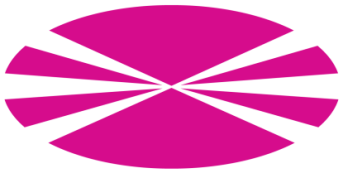

UNIVERSIDADE DA CORUÑA

Facultad de Informática

Departamento de

Tecnologías de la Información y las Comunicaciones

Modelado de un sistema celular artificial para generación de formas y procesado de información

Tesis Doctoral

Directores

Julián Dorado de la Calle

Cristian Robert Munteanu

Doctorando

Enrique Fernández Blanco

A Coruña, Junio 2010

UNIVERSIDADE DA CORUÑA

Dr. Julián Dorado de la Calle, Profesor Titular de Universidad en el área de Ciencias de la Computación e Inteligencia Artificial, perteneciente al Departamento de Tecnologías de la Información y las Comunicaciones

Y

Dr. Cristian Robert Munteanu, Contratado en el programa Parga Pondal en Biomedicina dentro el área de Ciencias de la Computación e Inteligencia Artificial, perteneciente al Departamento de Tecnologías de la Información y las Comunicaciones

HACEN CONSTAR QUE:

La memoria "**Modelado de un sistema celular artificial para generación de formas y procesado de información"** ha sido realizada por **D. Enrique Fernández Blanco**, bajo nuestra dirección en el Departamento de Tecnologías de la Información y las Comunicaciones, y constituye la Tesis que presenta para optar al Grado de Doctor en Informática de la Universidade da Coruña.

A Coruña, 5 de Abril de 2010

Fdo: Julián Dorado de la Calle Fdo: Cristian Robert Munteanu

Agradecimientos

Aunque habitualmente no se note, soy un sentimental empedernido. Esta faceta de mi carácter habitualmente no se vislumbra ya que me lo ha hecho pasar mal más de una vez. Pero, es en momentos como el de cubrir estas páginas de agradecimientos cuando surge, sin lugar a dudas, ese aspecto de mi carácter. Sólo espero que quién lea estas líneas me permita y comprenda esta pequeña licencia al final de un arduo camino lleno de obstáculos que ha habido que superar. Son un millar de recuerdos los que se agolpan en mi cabeza, por tanto, voy a intentar ordenarlos un poco en estas líneas. Ha pasado mucho tiempo desde que me trasladé a La Coruña para estudiar y pensaba que a lo mejor algún día podría llegar a ser doctor. Era casi como un sueño infantil, imposible y lejano. Pues bien, ahora me veo finalizando la tesis doctoral y, desde luego, es la mejor prueba de que, a veces, alguno de nuestros sueños puede hacerse realidad. Algo que veía tan lejano está a punto de tornarse palpable y real. Por el camino han quedado algunas cosas buenas, algunas malas y, sin lugar a dudas, un montón de trabajo.

El primero que me viene a la cabeza, a la hora de agradecer, es, sin lugar a dudas, mi director Julián, ya que fue él quien me permitió empezar a soñar con este día. Parece que fue ayer, cuando, hace cinco años, me comentó que tenía una idea pero que habría que probarla y trabajar muchísimo sobre ella. Pues bien, el resultado de esa idea y de nuestro trabajo ha sido una tesis doctoral y un puñado de publicaciones. Gracias por la paciencia, el apoyo, la guía, y en general, por confiar más en mí que yo mismo.

Agradecer a los "jefes" Alejandro, Juanra, Anuska y Nieves por su ejemplo, por ser más unos amigos que mis jefes. Gracias por vuestra comprensión, y, por tener siempre tiempo para escucharme. No puedo sino estar agradecido a todos y cada uno de ellos.

Qué puedo decir del RNASA, al que cariñosamente llamo "el agujero", pero que ha llegado a ser "mi agujero". Lo único que me he encontrado es gente con la que me he reído, he discutido y, en definitiva, he vivido. Todos los que han pasado por allí en estos años, y que no nombro por miedo a olvidarme a alguien, merecen todo mi agradecimiento. Permitidme sólo un recuerdo especial a la "vieja guardia", es decir, Mónica, Marcos, Jose Antonio, José Power, Juan Luis, Fran y Dani. Ellos son quienes más me han aguantado y apoyado, tarea nada fácil de vez en cuando. Por el camino, han pasado de ser mis compañeros de trabajo a mis amigos, acogiéndome desde el primer momento como una "rana" más.

Además quisiera recordar a los miembros del Intelligent Systems Research Group de York, por recibirme con las puertas abiertas, por hacerme sentir como en casa, y hacerme pasar dos de los mejores meses que recuerdo. Por eso, por sus consejos, y por esas conversaciones a orillas del lago sobre cualquier tema, quisiera mencionar especialmente a Tuzë, Martin, James y Julian.

Un recuerdo también para la gente del IEETA de la Universidade de Aveiro y, en especial, para los profesores José Luis Oliveira y Carlos Costa por ayudarme y acogerme en mi estancia con ellos con los brazos abiertos.

Otra gente que merece todo mi agradecimiento es ese pequeño grupo que por alguna razón incompresible me aguantan y al que llamó mis amigos. Sin ellos, desde luego, esto no sería una realidad. Son una de las razones por las que esto tiene sentido y, la verdad, no sé qué sería de mí muchas veces sin ellos. Mención especial para Curra y Reichi, por saber sacarme una sonrisa cuando más lo necesito; José y Diana, por saber escucharme y hacerme ver el lado positivo de todo por difícil que sea; Carlos, por soportar interminables horas de discusión y servirme de acicate intelectual para mirar hacia adelante y ver "The bright side of life"; finalmente, Cancela, León, Fernando y Simón, porque como mis compañeros de piso habéis aguantado más el proceso de esta tesis que ningún otro. Y en general, a todos los que, de vez en cuando, se refieren a mí como su amigo, gracias por demostrarme que aún se puede confiar siempre en alguna gente.

He dejado para el final precisamente a aquellos a los que estoy más agradecido. Esa gente a la que llamó mi familia. Mis primos, mis tíos, mis hermanos, mi sobrino y mi madre son, sin lugar a dudas, los que más se merecen este reconocimiento porque esta tesis es para ellos. Ellos son los que hacen que ponga los pies en la tierra, me centre y continúe mi camino. Son ellos los que soportan mis depresiones (a veces demasiado frecuentes), los que comparten mis alegrías y, en general, los que hacen que todo valga la pena. Quisiera que se sintiesen orgullosos porque una parte importante de todo esto también es suya. Especialmente quisiera mencionar a mi hermana, Amaya, y mi madre, gracias por escucharme, entenderme, apoyarme y, en definitiva, por hacer algo tan difícil como quererme tal y como soy.

A todos sin excepción desde el fondo de mi corazón, Muchas Gracias.

Enrique Fernández Blanco

A mi madre

Los que sueñan de día son conscientes de muchas

cosas que escapan a los que sólo sueñan de noche

Edgar Alan Poe

Resumen

En el ámbito de la informática se han modelado distintos procesos naturales para adaptar sus fundamentos en la resolución de problemas. En los últimos años algunos investigadores han centrado su atención en el comportamiento de las células no nerviosas. El motivo de este interés se debe a las características que presentan dichas células en cuanto a autoorganización y procesado de señales.

Las células naturales de un organismo son capaces de autoorganizarse usando unas pocas señales y la información contenida en el ADN de las mismas. Además, si se piensa en una célula, esta recibe múltiples señales de distintas fuentes y referidas a varios problemas, y, la célula, es capaz de dar una respuesta coordinada a todos, procesando la información en paralelo con otras células. Adaptar este comportamiento en un modelo artificial supondría una nueva herramienta que facilitaría afrontar problemas como los multiobjetivo.

El objetivo de esta tesis es realizar un nuevo paso para la consecución de ese objetivo. Así se busca estudiar e identificar los mecanismos más útiles del modelo biológico y crear un modelo artificial que los incluya.

Para comprobar el comportamiento de ese nuevo modelo, se plantea realizar algunas pruebas clásicas que se basan en la generación y autoorganización de distintas formas geométricas. Además, también se hace una primera incursión en el estudio de la aplicación de este tipo de modelos a la resolución de problemas de clasificación de entradas, que no se había hecho anteriormente con ningún otro modelo dentro de la Embriogénesis Artificial.

Abstract

Fundamentals of different natural processes in the field of Computer Science have been modelled in order to apply them in problem-solving situations. In recent years, the behaviour of non-nervous cells has been the focus of attention of some researchers. The main reason of this attention consists in the features shown by these cells in terms of self-organisation and signal processing capacities.

Natural cells of an organism are able to self-organise themselves by using a few signals and the information contained in their DNA. Moreover, cells receive many signals from different sources which are associated with several problems and they are able to process all those signals and coordinate their response at the same time as their neighbours, processing the signals and giving a coordinate response. The Artificial Models, which can adapt that behaviour, are the new tools facing challenges such as multi-objective problems.

This thesis is aimed at making another step towards this objective. Thus, the main focus of this work is to study and identify the most relevant mechanisms of the biological model and develop an artificial model by adapting these mechanisms.

In order to check the behavior of the development model, some standard assays based in the generation and self-organization of different geometrical forms were performed. Furthermore, the model presented herein is the first one of this kind of models applied to a new area such as the resolution of pattern classification problems, where no other Artificial Embryogeny model was applied before.

Índice

Índice de Figuras

Índice de Tablas

Que tu lengua no se adelante a tu mente

Quilón de Esparta

Capítulo 1

Introducción

Estudio, observación e inspiración son las bases para la ciencia. El estudio y la observación se basan en el esfuerzo que los investigadores emplean en el trabajo realizado. Pero, por el contrario, la inspiración puede provenir de prácticamente cualquier sitio, como pueden ser los distintos procesos naturales, el comportamiento de los animales o, simplemente, del uso de procesos, descubrimientos o conocimientos de un campo, en otro distinto. Thomas Alva Edison decía que:

"El éxito es una combinación de un diez por ciento

de inspiración y un noventa por ciento de esfuerzo"

Todas las áreas del conocimiento, sin excepción, han utilizado conocimiento, conceptos o descubrimientos de otras áreas como inspiración o para aplicarlas en las propias. El resultado de este trasvase de información, aparte de evidentemente hacer avanzar el campo que recibe la información, aporta nuevo conocimiento al área de la cual se facilita la información previa, o bien, también puede surgir un conocimiento totalmente nuevo a medio camino entre las dos áreas. Así, por ejemplo, el estudio de las aves, en particular del mecanismo y morfología de sus alas, aplicado en el campo de la ingeniería ha dado como resultado un campo como la aeronáutica.

La informática, a pesar de ser uno de los campos más jóvenes del conocimiento humano, puede que sea uno de los campos más influenciados por este tipo de trasvase de información. Esto se debe a que la informática ha sido tradicionalmente un campo que ha prestado apoyo a otros para automatizar sus procesos. Este contacto ha dado como resultado que, conocimientos provenientes de distintos campos, han sido adquiridos por la informática que se ve mejorada por la creación de nuevas técnicas.

Existen multitud de estas transferencias de conocimiento a la informática, pero hay un campo que destaca especialmente. Ese campo no es otro que la biología, ya sea con los procesos macroscópicos como puede ser el comportamiento de los animales, como los procesos microscópicos como puede ser el funcionamiento de las neuronas, la biología ha sido la fuente de inspiración de muchas de las técnicas de la informática que hoy en día se conocen. Así, técnicas como las redes de neuronas artificiales (McCulloch & Pitts, 1943), los algoritmos de enjambres de partículas (Kennedy & Eberhart , 1995), las colonias de hormigas artificiales (Dorigo et al., 1996), los autómatas celulares (Wolfram, 1994), los algoritmos genéticos (Holland, 1975), etc. son técnicas conocidas e utilizadas

por los expertos en informática que basan su funcionamiento en algún proceso existente en la biología.

El trabajo que aquí se presenta se puede incluir dentro de este grupo de trabajos que tienen su inspiración en elementos de la biología. En concreto, la atención de este trabajo se ha centrado en el comportamiento y funcionamiento de las células y particularmente, en las células embrionarias. Esta área ha recibido multitud de nombres como son Embriología Computacional, Embriogénesis Artificial, Sistemas Generativos y de Desarrollo, etc. Lo que todos tienen en común es que, en todos los casos, el trabajo trata de realizar un modelo artificial del proceso celular de desarrollo.

El desarrollo de los sistemas tanto software como hardware ha llegado a un punto que, cualquier modificación o problema puede llevarle a un experto solventarlo días, semanas o incluso no ser capaz de hacerlo. Por tanto, el modelo tradicional de desarrollo de sistemas parece tener los días contados. En este sentido, si se centra la atención en la naturaleza, se descubre que esta es capaz de desarrollar sistemas muy complejos a partir de un único punto con, relativamente, poca información. La adaptación de este comportamiento puede suponer un nuevo enfoque, ya que se dejaría de diseñar el sistema por completo por darle indicaciones y evolucionarlo.

En este sentido, las células embrionarias presentan algunas características que resultan interesantes desde el punto de vista de adaptación para las ciencias de la computación. Cualquiera de las células que compone un tejido, se comporta de manera similar a un procesador, ya que recoge las señales de su entorno y las procesa siguiendo un conjunto de instrucciones que se encuentran codificadas en el ADN de las células. Ese

procesado llevado a cabo por las células da, como resultado, una señal de salida que puede provocar distintos comportamiento en la célula. Esos comportamientos pueden ser: comunicar esa salida a las células que la rodean, almacenar esa salida para usarla posteriormente, puede provocar la mitosis de la célula (o división celular) o incluso la apoptosis celular (también conocida como muerte celular programada).

Ese comportamiento, descrito para cada célula, habría que multiplicarlo por cada una de las células que componen un tejido. Las células, por tanto, no exhibirían un comportamiento aleatorio, si no que cada una tendría un comportamiento coordinado con las demás células del tejido. Este comportamiento tiene como fin obtener un objetivo común, que no es otro que la supervivencia de la mayoría.

Por tanto, este proceso se puede interpretar como un conjunto de procesadores individuales que deben coordinar sus comportamientos para obtener un objetivo general más complejo. Este sistema tendría como particularidad que los elementos que lo componen se autoorganizarían, carecen de un control centralizado y que pueden resolver más de una función de manera simultánea. Las células determinarían su función según su posición relativa con respecto a las otras, las señales que reciben y la información codificada en el ADN.

1.1 Objetivos Generales

El principal objetivo de esta tesis es desarrollar un modelo que, teniendo como ideal el modelo biológico de célula, trate de adaptar ese modelo de proceso de datos en un sistema artificial.

Como ya se apuntó anteriormente, el modelo biológico tiene muchas propiedades interesantes desde el punto de vista de las ciencias de la computación. Adaptando el procesado de datos que llevan a cabo las células, se pretende incorporar algunas de las mencionadas características en un modelo que sea utilizable en un amplio espectro de problemas. Algunos ejemplos de las características que se quieren obtener son la autoorganización o el control distribuido de los elementos que lo compongan.

Esta no es para nada la primera vez que se intenta esta adaptación. Existen experiencias previas de distintas aproximaciones realizadas por distintos autores, las cuales, se discuten en el capítulo de la tesis dedicado a realizar un repaso en profundidad de la biografía del campo. El punto diferenciador del modelo propuesto en este trabajo con otros modelos existentes en la literatura, es que, desde un principio, se pensó en aplicar el mismo modelo a muchos tipos de problemas distintos. Normalmente, los modelos inspirados en las células que se han propuesto hasta ahora están centrados en algún tipo de problema específico, como puede ser la generación de una forma concreta (Kumar & Bentley, 2003).

Lo que se pretende en este trabajo es diseñar un modelo más abierto que sea aplicable a un conjunto mayor de problemas que los existentes y no encorsetarlo a un problema concreto. En cierta medida, la motivación es crear una nueva técnica que tome como ejemplo las Redes de Neuronas Artificiales (Rumelhart et al 1986; Arbib, 2002), que también son una técnica bioinspirada, pero que son aplicables a multitud de tipos de problemas.

Todo aquel que haya trabajado con las Redes de Neuronas Artificiales sabe que uno de los mayores problemas es encontrar la arquitectura adecuada para el problema a tratar. La mayoría de las redes de neuronas deben optar por una arquitectura para la red desde un primer momento, para un problema, antes de intentar solucionarlo. Desarrollar técnicas que incluyan la autoorganización podría solventar ese problema. Así el trabajo que se presenta en este documento puede ser visto como un paso en el sentido de conseguir estructuras que sean autoorganizativas y que procesen información, de manera que sean aplicables a un amplio espectro de problemas.

El modelo propuesto en este trabajo usa lo que se conoce como Redes Reguladoras de Genes (Gene Regulatory Networks). El sistema define un conjunto de reglas con la mencionada estructura que, como ya está demostrado en (Kauffman, 1969), tiene la misma capacidad para procesar información que el Algebra de Boole. Este conjunto de reglas controlan el comportamiento de las células artificiales que componen el sistema del mismo modo que el ADN controla el comportamiento de las células biológicas.

Ese conjunto de reglas, en forma de Red Reguladora de Genes, junto con un sistema para el paso de mensajes entre células forman la base del funcionamiento del sistema. A su vez, esa base plantea un sistema de procesado de información simple basado en la modelización del funcionamiento del ADN del sistema biológico, que es aplicable a varios problemas distintos que se muestran en el Capítulo 6.

Los problemas que se pueden encontrar en dicho capítulo se pueden dividir en dos grandes grupos: Generación de Formas y Procesado de Información. En el primer conjunto de problemas se verá cómo utilizar el modelo celular propuesto con una Red

Reguladora de Genes, que se ha encontrado mediante un Algoritmo Genético (Fernández-Blanco et. al., 2007), para generar estructuras con unas determinadas propiedades. El segundo conjunto de problemas tiene como objetivo estudiar la viabilidad de resolver problemas de procesado de información, como el problema de la clasificación de las flores Iris (Fisher, 1936), haciendo uso del modelo celular y una Red Reguladora de Genes.

1.2 Estructura de la Tesis

Esta tesis se estructura en capítulos de manera que, empezando desde una breve descripción de los procesos biológicos, continúa con una revisión de las distintas técnicas existentes en la literatura, para concluir con el desarrollo del modelo planteado en este trabajo. Dicho modelo, trata de alcanzar los objetivos descritos en el apartado 1.1. Más específicamente, la distribución de los capítulos de esta tesis quedaría como sigue:

- **Capítulo 1:** Este capítulo plantea el área de trabajo de esta tesis y los objetivos que se persiguen con su realización.
- **Capítulo 2:** El capítulo contiene todos los conceptos que se deben manejar para comprender el desarrollo completo de la tesis.
- **Capítulo 3:** En este capítulo se hace un repaso detallado de los trabajos más avanzados dentro de la Embriogénesis Artificial.
- **Capítulo 4:** Este capítulo contiene la hipótesis central de esta tesis y plantea los objetivos a conseguir mediante su realización.

- **Capítulo 5:** Aquí se detallan las partes del modelo propuesto de Embriogénesis Artificial de esta tesis.
- **Capítulo 6:** Contiene las pruebas realizadas sobre el modelo y que tratan de probar distintos aspectos del modelo propuesto.
- **Capítulo 7:** El capítulo contiene las conclusiones generales que se pueden extraer tras la realización de la tesis.
- **Capítulo 8:** El último de los capítulos de la tesis está dedicado a plantear las futuras líneas de investigación que se pueden afrontar tras la realización de la tesis doctoral.

Do not let one's tongue outrun one's sense.

Chilon of Sparta

Chapter 1

Introduction

The foundations of science are built on study, observation and inspiration. The study and observation are based on the effort that researchers invest in their work. But, on the other hand, inspiration can basically come from anywhere, such as so many natural processes, animals' behaviour or simply the application of processes, discoveries or knowledge from a field to another. Thomas Alva Edison used to say:

"Genius was one percent inspiration and ninety-nine percent perspiration."

All areas of knowledge, without exception, have used knowledge, concepts or discoveries from other areas as inspiration or application to their own. The result of this transfer of information, besides an obvious progress in the field receiving the information, provides a new insight into the area from which the prior information is originated, or entirely new data may also arise midway between the two areas. For example, the study of birds, specifically the functioning and morphology of their wings, applied in the field of engineering has resulted in a field such as aeronautics.

Computer science, despite being one of the newest fields of human knowledge may be one of the areas most affected by this type of information transfer. This is due to the fact that computing has traditionally been a field that provided support to others in order to automate their processes. This interconnection has led to the fact that knowledge from different fields was acquired by Computer Science, enhanced by the emergence of new techniques.

There are many of these knowledge transfers to computing, but there is a field that especially stands out. This field is none other than Biology, involving either macroscopic processes such as animals' behaviour or microscopic processes such as functioning of neurons; Biology has been the source of inspiration for many of the information techniques known today. Thus, techniques such as Artificial Neural Networks (McCulloch & Pitts, 1943), the particle swarm algorithms (Kennedy & Eberhart, 1995), artificial ant colonies (Dorigo et al., 1996), cellular automata (Wolfram, 1994), genetic algorithms (Holland, 1975), etc. are known techniques, used by computer experts, which base their functioning on some existing process in Biology.

The work presented here can be included within this group of works which have their inspiration in concepts of Biology. Specifically, this work has been focused on the behaviour and functioning of cells and particularly of embryonic cells. This area has received many names such as Computational Embryology, Artificial Embryogenesis, Generative and Development systems, etc. They all have in common that in all cases, the aim is to carry out an artificial model of cellular development.

The development of both software and hardware systems has reached such a complexity that any change or problem may occur, would involve days or weeks of work from an expert who may even not manage to solve it. Therefore, the traditional model of system development seems to have its days numbered. In this regard, if we focus our attention on the natural environment, we discover that it is able to develop very complex systems starting from a single point, with relatively little information. This behaviour adaptation may lead to a new approach, since the designing of the entire system would be given up by providing information or making it evolve.

Accordingly, embryonic cells have several features, interesting for the computer science from an adaptation viewpoint. Each cell is made up of a tissue and behaves similarly to a processor since it collects the signals from their environment and processes them using a set of instructions encoded in the DNA sequence of cells. This process carried out by cells results in an output signal that can cause different behaviours in the cell. These behaviours include: transmitting that output to the cells that surround it and storing that output for later use can cause cell mitosis (or cell division) or even cell apoptosis (also known as the process of programmed cell death).

This behaviour, described in each cell, would have to be reproduced by each of the cells in a tissue. The cells, therefore, do not exhibit a non-random behaviour, but each of them behaves coordinately with the other tissue cells. This behaviour is aimed at achieving a common goal, which is none other than the survival of the majority.

Consequently, this process can be interpreted as a set of individual processors which must coordinate their behaviours to achieve a more complex overall objective. This system's special feature is that the elements which make it up are able to self-organise themselves, lack a centralised control mechanism and are able to solve more than one function simultaneously. The cells determine their function according to their relative position regarding the others, the signals received and the information encoded in the DNA sequence.

1.1 General Objectives

Starting from the idea that the cell biological model is an ideal one, the main aim of this thesis is to develop a model that attempts to adapt this model of data processing to an artificial system.

As noted above, the biological model has many interesting characteristics from the point of view of Computer Science. When adapting the data processing carried out by cells, it is expected that some of these features were included in a model that could be used in a wide range of problems. Some examples of the features to be obtained are self-organisation or the distributed control of its component items.

This is not the first time at all that this adaptation is sought. There are previous attempts of different approaches performed by several authors, which will be discussed in the chapter of the thesis focused on reviewing in depth the literature in the field. The differentiating approach of the model proposed in this work with respect to other existing models in literature, is that the initial idea was to apply the same model to many different types of problems. The cell-inspired models that have been proposed so far, are usually focused on a specific problem, such as the generation of a specific geometric shape (Kumar & Bentley, 2003).

The aim of this work is to design an open model, applicable to a wider range of existing problems, not restricted to solving only one specific problem. To some extent, the incentive is to create a new technique that follows the example of the Artificial Neural Networks (Rumelhart et al 1986; Arbib, 2002), which are also a bio-inspired technique, but which can be applied to many types of problems.

Anyone who has employed Artificial Neural Networks knows that one of the biggest problems is finding the appropriate architecture for the problem to be solved. From the very beginning, before attempting to solve a problem most neural networks must choose an architecture for the network. The problem could be solved by developing techniques that include self-organisation. Thus, the work presented herein can be considered as a step forward in obtaining structures that are self-organising and able to process information, so they are applicable to a broad spectrum of problems.

The model proposed in this thesis uses what is known as Gene Regulatory Networks. The system defines a set of rules with the abovementioned structure that, as already

shown in (Kauffman, 1969), has the same ability to process information as the Boolean algebra. This set of rules controls the behaviour of artificial cells that make up the system in the same way that DNA controls the behaviour of biological cells.

This set of rules, represented as a Gene Regulatory Network, along with a system for the transmission of messages between cells, make up the operation basis of the system. In turn, this basis raises a simple information processing system depending on the modelling of the DNA functioning of the biological system, which is applicable to several different problems that are listed in Chapter 6.

The problems mentioned in this chapter can be divided into two groups: Generation of Forms and Information Processing. For the first set of problems, we will see how to use the proposed cell model with a Gene Regulatory Network, which has been obtained by means of a Genetic Algorithm (Fernandez-Blanco et. al., 2007), in order to generate structures with some specific properties. The second set of problems aims to study the feasibility of solving problems related to information processing, such as the Iris Flower classification problem (Fisher, 1936), using the cell model and a Gene Regulatory Network.

1.2 Structure of the Thesis

This thesis is divided into chapters so that starting from a brief description of the biological processes, it continues then with a review of the different techniques in the literature, concluding with the development of the model proposed in this work. This

model seeks to achieve the objectives described in section 1.1. More specifically, the distribution of the chapters of this thesis would be as follows:

- **Chapter 1**: This chapter presents the working range of this doctoral thesis and the objectives pursued with its implementation.
- **Chapter 2:** This chapter contains all the concepts that should be considered in order to understand the complete development of the thesis.
- **Chapter 3**: This chapter gives a detailed overview of the most advanced studies within the Artificial embryogenesis.
- **Chapter 4:** This chapter contains the central hypothesis of this thesis and outlines the objectives to be achieved through implementation.
- **Chapter 5:** It contains a detailed presentation of all the parts of the model on Artificial Embryogenesis proposed in this thesis.
- **Chapter 6:** It contains the assays carried out on the model and which try to test different aspects of the model proposed.
- **Chapter 7:** This chapter contains the general conclusions to be drawn after the completion of the thesis.
- **Chapter 8:** The last chapter of the thesis is aimed at setting out future lines of research that can be tackled after the completion of the doctoral thesis.

El gran libro de la Naturaleza está

escrito en símbolos matemáticos

Galileo Galilei

Capítulo 2

Fundamentos

En este capítulo de la tesis se comenzará discutiendo aquellos conceptos provenientes de la biología que es necesario conocer. Esta discusión no pretende ser un manual sobre los procesos biológicos sino un mero resumen que permita comprender la adaptación de los distintos conceptos de las células en el modelo artificial planteado en esta tesis.

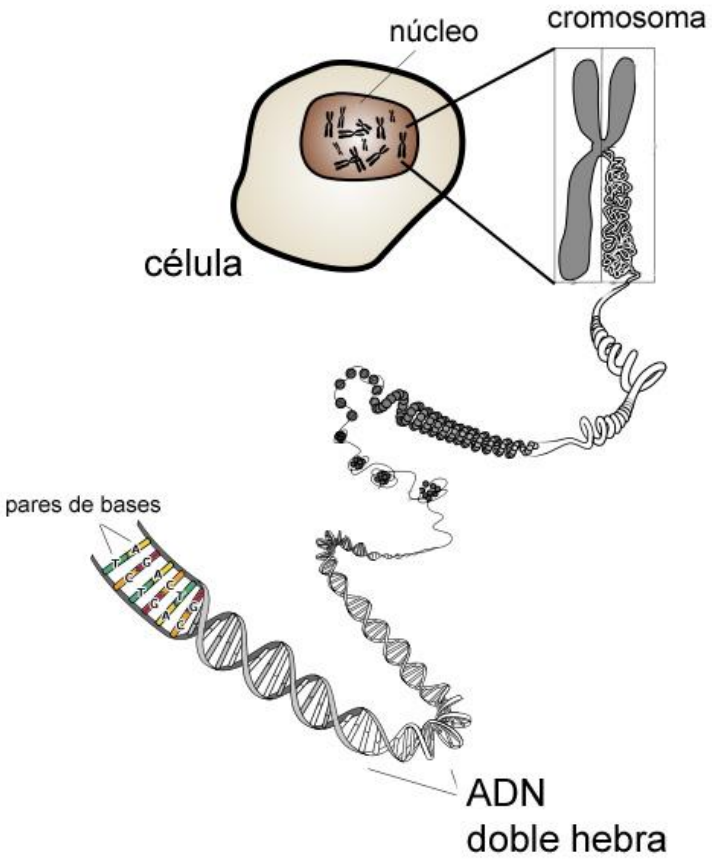

Figura 2.1 Esquema de ADN contenido dentro de una célula

Además de esos conceptos biológicos, en este capítulo, también se encuentra una explicación de lo que son las Redes Reguladora de Genes, los Algoritmos Genéticos y Embriogénesis Artificial. Si bien es evidente que estos no son conceptos fundamentales provenientes de la biología, son técnicas de las ciencias de la computación que resulta necesario manejar para entender el diseño del modelo de célula artificial y el desarrollo llevado a cabo durante las pruebas.

2.1 Conceptos Biológicos

El ADN es la molécula en forma de doble hélice, donde está almacenada la información genética de un individuo (ver Figura 2.1). Esta molécula contiene toda la información de un ser vivo tanto las características comunes de la especie a la que pertenece, como las particulares del individuo analizado. Esta información está presente en todas las células que forman el organismo, ya que cada célula contiene una copia completa del ADN. La diferencia entre las células de un organismo está en qué parte del código genético expresan. Este proceso se conoce bajo el nombre de diferenciación.

El ADN está compuesto de bases de Adenina, Citosina, Guanina y Timina. Estas bases se agrupan en lo que se conoce como genes, que son cada una de las secciones del ADN que codifican proteínas. En los genes se identifican principalmente dos secciones, la primera, llamada cis-region o región promotora, identifica tanto las proteínas que provocan la transcripción del gen como aquellas que limitan o impiden la transcripción del mismo. La segunda de las secciones del gen es la parte que identifica qué molécula de mRNA se genera al transcribirse el gen cuando se conectan todas las condiciones de la región promotora. Una molécula mRNA es una cadena que identifica los aminoácidos necesarios para que los orgánulos conocidos como ribosomas, traduzcan dicha cadena y generen la proteína que se encuentra codificada en la segunda parte del gen. Así, de manera simplificada, se puede decir que, cuando los genes se transcriben y después se traducen (en adelante al referirse a transcripción se referirá a ambos procesos), las células producen una proteína con la secuencia que se encuentra codificada en el gen. Este proceso puede llevar asociado distintos comportamientos por parte de la célula. Las cuatro principales acciones que pueden llevarse a cabo tras la

transcripción de un gen por parte de una célula son: almacenar la proteína generada, expulsarla al medio donde se encuentra la célula, activar la mitosis de la célula (división celular) o activar la apoptosis (muerte celular programada).

Cuando se examina la actividad celular se diferencian claramente dos procesos, uno constituye las actividades básicas de la morfogénesis, las cuales son todas las operaciones que tienen como resultado cambios internos de la célula. En estas se encuadran operaciones tales como la división de la célula, su muerte o el cambio de comportamiento en una especialización. En un segundo término se tienen las actividades de procesamiento de información proveniente, tanto del exterior, como generada por la propia célula. Estas dos actividades aunque se traten de manera diferenciada, tienen vínculos comunes entre ellas. Es decir, tanto las actividades de morfogénesis como las de procesamiento de información son el resultado de mensajes recibidos del exterior en forma de proteínas, las cuales se encuentran en el medio o bien son enviadas por otras células. Por tanto, a pesar de ser distintas, tienen procesos comunes.

Este es el que puede considerarse el comportamiento normal de los genes, pero existen particularidades dentro de lo que es el ADN. La primera es que existen genes llamados constitutivos, los cuales se transcriben siempre. Estos genes en lugar de necesitar proteínas para ser transcritos las necesitan para detener esa transcripción permanentemente o sólo durante un periodo de tiempo. La otra particularidad es la existencia de lo que se conoce como operón. El operón es un conjunto de genes que codifican una tarea más compleja. Cada operón tiene una parte que se corresponde con unas condiciones que deben ser satisfechas para que sus genes puedan expresarse. Así

cuando dichas condiciones son satisfechas, los genes que están contenidos dentro del operón pueden transcribirse. Un ejemplo clásico del operón es el operón lac, que se muestra en la Figura 2.2, y que es el responsable del procesamiento de la lactosa. El operón se activaría en el momento que la célula detectara la presencia de lactosa. En concreto la lactosa, uniéndose a la región promotora, permitiría la activación de los genes que se muestran en la Figura 2.2 (β-galactosidasa, β-galactopermeasa y β-galactósido transcetilasa), qué serían los responsable de generar las proteínas necesarias para convertir la lactosa en glucosa, para que pueda ser utilizada por las células.

