The Influence of Cell Type on Artificial Development

John Maher, Fearghal Morgan, Colm O'Riordan and Brian McGinley

National University of Ireland Galway, Ireland
john.maher@nuigalway.ie

Abstract

Two variants of biologically inspired cell model, namely eukaryotic (containing a nucleus) and prokaryotic(without a nucleus) are compared in this research. Experiments are designed to provide an understanding of how the evolved regulation of protein transport to and from the nucleus of the eukaryotic type cell gives rise to complex temporal dynamics that are not achievable in a prokaryoticcell.

A novel system of protein movement based on the process of nucleocytoplasmic transport observed in the biological eukaryotic cell is proposed. Nucleocytoplasmic transport is considered by biologists to be one of the most important factors when determining the developmental trajectory of a cell, as it allows for additional control of transcription factors entering the nucleus, thereby regulating gene activity.

Experiments contrast the ability of both cell models to generate protein patterns within the cytoplasm. Results demonstrate that the additional cell complexity of the eukaryotic does not impede the Gene Regulatory Networks control. For increasingly difficult tasks requiring precise temporal control the performance of the eukaryotic cell model outperforms the prokaryoticcell model. In addition, results demonstrate that the second level of regulation introduced by the transport process within the eukaryotic cell allows very precise control of gene activity and provides the EA with a source of heterochronic control not possible in prokaryotic-type cells.

Introduction

Cells are the fundamental building blocks of all biological life. Two distinct groups of cell exist, eukaryotic and prokaryotic. Eukaryotic cells are distinguished by the separation of the cell into compartments, the most pronounced compartment, the nucleus contains the genome (Figure 1). The genome is decomposed into genes, which encode the blueprint for creating the organism. During the process of development gene expression results in the creation of proteins levels within the cells. Proteins within the cell direct and dictate the cell fate and define its final role within the organism.

In order to enable gene expression, specialised proteins termed *Transcription Factors* (TF) must bind the gene cis-regulation sites. In response to TF binding, the gene expres-
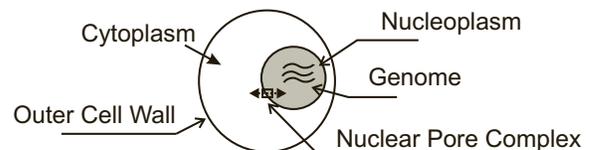

Figure 1: Architecture of the eukaryotic cell model. The nucleus is composed of the Nucleoplasm, Genome, and Nuclear Pore Complex (NPC)

sion rate is regulated, influencing protein levels within the cell. The presence of a nucleus within eukaryotic cells (Figure 1) restricts direct entry of transcription factors to the nucleus. Transport across into the nucleus through the *Nuclear Pore Complex* (NPC) is enabled by specialised *chaperons* proteins binding to TF proteins. The chaperon proteins are subdivided into two categories *importins* which enable import to the nucleus, and *exportins* which enable export from nucleus to cytoplasm across the NPC. *Importins* bind specific sites on the TF, named *Nuclear Localisation Sequence* (NLS). Export from the nucleus is enabled by the binding of *exportin* to *Nuclear Export Sequence* (NES) associated with the TF. This network of interacting genes and proteins is known as the Gene Regulatory Network (GRN).

Modifications to the relative timing of developmental events are termed *heterochronic*. Lee and Hannink (2003) report that the control of protein entry to the nucleus provides a powerful mechanism for the temporal regulation of gene expression. West-Eberhard (2003) highlights the importance of heterochronic control as a source of phenotypic novelty

> "if I could control the time of gene action I could cause the fertilised snail egg to develop in an elephant"

Heterochronic change is not a developmental process but rather an evolutionary process (West-Eberhard, 2003). Modifications to the timing of events are heritable and must be somehow linked to the genetic encoding.

Artificial developmental systems have been introduced as a technique aimed at increasing the scalability of Evolutionary Algorithms (EA) (Haddow and Hoye, 2007). These

systems seek to solve the problem of scale by replacing the linear genotype-phenotype mapping, with a non-linear mapping. However, developmental mappings have a low degree of evolvability caused by high degrees of epistatic interactions at the genotypic level (Van Remortel et al., 2003). Enhancing the ability of the genome to allow for heterochronic mutations increases the number of successful genotypes (Stanley and Miikkulainen, 2003) by offering a variety of paths to each successful phenotype.

