Generalized Stochastic simulation algorithm for Artificial Chemistry

Hedi A. Soula[1,2]

[1] INRIA EPI Beagle LYON, FRANCE
[2] INSERM INSA U1060 LYON, FRANCE
hedi.soula@inria.fr

Abstract

Artificial chemistries (AC) are useful tools and a simple shortcut for the study of artificial life. In many works, AC's are quite straightforward or simplistic or highly unrealistic (or all combined) but in several works AC are extremely complex. Among them, we focus of Hutton Artificial Chemistry HuAC where reactions act on the nodes of a graph (so-called the atoms) where the connected components composed the actual molecules of the environment. The main works from Hutton are based on a 2D simulator (squirm) with auto-replication and several other properties. This paper proposes a computation framework and software that cancel the need for 2d space simulation in the HuAC while keeping a lot of the features of this chemistry. It relies on the Stochastic Simulation Algorithm that has been here adapted to work on graph structure. In order to test it, we simulated Hutton's auto-replication – which relies heavily on strong spatial interactions – in a spaceless environment. In addition, due to the increase in performance, we develop some preliminary work on Random Chemical Worlds where reactions are randomly selected. We showed on simple metrics that the fraction of reactions among all possible is a general parameter that acts on the system similarly to a phase transition.

Introduction

Most of the artificial life questions and problematic revolve around understanding and providing clues to the origin of life and evolution of organism starting from scratch (Hutton (2003)). Because of the difficulties of real-life (or 'wet') experiments to address these questions in 'real' life, simulations and artificial modelling seems to be a most adequate tools (Dorin and Korb (2007)). Unfortunately, reproducing completely life-like systems are still not an option for understanding general features used by living system to adapt and develop. Few artificial systems can be designed without the need to fix rules between building blocks components of living organisms. It usually necessitates to design an artificial chemistry scheme (Dittrich et al. (2001)) and this need extends obviously to problems on artificial life and evolutionary strategies. Indeed put it simply, since real life rests on chemistry *artificial* life should rely on artificial chemistry (Suzuki and Dittrich (2009)).

What is done (usually) is that first most chemistry is prescribed: it has small dimension (small # of reactions), straightforward that is the chemistry graph is either extremely simplistic or small (Knibbe and Parsons (2014)) and it is somewhat unrealistic for example most transition energies are ignored. Several AC have been recently developed to tackle specifically this issue of energy transition (Benkö et al. (2005); Ducharme et al. (2012); Benkö et al. (2009)). Additionally, problems like mass conservation can arise e.g. $A + B \mapsto C$ and $C \mapsto A$. It is expected that any evolvable scheme will exploit this kind of easy shortcut.

So what properties an AC should possess for more general life-like and evolution-based framework experiments? AC must be complex, rich and generic. More precisely, AC should be large and have a huge number of reactions. Second, and this is related to energy transitions, all reactions should not be possible. This amounts to require that components (molecules) cannot be reactive with all others. Obviously, mass conservation is required. This problem usually arises when the chemistry of the system is procedurally generated – for example using artificial genome (Rocabert et al. (2015)). In other words, reversibility should not be hacked. Finally, AC should allow some form of open endedness: we don't know all reactions/molecules (Lenaerts and Bersini (2009)).

Several frameworks have been published that address some of the properties described above Tominaga et al. (2009); Oohashi et al. (2009)).

One of them in particular, Hutton's Artificial Chemistry (HuAC) own several of these properties (Hutton (2007, 2004, 2002)). The central feature (cf Model) is to describe the chemistry as reactions between atoms with a fixed type and changing state while describing molecules as connected graphs of atom. Note that the term atoms refers to the smallest structure in the system and can describe structures – or domains – that are larger than actual atoms. The chemistry is a set of reaction on pairs of atoms and strikingly *not* on molecules. Like in the classical sense, these atoms need to make an encounter to react with each others (as a classical bimolecular reaction). However, the originality of HuAC

relies on so-called conformation reactions: reaction that occurs between bounded atoms. HuAC possesses all the required elements to be of a wider use. It has some element of open-endedness but part of the chemistry reaction set has been finely tuned to obtain the expected outcome.

Understandably, hand designing was part of the proof of concept of the chemistry (mainly to display self-replication) but lacks of generalisations. The simple questions what are the 'best' set of chemistry rules to obtain a life-like chemical system where for example artificial evolution can be tested is still open. Also overall, the main drawback is that it is in essence a 2D construction with - without any explicit mention to it - a strong diffusion-limited component.

This is a drawback for two reasons: 2D systems are nice but long to simulate – most of the time time between reactions is juste to simulate diffusion. Also, 2D simulations are simple to simulate but introduce unnecessary topology constraints like chirality that cannot be overcome without strong tuning. Also, as we will see more in details afterward, all the tuning relied on strong spatial assumptions: correlation of positions ignoring mixing effects. Finally, and to for sake of completeness, HuAC does not introduce really reaction rates in a biochemical sense: bimolecular reactions occurred immediately upon collision - and conformation reaction occurred instantly. Thus, there were absolutely no notion of reactions rates (and affinity etc...).