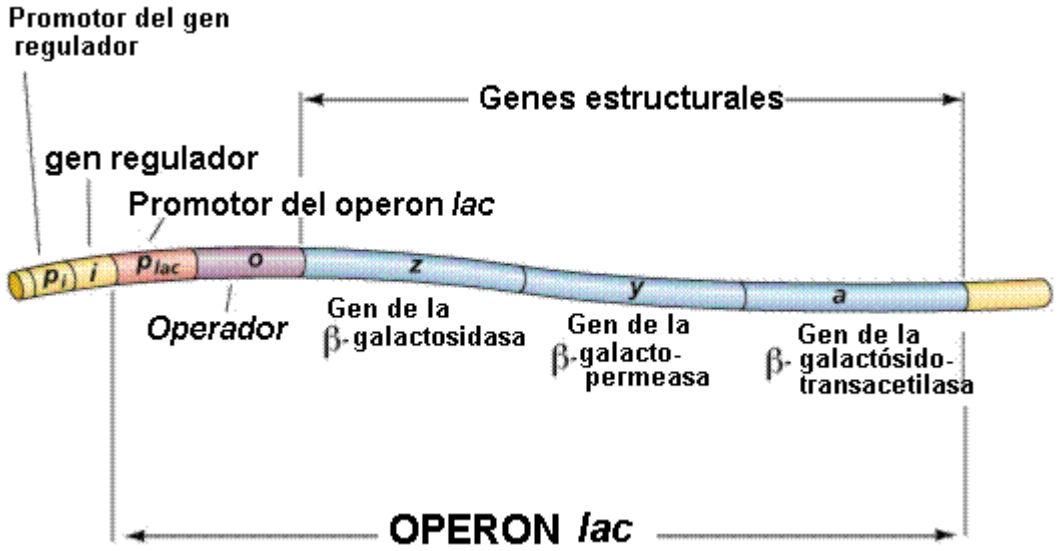

Figura 2.2 Esquema del operón lac

Todas estas reacciones de activación y transcripción se llevan a cabo, como ya se dijo antes, con la información que está contenida dentro de la célula. Esta información se encuentra codificada en proteínas que son el resultado tanto de la expresión del ADN de la propia célula como de mensajes recibidos de células vecinas. Estas proteínas se

encuentran en el citoplasma de la célula. Por citoplasma se suele entender todo aquello que no es núcleo pero que se encuentra en el interior de la membrana que separa la célula del entorno en el que se encuentra. En el citoplasma se encuentran distintos orgánulos que cumplen las distintas funciones que marca el ADN. Existen multitud de orgánulos pero todos ellos se pueden ver como los brazos ejecutores del ADN.

Así pues, las proteínas contenidas en una célula son lo que determina el comportamiento que esta va a expresar y, estas proteínas contenidas, vienen dadas por las proteínas generadas por la célula y las recibidas de sus vecinas, que están determinadas por la posición relativa que ocupa la célula en el sistema.

Finalmente, destacar que, con este sistema descrito de expresión del ADN y comunicación mediante proteínas, los organismos vivos son capaces de coordinar y hacer funcionar algunas de las estructuras más complejas que se conocen. Estando en todo momento el control de su desarrollo distribuido entre todos los elementos que componen el sistema.

Tras este resumen de las bases biológicas que rigen el funcionamiento de las células se puede concluir que los sistemas celulares cumplen todas las características propuestas por Nagl (Nagl, 2001) para definir qué es un sistema complejo, es decir:

- Un sistema complejo se compone de un gran número de elementos.
- Cada elemento solo responde a la información que tiene disponible localmente.
- Los elementos de un sistema complejo tienen interacciones dinámicas que cambian con el tiempo.

- Las interacciones entre elementos son de la forma que un elemento es influencia para y es influenciado por gran cantidad de elementos.
- Las interacciones entre elementos no son lineales.
- Las interacciones entre elementos son relativamente de corto alcance.
- Existen ciclos en estas interacciones entre elementos.
- Los sistemas complejos son normalmente sistemas abiertos.
- Los sistemas complejos operan bajo condiciones que distan del equilibrio.
- La complejidad surge como consecuencia de los patrones de interacción entre los elementos.
- Los sistemas complejos tienen memoria.

Un sistema celular cumple todos estos puntos, ya que las células son elementos relativamente simples y numerosos que componen el sistema. Además las células reaccionan en función de la información que tienen disponible. Dicha información proviene del medio, de la información procesada por las células vecinas cercanas o de la generada anteriormente por ella misma. Se puede observar, a su vez, que esa interacción con las células del vecindario cercano varía con el tiempo y es una de las claves que aporta complejidad. Esa interacción local, en lugar de un control global, unido la memoria interna de las células, que se almacena en forma de proteínas, identifica inequívocamente este sistema como complejo.

2.2 Red Reguladora de Genes

Desde que el ADN fuera descubierto por Watson y Crick en 1953, los investigadores han intentado entender cómo la expresión de este afecta al desarrollo del individuo.

Este ha sido un campo que ha generado multitud de teorías e investigaciones para desentrañar el misterio de cómo funciona el ADN. Se sabe que el ADN se estructura en base a genes como se comentó en el apartado 2.1. Cuando se comenzó a investigar se pensó que los genes tenían una función específica e influenciaban directamente el desarrollo y las operaciones de las células. Así, es frecuente oír expresiones del tipo "el gen que provoca…", pero realmente un gen, en lo que tiene influencia, es en otros genes y en la expresión de estos. Ahora se sabe que la mayor parte de los procesos dentro de las células lo que provocan es la expresión de conjuntos de proteínas expresadas por varios genes. Esos conjuntos de proteínas desencadenan combinaciones de procesos a nivel celular que son realmente los que generan el comportamiento observado finalmente. Así, la verdadera información para los distintos procesos no está tanto contenida en los genes como en las relaciones entre estos, que provocan una reacción en cascada.

Intentar profundizar en esta idea conlleva realizar modelos del mismo para intentar comprender y simplificar el grafo que forman las conexiones de los genes. Estudiar estas relaciones en el modelo biológico es sumamente complejo, es por ello que modelos y simulaciones son de gran ayuda para entender los principios de las redes reguladoras. Hay que tener en cuenta que los modelos no pueden capturar todos y cada uno de los detalles del organismo que tratan de representar. En su lugar tratan de capturar los elementos más importantes y simplificarlos para estudiar un punto concreto del comportamiento.

Una de las muchas alternativas para el estudio de las redes reguladoras es el uso de la teoría conocida como redes reguladoras de genes. En concreto, en este tipo de modelo,

cuando se trata de modelar los genes se identifican generalmente 2 partes (como se comentó en 2.1). La primera de las partes identifica los "productos" del gen, donde se identifica qué molécula se crea cuando el gen se expresa. La segunda de las partes de los genes es la "combinación de activación" o región promotora, que identifica los productos que se deben adherir al gen para que se produzca la molécula que tiene codificada en la parte de "productos". Cuando la "combinación de activación" tiene lugar entonces se activa el gen y se generan los productos que pueden formar parte de la "combinación de activación" otros genes.

Este esquema, que es sumamente potente, como se demostró en (Mjolness et. Al. 1991) tiene la misma capacidad de procesado que un algebra de Boole. Este hecho provoca que, en términos de computación, las redes reguladoras de genes sean capaces de resolver potencialmente cualquier problema.

Existen muchos ejemplos de distintos modelos para crear redes reguladoras de genes. La principal diferencia entre ellos es qué tipo de uso se le quiere dar. Por ejemplo, existen modelos más destinados a probar propiedades biológicas (Drennan & Beer 2006), mientras que, otros modelos, tratan de capturar únicamente características del modelo biológico para usarlos en otro tipo de problemas (Taylor 2004).

Los modelos más biológicos tienden a centrarse más en intentar modelar de la manera más fiel posible la activación de las distintas relaciones entre genes. Para ello se centran en modelar de manera matemática las activaciones, generalmente haciendo uso de ecuaciones diferenciales y funciones de transferencia (Schlitt & Brazma 2007).

En contrapunto, modelos más abstractos suelen ser los más apropiados para estudiar su aplicación a la resolución de distintos problemas. La manera más simple de representación de estos sería un grafo donde los genes son los nodos y el "producto" de cada uno determina las aristas con respecto a la "combinación de activación" de los otros. De esta manera se determinarían las influencias reguladoras. Comentar que la primera de estas aproximaciones es la que se puede encontrar en (Kaufmann 1969), donde se estudian las redes aleatorias booleanas que son el primer tipo de red reguladora de genes y, además, es el más sencillo. A partir de este, se han desarrollado otros modelos que tratan de incorporar interacciones más complejas como (Banzhaf 2003) donde la "combinación de activación" y el "producto" de los genes se encuentran mediante el uso de un algoritmo genético.

2.3 Algoritmos Genéticos

Los Algoritmos Genéticos son probablemente los representantes más característicos del conjunto de técnicas conocidas como computación evolutiva. Así, la computación evolutiva se podría definir en términos muy generales como una familia de modelos computacionales inspirados en la evolución.

Más formalmente el término de computación evolutiva se refiere al estudio de los fundamentos y aplicaciones de ciertas técnicas heurísticas basadas en los principios de la evolución natural (Alba & Tomassini, 2002).

Los Algoritmos Genéticos son métodos adaptativos, generalmente usados en problemas de búsqueda y optimización de parámetros, basados en la reproducción sexual y en el principio supervivencia del más apto (Goldberg, 1989b; Holland 1975).

Goldberg los definió, más formalmente, de la siguiente manera: "los Algoritmos Genéticos son algoritmos de búsqueda basados en la mecánica de selección natural y de la genética natural. Combinan la supervivencia del más apto entre estructuras de secuencias con un intercambio de información estructurado, aunque aleatorizado, para constituir así un algoritmo de búsqueda que tenga algo de las genialidades de las búsquedas humanas" (Goldberg, 1989b).

Para alcanzar la solución a un problema se parte de un conjunto inicial de individuos, llamado población, generado de manera aleatoria. Cada uno de estos individuos representa una posible solución al problema. Estos individuos evolucionarán tomando como base los esquemas propuestos por Darwin (Darwin, 1859) sobre la selección natural, y se adaptarán en mayor medida tras el paso de cada generación a la solución requerida.

Así, se podría tener una población de potenciales soluciones a un problema de las que se irían seleccionando las mejores hasta que se adaptasen perfectamente al medio, en este caso el problema a resolver.

El desarrollo de los Algoritmos Genéticos se debe, en gran medida, a John Holland, investigador de la Universidad de Michigan. A finales de la década de los 60 desarrolló una técnica que imitaba en su funcionamiento a la selección natural. Aunque originalmente esta técnica recibió el nombre de "planes reproductivos", a raíz de la

publicación en 1975 de su libro "Adaptation in Natural and Artificial Systems" (Holland, 1975) se conoce principalmente con el nombre de Algoritmos Genéticos. A grandes rasgos un Algoritmo Genético consiste en una población de soluciones codificadas de forma similar a cromosomas. Cada uno de estos cromosomas tendrá asociado un valor de bondad, ajuste o *fitness*, que cuantifica su validez como solución al problema. En función de este valor se le darán a cada individuo más o menos oportunidades de reproducción. Además, con cierta probabilidad se realizarán mutaciones de estos cromosomas.

Como se ha comentado anteriormente, la computación evolutiva tiene una fuerte base biológica. En sus orígenes los algoritmos evolutivos consistieron en copiar procesos que tienen lugar en la selección natural. Este último concepto había sido introducido, rodeado de mucha polémica, por Alfred Wallace (Wallace, 1855) y Charles Darwin (Darwin, 1859). A pesar de que aún hoy en día no todos los detalles de la evolución biológica son completamente conocidos, existen algunos hechos apoyados sobre una fuerte evidencia experimental:

- La evolución es un proceso que opera, más que sobre los propios organismos, sobre los cromosomas. Estos cromosomas pueden ser considerados como herramientas orgánicas que codifican la vida, o visto al revés, una criatura puede ser creada decodificando la información contenida en los cromosomas.
- La selección natural es el mecanismo que relaciona los cromosomas con la eficiencia respecto al medio, de la entidad que representan. Otorga a los

individuos más adaptados al medio un mayor número de oportunidades de reproducirse.

- Los procesos evolutivos tienen lugar durante la etapa de reproducción. Aunque existe una larga serie de mecanismos que afectan a la reproducción los más comunes son la mutación, causante de que los cromosomas de la descendencia sean diferentes a los de los padres, y el cruce o recombinación, que combina los cromosomas de los padres para producir la descendencia.

Sobre estos hechos se sustenta el funcionamiento de la Computación Evolutiva, en general, y de los Algoritmos Genéticos en particular.

Cualquier solución potencial a un problema puede ser presentada dando valores a una serie de parámetros. El conjunto de todos los parámetros (*genes* en la terminología de Algoritmos Genéticos) se codifica en una cadena de valores denominada *cromosoma*.

El conjunto de los parámetros representado por un cromosoma particular recibe el nombre de *genotipo*. El genotipo contiene la información necesaria para la construcción del organismo, es decir, la solución real al problema, denominada *fenotipo*. Por ejemplo, en términos biológicos, la información genética contenida en el ADN de un individuo sería el genotipo, mientras que la expresión de ese ADN (el propio individuo) sería el fenotipo.

Desde los primeros trabajos de John Holland la codificación suele hacerse mediante valores binarios. Se asigna un determinado número de bits a cada parámetro y se realiza una discretización de la variable representada por cada gen. El número de bits

asignados dependerá del grado de ajuste que se desee alcanzar. Evidentemente no todos los parámetros tienen por qué estar codificados con el mismo número de bits. Cada uno de los bits pertenecientes a un gen suele recibir el nombre de *alelo.*

```
Inicializar población actual aleatoriamente
MIENTRAS  no se cumpla el criterio de terminación
    crear población temporal vacía
    MIENTRAS  población temporal no llena
        seleccionar padres
        cruzar padres con probabilidad Pc
        SI se ha producido el cruce
            mutar uno de los descendientes con probabilidad Pm
            evaluar descendientes
            añadir descendientes a la población temporal
        SI NO
            añadir padres a la población temporal
        FIN SI
    FIN MIENTRAS
aumentar contador generaciones
establecer como nueva población actual la población temporal
FIN MIENTRAS
```

Figura 2.3 Pseudocódigo de Algoritmo Genético

Sin embargo, también existen representaciones que codifican directamente cada parámetro con un valor entero, real o en punto flotante (Wright, 1991). A pesar de que se acusa a estas representaciones de degradar el paralelismo implícito de las

representaciones binarias, permiten el desarrollo de operadores genéticos más específicos al campo de aplicación del Algoritmo Genético.

Los Algoritmos Genéticos trabajan sobre una población de individuos. Cada uno de ellos representa una posible solución al problema que se desea resolver. Todo individuo tiene asociado un ajuste de acuerdo a la bondad con respecto a la solución que representa (en la naturaleza el equivalente sería una medida de la eficiencia del individuo en la lucha por los recursos). El funcionamiento genérico de un Algoritmo Genético puede apreciarse en el pseudocódigo de la Figura 2.3.

Una generación se obtendrá a partir de los individuos genéticos pertenecientes a la anterior generación por medio de la aplicación de diferentes operadores de reproducción. Dentro de estos operadores genéticos pueden diferenciarse 2 tipos:

- Cruce. Se genera una descendencia a partir del mismo número de individuos (generalmente 2) de la generación anterior.
- Copia. Se trata de una reproducción de tipo asexual. Un determinado número de individuos pasa sin sufrir ninguna variación directamente a la siguiente generación.

Una vez generados los nuevos individuos se realiza la mutación con una probabilidad P_m. La probabilidad de mutación suele ser muy baja, por lo general entre el 0.5% y el 2%.

Se sale de este proceso cuando se alcanza alguno de los criterios de parada fijados. Los más usuales suelen ser:

- Los mejores individuos de la población representan soluciones suficientemente buenas para el problema que se desea resolver.
- La población ha convergido. Un gen ha convergido cuando el 95% de la población tiene el mismo valor para él, en el caso de trabajar con codificaciones binarias, o valores dentro de un rango especificado, en el caso de trabajar con otro tipo de codificaciones. Una vez que todos los genes alcanzan la convergencia se dice que la población ha convergido. Cuando esto ocurre la media de bondad de la población se aproxima a la bondad del mejor individuo.
- Se ha alcanzado el número de generaciones máximo especificado.

Sobre este algoritmo inicialmente propuesto por Holland se han definido numerosas variantes.

Quizás una de las más extendidas consiste en prescindir de la población temporal de manera que los operadores genéticos de cruce y mutación se aplican directamente sobre la población genética. Con esta variante el proceso de cruces varía ligeramente. Ahora no basta, en el caso de que el cruce se produzca, con insertar directamente la descendencia en la población.

Puesto que el número de individuos de la población se ha de mantener constante, antes de insertar la descendencia en la población se le ha de buscar una ubicación. Para ello existen diversas opciones, que se agrupan bajo el operador de remplazo.

2.3.1 Selección

Los algoritmos de selección serán los encargados de escoger qué individuos van a disponer de oportunidades de reproducirse y cuáles no. Puesto que se trata de imitar lo

que ocurre en la naturaleza, se ha de otorgar un mayor número de oportunidades de reproducción a los individuos más aptos. Por lo tanto, la selección de un individuo estará relacionada con su valor de ajuste.

No se debe, sin embargo, eliminar por completo las opciones de reproducción de los individuos menos aptos, pues en pocas generaciones la población se volvería homogénea (Cantú-Paz, 2000; Golberg & Deb, 1991).

Uno de los métodos más habituales y además el usado en esta tesis es la Selección por Ruleta. Propuesto por DeJong (De Jong, 1975), es posiblemente el método más utilizado desde los orígenes de los Algoritmos Genéticos.

A cada uno de los individuos de la población se le asigna una parte de una ruleta proporcional a su ajuste, de tal forma que la suma de todos los porcentajes sea la unidad. Los mejores individuos recibirán una porción de la ruleta mayor que la recibida por los peores. Generalmente la población está ordenada en base al ajuste por lo que las porciones más grandes se encuentran al inicio de la ruleta. Para seleccionar un individuo basta con generar un número aleatorio del intervalo [0..1] y devolver el individuo situado en esa posición de la ruleta. Esta posición se suele obtener recorriendo los individuos de la población y acumulando sus proporciones de ruleta hasta que la suma exceda el valor obtenido.

En mucha bibliografía se suele referenciar a este método con el nombre de Selección de Montecarlo.

2.3.2 Cruce

Una vez seleccionados los individuos, éstos son recombinados para producir la descendencia que se insertará en la siguiente generación (Booker et. al, 1997). Tal y como se ha indicado anteriormente el cruce es una estrategia de reproducción sexual.

Su importancia para la transición entre generaciones es elevada puesto que las tasas de cruce con las que se suele trabajar rondan el 90%.

Los diferentes métodos de cruce podrán operar de dos formas diferentes. Si se opta por una estrategia destructiva los descendientes se insertarán en la población temporal aunque sus padres tengan mejor ajuste (trabajando con una única población esta comparación se realizará con los individuos a reemplazar). Por el contrario utilizando una estrategia no destructiva la descendencia pasará a la siguiente generación únicamente si supera la bondad del ajuste de los padres (o de los individuos a reemplazar).

La idea principal del cruce se basa en que, si se toman dos individuos correctamente adaptados al medio y se obtiene una descendencia que comparta genes de ambos, existe la posibilidad de que los genes heredados sean precisamente los causantes de la bondad de los padres. Al compartir las características buenas de dos individuos, la descendencia, o al menos parte de ella, debería tener una bondad mayor que cada uno de los padres por separado. Si el cruce no agrupa las mejores características en uno de los hijos y la descendencia tiene un peor ajuste que los padres, eso no significa que se esté dando un paso atrás. Optando por una estrategia de cruce no destructiva se garantiza que pasen a la siguiente generación los mejores individuos.

Existen multitud de algoritmos de cruce. Sin embargo los más empleados y que se han usado en el desarrollo de esta tesis son los que se comentan en los siguientes apartados.

2.3.2.1 Cruce de 1 Punto

Es la más sencilla de las técnicas de cruce. Una vez seleccionados dos individuos se cortan sus cromosomas por un punto seleccionado aleatoriamente para generar dos segmentos diferenciados en cada uno de ellos: la cabeza y la cola. Se intercambian las colas entre los dos individuos para generar los nuevos descendientes. De esta manera ambos descendientes heredan información genética de los padres.

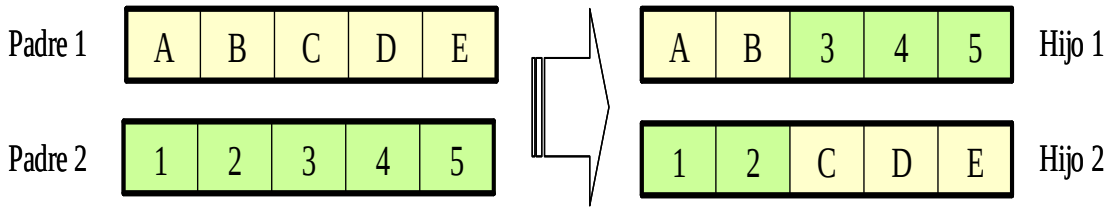

Figura 2.4: Esquema de cruce de un punto

En la Figura 2.4 se puede ver con claridad el proceso, al que en la bibliografía suele referirse con el nombre de SPX (Single Point Crossover).

2.3.2.2 Cruce de 2 Puntos

Se trata de una generalización del cruce de 1 punto. En vez de cortar por un único punto los cromosomas de los padres como en el caso anterior se realizan dos cortes.

Deberá tenerse en cuenta que ninguno de estos puntos de corte coincida con el extremo de los cromosomas para garantizar que se originen tres segmentos. Para generar la

descendencia se escoge el segmento central de uno de los padres y los segmentos laterales del otro padre.

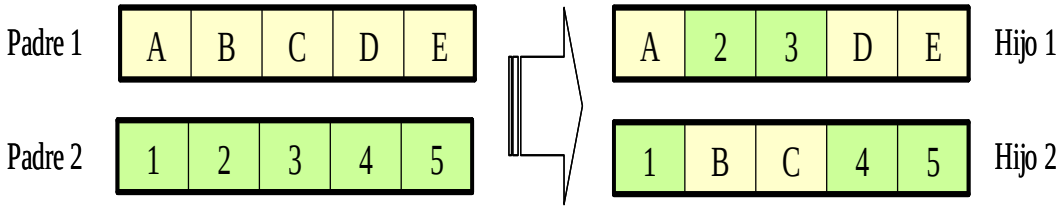

Figura 2.5: Esquema de cruce por dos puntos.

Generalmente se suele referir a este tipo de cruce con las siglas DPX (Double Point Crossover). En la Figura 2.5 se muestra un ejemplo de cruce por dos puntos.

2.3.3 Algoritmos de Reemplazo

Cuando en vez de trabajar con una población temporal se hace con una única población, sobre la que se realizan las selecciones e inserciones, deberá tenerse en cuenta que, para insertar un nuevo individuo, deberá de eliminarse previamente otro de la población. Existen diferentes métodos de reemplazo:

- **Aleatorio:** el nuevo individuo se inserta en un lugar cualquiera de la población.
- **Reemplazo de padres:** se obtiene espacio para la nueva descendencia liberando el espacio ocupado por los padres.
- **Reemplazo de similares:** una vez obtenido el ajuste de la descendencia se selecciona un grupo de individuos (entre seis y diez) de la población con un ajuste similar. Se reemplazan aleatoriamente los que sean necesarios.

- **Reemplazo de los peores:** de entre un porcentaje de los peores individuos de la población se seleccionan aleatoriamente los necesarios para dejar sitio a la descendencia.

Además, en todos los casos se puede realizar una estrategia de remplazo elitista, es decir, los mejores individuos de la población anterior pasan a la siguiente sin modificarse. Esto sirve para conservar la mejor solución obtenida.

2.3.4 Copia

La copia es la otra estrategia reproductiva para la obtención de una nueva generación a partir de la anterior. A diferencia del cruce, se trata de una estrategia de reproducción asexual. Consiste simplemente en la copia de un individuo en la nueva generación.

El porcentaje de copias de una generación a la siguiente es relativamente reducido, pues, en caso contrario, se corre el riesgo de una convergencia prematura de la población hacia los individuos copiados. De esta manera, el tamaño efectivo de la población se reduciría notablemente y la búsqueda en el espacio del problema se focalizaría en el entorno de los citados individuos.

Lo que generalmente se suele hacer es seleccionar dos individuos para el cruce y, si éste finalmente no tiene lugar, se insertan en la siguiente generación los individuos seleccionados.

2.3.5 Mutación

La mutación de un individuo provoca que alguno de sus genes, generalmente uno solo, varíe su valor de forma aleatoria.

Aunque se pueden seleccionar los individuos directamente de la población actual y mutarlos antes de introducirlos en la nueva población, la mutación se suele utilizar de manera conjunta con el operador de cruce. Así, en primer lugar se seleccionan los individuos de la población para realizar el cruce. Si el cruce tiene éxito entonces uno de los descendientes, o ambos, se muta con cierta probabilidad P_m. Se imita de esta manera el comportamiento que se da en la naturaleza, pues cuando se genera la descendencia siempre se produce algún tipo de error, por lo general sin mayor trascendencia, en el paso de la carga genética de padres a hijos.

La probabilidad de mutación es muy baja, como ya se ha comentado. Esto se debe sobre todo a que los individuos suelen tener un ajuste menor después de mutados. Sin embargo se realizan mutaciones para garantizar que ningún punto del espacio de búsqueda tenga una probabilidad nula de ser examinado.

Aunque no es lo más común, existen implementaciones de Algoritmos Genéticos en las que no todos los individuos tienen los cromosomas de la misma longitud. Esto implica que no todos ellos codifican el mismo conjunto de variables. En este caso existen mutaciones adicionales como puede ser la inclusión de un nuevo gen o la eliminación de uno ya existente.

2.3.6 Evaluación

Para el correcto funcionamiento de un Algoritmo Genético debe de proporcionarse un método que indique si los individuos de la población representan o no buenas soluciones al problema planteado. Por lo tanto para cada tipo de problema que se

desee resolver deberá derivarse un nuevo método, al igual que ocurrirá con la propia codificación de los individuos.

De esto se encarga la función de evaluación, que establece una medida numérica de la bondad de una solución. Esta medida recibe el nombre de ajuste o fitness. En la naturaleza el ajuste (o adecuación) de un individuo puede considerarse como la probabilidad de que ese individuo sobreviva hasta la edad de reproducción y se reproduzca.

En el mundo de los Algoritmos Genéticos se empleará esta medición para controlar la aplicación de los operadores genéticos. Es decir, permitirá controlar el número de selecciones, cruces, copias y mutaciones llevadas a cabo.

La aproximación más común consiste en crear explícitamente una medida de ajuste para cada individuo de la población. A cada uno de los individuos se les asigna un valor de ajuste escalar por medio de un procedimiento de evaluación bien definido. Tal y como se ha comentado, este procedimiento de evaluación será específico del dominio del problema en el que se aplica el Algoritmo Genético. También puede calcularse el ajuste mediante una manera 'co-evolutiva'. Por ejemplo, el ajuste de una estrategia de juego puede ser determinado mediante la aplicación de la estrategia codificada por un individuo contra la población entera (o en su defecto una muestra) de estrategias de oposición.

2.4 Embriogénesis Artificial

Han sido muchos los nombres por los que se ha dado a conocer la Embriogénesis Artificial (Stanley & Miikkulainen, 2003). Sistemas de generación y desarrollo (Stanley 2008) , embriología computacional (Bentley & Kumar 1999), codificación celular (Gruau 1994) o morfogénesis (Jakobi 1995) son sólo algunos de ellos. Todos estos nombres hacen referencia al mismo tipo de sistemas que, inspirándose en las células del cuerpo humano, tratan de conseguir mantener características como la autoorganización, la autoreparación, la tolerancia a fallos, etc. que están presentes en el modelo biológico.

Esta falta de unidad, incluso en la denominación del área de conocimiento, se debe a que los primeros trabajos se encuentran salpicados a lo largo de muchos años. Los orígenes del área se pueden establecer en las ideas enunciadas por Turing (Turing, 1952) y Lindenmayer (Lindenmayer, 1968) en las décadas de los 50 y 60. Esta tendencia cambia en los últimos años en los que ha empezado a haber un flujo continuo de trabajos sobre el tema. Un hecho que remarca el cambio de tendencia es la aparición de una metodología para la clasificación de los trabajos de esta área (Stanley & Miikkulainen, 2003).

2.4.1 Metodología de Clasificación

Según la metodología para la clasificación desarrollada por Stanley y Miikkulainen (Stanley & Miikkulainen, 2003), los trabajos desarrollados en la Embryogénesis Artificial se pueden clasificar en dos ramas bien diferenciadas: las aproximaciones gramaticales y las aproximaciones químicas. Estas dos ramas se explicarán en detalle, así como una descripción de los trabajos más importantes, en las secciones 3.1 y 3.2.

Además, esta metodología para la clasificación, establece cinco parámetros para clasificar cualquier sistema dentro de lo es la Embriogénesis Artificial. Los parámetros para la clasificación son: Ordenación del sistema, Selección de objetivos, Variabilidad temporal, Canalización y Complejidad de soluciones.

2.4.1.1 Ordenación del Sistema

El concepto de ordenación del sistema hace referencia a cómo una célula determina su posición final, sus conexiones o la función que va a desempeñar en el sistema global. En biología este hecho se determina por las distintas señales que recibe la célula. Las células precursoras, que constituyen el embrión, reciben señales externas que determinan las especializaciones de las células que constituyen el sistema. Por ejemplo, en un embrión de mamífero esto está relacionado con el gradiente de proteínas que forma el eje antero-posterior, el cual determina la simetría del futuro individuo (Gans & Northcut 1983; Lall S. & Patel N. 2001).

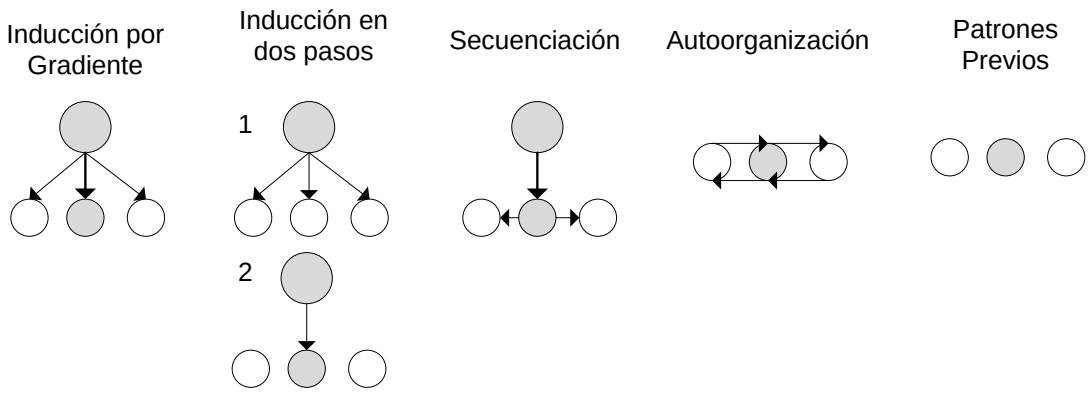

Figura 2.6: Estrategias de organización para sistemas

Se pueden identificar cinco estrategias básicas a la hora de la ordenación del sistema (Figura 2.6):

- **Inducción con gradiente**. En esta estrategia un elemento externo libera una señal en forma de gradiente y según la potencia de dicha señal el sistema asume una información u otra.
- **Inducción en dos pasos.** En esta un elemento organizador externo libera una señal que reciben todas las células y en un segundo momento libera otra señal que sólo reciben las células más cercanas a la señal que asumen un tipo distinto al resto.
- **Secuenciación.** En esta un elemento externo genera una señal que recibe una célula. Esta célula asume un tipo y a continuación transmite una segunda señal a las adyacentes para inducirlas a otro tipo distinto al que ella tiene.
- **Autoorganización.** En este sistema los componentes del sistema se intercambian mensajes a fin de negociar entre ellos los distintos roles necesarios.
- **Patrones previos.** Los roles de las células están determinados por el estado y la herencia recibida de sus progenitoras.

Aunque estas estrategias están presentes en todos los organismos vivos, los sistemas artificiales tienden a centrarse en uno o en una combinación de algunos. Por ejemplo, las aproximaciones gramaticales suelen usar más la estrategia de Patrones previos (Lipson & Pollack 2000), mientras que, los sistemas químicos suelen usar más las estrategias basadas en señales (Nolfi & Parisi 1991).

2.4.1.2 Selección de Objetivos

Con este término los autores hacen referencia a cómo las células determinan su posición dentro del sistema. La importancia de este hecho no reside tanto en posicionar la célula si no en cómo se relaciona con sus vecinas y en las comunicaciones con estas. Para esto existen dos modos básicos:

- **Identidad Especificada.** El objetivo de la célula está directamente codificado en el genoma.
- **Posición relativa.** En el genoma se encuentran codificadas sólo algunas indicaciones, como pueden ser el ángulo y la distancia a la que realizar una comunicación.

Existen múltiples variantes entre estas dos estrategias primarias como, por ejemplo, que lo que se encuentre codificado en un genoma sea la distancia del área a la que se quiere enviar una señal sin que sea una célula concreta sino más bien un grupo de células. El ejemplo más claro de estas estrategias mixtas se encuentra en la naturaleza, por ejemplo, en el sistema nervioso de las moscas conocidas como *Drosophila* existen muchas conexiones que no son fijas sino que emiten a las que estén cerca de su axón. Por otro lado, el sistema nervioso asociado al olfato de dichas moscas está conectado de manera específica con un grupo de neuronas (Marin et. Al 2002). Este tipo mixto es posible recrearlo en modelos asociados a la embriogénesis artificial, si bien, los investigadores deben encontrar el equilibrio entre coste de cómputo y la flexibilidad para especificar las entradas y las salidas por parte de los usuarios del modelo. Un ejemplo de esto se puede ver en los modelos que simulan el crecimiento de los axones y dendritas de un sistema celular. Este proceso puede simularse a bajo nivel, pero

tienen el problema de cómo los usuarios finales pueden especificar las entradas y las salidas del sistema. Este problema se suele dar en robótica y muchos investigadores lo eluden haciendo coevolucionar el cuerpo con la red neuronal que actúa como controlador (Bogard & Pfeifer 2001) si bien esto no siempre es posible, como por ejemplo en los robots prefabricados.

2.4.1.3 Variabilidad Temporal

Con este concepto la clasificación hace referencia a la variabilidad temporal y a los cambios en el desarrollo de un individuo a lo largo de su vida. Por tanto con este concepto se querría evaluar si el modelo presenta unas características similares a las del modelo natural. En la naturaleza los planes de desarrollo son asombrosamente flexibles. Existen múltiples ejemplos de organismos que desarrollan antes unas partes u otras. El ejemplo más evidente es el de los insectos que ejecutan una metamorfosis para pasar de larvas a adultos. Llegado un momento de desarrollo las funciones, apariencia y comportamiento del organismo cambian drásticamente. Existen trabajos que tratan de modelar este procedimiento como se pude ver en (Voss & Schaffer 1997). Este cambio de comportamiento con respecto del tiempo da a los modelos una flexibilidad en el desarrollo y poder explorar distintos caminos alternativos mientras el resultado final permanece invariante. Esta variabilidad en el desarrollo del individuo con el tiempo se ve en los sistemas que contienen una red reguladora de genes (Kauffman 1993) ya que la expresión del comportamiento codificado en cada gen requiere un tiempo hasta que se desenvuelve por completo.