This paper presents an Artificial EvoDevo (AED) platform capable modelling evolutionary and developmental processes within biologically plausible eukaryotic and prokaryoticcells. The prokaryoticcell model is based on the work of (Kumar and Bentley, 2003). Experiments compare the ability of the evolved GRNs within the cell models to control gene expression for a number of static, periodic and aperiodic objectives. Results demonstrate an increased evolvability of eukaryotic cells compared to the prokaryotic-cell model. Analysis of gene activity within the GRN shows that the transport process is instrumental in increasing evolvability by providing efficient hetrochronic control of gene activity.

The structure of the paper is as follows: Section 2 provides a review of the use of developmental processes by the ALIFE community. Section 3 describes the Artificial EvoDevo (AED) platform developed, the eukaryotic cell model and its associated transport process. Section 4 describes a series of experiments and analyses the GRN dynamics of both eukaryotic and prokaryoticcell models. Section 5 concludes the paper.

Background and Existing Research

This section provides a review of the use of developmental processes by the ALIFE community. Much of the research on developmental mappings has been motivated by the fact that the process of development is seen as a possible solution to the problem of scale in Evolutionary Algorithms (EA) (Bentley and Kumar, 1999), (Haddow and Hoye, 2007). By combining EAs with a developmental mapping between genotype and phenotype, the linear relationship between both is removed. Introducing developmental mappings also reduces the causality between genotype and phenotype spaces, which can reduce evolvability since regions of the search space become unreachable (Roggen and Federici, 2004).

The process of development is primarily a temporal one, where development starts from a single point and over time expands into a series of parallel pathways (Raff, 1996). Temporally shifting these processes relative to each other can give rise to phenotypic novelty. These shifts in the timing of events are termed *heterochronic mutations*, and must be heritable between generations. Efforts by the ALIFE community to identify sources of heterochrony within developmental encoding are limited. Matos et al. (2009) adapts

the framework proposed by Albrech et al. to quantify the degree of heterochrony achievable by both grammar based and cellular ontogenies. At the biologically plausible GRN level, Banzhaf and Miller (2004) illustrate a possible genetic mechanism by which heterchrony can be enacted. This is due to an encoding of time and strength of gene expression into the strength of interaction between transcription factors and the genes cis sites. Kumar and Bentley (2003) suggest that the modification to the diffusion rate of signalling proteins can also provide a degree of heterochrony.

The developmental model proposed by Kumar and Bentley (2003) is the primary source of inspiration for the AED platform described in this paper. Using a prokaryoticcell model (Kumar and Bentley, 2003) demonstrate how intricate gene regulatory networks can be evolved to establish and control protein concentrations within a single cell. Furthermore they demonstrate the ability of the system to evolve 3D multicellular spherical morphologies.

This paper recreates the work of Kumar and extends it by introducing a eukaryotic type cell within the AED platform. Experiments contrast the evolved developmental processes within both cells by comparing the ability of the GRN to regulate protein concentrations for a variety of increasingly difficult static, dynamic and aperiodic objectives.

The Artificial EvoDevo (AED) Platform

This section describes the developed Artificial EvoDevo (AED) platform, the eukaryotic cell model and its associated transport process. The section is decomposed into the following subsections - AED architecture and configuration, development and the cell cycle, protein model and classification, protein transport, mechanics of gene expression and the evolved genome structure.

The developmental algorithm within the AED platform captures the concepts of genes, proteins and cells. Similar to its biological counterpart, development proceeds along a time-line, where as a result of the GRN activity protein levels are established within the cell.

AED Architecture and Configuration

The AED comprises two main components (Figure 2) the Genetic Algorithm (GA) and developmental algorithm. Evolved genomes are supplied by the Genetic Algorithm (GA) to the developmental algorithm. The genome is then developed by placing it inside a user selected cell type (eukaryotic or prokaryotic) and returns a fitness to the GA.

The user configures the AED via a configuration file (Table 1). In order to start the developmental process, maternal proteins are placed inside the cell. The biological counterpart of this process is fertilisation of the embryo. Seeding involves placing a single TF inside the cytoplasm. In the reported experiments this has been arbitrarily chosen as protein 0 with a concentration of 0.5. Eukaryotic type cells
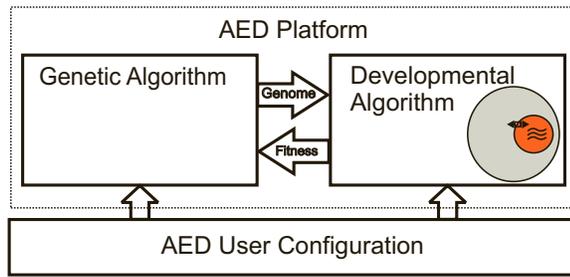
Figure 2: Architecture of the AED platform

are also seeded with *importins* (IM) and *exportins* (EX) type proteins at a concentration of 0.5.

