

## Artificial Life as an Aid to Astrobiology: Testing Life Seeking Techniques

Florian Centler<sup>1</sup>, Peter Dittrich<sup>1</sup>, Lawrence Ku<sup>2</sup>,  
Naoki Matsumaru<sup>1</sup>, Jeffrey Pfaffmann<sup>2</sup>, and Klaus-Peter Zauner<sup>1</sup>

<sup>1</sup> Jena Centre for Bioinformatics (JCB) and  
Department of Mathematics and Computer Science  
Friedrich-Schiller-University Jena  
D-07743 Jena, Germany

<http://www.minet.uni-jena.de/csb/>

<sup>2</sup> Department of Computer Science  
Wayne State University  
Detroit, MI 48202, U.S.A.

**Abstract.** Searching for signatures of fossil or present life in our solar system requires autonomous devices capable of investigating remote locations with limited assistance from earth. Here, we use an artificial chemistry model to create spatially complex chemical environments. An autonomous experimentation technique based on evolutionary computation is then employed to explore these environments with the aim of discovering the chemical signature of small patches of biota present in the simulation space. In the highly abstracted environment considered, autonomous experimentation achieves fair to good predictions for locations with biological activity. We believe that artificially generated biospheres will be an important tool for developing the algorithms key to the search for life on Mars.

### 1 Life Detecting Machines

A spot of rust on a sheet of steel has much in common with a spot of mold on a sheet of fabric. Discerning alive matter among dead matter is a challenge even in our familiar surroundings. The ability to draw this line with some certainty is a rather recent achievement [1]. The U.S. National Aeronautics and Space Administration outlined a program to detect signs of fossil or present life within our solar system and beyond [2]. Crucial to this endeavor is the capability of recognizing life in whatever form it may take:

“A strategy is needed for recognizing novel biosignatures. [...] For example, certain examples of our biosphere’s specific molecular machinery, e.g., DNA and proteins might not necessarily be mimicked by other examples of life elsewhere in the cosmos. On the other hand, basic principles of biological evolution might indeed be universal.” [2, p. 19]

This broad conception of biota was inherent in artificial life research from its very onset; in Langton’s words:

“[...] certainly, the dynamic processes that constitute life—in whatever material bases they might occur—must share certain universal features—features that will allow us to recognize life by its dynamic form, without reference to its matter.” [3, p. 2]

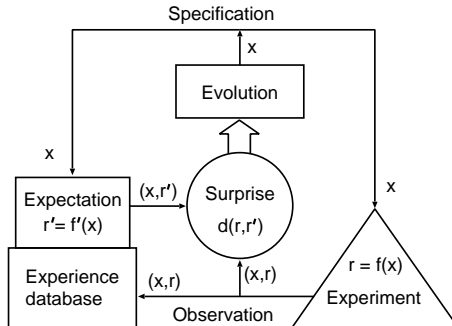
He went on to point to the usefulness of synthetic life to provide context for the known forms of life. Thus artificial life would substitute for the lack of samples of life forms based on other materials. At the time, biologists may have dismissed the need for a context as broad as Langton envisioned. But when it comes to the quest of searching real life beyond the scope of our immediate environment, a narrow perspective on biology may in fact preclude the discovery of those life forms that would allow us to strengthen our as of yet weak notion of what really constitutes life. Accordingly, the abstraction artificial life aims at can make a significant contribution to the recognition of unknown life forms.

In the following, we report on the application of artificial chemistry modeling [4] to evaluate an autonomous experimentation technique with regard to its capability to detect chemical signatures of life.

### 1.1 Autonomous Experimentation

The search for signs of extraterrestrial biota is characterized by vast, hard to access areas and severe restrictions on communication bandwidth. Instrumentation with a high degree of automation is required. Autonomous experimentation utilizes computational discovery methods [5, 6] to orchestrate the available sensor resources. In contrast to the majority of machine discovery algorithms that operate on a fixed dataset, autonomous experimentation goes beyond data analysis and include the experiments itself in the discovery process [7, 8]. Kulkarni and Simon [9] demonstrated that an algorithm can successfully emulate the interplay of adjusting hypotheses and modifying experiments characteristic of human experimenters [10]. The integration of such algorithms into space probes [11] and planetary rovers [12] is key to the search for life beyond earth.