2.4.1.4 Canalización

El término canalización hace referencia a la robustez de los modelos a las mutaciones (Waddington 1942). El modelo natural es muy robusto ante un número elevado de mutaciones, lo que le permite explorar nuevos caminos, mientras todo sigue funcionando de una manera similar, sin que se intuyan los cambios internos. Para ello existen tres mecanismos fundamentales que son:

- **Aleatoriedad de los eventos.** El rol de una célula no está predeterminado cuando la célula aparece, sino que es el resultado de que tras su aparición le ocurra algo. En este sentido no siempre teniendo toda la información del genotipo se puede saber cuál va a ser la función de una célula ya que depende de eventos externos. Como demostraron Kaneko y Furushawa en (Kaneko & Furushawa 1998) en organismos multicelulares las variaciones no lineales implican una mayor robustez del sistema.
- **Redistribución de los recursos.** Este término hace referencia al hecho que ocurre en biología de que los organismos son capaces de redistribuir los recursos que poseen. Así por ejemplo cuando a un insecto se le amputa una pata durante el desarrollo otra de sus extremidades se hace más grande para consumir el exceso de nutrientes (Nijhout & Emlen 1998).
- **Sobreproducción.** Con este concepto se hace referencia al hecho de que, las células, poseen buffers para gestionar posibles cambios en el comportamiento y mensajes de sus vecinas debidas a una mutación. Uno de esos sistemas de buffer es la sobreproducción de células que son necesarias así, ante cualquier cambio en el número de conexiones, las células puede solucionarlo con esta

sobre producción, o bien, si al final no son necesarias pueden ser eliminadas mediante la apoptosis (muerte celular programada).

2.4.1.5 Complejidad de las Soluciones

Con esta sección Stanley y Miikkulainen hacen referencia a que el sistema empieza desde un único elemento y se va complicando a medida que se desarrolla la solución. Esta complejidad creciente se consigue mediante mutaciones y combinaciones genéticas que van complicando las funcionalidades básicas del sistema. Para ese incremento de la complejidad, el mecanismo básico que utiliza la naturaleza es la duplicación de los genes para, posteriormente, diferenciar un grupo de ellos mientras se mantiene otro grupo que conserva la funcionalidad (Force et. Al 1999). En los sistemas de embriogénesis artificial esta duplicación implica el tener individuos de longitud variable. Este hecho puede provocar dos cosas:

- El resultado de los cruces del algoritmo genético, usado para la búsqueda, puede no ser válido, implicando incluso pérdida de información.
- Los conjuntos de genes que tienen un comportamiento dependiente entre ellos, tienen que ser duplicados todos para poder modificar una de las copias y que la otra se siga comportando de la manera habitual.

Es un hecho, el hombre tiene que controlar la ciencia y chequear ocasionalmente el avance de la tecnología

Thomas H. Huxley

Capítulo 3

Estado de la Cuestión

La tesis doctoral se puede encuadrar, en parte, dentro de lo que se conoce como Embriogénesis Artificial. Este capítulo de la memoria estará dedicado a una revisión de los trabajos más sobresalientes en el campo, prestando especial atención a los sistemas que se han encuadrado en la llamada aproximación química de la Embriogénesis Artificial.

Como se dijo en el punto 2.4, la Embriogénesis Artificial puede dividirse en dos grandes ramas: la aproximación gramatical y la aproximación química. Los trabajos más importantes de cada una de ellas se detallaran en los apartados 3.1 y 3.2.

3.1 Aproximaciones Gramaticales

Bajo el subtítulo de aproximaciones gramaticales, están todos aquellos sistemas que realizan un enfoque *top-down* en su aproximación a los sistemas celulares. Esta rama puede establecer su origen en las ideas enunciadas por Lindenmayer en 1968 (Lindenmayer, 1968), en donde se establecen las bases de los conocidos como L-systems. Estos sistemas plantean que cualquier objeto natural complejo puede ser descrito mediante una sustitución iterativa de partes simples por otras más complejas. En concreto este comportamiento está inspirado en el crecimiento de las plantas, las cuales empiezan con una estructura muy sencilla y, a partir de esa, la van complicando con ramificaciones.

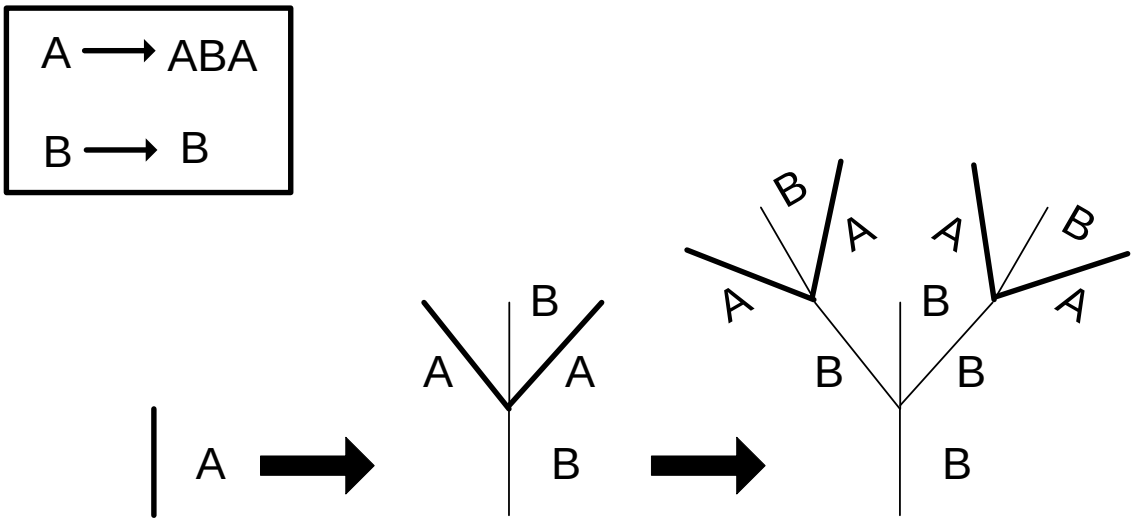

Figura 3.1: Ejemplo de L-system

El funcionamiento básico sería tener un conjunto de reglas de producción y un individuo. En dicho individuo se van sustituyendo las partes que coinciden con las condiciones de las reglas por los consecuentes. Por lo que, los L-systems, pueden ser vistos como una codificación indirecta de la morfología de ese individuo actuando de

una manera similar a como funcionan los fractales (ver Figura 3.1). Esa gramática o conjunto de reglas de producción puede incrementar su complejidad incluyendo parámetros en las reglas de producción (Lindenmayer 1968). Así, una misma regla de producción podría dar comportamientos completamente distintos según los valores de estos parámetros. Además, existen trabajos en los que las reglas están afectadas por el entorno y asignan una probabilidad (Prusinkiewicz & Lindenmayer 1990) a los distintos comportamientos que puede tener una regla. Finalmente, destacar el curioso hecho de que los L-systems no fueron pensados en un principio para ser evolucionados. Esta técnica se pensó para representar un hecho natural que parecía tener comportamiento fractal, pero son muchos los ejemplos de trabajos que, partiendo de un primer elemento, han evolucionado un conjunto de reglas para conseguir unos objetivos como en (Hornby & Pollack 2002).

El foco de atención de los trabajos incluidos bajo esta rama ha sido, generalmente, los sistemas Neuroevolutivos. Los sistemas Neuroevolutivos son sistemas destinados a desarrollar Redes Neuronales Artificiales. La bibliografía sobre la neuroevolución es muy extensa siendo utilizados diversos métodos para llevarla a cabo. Ejemplos como el uso de Algoritmos Genéticos, Programación Genética (Rivero 2007) o el sistema NEAT (Stanley et al 2009) son de sobra conocidos para los investigadores del campo. Los trabajos en los que se centrará la atención en esta sección son los que utilizan los L-systems u otra técnica gramatical para este hecho. En este sentido cabe destacar en primer lugar el trabajo presentado por Kitano en 1990 (Kitano 1990). Este es el primer trabajo constatado de neuroevolución, en el cual Kitano evoluciona las matrices de

conexión de las Redes de Neuronas Artificiales aplicando para ello un conjunto de reglas.

Dentro de lo que es la aproximación gramatical otro de los trabajos más importantes es el llevado a cabo por los investigadores Horby y Pollack (Hornby & Pollack 2001). En este trabajo, los investigadores utilizan un L-system con el fin de desarrollar la estructura del cuerpo de un robot en un entorno 3D simulado. Además, también usan el L-system para desarrollar una Red de Neuronas Artificiales que actuará como controlador del cuerpo en ese entorno simulado.

Finalmente, dentro de lo que son las aproximaciones gramaticales, destacar el trabajo efectuado por Gruau (Gruau 1994). El sistema propuesto en este trabajo almacena el desarrollo de una red neuronal en forma de árbol, donde este desarrollo se lleva a cabo desde un único elemento.

Mencionar que los trabajos en esta rama se han ralentizado bastante como se puede comprobar por la fecha de los últimos trabajos relevantes. Este hecho se debe a la dificultad de adaptar los modelos más abstractos a distintos problemas ya que, generalmente, tienen que ser diseñados de manera específica para un problema concreto.

3.2 Aproximaciones Químicas

El término aproximación química hace referencia a todos aquellos trabajos que, tomando como base los predicados de Turing (Turing, 1952) del año 1952, realizan una aproximación *bottom-up* del comportamiento de las células.

En este conjunto de trabajos el primero en ser nombrado, antes de hacer ninguna diferencia, debe ser el presentado por Kauffman en 1969 (Kauffman 1969). En este trabajo estudia las bases de expresión de los genes que componen el ADN y en él se enuncia la base de la teoría de las Redes Reguladoras de Genes (véase sección 2.2). Teoría que sirve como base para la investigación de cómo la expresión de los distintos genes está relacionada entre si, dando como resultado un comportamiento complejo, de estos es especialmente relevante el trabajo llevado a cabo por Mjolness y colaboradores en 1991 (Mjolness et al. 1991).

Una vez establecida la base, se puede decir que los trabajos existentes dentro de lo que es la aproximación química, se pueden dividir en dos grandes grupos según su objetivo.

El primero de estos grandes grupos estaría constituido por todos los trabajos que tienen como objetivo hacer modelos artificiales de lo que es la célula. Por tanto, lo que busca esta línea de trabajo es profundizar y mejorar el estudio de la célula. Estos modelos artificiales permitirían incrementar el conocimiento sobre los procesos que se llevan a cabo dentro de ellas. Un ejemplo de este tipo de trabajo es (Kitano 1994) presentado en 1994 y cuyo afán era precisamente realizar un modelo de célula lo más cercano a la realidad para permitir un mejor estudio de los procesos internos de la misma. Otros trabajos, por el contrario, en lugar de crear modelos para el estudio de la célula en su totalidad, se centran en modelar el metabolismo o alguno de sus procesos internos (Kaneko, 2006). Finalmente, comentar que algunos de estos modelos han servido para comprobar determinas reacciones que los neurocientíficos sospechaban, como ocurre en (Perea & Araque, 2007), donde un modelo computacional sirvió para

probar la influencia de las células gliales en la sinapsis nerviosa. Comentar que este trabajo sirvió posteriormente para desarrollar una mejora dentro de las redes de neuronas artificiales que es conocido como redes neurogliales artificiales (Porto et. al., 2007).

El segundo de los grandes grupos de trabajos, dentro de las aproximaciones químicas, son aquellos trabajos cuyo objetivo dista mucho de querer profundizar en el estudio de la célula. El objetivo de este conjunto de trabajos es adaptar el modelo de proceso de datos presente en todas las células para aplicarlo a otros problemas, pero con una técnica que, como se comentó anteriormente, tenga alguna de las propiedades de autoorganización, autoreparación, procesamiento paralelo, etc.

Algunos ejemplos de los problemas que se han intentado resolver con este tipo de técnicas son: la generación de distintas formas (Kumar & Bentley 2003), el guiado de robots (Kumar 2004a) o el desarrollo de hardware evolutivo (Kuyucu et al. 2009).

Este subconjunto de trabajos cuenta con muchos y buenos ejemplos de sistemas que han sido desarrollados en los últimos años. Autores como Miller, Tufte, Eggenberger o Beer son de sobra conocidos. En los siguientes párrafos se analizarán los modelos más representativos o de un mayor interés.

En 1996, Eggenberber (Eggenberger 1996) presentó un primer modelo que quería ser usado para la configuración de los sensores y el controlador de un agente. Es decir, el modelo artificial es utilizado para el desarrollo de una Red de Neuronas Artificiales. Este modelo se inspiraba en el crecimiento y funcionamiento de las células biológicas, ya que contaba con un ADN del cual se iban activando partes a medida que el sistema

se desarrollaba. Como mayores aportaciones de este trabajo, se puede citar el uso de los conceptos de diferenciación y movimiento, que son dos conceptos relativamente poco explotados dentro de lo que son los trabajos de embriogénesis artificial. Así, las células que componen el sistema artificial, actuarían como neuronas que se desplazan dentro de un espacio y las cuales van adquiriendo distintas funcionalidades a medida que se especializan. Para dicha especialización, las células guardarían el linaje del cual provenían y activaban, por tanto, una parte u otra del comportamiento codificado en el ADN. Este trabajo tiene dos puntos débiles, el primero de ellos es la dificultad de obtener una medida de bondad sobre el comportamiento de las células. Hecho que se plasma en la falta de resultados numéricos del mismo. Por otro lado, el diseño tan dependiente del problema a tratar por el modelo celular, hace que este sea difícil de generalizar para otro tipo de problemas.

También en 1996, Frank Dellaert y Randall D. Beer presentaron un trabajo, con dos modelos distintos, que tenía su base en el funcionamiento de las células. El objetivo de estos modelos era conseguir un sistema para el desarrollo de redes neuronales que autoorganizasen su estructura. El primero de los dos modelos es el que los autores denominan como "más complejo". Este modelo representa el ADN como un conjunto de funciones, con una o dos variables, que generan nueva información que se almacena en un citoplasma. El segundo de los modelos tiene un comportamiento similar, pero en lugar de utilizar un conjunto de funciones como ADN utiliza una red aleatoria de Boole (Dellaert & Beer 1996). El principal problema de este trabajo fue que el modelo estaba diseñado de tal manera que la configuración de los ADNs se realizaba a mano.

Este hecho imposibilita el desarrollo de estructuras más complejas que unos pocos elementos y que realmente sea aplicable a problemas del mundo real.

Uno de los problemas más recurrentes dentro de lo que son los sistemas de embriogénesis artificial es la generación de formas que tengan distintas propiedades. En este ámbito el trabajo presentado en 2003 por Julian Miller (Miller 2003) es una referencia. En este trabajo se puede encontrar la primera referencia al problema conocido como problema de la bandera francesa. Este modelo se basa en el crecimiento desde un primer elemento hasta la forma deseada. En este modelo cada elemento tiene como entradas las salidas de sus vecinos y como controlador, a modo de ADN, un programa desarrollado con programación genética. El gran valor de este trabajo reside en que, aparte de la prueba de la bandera de Francia, presenta pruebas de que este tipo de modelos pueden autorepararse. El principal problema de este trabajo reside en que los programas generados por el sistema de programación genética, no siempre son inteligibles. Además las soluciones también presentan problemas de estabilidad al llegar al punto deseado ya que no son capaces de detener su desarrollo.

En la misma línea de trabajo se pueden encontrar otros trabajos como (Federici 2004; Doursat 2008; Andersen et. Al. 2009), en los que los autores buscan desarrollar patrones con las células y que estos patrones presenten determinadas características.

Este último tipo de trabajos tiene su utilidad en distintos campos. Como se apuntó en esta misma sección, las aproximaciones químicas no sólo se utilizan para estudiar en profundidad la célula. Existen otros campos en los que este desarrollo de patrones con cierta funcionalidad es muy útil, como por ejemplo en el desarrollo de hardware

evolutivo o en el guiado de robots. Un buen ejemplo para el desarrollo de hardware evolutivo se puede encontrar en el trabajo de Tufte y Pauline (Tufte & Haddow 2005) y más recientemente, en (Tufte & Haddow 2008). En estos trabajos se trata de aplicar distintas propiedades de las células en el desarrollo de nuevo hardware que añada nuevo valor al hardware desarrollado por este método. Con respecto al guiado de robots el mejor ejemplo es el desarrollado por Sanjeev Kumar en (Kumar 2004a) en donde hace uso del modelo celular desarrollado por él mismo en (Kumar 2004b), que tiene como peculiaridad el uso de la teoría de las proteínas fractales para la comunicación entre las células desarrollada por Peter Bentley en (Bentley 1999).

Finalmente cabe destacar los esfuerzos que existen últimamente para extender este proceso de morfogénesis al entorno 3D. Cabe destacar entre estos trabajos el trabajo desarrollado por Fontana (Fontana 2009), donde el autor adapta un modelo en 2D al 3D para generar formas complejas a partir de primitivas. También debe mencionarse (Joachimczak & Wrotel 2009), que es un modelo pensado directamente en 3D. En este trabajo los autores no usan una cuadrícula para ordenar el sistema, como es habitual, si no que utilizan la física de esferas, este hecho tiende a hacer el sistema muy pesado en términos de eficiencia.

3.3 Consideraciones

Como bien argumenta Harding y Banzhaf en (Harding & Banzhaf 2008) y como ejemplifica el repaso llevado a cabo en estas páginas, el campo del desarrollo de sistemas de manera autoorganizativa está comenzando. Este hecho hace que muchas veces no sean evidentes las ventajas que pueden proporcionar este tipo de técnicas. Si

bien, la idea general de lo que pueden aportar estas técnicas es clara, su desarrollo y utilización no es evidente ni está extendido en la comunidad investigadora.

Tras el repaso llevado a cabo en este capítulo se hace evidente que, los modelos desarrollados anteriormente estaban focalizados en su aplicación en tareas muy específicas. Existen ejemplos que o bien para desarrollar algún tipo de red de neuronas artificiales, o bien para generar una forma.

Por tanto, en este campo faltan modelos celulares que sean más generales y que se puedan aplicar en distintos tipos de problemas. Además, se hace evidente tras repasar estos trabajos que aún no se han explorado todas las capacidades de procesado de información de este tipo de modelos. La aplicación de este tipo de sistemas en el campo de procesado de la información puede suponer, obviamente, una nueva vía de investigación.

La ciencia avanza a pasos no a saltos

Thomas B. MacAulay

Capítulo 4

Hipótesis

Como se apuntó en el Capítulo 3, el campo de la Embriogénesis Artificial está cobrando mayor importancia día a día. Este crecimiento se debe a la ya comentada caducidad del paradigma clásico de desarrollo para los sistemas. Según este paradigma, es el diseñador el que debe determinar e implementar todos y cada uno de los aspectos del comportamiento de un sistema. En sistemas muy grandes esto provoca grandes retrasos y problemas de mantenimiento (Kephart & Chess 2003). Así, una vía por la que esto se puede solucionar es el desarrollo de los sistemas generadores y desarrolladores (Stanley 2008). En este sentido, cada paso en este campo

puede ser un avance para solventar el problema que plantean los sistemas tradicionales.

El objetivo anterior es la utopía en la que se inspira el desarrollo de esta tesis, tratando de ser otro paso hacia ese ideal.

Como también se comentó en el Capítulo 3, los sistemas que se encuadran dentro de la embriogénesis artificial, tienden a estar muy centrados en un tipo de problema determinado. Además, como también se decía, en el citado capítulo, los sistemas existentes no han hecho sino que rascar la superficie de las posibilidades de estos sistemas.

Tras observar que las células no nerviosas de un organismo son capaces de autoorganizarse y además también tienen que procesar información en forma de señales químicas de manera coordinada, la hipótesis de esta tesis se puede establecer como: **"El modelo de procesado de información presente en las células no nerviosas se puede adaptar en un modelo artificial simplificado para efectuar reconocimiento de patrones, conservando las capacidades más importantes del modelo biológico como autoorganización o tolerancia a fallos"**.

Por tanto, los objetivos concretos que cubre esta tesis doctoral son:

- Estudiar el modelo biológico de célula y diseñar soluciones artificiales que mantengan las características deseables de estas.

- Analizar los trabajos previos y comprobar el comportamiento de los modelos artificiales desarrollados en esta tesis en pruebas clásicas de la Embriología Artificial para comprobar propiedades como la autoorganización.
- Aplicación de los modelos propuestos para problemas de reconocimiento de patrones.

El primero de los objetivos incluye el estudio de la bibliografía previa dentro de lo que es la embriogénesis artificial. Así, se pretende unificar muchas de las propuestas realizadas y diseñar otras nuevas, a fin de aprovechar experiencias previas y crear un modelo lo más generalista posible que tenga un amplio rango de aplicación. El término "rango de aplicación" se refiere a que sea aplicable a más tipos de problemas de lo que lo eran los modelos que se han visto previamente en la bibliografía del tema.

Es por esto que el segundo de los objetivos de la tesis es probar que las soluciones propuestas son aplicables en algunos de los problemas existentes en la bibliografía, como puede ser la generación de formas, en concreto, de un subconjunto de formas sencillas.

Finalmente, y como ya se ha apuntado varias veces en distintos capítulos de esta tesis, el potencial de los modelos celulares artificiales para procesado de información es, en principio, enorme. Es por ello que uno de los principales objetivos de esta tesis es explorar las capacidades del modelo en este sentido y determinar si los modelos celulares artificiales son aplicables en este tipo de problemas.

La respuesta a cualquier pregunta se encuentra en la naturaleza.
Sólo hay que saber cómo formular la pregunta y reconocer la respuesta
Bryce Courtney

Capítulo 5

Modelo propuesto

Este capítulo de la tesis abordará el primero de los objetivos citados anteriormente en el Capítulo 4. En concreto este capítulo contendrá la descripción pormenorizada del modelo teórico propuesto, el cual surge tanto del estudio del modelo biológico, como del análisis de los trabajos previos realizados en el área de la embriogénesis artificial, en concreto, en la rama denominada como aproximación química.

5.1 Descripción General

Todas las células de un tejido realizan un procesado conjunto de las señales que reciben y de la información que poseen. Esa información y señales están codificadas en moléculas conocidas como proteínas. A su vez, cada célula posee una copia del ADN, que es el conjunto de instrucciones para procesar las proteínas. El ADN está formado por genes, estos son tramos del ADN en los cuales se identifican una serie de condiciones necesarias (proteínas) y el resultado de la presencia de esas condiciones. Dicho resultado se codifica a su vez en forma de proteína, la cual puede ser almacenada por la célula como resultado intermedio, expulsada al entorno mediante osmosis, provocar la división de la célula (mitosis) o la muerte programada de la misma (apoptosis).

El procesado descrito en el párrafo anterior se produce a nivel individual en cada una de las células de un tejido, pero las células a su vez ejecutan un procesado conjunto. Este procesado tiene como objetivo resolver situaciones que, individualmente, no son asumibles. Dicho procesado conjunto es el resultado de que cada una de las células ejecute su rol dentro de las funciones generales codificadas en el ADN. Ese rol, es el resultado tanto de su posición relativa respecto de las otras células, como de la información que recibe. Es más, se debe destacar que la asunción de los distintos roles por parte de las células se hace de manera autoorganizativa entre ellas, ya que el sistema carece de un control centralizado que ordene ese reparto de funciones.

Estos dos tipos de procesado, el realizado por la célula y el realizado por el conjunto de células, son los que se pretenden adaptar en el modelo artificial que se desarrolla en

este capítulo. Para ello, se han diseñado una serie de estructuras de las células a fin de retener esos dos comportamientos. En los siguientes apartados se irá describiendo cada una de esas adaptaciones pormenorizadamente. Dicha descripción tratará de ir construyendo el sistema desde el interior del modelo artificial de célula, empezando desde los elementos más sencillos para concluir las cuestiones relativas a la manera de comunicarse las células entre sí.

5.2 Proteína

Las proteínas constituyen el elemento básico de información dentro del sistema celular. Cada una de ellas representa un dato o procesado previo realizado por la célula o un mensaje recibido de alguna de sus vecinas. Por lo tanto, cada una de las proteínas contribuye para identificar un estado concreto para la célula.

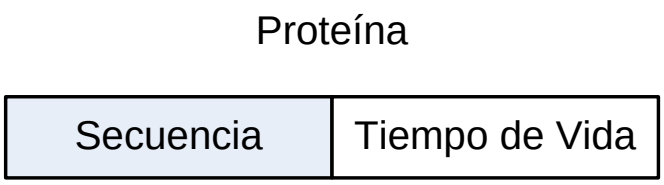

Figura 5.1: Esquema de la estructura de una proteína

En el sistema diseñado, las proteínas están compuestas de dos partes. La primera de las partes es una secuencia binaria que identifica la proteína en cuestión y la segunda es un campo que contiene un contador temporal (Figura 5.1). Este contador temporal funciona como el tiempo de vida de la proteína. Se establece con un valor y al llegar a cero la proteína es eliminada. Trata de simular el proceso de degradación de las proteínas si estas no son usadas. Este proceso es esencial a la hora procesar

información ya que determina cuánto tiempo van a ser válidos los datos y, por tanto, el tiempo de memoria que va a tener una célula. El establecer una ventana temporal de memoria es importante ya que depende de está, que los datos estén el tiempo suficiente en el sistema para ser útiles, pero no demasiado como para hacer errar otros por usar información obsoleta.

5.3 Citoplasma

Como se dijo, el conjunto de proteínas que contiene una célula constituyen el estado en el que se encuentra la misma. El encargado de gestionar esa información, tanto en el modelo biológico, como en el artificial aquí propuesto, es el citoplasma. Además de almacenar las proteínas el citoplasma también debe actualizar su tiempo de vida cuando es necesario, y, llegado el caso en que este alcance el 0, proceder a su borrado.

Finalmente, el citoplasma del sistema celular artificial propuesto cuenta con una última responsabilidad. Como se explicará más en profundidad en el punto 5.4, los genes necesitan proteínas para realizar su transcripción además de que estas deben de estar en una cierta concentración. El citoplasma es el encargado de proporcionarle al ADN las proteínas que necesita para la transcripción de los genes, siempre que estos cumplan las restricciones de concentración. Es el citoplasma el que se encarga de calcular esas concentraciones y transmitírselas al ADN para que este decida qué genes pueden ser activados.

5.4 Gen

En el modelo biológico, los genes son subsecciones del ADN que ordenan la transcripción de una proteína por parte de la célula cuando están presentes una serie de proteínas que se acoplan a esta sección del ADN.

De la misma manera, el gen del sistema artificial puede ser visto como una regla de reescritura basada en el estado de la célula y, más concretamente, en el contenido del citoplasma. Por tanto, como para cualquier regla, se deben cumplir una serie de condiciones a fin de que el gen ejecute el comportamiento que tiene almacenado.

La estructura de los genes, en el sistema celular propuesto, es un *string* de bits en el que se pueden identificar dos partes principalmente (Fig. 5.2):

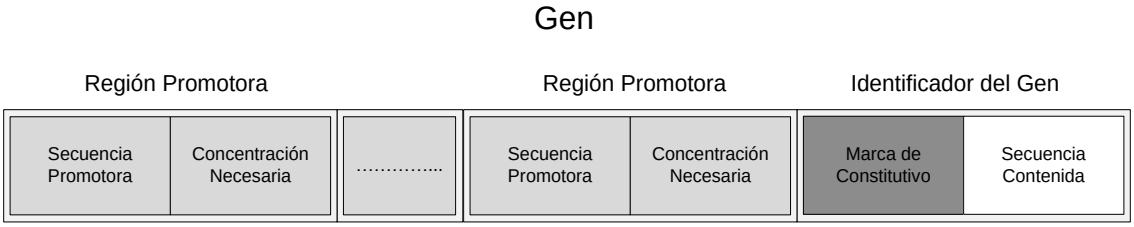

Figura 5.2: Esquema general de un gen

- **Regiones Promotoras**. El conjunto de regiones promotoras son la parte que adapta la conocida como *región cis* de los genes biológicos. Cada una de estas regiones promotoras identifica una proteína necesaria para poder activar el gen en cuestión. Es posible que aparezca más de una para un solo gen e identifica las distintas condiciones que se tienen que cumplir simultáneamente para que se active el gen. Cada una de las regiones promotoras tiene a su vez dos partes:

  - **Secuencia Promotora.** Esta subparte contiene el identificador binario de la proteína requerida.
  - **Concentración necesaria.** En esta subparte del gen se identifica la concentración mínima que debe haber en el citoplasma para que se transcriba el gen.

- **Identificador del Gen.** En esta parte del gen se codifica el consecuente de la regla. Así, cuando se satisfacen las condiciones enumeradas en la región promotora, esta parte marca el comportamiento que va a tener la activación del gen. Esta parte del gen está compuesta de dos subpartes que son:
  - **Marca de Constitutivo.** Esta marca identifica si el gen es constitutivo o no, lo cual cambia el comportamiento del mismo como se explica en el punto 5.4.1.
  - **Secuencia Contenida**. Esta subparte contiene la secuencia de la proteína que se va a generar cuando el gen se activa. Esta secuencia puede producir una proteína sin más, o bien tener algún comportamiento asociado que debe ejecutar la célula cómo la división celular (mitosis) o la muerte celular programada (apoptosis).

Para la activación de un gen es preciso que se cumplan todas las condiciones marcadas por la región promotora. En la Figura 5.3 se puede ver un ejemplo de activación de un gen. En dicha figura se ve que, cuando están presentes en el citoplasma el conjunto de proteínas que están marcadas en la región promotora, representadas en la figura por los rombos grises, el gen se activa y entonces produce una nueva proteína con la secuencia que tiene codificada en su interior.

Remarcar que las proteínas, que activan el gen, tienen que estar al menos en la cantidad identificada en la subsección *concentración necesaria* correspondiente. Cuando se cumplen todas las condiciones entonces, y sólo entonces, es cuando se genera una proteína con la secuencia contenida dentro de la parte *identificadora del gen*.

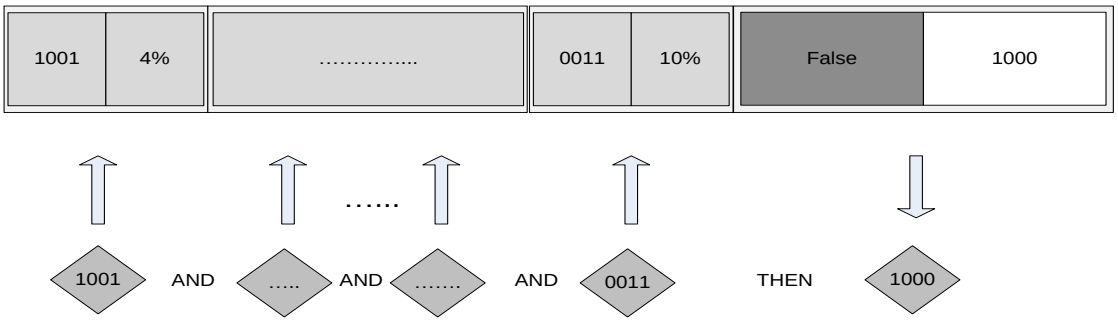

Figura 5.3: Ejemplo de activación de un gen.

Hay que resaltar que tras realizar unas pruebas iniciales, se optó por la inclusión de un método que flexibilizará la satisfacción de las condiciones. Es decir, no es necesario que la proteína con la secuencia exacta este presente. En la naturaleza, secuencias similares a las de las condiciones pueden provocar la transcripción de un gen si se encuentra en una cantidad elevada. Este hecho es el que hace que el modelo biológico sea flexible frente a ligeros cambios, ya que puede seguir funcionando aunque no sea eficientemente, mientras se adapta. Por ello, se ha articulado este mecanismo mediante la siguiente condición:

$$\text{Concentración de la Proteína } i >= (\text{Distancia Hamming} + 1) * \text{Concentración Necesaria} \quad (5.1)$$

Mediante esta condición se flexibiliza la posibilidad de que varias proteínas satisfagan una misma condición. Así, si alguna proteína cumple la condición enunciada por Eq. 5.1, se activaría el gen. En la condición, *Concentración de la Proteína i* se refiere a la concentración en el citoplasma de la proteína *i* que se está comprobando, *Distancia Hamming* es la distancia Hamming entre la secuencia de la proteína deseada y la proteína *i,* y, finalmente *Concentración Necesaria* es el valor del campo *Concentración Necesaria* de la región promotora que se está comprobando. Por lo tanto, la concentración necesaria para que una proteína *i* active una de las condiciones aumenta cuanto mayor es la diferencia de la secuencia de *i* con respecto de la identificada en la región promotora.

Así, si mediante el uso de la condición 5.1 todas las regiones promotoras son satisfechas, entonces, el gen es activado y se genera una proteína con la secuencia contenida en la subparte *identificador del gen*.

5.4.1 Gen Constitutivo

Lo explicado anteriormente es el comportamiento normal de un gen. Pero, en la naturaleza existen un conjunto de genes que se conocen como *genes constitutivos*. Estos genes tienen como particularidad que su comportamiento es exactamente el opuesto al de los genes ya comentados. Así, estos genes permanecen activos hasta el momento en el que se satisfacen todas sus condiciones de la *región promotora*, en cuyo caso dejan de estar activos por un periodo de tiempo. Este comportamiento en la naturaleza está asociado al metabolismo basal de las células, es decir, la base del funcionamiento, ya que, es necesario que existan algunos compuestos previos para que la célula funcione.

Así pues, el comportamiento de los genes constitutivos se ha adaptado también dentro de este modelo mediante la inclusión del campo *Marca de Constitutivo* dentro de lo que es el *Identificador de Gen*. Por tanto, el gen estaría activo siempre a menos que se cumpliesen las condiciones identificadas en la *Región Promotora*.

