

Using Artificial Epigenetic Regulatory Networks To Control Complex Tasks Within Chaotic Systems

Alexander P. Turner^{1,4}, Michael A. Lones^{1,4}, Luis A. Fuente^{1,4},
Susan Stepney^{2,4}, Leo S. Caves^{3,4} and Andy M. Tyrrell^{1,4}

¹ Department of Electronics, {apt503,ma1503,laf509,amt}@ohm.york.ac.uk

² Department of Computer Science, susan.stepney@cs.york.ac.uk

³ Department of Biology, lsd10@york.ac.uk

⁴ York Centre for Complex Systems Analysis (YCCSA)
University of York, Heslington, York YO10 5DD, UK

Abstract. Artificial gene regulatory networks are computational models which draw inspiration from real world networks of biological gene regulation. Since their inception they have been used to infer knowledge about gene regulation and as methods of computation. These computational models have been shown to possess properties typically found in the biological world such as robustness and self organisation. Recently, it has become apparent that epigenetic mechanisms play an important role in gene regulation. This paper introduces a new model, the Artificial Epigenetic Regulatory Network (AERN) which builds upon existing models by adding an epigenetic control layer. The results demonstrate that the AERs are more adept at controlling multiple opposing trajectories within Chirikov's standard map, suggesting that AERs are an interesting area for further investigation.

1 Introduction

Gene regulatory networks are complex structures which underpin an organism's ability to control its internal environment [23]. From a biological perspective, the study of gene regulation is of interest because it determines cellular differentiation which is responsible for the development of the different tissues and organs that underpin the structure of higher organisms [17]. From a computational perspective, gene networks are interesting because they are robust control structures, capable of dealing with serious environmental perturbations, and yet maintaining structure and order. Because of this, there has been significant interest in modeling gene regulatory networks *in silico* in order to capture these features [18, 21]. Due to the complexity of biological gene regulation, computational analogues are vastly simplified. However, research has shown that relatively simple networks such as the random Boolean network can exhibit emergent properties such as self organisation and robustness and, in addition to this, can model real regulatory circuits [15, 14]. It is the balance between complexity and function which is often the limiting factor in the creation of truly analogous computational models.

It is apparent that computational analogues of gene regulation fail to include models of one of the most pervasive methods of gene regulation, epigenetics. In this sense, the regulatory nature of these computational analogues may be limited in terms of complexity and performance. In this paper, we attempt to define a representation of epigenetic information which to some extent captures the useful properties of epigenetics for the control of a complex dynamical system.

2 Gene Regulation and Epigenetics

A gene is a unit of hereditary information within a living organism, most commonly considered to be a section of DNA that specifies the primary structure of a protein. The genetic code is a biological blueprint that details which proteins can be produced, and ultimately, the phenotypic space which the organism can exist within. The lowest known threshold on the number of genes required to naturally facilitate life is that of *Mycoplasma genitalium*, which has approximately 470 predicted genes [9]. Even in nature's most minimalist example of a gene regulatory network, 470 genes have to be coordinated in such a way to maintain the optimum internal environment of the organism, highlighting that even the simplest of gene regulatory networks are inherently complex.

Epigenetic mechanisms allow a further layer of control over gene regulation. This is commonly done by physically restricting the accessibility of the genes. One of the principal epigenetic mechanisms is DNA methylation which refers to the addition of a methyl group to either the cytosine or adenine nucleobase in DNA (Figure 1). It acts as an epigenetic marker which can regulate many physiological processes [3, 11]. With increasingly complex organisms, higher order structures such as chromatin have been shown to influence gene expression. Chromatin has two functions; Firstly, in the case of human cells, there are approximately 3400Mb of DNA of approximately 2.3m in length. Chromatin provides a structure for the condensation of DNA, so that it can be fully contained in a nucleus of approximately $6\mu\text{m}$ [1, 2, 6]. Secondly, chromatin modifications have the ability to control access to the DNA, which in turn acts as an additional level of genetic control. Generally, DNA methylation provides a more long term, stable effect on gene expression when compared to relatively short term reversible chromatin modifications [7]. However, research suggests that in some instances chromatin modifications and DNA methylation are intrinsically linked [12].

One of the more interesting aspects of epigenetics is that in certain instances, epigenetic traits can be inherited by successive generations of cells, and sometimes organisms [4]. In addition to this, epigenetic modifications can give the genetic code a relative memory [3], which can then be used in such processes as cellular differentiation [20]. However, the specific mechanisms and processes which control epigenetics, and in turn gene expression, are not yet wholly understood.

