VLSI design in biology

Unpredictable transgenesis
Design and simulation of biochemical networks
Challengers of synthetic biology
Acceleration of biological evolution by computation
DNA in bottleneck

http://obolab.euro.ru


How could we engineer living organisms?

• Minimal life? Programmable artificial cell?
  – Chris Langton’s Self-Reproducing Loop, 86 cells, 8 states
  – phiX174, 5386 nt, 11 genes

• Minimal cell, [~100, 265-350] genes
  – Top-down: reprogramming simple organisms
    • *Mycoplasma genitalium* G-37, 580 Kbp, 480 genes, Craig Venter
    • *Mesoplasma florum* L1, 793 Kbp, 517 genes, Tom Knight
    • Synthetic genomic Inc, 2005, Craig Venter
  – Bottom-up: creating cells from nonliving material
    • Los Alamos Bug, PNA, Steen Rasmussen
    • ProtoLife, 2005, Norman Packard, Mark Bedau

• Human, 3.4 Gbp, 20,000-25,000 genes
  – Genetic interface: biological VM on chassis cells (T7, SP6 DNA-dependent RNA polymerases), non-DNA VM (TNA), orthosome (HNA, CeNA), …

• Evolution under the control of a man or a computer?
Gene Composer™, GeneDesign - web-design of synthetic genes

- Reverse Translation
- Codon Juggling
- Silent Site Insertion/Removal
- Sequence Analysis

- Oligo Design
- Enzyme/Vector Choosing
- Short Sequence Removal
- Random DNA

More than 300 genomes in the global database! We could play God to try an “intelligent design”
The environment for DNA program installation

In *E. coli* the intracellular protein concentration is 200-320 mg/ml (100 nm window)

Database of Interacting Proteins (DIP)

**Benefits and limitations**

- very small system
- small number of molecules, intrinsic noise
- wires, diffusion, speed
- parallelism
- synchronization
- signal levels
# Molecular dynamics & simulation of biochemical networks

<table>
<thead>
<tr>
<th></th>
<th>Deterministic</th>
<th>Stochastic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non-spatial</strong></td>
<td>mass-action ODE solvers Gepasi</td>
<td>Gillespie’s first reaction algorithm Stochastirator</td>
</tr>
</tbody>
</table>
| **Spatial**      | PDE, reaction-diffusion equation Virtual Cell  
|                  | • exact                          | reaction events occur during collision          |
|                  | • approximate                    | MCell ray-tracing algorithms                    |
|                  | • hybrid approach                | ChemCell                                        |
|                  | • special-purpose hardware - field programmable gate arrays (FPGAs) | Moleculizer for protein complexes, the program automatically generates the network of reactions BioNetGen signaling pathways |

Gibson modified Gillespie algorithm allows to execute $10^{10}$ reactions per day on 800 MHz Pentium III Processor. A hypothetical simulation of *E.coli* will track $10^{14}$-$10^{16}$ reactions per cell.
Number of molecules vs. concentrations

- Plots of reaction probability density function $P(\tau, \mu)$
- Gibson modified Gillespie next reaction algorithm: only log of number of reactions
Enabling Biological Engineering (Drew Endy, 2005)

- **System abstraction**
  - Insulate relevant characteristics from overwhelming detail
  - Simple artifacts that can be used in combination
  - From Physics to EE, 1800s

- **Standardization of components**
  - Predictable performance
  - Off-the-shelf
  - ME, 1800s

- **Decoupling of design from fabrication**
  - Rules insulating design process from details of fabrication
  - Enable parts, device, and system designers to work together
  - VLSI electronics, 1970s

- **Hierarchy**
  - Systems
  - Devices
  - Parts
  - DNA

![Diagram of system abstraction, standardization of components, and decoupling of design from fabrication with DNA sequence: 5'-TAATACGACTCACTATAGGGAGA-3']
Registry of standard biological parts ([http://parts.mit.edu/](http://parts.mit.edu/))

- File parts in XML format or RDBMS
- A modular way of building and composing DNA parts with prefix & suffix (Tom Knight, MIT)
BioSPICE & BioJADE – tools for design and simulation of bio-systems

BioJADE

- Standard Simulators
  - Stochastirator
  - Tabasco
- Access Methods
  - JDBC
  - Perl/DBI
- BBDataStore
  - BBDBC (Oracle, MySQL)
  - BBXML
- BioBricks Collection
  - promoters, RBS, structural genes, terminators…
BioJADE for synthetic biologist
(Jonathan Goler, 2004)
DNA synthesis: long & cost

Carlson Curves:

- It is possible to order a 10 Kbp DNA fragment over the web: Codon Devices, Blue Heron
- $0.85 to $1.60 per bp
- Some researchers consider that it is now possible to build 200 Kbp DNA fragments and expect that a ~1 Mbp bacterial genome will be constructed within the next two years
- BioFAB™ production platform that is expected to accurately synthesize Kilo- to Mega-bp-length genetic code
Potential applications

• Programmable devices: sensors, control logic, effectors
• Bacterial robots
• Amorphous computing
• Smart materials and self-assembly
• Molecular medical devices
• Artificial metabolic pathways
  – *Artemisinin* (malaria drug)
  – Energy & Fuel productions (photosynthesis, ethanol, hydrogen)
  – Ecology (reduce greenhouse gas emissions)
• …
• Creation of artificial life for the industrial applications

*Levskaya et al.* (2005) Engineering *E.coli* to see light, Nature, 438: 441-442
Evolution *in silico* vs. design

- **Compiling, optimization**
  - The computation is cheaper and faster than the synthesis of DNA
  - Write DNA encoding the successful solutions and next transfer this design from model to life

- **Golem Project**: model-based evolution *in silico*
## Algorithmic paradigms of evolution

*(Richard Watson, 2006)*

<table>
<thead>
<tr>
<th>Dependency of variables</th>
<th>Few / weak interdependencies</th>
<th>Modular interdependencies</th>
<th>Arbitrary interdependencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landscape</td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
</tr>
<tr>
<td>Algorithmic paradigm</td>
<td>hill-climbing – accumulation of small variations</td>
<td>divide-and-conquer problem decomposition</td>
<td>exhaustive search, random search</td>
</tr>
<tr>
<td>Complexity</td>
<td>$KN$</td>
<td>$N^K$</td>
<td>$K^N$</td>
</tr>
<tr>
<td>Evolutionary analogy</td>
<td>gradual evolution</td>
<td>compositional evolution</td>
<td>“impossible” / “intelligent design”</td>
</tr>
</tbody>
</table>

N – # of variables, K – # of values for each variable
What is your dangerous idea?

• Start with fun and education
  – GEM-F Club
  – iGEM competition
• Continue with reliable infrastructure for the synthetic biology
• Finish with real applications to be a “Bill Gates”

A Scientist discovers that which exists.
An Engineer creates that which never was.
Theodore von Karman