Here we employ an autonomous experimentation technique based on evolutionary computation. This technique, named *scouting*, has initially been developed to characterize protein response with regard to chemical signals [13]. Figure 1 illustrates its operation. The control computer simulates an evolutionary population. Each genome in the population represents a specification of experimental conditions (labeled  $\mathbf{x}$  in Fig. 1). The first step toward the evaluation of the fitness of a genome is the generation of a prediction  $\mathbf{r}'$  for the outcome of the experiment specified by the genome under consideration. This prediction is based on a database of experiences in which the observations from all experiments that have been conducted up to this point are stored. Subsequently the experiment specified by the genome is actually performed and provides an observation  $\mathbf{r}$  for an experiment conducted under conditions  $\mathbf{x}$ . The more  $\mathbf{r}$  deviates from what was expected for such an experiment (i.e.,  $\mathbf{r}'$ ), the larger the fitness value that is assigned to the genome that specified the experiment. In other words, specifying experiments that yield a lot of information is rewarded. As a consequence



**Fig. 1.** Scouting combines the notion of information being equivalent to surprise value [14] with evolutionary computation [15] for autonomous exploration. See text for details

the next generation of the population in the evolutionary process will focus its attention on parameter combinations in the vicinity of the conditions that gave rise to the surprising experimental result.

The scouting method has been applied in wet-lab experiments in conjunction with a computer controlled fluidics system to autonomously investigate the response of the enzyme malate dehydrogenase to ion signals [16]. In this domain the scouting algorithm focuses measurements on areas of the parameter space that show unusual phenomena. However, the parameters set by the algorithm can also be spatial coordinates, in which case parameter space and real space coincide. The algorithm then probes the space for unusual observations.

The selection of measurement locations is of particular interest for instruments that can be applied locally only. Multi-touch surface sampling with altitude-controlled Montgolfiere balloons and air-born laser spectroscopy are examples. It is in this context that the artificial life simulation described in the next section provides a test scenario for autonomous experimentation.

## 2 Simulation of an Artificial Planetary Sphere

To create a test scenario for the study of the behavior of the scouting algorithm we simulated an abstract artificial planetary sphere with an inanimate chemistry that can be perturbed by locally introduced life forms. The assumption here is that life necessitates the synthesis of a larger variety of substances than commonly produced by inorganic reactions.

For simulating spatially inhomogeneous chemistry a two-dimensional cellular automata implementation of the chemical dynamics is convenient [17]. We employed an asynchronously updated, probabilistic variant in which the state of a cell represents a chemical component present at the location of the cell. All rules take the form  $A + B \xrightarrow{p} C + D$ . If component B is present within the set of four (von Neumann) neighbors of component A, then A may be substituted

by C and B substituted by D. The probability for this event to take place is proportional to  $p$ . In principle the states of the cells could represent a complex local chemical composition; for simplicity we use the abstraction of a single substance. Note that the defined order of substitution in the rules enables diffusion to be expressed by sets of rules of the form  $A + B \xrightarrow{p} B + A$ .

## 2.1 The Chemical Model

As the basis for our model we selected a randomly created artificial chemistry for which all modes of organization (see next section) are known [18]. In this abstraction of molecular interaction, substances are created by the cooperative action of two catalytically acting substances. This was motivated by biopolymer synthesis and entails the assumption of an inexhaustible pool of building blocks being available. In contrast, the present model does take substrates explicitly into account, such that all reactions are of the form  $X \xrightarrow{(E,F)} Y$ , where X is a substrate molecule, E and F are co-acting catalysts and Y is the reaction product. To realize reactions in which a substrate, a product, and two catalysts participate with the binary reaction scheme of the cellular automata, intermediate agents have been introduced for each reaction. This is in agreement with real chemistry, where elementary steps are assumed to be predominantly binary and at most ternary. Thus, above reaction would be represented in the cellular automata by two rules:

