Generalized Stochastic Simulation Algorithm for Artificial Chemistry
Gillespie in artificial chemistry

Hedi Soula

Sorbonne University

September 25 2018
Bacterial reaction graphs

(a) Step 1: From cDNA fasta file, get gene labels.

(b) Step 2: From KEGG gene entries, get EC codes.

(c) Step 3: From KEGG enzyme entries, get substrates and products.


(cDNA fasta file)

(KEGG gene entries)

(KEGG enzyme entries)

(L-aspartate ----> 4-phospho-L-aspartate)
Artificial Chemistry

Motivation

▶ (Real) Chemistry is ... difficult
  ● Maybe artificial chemistry is easier
  ● can model interesting properties (for e.g. space/diffusion)
▶ (Real) Life is ... chemistry
  ● Artificial Life should have ... artificial chemistry (AC)

What is done (usually)

▶ chemistry is prescribed : small dimension (small # of reactions)
▶ straightforward : chemistry graph is simple
▶ (somewhat) unrealistic
  ● transition energies are ignored (easier)
  ● mass conservation is sloppy
  ● \[A + B \leftrightarrow C\] and \[C \leftrightarrow A\]
AC: what is expected

1. large: should have a huge number of reactions
2. energy: all reactions are not possible
3. mass: reversibility should not be hacked
4. open: we don’t know all reactions/molecules
Hutton Artificial Chemistry (1)

Hutton J.

- several papers on his AC scheme
- only one follow up
- straightforward: chemistry is simple
- very complex: chemistry reactions network can be extremely complex
- however (somewhat) unrealistic

Main properties

- molecules are graphs
- nodes are domains
- edges are chemical bonding
- domains have a fixed type and a changing state
  - described by a pair \((t|s)\)
  - usually a letter and a number \(a0, b1 \ldots\)
- any domain can have any numbers of links
- the chemistry is composed of fully connected subgraphs
### Geometry and Physics

- HAC is 2D
- spatially resolved
- each domain has an id/position
- links are *springs* (however weird coding)
- (vaguely) Brownian in viscous medium
Reactions

Reactions are of the form:

\[(t_1|s_1).+(t_2|s_2) \mapsto (t_1|s_3).+(t_2|s_4)\]

- \((t_i|s_j)\) is a domain (within a molecule)
- \((.|+)\) is either linked or no (collision) +
- note that type is unchanged
- modification are only local (other links unmodified)
- conformation is performed instantaneously
- reaction \(+\) is performed instantaneously upon collision
- when conflict: chose at random
Examples

Starting a reactor with several $a_0$s and 1 (one) $a_1$ with

$$a_0 + a_1 \leftrightarrow a_1\cdot a_2 \text{ yields}$$

![Diagram 1](image1)

whereas $a_0 + a_1 \leftrightarrow a_2\cdot a_1 \text{ yields}$

![Diagram 2](image2)
Main properties

Using wildcards ...

R9 : $e^8 + e^0 \rightarrow e^4e^3$

R10 : $x^4y^1 \rightarrow x^2y^5$

R11 : $x^5 + x^0 \rightarrow x^7x^6$

R12 : $x^3 + y^6 \rightarrow x^2y^3$

R13 : $x^7y^3 \rightarrow x^4y^3$

R14 : $f^4f^3 \rightarrow f^8 + f^8$

R15 : $x^2y^8 \rightarrow x^9y^1$

R16 : $x^9y^9 \rightarrow x^8 + y^8$
# Hutton Artificial Chemistry (6)

## Pros and Cons

<table>
<thead>
<tr>
<th>Pros:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• very general</td>
</tr>
<tr>
<td>• 2D and nice</td>
</tr>
<tr>
<td>• mass conservation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cons:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• 2D and nice</td>
</tr>
<tr>
<td>• absurdly long</td>
</tr>
<tr>
<td>• no reaction rates</td>
</tr>
<tr>
<td>• <em>intelligently</em> designed</td>
</tr>
</tbody>
</table>
Getting rid of space

Same data structure (without space)

reactions with rates

SSA formalism (Gillespie)

slightly involved (but not too much)

github.com/hsoula/staarc
Gillespie in AC

**Modifications**

- Reaction have a rate:
  \[(t_1 | s_1)(.+)(t_2 | s_2) \leftrightarrow (t_1 | s_3)(.+)(t_2 | s_4) : \lambda\]

