PRION CRYSTALIZATION MODEL AND ITS APPLICATION TO RECOGNITION PATTERN

Paula Cordero, Rafael Lahoz-Beltra and Juan Castellanos

***Abstract**: This paper introduces APA ("Artificial Prion Assembly"): a pattern recognition system based on artificial prion crystalization. Specifically, the system exhibits the capability to classify patterns according to the resulting prion self- assembly simulated with cellular automata. Our approach is inspired in the biological process of proteins aggregation, known as prions, which are assembled as amyloid fibers related with neurodegenerative disorders.*

***Keywords**: pattern recognition, prion protein aggregation, crystal growth, simulation models.*

***ACM Classification Keywords**: F.1. Computation by abstract devices, F1.1. Models of Computation.*

Introduction

Molecular self-assembly is one of the most relevant biological mechanisms related with the self-organization and biological functions exhibited within cells [Lahoz-Beltra, 1997]. Cellular automata are able to capture the main features of biomolecules that form part of cellular structures and organelles. A breakthrough in the computer modeling and simulation of proteins took place when cellular automata modeling was applied to simulate the interaction of proteins during self-assembly [Lahoz-Beltra, 1999]. In a different realm spin glasses and the Hopfield content addressable memory exhibit emergent collective computational abilities. Such capabilities are related with phase transitions between the crystalline state (low temperature and energy) and liquid state (high temperature and entropy). In consequence, crystalization process may be a manifestation of the ability of Nature to process information, e.g storage and pattern recognition.

Cellular automata belong to a family of discrete, connectionist techniques being used to investigate fundamental principles of dynamics, evolution, and self-organization. In general, they constitute exactly computable models for complex phenomena and large-scale correlations that result from very simple short-range interactions. In this paper, a cellular automaton is designed to model the proteins behavior during self-assembly. The cellular approach makes it possible to achieve a theory-based view of morphogenesis detail to link the results directly to a classification model. In this paper we explore how prion protein aggregation or self-assembly could be used to design an artificial pattern recognition system. The proposed computational system has been dubbed as APA ("Artificial Prion Assembly") memory.

Prion protein aggregation and its simulation models

A prion is an infectious agent composed of protein in a misfolded form. Prion diseases are fatal neurodegenerative disorders associated with the polymerization of the cellular form of prion protein (PrPC) into an amyloidogenic β-sheet infectious form (PrPSc) [Fontaine & Brown, 2009].

Prions propagate by transmitting the misfolded protein state. When a prion enters a healthy organism, it induces existing properly folded proteins to convert into the disease-associated prion form; the prion acts as a template to guide the misfolding of more proteins into prion form. These newly formed prions can then go on to convert more proteins themselves; this triggers a chain reaction that produces large amounts of the prion form.

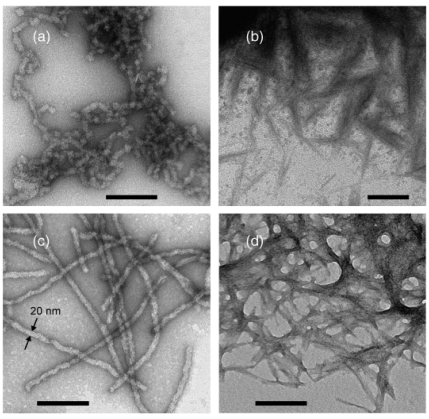

Figure 1. TEM images of prion-seeded Ure2p fibrils (a and b) and unseeded Ure2p fibrils (c and d) before PK treatment (a and c) and after PK treatment (b and d). TEM grids are negatively stained with uranyl acetate. Scale bars represent 200 nm. [Kryndushkin et al, 2011]

Under normal conditions, the high-energy barrier separates PrPC from PrPSc isoform. However, pathogenic mutations, modifications as well as some cofactors, such as glycosaminoglycans, nucleic acids, and lipids, could modulate the conformational conversion process. Abundant nonfibrillar oligomeric intermediates are a common feature of amyloid formation [Bemporad & Chiti, 2013], and these oligomers, rather than the final fibers, have been suggested to be the toxic species in some amyloid diseases.

