The Effect of Reactant and Product Selection Strategies on Cycle Evolution in an Artificial Chemistry

Thomas J. Young and Kourosh Neshatian

University of Canterbury, Christchurch, New Zealand thomas.young@pg.canterbury.ac.nz

Abstract. The molecules within an Artificial Chemistry form an evolutionary system, capable under certain conditions of displaying interesting emergent behaviours. We investigate experimentally the effect on emergence of the combinations of selected strategies for choosing reactants (Uniform and Kinetic selection) and products (Uniform and Least Energy selection) as measured by three measures of reaction cycle formation. Emergence is maximised by a Kinetic reactant selection strategy; the choice of product selection strategy has minimal effect.

Keywords: artificial chemistry, emergence, open-ended evolution.

1 Introduction

Artificial Chemistries of discrete atoms provide an interesting testbed for investigating various evolutionary phenomena. Fundamentally, they provide a tuneable evolutionary system, capable of highly complex behaviour, built around familiar metaphors (real-world Chemistry, and potentially Biology). A set of interaction rules describing how atoms interact gives rise to emergent forms—molecules. At a higher level, these molecules, under the same interaction rules, also interact in patterns—reactions.

Still higher emergent levels emerge under favourable conditions. Reactions may form cycles, where a sequence eventually returns to an earlier product. Our interest is in identifying the factors that influence the emergence of these higher levels. Cycles in particular are interesting as many biological processes are cyclical. Replication, resulting in an exact copy of an entity, is a macro-example of a cycle; metabolism is another. Building on the apparent correspondence between higher emergent levels in Artificial Chemistry evolution and Biology, we believe like others (e.g., [18]) that cycles, of some form, are a necessary building-block for more complicated structures again in Artificial Chemistries.

Unfortunately, although the emergence of cycles from a solely reaction-based system is certainly possible, and indeed likely under many conditions, more complicated structures are very rare. Perhaps it is simply a matter of probabilities—they may be theoretically possible, but in practice highly unlikely and so most

© Springer International Publishing Switzerland 2015

S.K. Chalup et al. (Eds.): ACALCI 2015, LNAI 8955, pp. 310-322, 2015.

reaction sequences do not do anything interesting: a strongly constructive chemistry can generate an infinite number of reaction types, and it is of course possible for these to combine in sequences that are interesting. But from other work the probability of this happening appears to be very low [3,16].

In an attempt to improve these odds we turn to heuristics, or strategies, inspired by our analogy, the real-world, to tune the Artificial Chemistry. We categorise these strategies into two types—those to do with the selection of reactants for the next reaction, and those to do with the selection of products from the set of all products possible given a particular set of reactants. The combination of the two completely describes a reaction. In this work we explore the following research questions:

- RQ1: Is there a quantitative difference between different reactant and product selection strategies?
- RQ2: Is there a combination of reactant and product selection strategies that leads to increased emergence as measured by cycles?
- RQ3: Is emergence significantly affected by the values of other parameters of an Artificial Chemistry, such as initial kinetic energy or bond energies?

To the best of our knowledge, this is the first time that reaction and product selection strategies in Artificial Chemistries have been experimentally compared. Instead, the general approach of previous work, where there has been a quantitative evaluation, has been to propose a particular strategy, build, and evaluate against the initial goals, rather than against alternatives.

2 Previous Work

Artificial Chemistries are often found employed in three main areas: as models for the study of real-world chemistry; to explore biochemical processes, often in connection with questions regarding the origins or life; and finally purely for the exploration of artificial life.

In the first two areas, the primary requirement is fidelity with real-world chemistry, which requires either a library of empirically derived reaction definitions and rates, or a model capable of accurately simulating quantum-mechanical processes. The latter approach has been taken by a family of Artificial Chemistries, beginning with Benkö et al [1], built on Extended Huckel Theory with parameters taken directly from chemical experiments and later extended (for example in [2]) to a general purpose model with parameters derived from theoretical chemistry. The model was used in [10] for the study of the behaviour and topology of chemical reaction networks, specifically Diels-Alder and Formose reaction networks, and in a series of papers (e.g, [7] and [20]) for the examination of the evolution of metabolic networks in early organisms using a simple model of RNA coding for catalysts.