5.4.2 Estructura de la Red Reguladora de Genes

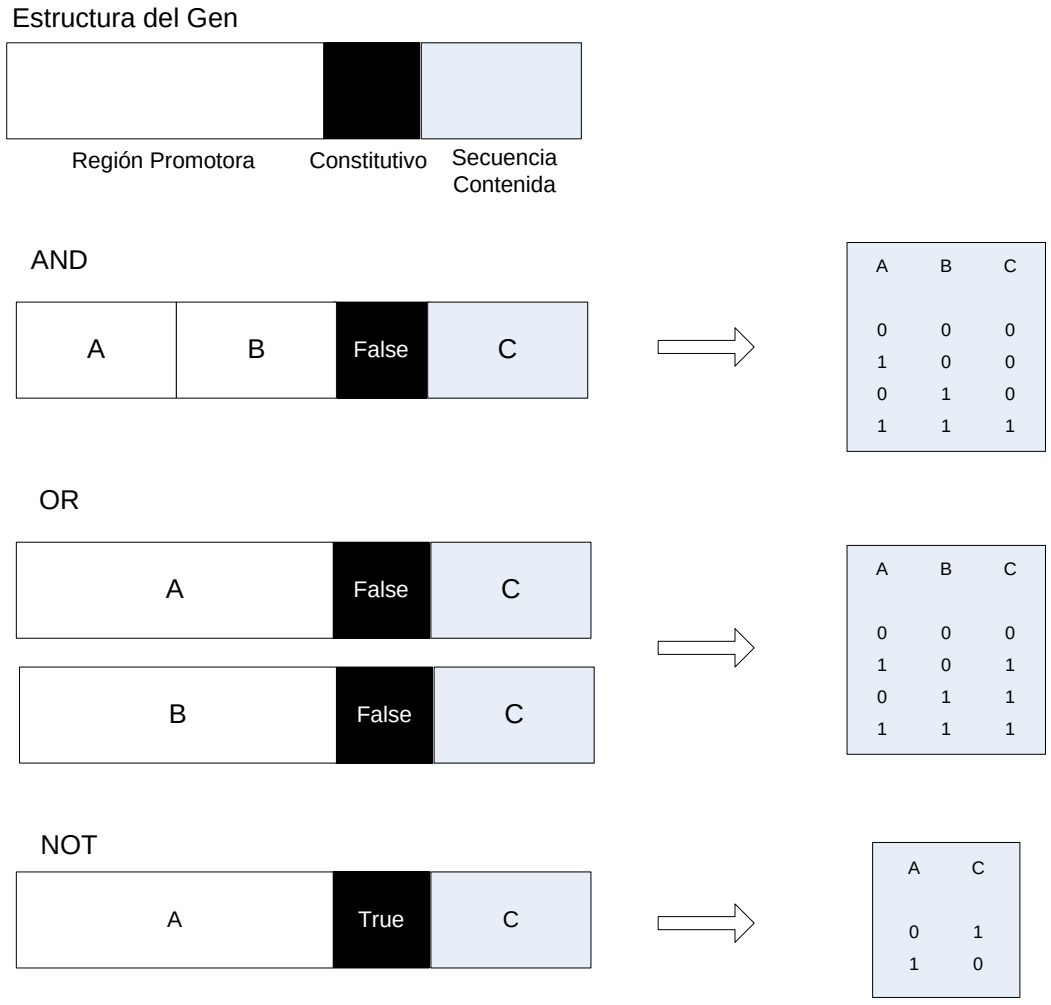

Figura 5.4: Operaciones lógicas mediante el uso de genes.

Con esta estructura en los genes y las proteínas, se puede afirmar que la célula cuenta con una Red Reguladora de Genes para su control. Como ya se comentó en el apartado

2.2 y, como demostró Kauffman (Kauffman 1968), las redes reguladoras de genes poseen la misma capacidad de procesado que un Algebra de Boole. La asunción en este caso sería interpretar la presencia de una proteína como el valor verdad en una operación booleana y, de manera opuesta, la ausencia de la proteína se interpretaría como el valor falso. De esta manera, la configuración de las distintas operaciones lógicas utilizando la red reguladora de genes de esta tesis quedaría como se muestra en la Figura 5.4.

Según esta Figura 5.4 la operación AND se puede asumir como un gen que tiene más de una proteína en la *Región Promotora.* A su vez, la operación OR se puede interpretar como dos genes que tienen la misma proteína en el *Identificador de Gen* pero distintas condiciones en la *Región Promotora*. Finalmente, la operación NOT, es asumida por los genes constitutivos. La proteína que codifican está presente a menos que se satisfagan las condiciones de su *Región Promotora*, en cuyo caso se inhibe la producción de la proteína como resultado de la presencia de otra u otras.

5.5 Operón

Como se comentó en el Capítulo 2, el concepto de operón se ha utilizado en diversos modelos. En biología, un operón representa generalmente una tarea compleja, la cual necesita la expresión de más de un gen para su ejecución. Así pues, los operones son conjuntos de genes a los cuales se les establecen una serie de condiciones que deben ser satisfechas para que cualquiera de ellos pueda ser expresado. Este mismo comportamiento se ha adaptado en el modelo artificial mediante la estructura que se muestra en la Figura 5.5.

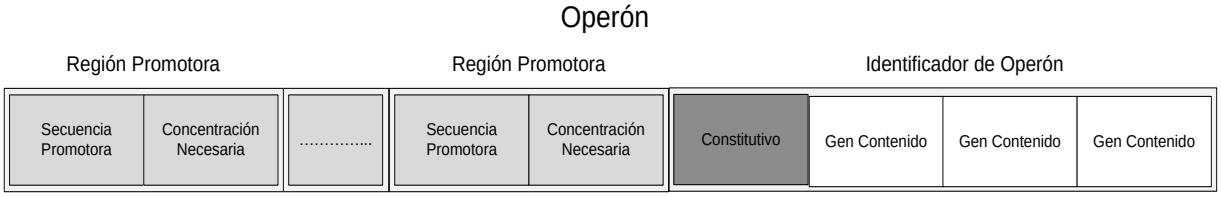

Figura 5.5: Esquema general de la estructura de un operón

La estructura mostrada en la Figura 5.5 tiene una composición similar a la mostrada en los genes. Al igual que los genes, en el operón se pueden identificar dos partes muy claramente:

- **Regiones Promotoras**. Se comportan de la misma manera que la de los genes. Identifica, por tanto, las proteínas que deben estar en el citoplasma en una cierta concentración para la activación del operón. Al igual que el gen también cuenta con dos subsecciones: *Secuencia Promotora* y *Concentración necesaria,* que se comportan de la misma manera que las del gen.
- **Identificador del Operón**. Esta sección contiene los genes codificados dentro del operón. Para ello cuenta con dos campos.
    - **Marca de Constitutivo.** Que funciona igual que la del gen, es decir, determina si se trata de un operón constitutivo.
    - **Genes Contenidos.** Esta subparte del *Identificador de Operón* contiene aquellos genes que no pueden ser expresados sin que se satisfagan las condiciones marcadas en la *Región Promotora*. Cada uno de estos genes tiene la misma estructura que la explicada en el apartado 5.4 que, a su vez, puede incluir nuevas condiciones para la expresión del gen.

Al igual que los genes, los operones pueden ser constitutivos lo que cambia su comportamiento de la misma manera. Así, un operón constitutivo permite la activación de los genes que contiene siempre hasta que se cumplen ciertas condiciones, momento en que, la posibilidad de la expresión de todos los genes contenidos es inhibida por un cierto periodo de tiempo.

5.6 ADN

Al igual que en el modelo biológico, el ADN de este modelo artificial está compuesto por operones y genes (Véase Figura 5.6). Por tanto, contiene toda la información que determina el comportamiento no sólo de la célula, sino también del organismo que contiene la célula.

Este a su vez no es un simple contenedor si no que determina la estrategia u orden de prioridad de los genes a la hora de expresarse. Por tanto es el que determina, con la información contenida dentro del citoplasma, qué genes pueden expresarse y cuáles van a ser los que se expresen de los que pueden hacerlo. Esta distinción se debe a que puede que se dé el caso que más de uno de los genes pueda expresarse ya que se cumplen las condiciones para su activación, pero no haya suficientes proteínas en el citoplasma. En este caso es el ADN el que debe determinar, según alguna estrategia, la prioridad entre los genes.

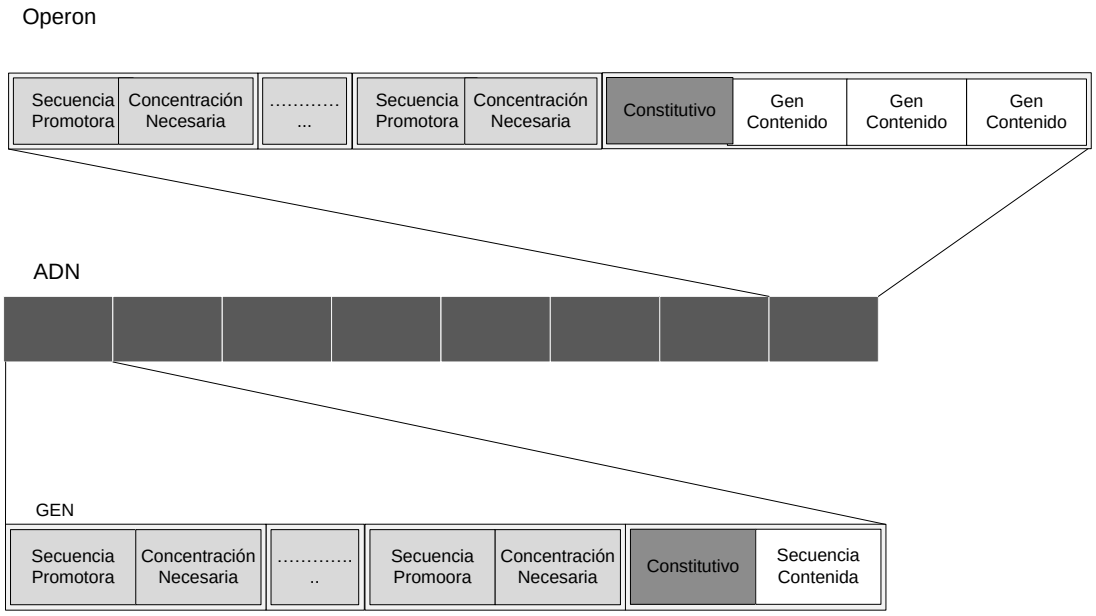

Figura 5.6: Ejemplo de ADN compuesto de genes y operones

5.7 Célula

En la naturaleza, las células son los elementos básicos que componen cualquier estructura biológica. Si se quiere ver así, se podría poner el ejemplo de que son los ladrillos que componen un sistema mayor. En el modelo artificial propuesto, las células también son los elementos básicos. Las células que aquí se proponen están compuestas por un ADN y un citoplasma como los descritos en este capítulo. Así pues, con esas herramientas las células son las unidades del sistema encargadas de procesar la información en forma de proteínas que contiene dicho sistema. Se puede ver su comportamiento como el de una máquina de Turing en la que, con el estado existente, actúa en consecuencia con las instrucciones del ADN para ofrecer una respuesta consecuente. Además esa respuesta debe estar coordinada con sus células vecinas para ofrecer la respuesta global que se espera del sistema.

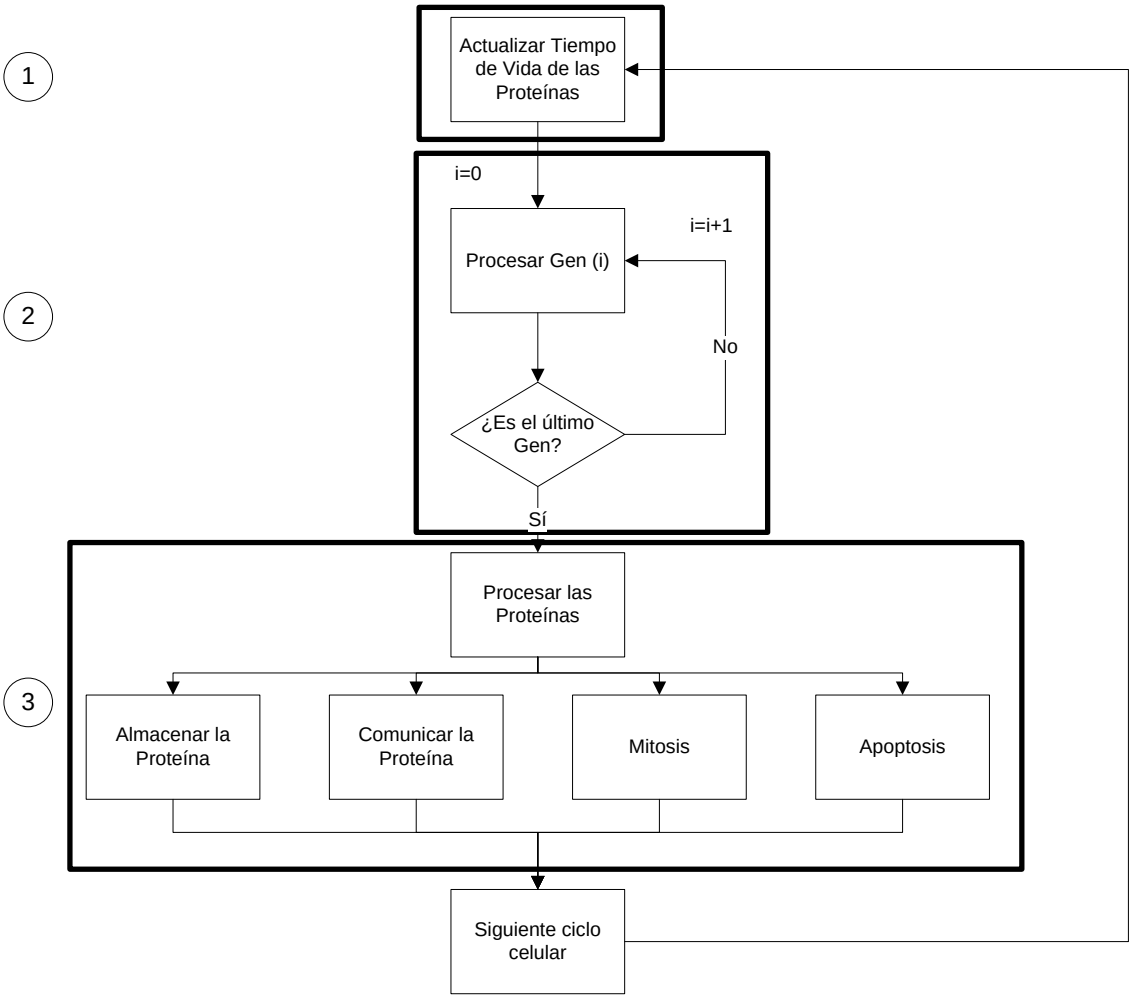

Figura 5.7: Esquema del procesado en un ciclo celular

Por tanto, las células realizarán una serie de acciones asociadas a la expresión del ADN, que se pueden resumir como:

- Almacenar la proteína generada en el citoplasma.
- Transmitir esa proteína a una célula vecina.
- Que la proteína provoque la mitosis de la célula.
- Que la proteína provoque la muerte de la célula.

Se hace necesario comentar en este punto, que la célula define lo que se conoce como "ciclo celular" para coordinar las distintas acciones. Este ciclo celular pretende simular el hecho que tiene lugar en el modelo biológico, por el cual, algunas de las acciones que puede ejecutar la célula están limitadas en tiempo. Así, por ejemplo, la célula puede comunicarse múltiples veces con otras células durante un ciclo celular pero, la célula, sólo se podría dividir una única vez. Así pues, un ciclo celular puede asumirse como la unidad de tiempo del sistema celular artificial desarrollado.

Utilizando este concepto de ciclo celular, se determinan qué operaciones se pueden realizar y, por tanto, el comportamiento de la célula. Así, las distintas acciones que se ejecutan en un ciclo celular se pueden ver en la Figura 5.7. En esta figura se muestra una imagen con un esquema básico del funcionamiento de la célula. Como se puede ver, el comportamiento dentro de un ciclo celular se puede dividir en pasos:

1. Actualización del Tiempo de vida de las proteínas y, en el caso de que llegue a cero, proceder a su borrado.
2. Determinar los genes que se van a transcribir y producir las nuevas proteínas.
3. Ejecutar las acciones asociadas a las proteínas producidas, ya sea almacenarlas, comunicarlas, dividir la célula o ejecutar la apoptosis.

Todas las células que componen un tejido realizan las mismas operaciones, los ciclos celulares tienen como finalidad establecer un marco temporal para que las células utilicen las proteínas correctas en cada instante sin usar información obsoleta o cuyo momento de creación está en un futuro.

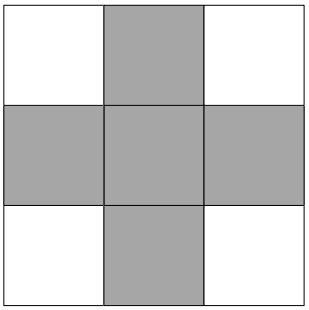

Figura 5.8: Vecindario de Von Neumann.

Además de esto hay que remarcar que para las acciones de mitosis y apoptosis celular se utilizan un grupo de proteínas, que se identifican como especiales. En el caso de la mitosis existe una proteína especial por cada una de las direcciones, por ejemplo, en un sistema 2D y que sólo utilice un vecindario de Von Neumann, como el que se muestra en las casillas sombreadas de la Figura 5.8, se tendrían 4 proteínas (Arriba, Abajo, Izquierda y Derecha). Cuando una célula produce una de estas proteínas se produce la mitosis en la dirección marcada a menos que, la célula, ya se haya dividido en ese ciclo celular o la posición ya esté ocupada. En el caso de que no se pueda ejecutar la mitosis, la proteína es almacenada y tratada como una más. Para la operación de apoptosis sólo se identifica una única proteína pero, al contrario que con la mitosis, no se ejecuta sólo con que aparezca la proteína. Para esta operación la proteína especial se interpreta como si fuese un producto tóxico. Por tanto se ha fijado un umbral por encima del cual la "proteína tóxica" es mortal para la célula.

5.8 Entorno

Una vez explicados los elementos que componen el sistema se hace necesario un último elemento que, por así decirlo, establece las reglas de juego para las células. En la

naturaleza ese elemento es el entorno donde se desarrollan las células. En la adaptación aquí presentada el entorno se encarga de almacenar, posicionar a las células y administrar las proteínas que no se encuentran dentro de ellas.

La tarea relacionada con almacenar la posición tiene sus implicaciones en que es el entorno el que determina el tipo de vecindario que tendrá. Por ejemplo, podría ser un vecindario en 2D, 3D, Cartesiano o Basado en elementos hexagonales (Véase Fig. 5.9). Esto, además de en la mitosis, también influye en los mensajes que recibe una célula, ya que determina su `posición relativa.

Figura 5.9: Ejemplos de posibles vecindarios para el entorno.

Por otro lado, como ya se dijo, esta estructura no sólo contiene las células sino que además determina cómo se tratan las proteínas libres, entendiendo por proteínas libres aquellas que no se encuentran en el interior de ninguna célula. Estas proteínas pueden haber sido expulsadas por las células a modo de comunicación, o ser proteínas ya presentes, introducidas por algún elemento externo. El entorno será el encargado de determinar cómo se mueven en su interior las proteínas libres y, además, tendrá que actualizar el tiempo de vida de estas proteínas. Finalmente, apuntar que, en el caso de

que el tiempo de vida de las proteínas llegue a cero, el entorno será el encargado de eliminarlas.

5.9 Modelo de Comunicación

Una vez definidos todos los elementos que componen el sistema celular, surge la necesidad de dotar al entorno con un mecanismo que permita comunicarse a las células entre sí. En el desarrollo de esta tesis se han explorado las dos alternativas posibles. Así se puede hablar de un modelo de comunicación basado en elementos discretos y, un segundo modelo, basado en probabilidades.

Aunque el funcionamiento de ambos se detallará a continuación, comentar que la exploración de un segundo modelo de comunicación surge como resultado de una necesidad detectada durante el desarrollo de las pruebas. Esta necesidad tenía su origen en el coste en recursos (tiempo y memoria) que necesitaba el modelo de elementos discretos cuando se trataba de afrontar problemas más complejos. Se presenta una comparativa en los apartados 6.2.2.1 y 6.2.2.2 del capítulo dedicado a las pruebas.

5.9.1 Comunicación Basada en Elementos Discretos

El primero de los modelos que se planteó durante la realización de esta tesis es un modelo basado en elementos discretos. En el entorno de este modelo, las células ocupan una posición dentro del mismo como se muestra en Figura 5.10.

El citoplasma de las células comprobará la concentración, tanto dentro, como en la posición del entorno donde se encuentra la célula cada ciclo celular. Cuando la

concentración de una proteína es mayor en el interior que en el exterior de la célula, entonces la célula reduce dicha cantidad colocando algunas de las proteínas que tiene en el citoplasma en la posición del entorno. Cuando la situación es la opuesta, la célula aumenta su concentración capturando algunas proteínas del entono y almacenándolas en el citoplasma.

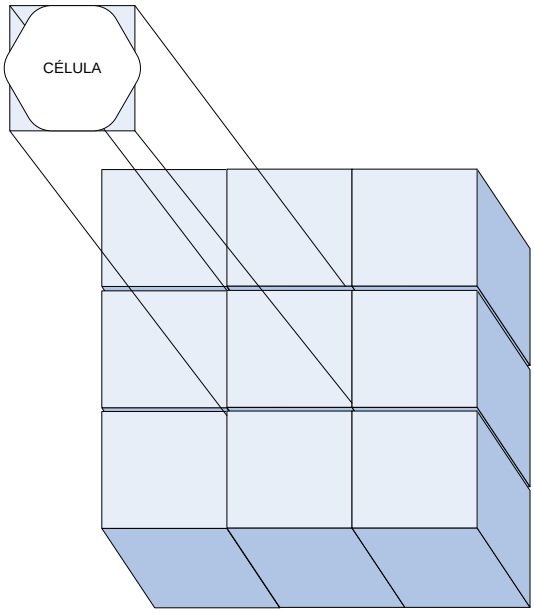

Figura 5.10: Una célula colocada en un entorno con vecindario cartesiano en 2D

Un hecho que sucede en la naturaleza, que también se ha querido incluir, es la resistencia que ofrece la membrana celular a la homeostasis de las proteínas. Es decir, las células están separadas del entorno por una membrana semipermeable compuesta de lípidos. Esta membrana ofrece una cierta resistencia a que las proteínas fluyan de un lado al otro de la misma. Con esto se busca que la célula pueda tener una concentración algo mayor de una determinada proteína que el entorno para así poder expresar un comportamiento diferenciado respecto de sus vecinas. En otro caso todas las células

tendrían los mismos datos y, por tanto, se comportarían exactamente igual. Lo que se busca es que exhiban un comportamiento diferenciado que se complemente con el de sus vecinas. Para esto, se le ha añadido un peso a las condiciones como se muestra en las Eq. 5.2 y Eq. 5.3. Con este peso se pretende simular la resistencia de la membrana.

$$\text{if (Concentración de la Proteína}_i \text{ en la célula >= Concentración de la Proteína}_i \text{ en el entorno + Factor Corrección de la Membrana) then Mover Proteínas de la Célula al Entorno} \quad (5.2)$$

$$\text{if (Concentración de la Proteína}_i \text{ en la célula + Factor Corrección de la Membrana <= Concentración de la Proteína}_i \text{ en el entorno) then Mover Proteínas del Entorno a la Célula} \quad (5.3)$$

La condición Eq. 5.2 modela el tránsito de las proteínas salientes de la célula y la condición Eq. 5.3 modela el tránsito de las proteínas entrantes. En estas condiciones, se comprueba la concentración de una proteína *i* tanto dentro como fuera de la célula corregida mediante la adición de un factor de corrección. Cuando se satisface una de las condiciones las concentraciones son reequilibradas mediante la recolocación de algunas proteínas de un lado en el otro.

Cuando una célula biológica posiciona nuevas proteínas en el entorno estas son expulsadas en una cierta dirección, quedándose estas proteínas en el espacio intercelular existente. Para representar este hecho, cuando una célula expulsa nuevas proteínas al entorno, estas llevan una etiqueta identificando la dirección en que deben ser posicionadas. Con esto se busca el efecto natural de la creación de gradientes de movimiento de las proteínas. Por ejemplo, en un entorno cuyo vecindario sea un vecindario de Turing (8 vecinos) las proteínas podrían ir a cualquiera de las 8 posiciones que rodean la célula (Figura 5.11).

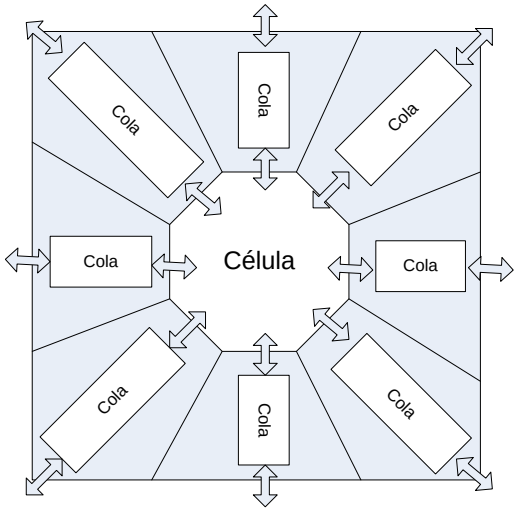

Figura 5.11: Estructura de comunicación

Por tanto, siguiendo el mismo ejemplo, se podría pensar que las posiciones del entorno están divididas en 8 subpartes, dónde la célula posiciona las proteínas que expulsa. Una vez las proteínas están en el entorno se comprueba cada ciclo celular las posiciones vecinas para igualar las concentraciones. Así, por ejemplo, como se muestra en la Figura 5.12 se muestran dos posiciones vecinas A y B. Cuando se comprueba la concentración de las posiciones se encuentra que, la concentración en A, es mayor que en B (5 proteínas frente a 1). El siguiente paso es comprobar el número de proteínas que tratan de ir de A a B frente al número que trata de ir de B a A. En el caso en que el número de proteínas que traten de ir de A a B sea mayor que el opuesto, entonces se mueve la diferencia entre las proteínas que tienden a ir de un lado al otro como se muestra en Figura 5.12. Estas proteínas que se desplazan se colocan en la posición con la qué están etiquetadas. Esto se hace para favorecer el desplazamiento de las proteínas a distancias mayores.

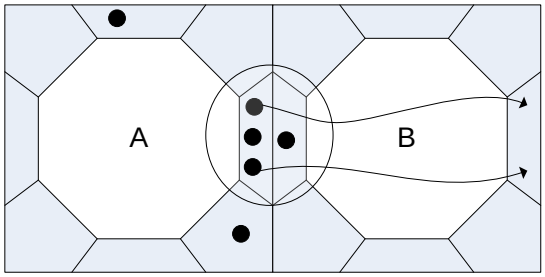

Figura 5.12: Ejemplo de comunicación discreta entre dos posiciones del entorno

Como se comentó anteriormente este modelo constituyó una primera aproximación para realizar las comunicaciones entre las células. Es un modelo que tiene una importante similitud con el modelo biológico del que adapta la mayoría de las ideas. El gran problema es su coste de ejecución en memoria. Esto provoca que no sea factible su uso en problemas complejos, por este motivo, se propone un segundo modelo de comunicación más ligero en el apartado 5.9.2, que trata de solventar este problema.

5.9.2 Comunicación Basada en Probabilidades

En el segundo de los modelos propuestos lo que se intenta es minimizar los cálculos y el uso de memoria pero intentando seguir siendo fiel al comportamiento observado en la naturaleza.

En la búsqueda de esta reducción de coste se optó por abandonar la idea de usar elementos individuales y usar en su lugar probabilidades de recepción de una determinada proteína por las células. Esto se basa en un hecho observado en la naturaleza. Según este hecho las proteínas que son expulsadas por las células son expulsadas en grupos a un medio acuoso. En este medio fluirán y lo más probable es

que tengan como receptor una de sus vecinas pero, alguna de las proteínas puede fluir y ser recibida por células más lejanas.

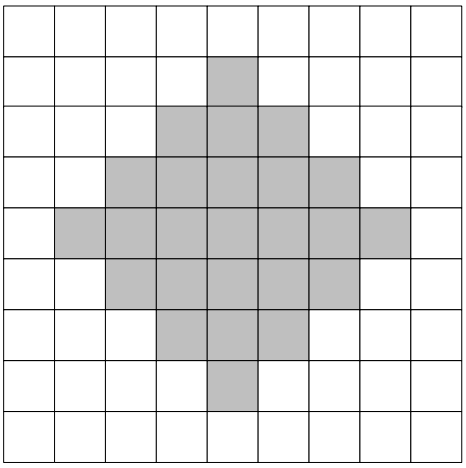

Figura 5.13: Ejemplo de vecindario Von Neumann con distancia manhattan 3

En este modelo cuando la concentración de cualquiera de las proteínas supera un límite preestablecido, la célula coloca algunas de las proteínas de su citoplasma en el entorno. Esto modifica la probabilidad de que una célula colocada en el vecindario de la célula emisora reciba una proteína de ese tipo. Para determinar el vecindario de una célula se utiliza un valor fijo como límite y la distancia Manhattan, lo que configura un vecindario del tipo de von Neumann (Figura 5.13).

Cuando una célula pone una nueva proteína en el entorno esto modifica una función de probabilidad asociada a la posición desde donde es emitida. Por tanto, cada posición del entorno tiene asociada una función de probabilidad para cada una de las proteínas existentes. El valor de dicha función de probabilidad decrece con la distancia del punto de emisión como se muestra en la Figura 5.14. En esta el valor de la probabilidad se representa por el grado de oscuridad de los mismos, a mayor

intensidad más probabilidades de recepción. En este ejemplo se ha utilizado una distancia 3 para el vecindario de von Neumann. Como se ve, la probabilidad de encontrar una proteína emitida por una célula que estuviese en el centro decrece con la distancia. Sobre esta función se ha modificado el valor para la distancia 0, ya que no se quiere permitir que las células que emiten una proteína recojan una proteína emitida por ellas mismas.

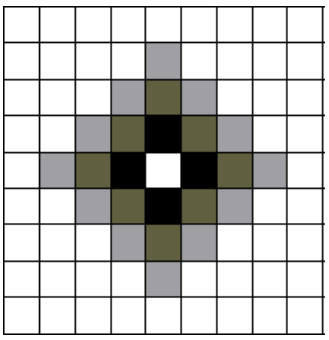

Figura 5.14: Ejemplo de función de probabilidad

Por tanto, el valor de la función dependerá de dos parámetros el número de proteínas posicionadas en un punto del entorno y la distancia al origen del punto que se está calculando.

En el ejemplo de la Figura 5.15 se puede ver una representación del valor de la función de probabilidad para las posiciones del vecindario de una célula.

Con estas probabilidades, las células comprobarán cada ciclo celular las probabilidades de las casillas de su vecindario y mediante una tirada de azar determinarán si la célula encuentra una de las proteínas que tiene como origen esa posición. En caso de ser hallada dicha proteína se reduce el número de proteínas en el origen y, por tanto, se modifica el valor de la función de probabilidad.

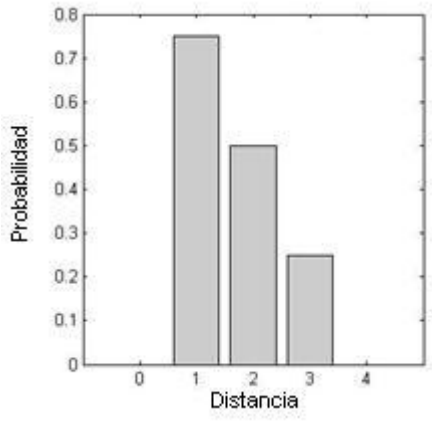

Figura 5.15: Probabilidad entre dos posiciones

5.10 Modelo de Búsqueda

Como ya se comentó en el apartado 5.6, el ADN es el conjunto de reglas codificadas dentro de la célula que determinan el comportamiento y desarrollo del sistema celular. Por tanto, el ADN será el objeto de la búsqueda dentro de estos sistemas teniendo como objetivo conseguir un determinado comportamiento. Es necesario remarcar que esta es una tarea extremadamente compleja ya que del ADN, para una determinada tarea, se desconoce el número de reglas, el número de condiciones de dichas reglas, el valor de las condiciones y los consecuentes, etc.

Una alternativa para la búsqueda de este tipo de sistemas es el uso de Algoritmos Genéticos (Fogel et al. 1966; Holland 1975; Goldberg 1989) en su variante con individuos de longitud variable (Deb 1991). La razón de porqué son una buena alternativa es que se trata de un sistema de búsqueda muy robusto y flexible, cualidades necesarias en este caso. Además, las operaciones de mutación y cruce han sido adaptadas al caso concreto de este tipo de búsqueda.

Comentar que el valor de ajuste de un individuo surge de transformar el individuo en un ADN para el sistema celular. Entonces se coloca ese ADN en una célula y se deja que desarrolle el comportamiento codificado por un periodo. Tras ese tiempo se evalúan las propiedades del tejido resultante del desarrollo y con eso se puede calcular el error cometido y, por consiguiente, un valor de ajuste.

5.10.1 Codificación de Individuos

En el algoritmo planteado cada uno de los individuos representa una posible solución al problema, por tanto, codifican una cadena de ADN completa. Esto presenta unas particularidades que deben ser tenidas en cuenta. Como se comentó, el número de los genes del ADN es desconocido, lo que lleva innegablemente al uso de individuos de longitud variable. Además, como también se dijo, la composición de los genes es desconocida (número de promotores).

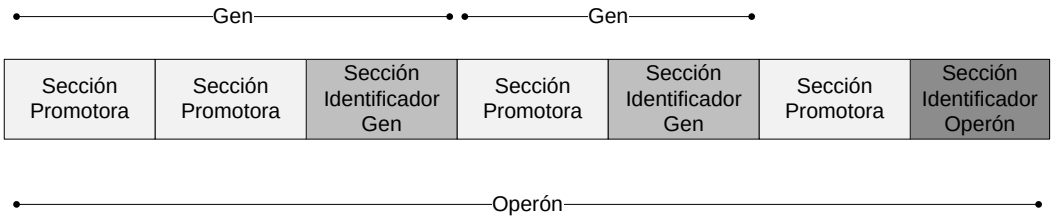

Figura 5.16: Codificación de un operón mediante la estructura de búsqueda propuesta para el algoritmo genético.

Finalmente, hay que tener en cuenta el uso en este modelo del concepto de operón que se vio en el punto 5.5. Los operones tienen, como los genes, el número de promotores variable pero, además, el número de genes que contiene también debe ser variable.