Evidence suggests that an aggregated form of PrPSc is in fact the key component in the disease [Stahl et al., 1987; Brown et al., 1997] but the precise character of the infectious aggregates is unclear. Therefore, the study of in vitro aggregation of recombinant PrP is instrumental in providing insight into the mechanisms behind conversion from PrPC to PrPSc and aggregate accumulation, as well as to determine the conformation and species that is actually responsible for prion pathogenesis.

Many efforts are dedicated to design aggregation models developed. Aggregation of PrP has been modelled using three kinetic theories: template assisted-aggregation, nucleation-elongation polymerization, and branched-chain polymerization [Fontaine & Brown, 2009]. Each of these theories has been reviewed elsewhere, but in brief summary all employ the idea that a smaller unit of PrP is responsible for further catalysing protein aggregation.

Mainly, models of aggregation processes in prion disease include one-dimensional, fibrillar aggregation-and-fission models, since aggregates grown in vitro are typically seen to be fibrillar. There are several approaches used to model this kind of processes. These models range from stochastically and deterministic approaches through known kinetic models based on differential equations [Greer et al, 2006] to cellular automata based models [Kulkarni et al, 2003].

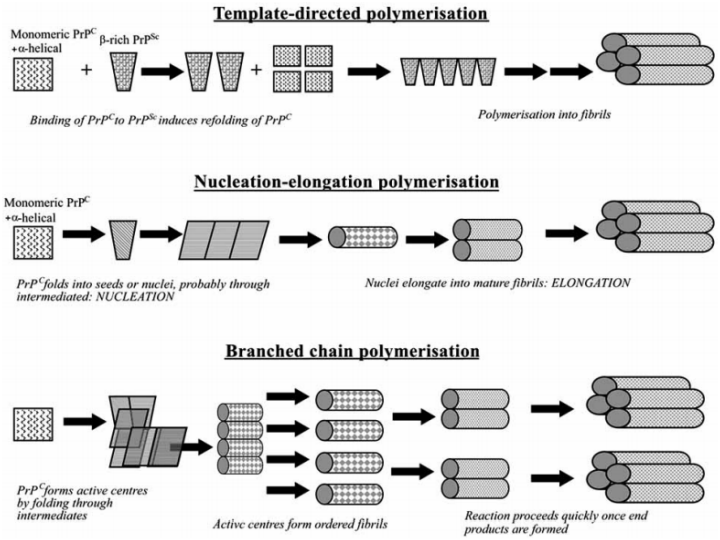

Figure 2. Prion protein aggregation kinetic models. Schematic illustrating the differences between the kinetic theories of PrP aggregation. [Fontaine & Brown, 2009]

Diffusion-Limited Aggregation and Crystal Growth models

Diffusion-limited aggregation (DLA) [Witten & Sander, 1981] is an idealization of the process by which matter irreversibly combines to form dust, soot, dendrites, and other random objects in the case where the rate-limiting

step is diffusion of matter to the aggregate. Diffusion is the movement of particles due to temperature fluctuations and seen in Brownian motion. By the other hand, an aggregate is a collection of particles that are connected together this growing process is called diffusion-limited when the aggregate increases in size by one particle at a time. A particle is appear from a random position far away and is allowed to diffuse. If it touches the seed, it is immobilized instantly and becomes part of the aggregate. This happens since the density of particles is low and thus the particles do not come into contact with each other before reaching the aggregate.

The growth processes [Levi & Kotrla, 1997] are described by nonlinear partial-differential equations and both the analytical and the numerical treatments of these equations are extremely difficult even on current computers. As a result, many of the questions concerning structure formation and transitions between different growth morphologies have not so far been satisfactorily answered. Much effort has especially been devoted to establishing the relationship between cluster morphology and the growth mechanism.