Current research demonstrates that epigenetic mechanisms allow for a level of genetic memory, and a method of genetic control above the genetic code itself.

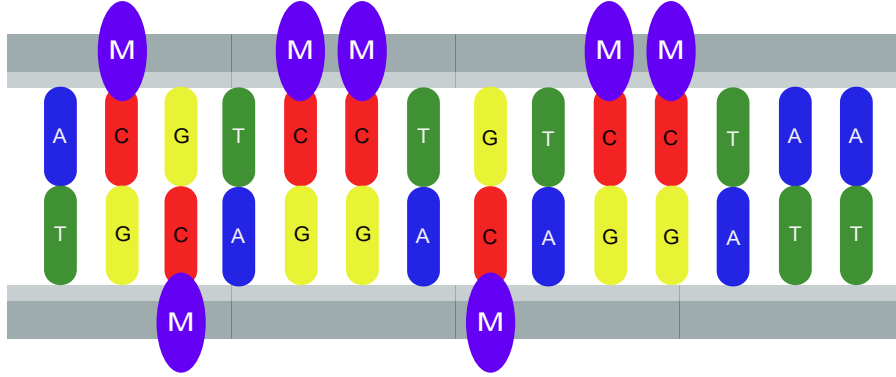


Fig. 1: An illustration of a highly methylated DNA region. The attachment of a methyl group (M) to the cytosine (C) nucleobase (DNA methylation) has been shown to be able to regulate physiological processes [3, 11]

This creates the ability for organisms to express high levels of adaptability via utilisation of the extra dynamical processing which epigenetics facilitates. In the following sections the idea of incorporating a level of artificial epigenetic information in pre-existing artificial gene regulatory networks is introduced. Simulations are conducted to ascertain if the addition of epigenetic information can yield any performance increases when controlling two trajectories within Chirikov’s standard map.

3 Artificial Genetic Regulatory Networks

Gene regulation in biology is a set of mechanisms which maintains control of an organism’s internal environment (homeostasis). The aim of artificial gene regulation is to create a computational model of genetic behaviour which exhibits the useful and interesting properties of gene regulation in nature, namely self organisation and robustness. The first such example was the random Boolean network (RBN) [15]. RBNs represent genes as Boolean expressions. These artificial genes are referred to as node’s. The network has a connectivity value (k) which specifies how many node’s influence a given nodes expression level. The state of a node is defined by randomly initiated state transition rules. Upon execution, the network is iterated over a number of time steps, during which each node modifies its value depending upon its connectivity and its state transition rules. These networks demonstrate that with a k value of 2 or 3, distinct order and repetitive patterns can be generated. Moreover, for certain parameter ranges, the RBNs expressed high levels of robustness, maintaining relative order when exposed to external perturbations [10].

Subsequent models of gene regulation draw inspiration from the RBN. However there has been a shift towards continuous models [16, 18], as they are com-

putationally more flexible. In addition, it has been shown that these models can be applied to the control of complex systems [18].

3.1 Artificial Epigenetic Regulatory Networks

This paper extends the model of artificial gene regulation (AGN) described in [18] by incorporating epigenetic information. The aim is to ascertain if an additional level of regulatory control will result in any performance benefits over the previous model. The Artificial Epigenetic Regulatory Network (AERN) uses an analogue of DNA methylation in combination with chromatin modifications as its epigenetic elements. This gives the network the ability to change its epigenetic information both during evolution, and within a single generation via the changing of epigenetic frames (\mathbf{E}_G in definition below). Table 1 gives an example of an evolved AERN.

The AERN can be formally described as: $\langle G, L_G, I_G, O_G, E_G \rangle$ where :

- \mathbf{G} = An indexed set of genes $\{g_0, \dots, g_n : g_i = \langle \lambda_i, R_i, f_i \rangle\}$, where:
 - $\lambda_i : \mathbb{R}$ is the expression level of a gene
 - $R_i \subseteq \mathbf{G}$ is the set of regulatory inputs used by the genes
 - $f_i : R_i \rightarrow \lambda_i$ is a gene's regulatory function
- L_G is an indexed set of initial expression levels, where, $|L_G| = |\mathbf{G}|$
- $I_G \subset \mathbf{G}$ are the external inputs applied to the network
- $O_G \subset \mathbf{G}$ are the outputs of the network
- $\mathbf{E}_G \subset \mathbf{G}$ is a data structure specifying which genes are *active* at a given instance

The AERN is executed as follows :

- G1. $\lambda_0 \dots \lambda_n$ are initialised from L_G (if AERN not previously executed).
- G2. Expression levels of enzymes in I_G are set by the external inputs.
- G3. At each time step, each *active* gene g_i applies its regulatory function f_i to the current expression levels of its *active* regulating genes R_i in order to calculate its expression at the next time step, λ'_i
- G4. After a given number of iterations, execution is halted and the expression levels of enzymes in O_G are copied to the external outputs.