Development and The Cell Cycle

Having seeded the cell with maternal proteins evolved genomes are developed by executing a *cell cycle* Figure 3. The developmental process is executed for a user defined number of *steps* (Table 1a), by iterating the cell cycle shown in Figure 3. At the end of development, protein levels within the cell cytoplasm are used to determine the genotype fitness. The calculated fitness value is subsequently fed back to the GA (Figure 2), where the corresponding genome is then subject to evolutionary control.
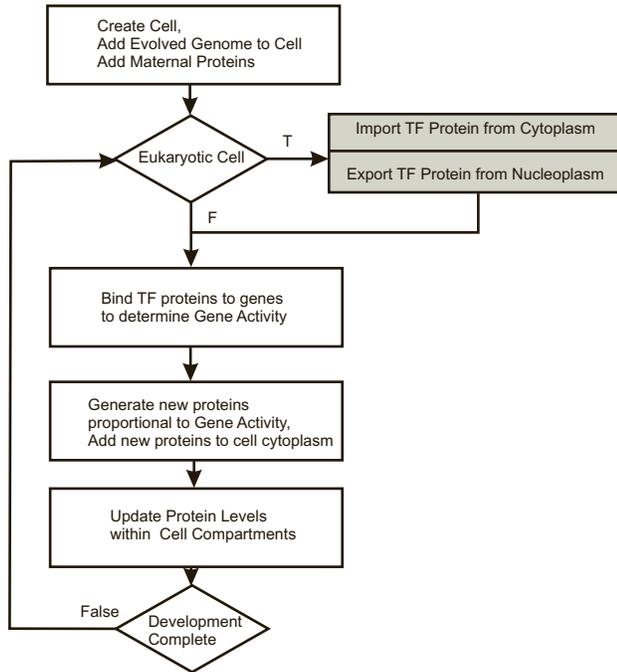

Figure 3: AED Cell Development Cycle

Protein Model and Classification

The AED eukaryotic cell model contains three proteins classes (Table 2) while the prokaryoticmodel contains only

steps	200	Defines the # Development Steps
ntf	3	Number of TF's
nim	3	Number of Importins
nex	3	Number of Exportins

(a) Developmental Algorithm Configuration

runs	50	Number of GA runs
gens	4000	Number of generations per run
psize	100	Population size
pmut	0.1	Mutation Rate
tsize	20	Tournament Size
pxo	0.6	Probability of crossover
obj	Sin	Objective
tps	2	Number of test proteins

(b) Genetic Algorithm Configuration

Table 1: Typical Configuration Parameters used by the AED platform. The settings listed are typical for the experiments reported in this paper.

a single TF type protein. Details of the three protein classes are listed in Table 2. Proteins are distinguished by a *protein code* derived from the gene code (Figure 4). There is a direct mapping between gene code and protein code, ie. gene code 1 maps to protein code 1 etc. The relationship between protein IDs and protein class (TF, IM, EX) is determined by the user setting of nim, nex and ntf in the configuration file listed in Table 1a.

Protein Class	Description	ID Range
Transcription Factor (TF)	regulates gene activity	$0 \rightarrow ntf - 1$
Importin (IM)	enables TF import	$ntf \rightarrow ntf + nim - 1$
Exportin (EX)	enables TF export	$ntf + nim \rightarrow ntf + nim + nex - 1$

Table 2: Protein Class Names and their corresponding function. The ID range of the protein types are calculated based on the user configuration of Table 1

Protein Transport

In eukaryotic cells TF type proteins must be first transported into the nucleus in order to regulate the rate of gene expression. TF proteins within the AED model include additional evolved NLS and NES regions, (Figure 4). Binding of IM/EX type proteins to these sites enables transport of the TF between compartments. During the transport phase of the cell cycle (Figure 3), each TF is selected and the proportion of protein exported $C_{ex}(tf)$ and imported $C_{im}(tf)$ is described by (1 and 2) respectively.

$$C_{ex}(tf) = TF_{\text{nuc}} * f(TF_{\text{nuc}} * W_{TF} + \sum_{i}^{n}(Ex_i * NES_i)) \quad (1)$$

$$C_{im}(tf) = TF_{\text{cyt}} * f(TF_{\text{cyt}} * W_{TF} + \sum_{i}^{n}(Im_i * NLS_i)) \quad (2)$$

where f is defined as the sigmoid function, with TF_{nuc} and TF_{cyt} being the concentrations of the selected TF in the nucleus and cytoplasm compartments respectively. W_{TF} is an evolved bias for the selected TF.