Teniendo en cuenta todas estas particularidades se propone un esquema para la codificación como el que se puede ver en la Figura 5.16

Como se puede ver en dicha figura, los individuos del Algoritmo Genético están compuestos por bloques que pueden ser de tres tipos distintos, siendo estos:

- **Sección promotora**. Estas secciones contienen el identificador y la concentración que figuran en los campos de la *Región Promotora* que se explicó en el apartado 5.4.
- **Sección identificadora de Gen**. Esta sección contiene los campos del *Identificador de Gen* explicado en el apartado 5.4. Esta sección además marca la creación de un gen del sistema celular y se asocia con las secciones promotoras previas hasta que se encuentra otro tipo de sección distinta o el comienzo del individuo como se muestra en la Figura 5.16.
- **Sección identificadora de Operón**. Esta sección marca la creación de un operón y contiene 2 valores: marca de constitutivo y número de genes contenidos. El primero de los campos contiene el valor para el campo *Marca de Constitutivo* explicada en el apartado 5.5. El segundo de los campos cuenta el número de genes de los identificados previamente que van a estar contenidos en su interior. Esta asociación se lleva a cabo hasta que se alcanza el número o se encuentra en el individuo otra sección identificadora de operón. Además, el propio operón se combina con las secciones promotoras previas hasta que se encuentra una sección de otro tipo o el comienzo del individuo. Un ejemplo de la asociación puede verse en Figura 5.16.

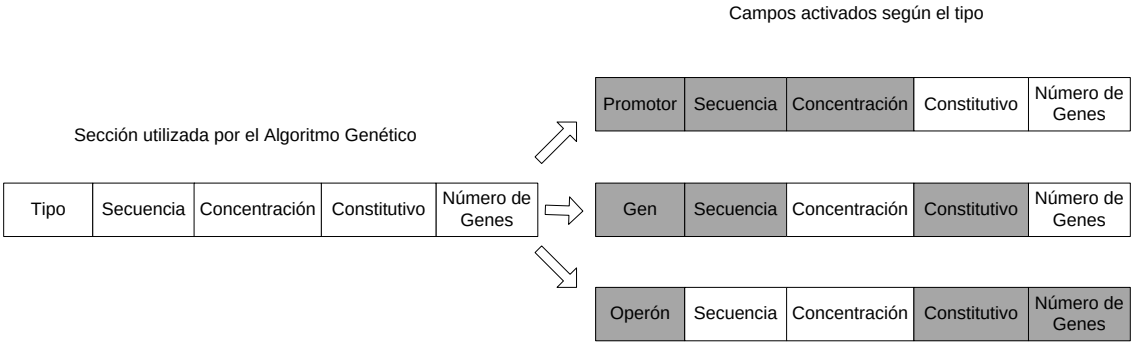

Figura 5.17: Estructura interna de las secciones del Algoritmo Genético y como el campo de tipo activa unos u otros campos.

Este esquema de codificación se puede hacer de tal manera que cualquier sección contenga todos los valores necesarios para cualquiera de los tipos de secciones (véase Figura 5.17). Añadiendo un campo que identifique el tipo se pueden tener activos sólo aquellos que se corresponden con el tipo seleccionado. Como se ve en la Figura 5.17 existen 3 tipos de sección: promotor, gen u operón. Según el tipo de sección se activarían unos u otros de los campos necesarios para codificar el contenido de ese tipo (en la figura los campos activados son los sombreados). Este hecho facilita la adaptación de las operaciones de cruce y mutación que se explicarán a continuación. Además, comentar que el esquema de codificación permite al Algoritmo Genético explorar todas las posibles combinaciones de reglas y comprime los campos comunes reutilizando la información.

El problema relativo que tiene este esquema de codificación es que las asociaciones de las distintas secciones pueden crear genes u operones que no se activen, es decir, que no son constitutivos y no tienen promotores. Este hecho, que un principio puede parecer un problema, se da también en la naturaleza (Force et. Al. 1999), su objetivo es

poder modificarlos y activarlos en un futuro o simplemente ser un legado genético. Este mecanismo es uno de los que provoca los conocidos como "saltos evolutivos". Estos saltos son los que hacen evolucionar a las especies y mejoran su adaptación al medio. Por tanto, su inclusión en este tipo de sistemas parece más que adecuado.

5.10.2 Operación de Cruce

El cruce como se explicó en el apartado 2.3.2, trata de simular la reproducción sexual. Por tanto, el Algoritmo Genético seleccionará 2 padres dentro de la población y generará 2 hijos que serán insertados en la misma. La diferencia es que, en este caso, los padres no son de longitud fija. Este hecho desemboca en la adaptación de esta operación ya que tiene que solventar esta dificultad.

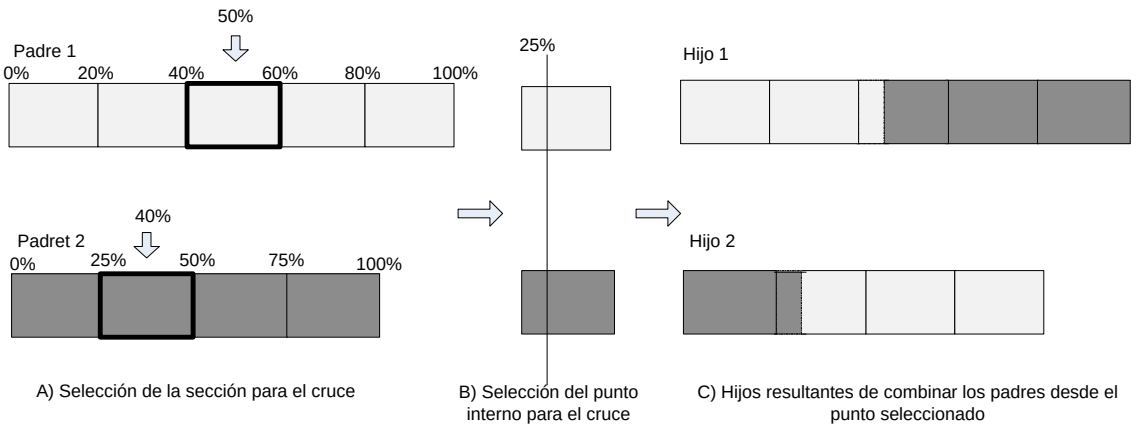

Figura 5.18: Ejemplo de cruce para individuos de longitud variable compuestos de secciones

Por lo enunciado anteriormente la selección del punto de cruce y la forma de realizar este, se convierte en un hecho no trivial y del que en la literatura no hay un consenso.

En este trabajo se ha usado el descrito en (Fernandez-Blanco et. al. 2009) para este tipo de sistemas celulares.

Según este método el punto de corte se selecciona en función de la longitud del individuo. Por tanto, se generan dos porcentajes aleatorios, uno por cada padre, y se selecciona la sección que se corresponde con el porcentaje de la longitud, como se ve en la Figura 5.18a. Remarcar que se entiende por longitud el número de secciones del individuo. Una vez seleccionadas las secciones se selecciona un nuevo punto dentro de las secciones, en este caso el mismo para cada una, a fin de crear individuos válidos (Fig.5.18b). Una vez hecho esto se combinan las secciones para crear dos nuevas secciones descendientes que remplazarán su posición en los hijos. El resto de las secciones se combina como se muestra en la Figura 5.18c. Así, los hijos contienen la primera parte de uno de los padres y un número aleatorio del segundo de los padres.

Este tipo de cruce mejora las prestaciones del cruce clásico utilizado en los Algoritmos Genéticos de longitud variable de la literatura (Deb 1991). Esta mejora viene del hecho que el cruce clásico tiende a mantener el mismo número de elementos en los hijos, mientras que, con el aquí propuesto, se exploran mucho más el número de reglas al poder modificarlo drásticamente. Este hecho es muy deseable en este tipo de sistemas ya que, como se dijo, el número de reglas es absolutamente desconocido.

5.10.3 Operación de Mutación

Al igual que la operación de cruce la operación de mutación también ha tenido que ser adaptada para trabajar con individuos de longitud variable. Así, cuando se ejecuta una

mutación sobre un individuo, el operador puede realizar tres acciones distintas sobre el mismo (Figura 5.19):

- **Añadir una nueva sección al individuo.** Además esta operación se subdivide en dos más, ya que la nueva sección puede ser la copia de una previa o una sección completamente nueva.
- **Borrar una sección.**
- **Cambiar el valor de una sección.** Esta puede o bien cambiar el valor de un campo o incluso cambiar el tipo de la sección provocando una reinterpretación del ADN.

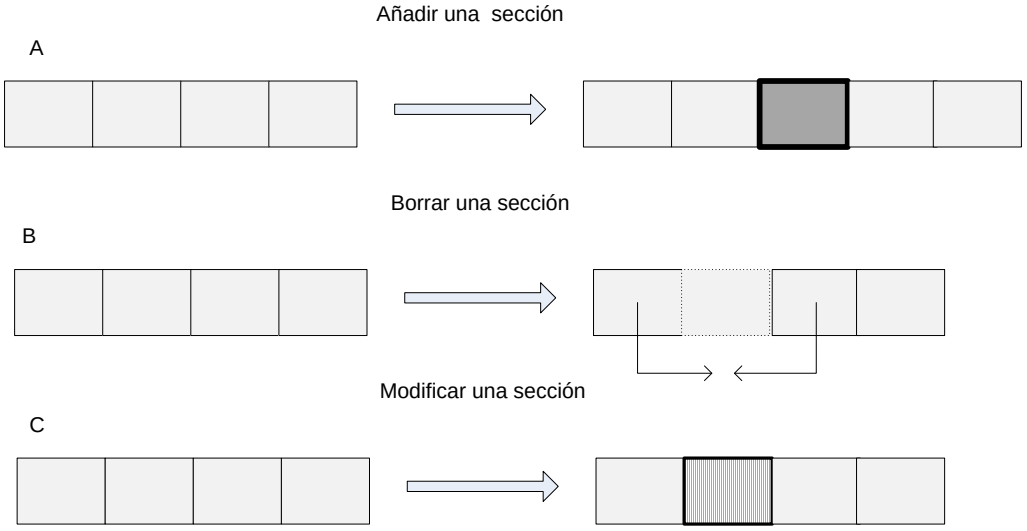

Figura 5.19: Tipos de mutación utilizadas por el Algoritmo Genético. (A) Añadir una sección, (B) borrar una sección y (C) Modificar el contenido de una sección

Comentar que la operación de copia esta bioinspirada y, normalmente, permite a los organismos vivos probar alternativas mientras mantienen copias funcionales de los genes (Force et. al. 1999). La operación de cambio requiere una explicación un poco

más profunda. La operación provoca un cambio en el valor de los campos que tiene activos una sección, pero entre los campos activos también está el de tipo de sección. Si se cambia este último eso hace que se activen los campos del nuevo tipo y se desactiven los del antiguo tipo de la sección. Además esto provoca que se tenga que reinterpretar el ADN con la nueva sección.

Produce una gran tristeza pensar que la naturaleza

habla mientras el ser humano no escucha

Víctor Hugo

Capítulo 6

Resultados

En este capítulo se agrupan el conjunto de resultados que se han extraído del uso de una implementación del modelo explicado en el capítulo 5. Dicha aplicación se ha utilizado sobre dos tipos distintos de problemas, en concreto, problemas de generación de formas y de procesado de información. En este capítulo se detallará la realización de estas y las conclusiones extraídas de cada una.

6.1 Consideraciones Generales

Las pruebas, que se presentan en este capítulo, han sido las realizadas durante el desarrollo del sistema. Por tanto, tienen como ambición comprobar determinados comportamientos del sistema. En el hilo de este razonamiento se adopta la decisión de utilizar un entorno 2D para las pruebas. Aunque el sistema es fácilmente exportable al 3D, se optó por simplificar el entorno donde se desarrollan las células a fin de poder realizar un análisis más claro. Además, tras analizar varios de los trabajos presentes en la literatura se comprobó que el desarrollo de las pruebas en 3D no aportaba nada salvo incrementar la complejidad del análisis.

Como ya se dijo, el conjunto de pruebas realizadas puede dividirse en dos grandes grupos. En primer lugar se tendrían las pruebas destinadas a probar si las células artificiales pueden generar una forma determinada y, por extensión, si son capaces de generar una población que dé una respuesta global coordinada. En este sentido se han desarrollado varias pruebas y se han analizado posteriormente los resultados que ofrecía el sistema. En el otro conjunto de pruebas se encuentran las realizadas con una única célula, cuyo objetivo es comprobar si este sistema puede entrenarse para resolver problemas de procesado de la información. Las pruebas de este segundo conjunto de problemas, se han limitado en gran medida a una única célula para facilitar el análisis de los resultados y debido a determinados problemas que se comentarán en el apartado 6.4.

Finalmente comentar que, como se apuntó en el apartado 5.10.3, la mutación en este tipo de sistemas realiza 3 operaciones distintas: añadir una sección, borrar una sección

o modificar el contenido de una sección. Tras varias pruebas se comprobó de manera empírica que el sistema ofrecía un buen comportamiento si cada vez que se ejecuta una mutación, existe una probabilidad del 20% de que se añada o borre una sección, respectivamente y, en el restante 60% de las posibilidades, modifica la sección. Esta configuración ha sido utilizada en todas las pruebas que se presentan en este capítulo, por tanto, en los siguientes apartados se eludirá comentarlo y sólo se hará alusión a la probabilidad de mutación total.

6.2 Desarrollo de Formas

Como se comentó, el primer conjunto de pruebas, que se han realizado con el modelo de Embriogénesis Artificial planteado en este documento, es el destinado a comprobar si una célula es capaz de generar un conjunto de células que expresen un comportamiento colectivo.

Teniendo en mente este objetivo las pruebas planteadas se basan en que el conjunto de células deben conseguir organizarse en una figura geométrica determinada. La razón para escoger una figura geométrica es que son figuras perfectamente reconocibles y que se conocen perfectamente que características tienen que cumplir.

Dentro de este punto se presentan dos aproximaciones distintas una basada en elementos discretos para la comunicación y otra basada en probabilidades. La primera de ellas se puede decir que constituye una aproximación más directa al problema. Esta aproximación plantea distintos problemas principalmente de rendimiento, que se

discutirán en el punto y que se solucionan con el planteamiento de la segunda de las aproximaciones presentada.

6.2.1 Método de Evaluación por Plantilla Correctora

Teniendo en cuenta que se pretende desarrollar formas geométricas, una manera de comprobar sus características es compararlos con un patrón. El ejemplo en el que se basa este método de cálculo es en la corrección de un examen tipo test con una plantilla con la que debe corresponderse. Este método fue propuesto en (Fernandez-Blanco et. al 2007) y simplifica el desarrollo de las pruebas.

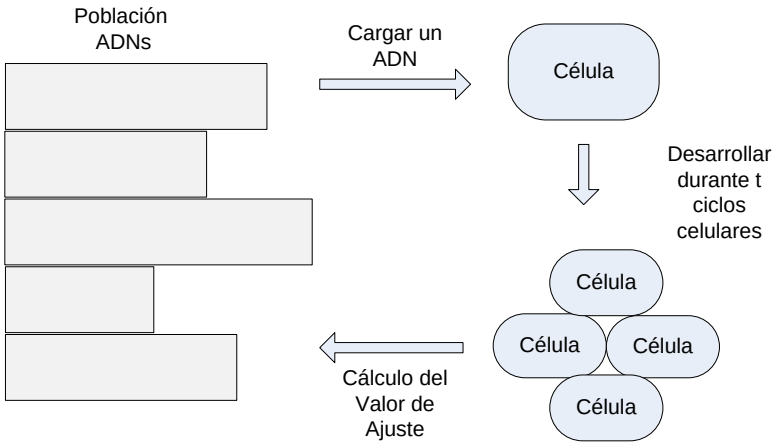

Figura 6.1: Evaluación de la población de ADNs encontrada por el Algoritmo Genético.

Con este método se establece sólo la plantilla en lugar de tener que formular todas las propiedades con ecuaciones matemáticas, cosa que a veces no resulta sencillo. Además de esta manera se evita el problema de que el individuo exprese un comportamiento que no se deseaba y que, por descuido, fue eliminado de las propiedades descritas en la fórmula matemática.

Según este método, el conjunto de reglas que representa un individuo del Algoritmo Genético se colocaría en una célula artificial. Esta célula se dejaría durante un tiempo en un entorno, tras el cual, se comprobarían las propiedades del tejido resultante (véase Figura 6.1).

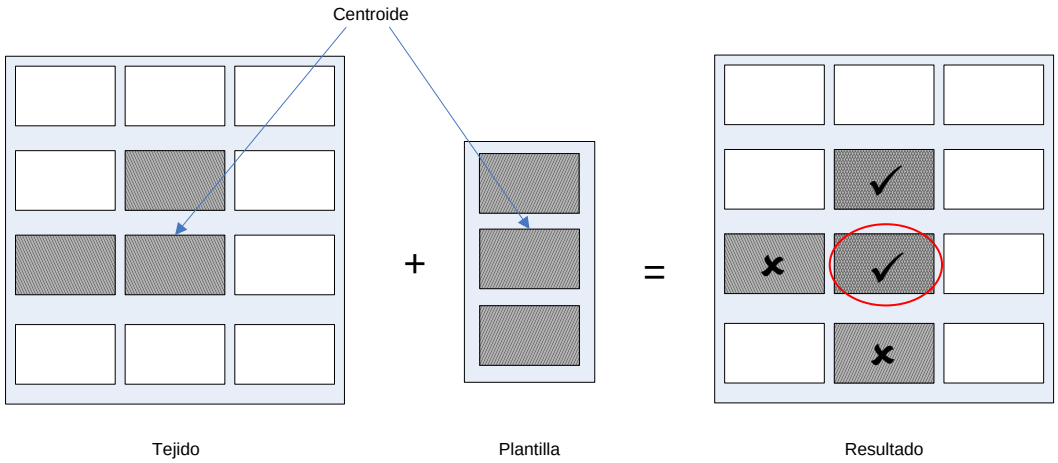

Figura 6.2: Ejemplo de evaluación mediante plantilla correctora.

Hasta este punto el proceso sería igual para cualquier tipo de propiedad que se quisiese comprobar pero, en este caso particular de comprobar la forma del tejido resultante, se puede definir una matriz Booleana que denota si las posiciones del entorno deben estar ocupadas o por el contrario dejarse vacías. Con esta matriz de posiciones (en adelante plantilla) se puede calcular la similitud de la forma generada a la forma generada por las células y la plantilla. Para calcular esa similitud basta con calcular la operación lógica NEXOR entre las posiciones del entorno y la plantilla, tal y como se muestra en Figura 6.2. En dicha figura se observa un tejido generado en la izquierda y una plantilla con la forma deseada en el centro, finalmente, en la derecha de la mencionada imagen figura el resultado de la operación NEXOR. Para realizar

dicha operación es necesario alinear ambas matrices por medio de su centroide de masas (posición rodeada con un círculo rojo). Se entiende como centroide de masas la posición que deja el mismo número de células tanto arriba como abajo, izquierda y derecha. Se puede ver un ejemplo en la Figura 6.3 donde la posición marcada con un círculo rojo deja 2 células tanto arriba como abajo y lo mismo a izquierda y derecha.

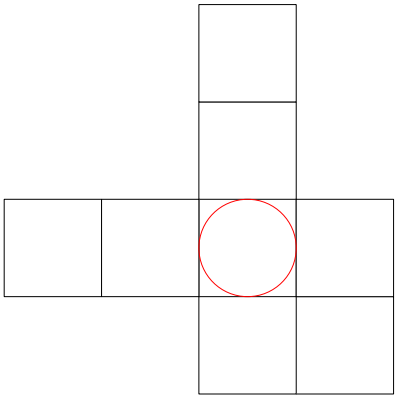

Figura 6.3: Ejemplo del cálculo de un centroide de masas.

La razón para este alineamiento es que la plantilla suele ser más pequeña que el entorno. Una vez alineados los centroides, se contabilizan las posiciones ocupadas que caen fuera de la plantilla y se toman como errores. Además, a esta cantidad hay que sumarle el número de errores cometidos dentro de la plantilla. Ese valor da el grado de similitud del tejido generado a la forma deseada.

Este método ha sido usado en todas las pruebas de desarrollo de formas que se explicarán en las secciones subsiguientes.

6.2.2 Aproximaciones Realizadas

Como se comentó previamente se realizaron dos aproximaciones distintas para modelar el comportamiento del sistema celular biológico. La diferencia entre ambas aproximaciones es la forma que tienen de representar las comunicaciones entre las células. La primera de las aproximaciones plantea un modelo de comunicación entre las células basado en la comunicación de elementos discretos (véase punto 5.9.1). Este modelo es más fiel al modelo biológico pero, como se comentará más en detalle, plantea serios problemas de rendimiento. Para solucionar este hecho se plantea un segundo modelo basado en probabilidades (véase punto 5.9.2).

6.2.2.1 Aproximación con elementos discretos

Esta sección contiene todas las pruebas realizadas con la primera de las aproximaciones que se planteó a la hora de implementar el modelo presentado en el capítulo 5. Esta aproximación puede verse más cercana al modelo natural. Entiéndase esta afirmación como que es más sencillo un paso desde modelo natural al modelo aquí utilizado. Esta percepción se da porque todas las proteínas que estén dentro del sistema artificial son representadas con un elemento que se desplaza y reacciona con los genes de manera individual, muy similar a como ocurre en el modelo natural.

En el conjunto de pruebas que se presentan en esta sección se utiliza el modelo de comunicación entre células que se detalla en 5.9.1 que, como ya se explicó, modela las comunicaciones cuando se están utilizando elementos discretos para representar las proteínas dentro del sistema.

En esta sección se pueden encontrar las pruebas realizadas para generar algunas estructuras sencillas así como una explicación detallada de cómo funcionan algunos ejemplos. Además de las conclusiones que se puedan extraer de las pruebas realizadas, también se plantean los problemas que se encuentran en la utilización de este tipo de sistemas que se intentan solucionar en la implementación que se presenta en 6.2.2.2.

Tabla 6.1: Parámetros de configuración de la prueba

Sistema celular			
Parámetros	Valor	Mínimo probado	Máximo Probado
Modelo de Comunicación	Basado en elementos discretos	-	-
Uso del Operón	No	-	-
Entorno	20x20 posiciones	10x10	40x40
Ciclos celulares antes del cálculo del fitness	50	5	50
Proteína para inducir la mitosis arriba	1010	-	-
Proteína para inducir la mitosis abajo	1000	-	-
Proteína para inducir la mitosis izquierda	1100	-	-
Proteína para inducir la mitosis derecha	0011	-	-
Proteína para inducir la apoptosis	0000	-	-
Número máximo de proteínas disponibles	16	8	32
Vida de las proteínas (ciclos celulares)	3	2	10
Algoritmo Genético			
Tamaño de la población	300	50	500
Número de generaciones	1000	300	2000
Operador de selección	Ruleta		
Tasa de cruce	80%	60%	90%
Probabilidad de mutación	20%	1%	30%

Finalmente, comentar que se realizaron distintas pruebas para conseguir la configuración del sistema celular y el algoritmo genético para el conjunto que se presenta. En la Tabla 6.1 se muestran los valores adoptados finalmente, además de los valores máximo y mínimo probados para cada uno de los parámetros del sistema.

El concepto de operón, que fue explicado en el apartado 5.5, no se utiliza porque fue un concepto incorporado posteriormente al incrementarse la complejidad de las pruebas.

Para el cálculo del ajuste se le permitieron al tejido 50 ciclos celulares. El motivo de esta cifra tan elevada es que se buscaba que el comportamiento fue estable a lo largo del tiempo. Este tiempo de 50 ciclos se considera más que suficiente para que el tejido desarrolle el comportamiento codificado en el ADN. Además, con tantos ciclos, se fuerza que una vez alcanzada la forma, esta se mantenga, mientras que con pruebas con un número menor esto no estaba garantizado.

Así mismo, como se comenta en 5.7, el sistema define unas secuencias especiales para las operaciones de mitosis y apoptosis. En concreto en esta prueba se utilizaron: 1010, 1000, 1100, 0011 y 0000 de las 16 proteínas disponibles en las pruebas. El número de proteínas utilizado se varió de 8 a 32 pero, en este caso, el mejor comportamiento fue observado con 16. Este número de proteínas es el resultado del número de bits de la secuencia elevado a 2. Finalmente, en el sistema celular se estableció un tiempo de vida para las proteínas de 3 ciclos ya que este es suficiente para que las mismas sean creadas, trasmitidas y procesadas por las células. Valores más elevados de este parámetro simplemente introducían una memoria mayor en el sistema y provocaban que los cambios en el mismo por una variación de la señal fuesen más lentos.

Además de los parámetros del sistema celular propiamente dicho hubo que ajustar los parámetros del algoritmo genético. En concreto, se utiliza una población estable de 300 individuos que evoluciona durante 1000 generaciones. Esta población utiliza una estrategia de remplazo de padres elitista con una población estable (véase 2.3.3). De los

operadores se puede destacar que, el operador de selección utilizado en este trabajo fue la selección por ruleta, que ya fue explicado en el punto 2.3.1. Además, en el algoritmo genético utilizado se estableció un ratio del 80% de cruces y un 20% de probabilidad para la mutación. Este último dato, aunque atípico, ya se ha observado en otros ejemplos de la literatura (Kumar 2004b). En este tipo de sistemas por el enorme espacio de búsqueda es necesario incrementar la exploración, lo que implica incrementar la probabilidad de mutación.

Para finalizar con las consideraciones sobre este grupo de pruebas, comentar que, para el cálculo del valor de ajuste, además del método basado en plantillas de corrección explicado en el apartado 6.1, la función de ajuste incluye dos objetivos más que tienen como función que la figura esté centrada en el entorno y, un segundo subobjetivo, que trata de minimizar el número de secciones utilizadas por el Algoritmo Genético, quedando, por tanto, la función de ajuste para estas pruebas definida como:

$$\begin{aligned}\text{Valor de Ajuste} = {} & \text{Número de diferencias con la plantilla} \\ & + 0.01\ |centro - centroide| + 10^{-4} * \text{Número de secciones}\end{aligned} \tag{6.1}$$

Como apunte final sobre la configuración, comentar que todas las pruebas fueron realizadas en ordenadores de Intel Xeon 5310 a 1,6GHz y 4GB de RAM pertenecientes al clúster SVG del Centro de supercomputación de Galicia.

6.2.2.1.1 Generar una barra vertical de 5 elementos

El primero de los test que se realizó con el sistema celular puede parecer demasiado sencillo en principio, ya que de lo que trata es de generar una barra de 5 células. Pero la motivación de esta prueba es otra. Al ser una prueba tan sencilla se esperaba que

tuviese un ADN lo suficientemente pequeño para que pudiese ser analizado en profundidad y ayudase a comprender mejor el funcionamiento del mismo.

Tabla 6.2: Datos medios de las pruebas y del mejor individuo

	Valor de ajuste	Numero de secciones utilizadas	Número de genes	Tiempo utilizado para la búsqueda
Media de los 10 mejores individuos	3,0822	22,8	15,8	5h 12min
Mejor individuo	0,0209	9	3	4h 02min

De esta prueba se realizaron 10 simulaciones distintas con poblaciones aleatorias. De estas simulaciones sólo en 2 ocasiones el sistema encontró la solución del problema. Los resultados de estas pruebas se pueden ver en la Tabla 6.2 incluyendo los datos del mejor ADN encontrado y la media de los mejores individuos de cada población.

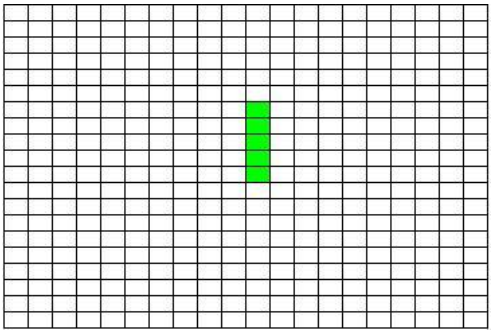

Figura 6.4: Solución para el problema de la barra vertical de 5 elementos

Comentar que de las 10 pruebas realizadas 4 de ellas no llegaron a ningún resultado válido. Es más, en esos 4 casos el ADN encontrado eliminaba todas las células. Los 4 casos restantes generaban alguna célula pero no llegaban al objetivo de 5 células. El resultado gráfico del mejor individuo para esta prueba puede verse en la Figura 6.4.

Este resultado es fruto del comportamiento codificado por un ADN de tres genes que se puede ver en la Tabla 6.3. En esta tabla se puede apreciar uno de los hechos más relevantes de este sistema, este es, que el sistema busca que los genes cumplan determinados roles en lugar de buscar una secuencia concreta. Si se estudia la Tabla 6.3 parece clara la existencia de tres roles bien definidos, uno por cada uno de los genes, en concreto: señal de reloj, contenedor de comportamiento y genes reguladores. Así en la primera línea de la Tabla 6.3, el algoritmo genético ha encontrado un gen que funciona como una señal de reloj introduciendo periódicamente una proteína en las células. Como se puede ver en la mencionada tabla, el primero de los genes es un gen constitutivo y que carece de sección promotora (apartado 5.4). Por tanto, este gen no puede ser inhibido nunca y cada ciclo celular introducirá una proteína con la secuencia 0100 en la célula. Así pues, parece que el sistema ha aprendido que necesita algún tipo de señal continua para poder expresar el resto del comportamiento. En la segunda de las filas de la Tabla 6.3, se encuentra un gen que asume el rol de contenedor de comportamiento, es decir, que contiene la información que posteriormente se quiere comprobar u observar del sistema. En el caso concreto que se está a tratar, este comportamiento es el crecimiento de una estructura. Para ello, el gen generará la proteína de división hacia arriba. Esta proteína sólo se podrá generar en los primeros ciclos de vida de una célula, ya que, como se puede ver en la Tabla 6.3, el gen requiere como promotor la proteína 0000 en una concentración de al menos 45.2 %. Esta proteína no está presente en el sistema, ya que sólo se encuentran como proteínas las 0100, 0101 y 1010. Según lo explicado en el apartado 5.4 para el uso de secuencias similares, el gen necesitaría una concentración de al menos el 90.4% de la proteína 0100

para expresar el comportamiento ya que las otras dos proteínas necesitarían una concentración de más del 100%. La mencionada concentración es sólo posible en los primeros ciclos del desarrollo del tejido mientras no se expresa el gen 3 y no se reciben mensajes de las células vecinas. Finalmente, el gen que aparece en la tercera fila de la tabla actúa como un regulador. Este gen expresa una proteína que hará que no se pueda expresar el gen 2 por que la proteína 0100 no puede alcanzar el nivel necesario de concentración al "diluirla" con la proteína expresada por este. Resaltar también que, como este gen tiene una concentración mínima muy baja, para su promotor es relativamente sencillo que consuma la proteína 0100 o la producida por el mismo para generar más con la secuencia 0101 que no le vale al gen 2 para expresarse.

Tabla 6.3: Genes que configuran la solución para la barra vertical de 5 elementos.

Nº	Secuencia Promotora	Concentración Mínima	Constitutivo	Secuencia Generada	Especial
1			true	0100	
2	0000	45,2033	false	1010	Mitosis Arriba
3	1000	3,8838	false	0101	

De la población resultante de esta prueba comentar que se alcanzó un mejor nivel de ajuste de 0.0209, ya que la solución se encuentra desplazada dos posiciones del centro (consultar Eq. 6.1). Además para encontrar la solución el algoritmo genético necesitó 9 secciones como las comentadas en el apartado 5.10.1. De estas 9 secciones, 5 dieron lugar a los 3 genes y las 4 restantes son secciones promotoras que aparecen al final del cromosoma del algoritmo genético y que no se asocian para crear ningún otro gen.

Para finalizar, las conclusiones que se pueden extraer de este experimento son:

- En primer lugar se constata que el sistema busca roles en las relaciones entre los genes no secuencias concretas.
- En segundo lugar, la identificación de los roles ha sido sencilla en este caso, pues se trata casi de un ejemplo de juguete. En problemas más complejos con más genes es mucho más difícil esta identificación ya que un gen puede tener varios roles según la relación en la que este influyendo.
- Finalmente, por las pruebas realizadas se puede asumir que no es sencillo que el Algoritmo Genético encuentre un ADN que ejecute el comportamiento deseado.

Además, sobre la población resultante de esta prueba se realizan un conjunto de pruebas para ver cuánto le costaba adaptar la solución. Estas pruebas se detallan en las secciones siguientes.

6.2.2.1.2 Ampliar la barra de 5 a 7elementos.

A partir de la prueba anterior se planteó, a su vez, cuanto tardaría el sistema en encontrar un ADN con una estructura similar pero ampliada. Además, se busca corroborar la conclusión anterior de que lo que busca el sistema son genes que cumplan unos roles. Por ello se propuso buscar, a partir de la población con la mejor solución de la prueba anterior, una cadena de ADN que desarrollase una barra de siete elementos. Por ello, se usó la misma configuración que en la prueba descrita en la Tabla 6.1 donde, simplemente, se cambió la plantilla utilizada por la función de ajuste.

Para hallar este comportamiento se ejecutaron 10 simulaciones sobre la población con el mejor individuo hallado en 6.2.2.1.1. De estas pruebas 8 de las 10 simulaciones

encontraron la solución, los detalles se muestran en la Tabla 6.4. Como se puede ver, el tiempo medio empleado es mucho más elevado que el del mejor individuo. Esto se debe a las dos simulaciones que no encontraron la solución, que elevan significativamente el tiempo medio de búsqueda.

Tabla 6.4: Datos del mejor individuo para la barra de 7 elementos

	Valor de ajuste	Numero de secciones utilizadas	Número de genes	Tiempo utilizado para la búsqueda
Media de los 10 mejores individuos	0,4812	12,3	8,2	01h 13min
Mejor individuo	0,0307	7	3	00h 15min

El resultado gráfico de esta prueba puede verse en la Figura 6.5 y es el resultado de expresar el comportamiento del ADN que se muestra en la Tabla 6.5.

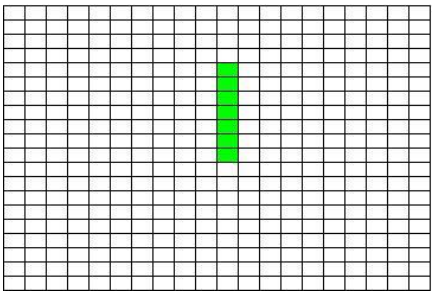

Figura 6.5: Resultado de la prueba de ampliación de la estructura

En la tabla 6.5se observan los mismos roles descritos en el apartado 6.2.2.1.1. Difiere, como era lógico pensar, en las secuencias utilizadas y las cantidades de las concentraciones mínimas. Por el resto, el análisis sería muy similar al enunciado en el apartado 6.2.2.1.1.

