A Molecular Approach to **Complex Adaptive Systems**

James Decraene, George Mitchell, Barry McMullin

Artificial Life Laboratory - Research Institute for Networks and Communications Engineering School of Electronic Engineering - Dublin City University, Ireland Email: james.decraene@eeng.dcu.ie

Abstract-Complex Adaptive Systems (CAS) are dynamical networks of interacting agents which as a whole determine the behavior, adaptivity and cognitive ability of the system. CAS are ubiquitous and occur in a variety of natural and artificial systems (e.g., cells, societies, stock markets). To study CAS, Holland [1], [2] proposed to employ an agent-based system in which Learning Classifier Systems (LCS) were used to determine the agents behavior and adaptivity. We argue that LCS are limited for the study of CAS: the rule-discovery mechanism is pre-specified and may limit the evolvability of CAS. Secondly, LCS distinguish a demarcation between messages and rules, however operations are reflexive in CAS, e.g., in a cell, an agent (a molecule) may both act as a message (substrate) and as a catalyst (rule). To address these issues, we proposed the Molecular Classifier Systems (MCS.b) [3], a string-based Artificial Chemistry based on Holland's broadcast language. In the MCS.b, no explicit fitness function or rulediscovery mechanism is specified, moreover no distinction is made between messages and rules. In the context of the ESIGNET project¹, we employ the MCS.b to study a subclass of CAS: Cell Signaling Networks (CSNs) which are complex biochemical networks responsible for coordinating cellular activities. As CSNs occur in cells, these networks must replicate themselves prior to cell division. In this paper we present a series of experiments focusing on the self-replication ability of these CAS. Results indicate counter intuitive outcomes as opposed to those inferred from the literature. This work highlights the current deficit of a theoretical framework for the study of Artificial Chemistries.

I. INTRODUCTION

Modeling and evolving CAS remains problematic as the traditional analytical and statistical approaches (coupled with some Evolutionary Algorithms) appear to limit the study of CAS [2]. To overcome these issues, Holland [1], [2] proposed to employ agent-based systems in which Learning Classifier Systems (LCS) were used to determine the agents behavior and adaptivity.

We note that for certain subclasses of CAS, Holland's approach should be refined. For example, an instance of CAS which is of current interest is Cell Signaling Networks (CSNs). CSNs are complex biochemical networks of interacting molecules (proteins, ions, secondary messengers, etc) occurring in living cells. Through complex molecular interactions (e.g., signal transduction), CSNs are able to coordinate cellular activities (cell differentiation, programmed cell death) in response to internal and external stimuli.

We suggest that for such CAS, Holland's LCS method has a number of limitations:

- Firstly, in traditional LCS [4], a distinction is made between messages and rules. However molecules can be considered *reflexive* in nature, in the sense that they can act as both messages (substrates) and rules (enzymes). This reflexivity property was partially addressed in the LCS's precursors: Holland's broadcast language (BL) [5]. However, due to the lack of study no implementation of the BL existed (until recently [6]). Although the BL proved to be inappropriate when applied to the modeling and evolving of biochemical networks, we employ a simplification of the BL in our approach which will be later described in this paper.
- Secondly, the LCS involves a credit assignment algorithm (such as the bucket brigade algorithm) which is employed to reward and strengthen efficient rules. Moreover a rulediscovery mechanism is specified which is responsible for generating potentially more efficient rules. These algorithms are pre-specified and do not evolve. These attributes may stifle the occurrence of "perpetual novelty" during evolution [7]. Therefore the performance of the system (as defined by the fitness function) may be limited during long term evolution. An alternative would be to define an Evolutionary Algorithm (EA) that would evolve these mechanisms, however this would present one with yet another problem: how to specify the EA fitness function? The latter is fixed and is potentially another point where novelty may be stifled, recreating the creditassignment and rule discovery mechanism problem. To avoid this infinitely recursive problem, we avoid defining any explicit fitness/rewarding functions but rather utilize an "implicit" function which will be discussed in the remainder of this paper.

To explore our "molecular" approach to Holland's LCS for the study of CAS, we employ the Molecular Classifier System version 2 called MCS.b. This Artificial Chemistry (AC) addresses the reflexive nature of molecular species and employs an implicit fitness function to drive the evolution of the system. The implicit fitness function is addressed with the "self-replication" ability of the individual molecules and of the system (cell) as a whole. We first introduce the MCS.b and then present a series of experiments focusing on the self-

¹ESIGNET: Evolving Cell Signaling Networks in silico, an EU FP6 project, contract no. 12789, http://www.esignet.net

replication ability which determines the survival/performance of the CAS.