The aim of this article is to provide a framework – a stochastic simulator that simulate Hutton's artificial chemistry with several properties. First it will describe this chemistry in a well stirred medium (which amounts to 'infinite diffusion') using stochastic methods known as the Gillespie algorithm. This algorithm will be modified to suit Hutton's artificial chemistry that keep track of graphs of atoms and not atom abundance only. Our framework also introduce reactions rate and other reactions that were not included in the descriptions of Hutton's original work. Also we will examine the difference with HuAC's dependence on 2D structural constraints and specially the impact of spatial correlation for the main Hutton's algorithm: auto-replication. Finally since this framework can handle a huge number of reactions and/or molecules we will present preliminary works on randomly generated Artificial Chemistry and study their properties on simple metrics.

Model: Hutton's Artificial Chemistry

We start this section with a brief refresher on Hutton's chemistry (HuAC). HAC has been published in several papers (Hutton (2009, 2007, 2004, 2002)) but to our knowledge only one follow-up (Lucht (2012)). The main advantage of HuAC is that the description is really straightforward. The chemistry's rules are very simple. However, it can grow extremely complex due its graph structure. Briefly, Hutton's chemistry is composed of molecules that are graphs of *atom* (in Hutton's words). The atom term do not encompass actual atomic structure but merely describe the smallest compound in the system. As described, atoms have a type and a state that are both integers and molecule are simply connected graphs whose nodes are atom and with connections (bonds). Atoms have a fixed type and a changing state that are both integers and described by a pair $(t|s)$ with $t, s \in \mathbb{N}$ (note: in Hutton's papers type is a letter and state is an integer e.g $a0, e8 \ldots$). Any atom can have any numbers of connections. Therefore the chemistry is composed of fully connected subgraphs which is the set of molecules. The chemistry relies on a physical simulator in 2D. All atoms are spatially resolved and each atom has its own id and position. Atoms are hard spheres (of equal size) that undergo some kind of brownian motion in viscous environment and links are coded using springs $-k(p-q)$.

Reactions are based only on atoms and are of the form:

$$(t_1|s_1)(.|+)(t_2|s_2) \rightarrow (t_1|s_3)(.|+)(t_2|s_4) \qquad (1)$$

where . design a link i.e one atom of $(t_1|s_1)$ and one of $(t_2|s_2)$ must be linked for the reaction to occur as they design a conformational change in the pair. Also $+$ design an encounter reaction (therefore $(t_1|s_1)$ and $(t_2|s_2)$ *must* not be linked together and the reaction occurs whenever the two atoms collide (whether they are otherwise linked to other atoms). Note that one way to ensure mass conservation is to never allow the type to be changed in the equations. All reactions occur locally so other links are not modified (they can be later if there is a reaction matching the new links in the graph). In the original papers, conformationnal reaction noted . is performed instantaneously and when several are possible they "are chosen at random" – presumably with uniform distribution.

This chemistry is extremely general and encompass very easily several classical equations such an enzyme-substrate-product $S + E \rightleftharpoons C \rightarrow P + E$ using

R1:	$s_0 + e_0 \rightarrow s_1 e_1$
R2:	$s_1 e_1 \rightarrow s_2 + e_0$
R3:	$s_1 e_1 \rightarrow s_0 + e_0$

and also the so-called *Hit & Run* reactions :
$S + E \rightleftharpoons C$ with $C + A \rightarrow C + B$:

R4:	$s_0 + e_0 \rightarrow s_1 e_1$
R5:	$e_1 + b_0 \rightarrow s_1 + b_1$
R6:	$s_1 e_1 \rightarrow s_0 + e_0$

note that in the last equation the complex the atom e_1 is the functional equivalent of the complex C. Technically any e_1 in the reactor can react with a b_0 to yield a b_1.

However, due to its atom/graph organisation, this chemistry can quickly give rise to complex structure. For example the simple reactions:

R7:	$a0 + a1 \mapsto a1.a2$
R8:	$a0 + a1 \mapsto a2.a1$

(where the only difference is a swap between target states) for a given number of $a0$'s and $a1$'s in the reactor yields two different graph structures as shown on Fig. 1.

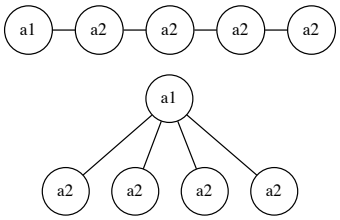

Figure 1: Examples of graph structure obtained using only R7 (top) and R8 (bottom) starting with 4 $a0$ and one (1) $a1$.