De la población final del algoritmo genético, comentar que el mejor individuo obtuvo un ajuste de 0.0307. Esto es fruto de que el centroide está una posición más lejos del centro que la prueba anterior.

Tabla 6.5: Genes que configuran la solución para generar una barra vertical de 7 elementos.

Nº	Secuencia Promotora	Concentración Mínima	Constitutivo	Secuencia Generada	Especial
1			true	0001	
2	0000	45,2926	false	1010	Mitosis Arriba
3	1110	3,6452	false	0010	

Además, también se comprueba que la solución utiliza sólo 7 secciones de las descritas en la sección 5.10.1, por tanto, en este caso sólo son 2 las secciones promotoras que no han encontrado una sección de identificación de genes para asociarse. Comentar que este ajuste fue alcanzado en tan solo 10 generaciones lo que parece indicar que se puede cambiar fácilmente la dimensión de las formas alcanzadas.

Por tanto, las conclusiones de esta prueba pueden resumirse como:

- Se comprueba que el sistema lo que busca son los roles que puedan jugar los distintos genes.
- El sistema apunta que es posible que la redimensión de las soluciones halladas sea, al menos en ejemplos pequeños, relativamente fácil.

6.2.2.1.3 Generación de un Cuadrado 3x3

Como se vio en la prueba descrita en 6.2.2.1.2., el sistema apuntaba que era sencillo redimensionar las soluciones usadas previamente. En este sentido, la siguiente prueba intenta desarrollar una estructura más compleja, como puede ser un cuadrado.

Tabla 6.6 Detalles de las pruebas para generar un cuadrado de 3x3

	Valor de ajuste	Numero de secciones utilizadas	Número de genes	Tiempo utilizado para la búsqueda
Media de los 10 mejores individuos	3,8125	25,8	13,37	06h 04min
Mejor individuo	0,0223	23	11	06h 30min

En ese sentido se planteó una prueba en la que se desarrollará un cuadrado de 9 elementos a partir de una población en la que se desarrolla una barra de tres elementos. Esta población para desarrollar la barra de 3 elementos surge como una redimensión parecida a la ampliación presentada en 6.2.2.1.2. No se detalla, pues el desarrollo es muy similar y carece de mayor interés. A su vez, el motivo para escoger la barra de 3 elementos como base en lugar de la barra de 5 elementos es para minimizar el procesado y simplificar esta primera estructura. Para esta prueba el sistema se configuró de manera como se muestra en la Tabla 6.1. Finalmente, y al igual que en los otros casos, se tuvo que configurar la plantilla para que coincidiese con la forma deseada.

Esta prueba fue ejecutada una decena de veces pero solamente una de las pruebas llegó a obtener la solución al problema. En la Tabla 6.6 se muestran los detalles generales de dichas pruebas.

El resultado gráfico para el mejor individuo de esta prueba puede verse en Figura 6.6. Para obtener este resultado el algoritmo genético ha necesitado un ADN que consta de 11 genes. Dicho ADN puede verse desglosado en la Tabla 6.7.

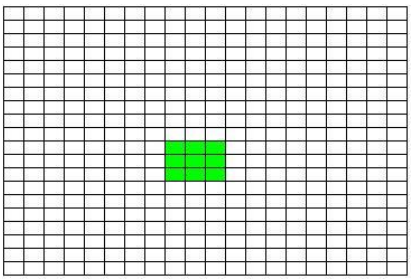

Figura 6.6: Resultado gráfico de la prueba para encontrar un cuadrado 3x3

Como se observa en la tabla, los 7 primeros genes son los que en esta ocasión funcionan a modo de señal de reloj, ya que cualquiera de ellos introduce cada ciclo celular una proteína con secuencia 1010 o 0100. Además de esto también se observan dos genes, en concreto (el número 8 y el número 10), que presentan un nivel de activación muy bajo y que funcionan en el rol de gen regulador explicado en 6.2.2.1.1. Así, aunque el gen número 8 necesita 2 promotores, la concentración necesaria de ambos es tan baja que la presencia de dos proteínas cualesquiera provocará su transcripción. Así, se puede alegar lo mismo para el gen 10, que tiene un promotor, pero la concentración necesaria es muy baja y se activará con casi cualquier proteína. Finalmente los genes 9 y 11 son los que contienen el comportamiento, en concreto, el gen 9 es el que provoca la división hacia abajo y el gen 11 el que provoca la división a la izquierda. Ambos genes tiene unas concentraciones relativamente altas, más del 40%. Este hecho provoca que necesiten concentraciones de al menos un 80% de una proteína que no difiera en más de una posición (vea la Eq. 6.1). Este hecho sólo puede darse en los primeros estadios del desarrollo ya que, con el paso de mensajes entre las células, lo producido por los genes constitutivos y lo generado por los genes reguladores hace que la concentración nunca puede alcanzar tales cotas.

Tabla 6.7: Genes para generar un cuadrado 3x3

Nº	Secuencia Promotora	Concentración Mínima	Constitutivo	Secuencia Generada	Especial
1-7		*	true	0100-1010	
8	0000-0011	2,0101E-13	false	0001	
9	0000	45,125	false	1100	Mitosis Izquierda
10	0010	1,4626E-29	false	0011	
11	0000	41,1894	false	1000	Mitosis Abajo

El valor de ajuste para el mejor individuo de la población resultante fue de 0.0223. Por tanto, se pude deducir de la fórmula que este error se debe a que la solución se encuentra desplazada dos posiciones del centro y, además, el algoritmo genético ha utilizado 23 secciones para conformar el ADN. De esas 23 secciones solo 16 han sido útiles finalmente para crear genes que fuesen utilizados por el sistema celular. Además comentar que, con respecto a lo que se apuntaba en la prueba 6.2.2.1.2, que parecía que era fácil partir de soluciones previas para obtener otras más complejas, es un tema que plantea incógnitas. En este caso, el algoritmo genético tardó en torno a 500 generaciones de las 1000 que tenía como límite en encontrar la solución. Esto identifica claramente que, a medida que las formas se hacen más complejas, se complican a su vez la comunicación entre las células. Esto hace que el ajuste de los términos de los genes no sea para nada tan trivial como apuntaba la prueba anterior.

Para confirmar la sospecha que se plantea, según la cual el partir de soluciones previas no garantiza el reducir el tiempo de búsqueda para estructuras más complejas, se planteó realizar la misma prueba pero partiendo de una población aleatoria de ADNs.

La nueva prueba se ejecutó 10 veces. De estas ejecuciones, sólo 2 encontraron alguna respuesta dentro de las 1000 generaciones aunque con un nivel de error mayor que el

obtenido por la prueba anterior. Los datos de estas pruebas pueden verse en la Tabla 6.8.

Tabla 6.8 Datos de las pruebas para el cuadrado 3x3 partiendo de una población aleatoria

	Valor de ajuste	Numero de secciones utilizadas	Número de genes	Tiempo utilizado para la búsqueda
Media de los 10 mejores individuos	7,7111	65,6	22,5	10h 12min
Mejor individuo	6,003	30	13	7h 30min

En concreto el mejor individuo obtenido por estas pruebas tenía un valor de ajuste de 6.003. Este valor supone que el individuo comete un error en seis posiciones y para ello utiliza 30 secciones. Lo cual a todas luces es un peor resultado que las pruebas realizadas partiendo de una población con un resultado más simple.

Por tanto las conclusiones de esta prueba puede ser resumidos como:

- La explicación de los genes se hace más compleja a medida que aumentan tanto el número de elementos como la complejidad de la forma desarrollada.
- El uso de una población que soluciona una forma más simple, aunque no garantiza que la solución vaya a ser encontrada más fácilmente, parece simplificar las búsquedas. El aumento de la complejidad de las comunicaciones implica que el ajuste de los genes no sea fácil pero, como norma general, es más sencillo que una búsqueda colocando los individuos de manera aleatoria.

6.2.2.1.4 Generación de un Cuadrado de 5x5

Otra de las pruebas que se realizó fue la búsqueda de una estructura algo más compleja que las anteriores partiendo, tanto desde una población aleatoria, como desde la mejor

población hallada en 6.2.2.1.3. La estructura que se decidió afrontar en esta prueba es un cuadrado de 5x5 elementos. Puede que no parezca una estructura mucho más compleja, pero el incremento de elementos de procesado (células) y la cantidad de mensajes entre estas para mantener esa forma hacen que sea una prueba perfecta para testear el sistema. Para la realización de esta prueba la configuración del sistema quedo como se puede consultar en la Tabla 6.1.

La prueba fue ejecutada con 10 poblaciones distintas donde ninguna de ellas llegó al resultado óptimo. El resumen de dichas pruebas puede verse en la Tabla 6.9.

Tabla 6.9 Resultados para la generación de un cuadrado 5x5

	Valor de ajuste	Numero de secciones utilizadas	Número de genes	Tiempo utilizado para la búsqueda
Media de los mejores individuos de las 10 pruebas partiendo de una solución previa	14,9681	81,2	34,2	26h 01min
Mejor individuo desde población previa	8,02	62	27	24h 11min
Media de los mejores individuos de las 10 poblaciones aleatorias	16,7814	73,1	30,8	26h 04min
Mejor individuo desde población aleatoria	2,0245	45	16	23h 15min

Como se puede observar, por los datos mostrados en la Tabla 6.9, en esta ocasión los mejores resultados son los ofrecidos por un individuo encontrado a partir de una población aleatoria. Uno de los motivos para este comportamiento puede ser que, a medida que las comunicaciones se hacen más complejas, al aumentar el número de células, es más difícil para el sistema reconfigurarlo. Si bien, en media, los resultados

son mejores, parece así mismo claro, que los dos mejores resultados fueron obtenidos partiendo de una población aleatoria. El resultado gráfico del mejor individuo puede verse en la Figura 6.7. La ecuación utilizada para el cálculo del valor de ajuste es la Eq. 6.1, pero adaptando la plantilla para que se corresponda con la forma deseada.

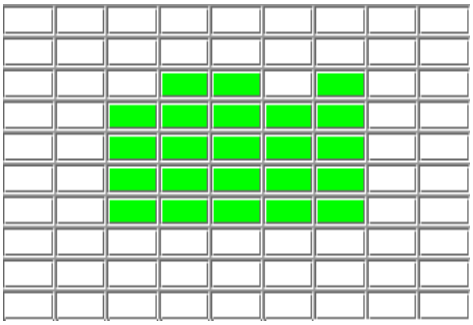

Figura 6.7: Solución gráfica para el cuadrado 5x5

Para hallar esta solución el algoritmo genético encontró un ADN de 16 genes que se muestran en la Tabla 6.10. Este ADN, aunque mayor que los mostrados anteriormente, sigue siendo asumible el analizarlo. Si se centra la atención en la Tabla 6.10 se detecta que los genes 7, 8, 11, 13 y 16 contienen información que no se expresará nunca. Esto se debe a que no son genes constitutivos, y además no tienen promotores que los activen. Aun así se han conservado durante la simulción para tener diversidad en los cruces y mutaciones.

El gen 14 también es casi imposible que llegue a activarse, ya que necesita una concentración muy alta de la proteína que figura en su parte activa.

Una vez planteado esto, se identifican los genes 2, 4 y 9 como los constitutivos del sistema. Esto implica que serán los que, en ausencia de las proteínas que contienen en la parte promotora, se expresarán siempre. Cabe destacar que, tanto el gen 2, como el

gen 4, son el mismo pero con distinto nivel de activación. Esto implica que, ante la ausencia de la proteína 0011 se expresen ambos. Si está presente dicha proteína, es probable que sólo se exprese el gen 2, mientras que, si su presencia es notable, ninguno de los dos se expresará. Además, resaltar que la producción de la proteína 0111 por estos dos genes puede provocar la desactivación del gen 4 ya que, la distancia Hamming entre la proteína 0011 (que inhibe al gen 4) y la 0111 no es alta y el umbral de activación es bajo, lo que facilita la activación como se explica en el punto 5.4.

Tabla 6.10: Genes necesarios para generar el cuadrado 5x5

Nº	Secuencia Promotora	Concentración Mínima	Constitutivo	Secuencia Generada	Especial
1	0010	13.5990	false	0110	
2	0011	59.2441	true	0111	
3	0111	36.8929	false	1001	Mitosis Derecha
4	0011	1.4519	true	0111	
5	0001	39.8667	false	0011	Mitosis Izquierda
6	0001	16.6105	false	0110	
7			false	1100	
8			false	0001	Mitosis Abajo
9	1101	58.6701	true	0010	
10	0000	42.2078	false	0110	
11			false	0100	
12	1110	23.8562	false	0101	Mitosis Arriba
13		56.1970	false	1010	
14	0101	91.6035	false	1100	
15	0011	65.6439	false	0011	Mitosis Izquierda
16			false	0110	

Teniendo, por tanto, las proteínas 0111 y 0010 en el sistema, se pueden activar también los genes 1 (directamente), 5, 6, 10 y 15 (por promotores parecidos a la proteína 0010: 0000, 0001 y 0011). Sin embargo, el gen 15 tiene una concentración mínima tan alta que no es probable que se active. La expresión del gen 5 provocaría que la célula se dividiese hacía la izquierda y la activación del gen 3 hacia la derecha. El crecimiento

hacia arriba se da por activación del gen 12 debido a proteínas parecidas a la 1011 que no se genera directamente.

El crecimiento del sistema celular se detiene en un cierto momento ya que, al pasar los ciclos, es más difícil conseguir concentraciones altas de las proteínas de las partes activas que permitan activar los genes de crecimiento.

Como se puede comprobar en la Figura 6.7 la solución no es perfecta. El mejor ajuste alcanzado para esta prueba fue de 2.0045. Cabe destacar que el algoritmo genético necesita 45 secciones para hacer uso de 36 para encontrar la solución. Finalmente sobre esta prueba se puede decir, además, que el tiempo que tardó en realizar las 1000 generaciones fue de aproximadamente unas 23 horas. Estos hechos plantearon un punto de inflexión en el desarrollo del modelo ya que, ni el tiempo, ni el resultado, eran los esperados. Como conclusiones de estas pruebas se puede destacar:

- Con los resultados de las pruebas partiendo de una solución previa y partiendo de poblaciones aleatorias. Se puede afirmar que, partir de una población previa, no garantiza el obtener los mejores resultados, si bien en media tienden a ser mejores que los obtenidos a partir de poblaciones aleatorias.
- A medida que el número de células aumenta su control se hace más complejo y requiere más información.
- Se requerían adaptaciones para hacer el modelo más ligero para afrontar pruebas con más células y que permitieran hacer desarrollos más complejos.

6.2.2.2 Aproximación con comunicaciones basadas en probabilidades

De las conclusiones de los apartados anteriores surgen dos ideas principales, la primera de ellas es que se hace preciso el hacer el modelo celular más ligero en términos de cómputo para poder ejecutar pruebas más complejas y, la segunda de las ideas, es la necesidad de incluir mecanismos que faciliten manejar el aumento de la complejidad del ADN. Para abordar estas dos ideas se plantea, en primer lugar, la simplificación de las comunicaciones entre células mediante el modelo basado en probabilidades (punto 5.9.2) que al reducir el número de elementos a ser manejados hace que el sistema sea más ligero. Por el otro lado, se volvió a examinar el modelo biológico y se decidió la adaptación del Operón (punto 5.5 de la presente tesis) que en la naturaleza es el mecanismo que fácilita a las células realizar tareas más complejas.

De las mejoras anteriores el primer punto a tratar sería sobre el sistema de comunicación entre las células ya que el basado en elementos discretos se hace sumamente pesado en términos de coste de cómputo. Así, se plantea ejecutar nuevamente las mismas pruebas que en el apartado 6.2.2.1, pero esta vez utilizando el modelo de comunicaciones basado en probabilidades explicado en el punto 5.9.2. Así mismo, también se incluyó el concepto de operón explicado en el punto 5.5. Por ello, la configuración queda como se muestra en la Tabla 6.11.

Como se puede ver, los parámetros son los mismos mostrados en la Tabla 6.1 para las pruebas del apartado 6.2.2.1, pero se ha cambiado el modelo de comunicación entre las células del sistema celular. En este caso se ha sustituido el sistema basado en elementos discretos por uno basado en probabilidades (véase el punto 5.9.2) y se ha establecido

un rango máximo para las comunicaciones de 3 unidades según una distancia Manhattan.

Tabla 6.11 Parámetros para la prueba

Sistema celular			
Parámetros	Valor	Mínimo probado	Máximo Probado
Modelo de Comunicación	**Basado en Probabilidades con distancia 3**	1	5
Uso del Operón	**Si con entre 2 y 5 genes**	2	10
Entorno	20x20 posiciones	10x10	40x40
Ciclos celulares antes del cálculo del fitness	50	5	50
Proteína para inducir la mitosis arriba	1010	-	-
Proteína para inducir la mitosis abajo	1000	-	-
Proteína para inducir la mitosis izquierda	1100	-	-
Proteína para inducir la mitosis derecha	0011	-	-
Proteína para inducir la apoptosis	0000	-	-
Número máximo de proteínas disponibles	16	8	32
Vida de las proteínas (ciclos celulares)	3	2	10
Algoritmo Genético			
Tamaño de la población	300	50	500
Número de generaciones	1000	300	2000
Operador de selección	Ruleta		
Tasa de cruce	80%	60%	90%
Probabilidad de mutación	20%	1%	30%

Como ya se comentó en el punto 5.9.2 el modelo cambia el cálculo de las concentraciones por una probabilidad de recepción de las proteínas. Esto provoca que, a veces, 2 ejecuciones de un mismo ADN puedan no dar el mismo resultado. Aún así, y a pesar del incremento de la complejidad de las soluciones, la ejecución de las pruebas es más rápida, como se verá en las pruebas realizadas a continuación, que era una de las cosas que se buscaba.

También se ha añadido el operón que fue explicado en 5.5. Comentar que, el número de genes que puede contener un operón se fijó entre 2 y 5, ya que tras diversas pruebas se observó que era el rango en el cual ofrecía mejores resultados.

Para comprobar el funcionamiento con las incorporaciones se ejecutaron los mismos test que en la sección 6.2.2.1 que se detallan a continuación. Además también se presentan algunas pruebas que no se habían podido realizar con la configuración utilizada en 6.2.2.1, pues por problemas de complejidad el sistema no era capaz de solventarlas en modo alguno.

6.2.2.2.1 Barra vertical de 5 elementos

Se repitió la misma prueba que la explicada en el apartado 6.2.2.1.1 pero con la configuración que se muestra en la Tabla 6.11. Se recuerda que la prueba consiste en generar una barra vertical de 5 células. Esta prueba se ejecutó sobre 10 poblaciones aleatorias distintas. El resultado fue que, en todos los casos, el algoritmo genético encontraba la solución a diferencia de la prueba 6.2.2.1.1 donde sólo en dos de las poblaciones se encontraba la solución. El resumen de estas pruebas puede verse en la Tabla 6.12.

Tabla 6.12 Resultados medios y mejor individuo para la barra vertical de 5 elementos

	Valor de ajuste	Numero de secciones utilizadas	Número de genes	Tiempo utilizado para la búsqueda
Media de los mejores individuos de las 10 poblaciones	0,0140	30,6	10,2	1h 15min
Mejor individuo	0,0022	22	4	1h 05min

Al examinar la Tabla 6.12 lo primero que se aprecia es que las soluciones contienen una complejidad similar en número de secciones, pero el tiempo para encontrar dichas soluciones es sensiblemente inferior. Así, por ejemplo, el mejor individuo fue encontrado en 1 hora y 5 minutos mientras que, la prueba 6.2.2.1.1, necesitó 4 horas y 2 minutos. Además, también llama la atención que, en media, requiere más secciones para codificar menos genes que la prueba 6.2.2.1.1. Mientras que en esa prueba se requerían 22,8 secciones para codificar 15,8 genes, en este caso se requieren 30,6 para codificar 10,2 genes.

El ADN del mejor individuo encontrado para esta prueba puede verse en la Tabla 6.13. Esta tabla muestra un identificador (ID) para referirse al gen u operón, las secuencias promotoras, la concentración mínima de cada secuencia promotora, si es un gen u operón constitutivo, la secuencia que generan los genes cuando se transcriben, si tienen un comportamiento asociado o no (Especial), y, en la última columna, si es un operón o el operón al que pertenece el gen en cuestión. Así mismo sobre la Tabla 6.13 se puede observar un primer elemento que es un operón, el cual contiene 2 genes (líneas 1.1 y 1.2) que sólo se podrán activar cuando se satisfagan las condiciones del operón. A continuación en la misma Tabla se encuentran dos genes (identificados con los ID 2 y 3) que completan el ADN necesario para dar la solución.

Comentar que los valores de concentración mínima menores de 1E-04 han sido aproximados como 0.

Tabla 6.13 Genes para la barra vertical de 5 elementos

ID	Secuencia Promotora	Concentración Mínima	Constitutivo	Secuencia Generada	Especial	Operon
1	0110-0101	≈0,0 – 70,12	NO			SI
1.1	-	-	SI	0110	NO	1
1.2	0001	2,14	SI	1000	Mitosis hacia Abajo	1
2	0111	98,34	NO	1010	Mitosis hacia Arriba	NO
3	-	-	SI	0110	-	NO

Lo primero que se puede observar de estos resultados es que ahora el algoritmo genético es capaz de encontrar soluciones mejores, más fácilmente y en menos tiempo. Por otro lado, el segundo de los hechos que saltan a la vista es que esta configuración utiliza un mayor número de genes y secciones para encontrar la solución al mismo problema. Por tanto como conclusión de esta prueba se puede argumentar:

- El sistema de comunicación por probabilidades (apartado 5.9.2) es capaz de encontrar la solución en menos tiempo para este problema.
- En media requiere de más secciones para codificar menos genes como efecto secundario de que ahora sean 3 tipos de secciones en lugar de 2, en la codificación del Algoritmo Genético (véase 5.10).

6.2.2.2.2 Generación de un cuadrado 3x3

Esta segunda prueba es la repetición de la prueba comentada en 6.2.2.1.3. En esta prueba se intentará busca un ADN capaz de generar un cuadrado 3x3 partiendo de una población aleatoria. Para ello se configura el algoritmo genético y el sistema celular como se muestra en la Tabla 6.11 y se ejecuta la prueba 10 veces, de las cuales encontró la solución en 4 ocasiones. El resumen de los resultados para estas pruebas puede verse en la Tabla 6.14

Tabla 6.14 Resultados para generar un cuadrado 3x3

	Valor de ajuste	Numero de secciones utilizadas	Número de genes	Tiempo utilizado para la búsqueda
Media de los mejores individuos de las 10 poblaciones	2,63	121,8	43,5	2h 37min
Mejor individuo	0,0352	152	52	2h 50min

Si se comparan los resultados con los mostrados en la Tabla 6.8 se pude ver fácilmente que los valores de ajuste son mejores en todos los casos y el tiempo empleado en la búsqueda es menor. Además, como en la prueba anterior, se constata que el algoritmo necesita más información ya que tanto el número de secciones utilizadas, como el número de genes, es mayor. Esta es la razón por la que el valor de ajuste del mejor individuo es peor que el de la prueba 6.2.2.1.3, ya que, en este caso, se usan 152 secciones para encontrar la solución. Si no se tuviesen en cuenta el número de secciones, en el cálculo del ajuste, el resultado sería incluso igual al mejor resultado de la prueba presentada en 6.2.2.1.3 partiendo de una población previa.

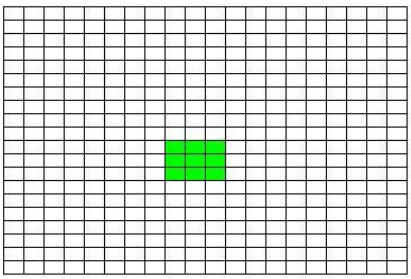

Figura 6.8 Resultado de generar un cuadrado 3x3

El resultado gráfico del mejor de los individuos puede verse en la Figura 6.8. Se puede observar aparentemente el mismo comportamiento que en la prueba descrita en

6.2.2.1.3, pero, internamente, el sistema celular utiliza muchos más datos ya que cuenta con 50 genes.

Las conclusiones que se pueden extraer tras esta prueba son:

- Que la inclusión del operón y del sistema de comunicación basado en probabilidades consigue en las pruebas realizadas, al menos, un nivel de ajuste igual a las pruebas anteriores, e incluso mejorando el ajuste en el caso medio.
- En las pruebas realizadas las incorporaciones consiguen que el sistema maneje más información y encuentre las soluciones en un tiempo sensiblemente inferior.

6.2.2.2.3 Generar un cuadrado 5x5

Finalmente entre las pruebas que se repitieron del modelo utilizado en 6.2.2.1 está la generación de un cuadrado 5x5, que se puede ver en 6.2.2.1.4. Para esta prueba se configuró el sistema con los parámetros que se muestran en la Tabla 6.11 y se utilizó la misma ecuación que en el resto de las pruebas.

La prueba fue ejecutada 10 veces de las cuales 2 obtuvieron la solución al problema. Los datos generales de las pruebas pueden verse en la Tabla 6.15

Tabla 6.15 Detalles de las pruebas para generar un cuadrado 5x5

	Valor de ajuste	Numero de secciones utilizadas	Número de genes	Tiempo utilizado para la búsqueda
Media de los mejores individuos de las 10 poblaciones	2,5427	270,4	42,4	6h 10 min
Mejor individuo	0,0580	280	47	10h 22min

La solución gráfica del mejor ADN encontrado para esta prueba se muestra en la Figura 6.9.

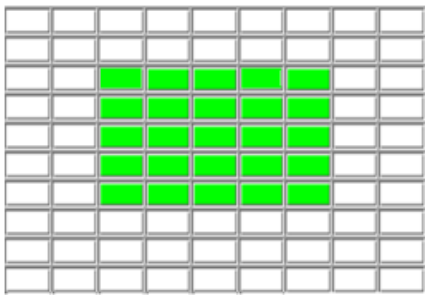

Figura 6.9: Solución del cuadrado 5x5 con las comunicaciones basadas en probabilidades

La solución tiene un valor de ajuste de 0,0580 ya que, como se ve, es capaz de generar el cuadrado, cosa que no se había logrado en la prueba 6.2.2.4, incluso con un número mayor de pruebas. Esta solución tiene el problema de salir descentrada y que utiliza muchas más secciones, en concreto 280 para generar 47 genes. Un punto importante de esta prueba es que el tiempo se redujo de las más de 23 horas utilizadas en la prueba que se muestra en 6.2.2.1.4 a un poco más de 10 horas. Lo que supone una importante reducción en tiempo.

Por tanto, como conclusiones se puede destacar:

- De esta, y de las otras pruebas, se puede afirmar que, el sistema de comunicaciones basado en probabilidades con el uso del operón es más rápido que el modelo basado en elementos discretos, como se comprueba por el tiempo que tarda en ejecutar las mismas pruebas, tanto para el mejor individuo, como en el caso medio.

- Además, el modelo celular puede tratar con pruebas que requieren más genes para solucionar el problema. Una de las razones es que este modelo de comunicación es más ligero, en términos de cómputo, que el modelo de comunicación basado en elementos discretos.

6.2.2.2.4 Influencia del Operón en la búsqueda

Como se comentó al principio del punto 6.2.2.2, se detectó la necesidad de introducir nuevos mecanismos en lo que era el modelo básico que se había desarrollado para realizar tareas más complejas y agilizar las búsquedas. Los dos mecanismos introducidos en este apartado son: la comunicación basada en probabilidades y el operón. La influencia del primero de los elementos parece más clara al reducir el número de elementos que deben ser manejados como se demuestra en las pruebas de los apartados anteriores. Lo que no resulta tan evidente es la influencia del operón en la mejora de los resultados y las búsquedas.

Para comprobar la influencia de este concepto se esperó hasta este punto ya que, para que fuese más clara, era necesario poder ejecutar una prueba con una cierta complejidad. Por tanto se ha escogido la prueba del cuadrado 5x5, la cual se ejecutará en las mismas condiciones salvo que en uno de los casos se usará el operón y en otro no. La configuración del resto de parámetros se estableció como se muestra en la Tabla 6.11 a excepción del operón que es el objetivo de esta prueba y el número máximo de generaciones que se fijó en 2000. El operón se establece de manera que puede contener entre 2 y 5 genes según lo explicado en el punto 5.5. Además, también se le estableció un límite de 20 horas al total de las generaciones. Con esta medida se espera utilizar el

mismo esfuerzo computacional en todas las pruebas, es decir, el mismo tiempo y en las mismas condiciones para todas las ejecuciones de la prueba.

Esta prueba fue ejecutada cinco veces con el operón y otras cinco veces sin el operón. En Figura 6.10 se muestra la evolución del valor de ajuste medio de las cinco ejecuciones iteración por iteración para cada uno de los casos. En esta gráfica se observa cómo la inclusión del operón mejora claramente los resultados obtenidos en media de las pruebas.

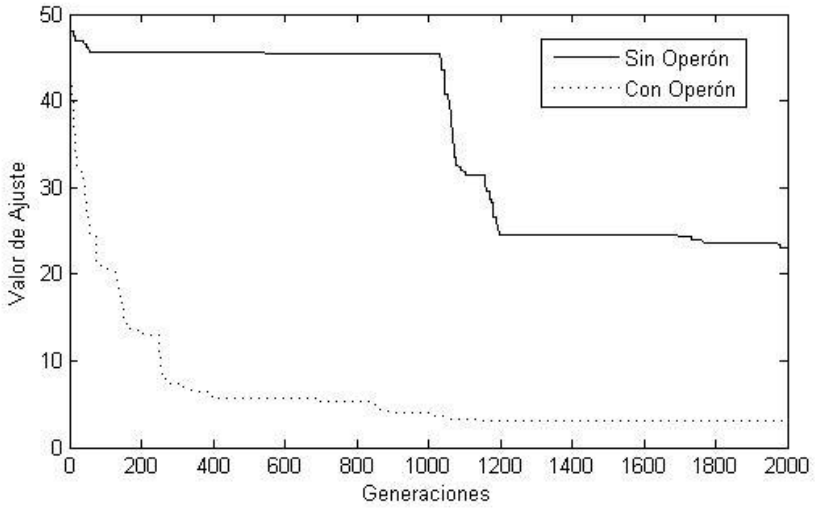

Figura 6.10: Evolución del valor de ajuste

Finalmente comentar que las pruebas que no utilizaban el operón tendían a terminar por superar el límite de tiempo establecido o bien por encontrar soluciones con un ajuste muy malo. Así, por término medio, las pruebas que incluyen el operón tienden a acabar entorno a las 10h y siempre mejoran los resultados de los individuos que son ejecutados sin el operón, como se muestra en la Tabla 6.16.

Comentar que muchas de las poblaciones acababan en un tiempo pequeño porque, al ser individuos de longitud variable, generan ADNs más pequeños que se ejecutan más rápido. Estas poblaciones tienden a acabar al superar el límite de generaciones, pero estas poblaciones no encuentran la solución.

Tabla 6.16 Datos de la ejecución de las distintas poblaciones

Identificador	Mejor Ajuste	Tiempo	Número Generaciones
Ind. Con Operon 1	2,5326	20 h 00min 00s	1321
Ind. Con Operon 2	8,9172	7h 00min 48s	2000
Ind. Con Operon 3	0,9000	6 h 44 min 48s	2000
Ind. Con Operon 4	0,9302	19h 14min 52s	2000
Ind. Con Operon 5	1,9183	11h 02min 26s	2000
Ind. Sin Operon 1	10,5000	20h 00min 00s	1810
Ind. Sin Operon 2	45,5000	00h 45min 23s	2000
Ind. Sin Operon 3	38,3000	00h 48min 21s	2000
Ind. Sin Operon 4	10,5328	16h 56 min 45s	2000
Ind. Sin Operon 5	10,5273	20h 00min 00s	1847

Como conclusiones de estas pruebas se puede extraer:

- El operón mejora el rendimiento tanto en tiempo de búsqueda como en calidad de las soluciones halladas.

6.2.2.2.5 Prueba de Generalización

Llegado a este punto se plantea comprobar cómo de generales son las soluciones encontradas por el sistema. Así, en este caso, se trata de comprobar si la solución que se halló en un entorno, con unas condiciones, sigue desarrollando la misma estructura cuando éstas se cambian. Para ello, se incluyen en el entorno determinadas posiciones que no pueden ser ocupadas por las células. En esta prueba se coge el mejor individuo de la prueba 6.2.2.2.3 que se coloca en una célula. En esta prueba no se evolucionará la

población sino que solamente se comprobará el funcionamiento del ADN en unas nuevas condicones. La configuración para la parte del sistema celular quedaría como en la Tabla 6.11.

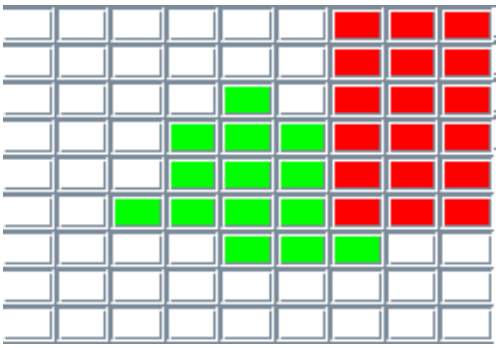

Figura 6.11: Prueba de desarrollo con un gran obstáculo

En estas condiciones lo primero que se intenta realizar es que se desarrolle un cuadrado similar al encontrado en la prueba 6.2.2.2.3, pero para la primera de las pruebas en este sentido se estableció que el cuarto superior derecho del entorno no fuese accesible mediante un gran obstáculo como se muestra en la Figura 6.11. En esta misma figura se puede ver como el ADN, aún a pesar de que no ha sido entrenado en estas condiciones, intenta desarrollar el cuadrado que había desarrollado previamente. Como era de esperar la solución no es perfecta, pero el sistema celular trata de solucionar el problema de la mejor manera. Así, se puede ver cómo, la solución desarrollada comete un error de 7 células en comparación con el cuadrado sin errores que se muestra en 6.2.2.2.3, pero tiene una estructura similar al cuadrado deseado. Los errores cometidos son principalmente debidos al bloqueo establecido en el entorno y a cierta descoordinación por la falta de estas células. Si bien la solución ofrecida es peor

que la que se puede ver en 6.2.2.2.3 era de esperar, al no tener ninguna clase de búsqueda en las condiciones planteadas y sólo usar el ADN encontrado en 6.2.2.2.3.

