Nonlinear estimation in a class of gene transcription process

Ricardo Aguilar-López [a], M. Isabel Neria-González [b], Rafael Martínez-Guerra [c], Juan L. Mata-Machuca [d],*

[a] Department of Biotechnology and Bioengineering, CINVESTAV-IPN, Av. Instituto Politecnico Nacional, No. 2508, San Pedro Zacatenco, 07360 D.F. México, Mexico
[b] Chemical and Biochemical Engineering Division, TESE, Av. Tecnologico at Av. Carlos Hank Gonzalez, 55210 Ecatepec de Morelos, Mexico
[c] Department of Automatic Control, CINVESTAV-IPN, Av. Instituto Politecnico Nacional, No. 2508, San Pedro Zacatenco, 07360 D.F. México, Mexico
[d] Instituto Politécnico Nacional, Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas, Academia de Mecatrónica, Av. Instituto Politécnico Nacional, No. 2580, 07340 D.F. México, Mexico

ARTICLE INFO

ABSTRACT

In this work the Goodwin model applied to gene transcription is employed as a benchmark system for estimation purposes, considering two dynamic behaviors, monotone decreasing and sustained oscillations, each one under a specific parameter's set. The preceding observability analysis of the Goodwin model was done via linear observability and the differential–algebraic framework, where is proved that the system is fully observable from mRNA concentration measurements. Therefore a class of nonlinear observer which considers a class of sigmoid and linear functions of the output feedback, considering model uncertainties, is proposed and a sketch of proof of the observer's convergence is provided under the background of the Lyapunov theory, in order to demonstrate asymptotic convergence. Numerical experiments are carrying out in order to show the performance of the proposed methodology which is compared with a standard Luenberger (Proportional) observer and a proportional sliding-mode observer (PSMO).

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Gene transcription roughly speaking is a process that consists on transcribing of genetic information from DNA nucleotide sequence into a RNA nucleotide sequence. The RNA nucleotide sequence is used as a code to synthesize polypeptides, the basic component of all proteins [1–3]. DNA is retained mainly within the nucleus chromosomes in eukaryotic cells. It controls cellular activity by coding for the production of enzymes and proteins. The DNA information is not directly converted into proteins, an intermediate step is carry out where DNA is copied into messenger RNA (mRNA), process knowing as "central dogma of molecular biology" [4,5]. This mechanism ensures that the information contained within the DNA is conserved [6].

In-depth understanding of the genetic transcription process is a big challenge, it is well known that this process is related with the main steps which keeping the cellular life, for example, the several cellular activities concerning survival, growth and differentiation are reflected in altered states of gene expression, also the ability to quantitative and predictive understanding of transcription levels of specific genes has always been central to any research into encoded DNA in order to understand how the biological machinery works [7,8]. In recent years, the emergence of molecular biology has resulted in the increased use of techniques able to identify and measure the DNA/RNA levels. Such applications are extensive in a

---

* Corresponding author.
  E-mail addresses: raguilar@cinvestav.mx (R. Aguilar-López), rguerra@ctrl.cinvestav.mx (R. Martínez-Guerra), jmatam@ipn.mx (J.L. Mata-Machuca).

lot of fields for example: in diagnostic microbiology for the identification of pathogen microorganisms and virus (*C. trachomatis, N. gonorrhoeae, C. pneumoniae, Legionella* spp., *Plasmodium* spp., *Aspergillus* spp., Herpes simplex virus, human papilloma virus, influenza virus, HIV, hepatitis B, etc.) to detect antimicrobial resistance genes (methicillin resistant *S. aureus*, vancomycin resistant enterococci) to improved the treatment by viral resistance detection (lamivudine resistance in hepatitis B virus) [9]; identification of genes responsible for the formation of carcinoma of prostate [10]; in malign melanoma detection and therapy [11]; for the comprehension of the mechanism of rheumatoid arthritis disease and the creation of novel treatments [12]; in the study of brain diseases as amyotrophic lateral sclerosis, Huntington, Alzheimer and Parkinson [13–16].

The principal methods for the quantification and dynamical variation of DNA/RNA during the gene transcription are the reverse transcription polymerase chain reaction (RT-PCR) [17–19], cDNA arrays [20], northern blotting [21], RNAse protection assays [22] and fluorescence in situ hybridization [23].