Hutton's original papers also come with an auto replication scheme where a given arbitrary single strand molecule $e_8.x_1^1 \dots x_1^k.f_1$ can automatically be duplicated using a given set of reactions:

R9:	$e8 + e0 \rightarrow e4e3$
R10:	$x4y1 \rightarrow x2y5$
R11:	$x5 + x0 \rightarrow x7x6$
R12:	$x3 + y6 \rightarrow x2y3$
R13:	$x7y3 \rightarrow x4y3$
R14:	$f4f3 \rightarrow f8 + f8$
R15:	$x2y8 \rightarrow x9y1$
R16:	$x9y9 \rightarrow x8 + y8$

Here is introduced the wild-card reaction using x and y. For example, $x4y1$ refers to linked atoms of *any* type but in state 4 and 1 whereas $x4 + x0$ refers to the collision of two atoms of the same type (but of any type) in state 4 and 0. One can immediately see that this does lead to a replication of any sequence $e_8.x_1^1 \dots x_1^k.f_1$. Elements are added and linked using reaction R11 and the splicing is initiated by R14 and propagated via R15/R16 (see 2 for a walkthrough).

This AC has several wanted features for artificial chemistry. It is very general and allows for very complex feature emergence (as attested by Fig. 1). Namely, molecules can be very complex. Also this comes with default embedded mass conservation. Due to a complicated graph geometry with 2D features, there are some chirality issues (due to 2D) and of course this AC is computationally demanding. Also as mentioned, from a chemical standpoint, there is no reaction rates.

The final remark is that most results that have been published on this AC has used hand-crafted and well designed set of reactions. In particular, there was no clear thought process described that explained how the replication system could work. It seems to us that it was drawn by hands and graph evolution was constructed with implicit bias due to proximity. However cells and more generally biological medium are disordered and can be highly diffusive. And of course processes takes places, mostly, in three dimensions.

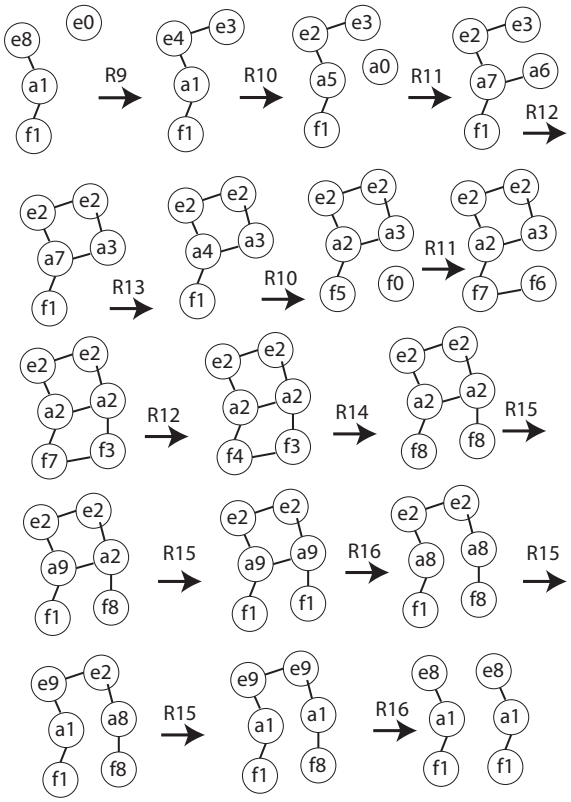

Figure 2: Walkthrough of the replication R9-R16 with an initial seed of $e8a1f1$ reproduced from Hutton (2002)

The very complexity of the HuAC lead us to design a system that would be more tractable and keep the essence of the original chemistry. We are also able to experiments with a large number of particles and compounds and deal with a large number of possible reactions to obtain more reliable results on the potential of HuAC.

STAARC: STochastic Atom-based ARtificial Chemistry

We developed the STAARC framework to use the HuAC in a virtual reactor that completely eliminate spatial location. Since due to the particle-based nature of HuAC and its graph structure, reactions are identified and occur concurrently but asynchronously at a given speed. Therefore, the ODE formalism seems not realistically tractable for this problem. This framework is not based on differential equations but on the Stochastic Simulation Algorithm (SSA) as it was originality developed by Gillespie Gillespie et al. (2013); Gillespie (2007).

Stochastic Simulation Algorithm

The important feature of the Gillespie algorithm is that we simulate the world only at times when a reaction occurs and not in between. Contrary to ODE formalism where description can occur for arbitrarily small time step, in SSA, all that is needed is to estimate which will be the next reaction, when it will occur and to actually implement the reaction. Since both the time and the reaction will be drawn according to a given diffusion, it is a simple mean to obtain noise and variability on molecular reactions. Note that, however, both the ODE and its SSA counterpart describe the same system with the same hypotheses and therefore yield the same results: the ODE being the description of the dynamics of the average.