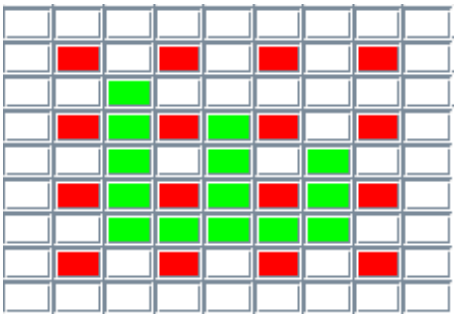

Figura 6.12: Prueba de desarrollo con múltiples obstáculos

La segunda de las pruebas realizadas, tiene como objetivo desarrollar el mismo cuadrado que la prueba anterior, con el mismo ADN que se encontró en 6.2.3.3 pero, esta vez, los obstáculos se establecen en el entorno ocupando un cuarto de las posiciones como posiciones aisladas rodeadas de posiciones válidas. Esta configuración puede verse en la Figura 6.12, donde también aparecen las posiciones rojas que son las que no pueden ser ocupadas por células. Las posiciones coloreadas en verde son las células desarrolladas por el ADN. Al igual que la otra prueba descrita en esta sección se puede ver como el ADN trata de desarrollar el cuadrado pero en un entorno con unos obstáculos para los que no ha sido entrenado. Aunque la solución no es perfecta se aproxima a la forma deseada. Esta solución comete un error de 7 células al igual que la prueba mostrada con anterioridad y que son 7 errores que no se cometína en 6.2.2.2.3. Estos errores pueden ser achacados, además de a las posiciones que no se pueden ocupar debido a los obstáculos, a cierta descoordinación por parte de las células por la falta de las señales de las células que deberían ocupar células en las posiciones de los obstáculos.

Por tanto como conclusión se extrae:

- El modelo celular es capaz de comportarse de manera robusta ante situaciones en las que no ha sido entrenada y da una respuesta consecuente a la solución codificada.

6.2.2.2.6 Incorporar elementos externos a las soluciones

La siguiente prueba a realizar con el sistema celular tenía como objetivo comprobar si las células artificiales podían incluir elementos en las estructuras desarrolladas, que no interaccionarían con ellas salvo para impedirles el paso. Así se plantea una prueba en que a partir de una célula, esta tiene que generar un tejido que envuelva un obstáculo, un cuadrado de 3x3, y generar un cuadrado de 5x5. Para esta prueba se configuró el sistema como se muestra en la Tabla 6.11. Además, la función de ajuste ha sido, como en el resto de los casos planteados en el apartado destinado a la generación de formas geométricas, una plantilla.

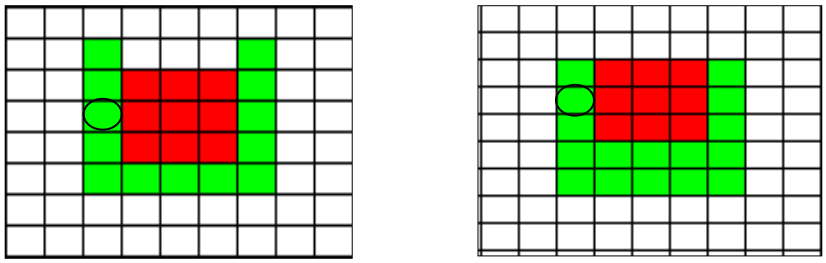

Figura 6.13: Dos pruebas de los ADNs para integrar un obstáculo en la solución

El resultado de esta búsqueda se puede ver en la Figura 6.13. En dicha imagen se muestran los tejidos desarrollados por dos individuos distintos que tratan de solventar el problema. Como se puede ver, cada uno de los individuos trata de solucionar el problema de una manera. El individuo que se muestra en la izquierda de la Figura 6.13

tiene un valor de ajuste de 3.0442. En este caso el individuo intenta rodear el obstáculo por todos los lados para minimizar el error que provoca el que el resultado esté desplazado del centro, pero como se puede ver le faltan 3 células para solucionar el problema. Este individuo codifica 103 de genes y para componerlos hicieron falta 442 secciones. Como se muestra en la Tabla 6.17.

Tabla 6.17 Datos de los mejores individuos para la prueba

	Valor de ajuste	Numero de secciones utilizadas	Número de genes	Tiempo utilizado para la búsqueda
Individuo Izquierda	3,0442	442	103	12h 36min
Individuo Derecha	0,0480	380	70	14h 24min

Por otro lado, en la misma Figura 6.13, se puede ver como el individuo que se muestra a la derecha ha utilizado una estrategia alternativa para solucionar el problema. En este caso el individuo ha optado por rodear solo 3 de los bordes del obstáculo y completar la figura con una hilera más de células. En este caso la solución utiliza más células de las estrictamente necesarias y está desplazada dos posiciones del centro del entorno, pero el individuo soluciona el problema planteado. El valor de ajuste para este es de 0.0480. Finalmente comentar que este individuo utiliza 380 secciones para codificar los 70 genes que emplea para generar esta solución, como también se muestra en la Tabla 6.17.

Como conclusiones de esta prueba:

- Se puede afirmar que el sistema se puede entrenar con elementos en el entorno que no son células propiamente.

- Además, puede hacerse que estos elementos se unan con el sistema celular a fin de que estén incluidos en la solución.

6.2.2.2.7 Prueba de Reducción

Llegado a este punto, se han visto pruebas de crecimiento de estructuras, pero es interesante saber si el modelo planteado, llegado el momento, podría realizar una reducción. Este es un proceso muy común en la naturaleza, por ejemplo, los vestigios de cola en la formación de los embriones humanos, que posteriormente desaparecen. Así en esta prueba se plantea establecer de salida, en vez de una célula, un cuadrado de 5x5 células. El objetivo sería quedarse sólo con el marco exterior del citado cuadrado. Para ello, la configuración del sistema quedó igual que la que se muestra en la Tabla 6.11.

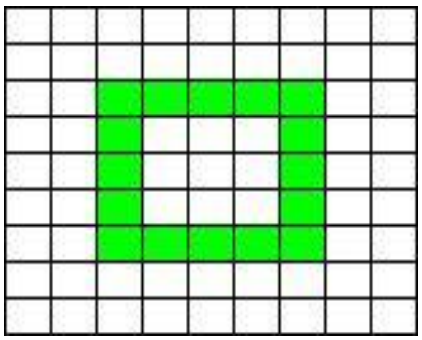

Figura 6.14: Prueba de reducción de una forma previa

La función de ajuste se estableció con la Eq. 6.1 y adaptando la plantilla utilizada a lo que se quería. En la Figura 6.14 se puede ver la solución a la que llegó el sistema. Como se puede ver, la solución es la ideal. En este caso el valor de ajuste del mejor individuo fue de 0.0016. Esta solución usa 16 secciones del algoritmo genético (véase el punto

5.10.1) para hallar los 2 genes que coordinan la obtención de esta forma y que se pueden consultar en la Tabla 6.18.

Tabla 6.18 Genes para quedarse sólo con los márgenes del cuadrado

ID	Secuencia Promotora	Concentración Mínima	Constitutivo	Secuencia Generada	Especial
1	0100 - 0010	99,98 – ≈0,0	Si	1111	
2	0000 – 1000 – 1011 – 0100 - 0000	≈0,0 - ≈0,0 – 40 - ≈0,0 - ≈0,0	No	0000	Apoptosis

Como conclusión de esta prueba se puede extraer que:

- El sistema está capacitado para ser entrenado para una reducción y no sólo para el crecimiento.

6.2.2.2.8 Prueba para una estructura estable pero con una comunicación mínima

Finalmente, en el apartado dedicado a las pruebas destinadas a la generación de formas geométricas, se presenta una prueba en la que el sistema celular tiene que generar una cruz en el entorno. La dificultad de esta prueba radica en que las células de los extremos reciben muy poca información ya que, el radio de comunicación se establecerá en un rango corto. El objetivo de esta prueba es comprobar si el sistema celular puede generar una estructura estable, pero con una cantidad mínima de información para la coordinación. Para ello se fija la configuración del sistema con la configuración de la Tabla 6.11.

Como en los casos anteriores, la única adaptación de la función de ajuste fue modificar la plantilla (véase 6.1) consecuentemente, para la forma deseada.

El resultado de esta búsqueda se puede ver en la Figura 6.15. Como se observa en la imagen, el ADN encontrado genera la forma deseada. El individuo que se muestra es el

mejor individuo encontrado que tiene un valor de ajuste de 0.0470. Este valor de ajuste se debe a que está desplazada 1 posición del centro del entorno y que usa 370 secciones (apartado 5.10.1) para generar 46 genes. Las conclusiones que se pueden extraer son:

- Incluso reduciendo la cantidad de información que reciben las células artificiales, el modelo es capaz de encontrar soluciones.
- La parte más importante del modelo es el ADN ya que, estando entrenado, es capaz de que, aún con muy poca información, desarrollar una solución al problema propuesto.

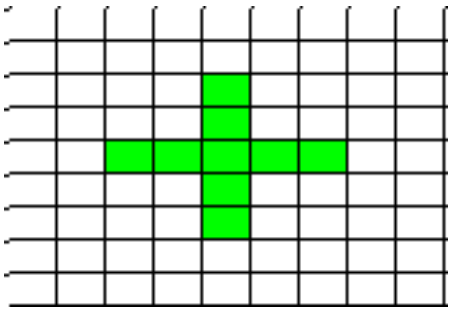

Figura 6.15: Prueba para generar una cruz con células

6.3 Procesado de Información

En el apartado 6.2 se han visto pruebas que se podrían considerar clásicas dentro de lo que es la Embriogénesis Artificial como la generación de formas sencillas. Además, algunas de ellas se han analizado detenidamente y han servido para mejorar el modelo celular propuesto.

Como ya se comentó en el capítulo 4, entre los objetivos de esta tesis está, el utilizar el modelo presentado en entornos lejos de las pruebas clásicas y que fuera aplicable en

distintos problemas. En concreto, las pruebas que se presentan en este apartado tienen como objeto mostrar la aplicación del modelo en algunos problemas sencillos de procesado de información.

Los problemas que se han seleccionado han sido: la simulación de una puerta lógica XOR y el problema de clasificación de las flores Iris (Fisher 1936).

6.3.1 Adaptaciones para el Procesado de Información

Para ejecutar estas pruebas el modelo propuesto necesitó dos sencillas adaptaciones. En primer lugar, se desactivaron las proteínas especiales que provocan la apoptosis o la mitosis de la célula (véase el punto 5.7). La razón para esto es que el análisis de la prueba tiene como objetivo concentrarse en las capacidades de procesado de una célula y, por tanto, permitir que se dividiese o muriese, sólo entorpecería el análisis en estas primeras fases del estudio.

La segunda adaptación que se requirió fue incluir dos nuevos elementos en el modelo, en concreto, puntos que permitiesen la emisión de una señal en forma de proteínas y puntos que permitiesen realizar una lectura de las proteínas de salida. El primero de los elementos, los puntos emisores, reciben el nombre de fuentes. Las fuentes son puntos que se establecen en el entorno y que, en cada ciclo celular, introducen una cantidad de proteína o lo que es lo mismo modifican la probabilidad de recibir una proteína del tipo que introducen (veasé 5.9.2). Estos puntos funcionan como las entradas del sistema, donde los valores que se pondrán en ellos son proteínas que representan cada uno de los conceptos. Los puntos lo que establecerían sería la probabilidad mínima en la posición que ocupan y, después, se calcularía la

probabilidad de recepción de esa proteína por parte de las células, igual que si la hubiese emitido otra célula cualquiera, como se explica en el apartado 5.9.2. El segundo de los elementos son puntos que reciben el nombre de sumideros. Estos sumideros funcionan como las salidas del sistema ya que su función es leer los datos existentes en el entorno que procedan de las células. Así, esta lectura establece la salida del sistema.

Tabla 6.19 Parámetros para configurar las pruebas de procesado de información

Sistema celular			
Parámetros	Valor	Mínimo probado	Máximo Probado
Modelo de Comunicación	Basado en Probabilidades con distancia máxima de 1	1	5
Uso del Operón	Si con entre 2 y 5 genes	2	9
Entorno	20x20 posiciones	10x10	40x40
Ciclos celulares antes del cálculo del fitness	10	5	50
Proteína para inducir la mitosis arriba	-	-	-
Proteína para inducir la mitosis abajo	-	-	-
Proteína para inducir la mitosis izquierda	-	-	-
Proteína para inducir la mitosis derecha	-	-	-
Proteína para inducir la apoptosis	-	-	-
Número máximo de proteínas disponibles	16	8	32
Vida de las proteínas (ciclos celulares)	3	2	10
Algoritmo Genético			
Tamaño de la población	50	50	500
Número de generaciones	1000	300	2000
Operador de selección	Ruleta		
Tasa de cruce	70%	60%	90%
Probabilidad de mutación	30%	1%	30%

Una vez establecido cómo se pueden configurar las entradas y salidas del sistema se puede hablar del concepto de patrón. Al igual que ocurre en las Redes de Neuronas Artificiales, se denomina patrón a una tupla de valores para las entradas del sistema y del cual se conoce el resultado deseado en la salida. Así, se puede establecer esos

valores en las entradas del sistema y establecer el error que comete en la salida respecto de la salida deseada para evaluar el ajuste del sistema.

El resto de parámetros del sistema se estableció como se muestra en la Tabla 6.19.

6.3.2 Pruebas de Procesado

Esta subsección contiene las pruebas realizadas con el modelo celular pero centrándose en la capacidad de procesado de información de las células. En concreto en esta sección se pueden encontrar las pruebas efectuadas para que la célula simule una puerta XOR y para resolver el problema de clasificación de flores Iris (Fisher 1936).

6.3.2.1 Simulación de una Puerta XOR

La primera de las pruebas que se presentan pretende saber si el modelo celular es capaz de realizar clasificaciones no lineales.

La prueba del XOR tiene su origen en las pruebas realizadas con los perceptrones en la década de los 60 (Minsky 1961). Esta es la prueba que detuvo el progreso de las redes de neuronas artificiales durante 20 años hasta la aparición de la regla de retropropagación del error (Rumelhart et al. 1986). Por tanto, parece un buen primer test para el modelo propuesto en este documento. Así, lo que se tratará de hacer es que una célula aprenda a discriminar dos señales de manera que sólo produzca una respuesta cuando una sólo una de ellas está activa. Para esta prueba la configuración del sistema queda como se puede ver en la Tabla 6.19.

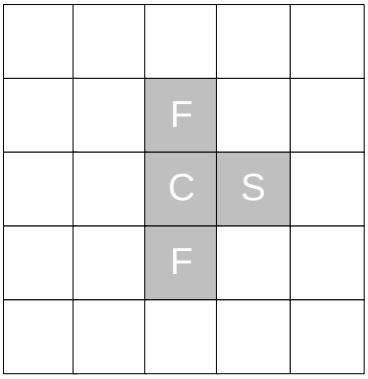

Figura 6.16: Configuración del entorno para probar el XOR

Para ejecutar las pruebas se ha configurado un entorno como el que se muestra en la Figura 6.16. En este entorno se pueden identificar dos fuentes, que se encuentran en las posiciones identificadas con la letra F. Cada una de estas fuentes introduce una proteína distinta.

Tabla 6.20: Casos del XOR y su traducción a proteínas

ID	Entrada 1	Entrada 2	Proteína 1	Proteína 2	Deseada
A	Falso	Falso	-	-	-
B	True	False	0001	-	0010
C	False	True		0011	0010
D	True	True	0001	0011	-

En concreto, para la prueba que se presenta en la Tabla 6.20, una de las fuentes introduce la proteína 0001 y la otra la 0011. Además también se ha colocado un sumidero, que se encuentra en la posición identificada con una S, que es donde se realizará la lectura de la salida del sistema. Más específicamente cuando sólo una de las proteínas que se introduce está presente, en el sumidero se desea recibir la proteína

0010. Finalmente en la posición que está identificada con una C, es donde se posicionará la célula con el conjunto de instrucciones halladas por el algoritmo genético.

Para saber si la respuesta es estable se deja la señal de entrada durante 10 ciclos celulares antes de proceder a la lectura de la salida en el sumidero. En la Tabla 6.20 se presenta la configuración para esta prueba. Para su ejecución se presentaría cada uno de los patrones de entrada en las fuentes y tras diez ciclos celulares se realizaría la lectura de la salida en el sumidero. Así, por ejemplo, el patrón A de la Tabla 6.20 no presentará ningún patrón en la en la entrada y tras 10 ciclos celulares se espera que en el sumidero no haya ninguna proteína. Por el contrario en los patrones B y C se activa una de las fuentes que emite la proteína de entrada correspondiente en cada caso y, tras diez ciclos celulares presentando esta proteína, se espera tener en el sumidero la proteína 0010 que es la deseada en este caso. Finalmente para el patrón D se activan las dos fuentes, donde cada una introduce una proteína distinta y, tras los 10 ciclos celulares, se espera que no haya nada en el sumidero.

$$fitness = \sum_{1}^{4} \sum_{i} f(i) \quad i \in \text{Proteinas}$$
$$f(i) = \begin{cases} r_i * \text{Hamming}(i,x) & \text{salida deseada } x \\ r_i & \text{sin salida deseada} \end{cases} \tag{6.2}$$

Para el cálculo del error cometido para un patrón se utiliza la función *f(i)*, que se muestra en la Eq. 6.2. Dicha función tiene dos formas según si existe una salida deseada para el patrón o, por el contrario, no debe recibirse nada. Además, se hace uso de la relación $r_{i,}$ que representa la relación entre las proteínas del tipo *i* respecto del

total de proteínas recibidas en el sumidero. En el caso de que la salida deseada sea nula, la función *f(i)* toma la forma de la relación r_i para la proteína de secuencia *i*. En el caso de que sí se desee una determinada proteína en la salida, la función toma la forma de la relaciones $r_{i,}$ pero esta vez ponderada por la distancia Hamming entre la proteína *i* y la proteína deseada. De esta manera se busca favorecer que si el sistema emite más proteínas que la deseada, al menos sean lo más parecidas a la secuencia deseada. La suma del valor de esta función para todas las posibles secuencias de proteínas en el sistema (en este caso 16) es lo que constituye la función de error para un patrón. Finalmente, la suma del valor de la función de error para todos los patrones que se muestran en la Tabla 6.20 constituye el valor de la función de ajuste, el cual se utiliza en el algoritmo genético para encontrar los genes que expresan ese comportamiento.

Tabla 6.21 Genes para realizar el XOR

ID	Secuencia Promotora	Concentración Mínima	Constitutivo	Secuencia Generada
1	0010	≈0.0	SI	1010
2	0010 -0010-0010	≈0.0- ≈0.0- ≈0.0	NO	1010
3	-	-	SI	0010

Esta prueba fue ejecutada 20 veces y en todos los casos el valor de ajuste alcanzado por el algoritmo genético fue de 0.0. Así que se puede afirmar que el sistema ha aprendido a realizar la operación de XOR para las secuencias mostradas en la Tabla 6.10. Para esta prueba, además, el individuo que se obtuvo de una menor longitud hacía uso de 7 secciones para generar 3 genes. Estos pueden parecer muchos pero hay que recordar que como se explica en 5.4.2 la operación lógica NOT no está codificada directamente si no que el sistema tiene que buscar rodeos para expresarla mediante genes constitutivos. Los genes necesarios para la solución se muestran en la Tabla 6.21.

Como se observa el ADN es bastante sencillo, el primero de los genes está activo a menos que esté presente cualquier proteína en el citoplasma. El segundo de los genes necesita 3 proteínas para expresarse y, finalmente, el gen 3 es un gen que funciona como señal de reloj, ya que introduce cada ciclo celular la misma proteína pase lo que pase. En el caso en que no se tenga ninguna probabilidad de recibir una proteína externa, todo lo producido por el gen 3 sirve para inhibir la expresión del gen 1. Con lo que el comportamiento externo de la célula es no emitir nada. En el caso de que sólo una de las proteínas de entrada tenga posibilidades de ser recibida cada 3 ciclos se activará el gen 2 lo que provocará que se introduzca la proteína 1010 que inhibirá el gen 1 en lugar de lo producido por el gen 3. Esta proteína repartirá su concentración con lo recibido del exterior en la célula. Si supera el límite establecido para las comunicaciones en el modelo de probabilidad, establecido en el 10%, será expulsada y establecerá una probabilidad de que se tenga en la salida la proteína producida por el gen 3. Finalmente en el caso que se tengan dos proteínas externas con probabilidades de ser recibidas por la célula, aunque la activación es como en el caso con una sola proteína con probabilidades, al recibir proteínas de dos fuentes hace que la concentración se divida mucho más, complicando el hecho de que se llegue a alcanzar el mínimo para comunicarse al exterior. Por lo que, la célula, se queda con las proteínas o emitirá muy pocas al exterior y tendrá una probabilidad muy baja de ser recibida en el sumidero de salida. Por tanto, el comportamiento externo es no emitir nada, que es el que se buscaba.

Además se hicieron otras pruebas con diferentes secuencias escogidas aleatoriamente para las entradas y la salida. El resultado fue que la secuencia no influía ya que el

ajuste alcanzado por el sistema siempre fue de 0.0. Como era de esperar, y como ya se había comentado en las pruebas 6.2.2.1.1 y 6.2.2.1.2, la operación es independiente de las secuencias seleccionas ya que lo que busca el sistema es que se cumplan determinados roles.

Finalmente comentar que se realizaron dos pruebas más para probar la generalidad de la solución encontrada, es decir, que si variando el número de fuentes de las proteínas de entrada seguía dando la misma respuesta. Estas pruebas se realizaron sólo con el patrón D de la Tabla 6.20, ya que es el único que tiene 2 entradas diferentes. Así, en la primera de las pruebas, se aumentó el número de fuentes de una de las secuencias quedando la primera de las secuencias con 2 fuentes y la segunda con 1. La segunda de las pruebas, aumenta todos los números de fuentes para comprobar que era escalable. De esta manera también se ejecutó la prueba con 3 fuentes para el primer tipo y 2 para el segundo. Estas pruebas fueron ejecutadas haciendo uso del mismo ADN encontrado para los casos sencillos. El resultado de estas pruebas fue que todo funcionaba exactamente igual y en el sistema no se registró ninguna salida en el sumidero. Por lo que se puede decir que, al menos en este problema, el resultado es escalable a un incremento de información.

Como conclusiones de estas pruebas se puede decir:

- El sistema celular puede ser adaptado para resolver problemas de procesado de información de un tipo similar al del XOR y no sólo para problemas de generación de formas.

- La inexistencia de un NOT explícito puede dificultar determinadas operaciones. Por tanto, debería de valorarse la inclusión del mismo como una posibilidad en la sección promotora.
- Finalmente, un mayor número de fuentes parece no tener influencia en la respuesta del sistema siempre que estos lleguen al elemento de procesado. Es decir, un incremento del número de fuentes, si no es para aportar datos a otros elementos de procesado que no eran influenciados por los primeros, parece no suponer ninguna ventaja sino, más bien, un simple incremento de la complejidad.

6.3.2.2. Problema de Clasificación de las Flores de Iris

El siguiente problema es uno de los más conocidos y estudiados como ejemplo de sistema no linealmente separable. El problema en cuestión es el de clasificación de las flores iris. Este problema fue inicialmente propuesto por Fisher en el año 36 (Fisher 1936) como ejemplo de análisis discriminante. El problema tiene como objeto clasificar un conjunto de flores de la especie Iris en las tres subclases existentes: Iris Setosa, Iris Versicolor e Iris Virgínica. Esta clasificación se efectúa haciendo uso de 4 parámetros morfológicos de las flores, medidos en milímetros, para 50 flores de cada una de las especies. Los parámetros de entrada son: longitud del pétalo, ancho del pétalo, longitud del sépalo y ancho del sépalo.

La razón para escoger este problema para probar el sistema, como primer problema complejo, fue porque está muy estudiado y documentado. En concreto, se trata de un problema que ha sido afrontado con diversas técnicas de Inteligencia Artificial.

Además, el conjunto de datos utilizado pertenece a la base de datos UCI (Asuncion & Newman 2007), una de las bases de datos más utilizadas por investigadores de todo el mundo. Este hecho provoca que los resultados de distintas técnicas sean comparables ya que todos usan los mismos datos. Por ejemplo, uno de los últimos trabajos que ha utilizado los mismos datos plantea utilizar Redes de Neuronas Artificiales diseñadas y entrenadas con Programación genética y consigue un ratio de acierto del 99.3% (Rivero 2007) (149 patrones de 150).

6.3.2.2.1 Clasificación con Todos los Patrones

En la primera de las pruebas efectuadas con el sistema se busca presentar los patrones en la entrada del sistema y que una célula decida a que tipo pertenece el patrón deseado. Para esto el sistema se configuró como se muestra en la Tabla 6.19.

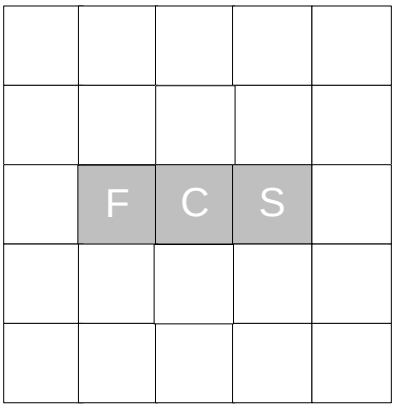

Figura 6.17: Configuración del entorno para la prueba de clasificación de flores Iris

Para la ejecución de las pruebas para el cálculo del valor de ajuste, además se configuró el entorno como se muestra en la Figura 6.17. En una de las posiciones se establecen 4 fuentes que emitirán una proteína distinta cada una de ellas. Cada una de estas proteínas identifica una de las mediciones mencionadas en el apartado 6.3.2.2. Las 4

fuentes han sido colocadas en la misma posición que está marcada con una F. Además también se situó una célula en el centro en la posición marcada con una C y un sumidero posicionado en la S para realizar la lectura de la salida tras 10 ciclos celulares.

Como se explica en el apartado 5.9.2 del presente documento, el modelo de comunicación utilizado está basado en una función, la cual establece la probabilidad de recibir una proteína del tipo emitido teniendo en cuenta el punto de emisión y la distancia a la que se encuentra el punto receptor. Esa probabilidad decrece con la distancia desde un valor máximo. Teniendo esto en cuenta para las fuentes del problema, se establece el valor máximo de la probabilidad desde las fuentes como el valor máximo normalizado de la variable en cuestión. Así, ante un valor más elevado en una determinada proteína las células tendrán más probabilidades de capturar proteínas de ese tipo. Por tanto, los valores de las mediciones para las cuatro variables de la flor Iris han sido normalizados entre 0 y 1. Esos valores normalizados se establecen para cada una de las fuentes como la probabilidad máxima durante los 10 ciclos antes de la lectura de la salida en el sumidero para un patrón determinado. Comentar además que, tanto el entorno, como el citoplasma de la célula, se vacían de proteínas tras cada patrón para no interferir en los resultados del siguiente.

Por tanto, por lo comentado anteriormente, está claro que hay que identificar 7 secuencias, 4 para representar las entradas y 3 para las salidas deseadas. En concreto en los ejemplos que se muestran se identificaron como:

- Longitud de sépalo: 0001
- Ancho de sépalo: 0010

- Longitud de pétalo: 0100
- Ancho de pétalo: 1000
- Tipo Iris Setosa: 1111
- Tipo Iris Versicolor: 1001
- Tipo Iris Virgínica: 1010

Finalmente para la configuración del sistema queda por fijar la función de ajuste. Para realizar el cálculo del error de un patrón se establecen los valores de entrada en las fuentes y, tras 10 ciclos celulares, se leen en el sumidero las proteínas recibidas. Con estas proteínas se crea una lista ordenada según su concentración. Esta lista se utiliza para determinar la posición que ocupa la proteína deseada para el patrón. Esta posición marcará el valor del error que ese patrón aportará al error total, que constituirá el valor de ajuste del ADN.

$$fitness = \sum_{i} f(i) \tag{6.3}$$

Valor de *f(i)*	Cuando se aplica
0	La proteína deseada es la 1ª de la lista ordenada de proteínas recibidas
0,5	La proteína deseada es la 2ª de la lista ordenada de proteínas recibidas
0,75	La proteína deseada es la 3ª de la lista ordenada de proteínas recibidas
0,85	La proteína deseada está en la lista ordenada de proteínas recibidas
1	La proteína deseada no figura entre las recibidas
10	No se ha recibido ninguna proteína en el sumidero

El ajuste de un ADN se mide utilizando la Eq. 6.3. Esta función representa la suma de los errores para cada patrón *i* del conjunto, donde el error de cada patrón se calcula con la función *f(i)*. Dicha función *f(i),* tal cual como se comentó antes, utiliza la lista ordenada de concentraciones de proteínas recibidas en el sumidero. Según la posición que ocupa en dicha lista la proteína deseada para cada patrón se añade un error como muestra la Eq. 6.3. Comentar que los valores de *f(i)* han sido hallados de manera empírica para este problema concreto. Finalmente comentar que se han establecido dos casos extra. El primero le asigna un error 1 al patrón si ninguna de las proteínas recibidas en el sumidero es la proteína deseada. El segundo establece un valor de penalización de 10 si, para un patrón, la célula no emite ninguna respuesta.

Con esta configuración el sistema halló la solución que se muestra en la Figura 6.18 y cuyo desglose de aciertos se muestra en la Tabla 6.22. En concreto, en la citada figura se muestra en la izquierda el conjunto de patrones acertados y en la derecha el conjunto de patrones que se han clasificado incorrectamente.

Tabla 6.22: Desglose de aciertos del sistema celular por los distintos tipos

Tipo	Número Total	Aciertos
Iris Setosa: 1111(*)	50	50
Iris Versicolor: 1001 (+)	50	46
Iris Virgínica: 1010(□)	50	44

Comentar que el sistema ha sido capaz de clasificar 139 patrones de 150 lo que supone un 92.67% del total. La Figura 6.18 muestra además los patrones clasificados por la

longitud del sépalo y el ancho del pétalo. Se realiza la representación de esta manera para facilitar el visionado de los patrones. La razón para escoger estas dos características es que estudios previos (Rivero 2007) han demostrado que estos son los dos parámetros más influyentes en la clasificación y, por tanto, son una referencia para la representación. Como se puede ver la clasificación se hace correctamente excepto en aquella zona en que dos de las clases entran en conflicto, donde ocurren algunos errores.

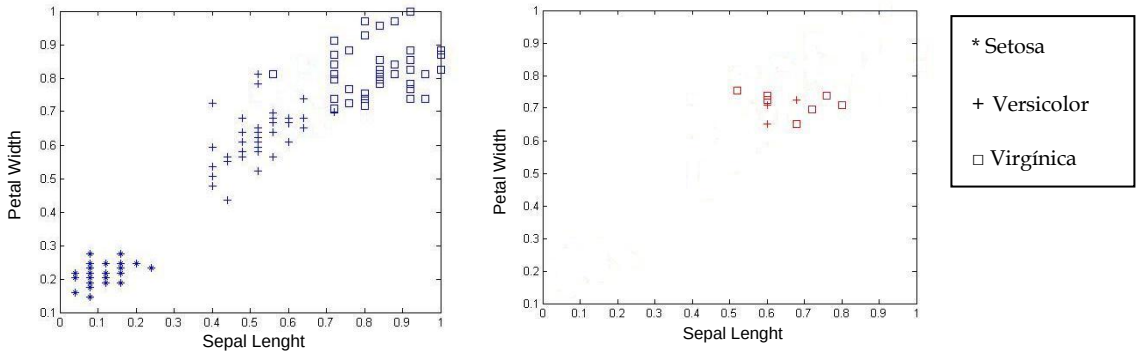

Figura 6.18: Clasificación de las flores Iris. A la izquierda los aciertos y a la derecha los fallos.

Finalmente, comentar que para estos resultados el individuo hallado utilizó 483 secciones del algoritmo genético (ver el punto 5.10.1) para codificar 108 genes. Comentar que estos resultados parecen bastante satisfactorios si se tiene en cuenta que se ha utilizado un solo elemento de procesado (célula).

Como conclusiones de esta prueba se puede extraer que:

- El sistema es apto para resolver problemas de clasificación de patrones con características similares a las del problema de las flores Iris.

- Los resultados que se observan con un único elemento de procesado son esperanzadores para que, con más elementos, se muestre un mejor comportamiento

6.3.2.2.2 Clasificación Utilizando Entrenamiento y Test

Una vez probado el sistema con todos los datos se planteó la pregunta de cómo de general serían las soluciones encontradas para un elemento de procesado. En concreto se planteó dividir el conjunto de datos de las flores Iris en dos subconjuntos el primero de ellos para el entrenamiento y el segundo para test.

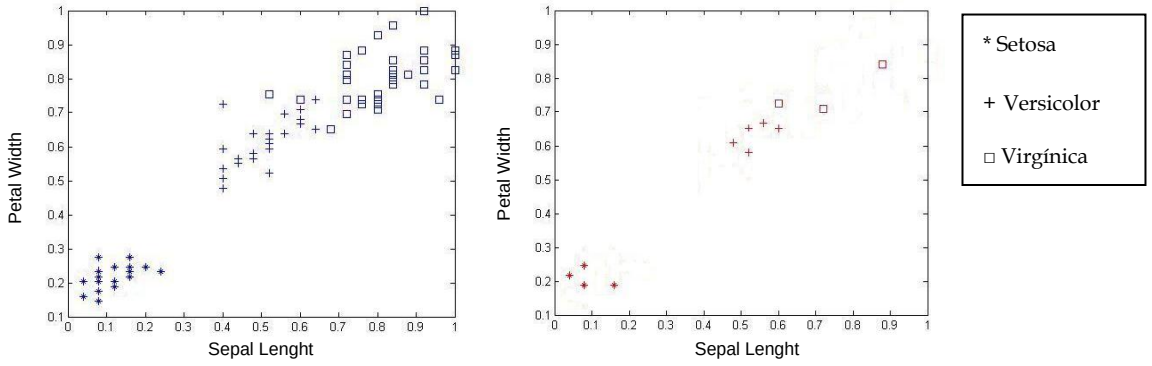

Figura 6.19: Aciertos (izquierda) y fallos (derecha) durante el entrenamiento con 110 patrones de Iris

Así, la división que se plantea reserva 110 patrones para el entrenamiento y 40 para el test. Repartidos de manera que las clases estén representadas de la manera más equitativa posible. La configuración del sistema es la misma que en el caso anterior (ver Tabla 6.19) incluida la función de ajuste.