Real-world chemical processes are also important to modelling scenarios for the origin of life or of other related areas such as the formation of metabolic networks in the earliest protocells. In many cases though the specific focus is less on the bottoms-up model from the most elementary elements, and more on taskbased models of processes where the particular starting point is predetermined by the researcher.

For example, in Lattice Artificial Chemistry [17,15], the study of membrane formation and cell division assumes five different types of particles (some hydrophilic and some hydrophobic) that together form an autocatalytic cycle similar to those observed in biological cells. Three types of particle are employed by the Substrate-Catalyst-Link (or SCL) chemistry of [21,19]: the eponymous Substrate, Link and Catalyst. Cells are formed from links around a catalyst, with a single predefined reaction rule $S + S + C \Rightarrow L + C$ and some straightforward constraints on movement of the particles in the matrix (for example, bonded Link particles cannot cross each other.)

Most such work concentrates on the behaviour of individuals; by contrast in [5] the focus is on an ecosystem, based on a set of atoms interacting in pre-specified ways that represent biological photosynthesis, respiration and biosynthesis (or growth). The goal is to explore the interactions in an ecosystem made up of a set of organisms pre-built to perform various defined roles.

Finally, in Artificial Life, Artificial Chemistries have been used in the exploration of open-ended or creative evolution. Squirm3 [11,12,14] adopts fixed molecule types, and pre-defined reactions for replication and gene-sequence transcription, and so although capable of interesting behaviour is not capable of unlimited extension. Stringmol [9] - a bacterial inspired microprogram chemistry - though does demonstrate a rich heredity for open-ended evolution using string-matching to model binding between sequences, and RBN-World [6] shows that a form of Random Boolean Network, with the addition of a bonding mechanisms to allow for composition and decomposition of RBNs, can be used to build a chemistry capable of almost limitless extension out of non-traditional components.

3 The ToyWorld Artificial Chemistry

ToyWorld, our Artificial Chemistry for the exploration of emergent behaviours, was first introduced in [22]. The elements of the model - Atoms, Molecules, Reactions, a Reaction Vessel - are recognisable from real-world chemistry, but in highly simplified forms. Familiarity is important for understanding, but only in so far as the analogy is consistent, and therefore we endeavour to maintain a basic correspondence wherever possible. However, there is no requirement to provide chemically-realistic results - our model cannot be used to investigate real-world chemical behaviours.

The lowest level component in the ToyWorld model is the atom, and atoms can be joined by bonds to form molecules. Reactions between molecules are the only mechanism in the model to modify molecules; a reaction is simply the addition or subtraction of a single bond between any two atoms in two molecules. ToyWorld provides a strongly constructive chemistry ([8]) where completely new forms of molecules may be generated by reactions, and where the new molecules may in turn take part in further reactions: the chemistry emerges from the lower level atomic properties.

All atoms, and therefore molecules and reactions, are contained within a reaction vessel. ToyWorld provides a basic energy model, where molecules have kinetic energy and bond breaking requires energy input and bond formation releases energy. The reaction vessel, which provides the strategies by which reaction reactants (or input molecules) and products (output molecules) are determined, is described in detail in the following section.

4 The Reaction Vessel – Reactant and Product Selection Strategies

Importantly for this work, a reaction may be seen as two stages in sequence: first, the choice of reactants from a population of possible reactant molecules (the Reactant selection strategy, denoted here by S_{Reactant}), and second, the determination of products given that set of reactants (Product selection strategy, denoted S_{Product}).

4.1 Selecting Reactants for a Reaction

Two generic strategies are described in [6] for the selection of reactants—spatial and aspatial—where the primary difference is whether molecular position is a factor in reactant selection. It is possible to further generalise this scheme by considering other differentiating factors. Analogous with real-world chemistry, a cumulative scheme presents itself starting with the pure aspatial, or uniform probability strategy, and then proceeding through the spatial strategy, based on molecular kinetics, to kinetics plus intra-molecular and external forces such as electromagnetism. These strategies can be viewed as being based on increasing derivatives of position or location in the reaction vessel; from no position (uniform selection), through fixed position (uninteresting as we cannot have a sequence of reactions without motion) to the first derivative (velocity or kinetic selection) and finally to the second derivative (acceleration, or force selection.) Accordingly we adopt this more detailed classification for the descriptions below.