The crystal growth is a phase transition process with sharp border between it and initial feeding phase like a liquid, gas or plasma. The structure element, molecule, of the crystal could be determined as a minimum part of it when a reaction of incorporating itself in the crystal will effect with changing energy of the whole system that will be equivalent to the condensation energy of the corresponding mass of crystal. The molecule is the minimum part of the crystal that behaves as a whole crystal. The principal difference of behavior of molecules in liquids is based on the principle of the long order in the crystals structure. Each molecule has exact position relatively to the other in crystal.

In the every moment of time one molecule on the surface of a crystal have two options: to get out to initial matter or stay incorporated as a part of crystal. The same choice is true for the molecule outside of crystal in direct closeness to its surface. It can be incorporated into body of crystal or stay outside. Most of simulation models implement these phenomena by calculating these probabilities for each position on the surface of the crystal and comparing with the random number to decide what one of the possible events will happen.

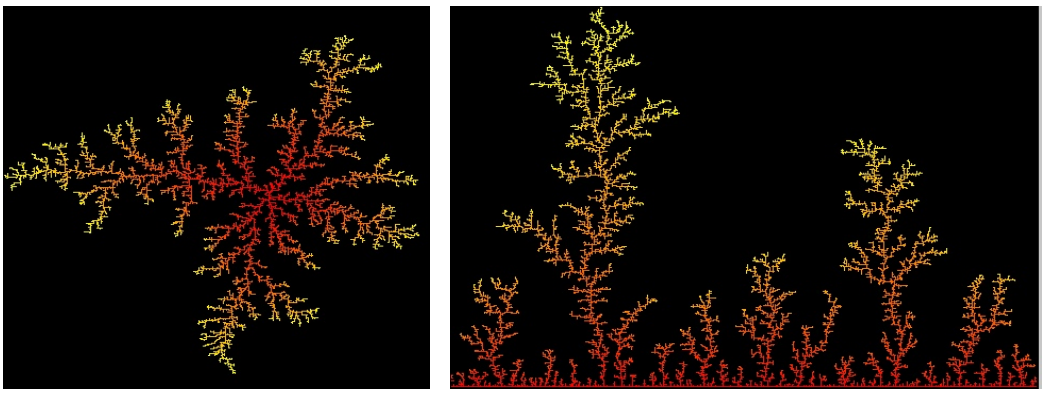

Figure 3. Sample DLA images from java applet: http://www.joakimlinde.se/java/DLA/

Spin glass models and addressable memories

At this point it should be noted that many studies find that some types of addressable memory as Hopfield's neural network [Hopfield, 1982] was inspired by analogies with the physics of magnetism the same as crystallization kinetics. Main characteristic of the Hopfield's model is the recurrence of the network with total connectivity and a symmetric weight matrix; binary valued outputs, which provide a simple prescription for the weights, with no training needed; output settles down to a steady state.

As it was mentioned, there are some relations between the behavior of addressable memories and the physics of magnetic spin systems [Edwards & Anderson, 1975]. In particular, phase transitions represent a competition between minimizing the energy (usually producing an ordered state) and maximizing the entropy (increasing the

disorder). This can be understood in terms of the behavior to the free energy, F=E-TS, which tends to a minimum for a system in thermal equilibrium. At low values of the temperature T, is more important to minimize the energy E. When T is high, then large values of the entropy S will make F smaller. The most interesting aspect is the abrupt, discontinuous nature of the transition between the ordered state and the disordered state occasionally attributed to emergent collective behavior. This behavior depends on the values of the positions and momentum of each of the molecules.

The key to being able to quantitatively describe the free energy and related thermodynamic quantities is to express the total energy of the system in terms of the states of the atoms. In this sense, it is described the Hopfield's model behavior where, in the presence of more than one pattern, the weights aren't optimum for the retrieval of any one pattern, but represent an average or compromise over the set of patterns. Instead of having a single minimum energy state, we will have a local minimum for each pattern. If the initial state of the network isn't too far from the stored state, the system will slide into the nearest local minimum, which will be the desired output state. When it try to minimize the error in a feed-forward net by using the back-propagation algorithm, getting stuck in a local minimum can keep us from finding the global minimum that produces the best set of weights. In this case, local minima are desirable.

