Bio-design automation and synbio tools

Melanie Swan, MS Futures Group, m@melanieswan.com, 415-505-4426
July 28, 2009; http://melanieswan.com/documents/biodesign_automation.pdf

The ability to write DNA could have an even greater impact than the ability to read it. Synthetic biologists are developing standardized methodologies and tools to engineer biology into new and improved forms, and presented their progress at the first-of-its-kind Bio-Design Automation workshop (agenda, proceedings) in San Francisco, CA on July 27, 2009, co-located with the computing industry's annual Design Automation Conference. As with many areas of technological advancement, the requisite focus is on tools, tools, tools!

Experimental evidence has helped to solidify the mindset that **biology is an engineering substrate like any other** and the work is now centered on creating standardized tools that are useful and reliable in an experimental setting. The metaphor is very much that of computing: just as most contemporary software developers work at high levels of abstraction and need not concern themselves with the 1s and 0s of machine language, in the future, synthetic biology programmers would not need to work directly with the Ac, Cs, Gs and Ts of DNA or understand the architecture of promoters, terminators, open reading frames and such. However, with synthetic biology being in its early stages, the groundwork to define and assemble these abstraction layers is currently at task.

Status of DNA synthesis

At present, the DNA synthesis process is relatively unautomated, unstandardized and expensive ($0.50-$1.00 per base pair (bp)); it would cost $1.5-3 billion to synthesize a full human genome. Synthesized DNA, which can be ordered from numerous contract labs such as DNA 2.0 in Menlo Park, CA and Tech Dragon in Hong Kong, has been following Moore's Law (actually faster than Moore's Law Carlson Curves doubling at 2x/yr vs. 1.5x/yr), but is still slow compared to what is needed. Right now short oligos, oligonucleotide sequences up to 200 bp, can be reliably synthesized but a low-cost repeatable basis for genes and genomes extending into the millions of bp is needed. Further, design capability lags synthesis capability, being about 400-800-fold less capable and allowing only 10,000-20,000 bp systems to be fully forward-engineered at present.

So far, practitioners have organized the design and construction of DNA into four hierarchical tiers: DNA, parts, devices and systems. The status is that the first two tiers,

DNA and parts (simple modules such as toggle switches and oscillators), are starting to be consistently identified, characterized and produced. This is allowing more of an upstream focus on the next two tiers, complex devices and systems, and the methodologies that are needed to assemble components together into large-scale structures, for example those containing 10 million bp of DNA.

Standardizing the manipulation of biology
A variety of applied research techniques for standardizing, simulating, predicting, modulating and controlling biology with computational chemistry, quantitative modeling, languages and software tools are under development and were presented at the workshop.

Models and algorithms
In the models and algorithms session, there were some examples of the use of biochemical reactions for computation and optimization, performing arithmetic computation essentially the same way a digital computer would. Basic mathematical models such as the CME (Chemical Master Equation) and SSA (Stochastic Simulation Algorithm) were applied and extended to model, predict and optimize pathways and describe and design networks of reactions.

Experimental biology
The experimental biology session considered some potential applications of synthetic biology, first the automated design of synthetic ribosome binding sites to make protein production faster or slower (finding that the translation rate can be predicted if the Gibbs free energy (delta G) can be predicted). Second, an in-cell disease protection mechanism was presented where synthetic genetic controllers were used to prevent the lysis normally occurring in the lysis-lysogeny switch turned on in the disease process (lysogeny is the no-harm state and lysis is the death state).

Tools and parts
In the tools and parts session, several software-based frameworks and design tools were presented, many of which are listed in the software tools section below.

Languages and standardization
The languages and standardization session had discussions of language standardization projects such as the BioStream language, PoBol (Provisional BioBrick Language) and the BioBrick Open Language (BOL).

Software tools: a SynBio CrunchUp
Several rigorous computer-aided design and validation software tools and platforms are emerging for applied synthetic biology, many of which are freely available and open-source.

- Clotho: An interoperable design framework supporting symbol, data model and data structure standardization; a toolset designed in a platform-based paradigm to consolidate existing synthetic biology tools into one working, integrated toolbox

- SynBioSS - Synthetic Biology Software Suite: A computer-aided synthetic biology tool for the design of synthetic gene regulatory networks; computational synthetic biology

- RBS Calculator: A biological engineering tool that predicts the translation initiation rate of a protein in bacteria; it may be used in Reverse Engineering or Forward Engineering modes

- SeEd - Sequence Editor (work in progress): A tool for designing coding sequence alterations, a system conceptually built around constraints instead of sequences

- Cellucidate: A web-based workspace for investigating the causal and dynamic properties of biological systems; a framework for modeling modular DNA parts for the predictable design of synthetic systems

- iBioSim: A design automation software for analyzing biochemical reaction network models including genetic circuits, models representing metabolic networks, cell-signaling pathways, and other biological and chemical systems

- GenoCAD: An experimental tool for building and verifying complex genetic constructs derived from a library of standard genetic parts

- TinkerCell: A computer-aided design software for synthetic biology

Future of BioCAD
One of the most encouraging aspects in the current evolution of synthetic biology is the integration the field is forging with other disciplines, particularly electronics design and manufacture, DNA nanotechnology and bioinformatics.

Scientists are meticulously applying engineering principles to synthetic biology and realize that novel innovations are also required since there are issues specific to engineering biological systems. Some of these technical issues include device characterization, impedance, matching, rules of composition, noise, cellular context, environmental conditions, rational design vs. directed evolution, persistence, mutations, crosstalk, cell death, chemical diffusion, motility and incomplete biological models.

As it happened in computing, and is happening now in biology, the broader benefit of humanity having the ability to develop and standardize abstraction layers in any field can be envisioned. Clearly there will be ongoing efforts to more granularly manipulate and create all manner of biology and matter. Some of the subsequent areas where standards and abstraction hierarchies could be useful, though not immediate, are the next generations of computing and communications paradigms, molecular nanotechnology (atomically precise matter construction from the bottom up), climate, weather and atmosphere management, planet terraforming and space colony construction.