Figure 4: The evolved NLS and NES sequences associated with each TF

Mechanics of Gene Expression

Upon entering the nucleus the rate of gene expression (GE) is regulated by the binding of TF protein to the cis-regulatory sites of individual genes. For each gene encoded on the genome the expression rate GE_n is described as (3).

$$GE_n = SR_n * f(\sum_{i=0}^{n} I_n * TF_n - TH_n) \qquad (3)$$

where SR_n is the evolved max synthesis rate for the gene, TH_n and I_n are the evolved gene threshold and interaction levels respectively. All protein produced during the gene expression phase is placed within the cytoplasm and any existing protein concentration (C_{n-1}) is updated (4).

$$C_n = C_{n-1} - (C_{n-1} * DR) + GE_n \qquad (4)$$

where DR is the evolved decay rate for this protein.

The Evolved Genome Structure

The role of the GA is to provide candidate configurations (genomes) to the developmental algorithm. Following the development process these configurations are assigned a fitness. The GA is a derivative of the standard generational GA with elitism, gaussian mutation, uniform crossover and tournament selection. All parameters for the GA and development algorithms are user selectable via the configuration file (Table 1). The genome is subdivided into two chromosomes, with each chromosome subsequently decomposed into genes. The genes contained on the first chromosome exclusively encode the protein information for all 3 protein classes (TF, IM, EX). Genes on the second chromosome encode the transport specific information (NLS/NES) for each of the TF proteins (Figure 4). Figure 5 illustrates the structure of the genes contained on each chromosome.

Experiments and Results

This section describes a series of experiments and analyses of the GRN dynamics of both the eukaryotic and prokaryoticcell models. The experiments compares the abilities of the GRNs within eukaryotic and prokaryoticcell models to
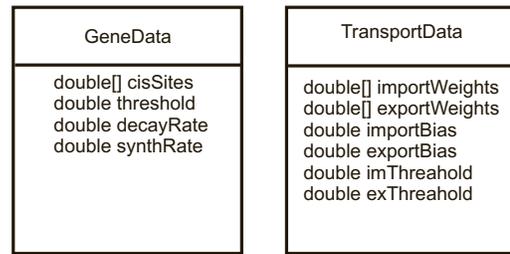

Figure 5: UML diagram of genes encoded on the 2 chromosome within eukaryotic cells.

evolve and follow defined protein patterns during their development. The objectives selected for the comparison are divided into three categories in order of increasing difficulty, namely static, periodic and aperiodic (Table 3). By contrasting the fitness achieved for the three classes of objective, allows a preliminary assessment of the contribution of transport within eukaryotic cell types.

Objective Name	Class
Lin	Static
Sin, Rect,	Perodic
Gauss, SinOff, GaussOff, AmpGauss	Aperodic

Table 3: Objective Names and their corresponding classification

The simplest static objectives determine the fitness at the end of the developmental cycle, ignoring the concentration profile of the protein during the developmental cycle. For the static objective the fitness function is similar to that used by Kumar (5).

$$Fitness = \sum_{j=0}^{M-1} (Cj - (1+j)/M)^2 \qquad (5)$$

where M is the total number of transcription factor proteins under test and Cj is the concentration of protein j at the end of development.

In contrast both periodic and aperiodic objectives assess the fitness over the entire developmental time. Periodic objectives are designed to mimic their biological equivalent, termed circadian rhythms, where two proteins oscillate in lockstep. The aperiodic objectives, represent another biologically plausible objective as they place a precise temporal dependence on gene expression. In all test cases the number of proteins tested against the fitness function can be specified, up to the maximum number of transcription factors. Thus the fitness function for dynamic and aperiodic objectives is defined as (6).

$$Fitness = \sum_{j=0}^{M-1} \sum_{i=0}^{n} (Oj_i - Cj_i)^2 \qquad (6)$$

where Cj_i is the concentration of protein j at time i, and Oj_i is the objective concentration.

For each of the seven objectives tested the AED platform, was typically configured as per Table 1.

