essential knowledge from this domain was gathered. Together with an abstract of this, a reflection on the relationship of this project with synthetic biology is given, motivating the value of the effort that this project makes at applying computer science in the field of synthetic biology. Along we provide a reflection what we think is the role of computer science with biology in general.

1 Abstracts of the seminar

For each seminar a summary was made. This provides an abstract of each summary in chronological order. The summaries themselves are provided in appendix **??**.

1.1 Introduction to molecular cell biology

DNA is a molecule that provides the information for the construction of proteins. This molecule can be divided in several ways to provide pieces to describe the buildup of DNA. Messenger RNA (mRNA) relays the information in DNA to the protein synthesis machinery. mRNA shares a similarity with DNA but is directly used in protein construction, because mRNA encodes for the amino acids that make up the protein. In this chain of translation, the expression of a gene, several factors that interact play a role in promoting or suppressing of this process. The transfer, expression of DNA varies based on the the type of cell and the life form. Based on the study of this, the ability to influence this arises, and allows for manipulation of DNA and the partial prediction of it's outcome.

1.2 Math in biology

The math in biology centers very much around the concentration of molecules in the cell or cell compartments. To determine these concentration the changes in these is studied. These changes are determines by the speed of reactions that create or break down these molecules. This can be modeled in differential equations for process which are composed to describe the interaction of several molecules in a cell. Because of the interaction between these organic

molecules, some processes are catalyzed and others expressed. Enzymes play a big role in this and the reactions of these are described by enzyme kinetics. With these models that describe the expression and interaction between genes, gene networks can be imagined and described in genes. By placing these in a cell, bacterial response regulators or activators can be used to target promoters or suppressors. This way input from outside the cell can be provided to these gene networks, to create a information processing system. Such systems can implement signal generating components or logical gates.

1.3 iGEM contest

To motivate students and bring advances in synthetic biology, the iGEM contest challenges teams from universities all around the world to create and study systems in synthetic biology.

The team from University Groning attempted to tackle a problem in the watersupply of areas where it is unsure wether the water is polluted. To determine this pollution they implemented a mechanism in E. Coli to detect certain pollutants and create a visible reaction. The aim is to change the buoyancy of the cell by producing gas, so it will float to the surface.

The team of Imperial College London created a framework for biosensors. Their aim was to create better reaction biosensors, so they react faster and more noticeable. For this they modified *B. subtilis*, and as an example made it reacting to a harmful substance, such that it is visible for the human eye. This could be applied in crisis situations to detect contaminated water.

The aim of Cambridge was to create glowing *E. Coli* that produces a large amount of light. For this the bioluminescence in fireflies and several light producing bacteria. These processes were optimized and adapted for *E. Coli*. Maximizing the light production by ten times brought new insights on the application of this.

1.4 Bioinformatics

The increase in data that biology produced opened a lot of possibilities in analysis. Among fields of interests are the properties of genes, the prediction of the functioning of genes, research on the relation of genes among species, the integration of databases with biological information, and the search for molecules that could serve a wanted purpose. In medicine this knowledge opens up the possibility for personalized treatment based on your genome. Other trends are analysis of RNA-splicing, and the development of a sort of *web 2.0* for bioinformatics to share the knowledge.

2 Relationship between this project and synthetic biology

In computer science one could argue that programming lies at the very foundation of it. Since the discovery of DNA in biology, the knowledge about it's working gave the insights on how to DNA translates to complex effects and behaviour very similar as to a programming language. Analogous to how the simple languages in computer programming that spawned great complexity in

this field of science gave raise to the demand of tools to aid this development, this also goes for DNA in biology. And this parallel with computer science makes it inviting to use the same methods and paradigms, to build layers of abstraction and functionality.

In this project we attempt apply a selection of paradigms of computer science, and use some of the technologies that have proven successful in computer science, and bring them in the field of synthetic biology. The migration of these to a field of science that is quite a foreign domain for computer science deserves careful implementation and an open mind to new problems.

We will create a modelling and simulation package for BioBricks. The power of a modelling and simulation suite has significantly aided and accelerated the development of numerous technologies in computer science and other fields, and the aim is to bring these same benefits to those working beyond the border between computer science and biology. This will apply biological math to model the functioning of biological systems, and a GUI and simulation server environment in modern, mature technologies from computer science.

3 The role of computer science in biology

Computer science came about not with the purpose to study itself. It was a means to and end, a aid to accelerate technology and fields of science where it could be applied at the time. Biology was so much not among at first. Since the discovery of DNA and the development of technologies to read and synthesize these, together with the development of technologies to observe and analyse molecular processes such as cells, this exploded the bandwidth of digital data that biology produces constantly. Exactly like how computers were developed for, and deployed in fields of science to work on their data, now computer science will have to make a quick catch-up in this novel domain. We think that in order to do this, it should not have to walk the same path as other fields did and work with antique technologies, but should immediately benefit from the latest improvements in computer science. We also think that the marriage of computer science and biology holds a very exiting future with it's application in society, and with a lot of potential which will probably spawn a lot of new fields in science. We think it could very well play a key role in solving the most important problems facing humanity, and could greatly determine the shape of our future, and therefore we think that it deserves special attention. This could be in the form of contests to motivate students to pick up this field, or consciousness raising of the general public to create more support for this field among policy makers.