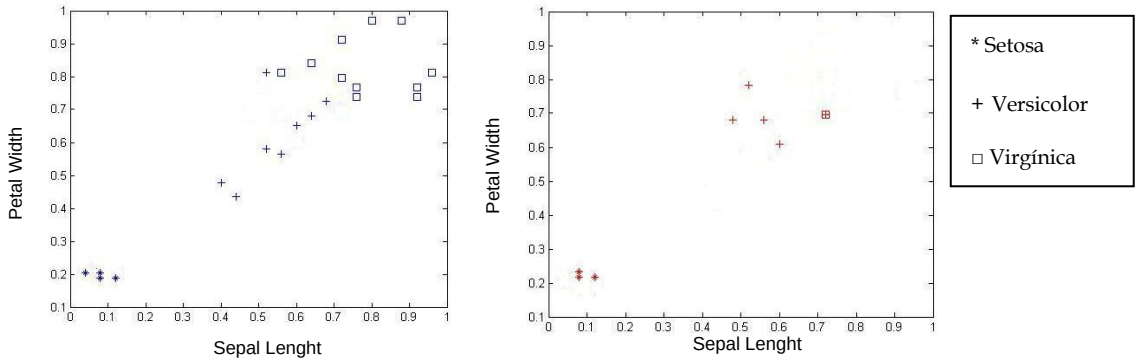

Figura 6.20: Aciertos (izquierda) y fallos (derecha) durante el test con 40 patrones de Iris

El mejor resultado de esta búsqueda fue un individuo que en la fase de entrenamiento tiene un 88.18% de acierto (97 de 110 patrones) como se puede ver en la Figura 6.19 y que, en la fase de test, se queda en un 75% de acierto (30 de 40 patrones), que se muestra en la Figura 6.20. En ambas figuras se muestran los patrones clasificados correctamente en la izquierda y los incorrectos en la derecha.

Además de esto se muestra en la Figura 6.21 la evolución del valor de ajuste para el mejor individuo tanto en entrenamiento (izquierda), como en test (derecha). En esta imagen se ve el clásico patrón para los problemas de aprendizaje máquina, donde el error en entrenamiento decae continuamente a lo largo de las iteraciones, en cambio el error en test tiende a caer pero con repuntes puntuales debido al posible sobre entrenamiento.

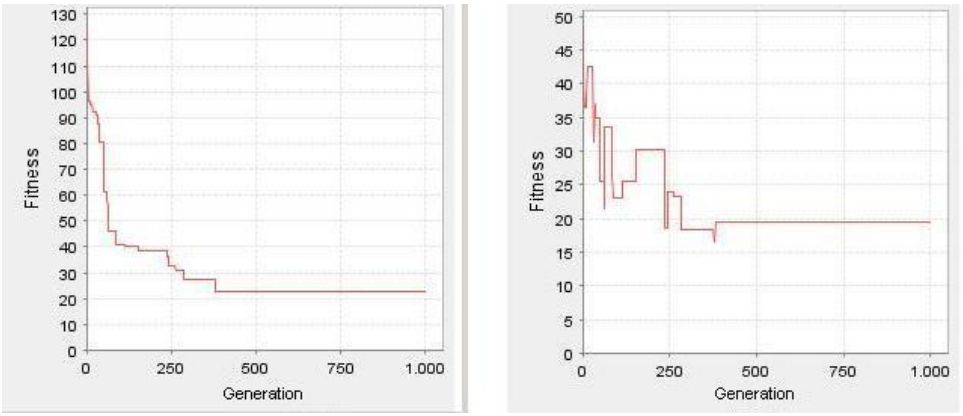

Figura 6.21: Evolución del valor de ajuste del mejor individuo durante el entrenamiento (izquierda) y el test (derecha)

Comentar, además, que el ADN está compuesto de 120 genes para lo que se necesitaron 543 secciones (ver el punto 5.10.1). Finalmente como conclusiones:

- El sistema puede ser entrenado de manera que sus soluciones son generalizables ante datos que no ha visto nunca.
- Se supone que el empeoramiento de los resultados es debido a la disminución del número de datos ya que, aunque es más complejo que el problema del XOR, sigue siendo un problema relativamente sencillo.
- Finalmente, comentar que se realizaron algunas pruebas con más de un elemento de procesado. Para ello se utilizó el mismo problema de clasificación de flores Iris con la configuración de la Tabla 6.19 de la que se obtuvieron algunos resultados preliminares. Estos resultados eran peores a todas luces que los obtenidos con un único elemento. Por esta razón se llegó a la conclusión de que el modelo planteado presenta carencias en la comunicación para la coordinación de varios elementos. Es por ello que se plantea como una futura

línea de investigación mejorar el estudio sobre el mantenimiento de las proteínas que está relacionado con el principio de conservación de la energía dentro del sistema.

6.4 Discusión

Finalmente y como resumen de las pruebas presentadas en este capítulo, se puede afirmar que los modelos creados para la embriogénesis artificial guardan un potencial mayor del que hasta ahora se ha explotado. Posiblemente, quede aun mucho camino que recorrer hasta que el ideal que persiguen se alcance, pero los indicios parecen esperanzadores.

En estas pruebas se ha comprobado que, un sistema celular como el presentado, busca una estructura que cumpla determinados roles para la solución de un problema determinado, como se puede ver en las pruebas 6.2.2. Así se demuestra que el conocimiento estaría realmente en la estructura de los genes y no en las relaciones concretas.

Además, una cosa que se ha comprobado durante la realización de las pruebas y que hizo cambiar el modelo de comunicación utilizado (véase las pruebas 6.2.2.1.3 y 6.2.2.1.4), es, por lo que se constata, que hay que llegar a un compromiso entre acercarse al modelo biológico y que el modelo resultante sea eficiente. Además, al realizar pruebas más complejas el sistema incrementa también en gran medida la complejidad del ADN que lo controla. Para facilitar esta tarea, se comprueba en

6.2.2.2.4, que el concepto de operón es una incorporación valiosa para un sistema como este.

Otras propiedades que se han probado sobre el sistema son: su capacidad de adaptación a un entorno donde no ha sido entrenado (prueba 6.2.2.2.5), que es capaz de incorporar elementos externos a la solución (prueba 6.2.2.2.6) y que es capaz de realizar figuras cuando la comunicación entre elementos es mínima (pruebas 6.2.2.2.7 y 6.2.2.2.8).

Lo anteriormente comentado es válido para las pruebas de generación de formas pero, además, este modelo propuesto ha sido utilizado en un campo completamente nuevo para modelos de embriogénesis artificial. Las pruebas desarrolladas en 6.3, muestran que el modelo propuesto puede ser utilizado en tareas de procesado de información. En concreto en esta tesis se ha demostrado que este tipo de modelos puede usarse para solucionar problemas como la clasificación de las flores iris. Es cierto que los resultados no son los mejores, pero ese no era el objetivo, si no sólo probar que es una técnica perfectamente aplicable.

Con este último conjunto de pruebas se han demostrado dos cosas:

- Que los modelos celulares son perfectamente aplicables a la resolución de problemas lejos de los clásicos de generación de formas.
- Que el potencial de este tipo de sistemas es mayor que el que se les presupone debido a las computaciones que puede realizar de manera interna.

En definitiva, y como conclusión general de las pruebas realizadas, las pruebas presentadas dan fe del potencial para crear un modelo a partir del modelo biológico de célula de modo que este modelo se autoorganice para dar la mejor respuesta a distintos problemas. Así, esta tesis se puede ver como un nuevo paso en ese ideal.

Estudia el pasado si quieres pronosticar el futuro

Confucio

Capítulo 7

Conclusiones

Este capítulo está destinado a discutir las conclusiones que se pueden extraer del desarrollo del modelo y las pruebas presentadas. De esta discusión se verá como los objetivos planteados en el Capítulo 4 se cumplen totalmente.

El primero de los objetivos de la tesis era realizar un estudio del modelo biológico de la célula y el proceso de morfogénesis, para estudiar su adaptación en un modelo artificial. Este objetivo ha sido cubierto ya que no sólo se ha procedido a estudiar este modelo si no que se ha diseñado un modelo artificial, completamente nuevo y distinto de los existentes en la literatura. Este hecho se atestigua con la descripción del mismo

en el Capítulo 5 de esta tesis donde se describe el modelo propuesto. En esa descripción se encuentran elementos como, por ejemplo, los operones que no han sido usados previamente en la literatura en su sentido biológico, de conjunto de genes que deben cumplir unas condiciones todos para activarse. Además, por lo mostrado en la prueba 6.2.2.2.4 pueden suponer una mejora en este tipo de sistemas.

El segundo de los objetivos propuestos era analizar los modelos existentes y tratar de afrontar pruebas que se planteasen en estos. Este objetivo queda cubierto por el estudio realizado previamente que se muestra en el Capítulo 2 y por las pruebas realizadas para generar estructuras que se encuentran en el apartado 6.2 de la tesis.

Para concluir el repaso de los objetivos de la tesis, el último punto era ampliar el estudio de este tipo de técnicas para comprobar su aplicación a distintos tipos de problemas. En concreto en 6.3 se ha probado que el sistema es perfectamente aplicable para resolver problemas de procesado de información como el de la clasificación de las flores Iris.

Además de estos objetivos también se ha profundizado en el conocimiento del funcionamiento de este tipo de modelos. Una de las conclusiones más importantes es que el conocimiento en este tipo de sistemas no reside en los valores que tenga un determinado gen si no en su relación con el resto. En esas conexiones es donde reside el verdadero conocimiento en estos sistemas. Es más, según se ha comprobado y explicado en 6.2.2.2.1, 6.2.2.2.2 y 6.2.2.2.3, lo que realmente busca el sistema es cumplir una serie de roles que determinan su comportamiento.

También se ha comprobado que en este tipo de sistemas se hace necesario un equilibrio entre similitud con el modelo biológico y eficiencia. Esto queda demostrado ante el cambio que se tuvo que hacer en el sistema de comunicación (ver 6.2.2.1 y 6.2.2.2) pues el sistema basado en elementos discretos, que se proponía, era demasiado pesado, en términos de cómputo, para abordar pruebas más complejas.

De la prueba 6.2.2.2.5 se puede extraer que el sistema es, en cierta medida, tolerante a pequeños errores y se comporta de manera flexible para tratar de dar una solución coherente. Además como se mostró en 6.2.2.2.6 no sólo es flexible en la respuesta si no que el sistema es capaz de incorporar obstáculos existentes para dar la solución.

Finalmente, de las pruebas de procesado se pueden extraer dos conclusiones, en primer lugar que el sistema celular puede realizar cualquier tipo de operación lógica lo que le posibilita potencialmente para resolver cualquier problema. Además, se ha comprobado que puede resolver problemas complejos. Aún se está empezando a estudiar el crear redes complejas de células para mejorar los resultados. Esto se debe principalmente al espacio de búsqueda tan grande en el que tiene que trabajar el sistema para coordinar las respuestas de las células.

Como conclusión general de todo el trabajo visto en esta tesis, se puede afirmar que, los sistemas inspirados en el sistema celular constituyen una base para el desarrollo de técnicas de Inteligencia Artificial con capacidades de autoorganización, control distribuido y procesado de información.

Study the past if you would define the future

Confucius

Chapter 7

Conclusions

This chapter is intended to discuss the conclusions that may be drawn from the model development and the assays presented. This discussion will show how the objectives outlined in Chapter 4 are fully met.

The first objective of the thesis was to carry out a study on the biological model of the cell and the morphogenesis process, in order to study their adaptation to an artificial model. This objective has been achieved since not only we have studied this model but also have we designed an artificial model, completely novel and different from those existing so far in the literature. This fact is proven by the presentation of its

development in Chapter 5 of this thesis which describes the proposed model. This characterisation includes elements such as the operons that have not been used previously in the literature in their biological sense, as a set of genes that must simultaneously meet certain conditions in order to get activated. Furthermore, as shown in the test 6.2.2.2.4, they can lead to an improvement in such systems.

The second objective proposed was to analyse the existing models and try to deal with any assays that may be necessary. This objective is met by the study previously carried out and shown in Chapter 2 and by the assays conducted in order to generate structures that are mentioned in section 6.2 of the thesis.

To conclude the review of the objectives in the thesis, the last issue was to extend the study of such techniques so to verify their application to various types of problems. Specifically, in section 6.3, it has been proved that the system is perfectly adequate to solve problems related to information processing such as the Iris flower classification.

Besides these objectives, we have also studied in depth the knowledge on the functioning of this type of models. One of the most important conclusions is that knowledge in such systems does not lie in the values that a particular gene has but in its relationship with the other genes. The real knowledge on these systems lies in these very connections. Moreover, as it has been shown and explained in sections 6.2.2.2.1, 6.2.2.2.2 and 6.2.2.2.3, what the system really seeks is to meet a set of roles that determine its behaviour.

It has also been found that in such systems it becomes necessary to ensure a balance between the similarity to the biological model and efficiency. This is evidenced

considering the change that had to be done in the communication system (see sections 6.2.2.1 and 6.2.2.2) since the system based on discrete series, previously proposed, turned out to be too difficult to use in terms of computation, to deal with more complex assays.

From the assay shown in section 6.2.2.2.5, it can be drawn the conclusion that the system is, somehow tolerant to small errors and behaves in a flexible way in order to find a coherent solution. Furthermore, as shown in section 6.2.2.2.6, the system is not only flexible when providing a response but it can incorporate real obstacles so that a solution could be found.

Finally, from the processing attempts, we can draw two conclusions: first, that the cellular system can perform any logical operation enabling it to potentially solve any problem. Furthermore, it has been proven to be effective in solving complex problems. We are witnessing the first steps towards the creation of complex networks of cells in order to improve the results. This fact is mainly due to such a large search space in which the system has to work in order to coordinate the responses of cells.

As a general conclusion of the entire work presented in this thesis, it can be stated that the systems inspired in the cellular system are a basis for the development of Artificial Intelligence techniques with abilities of self-organisation, distributed control and information processing.

El pasado es un prólogo

William Shakespeare

Capítulo 8

Futuros desarrollos

Existen multitud de puntos en esta tesis que abren nuevas vías de investigación. Esta tesis no es más que un primer paso en el desarrollo de un sistema más complejo que se acerque más al ideal ya enunciado.

Uno de los primeros puntos que requeriría la atención es mejorar la conservación de la energía en el modelo. Esta afirmación se refiere a mejorar el modelo por el cual se introducen proteínas y se consumen o degradan para que llegue a un equilibrio. Hasta el momento, el tiempo de vida ha sido el mismo para todas las proteínas y la distancia a la que pueden comunicarse, también. Esto no es realista ya que distintos compuestos

químicos tienen propiedades distintas. Introducir esto en el modelo puede mejorar sus prestaciones. Por tanto, se precisaría introducir la búsqueda de estos parámetros en el Algoritmo Genético. Esto desemboca en que el sistema tiene que buscar 2 conjuntos de datos distintos pero interrelacionados, por un lado, las reglas del ADN y, por el otro lado, las reacciones de las proteínas. El tener que buscar ambos conjuntos nos lleva inevitablemente a fijarnos en los sistemas de búsqueda cooperativa como los algoritmos genéticos cooperativos (Potter & Jong 2000).

Otro mecanismo que existe en el modelo biológico y no se ha implementado es la especialización. Llegado un determinado momento del desarrollo, las células biológicas bloquean su comportamiento y se especializan para realizar un trabajo más eficiente. Aunque poseen toda la información necesaria en el ADN, sólo utiliza parte de ella. Adaptar este proceso puede ayudar a desarrollar estructuras más complejas con el modelo artificial, ya que, como se vio en las pruebas, a medida que la complejidad del problema aumenta también lo hace la complejidad del ADN. Este mecanismo serviría para simplificar esa complejidad al dejar sólo una parte expresándose en cada célula.

Finalmente y entre las adaptaciones que se pueden hacer en el modelo de célula artificial estaría el que el entorno donde estas se desarrollan también evolucionase. Si se piensa en cómo funciona en la naturaleza el entorno de las células no suele ser un ambiente estéril como una placa de cristal en un laboratorio. El entorno suele tener sustancias y jugar un papel en el desarrollo del individuo. Evolucionar el entorno a su vez podría aportar nuevas funcionalidades. Por ejemplo, se podrían introducir elementos pasivos como se vio en 6.2.2.2.5, donde se colocaban una serie de posiciones

que no podían ser ocupados a modo de guía. Otros elementos que se podrían considerar son:

- Puertas Lógicas o funciones que combinasen las proteínas que pasen por la posición para dar nuevos compuestos.
- Funciones que aumenten la cantidad de una proteína si está presente en la posición a modo de catalizador.
- Puertas selectivas, que sólo dejan pasar determinado tipo de proteínas

Estos son algunas de las posibles modificaciones desde el punto de vista de mejorar el modelo de célula artificial propuesto en este documento.

Las Redes de Neuronas Artificiales son una gran herramienta para resolver los problemas para el cual han sido entrenadas, pero siempre que no sean problemas multimodales. Los problemas multimodales son aquellos problemas que presentan múltiples óptimos globales en su espacio de búsqueda. La resolución mediante Redes de Neuronas se hace muy complicada ya que estas están preparadas para dar una única respuesta para un problema. Si el problema tiene dos o más posibles soluciones validas, su uso, se hace muy complicado (Gestal et. al, 2009). Esta deficiencia no sería, en principio, un problema para un modelo celular ya que podría dar más de una salida a la vez. Además, los sistemas celulares artificiales podrían tratar con más de un problema simultáneamente ya fuesen independientes entre ellos o no.

Como se ha visto en las pruebas realizadas el sistema es capaz de afrontar tanto problemas de generación de formas como de procesado de información, pero aun quedaría por explorar la combinación de ambas facetas de manera conjunta que, como

ya se ha apuntado, es el ideal que se persigue. Esto extendería las capacidades de las redes de neuronas artificiales ya que, aunque existen modelos de redes de neuronas que se reconfiguran en tiempo de ejecución, su uso se restringe casi exclusivamente a la clasificación. Un modelo celular, en principio, no tendría esa laguna ya que la solventaría de manera natural por cómo funciona. Si la solución no es satisfactoria, se pueden utilizar los mecanismos de crecimiento para tratar de solventarlo. Estudiando los mecanismos que hacen que provocan el crecimiento de un tejido se pueden aplicar si la solución que está dando no es adecuada a fin de aumentar la complejidad del procesado llevado a cabo en tiempo de ejecución.

Uniendo las dos ideas se puede fijar como objetivo ideal para el futuro un modelo celular que evolucione para resolver uno o varios problemas simultáneamente. Esa evolución llevaría asociada la selección, tanto el número de elementos de procesado, como las conexiones entre los elementos más adecuados para la resolución, pero todo en tiempo de ejecución.

8.1 Posibles Aplicaciones

En otro orden de cosas, existen multitud de aplicaciones para los modelos resultantes. En el caso de generación de formas se podría plantear el uso del sistema para compresión de imagen de manera similar a como funciona la compresión fractal. Así, el sistema buscaría un ADN que generase un conjunto de células que representasen la imagen deseada.

Otra de las posibles aplicaciones es para el desarrollo y diseño de hardware autoconfigurable con capacidades de autoreparación, que es una de las aplicaciones más perseguidas desde hace años con este tipo de sistemas.

Si bien el caso de generación de formas está más estudiado, el campo de procesado de información aún está casi por completo inexplorado. Como se ha apuntado en esta tesis los modelos de embriogénesis artificial pueden ser adaptados para resolver problemas de procesado de información. En concreto se han visto problemas de clasificación, pero donde es posible que haya que profundizar más el estudio es en la predicción de series temporales. Las células llevan de por si incluido el concepto de tiempo con la definición de los ciclos celulares. Por tanto, pueden tener una ventaja a la hora del tratamiento de series temporales respecto de otras técnicas como las Redes de Neuronas que precisan de ciertos artificios para tal fin (ventanas temporales de introducción de datos, conexiones recurrentes, etc.).

En este último sentido una de las aplicaciones que se persigue es el desarrollo para controladores de robots. En este caso serían especialmente útiles ya que, por ejemplo en el caso de un robot bípedo se tienen varios problemas para el movimiento que deben ser resueltos de manera conjunta y que son dependientes del tiempo.

Algo he aprendido en mi larga vida: que toda nuestra ciencia contrastada con la realidad, es primitiva y pueril; y, sin embargo, es lo más valioso que tenemos

Albert Einstein

Referencias

Bibliografía

Alba E. & Tomassini M. (2002) "A connectioist machine for genetic hiiclimbing" *Kluwer Academic Publishers*.

Andersen T., Newman R. & Otter T. (2009) "Shape Homeostasis in Virtual Embryos". *Artificial Life, Vol.15(2) Pp. 161-183. MIT Press.*

Arbib M.A. (Ed.) (2002) "The handbook of brain theory and neural networks". *MIT Press*.

Asuncion, A. & Newman, D.J. (2007) "UCI Machine Learning Repository" *University of California, School of Information and Computer Science. http://www.ics.uci.edu/~mlearn/MLRepository.html*

Bentley P.J. (1999) "Digital Biology" *Simon and Shuster.*

Bentley P. J. & Kumar S. (1999) "The ways to grown designs: A comparison of embriogenies for an evolutionary design problem" *In Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 1999) Pp. 35-43. Morgan Kauffmann.*

Bogard J. C. & Pfeifer R. (2001) "Repeated structure and dissociation of genotypic and phenotypic complexity in artificial ontogeny" *In Spector L. Goodman E. D., Wu A., Langdon W. B., Voight V. M, Gen M, Sen S., Dorigo M., Pezeshk S., Garzon M. H. & Burke E. (Eds) Proceedings of the Genetic and Evolutionary Computation Conference. Pp. 829-836. Morgan Kaufmann.*

Booker L.B., Fogel D.B., Whitley D. L. & Angeline P. J. (1997) **"**Recombination". *In Bäck T., Fogel D. B. & Michalewicz Z. (Eds.) Handbook of Evolutionary Computation. Pp. 1-27. Institute of Physics Publishing and Oxford University Press.*

Cantú-Paz E. (2000) "selection Intensity in Genetic Algorithms with Generation Gaps" *Paper presented at the Genetic and Evolutionary Computation Conference 2004, Las Vegas, USA.*

Darwin C. (1859) "On the Origin of Species by Means of Natural Selection". *John Murray.*

Deb K. (1991) "Binary and floating-point function optimization using messy genetic algorithms" *Report nº 91004. University of Alabama.*

de Jong, K. A. (1975) "An analysis of the behavior of a class of geneticadaptative systems" *PhD. Thesis, University of Michigan.*

Dellaert, F. & Beer, A. D. (1996) "A developmental model for the evolution of complete autonomous agents", *In Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior, Pp. 393–401. MIT Press.*

Dorigo M., Maniezzo V. & Colorni A. (1996) "The Ant system: Optimization by a colony of cooperating agents". *IEEE Transactions on Systems, Man, and Cybernetics. Vol.26 (1) Pp. 1-13. IEEE Press.*

Doursat R. (2008) "Organically Grown Architectures: Creating Descentralized, Autonomous Systems by Embryomorphic Engineering" *In Würtz R. P. (Ed.) Organic Computing. Springer*

Drennan, B. & Beer, R. D. (2006) "Evolution of repressilators using a biologically-motivated model of gene expression", *in L. M. Rocha, L. S. Yaeger, M. A. Bedau, D. Floreano, R. L. Goldstone & A. Vespignani, eds, Artificial Life X, International Society for Artificial Life, Pp. 22–27. The MIT Press (Bradford Books)*

Eggenberger P. (1996) "Cell interactions as a control tool of developmental processes for evolutionary robotics", *In Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior, Pp. 440–448. MIT Press.*

Federici D. (2004). "Using methabolic stages to increase the evolvality of development" *In Proceedings of the Genetic and Evolutionary Computation Conference. ACM Press.*

Fernandez-Blanco E., Dorado J, Serantes J. A., Rivero D., Rabuñal J. R. (2009) "Artificial cells for information processing: Iris Classification. *Lecture Notes in Artificial Intelligence. Lecture Notes in Artificial Intelligence. Vol. 5777. Springer Press.*

Fernandez-Blanco E., Dorado J, Rabuñal J.R., Gestal M., Pedreira N. (2007) "A Computational Approach to Simple Structure Development". *Advances in Artificial Life edited by Almeida e Costa et. al.. Lecture Notes in Artificial Intelligence. Vol. 4648 Pp. 825-834. Springer Press.*

Fisher R. A. (1936) "The use of multiple measurements in taxonomics problems". *Annals of Eugenics. Vol. 7 Pp. 179-188.*

Fogel L. J., Owens A. J., Walsh M. J. (1966) "Artificial intelligence through simulated evolution" *Jonh Wiley & Sons*

Fontana A. (2009) "Epigenetic Tracking: Biological Implications" *In Porceedings of European Conference on Artificial Life 2009. Lecture Notes in Artificial Intelligence. Vol. 5777. Springer Press.*

Force A., Lybch M., Pickett F. B., Amores A., Lin Yan Y. & Postlethwait J. (1999) "Preservation of duplicate genes by complementary, degenerative mutations" *Genetics. Vol. 151. Pp. 1531-1545*

Gans C. &Northcut R. G. (1983) "Neural crest and the origin of vertebrates: A new head" *Science Vol.220 (4594). Pp 268-274.*

Gestal M., Rivero D., Fernandez-Blanco E., Rabuñal J.R. & Dorado J. (2010) "Two-Population Genetic Algorithm: an Approach to Improve

the Population Diversity" *In Proceeedings of International Conference on Agents and Artificial Intelligence 2010. Pp. 635-639*

Goldberg D.E. (1989) "Genetic algorithms in search, optimization, and machine learning" *Reading, Massachussetts. Addison-Wesley*

Goldberg D.E. & Deb K. (1991) "A Comparative Analysis of Selection Schemes Used in Genetic Algorithms" *Urbana. Vol. 51*

Gruau F. (1994) "Neural network synthesis using cellular encoding and the genetic algorithm" *Doctoral dissertation, Ecole Normale Superieure de Lyon, France.*

Harding S. & Banzhaf W. (2008) "Artificial Development" *In Würtz R. P. (Eds.) Organic Computing. Pp. 201-220. Springer Press.*

Holland J. H. (1975) "Adaptation in natural and Artificial Systems". *MIT Press.*

Horby G.S. & Pollack J.B. (2001) "Body-Brain co-evolution using L-sytems as a generative encoding" *In Spector L., Goodman E. D., Wu A., Langdon W. B., Voight V. M, Gen M, Sen S., Dorigo M., Pezeshk S., Garzon M. H. & Burke E. (Eds) Proceedings of the Genetic and Evolutionary Computation Conference. Morgan Kaufmann.*

Horby G.S. & Pollack J.B. (2002) "Creating high-level components with a generative representation for body-brain evolution" *Artificial Life Vol. 8(3).*

Jakobi N. (1995) "Harnessing morphogenesis" *In Proceedings of Information Processing in Cells and Tissues. Pp 29-41. University of Liverpool*

Joachimczak M. & Wrobel B. (2009) "Evolution of the morphology and patterning of artificial embryos: scaling the tricolor problem to the third dimension" *In Proceedings of European Conference on Artificial Life 2009. Lecture Notes in Artificail Intelligence. Vol. 5777. SpringerPress.*

Kaneko K. (2006) "Life: an introduction to Complex Systems Biology". Springer-Verlag.

Kaneko K. & Furushawa C. (1998) "Emergence of multicellular organisms with dynamic differenciation and special pattern". *Artificial Life 4 (1).*

Kauffman S. A. (1969) "Metabolic stability and epigenesist in randomly constructed genetic nets". *Journal of Theorical Biology. Vol. 22 Pp. 437-467. Elsevier Press.*

Kauffman S. A. (1993) "The origins of order" *Oxford University Press.*

Kephart J. O., & Chess D. M. (2003) "The vision of Autonomic Computing" *IEEE Computer Magazine, January 2003. (Pp. 41-50) IEEE Press.*

Kennedy, J. & Eberhart R. (1995) "Particle Swarm Optimization". *In Proceedings of IEEE International Conference on Neural Networks IV, Piscataway, NJ. Vol.4 Pp. 1942-1948. IEEE Press.*

Kitano, H. (1990) "Designing neural networks using genetic algorithms with graph generation system", *Complex Systems Journal 4, Pp. 461–476.*

Kitano H. (1994) "Evolution of metabolism for morphogenesis" *In proceedings of Artificial Life IV. MIT Press.*

Kumar S. & Bentley P.J. (Eds.) (2003) "On Growth, Form and Computers" *Elsevier Academic Press.*

Kumar S. (2004a) "A developmental biology inspired approach to robot control". *In Proceedings of Artificial Life IX. MIT Press.*

Kumar S. (2004b) "Investigating computational model for the construction of shape and form" *PhD Thesis. University Collage of London*

Kuyucu T., Trefzer M. A., Miller J. F., Tyrrell A. M. (2009) "On the Properties of Artificial development and Its Use in Evolvable Hardware" *In Proceeddings of the IEEE Symposium on Artificial Life 2009. Pp. 108-115. IEEE Press.*

Lall S. & Patel N. (2001) "Conservation and divergence in molecular mechanisms of axis formation" *Annual Review of Genetics. Vol. 35 Pp. 407-447*

Lindenmayer. **A. (1968)** "Mathematical models for cellular interaction in development: Part I and II". *Journal of Theorical Biology. Vol. 18 pp. 280-315*

Lipson H. & Pollack J. B. (2000) "Automatic design and manufacture of robotic lifeforms". *Nature Vol. 406 Pp. 974-978*

Marin E., Jeffries G. S., Komiyama T., Zhu H. &Luo L. (2002) "Representation of the glomerular olfactory map in the Drosophila brain" *Cell Vol. 109(2) Pp. 243-255.*

McCullogh W.S. & Pitts W. (1943) "A logical calculus of the ideas immanent in nervous activity". *Bulletin of Mathematical Biology Vol.5 (4) Pp. 115-133. Springer Press.*

Miller, J. F. (2003) "Evolving developmental programs for adaptation, morphogenesis, and self-repair", *in Advances in Artificial Life, 7th European Conference, ECAL'03, Vol. 2801 of Lecture Notes in Artificial Intelligence, Pp. 256–265. Springer Press.*

Minsky M. (1961) "Steps toward Artificial Intelligence" *In Proceedings of the IRE 49(1) Pp. 8-30.*

Mjolsness, E., Sharp, D. H. & Reinitz, J. (1991) "A connectionist model of development". *Journal of Theoretical Biology 152(4),Pp. 429–453.*

Nagl S.B. (2001) "Can correlated mutations in protein domain families be used for protein desigh?" *Briefings in Bioinformatics Vol.2(3) Pp. 279-288. Oxford University Press.*

Nijhout H. F. & Emlen D. J. (1998) "Competition among boty parts in the development and evolution of insects morphology". *Proceedings of the National Academy of Science of the USA. Vol. 95. Pp. 3685-3689.*

Nolfi S. & Parisi D. (1991) "Growing neural networks" *Technical Report PCIA-91-15. Rome Institute of Psichology, C. N. R.*

Perea G. & Araque A. (2007) "Astrocytes Potentiate Transmitter Release at Single Hippocampal Synapses." *Science, Vol. 317(5841), Pp. 1083 – 1086.*

Porto A., Araque A., Rabuñal J.R., Dorado J. & Pazos A (2007) "A New Hybrid Evolutionary Mechanism Base on Unsupervised Learning

for Connectionist Systems". *NeurocomputingVol. 70 Pp. 2799-2808. Elsevier B.V.*

Potter M. A. & Jong K. A. D. (2000) "Cooperative coevolution: an architecture for evolving coadapted subcomponents" *IEEE Transactions on Evolutionary Computation Vol. 8(1) Pp. 1–29. IEEE Press*.

Prusinkiewicz P. & Lindenmayer A. (1990) "The algorithmic beuty of plants" *Springer-Verlag*

Rivero D. (2007) "Desarrollo y simplificación de Redes de Neuronas Artificiales mediante el uso de Técnicas de Computatión Evolutiva" *Tesis Doctoral. Universidad de A Coruña.*

Rumelhart D. E., Hilton G. E. & Williams R. J. (1986) "Learning internal representations by error propagation" *In Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1. Pp. 318-362. MIT Press.*

Schlitt, T. & Brazma, A. (2007) "Current approaches to gene regulatory network modeling", *BMC Bioinformatics 8(6)*.

Stanley K. (2008) "Generative and Developmental Systems" *In Proceedings of Genetic and Genetic and Evolutionary Computation Conference (GECCO 1999) Pp.2849-2863. ACM Press.*

Stanley K., D'Ambrosio D. B. & Gauci J. (2009) "A Hypercube-Based Indirect Encoding for Evolving Large-Scale Neural Networks" *Artificial Life, vol. 15(2). In Press.*

Stanley K. & Miikkulainen R. (2003). "A taxonomy for artificial embryogeny" *In Proceedings of Artificial Life IX. Pp. 93-130. MIT Press.*

Taylor, T. (2004) "A Genetic Regulatory Network-Inspired Real-Time Controller for a Group of Underwater Robots". *In Intelligent Autonomous Systems 8, Pp. 403–412. IOS Press.*

Tufte G. & Haddow P. (2005) "Towards developmental on a silicon based cellular computer machine" *Natural Computing. Vol. 4(4)Pp. 387-416.*

Tufte G. & Haddow P. (2008) "Achieving environmental tolerance through the initiation and exploitation of external information" *In Proceedings of IEEE Conference on Evolutionary Computation 2007. Pp 2485-2492. IEEE Press.*

Turing A. (1952) "The chemical basis of morphogenesis" *Philosophical transactions of the Royal Society B. Vol. 237 Pp. 37-72.*

Voss S. R. & Schaffer H. B. (1997) "Adaptative evolution via a mayor gene effect: Paedomorphosis in the Mexican axolotl" *Proccedings of the National Academy of Science of the USA. Vol. 94. Pp. 14185-14189.*

Waddington C. H. (1942) "Canalization of development and the inheritance of adquired characters" *Nature Vol. 150(563).*

Wallace A.R. (1855) "On the law which has regulated the introduction of new species". *Annals and Magazine of Natural History. Vol. 16 (2nd series) Pp. 184-196.*

Wright A.H. (1991) "Genetic algorithms for real parameter optimization". *Paper presented at the Foundations of Genetic Algorithms.*

Wolfram S. (1994) "Cellular Automata and Complexity – Collected Papers". *Westview Press.*