The RT-PCR is a sensitive method for the quantification of equilibrium mRNA levels, principally in small samples of RNA, or for analysis of low level transcripts also it can be used to determine and compare the concentration of gene and/or transcript numbers in different sample populations, to discriminate between closely connected mRNAs, to characterize patterns of mRNA expression and to analyze RNA structure [17,24]. The cDNA array analysis has 15 years of launched, it allows analyzing the gene expression data of thousands of genes simultaneously although with a very high cost [20,25,26]. Northern blotting analysis is a classical method that gives a snapshot of the size, abundance and steady-state level of a specific RNA transcript in a complex sample. It is relatively simple to perform, quantitative, relatively inexpensive, and not plagued by artifacts, it is very used in the specific detection, validation and study of the expression of small RNAs (20–24 nucleotides long and nonprotein coding). The main disadvantages are the high time consumption and its reduced sensitivity and reproducibility [21,27–29].The RNA protection assay (RPA) provides a sensitive test for the detection and simultaneously quantification of multiple mRNA targets for different tissues, developmental stages, or times of the day with the disadvantage of to be short of information on transcript size, it is very used in the study of circadian genes and its oscillations altered by genetic manipulation [22,30,31]. Fluorescence in situ hybridization (FISH) is a powerful technique to detect and localize DNA or RNA sequences in cells environment, tissues and tumors. It is very useful to detect chromosome aberrations but it low sensitive and gives high false positive [23,32].

However, the above methods are still limited in measure online a lot of different metabolites and genetic material necessary to fully understand the bio-system principally by the insufficient suitable analytical sensor techniques and the long reaction times, also, it is clear that is remotely achievable use as many sensors as signals of interest characterizing the bio-system behavior due to technological constraints, cost reasons, and so on. One way to around the previous problems is employing mathematical structures to predict or estimates the unknown system variables of this kind of systems. Mathematical models have been employed to describe rigorously many biological systems for last few decades, for example, its interest has grown a lot since it was development that the gene expression is the nexus for many human diseases, population differences and evolution, although still limited by its poorly quantitative understanding (there are many metabolites and molecules impossible to measure online) of the process at a molecular level. A classification of the type models that have been employed en bio-systems to describe genetic regulatory systems are the following: directed graphs, Bayesian, Boolean and Generalized logical networks, nonlinear ordinary differential equations and partial differential equations. Ordinary differential equation models are broadly use as these frequently include time and/or space-dependent variables such as protein and mRNA concentrations and parameters such as production and degradation constant rates. These equations specify the levels of each protein or mRNA as a function of the other components as the system evolves [33].

The Goodwin model has been earliest predicting model employed to describe the negative feedback oscillation on gene expression [34,35]. Broadly speaking it has been employed as a negative feedback regulatory system until today, it was designed in order to explain via differential equations formalism important biological processes, for example the molecular mechanism of circadian rhythms or enzymatic regulation (such as lactose in bacteria), and until today it has been a good and easy example of the homeostatic behavior at cellular level and for gene expression.

A description and discussion of the regulation of gene expression models including the Goodwin model has been given by [33]. In the three-dimensional model for circadian clocks that we derived from Goodwin, the considered states variables are related with clock mRNA concentration, the clock protein concentration and the inhibitor concentration of transcription. The characteristic feature of the Goodwin model is that the production rate of intermediates (clock protein and inhibitor) is a (linear) function of the concentrations of the preceding intermediates (clock mRNA and clock protein, respectively), while regular transcription (production of mRNA without inhibition) occurs at a constant rate. However, the production of mRNA is inhibited by increasing concentrations of the inhibitor protein.

During the process mRNA, clock protein and clock protein inhibitor approach high- or low-steady-state values depending on whether synthesis reactions are turned on or, due to inhibition, are turned off. In this model the degradation rate constants (α, β, γ) plays an important role in the determination of the dynamic behavior in this model.

Dynamic of coupled genetic systems has important biological implications and potential engineering applications from both theoretical and experimental viewpoints, and it is also essential for the understanding of the rhythmic phenomena of living organisms for both molecular and cellular levels.

From the above, the problem of observer design arises naturally in genetic systems as soon as it is necessary to know about unmeasured internal information (like proteins concentration impossible to measure) from external measurements (for example mRNA levels) in order to a whole understanding of the biological process. An observer is a mathematical