Comparison of Eukaryotic and prokaryoticCell Dynamics

Results for each of the seven objectives (Table 3) are listed in Table 4. These results illustrate that both cell configurations can solve static and dynamic tasks with a high degree of accuracy. For the aperiodic objectives the eukaryotic type cells configurations show a considerable improvement over the prokaryoticcells. This improvement has its origins in the ability of the eukaryotic cell to limit the target protein activity to the specific times during the development. In contrast, the prokaryoticcells tends to have a continuous level of protein present at all times. Figure 7 illustrates the phenotype (protein levels within the cytoplasm) associated with each of the aperiodic objectives.

Fitness Results			
Objective	Nucleus	Best	Mean
Lin	T	0.0	0.092
	F	0.0	0.041
Sin	T	0.046	199.7
	F	0.058	250.95
SinOff	T	1.05	1288.71
	F	20.67	1415.8
RectSin	T	0.802	142.2
	F	1.31	131.13
Gauss	T	0.60	601.6
	F	4.765	785.39
GaussOff	T	0.855	590.25
	F	9.529	722.57
GaussAmp	T	0.581	391.78
	F	3.83	432.05

Table 4: Results for Objectives, the presence of a nucleus indicating a eukaryotic type cell is shown by the Boolean [T]rue.

Results Analysis: Gene Activity

The section illustrates how the process of transport within the eukaryotic cell is instrumental in generating the protein profiles associated with the aperiodic genome solutions. The level of gene activity for each of the individuals developed in Figure 7 is plotted in Figure 8. Because the gene ID corresponds to the proteins ID any variation in gene activity results in a corresponding change in its protein level.

Gene activity for the prokaryoticcell configuration (Figure 8 b, d, f) shows continuous activity over the entire development time, which is penalised by the aperiodic objectives. The close coupling between the genes and proteins in the prokaryoticcell makes it difficult to generate the isolated the gene activity required for the aperiodic objectives. This

coupling arises as a consequence of the genome and proteins being contained in the same cell compartment.

In contrast the eukaryotic cell achieves very specific regions of gene activity for genes 0 and 1 with relative ease (Figure 8 a, c, e). The regions of activity for genes 0 and 1 are very localised to the required times for peak protein activity within the cytoplasm. Inspecting the gene activity for the IM genes, (IDs $\{3, 4, 5\}$) and EX genes (IDs $\{6, 7, 8\}$) shows very broad and intense levels of activity, indicating that the process of transport is very heavily involved in the generation of the target protein profile.

The dynamics of the *Gauss* profile in Figure 8 e deserve special mention, as there is little or no activity on Gene 0 around the peak presence of protein 0 in the cytoplasm. Figure 6 illustrates that the protein profile is generated as a result of exporting the stored protein 0 from the nucleus at the appropriate time, Figure 6.
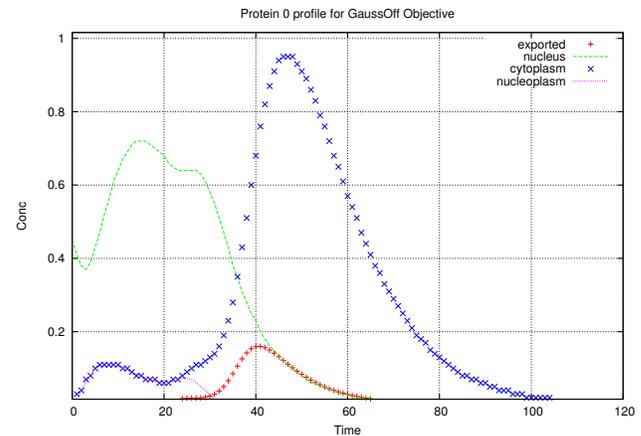

Figure 6: Protein 0 Dynamics for GaussOff Objective, illustrates the accumulation of Protein 0 in the nucleus prior to time step 20, while there is no gene activity around peak evaluation time, the cytoplasm protein profile for protein 0 is generated by exporting proteins from nucleus during this time.

Results Analysis : Transport as a source of Heterochrony

This section investigates how mutations to the transport chromosome affect the quality of the aperiodic solutions. Selecting the best eukaryotic individuals from the *gaussOff* and *ampGauss* objectives illustrated in Figures 7 c, e the activity of the transport specific genes (ID 3-8) is reduced ('*knocked out*'), according (7) -

$$GeneActivity = KO * SR * sigmoid(activity)$$
$$\text{where } KO \; \epsilon \; A = \{0.9, 0.8, 0.5, 0.1\} \quad (7)$$

Figure 9 illustrates the full spectrum of heterochronic mutations are possible. For the *ampGauss* objective the level

of protein 0 (Figure 9c), increases its level and duration in the cytoplasm in response knock out. The level of protein 1 remains relatively unaffected but its onset is delayed in response to knock out (Figure 9d) .