To function, Gillespie algorithm introduces propensities a_i: the average rate at which the reaction i can occur in the medium. For a bimolecular reaction $X + Y \xrightarrow{k}$ this rate is equal to kxy when x and y are the *number* of molecule X and Y respectively in the reactor. Similarly, unimolecular reaction $X \xrightarrow{k}$ propensity is kX. Note that if $X = Y$ i.e bimolecular reaction involving the same type of molecule the propensity is equal to $kx(x-1)$. Propensities are the speed for one reaction to occur and are therefore of unit $time^{-1}$. The algorithm is as follow: When all reactions propensities are computed we compute the propensity of the system :

$$a = \sum_{i \in Reactions} a_i$$

and Gillespie showed that the time for the next reaction to occur is exponentially distributed with this parameter a.

Since reactions occur proportional to their rate, a simple random selection biased by relative rate proportion of rate yields the next reaction. Once the reaction is selected, it is applied i.e the reactants are removed and the products are added. This in turn modifies propensities - that need to be reevaluated and so forth. Note that in this algorithm, propensities are only calculated when reactions are occurred, propensities depends only on the number of the reactants involved in actual reactions. Only one set of reactants is updated at each step of the algorithm and finally it involves only the drawing of two random numbers: One to find the next reaction time and another to find the reaction itself.

This description of Gillespie can be applied as this to the HuAC scheme with few tweaks. First we simply add to each reaction a rate k_r for each reaction r.

Now the main hurdle is to obviously have a data structure that contains of all the information about the atoms and molecules (which are the connected components). Therefore the first – and important – difference with the SSA scheme is that we need to keep track of each individual particles $p_i = (t_i, s_i)$ for $i \in [1, N]$ (N being the number of particles) and we keep track of all the edges (link) between particles $\mathcal{E} = \{(p_i, p_j) \text{where} p_i \text{linked} p_j\}$. Note that in the SSA, we normally keep track of the *number* only of each reactive particles.

For a given pair $(t|s)$ let's set $\mathcal{A}(t,s) = \{p_i | t_i = t, s_i = s\}$ the set of particles of this given type and state. Let's note $a(t,s) = |\mathcal{A}(t,s)|$. Also let $\mathcal{B}(t_1,s_1,t_2,s_2) = \{(p_i,q_j) | p_i \in \mathcal{A}(t_1,s_1), p_j \in \mathcal{A}(t_2,s_2), (p_i,p_j) \in \mathcal{E}\}$ and finally let $b(t_1,s_1,t_2,s_2) = |\mathcal{B}(t_1,s_1,t_2,s_2)|$.

For the bimolecular reactions of the type:

$$\begin{array}{ll} \text{Ra:} & (t_1|s_1) + (t_2|s_2) \xrightarrow{k_a} \dots \\ \text{Ra':} & (t_1|s_1) + (t_1|s_1) \xrightarrow{k_{a'}} \dots \end{array}$$

the rates will be equal to

$$\begin{array}{ll} \text{Ra:} & k_a \left(a(t_1,s_1)a(t_2,s_2) - b(t_1,s_1,t_2,s_2) \right) \\ \text{Ra':} & k_{a'} \left(a(t_1,s_1)(a(t_1,s_1)-1) - b(t_1,s_1,t_1,s_1) \right) \end{array}$$

We basically need to remove atoms that are already link for the equation since it concerns only unlinked atoms. We also need to take care of the case when atoms have the same type and state.

Also for a reaction of the type:

$$\text{Rb:} \quad (t_1|s_1).(t_2|s_2) \xrightarrow{k_b} \dots$$

the rate will be equal to

$$\text{Rb:} \quad k b(t_1,s_1,t_2,s_2)$$

We compute the total propensity as the sum of all the rates of all reactions using the formulas above and draw our two random numbers. The first one $r \in [0,1]$ uniformly to compute the next time $\tau = -\log(r)/a$ and the other s for determining the reaction i such that:

$$\sum_{k \in reaction, k \leq i} a_i \leq s < \sum_{k \in reaction, k \leq i+1}$$

Once a reaction is selected we apply the modification to a given pair (selected at random uniformly). Due to the SSA scheme only one reaction is applied between each step so only one update to a maximum.

Properties

Akin to a more realistic chemistry, all reactions now have rate. This simulates a well mixed 3D reactor without any chirality issue that was only related to 2D. Almost all the HuAC properties are conserved in particular complex graphs. We can mimic diffusion to a certain point by modifying the actual rate of bimolecular reactions. Indeed, since Gillespie is the limit at infinite diffusion but reaction rate can be modified by diffusion using Smoluchowski equation (Szabo (1989)):

$$k = \kappa \frac{D}{A + D} \tag{2}$$

κ being the thermodynamic rate (when $D \mapsto \infty$).