II. MOLECULAR CLASSIFIER SYSTEMS

Molecular Classifier Systems are a class of string-rewriting based AC (inspired by LCS) that was proposed to investigate protocell computation [8]. As opposed to traditional stringrewriting systems, operations are stochastic and reflexive (no distinction made between operands and operators). The behavior of the condition (binding) properties and action (enzymatic functions) is defined by a language specified within the MCS. This "chemical" language defines and constrains the complexity of the chemical reactions that may be modeled and simulated. In this AC, a stochastic flow reactor is employed, i.e., reactions occur at random and there is a constant inflow and outflow of molecules.

A reaction between molecules occurs if the conditional part of a molecule A matches another molecule B. A is regarded as an enzyme whereas B is regarded as a substrate molecule. When a reaction occurs, the action part of molecule A is utilized to perform the enzymatic operations upon the bound substrate molecule B. This operation results in the production of another offspring (product).

As introduced earlier, we proposed a simplification of the broadcast language which is used as the MCS chemical language (MCS.b). A number of differences exist between the BL and the LCS, e.g., the LCS's alphabet is $\lambda = \{1, 0, \#\}$ whereas the BL includes additional symbols $\Lambda = \{1, 0, p, *, :, \Diamond, \nabla, \Psi, \triangle, '\}$. The basic elements of the BL are strings made from Λ called *broadcast devices*. A broadcast device is parsed into zero, one or more *broadcast units*, where each unit represents a single condition/action rule. The symbol * separates broadcast units within a broadcast device. The symbol : separates a condition from an action within a single broadcast unit. $\{\Diamond, \nabla, \Psi, \triangle\}$ are single/multiple character(s) wildcards that may also transpose matched strings into output strings. Table I presents an analogy between the biological and broadcast language terminology.

TABLE I Comparison of biological and broadcast language terminology

Biology	Broadcast Language
sequence of amino acids from	string of symbols from Λ =
$\{A, R, N, D, C, E, \ldots\}$	$\{0, 1, *, :, \diamond, \nabla, \nabla, \Delta, p, '\}$
substrate	input signal
product	output signal
protein with no enzymatic function	null unit
enzyme	broadcast unit
protein complex	broadcast device
cellular milieu	list of strings from Λ

A detailed description is omitted in this paper, see [9] for full specification of the BL implementation.

III. SELF-REPLICATION IN COMPLEX ADAPTIVE SYSTEMS

A series of experiments that have been conducted on Artificial Cell Signaling Networks (ACSNs) is now outlined to illustrate this molecular approach to CAS. To initiate this investigation on CSNs using the MCS approach, the self-replication ability of these classes of CAS is examined. As CSNs occur in cells, these networks have to replicate themselves prior to the cellular division. This allows the replicated CSNs to be "distributed" to the offspring cells. Errors may occur during this replication process, e.g., an offspring cell may inherit only a partial CSN. Thus resulting in a potentially defective cell which would lead to a variety of undesired effects (e.g., premature cell death). As a result, the "fitness" of a cell is implicitly represented by the *survival* and *performance* of a cell in achieving cell-level replication.

Examining such replication (or autocatalytic) phenomena closely relates to other studies which have been conducted on Holland's alpha-universe [10], Tierra [11] and Alchemy [12]. Although these ACs were developed for different purposes and were implemented differently, these systems exhibited common evolutionary phenomena such as the emergence of (collectively) replicating reaction networks [13], [14]. In this investigation, such classes of network are of interest as they would allow CSNs to maintain and replicate themselves. Moreover, as demonstrated in several ACs, it is commonly accepted that the emergence of these collectively autocatalytic reaction networks is trivial.

To explore this issue with this system, it is first examined whether it is possible to observe the spontaneous emergence of autocatalytic molecules (i.e., molecules that can self-replicate) which was obtained in Alchemy but not in Tierra.

A. Specificity and domination of the self-replicators

An artifact of the BL's syntax is that it is very difficult to observe the emergence of an autocatalytic molecule: e.g., if we consider a population containing all broadcast devices (molecules) of length 5 symbols, there would be 5^{10} possible combinations of molecules, in which only 5 molecules in this entire population would be autocatalytic. The probability of obtaining a replicator given a random population of molecules increases linearly with the length of the molecules and size of the population.