Hopfield found in his computer experiments that the ability of the network to retrieve patterns fell dramatically when the number of stored patterns approached 15% of the number of neurons. As the number of patterns becomes large, the weight begins to look like a random variable. The term in the sum that favors a particular pattern is greatly outweighed by all the others, and the associative memory begins to look like a true spin glass.

Analogously to the above approach, the work presented in this paper is based on parallelism between the phenomena of crystallization (in the case of prions) and addressable memories behavior that implement pattern recognition capabilities as it is described below.

Pattern recognition system based on Prion Crystalization

This work presents a pattern recognition system based on a model of prion crystallization. This proposed model represents an unsupervised learning system. The theoretical model is composed of two phases: Phase-I) in which is modelled the winged-helix dimerization process through a probabilistic cellular atuomata approach in order to obtain the nucleation seeds of the prion-crystallization process. Phase-II) in which crystallization process will model prion dynamics from the nucleation centers obtained. The system will be able to classify input patterns by decoding implicit information from the morphology of the crystal-prions, which will be obtained as a result of the application of the two phases of the model.

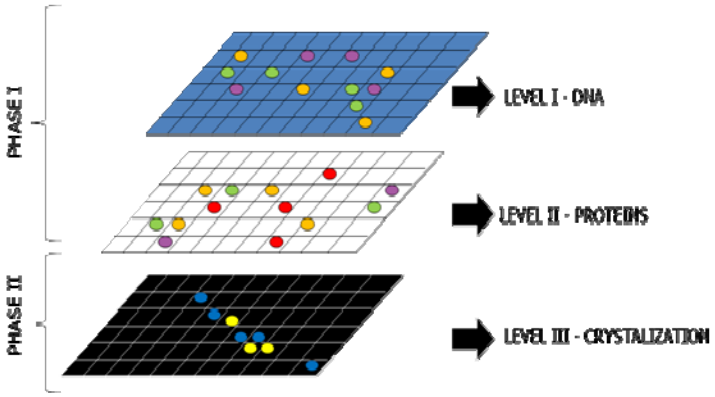

Figure 4. Schema of the Probabilistic Cellular Automata proposed.

As mentioned, the proposed system is inspired by the biochemical process by which certain strands of DNA in vitro affect the process of prion formation. Specifically, this process is described in the work "A DNA promoted amyloid proteinopathy in Escherichia coli" [Fernandez et al, 2011]. This research team goes into detail about this phenomena, core of proposed model in its Phase-I, and states that similarly to the mammalian proteins PrP and a-

synuclein, the winged-helix dimerization (WH1) domain of the bacterial plasmid-encoded RepA protein can assemble into amyloid fibers upon binding to DNA in vitro.

In the proposed model, this phenomenon is simulated by a two-level automaton. At first level (Level-I) each occupied cell represents one of the three different DNA sequences (ADN_1, ADN_2, ADN_3) that can interact with the protein subunits (PrP). At second level (Level-II) proteins are represented. These cells can transit, with some probability defined at probability vector (P_v), from six different states according to the transition rules defined in each case. The formalization of the proposed cellular automata is described below. Our cellular automata is defined by M= (G, G_0, N, Q, δ, T), where:

- G: matrix automata (Q-dimensional);
- G_0: set of initial values of the automata (G) states;
- N: function that assigns each automaton the set of neighbors (Neighborhood function);
- Q: sets of possible states;
- δ: transition function that assigns a new state to an automaton having into account the state of all its neighbors;
- T: set of final states.