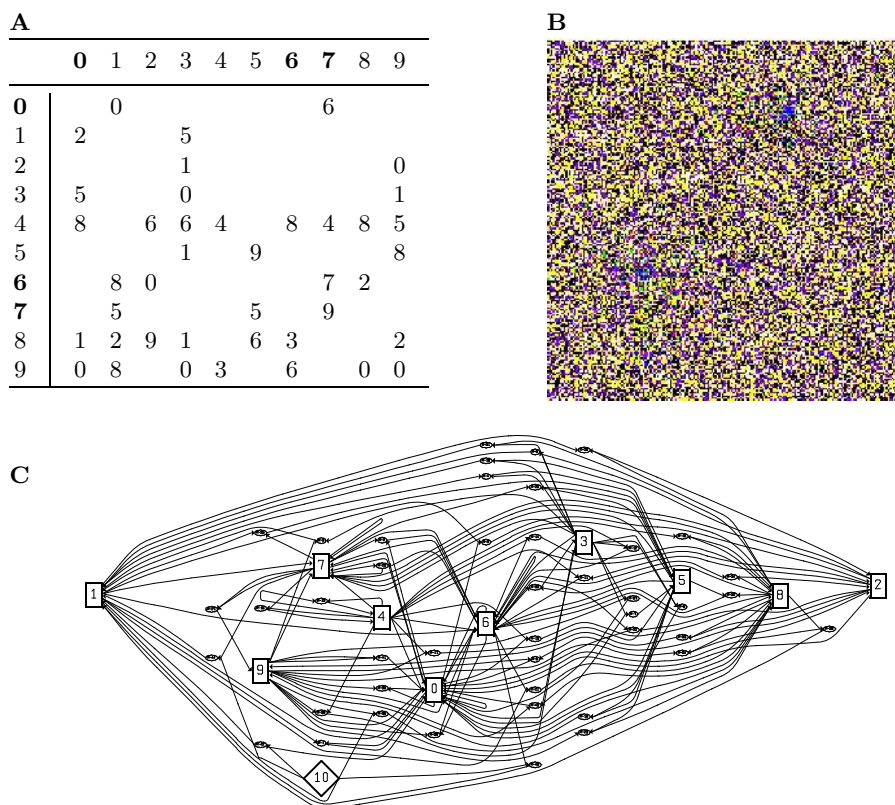


The intermediate agent I is assumed to be highly reactive and leads to a rapid transformation of substrate X to product Y, i.e.,  $p_1 \ll p_2$ . Any one of a set of three arbitrarily selected special substances  $\{0, 6, 7\}$  can serve as substrate. Figure 2A shows the reaction network taken from [18]. Note that the reaction matrix is asymmetric, which effectively corresponds to the possibility of two products being catalyzed by the cooperative action of the two catalysts.

For the present purpose the reaction network shown in Fig. 2A was augmented by decay reactions that transform any substance other than intermediates into the substrate substances through rules of the form  $A + B \xrightarrow{p} X + B$ , where  $A \in \{0, 1, \dots, 9\}$ ,  $X \in \{0, 6, 7\}$  and B can take any value. Decay reactions for all but substance 1 have a probability of 0.075%. Substance 1 is instable and decays with the normal reaction probability of 0.75%. The probability for diffusion is 25%. Finally, an additional catalyst, number 10, was introduced as *biota*. It consumes the substrate molecules  $\{0,6,7\}$  and excretes the instable substance 1. All substances, but not the biota had the ability to diffuse. The resulting reaction network, illustrated in Fig. 2C, was simulated with the cellular automata.

## 2.2 Simulation of Spatial Chemical Composition

A cellular automata implementation of the model described above was run in a bounded simulation space of  $200 \times 200$  cells. This space was initialized with a



**Fig. 2.** (A) The network of potential reactions. Numbers in the leftmost column refer to substances that take the place of E in equation 1, numbers in the top row correspond to F in equations 1 and 2, and entries in the table correspond to Y in equation 2. Substances that can serve as substrates are marked with bold face. (B) Snapshot of the states of the  $200 \times 200$  cell automaton after  $10^8$  individual updates. The different shades indicate cell states corresponding to different chemical compositions. All intermediate products are mapped to the same shade. (C) The reaction network used to model the planetary sphere. Numbered boxes represent the substances. The diamond shape (10) indicates the biota's interaction with its chemical environment. Reactions are shown as tiny nodes; for clarity the 123 fast intermediate reactions are not shown

pseudo-random distribution<sup>3</sup> of four of the ten substances, namely  $\{0, 6, 7, 9\}$ . These four substances form a closed and self-maintaining reaction system. In other words, they do not catalyze the production of any substance not present in the system (property of closure) and every substance is catalytically produced within the reaction system (property of self-maintenance). A reaction system that is closed and self-maintaining is called *organization* [19, 20]. The chemical

<sup>3</sup> For our simulations, we initialized always: 12762 cells with substance 0, 12556 cells with substance 6, 13419 cells with substance 7, and 1263 with substance 9.

organization of the reaction system may change if it is perturbed by the introduction of new chemical species, such as substance 1 excreted by the biota. For all simulations in which the biota was present, exactly six cells were set to the biota state.