Although the probability of obtaining such autocatalytic molecules is low in the MCS.b, the intuition was: since such autocatalytic molecules are able to self-replicate, they would have a greater fitness than other non-autocatalytic molecules. If such a molecule appears, it would therefore be able to out compete other molecules and fill rapidly the reaction space. This phenomenon was indeed observed by Fontana in Alchemy and was expected to occur in the MCS.b. In the remainder of this section, a series of experiments is discussed to examine this domination phenomenon of autocatalytic molecules in the MCS.b. The simplest form of self-replicator that may be built with the BL is first presented:

$SR_0 = *\nabla : \nabla$

The behavior of SR_0 is now detailed: The matching condition of SR_0 is defined by a single symbol: ∇ which designates a multiple character wildcard. This indicates that SR_0 may bind to any molecule. In addition when a reaction occurs between SR_0 and a substrate molecule I_0 , ∇ is assigned a value: the informational string of I_0 . A unique symbol ∇ also constitutes the action part of SR_0 , this denotes that the output of SR_0 would be equal to the value of ∇ that is located in the condition part, i.e the output string of SR_0 has for value the input string. Therefore the broadcast device SR_0 is not only a self-replicator but a "universal" replicator meaning that it would replicate any binding molecules. The "specificity" of SR_0 is said to be *null*.

Fig. 1 presents a first experiment examining the behavior of SR_0 averaged over 30 simulation runs. In this initial series of experiments, the broadcast "universe" is initialized with the following parameters:

- The system is seeded with 90 randomly generated molecules, each of length 10 symbols.
- In addition, 10 replicators SR_0 are inserted.
- n_{max} designates the maximum number of molecules that may be contained in the universe, $n_{max} = 1000$.
- 50 molecular interactions occur at each time step.
- An interaction occurs when: two molecules A and B are picked at random, A is considered as a catalyst and B as a substrate. If A can bind and react with B then a molecule C is produced. If the current size of the population, n, is less than n_{max} then C is simply added to the population (and n increases by 1); otherwise a molecule is picked at random and is replaced with C (and the population size remains fixed at n).
- No mutation may occur in these experiments.

A high relative concentration (0.1) of SR_0 was selected to discourage an "accidental" early extinction. The latter may be caused by a high reaction rate of the side reactions.

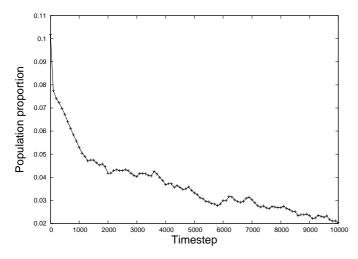


Fig. 1. Relative population growth of replicators SR_0 averaged over 30 simulation runs.

From Fig. 1 we note that SR_0 replicators do not take over the population but eventually diminish—which contradicts the result that would be naively expected from the presence of a single, unique, self-replicating molecular species in the population. The explanation is that the *binding specificity* of self-replicators can strongly influence the system dynamics: a self-replicase molecule having low or zero specificity will not only replicate itself but also replicate a more or less wide variety of other molecules. These experiments demonstrate that such a species, although autocatalytic, will not *necessarily* grow to dominate the molecular population; on the contrary, it may even go extinct.

Self-replicator	Informational string
SR_0	$* \nabla$: ∇
SR_1	$* \nabla 1 : \nabla 1$
SR_2	$* \nabla 01 : \nabla 01$
SR_3	$* \nabla 101 : \nabla 101$
SR_4	*⊽0101 : ⊽0101

TABLE II

SPECIFICATION OF SELF-REPLICATORS WITH INCREASING SPECIFICITY

To verify this hypothesis, we proceeded to a series of experiments in which we incrementally increased the specificity of the self-replicators (see Fig. 2). In Table II, the different selfreplicators employed in these experiments are presented. SR_1 designates a molecule that would only react with molecules ending with the symbol 1. As the latter occurs at the rightmost position of SR_1 , it may react with itself, producing another instance of SR_1 . Similarly, SR_2 only binds to molecules containing the suffix 01. This "signature" forms a constraint on the replicators, specifying them to react only with a specific subset of molecules. This impacts directly on these molecules' binding specificity.