The automata consists of a set of six possible states (Q= {S_0, S_1, S_2, S_3, S_4, S_5}), it implements the well-known function of Von Neumann neighborhood [Kennedy & Mendes, 2003] and has a unique final state, S_4. Each of the possible states represents a different physicochemical prion formation state during the process, thus:

- S_0= represents the absence of protein in the cell;
- S_1= represents a stable state of the protein. This state is not affected by the interaction with the DNA grid (Level-I);
- S_2= molecular state of instability. Cells in this state are potentially exposed to interaction with Level-I. The neighborhood function is applied and executed transition rules;
- S_3= represents a functional protein;
- S_4= represents a prion infectious form. Seed of the nucleation center;
- S_5= represents a nonfunctional protein;

The transition rules, TR = {$p_{0,1}$, $p_{1,2}$, $p_{2,3}$, $p_{2,4}$, $p_{2,5}$, $p_{3,0}$, $p_{5,0}$}, are applied in each case with a certain probability. In particular, the rules $p_{0,1}$ and the set {$p_{3,0}$,$p_{5,0}$} represent the transition probability $S_0 \rightarrow S_1$, $S_3 \rightarrow S_0$ y $S_5 \rightarrow S_0$, these probabilities represent the rate of molecular generation and degradation rates respectively. In the case of the rule $p_{1,2}$ represents the transition probability $S_1 \rightarrow S_2$ in which the cell moves from one stable state to a state of molecular instability. Finally, the rule set {$p_{2,3}$,$p_{2,4}$, $p_{2,5}$} implements the transition functions related to the interaction with DNA sequences (Level-I) by applying the selected neighborhood function ($S_2 \rightarrow S_3$ binding to ADN_1, $S_2 \rightarrow S_4$ binding to ADN_2, $S_2 \rightarrow S_5$ binding to ADN_3). This behavior is shown in FIGURE 5.

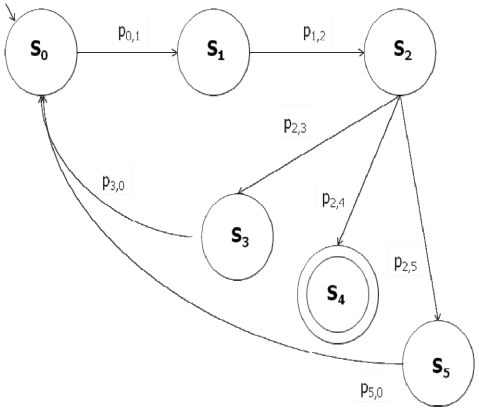

Figure 5. Pion-crystalization Automaton. State-transition detail.

In this paper are shown two simulations of the proposed model. The first one corresponds to the simulation without the presence of input pattern. By this simulation it can appreciate the molecular concentrations of different states according to the underlying model at Phase-I. The second simulation presents the results obtained during the training phase of the system. Note that in this work the results related to Phase-II of the model (algorithm crystallization) are not shown. The results obtained by applying the model of crystallization and the appropriate decoding for final classification of the patterns will be shown in the near future.

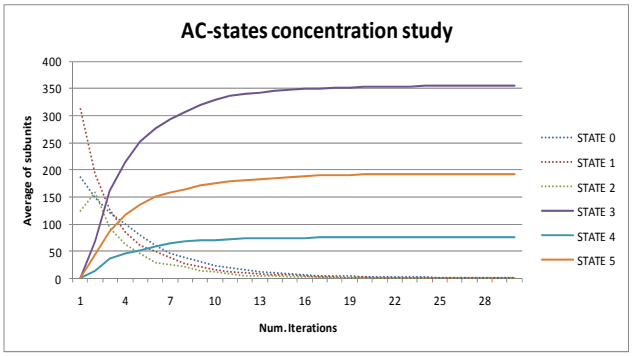

Figure 6. Average of states concentration chart.

At first simulation, the initial population of items, molecules, is randomly distributed on a two-dimensional square lattice with periodic boundary conditions, $N_X \times N_Y = 50 \times 50$. The initialization of the Level-I grid corresponds to three-type DNA initial concentration (C_{DNAn}). These values have been calculated according to the following proportions: $C_{DNA3} > (C_{DNA1} + C_{DNA2})$ and $C_{DNA1} > C_{DNA2}$. In FIGURE 6 results over 10 simulations with same settings can be observed. These average concentrations were obtained after 30 system iterations and with rates of generation and degradation of 0,15 and 0,45 respectively, for each simulation. As you it can see the concentration of state 3 (functional proteins triggered by DNA_1) is higher despite the initial conditions of concentration of DNA sequences, where the proportion of type sequences DNA_3 is majority. This behavior corresponds proportionally to results observed in real (in-vitro) prion formation processes.