Uniform Selection. In a uniform selection strategy ($S_{\text{Reactant}} = \text{Uniform}$), reactants are chosen at random with equal (uniform) probability from the population: no property of a molecule has an effect on the selection. Conceptually we have a well-stirred reaction container with no intra-molecular forces.

Kinetic Selection. By contrast, in a kinetic selection strategy ($S_{\text{Reactant}} =$ Kinetic) molecules have spatial position (and implicitly, velocity) within some assumed reaction vessel, and selection is determined by molecular position—molecules which are spatially co-located (that is, in collision) form a reactant set. Molecules move at constant velocity until they collide with something else

(either another molecule or possibly a boundary of an explicit reaction vessel) and then either react, or bounce. Currently in our work we assume that all molecules have a fixed and common size and shape (circular in two-dimensions), irrespective of molecular formula.

Intra-molecular Selection and External Force Selection. More complicated forms, where molecular velocities are not constant, can be generated by the introduction of some combination of intra-molecular forces (such as electromagnetism) or external forces (such as gravity or heat.)

4.2 Determining the Products of a Reaction

For each reaction we can generate a number of alternative product sets [22,2] by enumerating all possible single bond additions, bond subtractions, and changes in bond type between the reactants (more complicated alternatives can be generated by combining these single bond changes).

How should we choose between alternative sets of possible products for the same reactants? Various product strategies appear plausible: the random choice of an alternative; the most complex alternative; least complex; rarest; most common, and so on, but each strategy requires effort to develop and evaluate. We have chosen to focus in this work on a strategy which supports an argument by analogy, where there is a reasonable parallel between the strategy and real-world chemistry: the strategy of Least Energy.

When following a Least Energy strategy ($S_{Product}$ = LeastEnergy) we select a reaction by choosing with uniform probability from a distribution of reaction alternatives weighted by the total of the energy changes associated with the bond changes. This biases selection towards the Least Energy alternative; the strength of the bias is determined by the degree of the weighting.

As an experimental control, we also evaluate a strategy with minimal bias: a Uniform selection strategy ($S_{Product} = Uniform$), where every alternative product set has equal probability of selection.

5 Evaluation

Following on from the research questions, our two primary factors, or independent variables, are S_{Reactant} and S_{Product} . We also introduce two secondary factors, overall reaction vessel energy (E_{Vessel}) and bond energy (E_{Bonds}), to assess the sensitivity of the simulation to other parameters. For simplicity of analysis, all of our factors are two-level, meaning they take one of two possible levels, or values, in each run. The parameter values chosen for each level of E_{Vessel} and E_{Bonds} were chosen as representative from a set of alternatives used in initial exploratory experiments; in each case they allowed the simulation to run for an extended period without running out of possible reactions (from lack of energy for example.)

Factor	+1 value	-1 value	Description
$S_{ m Reactant} \ S_{ m Product} \ E_{ m Vessel}$	Kinetic LeastEnergy 300	Uniform Uniform 100	See Section 4.1 See Section 4.2 Initial kinetic energy of each molecule in the meeting model
$E_{ m Bonds}$	Single=50, Double=100, Triple=200	Simplified real-world chem- istry. Average values for Single=77.7, Double=148.2, and Triple=224.3	Energy required to break a bond of the given type

Table 1. Factors, or independent variables

We concentrate on three related response, or dependent, variables—Number of cycles, Length of longest cycle, and Count of most common cycle. All three are derived from a reconstruction of the network of reactions that occur during each experiment run, where every edge represents a specific reaction connecting a particular set of reactants with a particular set of products. Note that the nodes in the constructed network capture specific molecules, rather than molecular types or species that share the same chemical formula (as would be more usual in the construction of a Reaction Network for real-world chemistry.)

We exclude all unique cycles, and all cycles with three or fewer elements (for example, where a molecule loses, then regains, an atom repeatedly). Unique cycles by nature are unlikely to be representative; very short cycles on the other hand are so common as to dominate other more interesting cycles in any analysis.

5.1 Experiment Design

The experiments follow a full factorial design over four factors ($S_{\text{Reactant}}, S_{\text{Product}}, E_{\text{Vessel}}$ and E_{Bonds}), each at two levels, run in a randomized order, with three (3) replicates of each combination of factors executed in sequence before beginning the next combination. The first replicate of each combination starts with a predefined random seed incremented by one for each successive replicate of the same combination. The factor levels used are given in Table 1.