Diffusion rules ( $A+B \xrightarrow{p_D} B+A$  with  $p_D = 0.25\%$ ), here an abstraction of all transport phenomena, were implemented for all substances but not for the biota. The total number of rules, including diffusion rules for all substance combinations (inclusive intermediate substances) is 2873. To update the simulation space a cell and one of its von Neumann neighbors are pseudo-randomly selected. The subset of rules applicable to the two chemical components located in the two cells is determined and in accordance with their probabilities one of these rules may be applied to transform the content of the cells. Figure 2B indicates the distribution of chemical compounds in the simulation space after an average number of 2500 updates per cell. The grey level distribution is indicative of the complexity of spatial distribution of substances. The chemical composition of local areas in the simulation space shows considerable fluctuation. It provides a rich testbed for the scouting algorithm because in general no accurate prediction of the local chemical composition is possible. The dynamics of the artificial chemistry model was simulated and snapshots of the cellular automata state, such as depicted in Fig. 2B, were saved during the run to serve as input for the scouting algorithm.

### 3 Scouting Experiments

In the experiments described here the scouting algorithm is applied to detect unusual chemical signatures in a complex background chemistry [21]. It is assumed that the chemical composition can be sampled locally, but no a priori knowledge of the chemical effects of potentially present biota is available to the algorithm. For sampling the algorithm has to choose a location  $\mathbf{x} = (x, y)$  in the simulation space. The measured data at this location is a vector  $\mathbf{r} = (r_1, \dots, r_n)$  containing for each of the  $n$  substances the fraction of cells in a  $\pm 2$  pixel vicinity that hold this substance. Correspondingly the entries  $(\mathbf{x}, \mathbf{r})$  in the experience database and the expectations  $(\mathbf{x}, \mathbf{r}')$  formed for a location prior to its sampling are such vectors (cf. Fig. 1). The surprise value  $d$ , which constitutes the fitness criterion, is computed as the Shannon entropy [14] of the difference between the actual response  $\mathbf{r}$  and response that was expected  $\mathbf{r}'$ . The expectation  $\mathbf{r}'$  for a location  $\mathbf{x} = (x, y)$  is computed as distance-weighted average over the (up to) 25 measurements nearest to  $\mathbf{x}$  available in the experience database. A population size of 10 genomes, all offspring of the single best parent was used for evolution (i.e., a (1, 10)-strategy [15]). We mutate an offspring by adding an equally distributed random number taken from within a radius of  $\delta$ . The mutation strength  $\delta$  depends on the previous surprise

$$d(\mathbf{r}, \mathbf{r}') = - \sum_{i=1}^n |r_i - r'_i| \ln |r_i - r'_i|, \quad \delta = \begin{cases} 0.03 & \text{if } \sqrt{4.2} \leq d(\mathbf{r}, \mathbf{r}'), \\ 0.04 & \text{if } 2.0 \leq d(\mathbf{r}, \mathbf{r}') < \sqrt{4.2}, \\ 0.1 & \text{if } \sqrt{3.5} \leq d(\mathbf{r}, \mathbf{r}') < 2.0, \\ 0.99 & \text{else .} \end{cases} \quad (3)$$

In the scouting algorithm applied for the experiments described below, these parameters have been established empirically and kept constant for all experiments. Methods to adapt the mutation strength dynamically during the sampling are under development. It should be noted that due to the highly dynamic fitness landscape, adaptation methods commonly used in evolutionary computation do not perform well in scouting (unpublished results).