Finally, it is shown the results obtained in the training phase of the proposed system, FIGURE 7. This training experiment aims to get the system to recognize alphanumeric characters. Specifically, in the case above, is introduced an input pattern which codifies the character 'A' (FIGURE 7-a), the input pattern is encoded in the Level-I of the model. The DNA sequences of this level are initialized with concentrations according to the codification of the input pattern. As a result of applying the underlying algorithm it is obtained the corresponding Level-II, state level, configuration (FIGURE 7-b) and also the resulted configuration of Level-III, nucleation seed level, (FIGURE 7-c); starting point of crystallization process. As mentioned above, this process will result in encoded input patterns implicitly in the resulting morphology of prion forms. By applying the decoding algorithm designed for the system will be able to classify unknown input patterns.

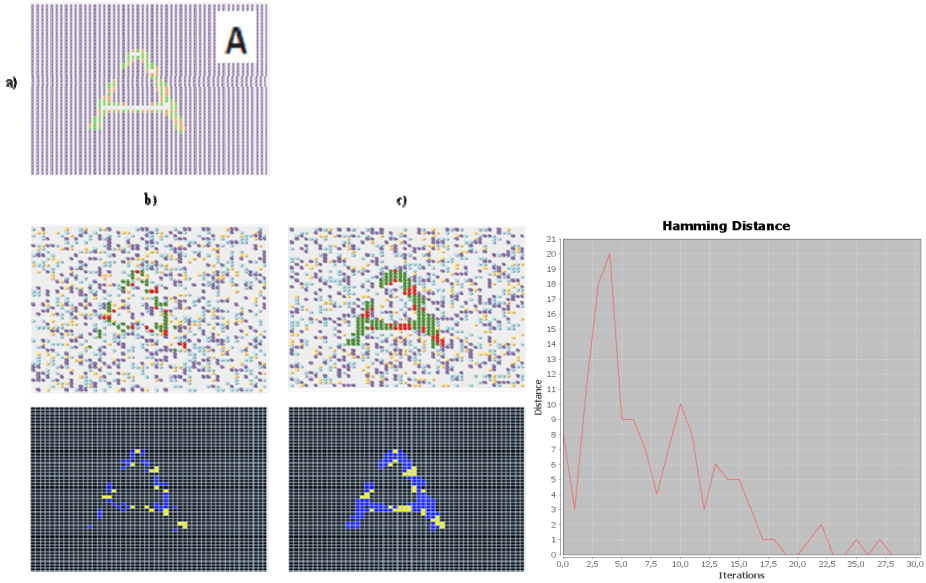

Figure 7. Pion-crystalization Automaton. State transition detail. a)encoded pattern in Level-I b) Level-II and Level-III at 30% simulation c) Level-II and Level-III at hundred 100% simulation d) Hamming distance beyween input pattern and resulted outputs.

Conclusion

This paper introduces APA: a novelty artificial pattern recognition system based on prion crystalization. At present we conducted the modeling and simulation experiments showing the plausibility of a memory based on prion self-assembly. Thus, we studied the main steps and features of the training step. The system is able to memorize patterns into the resulting prion self- assembly. Such memory is implemented as a hierarchical (DNA, proteins, crystal assembly) bioinspired 2D cellular automata. Our approach opens the possibility of designing pattern recognition systems inspired by the phenomenon of crystallization in biology.

Bibliography

[Bemporad & Chiti, 2013] Bemporad, F., & Chiti, F. Pathways of Amyloid Formation. In: Amyloid Fibrils and Prefibrillar Aggregates: Molecular and Biological Properties, 2013.