The expectation is that epigenetic control will enable the network to evolve in such a way that it will have certain genes that are more able to perform a given objective. The inclusion of epigenetic information would give the network the ability to allocate different genes to different tasks, effectively regulating gene expression according to the environment in which it is operating.

Each gene uses a parameterisable sigmoid function which has been shown in [18] to be the most effective in traversing the complex. The parameter settings are summarised in Table 2.

Variable	External Inputs (\mathbf{I}_G)		Genes					Outputs ($\mathbf{O}_G \subset G$)
Gene Expression Values (\mathbf{L}_G)	0.18	0.81	0.54	0.38	0.95	0.14	0.05	0.47
Weights	0.47	-0.27	0.24	0.99	-0.87	-0.02	-0.47	0.97
Sigmoid Offset	-0.18	0.24	0.14	-0.50	-0.21	0.57	0.31	0.38
Sigmoid Slope	1	10	5	19	2	14	3	7
Connections	5	2	1	5	7	3	2	3
	7	4		5	2	7	1	1
	5	2		5		3	2	3
		4		4		1		7
Epigenetic Frame A ($\mathbf{E}_G \subset G$)	1	0	1	1	0	0	0	1
Epigenetic Frame B ($\mathbf{E}_G \subset G$)	0	1	1	0	1	1	1	1
Network Iterations	15							

Table 1: Example data attributes for an AERN of size 8. The only difference between the AERNS and the AGNs is the introduction of 2 epigenetic frames, which specify which genes will be active for each objective

4 Dynamical Systems

A dynamical system is a system whose current state is a product of its previous state and an evolution rule. For any given point within a dynamical system, its trajectory through the space is governed by the iteration of this rule over a given time frame. One of the most interesting groups of dynamical systems are those that exhibit chaotic dynamics. Chaos can be recognised as irregular and unpredictable behavior within a system, due to extreme sensitivity to initial conditions [22]. However, unlike random behavior, this is the result of applying deterministic evolution rules.

4.1 Standard Map

Chirikov’s standard map [8] is a dynamical system which under certain conditions expresses chaotic dynamics. Its behaviour results from two difference equations :

$$x_{n+1} = (x_n + y_{n+1}) \bmod 1 \quad y_{n+1} = y_n - \frac{k}{2\pi} \sin(2\pi x_n) \quad (1)$$

The modification of parameter k has a direct effect on the dynamics of the system (Figure 2). In figures 2a and 2c it can be seen that the dynamics of the map are for the most part homogeneous (with slight differences in the corner of the map). However, when k reaches a critical value of 0.972, there is a distinct increase in the number of elliptic islands, and the distance change at a given co-ordinate is no longer consistent over the majority of the map [22] (Illustrated in figure 2b, where $k = 1$).

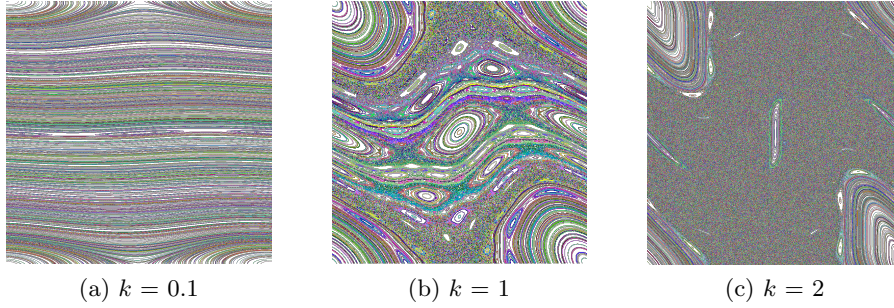


Fig. 2: Sample of Chirikov’s Standard Map during ordered (a and c) and chaotic (b) dynamical phases

4.2 State Space Targeting

The presence of chaos makes it difficult to predict subsequent points in a trajectory. However, research has shown that targeting within the standard map during periods of chaos is possible via the use of perturbations, allowing navigation to different regions of the state space[18, 5].