We obtain an extremely fast computation where the only simulated moments are whenever a reaction occurs. We also

update the data structure that keeps track of the graph for only one pair at a time. The addition allows to simulate also a wider variety of reactions. Indeed, several reactions can be added to the simulator to allow even more realism. Equation such as production $\varnothing \xrightarrow{k_p} (t|s)$, degradation $(t|s) \xrightarrow{k_d} \varnothing$ and 'auto-conformation' $(t|s_1) \xrightarrow{k_c} (t|s_2)$ with respective rates as k_p, $k_d a(t,s)$ and $k_c a(t,s_1)$ respectively. The simulator was written in C# and is available with MIT licence at `https://github.com/hsoula/staarc`.

Results

We provide in this section two examples of the output of the simulator: we reproduce the auto-replication scheme of Hutton and we create Random Chemical Worlds to study their properties. Because of its flexibility it allows us to test several possibilities in the set of equations rules and rates and also compare with the 2D case of HuAC. Due to the local nature of the replication process, we expected it to fail completely in a infinite diffusion medium.

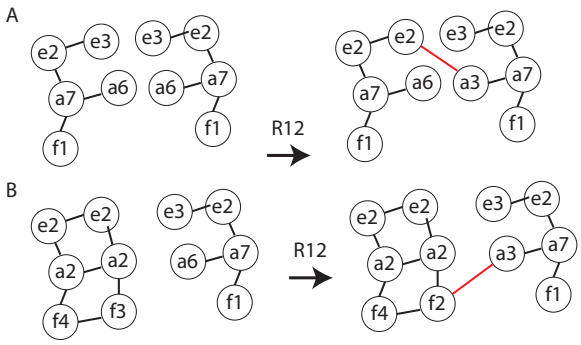

Figure 3: Examples of replication errors A) Due to no local constraints and concurrent replication bimolecular reactions can occur from other replication process B) Due to race condition – conformation change – occurring too late to prevent another bimolecular reaction.

Vanilla Replicator

We submitted our simulator the restricted set of equations (R9-R16) described above to check how replication occurs. When drawn by hands the replication seems it can occur flawlessly as shown on Fig. 2 but inspection of equation R13 shows a race condition. For the replication to continue atoms $a6$ must encounter (any) atome $x3$ (here $e3$). In a 2D and diffusion limited environment, the closest possible atom would be the $e3$ but it could theoretically be any atom from a concurrent replication elsewhere (see Fig.3A). This feature happened in the original Hutton's simulation occasionally and as this is the case here does not impact the stability of the replication it only changes the sequence in the replicated molecules. This deviation in replication is expected because

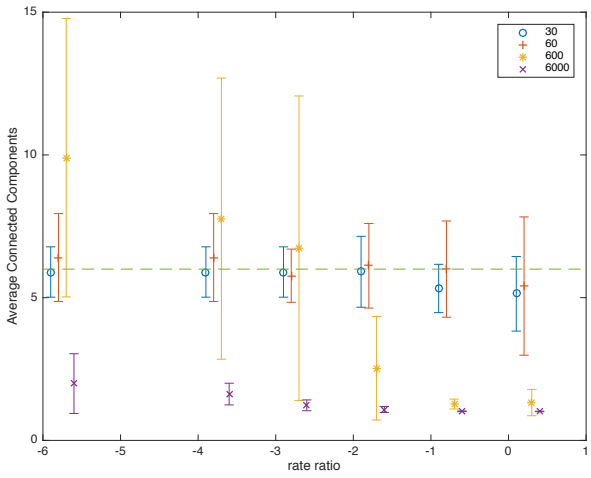

Figure 4: Number of connected components – molecules – at the end of the simulation of the replicator R9-R16. Two parameters were varied; first the rate of conformation versus bimolecular reaction (rate ratio) to simulate the impact of diffusion. The other parameter is the number of particles. The simulator was tested with an initial replicator seed: $e8a1b1c1d1f1$ of length 6 in addition with a number $x0$ with $x \in \{a,b,c,d,e,f\}$ varying from 60 to 6000 (from 5 to 1000 for each type). Results are displayed as mean $\pm$ standard-deviation (computed on 20 runs). For a perfect replication, molecules size average should be 6 – plotted as a dash line to guide the eye. A small jitter has been added on x values for clarity.

local interaction and spatial correlation can strongly modify bimolecular reactions either transiently (Van et al. (2014)) or at equilibrium (Caré and Soula (2011, 2013)) by modifying encounter probabilities.

In addition, there is another race condition on reaction R13. If R3 does not occur fast enough the $a3$ molecule can become linked with another $x6$ atom and turned into an $a2$ blocking the replication (see Fig.3B). This second race condition occurs because in the original scheme conformation reaction where instantaneous which is not the case anymore.

As we mentioned, this replication scheme is highly local and space dependent and can probably fail when confronted to well stirred and infinite crowding medium. In order to test the resilience of the replication, we created a reactor with an initial molecule $e8a1b1c1d1f1$ of size 6 and provided the medium initially with N of each type at state 0. This should start the replication process. All conformation reactions had given rate κ and bimolecular reaction of rate $\lambda\kappa$ (with $\lambda < 1$ the rate ratio to take into account diffusion).