Each replicate used the same initial population of 800 molecules, made up of 100 molecules each of [H][H], O=O, [O-][N+](=O)[N+]([O-])=O, and N(=O)[O] and 200 molecules each of O and O=C=O (all represented in SMILES [4].) This initial population is somewhat arbitrary, although reasonable; given that Toy-World is a strongly constructive chemistry, we would expect that any differences between initial populations would reduce as the simulation proceeds.

For details of the Artificial Chemistry, see [22]. The chemistry makes use of some low-level components from RDKit [13], open-source software for cheminformatics. RDKit provides a number of useful capabilities, including format conversions to and from SMILES and graphical forms of molecules; standard sanity checks for molecular structure, and molecular manipulations. In Toy-World, atoms are closely based on real-world chemistry atoms, and in fact are implemented as wrappers around the Atom definitions provided by RDKit; we allow any atom type provided by RDKit. Bonds in ToyWorld are represented by RDKit bonds, but the addition or subtraction mechanism makes use of the parameterised ToyWorld energy model.

6 Results

All replicates completed a set of 20,000 reactions; given the initial population size of 800 molecules, and from the summary of results below, we believe that this captures a representative set of reactions. This also simplifies the analysis as we can assume a balanced set of treatments in the statistical sense (that is, the sample sizes for all treatments are equal).

Statistic	Number of cycles	Length of longest.cycle	Count of most common cycle		
Reactions 4750 to 5000					
Min.	0.00	0.00	0.00		
1st Quartile	0.00	0.00	0.00		
Median	1.50	3.50	2.50		
Mean	219.06	5.04	215.80		
3rd Quartile	91.25	7.75	96.00		
Max.	5704.00	20.00	2728.00		
Reactions 9750 to 10000					
Min.	0.00	0.00	0.00		
1st Quartile	0.00	0.00	0.00		
Median	6.00	4.00	6.00		
Mean	62.10	4.65	169.21		
3rd Quartile	68.75	8.25	27.75		
Max.	526.00	13.00	6684.00		
Reactions 14750 to 15000					
Min.	0.00	0.00	0.00		
1st Quartile	1.00	3.00	2.00		
Median	5.00	4.50	5.00		
Mean	27.17	4.79	42.27		
3rd Quartile	34.50	7.00	16.25		
Max.	237.00	12.00	862.00		
Reactions 19750 to 20000					
Min.	0.00	0.00	0.00		
1st Quartile	0.00	0.00	0.00		
Median	3.50	4.00	4.00		
Mean	20.04	3.90	14.62		
3rd Quartile	20.25	6.00	13.25		
Max.	199.00	12.00	216.00		

Table 2. Summary of results over 2^4 runs of 3 replicates

A view of the results is given in Table 2: reaction networks built from the full dataset of 20,000 reactions can be too large for easy analysis. Instead, we choose to partition the reaction data into four equally spaced blocks of 250 reactions each and analyse each block independently.



Fig. 1. Cycles by reaction partition (starting reaction number for each partition along x-axis)

7 Analysis and Discussion

Figure 1 suggests that the first partition, representing the vessel a quarter of the way into its lifespan, is quantitatively different from the other three partitions, with a significantly greater range for all three response variables. Intuitively this corresponds with an initial period where the diversity in the reaction vessel rapidly increases from the limited starting set of molecules, as seen in some (e.g, Figure 2) but not necessarily all of the replicates. Diversity here is measured by (average molecular quantity)⁻¹. All following sections therefore exclude data from the first partition of reaction numbers from 4750 to 5000.

7.1 RQ1: Is There a Quantitative Difference between the Different Reactant and Product Selection Strategies?

From visual inspection of Figure 3, there appears to be a significant difference between the Uniform and Kinetic reactant selection strategies for number and length of cycles. Kinetic reactant selection seems to result in significantly higher levels of emergent behaviour than Uniform reactant selection. Similarly, from Figure 4, there is very little apparent difference between the two product strategies, Uniform selection and Least Energy selection.