For the *GaussOff* objective (Figure 9a) the level of protein 0 is reduced in response to increasing knock out while the onset of protein 1 occurs earlier in the development (Figure 9b).

In addition to reducing the activity level the transport specific genes, the NLS/NES interaction levels could also have targeted to give similar results. From an evolutionary perspective, mutations to the transport chromosome have a causal relationship to the generated phenotype allowing the conclusion that the addition of the transport process, which is in effect a second level of regulation tends to provide a smoothing of the phenotype landscape.
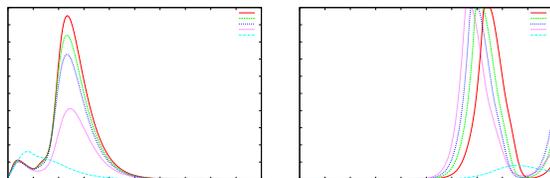
(a) Protein 0 GaussOff Objec-(b) Protein 1 GaussOff Objec-
tive, eukaryotic Cell Configu-tive, eukaryotic Cell Configu-
ration   ration
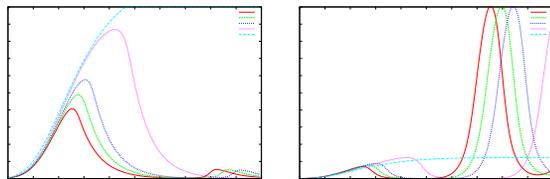
(c) Protein 0 AmpGauss Ob-(d) Protein 1 AmpGauss Ob-
jective, eukaryotic Cell Con-jective, eukaryotic Cell Con-
figuration   figuration

Figure 9: Heterochronic Mutations to the best eukary-otic cell individuals from the *GaussOff* and *AmpGauss* objectives, realised by reducing the activity of Importin and Exportin genes.

Discussion and Conclusions

The regulated entry of *transcription factors* to the nucleus of eukaryotic type cells has been shown to have a major influence on the direction of biological development. This paper has reported a biologically inspired eukaryotic cell model that captures the concept of regulated protein transport to and from the cell nucleus. Tests on the evolvability of the GRN indicate that the addition of this level of complexity does not prevent the cell successfully generating GRN dynamics. Indeed, it serves to improve the GRNs ability to

evolve aperiodic objectives. Analysis of the gene activity within the eukaryotic cell shows that it relies heavily on the transport of TF to and from the nucleus to control gene activity. In particular it is observed that for aperiodic tasks TF protein is only present in the cytoplasm at the required development time intervals. In contrast, while the prokaryotic-cell model fared well for static and periodic tasks, its performance suffered significantly for aperiodic objectives. An examination of the gene activity within the prokaryoticcell model has shown continuous levels of gene activity during development time. In contrast, the eukaryotic cell isolates its gene activity to very specific regions of the development time. The high levels of activity for the IM/EX genes indicates their importance in generating the protein dynamics. The eukaryotic cell model demonstrates the potential for heterchronic mutations to arise by scaling the activity of transport specific genes. Moreover a high degree of correlation between the level of disruption to these genes and the resulting change in protein profile has been observed.

References

Banzhaf, W. and Miller, J. (2004). The challenge of complexity. In *Frontiers of Evolutionary Computation*, pages 243–260. Springer.

Bentley, P. and Kumar, S. (1999). Three ways to grow designs: A comparison of embryogenies for an evolutionary design problem. In *Proceedings of the genetic and evolutionary computation conference*, volume 1, pages 35–43.

Haddow, P. C. and Hoye, J. (2007). Achieving a simple development model for 3d shapes: are chemicals necessary? In *Proceedings of the 9th annual conference on Genetic and evolutionary computation*, pages 1013–1020.

Kumar, S. and Bentley, P. J. (2003). Biologically inspired evolutionary development. In *Evolvable Systems: From Biology to Hardware*, pages 57–68. Springer.

Matos, A., Suzuki, R., and Arita, T. (2009). Heterochrony and artificial embryogeny: A method for analyzing artificial embryogenies based on developmental dynamics. *Artificial life*, 15(2):131–160.

Raff, R. A. (1996). The shape of life: genes, development, and the evolution of animal form.