[Brown et al, 1997] Brown, D. R., Schulz-Schaeffer, W. J., Schmidt, B., & Kretzschmar, H. A. Prion protein-deficient cells show altered response to oxidative stress due to decreased SOD-1 activity. In: Experimental neurology, 1997.

[Edwards & Anderson, 1975] Edwards, S. F., & Anderson, P. W. Theory of spin glasses. In: Journal of Physics F: Metal Physics, 1975.

[Fernández et al, 2010] Fernández Tresguerres, M. E., Moreno Díaz de la Espina, S., Gasset Rosa, F., & Giraldo, R. A DNA-promoted amyloid proteinopathy in Escherichia coli. In: Molecular microbiology, 2010.

[Fontaine & Brown, 2009] Fontaine, S. N., & Brown, D. R. Mechanisms of prion protein aggregation. In: Protein and peptide letters, 2009.

[Greer et al, 2006] Greer, M. L., Pujo-Menjouet, L., & Webb, G. F. A mathematical analysis of the dynamics of prion proliferation. In: Journal of theoretical biology,2006.

[Hopfield, 1982] Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. In: Proceedings of the national academy of sciences, 1982.

[Kennedy & Mendes, 2003] Kennedy, J., & Mendes, R. Neighborhood topologies in fully-informed and best-of-neighborhood particle swarms. In: Soft Computing in Industrial Applications, 2003.

[Kryndushkin et al, 2011] Kryndushkin, D. S., Wickner, R. B., & Tycko, R. The core of Ure2p prion fibrils is formed by the N-terminal segment in a parallel cross-β structure: evidence from solid-state NMR. In: Journal of molecular biology, 2011.

[Kulkarni et al, 2003] Kulkarni, R. V., Slepoy, A., Singh, R. R. P., Cox, D. L., & Pázmándi, F. Theoretical modeling of prion disease incubation. In: Biophysical journal, 2003.

[Lahoz-Beltra, 1997] Lahoz-Beltra, R. Molecular automata assembly: principles and simulation of bacterial membrane construction. In: Biosystems. 1997.

[Lahoz-Beltra, 1999] Lahoz-Beltra R. Molecular automata modeling in structural biology. In: Advances in Structural Biology, 1999.

[Levi & Kotrla, 1997] Levi, A. C., & Kotrla, M. (1997). Theory and simulation of crystal growth. In: Journal of Physics: Condensed Matter, 1997.

[Slepoy et al, 2001] Slepoy, A., Singh, R. R. P., Pazmandi, F., Kulkarni, R. V., & Cox, D. L. Statistical mechanics of prion diseases. In: Physical review letters, 2001.

[Stahl et al, 1987] Stahl, N., Borchelt, D. R., Hsiao, K., & Prusiner, S. B. In: Scrapie prion protein contains a phosphatidylinositol glycolipid. In: Cell, 1987.

[Witten & Sander, 1981] Witten Jr, T. A., & Sander, L. M. (1981). Diffusion-limited aggregation, a kinetic critical phenomenon. In: Physical Review Letters, 1981.

Authors' Information

Paula Cordero – Natural Computing Group, Facultad de Informática,Universidad Politécnica de Madrid, Campus de Montegancedo s.n., 28660 Boadilla del Monte, Madrid, Spain; e-mail: paula.cormo@gmail.com

Major Fields of Scientific Research: complex systems, natural computing, evolutionary computation and the design of bioinspired algorithms.

Rafael Lahoz-Beltra – Department of Applied Mathematics, Faculty of Biological Sciences, Complutense University of Madrid, 28040 Madrid, Spain ; e-mail: lahozraf@ucm.es

Major Fields of Scientific Research: evolutionary computation, embryo development modeling and the design of bioinspired algorithms

Juan Castellanos – Head of Natural Computing Group, Facultad de Informática,Universidad Politécnica de Madrid, Campus de Montegancedo s.n., 28660 Boadilla del Monte, Madrid, Spain; e-mail: jcastellanos@fi.upm.es

Major Fields of Scientific Research: natural computing, formal language and automata theory.