We use ANOVA (Analysis of Variance) to further examine the relationship of S_{Reactant} and S_{Product} to the response variables using a two-factor with twolevels (2x2) model (degrees of freedom=1) with interaction effects. There is a



Fig. 2. Diversity for two example replicates (12-0 and 16-1)



Fig. 3. Response by S_{Reactant}

highly significant difference (p<0.001) between the Uniform and Kinetic reactant selection strategies when comparing the number of cycles (f-value=40.442) and length of cycles (f-value=361.891) (confirming the impression given by Figure 3), although again without difference for the count of the most common cycle. The effect of S_{Product} on cycle number and length is also significant (f-value=4.050 and 5.705 respectively, p<0.05) and there is a first-order interaction between S_{Reactant} and S_{Product} for number of cycles (f-value=4.011, p<0.05).



Fig. 4. Response by S_{Product}

7.2 RQ2: Is There a Combination of Reactant and Product Selection Strategies That Leads to Increased Emergence as Measured by Cycles?

From Figure 5c it is clear that there is no significant relationship between strategy and the number of occurrence of the most common cycle. However, from Figures 5a and 5b it seems that such a relationship does exist for the number and length of cycles, with the strongest effect as a result of S_{Reactant} , and a lesser effect from the choice of S_{Product} .

We conclude that the greatest levels of emergence are likely to be seen with the combination of $S_{\text{Reactant}} = \text{Kinetic}$ and $S_{\text{Product}} = \text{LeastEnergy}$.

7.3 RQ3: Is Emergence Significantly Affected by the Values of Other Parameters of an Artificial Chemistry, Such as Initial Kinetic Energy or Bond Energies?

We constructed a two-factor with two-levels (2x2) ANOVA model (degrees of freedom=1) with interaction effects to examine the relationship of the independent variables E_{Vessel} and E_{Bonds} to the response variables, and applied it to our dataset (summarised in Table 2). E_{Bonds} is significant (f-value=4.221, p<0.05) to number of cycles. No other significant relationships exist.



(a) Cycle Count by Strategy Combination $(S_{\text{Reactant}}:S_{\text{Product}})$

(b) Cycle Length by Strategy Combination $(S_{\text{Reactant}}:S_{\text{Product}})$



(c) Count of Most Common Cycle by Strategy Combination (S_{Reactant}:S_{Product})



8 Conclusions

The choice of S_{Reactant} is critical to the behaviour of an emergent Artificial Chemistry; S_{Product} on the other hand appears to have a lesser effect on the emergence of cycles in our experiments. Furthermore, S_{Reactant} = Kinetic is more effective for cycle emergence than S_{Reactant} = Uniform.

The most significant limitation of our analysis overall is that the values chosen for the high and low values of E_{Bonds} make it impossible to determine the cause of the difference observed in RQ3. There are two alternative explanations: first, the energy required to make or break bonds is simply different between the two factor levels; second, in the low factor level, based on real-world values, the bond make and break energies for even a single bond vary depending on the atoms involved, while in the high factor level these values are consistent for all bonds of the same degree. To distinguish between the two explanations we would need at least the average levels at each degree to be the same for each factor; this is a suggestion for a future experiment.

In future work we intend to examine the sensitivity of the results to parameter selection, and to extend the ToyWorld Reaction Vessel to include the option of intra-molecular forces such as are seen in real-world chemistry between charged regions on adjacent molecules. These forces give rise to accelerations, which would allow us to continue the Reactant selection strategy series that extends from uniform selection (no physics) through location then velocity and finally to acceleration.