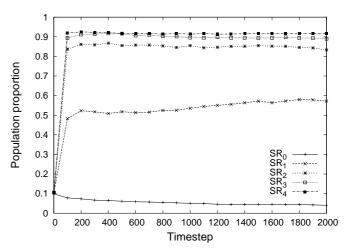


Fig. 2. Population growth of replicators SR_0 , SR_1 , SR_2 , SR_3 and SR_4 . Each series is averaged over 30 simulation runs.

The results depicted in Fig. 2 confirm the importance of specificity upon the system dynamics: the tendancy for a self-replicator to dominate the population increases with (and is ultimately limited by) its binding specificity. Therefore in this system, for autocatalytic molecules to successfully sustain themselves and/or dominate the molecular population, a significant binding specificity is required. This phenomenon may well have been implicated in previous results reported for a variety of artificial chemistries; but, to our knowledge, it has

not previously been *explicitly isolated* in the manner presented here.

B. Spontaneous emergence of self-replicators

In the previous set of experiments, mutation was turned off in order to facilitate our investigation on self-replicators, which were hand-designed and inserted into the initial population. This led to a limited diversity in the population. To examine the spontaneous emergence of autocatalytic molecules, we performed a second series of experiments in which no selfreplicators are specified and molecular mutation could occur. The latter is implemented as follows:

- When a new molecule is produced, a mutation with probability $p_{sym} = 0.001$ may be applied upon each symbol contained in the molecule. Therefore, the longer the molecule, the higher the probability of mutation occurring.
- Three types of mutation are distinguished and may be applied with equal chances:
 - Symbol flipping: The current symbol is replaced with a symbol picked at random from Λ .
 - Symbol insertion: A symbol from Λ is picked and inserted after the current symbol.
 - *Symbol deletion:* The current symbol is removed from the molecule's informational string.
- In addition, to encourage diversity over time, a global mutation (per molecule) is defined: every 100 timesteps, a subset ($r_{mut} = 0.01$) of the population is selected at random. A mutation is then applied to a single symbol picked at random in each molecule present in this subset.

As mutation now occurs, diversity is increased during long term evolution. It was expected to observe the spontaneous appearance of replicators in this system. Results obtained from multiple series of experiments indicated that self-replicators do emerge but they never manage to self-sustain. The reasons for this behavior could be explained as follows:

- As indicated earlier, the BL's syntax does not strongly facilitate the spontaneous emergence of replicators. This syntactical constraint may discourage the spontaneous emergence of self-replicators. The BL's syntax may also have an impact on the robustness of these self-replicators against mutation effects.
- Secondly if self-replicators replicators do emerge, they would be required to possess a specificity higher than null to sustain themselves.
- Finally, replicators are likely to possess a low molecular concentration when emerging. This low concentration diminishes the capacity of these molecular species to persist against side reactions and mutation events.

These three factors, when combined, may significantly lower the probability of having a replicator spontaneously emerge and self-sustain in the MCS.b.

We examined the nature of the self-replicators that may emerge during evolution. An additional set of experiments was specified as follows:

- 30 simulations were performed (each for 100000 timesteps).
- The different populations were seeded with randomly generated molecules (with length 10 and relative concentration 0.1).

During these experiments, self-replicators emerged and presented the following properties:

- 13 distinct self-replicators appeared.
- A common broadcast unit of the form *∇ : ∇ was found in all self-replicators. This broadcast unit refers to the universal replicator SR₀ (having zero specificity) presented earlier.
- Finally, the highest relative concentration ever achieved by these universal replicators was $[SR]_{max} = 0.0014$.

These observations support the explanation provided earlier. As a result, these self-replicators could not sustain themselves and consequently go extinct.

Although it is *still* theoretically possible (given enough time) to observe the spontaneous emergence and domination of a self-replicator in the MCS.b (as in Alchemy), multiple findings suggest that this system shares a common property with Tierra: the spontaneous emergence of self-replicators is *unlikely* to be observed. Our next logical step was to mirror the Tierra system and introduce an "ancestor". The latter was designed and employed to counter balance this effect. This is presented in the next section.

C. Rise and fall of the fittest

In the Tierra system, a hand-designed molecule called *the* ancestor was introduced into the population which encouraged the emergence of collectively autocatalytic reaction networks. We adopted a similar approach in which we introduced a hand-designed autocatalytic molecule (having a high binding specificity such as SR_4). Results indicated that the MCS does not exhibit the same evolutionary dynamic as Tierra:

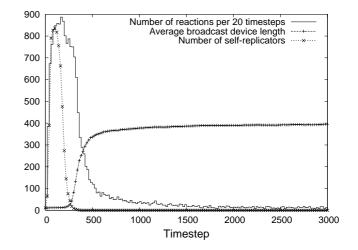


Fig. 3. Effects of molecules length growth upon overall system reactions rates. In this experiment, an ancestor $(SR_4 = \nabla 0101 : \nabla 0101)$ is inserted (with initial relative concentration $[SR_4] = 0.1$) in addition to randomly generated molecules. Moreover mutation per molecule and per symbol is turned on.