5 Experimentation

In order to test the relative performance increases of the AERNs, they have been evaluated against the model from [18] when controlling two opposing trajectories within the standard map.

5.1 Evolution Of The Networks

A genetic algorithm (GA) is used to evolve the network. GAs are population based search algorithms [19], which often find near optimal solutions within a tractable time frame, which makes it a favorable approach for this experiment. The GA uses a crossover rate of 0.5, a mutation rate of 0.001 and tournament selection of size 4 to evolve networks containing 20 genes. The simulation is run over 50 generations, and the fittest individual at the 50th generation is the score for that run. 50 runs were carried out for each network type.

5.2 State Space Targeting Using Artificial Regulatory Networks

Previous work demonstrates that artificial gene regulatory networks (AGN) can be used to target specific regions of Chirikov’s map [18]. This paper builds upon this by using the networks to control two opposing trajectories within the standard map. There were two objectives, which specify that the networks have to guide a trajectory from the bottom to the top of the complex map, and then from the top to the bottom in the lowest number of iterations.

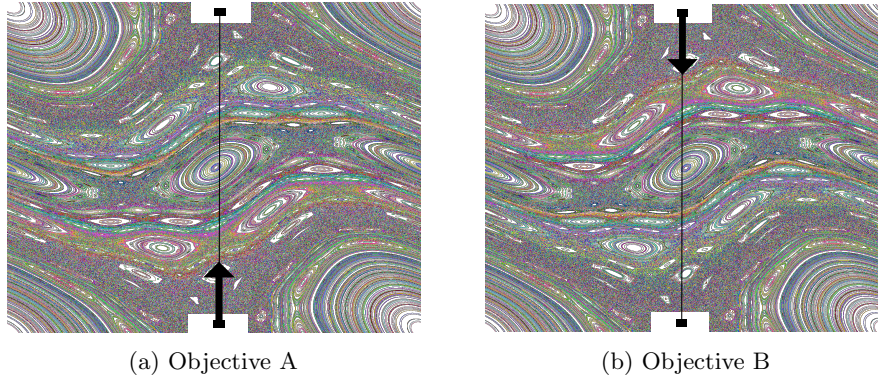


Fig. 3: Chirikov's standard map showing two start and finish areas for each objective

Each network received 3 inputs and produced 1 output. The three inputs were the x co-ordinate, the y co-ordinate and the distance from the center of the target. The output of the network was the new k value (see Eqn 1), which is used in the next iteration of the equation. Since the output of the network is a real value (Table 2), it must be scaled to the interval $[1,1.1]$ to ensure that the map remains in its chaotic phase. Upon initialisation, the starting point for the trajectory is a randomly initiated point within the starting region, and the networks are randomly initiated with values shown from Table 2. The epigenetic frame changes when either the first objective is completed, or the the maximum number of steps for that objective has been reached.

For the objective specifying the trajectory is to travel from the bottom to the top of the map, the initial start point is $[0,0.05]$ for the y co-ordinate, and $[0.45,0.55]$ for the x co-ordinate. For the second objective, the randomly initialised start point is between $[0.95,1]$ for the y co-ordinate, and $[0.45,0.55]$ for the x co-ordinate. The score for that run is the average number of steps to traverse the map in both directions, up to a maximum number of 1000 steps. During execution, the average of the 40 runs (20 for each objective) are collated together, and if there is an instance where the network completed both objectives, the fitness is rewarded by 250 steps. If only one objective is completed, the score is decremented by 100 steps up to a maximum of 1000 steps (the final results will not take into consideration rewards and punishments, and only solutions that complete both objectives are able to

Variable	Type	Range
Gene Expression	Real	0;1
Weights	Real	-1;1
Sigmoid Offset	Real	-1;1
Sigmoid Slope	Int	0;20
Epigenetic Objective	Int	0;1
Network Iterations	Int	1;20

Table 2: Parameters of the variables within the AERN

achieve a score less than 1000). This is to place emphasis on the networks completing both objectives, whilst allowing strong solutions that only complete one objective to remain in the population.

6 Results

The results show that both the AGNs and the AERNs were able to produce solutions that could control both trajectories. However, the results indicate that the tasks were difficult (Figure 4), as only 35% of the AGN's and 52% of the AERNs were able to complete both objectives. The results demonstrate that the AERNs perform significantly better than the AGNs. The trend is more evident when the unsuccessful instances are removed, as can be seen in Figure 5.