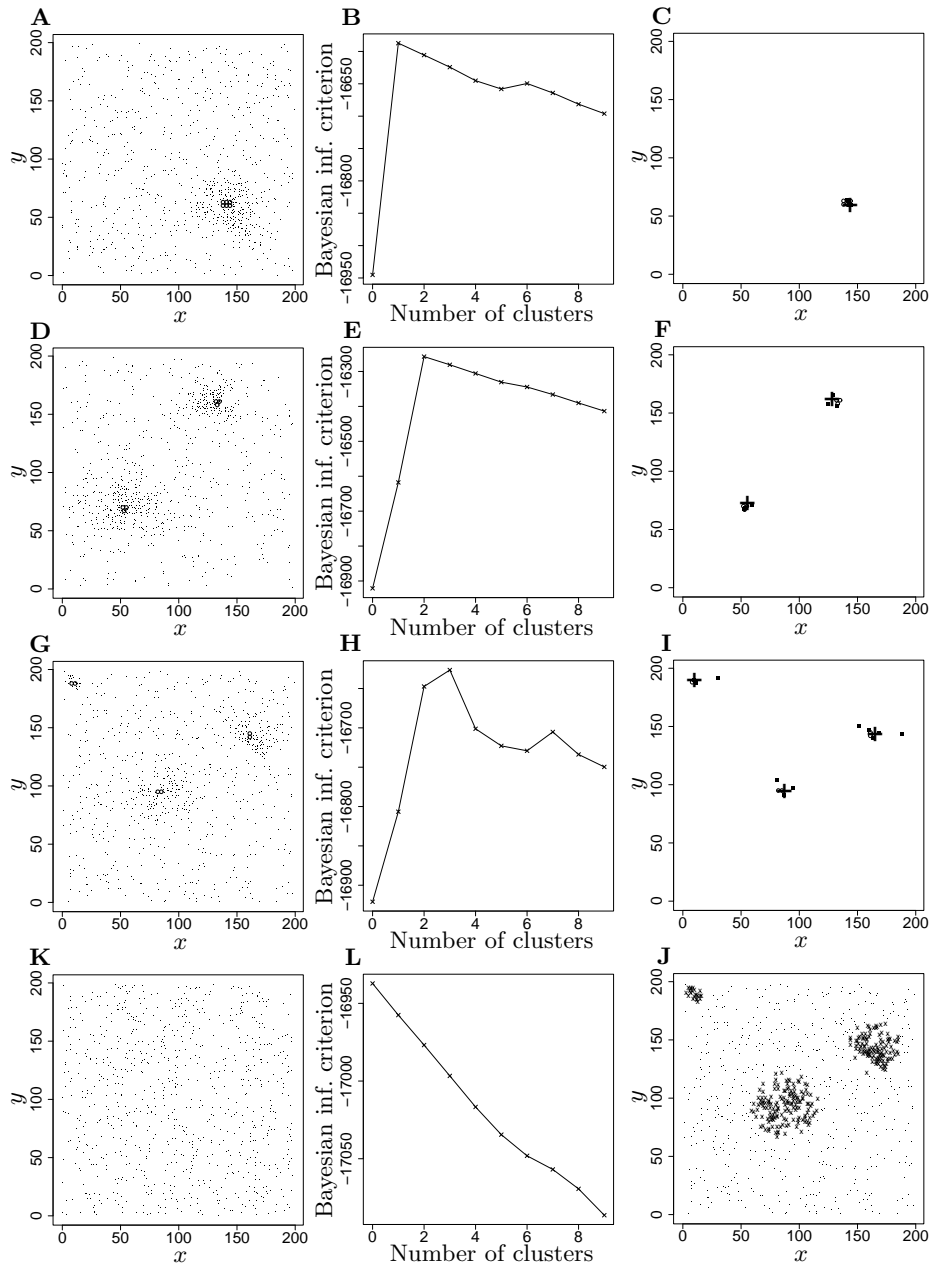
Simulations were conducted for four situations, one patch with biota (6 cells), two ( $2 \times 3$  cells), and three patches with biota ( $3 \times 2$  cells). After  $10^8$  updates (2500 updates per cell) of the cellular automata a snapshot of the distribution of chemicals in the simulation space was saved.