Roggen, D. and Federici, D. (2004). Multi-cellular development: is there scalability and robustness to gain? In *Parallel Problem Solving from Nature-PPSN VIII*, pages 391–400.

Stanley, K. O. and Miikkulainen, R. (2003). A taxonomy for artificial embryogeny. *Artificial Life*, 9(2):93–130.

Van Remortel, P., Ceuppens, J., Defaweux, A., Lenaerts, T., and Manderick, B. (2003). Developmental effects on tuneable fitness landscapes. In *Evolvable Systems: From Biology to Hardware*, pages 117–128. Springer.

West-Eberhard, M. J. (2003). *Developmental plasticity and evolution*. Oxford University Press, USA.
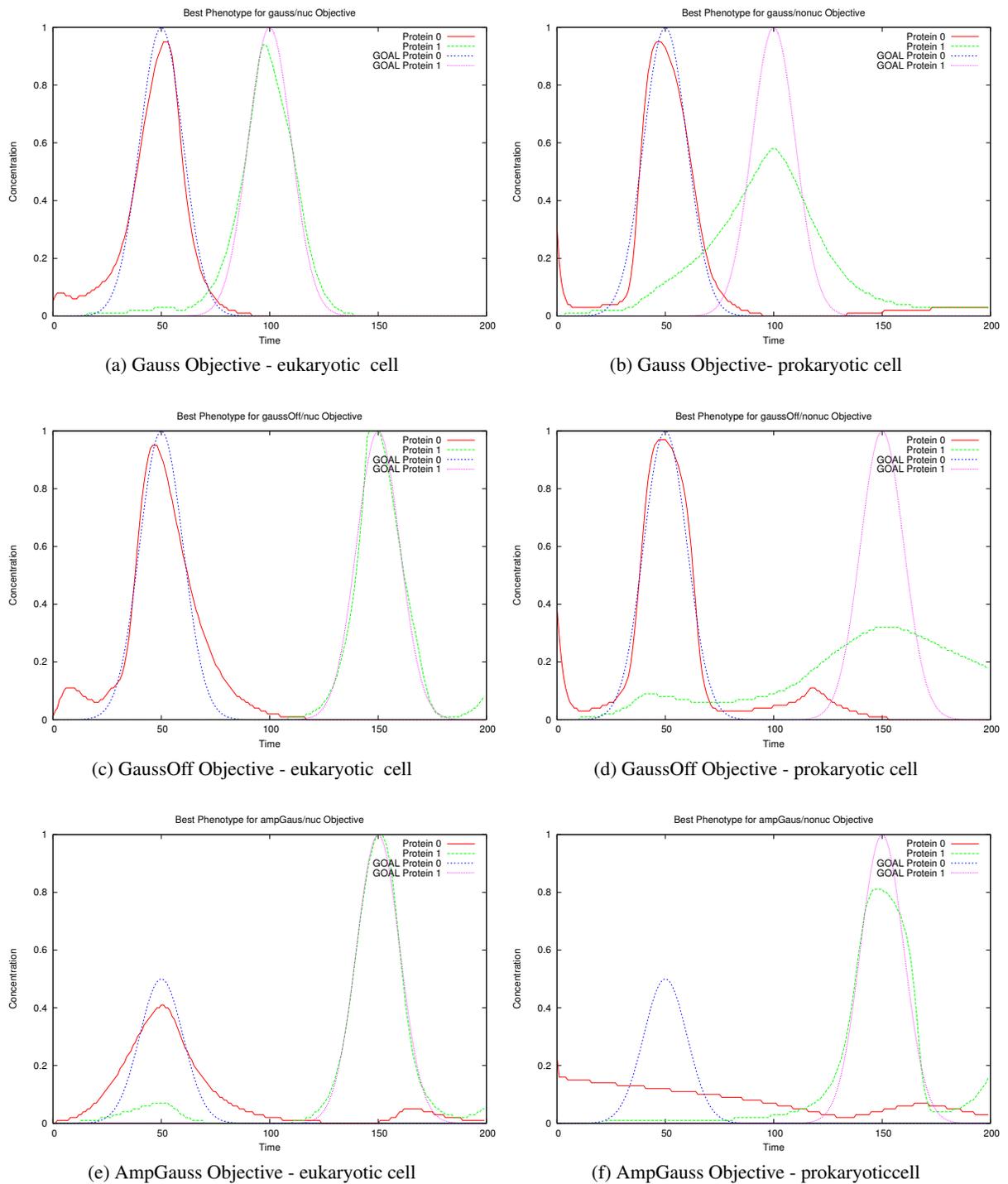