A Seminar summaries

Here the summary of each seminar is given in chronological order.

14 februari 2012
Groep 5

TI2800 Programming life

Seminar 1: Introduction molecular cell biology

DNA contains information that describes the construction of proteins. Proteins serve structural and functional roles that give individual cells—and by extension whole organisms—specific structures and functional characteristics. Each of our cells (with a few exceptions like red blood cells, eggs, and sperm) contain all the DNA required to code our genetic features. Individual regions of DNA that confer traits are called genes. The word “genome” refers to all the DNA in an organism. The human genome contains approximately 30,000 genes arrayed on 46 long stretches of DNA called chromosomes. The two strands of DNA are said to be complementary because the sequence of one strand indicates the sequence of the opposite strand. These four chemical units, adenine, cytosine, guanine, and thymine, are often abbreviated as A, C, G, and T, respectively. The base-pairing follows a certain set of rules, now called Watson-Crick base-pairing, in which a specific purine pairs only with a specific pyrimidine: A with T, and C with G.

Messenger RNA (mRNA) is used to relay information from genes in DNA to the protein synthesis machinery. Unlike DNA, which is usually double-stranded, mRNA is single stranded. When a gene is being transcribed into RNA, which is in turn directing protein synthesis, the gene is said to be expressed. The protein-synthesis machinery “reads” the nucleotides three at a time, assembling amino acid chains that correspond to the mRNA sequence.
At any one time, a cell is only expressing a few thousand of the genes in its genome. Proper regulation of gene expression is essential. Under- or over-expression of genes can have deleterious effects. For example, many forms of cancer are caused by mis-regulation of gene expression that results in uncontrolled cell division. To accomplish this regulated gene expression, the DNA contains control sequences in addition to coding regions. The control regions in DNA at which RNA polymerase binds to initiate transcription are called promoters. RNA polymerase binds more tightly to these regions than to the rest of the DNA and this triggers the start of transcription.

Four different bases in nucleic acids can specify proteins made up of 20 different types of amino acids. Each amino acid is encoded by a set of three consecutive bases. The three-base sets in RNA are called codons. There are three codons that do not encode an amino acid but instead signal the end of a protein-coding sequence. These are generally called the stop codons.

The prokaryotes are a vast group of unicellular microorganisms. Their cells are simple in structure, lacking a nucleus and other intracellular organelles such as mitochondria and chloroplasts. All animals, plants, fungi, algae, and protozoa are eukaryotes.
Molecular biologists can transfer genes from humans and other animals into bacteria, yeast, and other organisms to confer the ability to produce specific proteins that may be extracted for therapeutic use. Rather than thinking of genes as determinants of physical characteristics, they should be regarded as potentials or predispositions for characteristics. The ability to modify characteristics of cells is similarly limited by biological and physical constraints.
Since some cells are rapidly replaced, induced changes will be quickly lost. Because it is not possible to fully predict the outcome of these procedures, scientists must perform experiments, take observations, refine theories, and finally develop functional applications.

Groep 5
Felix Akkermans, Albert ten Napel, Niels Doekemeijer, Thomas van Helden, Jan Pieter Waagmeester

Rates of Chemical Reactions

The “Rates of Chemical Reactions” piece is about how chemical reactions regularly work. For a complete detailed description read that entire section. So summarize it a little bit:

Chemical reactions happen at a certain rate. A + B -> X + Y in a uni-directional reaction. It will stop once there is not enough fuel (A or B). The rate at which this conversion happens can be calculated. This is depicted in this chapter.

Enzyme Kinetics

Enzymes are catalysts in living systems. They make complex reactions happen at lower temperature and a lot faster. Proteins used as catalyst will first be used and then regenerated with during a multi-step process. Enzyme catalyzed reactions are normally rate-limited by enzyme saturation, meaning their limit depends on how many enzymes you got. Afterwards there is a whole example on how to calculate this rate etc…

Synthetic gene networks

Due to advanced molecular manipulation techniques and more knowledge on genetics we are slowly getting closer to using synthetic networks to manipulate cellular behavior. This chapter is about building genetic networks and its modular genetic components. In many ways these synthetic networks have analogies with electrical circuits. To understand how this synthetic network works we need to understand the components and how they interact.

Transcriptional control operates at the level of mRNA synthesis through the use of inducible transcriptional activators and repressors that are capable of binding naturally occurring or specifically engineered promoters. The majority of systems utilize bacterial response regulators or activators that, upon binding to a target promoter, inhibit or activate transcription respectively. Binding of a specific molecule to the response regulator induces an allosteric change leading to disassociation of the regulator from its cognate promoter.