Scouting experiments were run on these snapshots to sample 800 locations in the simulation space. Figure 3 shows the probed locations in the left column. Their density indicates areas of high interest to the scouting algorithm. We used hierarchical clustering with subsequent expectation maximization (performed with *mclust* in *R* [22, 23]) to automatically identify clusters in the sampling positions. The number of clusters identified during expectation maximization can be seen in the middle column as peaks of the Bayesian information criterion. We used the samples allocated to a cluster (shown in Fig. 3 only for the case of three biota patches: panel J) and calculated their mean position as prediction for the location of biota in the simulation space. These predicted positions are marked with + in panels C, F and I. If the scouting is repeated with an alternate seed value for its random generator, the evolution of the sampling positions will take a different course; predictions from four additional runs are marked with • for comparison. The localization is fairly good, despite the complex chemical background (panel D, for example, shows the scouting of Fig. 2B).

Detection of the biota is possible with considerably less than the 800 samples used in Fig. 3, albeit with false positives and less accuracy. In Fig. 4 bars represent clusters, the solid bars correspond to clusters at biota patches, striped bars are clusters close to the correct location and white bars are false positives. The top graph (I) shows an experiment with a single biota patch that is detected with less than 100 samples and located with 300 samples. The middle graph (II) shows scouting with two biota patches, more samples are needed to find both locations. With three biota patches (graph III) one is detected early, but several false positives appear and one of the three patches is only approximately located (striped bar) with 800 samples.

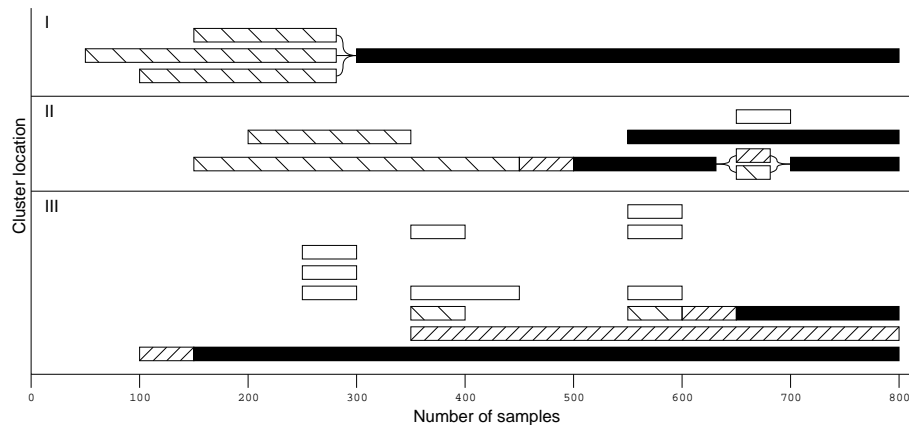
## 4 Concluding Remarks

We have applied an autonomous experimentation technique for the search of unusual chemical signatures in a complex environment. Scouting was able to consistently detect biota signatures without any domain knowledge regarding the chemistry of the environment or the biota. This performance of course has to be seen in the light of the relative small search space and the highly abstracted environment applied here.



**Fig. 3.** Detection of an unusual chemical composition. Actual location of biota ( $\circ$ ) and the 800 locations ( $\cdot$ ) probed by scouting, are shown for 1, 2, 3, and 0 biota patches in (A, D, G, K). By clustering the biota can be localized, marked + (and  $\cdot$  for additional scouting runs) in C, F, I. In J the allocation of sample positions to clusters is shown for the case of three biota patches. A run without biota (K, L) does not lead to clusters. For clarity symbols for biota and cluster centers have been enlarged but refer to only one point in the plane. See text





**Fig. 4.** Development of clusters for one, two and three biota patches (I, II, III). Clustering was performed every 50 samples, solid bars represent correctly localized biota, white bars false positives; see text for details

The path to algorithms that confidently can be entrusted with the search of life in remote locations of our solar system is still long. We believe that the broad variety of artificial life is an important resource to draw on for test scenarios during the development of such algorithms.

## Acknowledgments

The authors thank E. G. Schukat-Talamazzini for helpful discussion. This article is based upon work supported by BMBF (Federal Ministry of Education and Research, Germany) under grant No. 0312704A to PD, and NASA under Grant No. NCC2-1189 to KPZ.