- For each `+` reaction:
  1. We count the number of `n_1 = (t_1 | s_1)` and `n_2 = (t_2 | s_2)`
  2. minus the number of `n_{12} = (t_1 | s_1)(t_2 | s_2)` already linked
  3. the propensity is `a = \lambda (n_1 * n_2 - n_{12})`

- For each `.` reaction:
  1. we count the number of `n_{12} = (t_1 | s_1)(t_2 | s_2)` already linked
  2. the propensity is `a = \lambda n_{12}`

- then classical Gillespie algorithm

- once a reaction is selected we apply the modification to a given pair (selected at random uniformly)
First step

compute all propensities $a_i$ Compute the combined rates of all reactions

$$a_0 = \sum_{i=1}^{R} a_i$$

To compute the time of the next reaction, draw a random number

$$\tau = -\log(rn)/a_0$$

This answers the when
To compute the what, we choose randomly the equation weighted by their relative weight i.e

$$Pr(\text{next reaction is } i) = \frac{a_i}{a_0}$$

in practice, shoot a random number uniform $rn$ in $[0, 1]$ and find $r$ such as:

$$\sum_{i=1}^{r-1} a_i \leq a_0 rn < \sum_{i=1}^{r} a_i$$
Properties

1. now all reactions have rates: real chemistry
2. simulated 3D well mixed medium
3. almost all HAC properties conserved
4. we can simulate diffusion (reaction rate for collision)
5. way faster computation

This turn a local resolution to a global one
- well mixed medium
- with infinite crowding

not intuitive ...
Simple experiment: the replication according to the initial number of particles \( n \in \{30, 60, 600, 6000\} \) and \( \lambda \) ratio (between collision reaction \(+\) and conformation reaction \(-\)).
STAARC : replication

Properties
- Not very resilient - original sequence quickly lost
- Scales very badly

Not that intelligently designed :)

Race conditions
- Race condition – diffusion vs conformation
- Race condition – concurrent replications
STAARCR : replication

A

B

R12

Hedi Soula

Staarc
STAARC : long replication

Long replication experiment

- Starting with size 6 replication seed
- Production and degradation
  - \((x|0) \rightarrow \emptyset\)
  - \(\emptyset \rightarrow (x|0)\)
- Let the simulation for 450,000 reactions
- Compute average size and number of divisions
Long replication: replication event
Long replication : average size

![Graph showing Long replication: average size](image-url)
Long replication: standard deviation

![Graph showing standard deviation of molecule size over time.](image)
STAARC : long replication

Long replication experiment

- Replication is extremely stable
- Division occurs in bumps
- The limit size is ... 6
- Big compounds created transiently
Random Chemical World

Random generation of reaction

- available types: \{a, b, c\}
- maximum number of state is 5.
- compute all the possible reactions
  - with no production nor degradation.
  - keeping only a fraction \( p \in [0, 1] \) randomly
- starting \( N = 10,000 \) particles \((t|s)\) with \( t \in \{a, b, c\} \) and \( 0 \leq s \leq 4 \)

Experiment

- Maximum of 2,000 reactions.
- Compute the time needed
- The ratio of number of molecules
Random Chemical World

![Graph showing components ratio against log fraction with different symbols for -6.0, -5.0, and -4.0.]
Phase transition

- Huge variability in the middle
- in molecules created
- in the time to the end of reactions
- this middle is for a low number of reactions

Drawbacks

- Results not conclusive
- Better random selection of reactions
- Common elements
STAARC: STochastic Atom-based ARtificial Chemistry

STAARC: STochastic Atom-based ARtificial

github.com/hsoula/staarc
Perspectives

Code
- Open-Endedness procedural enzyme
- Tri-molecular reactions
- Graph-based random chemistry
- ’Metabolic network’-like graph

Projects
- Explore Random Chemistry
- Use it in an evolutionnary set-up
- Artificial ’bacterial world’
thanks

hedi.soula@upmc.fr // sites.google.com/site/hsoula/

Soula HA, Generalized Stochastic simulation algorithm for Artificial Chemistry Proceedings of ALIFE XV, 2016 Cancun, Mexico