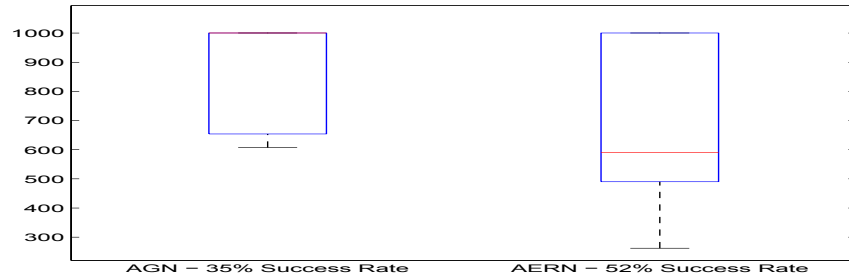


Fig. 4: Summary of the fitness distributions (low numbers are better). The results for the AGN show less than a 50% success rate, and a best value of approximately 600 steps. The AERNs show a significant increase in success rate, and better solutions overall.

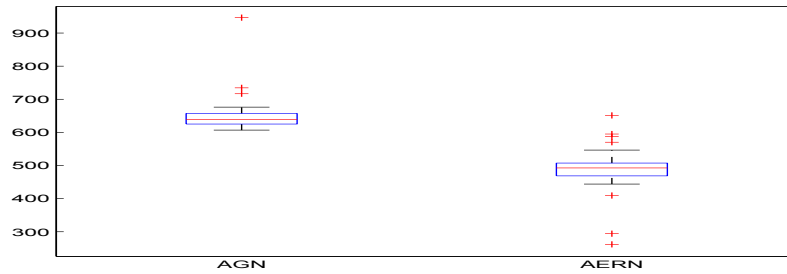


Fig. 5: Data from Figure 4 with the unsuccessful results omitted (low numbers are better). This shows a much clearer trend in the data, and the addition of the artificial epigenetic information can be seen to increase performance by approximately 150 steps on average.

One of the more interesting aspects of the results is that the data in Figure 5 shows that the vast majority of the successful AERNs outperform almost all of the successful results from the AGNs, and yet the trends of the data are quite similar. On average the AERNs produce an improvement of approximately 150 steps in terms of mean path length.

Aside from increased performance, the AERNs demonstrate further benefits. First, with the ability to inactivate genes comes the ability to increase the efficiency of the networks. Hence, with each inactive gene for each objective, there is less computational effort required to complete a single iteration of the network. Over the entirety of a simulation, this could lead to significant performance increases. A further advantage is that the epigenetic layer of the networks provides a source of qualitative analysis. By dissection of the network structures, it could be seen that certain input variables are not needed in the navigation of the complex map. In some of the best solutions, the input variable for the x co-ordinate was not required, which provides additional information about the problem space that otherwise would not have been readily available.

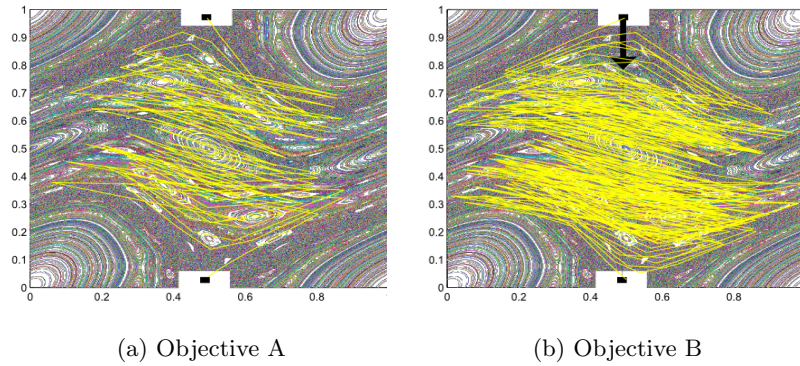


Fig. 6: Example behaviour of the solutions produced by the AERNs. Objective A was completed in 113 steps, and objective B was completed in 329 steps

7 Conclusions

This paper has illustrated the potential for incorporating epigenetic information in computational models of gene regulation, and the initial results are highly promising. The results demonstrate that evolved AERNs are able to assign certain genes to certain tasks, improving functionality and efficiency. This ties in well with the biology of epigenetics, which allow for a higher level of genetic control without compromising efficiency [13].