## References

1. Toulmin, S., Goodfield, J.: *The Architecture of Matter*. Harper & Row, New York (1962)
2. Anonymous: *Astrobiology roadmap*. U.S. National Aeronautics and Space Administration (NASA) (November 2002) Available at: <http://astrobiology.arc.nasa.gov/roadmap/roadmap.pdf>
3. Langton, C.G.: *Artificial life*. In Langton, C.G., ed.: *Artificial Life*. SFI Studies in the Science of Complexity, Addison-Wesley, Redwood City, CA (1989) 1–45
4. Dittrich, P., Ziegler, J., Banzhaf, W.: *Artificial chemistries—a review*. *Artificial Life* **7** (2001) 225–275
5. Langley, P., Simon, H.A., Bradshaw, G.L., Żytkow, J.M.: *Scientific discovery: Computational exploration of the creative processes*. MIT Press, Cambridge, MA (1987)

6. de Jong, H., Rip, A.: The computer revolution in science: steps towards the realization of computer-supported discovery environments. *Artificial Intelligence* **91** (1997) 225–226
7. Rajamoney, S.A.: A computational approach to theory revision. In Shrager, J., Langley, P., eds.: *Computational Models of Scientific Discovery and Theory Formation*. Morgan Kaufmann, San Mateo, CA (1990) 225–253
8. Karp, P.D.: Hypothesis formation as design. In Shrager, J., Langley, P., eds.: *Computational Models of Scientific Discovery and Theory Formation*. Morgan Kaufmann, San Mateo, CA (1990) 275–317
9. Kulkarni, D., Simon, H.A.: Experimentation in machine discovery. In Shrager, J., Langley, P., eds.: *Computational Models of Scientific Discovery and Theory Formation*. Morgan Kaufmann, San Mateo, CA (1990) 255–273
10. Gooding, D.: *Experiment and the Making of Meaning*. Kluwer Academic, Dordrecht (1990)
11. Stolorz, P., Cheeseman, P.: Onboard science data analysis: Applying data mining to science-directed autonomy. *IEEE Intelligent Systems & Their Applications* **13** (1998) 62–68
12. Gilmore, M.S., Castano, R., Mann, T., Anderson, R.C., Mjolsness, E.D., Manduchi, R., Saunders, R.: Strategies for autonomous rovers at mars. *J. Geo. Phys. Res. Planets* **105** (2000) 29223–29237
13. Pfaffmann, J.O., Zauner, K.P.: Scouting context-sensitive components. In Keymeulen, D., Stoica, A., Lohn, J., Zebulum, R.S., eds.: *The Third NASA/DoD Workshop on Evolvable Hardware*, IEEE Comp. Soc., Los Alamitos (2001) 14–20
14. Cherry, C.: Chapter 5. In: *On Human Communication: A Review, a Survey, and a Criticism*. 2nd edn. MIT Press, Cambridge, MA (1966)
15. Beyer, H.G.: *The theory of Evolution Strategies*. Springer, Berlin (2001)
16. Matsumaru, N., Colombano, S., Zauner, K.P.: Scouting enzyme behavior. In Fogel, D.B. et al., eds: *2002 World Congress on Computational Intelligence*, IEEE, Piscataway, NJ (2002) CEC 19–24
17. Adamatzky, A.: *Identification of Cellular Automata*. Taylor and Francis, London (1994)
18. Speroni di Fenizio, P., Dittrich, P.: Artificial chemistry’s global dynamics. movement in the lattice of organisation. *The Journal of Three Dimensional Images* **16** (2002) 160–163
19. Fontana, W., Buss, L.W.: ’The arrival of the fittest’: Toward a theory of biological organization. *Bull. Math. Biol.* **56** (1994) 1–64
20. Speroni di Fenizio, P., Dittrich, P., Ziegler, J., Banzhaf, W.: Towards a theory of organizations. In: *German Workshop on Artificial Life*, in print, Bayreuth, 5.-7. April, (2000)
21. Yung, Y.L., DeMore, W.B.: *Photochemistry of Planetary Atmospheres*. Oxford University Press, New York (1999)
22. Ihaka, R., Gentleman, R.: R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **5** (1996) 299–314
23. Fraley, C., Raftery, A.E.: MCLUST: Software for model-based clustering, discriminant analysis and density estimation. Technical Report 415, Department of Statistics, University of Washington, Seattle, WA (2002) Available at: <http://www.stat.washington.edu/www/research/reports/>