We computed the average size of the molecules at the end of the process: when no reactions can occur i.e when $a = 0$. The results are displayed on Fig.4 with mean $\pm$ standard

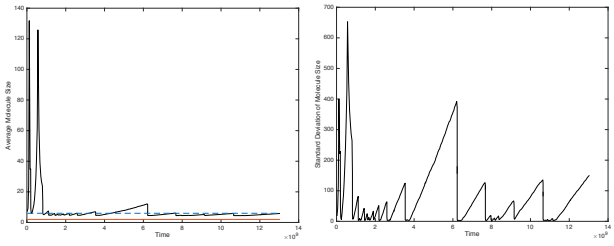

Figure 5: Results of long replication. Left: Average molecules size throughout time. Right: Standard-deviation of molecules size.

deviation (on 20 runs) for various N (so $6N$ total atoms) and various λ.

Normally when replication is occurring correctly the average size of the molecules at the end of the process should be 6 (dashed line displayed to guide the eye). For low values of N everything works correctly - except if λ is close to 1 and race conditions occur frequently stopping the replication earlier. This race condition 'error' occurs more and more frequently with increasing N because bimolecular reactions occurs more frequently with increasing concentrations. When λ is very small however most errors come from intertwined replication creating either big molecules ($N = 100$ and $\lambda = 10^{-6}$) or the minimal replicator $e8f1$ of size 2 ($N = 100$ and $\lambda = 10^{-6}$). Minimal replicators are the only stable replication seed that will ignore other replication interferences. In Hutton's original papers, he performed environmental 'wash' by setting all atoms to a zero state and unlinks them in a quadrant of the environment. He showed that the minimal replicator was 'selected' by this wash. We show here that it is enough to have a high population and ideal mixing to achieve the same result.

Long Replicator

We tested a longer episode of replications with initially 10 replicator seed $e8a1b1c1d1f1$ (of length 6) and no other particules. We've added the reactions of production/degradation for particles $x0$:

R17:	$(x\|0) \rightarrow \varnothing$
R18:	$\varnothing \rightarrow (x\|0)$

and let the system go for 450,000 reactions to occur (around 1.210^9 time steps). Here conformational rate were equal to 1 and bimolecular reaction's rate was 10^{-4}. Due to both degradation and production (note that once bound, an atom does not degrade anymore), the system slowly feeds atoms to the various replications seed. Since λ is low, most 'errors' are from entwined replications that create bigger and bigger molecules as seen on Fig.5. The average size increases at the start of the experiment and slowly settled to the actual size of the initial replication seeds: 6 (a dashed line is displayed to guide the eye; another line is for the size 2 of the minimal replicator). Since the number of particle grows through time averages are deceiving. Also on Fig.5, the standard deviation of the molecules size is displayed and showed that, at times, some molecules becomes extremely big compounds. As long as the replication process goes on they decreases to smaller size via separation. This alternation is particularly illustrated on Fig.6 which shows the number of replication according to time. High variation corresponds to big compounds slowly building – few replications per time unit – followed by quick disintegrations.

These results suggest that, when fed with atoms in a steady state manner, the replication scheme is fairly stable and in the end produces molecules whose sizes are the same as the initial replication seed.

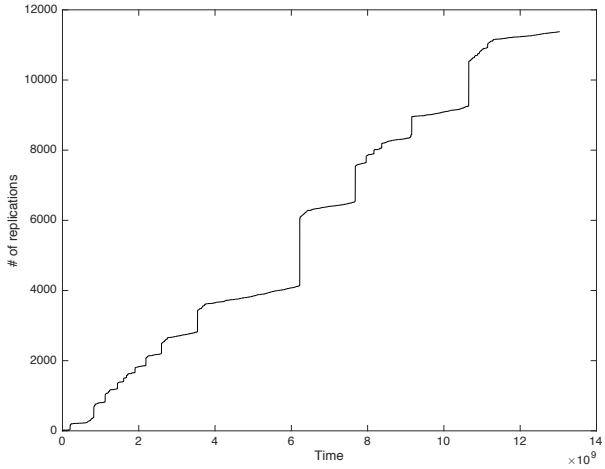

Figure 6: Number of replication through time (MC time) for the long replication events. Replication are obtained when rule R16 happened with an atom of type 'e'. Alternation of periods between extreme and slow replications are clearly visible.

Random Chemical World

The simulator allows us to compute several thousands of atoms and thousands of reactions. It is therefore possible to test use-cases where the reactions are drawn and chosen randomly and vary the amount of reactions available. Let the set of available set be $\{a, b, c\}$ and the maximum state being 5. We can compute all the possible rules in reactions, conformations, with no production nor degradation. We start with N particles $(t|s)$ with $t \in \{a, b, c\}$ and $0 \leq s \leq 4$ and $p \in [0, 1]$ describe the fraction of the reactions kept.