Fig. 3 presents an experiment in which we may observe that the self-replicators first quickly fill ($n_{max} = 1000$) the reaction space as expected. This led to a high overall reaction rate of the system. However we note then that along with the decrease in the reaction rates and in the number of self-replicators, the average length of the molecules increase. In the current experiment settings, we limited the maximum length of the molecules to $BD_{lmax} = 500$ due to computational constraints. Other experiments showed that with no limit upon the length of the molecules, an indefinite length growth would be observed leading to a critical impact on the system performance.

In keeping with the previous sets of experiments which were discussed earlier, these results were not expected. In fact, certain mutants of the original autocatalytic molecule developed a distinct advantage over the ancestor. That is, these mutants could be replicated by the ancestor molecules but only to the cost of these ancestors, i.e., an asymmetric relationship. Moreover, some of these mutants also lose their ability to selfreplicate, explaining the rapid decrease in the global number of self-replicators. By exploiting their molecular signature and the ancestors, these non-autocatalytic molecules succeed in displacing the dominant ancestors.

During evolution, a series of such parasitic displacements is observed leading to longer molecules. However, as the mutation effect increases (with the length of a molecule), disruptive structural changes may occur more frequently. The consequences of these structural changes are twofold:

- Molecules may become inactive, being therefore considered only as substrate molecules.
- The binding specificity may be increased (reducing connectivity).

Ultimately, this leads to a decrease in the overall reaction rates, (until reactions cease completely, i.e., system death). In this approach, the stepwise emergence of "fitter" molecules leads effectively to system- or population-level extinction.

IV. FUTURE WORK

These preliminary experiments presented unexpected outcomes, this highlights the fact that there is currently no theoretical framework for the study of ACs. Future work includes the development of this theoretical foundation which would allow a better understanding of evolutionary dynamics in a given AC. Supplementary experimental work will be conducted, involving the exploitation of our implicit function to evolve CAS to carry out pre-specified tasks: To evolve these artificial biochemical networks toward a pre-specified goal, it is proposed to extend this system in which a second level of selection would be introduced. In the system presented in this paper, molecules were competing with each other, referring to a single level of selection. Our extended system involves a second level of selection: molecules (still competing with each other) would be placed in containers (cells). These artificial cells would also compete with each other: in keeping with molecules, this competition would be based on the cells' ability to self-replicate. Cells would self-replicate if their internal

energy reservoir reach a certain threshold (concentration of specific molecules).

As was recently demonstrated [15], a second level of selection may prevent parasitism from occurring in Molecular Classifier Systems. Moreover, this future system shares a number of common features with Holland's approach [1], [2] but differs on several points. Let us examine how these might differ in our future evolutionary system.

Our system will probably differ from Holland's approach on the following points:

- Broadcast devices employ an adaptable representation as opposed to classifiers which utilize a *fixed* representation (that do not change over time).
- No distinction is made between messages and classifiers. Both are specified as broadcast devices that may be active (classifiers) or not (messages).
- Broadcast devices are also employed to specify detectors. These detectors may be satisfied by broadcast devices detected in the environment. When satisfied, detectors generate messages (signaling molecules) that would be inserted in the cell. These inserted broadcast devices may then trigger a cascade of reactions within the cell.
- Similarly, effectors are specified as broadcast devices. Effectors can only be satisfied by molecules originating from the cells. Molecules resulting from the effectors' action statements are inserted in the environment. The latter would then react in accord with the molecules generated by the effectors (e.g. update the agent's location).

This extended system refines Holland's proposal and improves biological plausibilities: molecules are reflexive (acting as messages and/or classifiers) and are evolvable (representation is dynamic). Secondly, no explicit fitness function is defined preventing the credit assignment problem. These refinements may encourage open-ended evolution and improve the system's performance during long term evolution.