There is a significant amount of further research required to assess the full functionality of the AERNs. In future work, the AERNs will be applied to a range of tasks to best evaluate their strengths. Additionally, the topologies of

the networks will be looked at in more detail to ascertain the role and function of varying epigenetic mechanisms. Furthermore, research will be conducted into introducing a metabolic network to attempt to best utilise the efficiency potential of the AERs.

Acknowledgments

This work is supported by the EPSRC funded project (ref:EP/F060041/1), Artificial Biochemical Networks: Computational Models and Architectures.

Thank you to the reviewers for their comments and feedback.

References

1. Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., Watson, J.: Molecular Biology of the Cell, 3rd edition. Oxford Univ Press (1994)
2. Allis, C., Jenuwein, T., Reinberg, D.: Epigenetics. Cold Spring Harbor Laboratory Press (2007)
3. Bird, A.: DNA methylation patterns and epigenetic memory. *Genes & development* 16(1), 6 (2002)
4. Bird, A.: Perceptions of epigenetics. *Nature* 447(7143), 396 (2007)
5. Bollt, E., Meiss, J.: Controlling chaotic transport through recurrence. *Physica D: Nonlinear Phenomena* 81(3), 280–294 (1995)
6. Bushman, F.: Lateral DNA transfer: mechanisms and consequences. Cold Spring Harbor Laboratory Press (2002)
7. Cedar, H., Bergman, Y.: Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet* 10(5), 295–304 (2009)
8. Chirikov, B., Sanders, A.: Research concerning the theory of non-linear resonance and stochasticity. Nuclear Physics Institute of the Siberian Section of the USSR Academy of Sciences (1971)
9. Fraser, C., Gocayne, J., White, O., Adams, M., Clayton, R., Fleischmann, R., Bult, C., Kerlavage, A., Sutton, G., Kelley, J., et al.: The minimal gene complement of mycoplasma genitalium. *Science* 270(5235), 397 (1995)
10. Harvey, I., Bossomaier, T.: Time out of joint: Attractors in asynchronous random boolean networks. In: Proceedings of the Fourth European Conference on Artificial Life. pp. 67–75. MIT Press, Cambridge (1997)
11. Hattman, S.: Dna-[adenine] methylation in lower eukaryotes. *Biochemistry (Moscow)* 70(5), 550–558 (2005)
12. Jackson, J., Lindroth, A., Cao, X., Jacobsen, S.: Control of CpNpG DNA methylation by the kryptonite histone h3 methyltransferase. *Nature* 416(6880), 556–560 (2002)
13. Jeanteur, P.: Epigenetics and Chromatin. Progress in Molecular and Subcellular Biology, Springer (2008)
14. Kauffman, S., Peterson, C., Samuelsson, B., Troein, C.: Random Boolean network models and the yeast transcriptional network. *Proceedings of the National Academy of Sciences of the United States of America* 100(25), 14796 – 14799 (2003)
15. Kauffman, S.: Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of theoretical biology* 22(3), 437–467 (1969)

16. Kumar, S.: The evolution of genetic regulatory networks for single and multicellular development. In: GECCO. Citeseer (2004)
17. Latchman, D.: Gene regulation: a eukaryotic perspective. Advanced text, Taylor & Francis (2005)
18. Lones, M.A., Tyrrell, A.M., Stepney, S., Caves, L.S.: Controlling complex dynamics with artificial biochemical networks. In: Esparcia-Alczar, A.I., et al. (eds.) Proc. 2010 European Conference on Genetic Programming (EuroGP 2010). Lecture Notes in Computer Science, vol. 6021, pp. 159–170. Springer Berlin / Heidelberg (2010)
19. Mitchell, M.: An introduction to genetic algorithms. Complex adaptive systems, MIT Press (1998)
20. Mohn, F., Schübeler, D.: Genetics and epigenetics: stability and plasticity during cellular differentiation. Trends in Genetics 25(3), 129–136 (2009)
21. Quick, T., Nehaniv, C.L., Dautenhahn, K., Roberts, G.: Evolving embodied genetic regulatory network-driven control systems. In: Proc. European Conference on Artificial Life (ECAL'03). vol. 2801, pp. 266–277. Springer (2003)
22. Tél, T., Gruiz, M.: Chaotic dynamics: an introduction based on classical mechanics. Cambridge University Press (2006)
23. Turner, B.: Chromatin and gene regulation: mechanisms in epigenetics. Blackwell Science (2001)