We simulate for a maximum of 2,000 reactions. We chose to have a rate of $1time^{-1}$ for conformation reactions and $1^{-2}time^{-1}$ r collision reactions. We simulated 20 different AC's for increasing fraction of p simulating the first 2,000 reactions for $N = 10,000$ initial atoms. In addition, we

provided larger molecule by adding random links to atoms (with increasing probability). We used a very simple metric by computing the ratio of the number of molecules between the start and the end of the experiment. These results are displayed on Fig. 7. When the number of chemical reactions available is low, the AC does not modify the structure of the particles graphs - even when it is already structured (orange star). Whereas a big number of reactions yields standardised particles graph with very low variability among chemistries. The interesting behaviour occurs at the transition where variability is at the highest between 10^{-4} and 10^{-3} fraction of possible reactions.

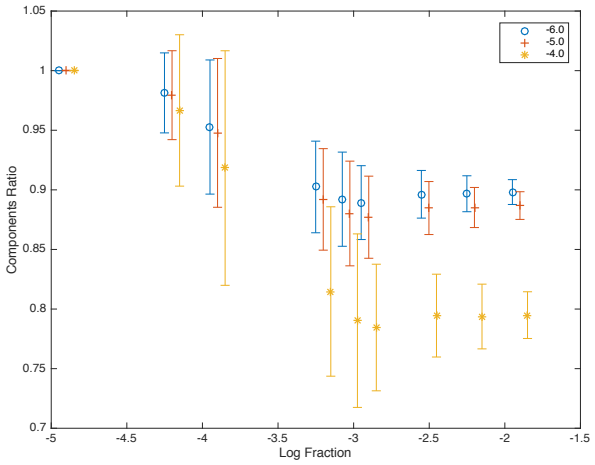

Figure 7: Results on the ratio of the number of connected components in the particles graph between the start and the end of the simulation as a function of the fraction (in log) of the reactions. Values are mean $\pm$ standard deviation. A small jitter has been added on x values for clarity.

The same results hold when looking at the dynamics i.e the actual time (in time steps) it took to complete the 2,000 reactions (see Fig 8). Conformation are the fastest reactions but must occur when molecules are already formed. Therefore interesting situation occurs when there is a mix of bimolecular and conformation to keep the system going.

Discussion

Artificial chemistries are extremely useful to understand and develop artificial life simulations. It will prove to be undoubtedly interesting in the future in the context of metabolic networks – either for theoretical considerations or for artificial recreation of entire cells. In this context, intricate and complex chemistries will be needed to create complex and open-ended simulated environment that could yield non trivial and emergent properties.

Among complex chemistries, we used the Hutton Artificial chemistry (HuAC) that, while very general, can generate complex chemistries by acting on nodes on molecules

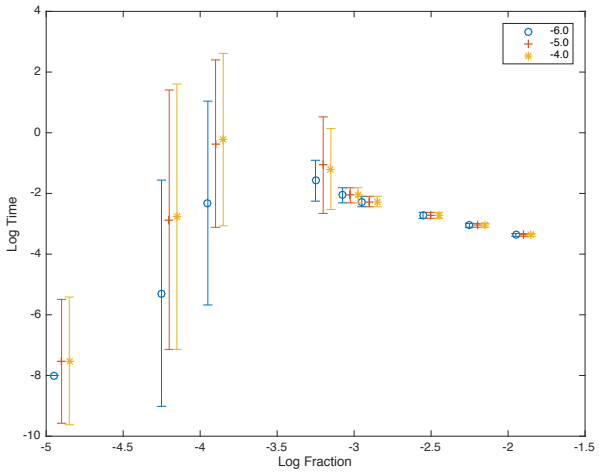

Figure 8: Total time (in log) at the end of the simulation (2,000 reactions) as a function of the fraction (in log) of the numbers of reactions. Values are mean $\pm$ standard deviation. A small jitter has been added on x values for clarity.

considered as connected graphs. The drawbacks of HuAC is linked to its spatial nature dependency – both in computational power and the unrealistic dependence on 2D constraints to function.

We provided here a Stochastic Simulation Algorithm that recreates HuAC in a fully stirred 3D medium. HuAC has problems of being locally constrained and had spatial correlation dependencies that do not exist in well- stirred medium. Even though we have argued and showed elsewhere that spatial correlations should not be ignored in the context of real cell signalling we contend that the loss of spatial correlation in the context of HuAC is a net gain due to its strongly 2D dependence.

We tested the impact on the Hutton's original replicator. We show that simple diffusion in a high concentration medium achieves the same result as his environmental 'wash'. We showed also that race conditions between reactions can stop the replication process altogether.

Our simulator allows to perform long simulation and we could test the replication process to its limit. By adding atoms stochastically, we showed that the result replication process was able to stabilise and provide correct replication size (on average) provided we waited long enough for episode of enormous compound to subside.