References

- Benkö, G., Flamm, C., Stadler, P.F.: A graph-based toy model of chemistry. Journal of Chemical Information and Computer Sciences 43(4), 1085–1093 (2003), http://pubs.acs.org/doi/abs/10.1021/ci0200570, PMID: 12870897
- Benkö, G., Flamm, C., Stadler, P.F.: The toychem package: A computational toolkit implementing a realistic artificial chemistry model (2005), http://www.tbi.univie.ac.at/~xtof/ToyChem/
- Channon, A.: Unbounded evolutionary dynamics in a system of agents that actively process and transform their environment. Genetic Programming and Evolvable Machines 7(3), 253–281 (2006), doi:10.1007/s10710-006-9009-3
- 4. Daylight Chemical Information Systems, Inc. Daylight theory manual (2011), http://www.daylight.com/dayhtml/doc/theory/index.html
- Dorin, A., Korb, K.B.: Building virtual ecosystems from artificial chemistry. In: Almeida e Costa, F., Rocha, L.M., Costa, E., Harvey, I., Coutinho, A. (eds.) ECAL 2007. LNCS (LNAI), vol. 4648, pp. 103–112. Springer, Heidelberg (2007)
- Faulconbridge, A.: RBN-World: sub-symbolic artificial chemistry for artificial life. Ph.D. thesis, University of York (2011)
- Flamm, C., Ullrich, A., Ekker, H., Mann, M., Hogerl, D., Rohrschneider, M., Sauer, S., Scheuermann, G., Klemm, K., Hofacker, I., Stadler, P.: Evolution of metabolic networks: a computational frame-work. Journal of Systems Chemistry 1(1), 4 (2010), http://www.jsystchem.com/content/1/1/4
- Fontana, W., Wagner, G.P., Buss, L.W.: Beyond digital naturalism. Artificial Life 1(2), 211–227 (1994)
- Hickinbotham, S., Clark, E., Stepney, S., Clarke, T., Nellis, A., Pay, M., Young, P.: Molecular microprograms. In: Kampis, G., Karsai, I., Szathmáry, E. (eds.) ECAL 2009, Part I. LNCS, vol. 5777, pp. 297–304. Springer, Heidelberg (2011), http://dx.doi.org/10.1007/978-3-642-21283-3_37
- Högerl, D.: Simulation of prebiotic chemistries. Master's thesis, Institute for Theoretical Chemistry, University of Vienna (2010)
- Hutton, T.J.: Evolvable self-replicating molecules in an artificial chemistry. Artificial Life 8(4), 341–356 (2002),

http://www.sq3.org.uk/Evolution/Squirm3/EvSelfReps/

 Hutton, T.J.: The organic builder: A public experiment in artificial chemistries and self-replication. Artificial Life 15(1), 21-28 (2009), http://dx.doi.org/10.1162/artl.2009.15.1.15102 (October 9, 2012)

- 13. Landrum, G.: Rdkit: Open-source cheminformatics (2013), http://www.rdkit.org
- Lucht, M.W.: Size selection and adaptive evolution in an artificial chemistry. Artificial Life 18(2), 143–163 (2012)
- Madina, D., Ono, N., Ikegami, T.: Cellular evolution in a 3D lattice artificial chemistry. In: Banzhaf, W., Ziegler, J., Christaller, T., Dittrich, P., Kim, J.T. (eds.) ECAL 2003. LNCS (LNAI), vol. 2801, pp. 59–68. Springer, Heidelberg (2003)
- Maley, C.: Four steps toward open-ended evolution. In: GECCO-99: Proceedings of the Genetic and Evolutionary Computation Conference, pp. 1336–1343. Morgan Kaufmann (1999)
- 17. Ono, N., Ikegami, T.: Self-maintenance and self-reproduction in an abstract cell model. Journal of Theoretical Biology 206(2), 243-253 (2000), http://www.sciencedirect.com/science/article/pii/S0022519300921210
- Steel, M., Hordijk, W., Smith, J.: Minimal autocatalytic networks. Journal of Theoretical Biology 332(0), 96-107 (2013), http://www.sciencedirect.com/science/article/pii/S0022519313002002
- Suzuki, K., Ikegami, T.: Shapes and self-movement in protocell systems. Artificial Life 15(1), 59-70 (2008), http://dx.doi.org/10.1162/artl.2009.15.1.15104 (October 9, 2012)
- Ullrich, A., Flamm, C., Rohrschneider, M., Stadler, P.F.: In silico evolution of early metabolism. In: Fellermann, H., Dorr, M., Hanczyc, M., Laursen, L.L., Maurer, S., Merkle, D., Monnard, P.A., Sta, Y.K., Rasmussen, S. (eds.) Proceedings of the Twelfth International Conference on the Synthesis and Simulation of Living Systems (Artificial Life XII), The MIT Press, Cambridge (2010)
- 21. Varela, F.G., Maturana, H.R., Uribe, R.: Autopoiesis: The organization of living systems, its characterization and a model. Biosystems 5(4), 187–196 (1974)
- Young, T.J., Neshatian, K.: A constructive artificial chemistry to explore openended evolution. In: Cranefield, S., Nayak, A. (eds.) AI 2013. LNCS, vol. 8272, pp. 228–233. Springer, Heidelberg (2013)