V. CONCLUSION

A molecular approach to study a subclass of CAS (CSNs) was presented. We first discussed the limitations of Holland's proposal which was based on LCS. Following this, the MCS.b, an Artificial Chemistry based on Holland's broadcast language, was then briefly depicted. The MCS.b accounts for the reflexive nature of molecules. Moreover, to encourage openended evolution, this system employs a fitness function that is not explicit but implicit. This implicit fitness function was addressed with the self-replication ability of CSNs. A number of preliminary experiments focusing on self-replication was then discussed. This experimental work was closely related to studies conducted in other ACs. However, these experiments exhibited unexpected results as opposed to those inferred from the literature. We identified a major cause which led to these unexpected outcomes: the molecules' binding specificity. We demonstrated that molecular specificity plays an important role and may influence significantly the system dynamics. This work highlights the current deficit of a theoretical framework for the study of ACs. The latter is part of this project's future

work. In addition, further experiments will be conducted to evolve molecular CAS to carry out pre-specified tasks. An extended system was proposed and introduced a second level of selection. Moreover, this proposed system would exploit our implicit fitness function by adding constraints on the selfreplication process.

ACKNOWLEDGMENT

This work was funded by ESIGNET (Evolving Cell Signaling Networks in Silico), an European Integrated Project in the EU FP6 NEST Initiative (contract no. 12789).

REFERENCES

- J. Holland, "Exploring the evolution of complexity in signaling networks," *Complexity*, vol. 7, no. 2, pp. 34–45, 2002.
- [2] —, "Studying Complex Adaptive Systems," Journal of Systems Science and Complexity, vol. 19, no. 1, pp. 1–8, 2006.
- [3] J.Decraene, G.G. Mitchell, B. McMullin, "Evolving Artificial Cell Signaling Networks: Perspective and Methods," *Advances in Biologically Inspired Information Systems: Models, Methods, and Tools*, vol. 69, 2007, to appear.
- [4] J. Holland and J. Reitman, "Cognitive systems based on adaptive algorithms," SIGART Bull., no. 63, pp. 49–49, 1977.
- [5] J. Holland, Adaptation in natural and artificial systems. MIT Press Cambridge, MA, USA, 1992.
- [6] J. Decraene, G. G. Mitchell, B. McMullin, and C. Kelly, "The holland broadcast language and the modeling of biochemical networks," in *Proceedings of the 10th European Conference on Genetic Programming*, ser. Lecture Notes in Computer Science, M. Ebner, M. O'Neill, A. Ekárt, L. Vanneschi, and A. I. Esparcia-Alcázar, Eds., vol. 4445. Valencia, Spain: Springer, 11 - 13 Apr. 2007.
- [7] D. Groß and B. McMullin, "The creation of novelty in artificial chemistries," in *ICAL 2003: Proceedings of the eighth international conference on Artificial life*. Cambridge, MA, USA: MIT Press, 2003, pp. 400–409.
- [8] B. McMullin, C. Kelly, D. OBrien, G. G. Mitchell, and J. Decraene, "Preliminary Steps toward Artificial Protocell Computation," in *Proceedings of the 2007 International Conference on Morphological Computation*, Venice, Italy, March 2007.
- [9] J. Decraene, "The Holland Broadcast Language," Artificial Life Lab, RINCE, School of Electronic Engineering, Dublin City University, Tech. Rep. ALL-06-01, 2006.
- [10] J. Holland, "Studies of the spontaneous emergence of self-replicating systems using cellular automata and formal grammars," *Automata, Languages, Development*, pp. 385–404, 1976.
- [11] T. Ray, "An approach to the synthesis of life," *Artificial Life II*, vol. 10, pp. 371–408, 1991.
- [12] W. Fontana and L. Buss, "What Would be Conserved if" the Tape were Played Twice"?" *Proceedings of the National Academy of Sciences*, vol. 91, no. 2, pp. 757–761, 1994.
- [13] P. Dittrich, J. Ziegler, and W. Banzhaf, "Artificial Chemistries-A Review," Artificial Life, vol. 7, no. 3, pp. 225–275, 2001.
- [14] B. McMullin, "Remarks on Autocatalysis and Autopoiesis," Annals of the New York Academy of Sciences, vol. 901, no. 1, pp. 163–174, 2000.
- [15] B. McMullin, C. Kelly, and D. O'Brien, "Multi-level selectional stalemate in a simple artificial chemistry." in Advances in Artificial Life, 9th European Conference, ECAL 2007, Lisbon, Spain, September 10-14, 2007, Proceedings, ser. Lecture Notes in Computer Science. Springer, 2007.