We finally tried to build Random Chemical World and see some of their simple properties when the reactions are chosen randomly. However this work is preliminary and random selection of reaction will not be susceptible to provide complex behaviours. A higher number of reaction that are linked together (e.g by being a chain between reactants and products) will probably yield more reactions time and

more molecules graphs modifications. In addition, the starting 'soup' of randomly connected atoms is also unlikely to provide interesting situations and we should investigate the impact on a initial set of bigger molecules. Not surprisingly, both these requirements are indeed met in the replicator situation. Therefore future works should include the study of random 'graphs' of reaction instead of random reactions.

The simulator, which that we called STAARC, can be used to scale up to very general and procedural chemistry. Scale up in terms of size and time but also in the number of reactions that can be handled. Our hope here is that it will prove to be well suited for study of artificial evolution.

Acknowledgments. I would like to thank V. Liard, D. Parsons and C. Lothe of the INRIA Team Beagle for their help.

References

Benkö, G., Centler, F., Dittrich, P., Flamm, C., Stadler, B. M., and Stadler, P. F. (2009). A topological approach to chemical organizations. *Artificial Life*, 15(1):71–88.

Benkö, G., Flamm, C., and Stadler, P. F. (2005). Explicit collision simulation of chemical reactions in a graph based artificial chemistry. In *Advances in Artificial Life*, pages 725–733. Springer.

Caré, B. R. and Soula, H. A. (2011). Impact of receptor clustering on ligand binding. *BMC Syst Biol*, 5:48.

Caré, B. R. and Soula, H. A. (2013). Receptor clustering affects signal transduction at the membrane level in the reaction-limited regime. *Phys Rev E Stat Nonlin Soft Matter Phys*, 87(1):012720.

Dittrich, P., Ziegler, J. C., and Banzhaf, W. (2001). Artificial chemistries—a review. *Artificial life*, 7(3):225–275.

Dorin, A. and Korb, K. B. (2007). Building virtual ecosystems from artificial chemistry. In *Advances in Artificial Life*, pages 103–112. Springer.

Ducharme, V., Egli, R., and Legault, C. Y. (2012). Energy-based artificial chemistry simulator. In *Artificial Life*, volume 13, pages 449–456.

Gillespie, D. T. (2007). Stochastic simulation of chemical kinetics. *Annual review of physical chemistry*, 58:35–55.

Gillespie, D. T., Hellander, A., and Petzold, L. R. (2013). Perspective: Stochastic algorithms for chemical kinetics. *J Chem Phys*, 138(17):170901.

Hutton, T. J. (2002). Evolvable self-replicating molecules in an artificial chemistry. *Artificial life*, 8(4):341–356.

Hutton, T. J. (2003). Simulating evolution's first steps. In *Advances in Artificial Life*, pages 51–58. Springer Berlin Heidelberg.

Hutton, T. J. (2004). A functional self-reproducing cell in a two-dimensional artificial chemistry. *Proceedings of the Ninth International Conference on the Simulation and Synthesis of Living Systems (ALIFE9)*, pages 444–449.

Hutton, T. J. (2007). Evolvable self-reproducing cells in a two-dimensional artificial chemistry. *Artificial life*, 13(1):11–30.

Hutton, T. J. (2009). The organic builder: A public experiment in artificial chemistries and self-replication. *Artificial life*, 15(1):21–28.

Knibbe, C. and Parsons, D. P. (2014). What happened to my genes? insights on gene family dynamics from digital genetics experiments. In *ALIFE 14: The Fourteenth Conference on the Synthesis and Simulation of Living Systems*, volume 14, pages 33–40.

Lenaerts, T. and Bersini, H. (2009). A synthon approach to artificial chemistry. *Artificial life*, 15(1):89–103.

Lucht, M. W. (2012). Size selection and adaptive evolution in an artificial chemistry. *Artif Life*, 18(2):143–63.

Oohashi, T., Ueno, O., Maekawa, T., Kawai, N., Nishina, E., and Honda, M. (2009). An effective hierarchical model for the biomolecular covalent bond: An approach integrating artificial chemistry and an actual terrestrial life system. *Artificial life*, 15(1):29–58.

Rocabert, C., Knibbe, C., and Beslon, G. (2015). Towards a Integrated Evolutionary Model to Study Evolution of Evolution. In *EvoEvo Workshop (Satellite workshop of ECAL 2015)*, York, United Kingdom.

Suzuki, H. and Dittrich, P. (2009). Artificial chemistry. *Artificial life*, 15(1):1–3.

Szabo, A. (1989). Theory of diffusion-influenced fluorescence quenching. *The journal of physical chemistry*, 93(19):6929–6939.

Tominaga, K., Suzuki, Y., Kobayashi, K., Watanabe, T., Koizumi, K., and Kishi, K. (2009). Modeling biochemical pathways using an artificial chemistry. *Artificial life*, 15(1):115–129.

Van, A. L., Soula, H. A., and Berry, H. (2014). Space-induced bifurcation in repression-based transcriptional circuits. *BMC Syst Biol*, 8:125.