Bacterial response regulators also form the basis of synthetic eukaryotic gene regulation systems although given transcriptional differences they require adaptation. This has been successfully achieved for many bacterial response regulators by placing the operator for the response regulator adjacent to an eukaryotic compatible promoter. The response regulator thus acts as a heterologous DNA-binding protein (DBP) whose association with the desired promoter can be controlled through addition of an appropriate inducer. If the operator is placed close to an strong constitutive promoter, DBP binding can prevent the initiation of transcription by RNA polymerase II machinery. Alternatively, transcription can be actively repressed by fusing a eukaryotic transcriptional silencer, such as the Kruppel-associated box protein (KRAB), to the DBP. Such systems are referred to as ON-type systems, as the addition of an inducer leads to de-repression of transcription. In an OFF-type configuration, in which addition of inducer leads to transcriptional silencing, a transcriptional activation domain is fused to the DBP. By placing the corresponding operator site adjacent to the minimal promoter, DBP binding activates

TI2800 Contextproject: Samenvatting week 3

Groep 5/E

22 februari 2012

1 iGEM: Team Groningen

Het team van Groningen heeft geprobeert een probleem in watervoorziening op te lossen. Hier is vervuiling door zware metalen (arsenicum, zink en koper) een probleem. Zij hebben geprobeert een mechanisme in *E. Coli* in te bouwen dat dat deze metalen importeert en hier ionengas van vormt. Bij het opnemen van een metaal en het verwerken er van moet de bacterie gaan drijven. Omdat *E. Coli* geen drijfsysteem heeft ingebouwd bij het produceren van de gassen, hebben ze hiervoor de genen geleend van *Bacillus Magaterium* die hiervoor verantwoordelijk zijn.

Het team is er in geslaagd volgens de BioBring Standard Assembly 10 de genen te clonen in een herhalingspatroon, en deze in de baterie te plaatsen. Er waren gemixde resultaten. Er was was een meetbaar verschil in de drijfkracht van de baterie, en het bleek ook dat de bacterie gevoelig was voor metaal en deze opnam. Vooral individueel bleken de modules goed te werken.

2 iGEM: Imperial College

Het team van het Imperial College heeft geprobeert een cell te maken die ingezet kan worden als biosensor. Het probleem op dit moment is dat veel biosensoren te langzaam werken. Daarnaast is het mogelijk met het framework dat zij ontwikkeld te hebben verschillende parasiten waar te nemen. Met hun deelname hebben ze zich gericht op de parasiet *Schistosoma*.

Het voornaamste werk is gegaan in het ontwerpen van een protiene bindende site op de cellwand. Hier voor is *B. subtilis* omgebouwd, en kan de surface site de larven van de parasiet waarnemen. Het signaal dat deze zijn waargenomen wordt dan doorgegeven door autoinducing peptides. Het input signaal activeert een genexpressie. Hierdoor wordt een geel protease gevormd dat zelfs voor het blote oog zichtbaar is. Dit is een nieuw systeem, en is vele malen gemakkelijker dan wat voorheen gebruikt werdt, waarbij fluoriserende proteine werden aangemaakt en deze vaak met ingewikkelde metingen pas konden worden vastgesteld. Dit moet mede in praktijk helpen om in crisis situaties het systeem effectief in te zetten

Uit experimenten bleek het systeem zeer gevoelig en snel reagerend.

TI2800 Contextproject: Samenvatting week 4

Groep 5/E

7 maart 2012

1 Wat is Bioinformatica?

In de *review paper* What is bioinformatics? An introduction and overview uit 2001 wordt een overzicht gegeven van de uitdagingen die aanwezig zijn in het opslaan en analyseren van de datastroom die biologen in hun werk genereren. Om de grote datastroom overzichtelijk te kunnen houden zijn verschillende soorten databases ontwikkeld. Van het opslaan van afzonderlijke eiwitten met hun eigenschappen en functie, tot het opslaan van de volgorde van de baseparen in de het hele DNA van verschillende organismen, de zogenaamde *whole-genome*-databases. Door toenemende integratie van deze verschillende databases is de informatie veel beter in zijn context te plaatsen en kunnen waardevoller resultaten worden behaald.

Nu de gegevens goed kunnen worden opgeslagen en geraadpleegd kan er worden nagedacht over wat voor analyeses we erop kunnen loslaten. Er wordt bijvoorbeeld gekeken naar eiwitvolgordes, 3D-stucturen en uitlijningen, welke delen van het DNA encoderen voor welke eiwitten en het voorspellen van secundaire en tartiare structuur van eiwitsequenties.

Om al deze zaken mogelijk te maken zijn verschillende informatietechnieken nodig: simulaties en voorspellingen leunen zwaar op beschikbare kennis uit de informatica en de statistiek.

Tot slot worden enkele toepassingen opgesomd: Het zoeken naar vergelijkbare biomoleculen van een molecuul waar nog weinig van bekend is om zo meer te weten te komen over het onbekende molecuul. Het ontwerpen en produceren van medicijnen.