Figure 7: Protein profiles generated by the best Individuals as reported in 4 - Aperiodic objectives only, contrasted against the Objective. The AED configuration of Table 1 configures proteins IDs (0,1) to be used in the fitness calculation.
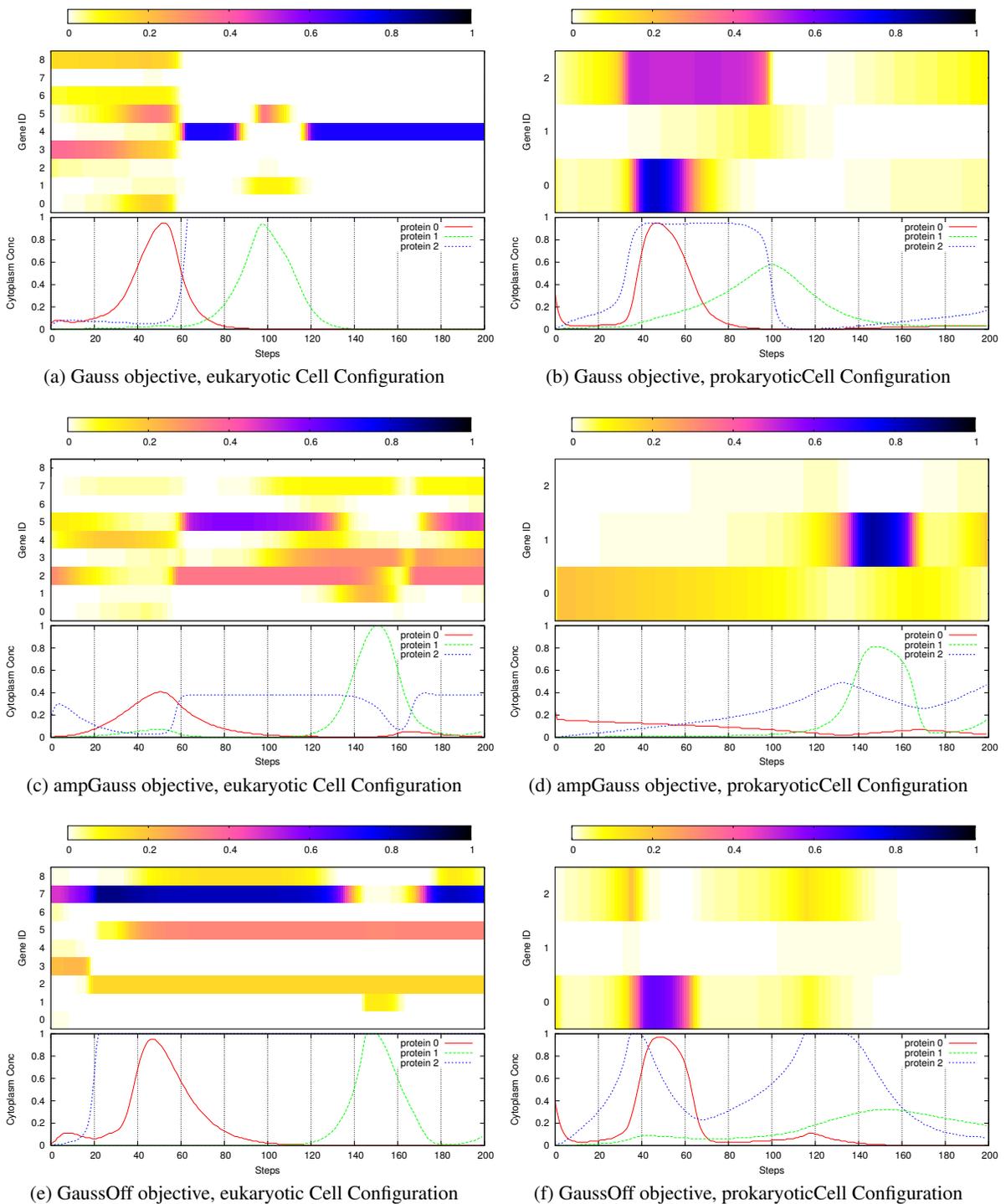
Figure 8: Gene Activity for best evolved individuals for aperiodic objectives. The top section of each plot maps the gene activity to a colour intensity, while the bottom section shows the corresponding protein profile in the cytoplasm. For subplots (a,c,e) the configuration listed in Table 1 configures the import proteins IDs range from 3-5, and export proteins IDs range from 6-8