Intelligent Control and Biological Regulation For Bioinformatics

Aboubekeur Hamdi-Cherif

***Abstract*—**Regulation and control in biological processes are the center of life. Living organisms grow and reproduce. They maintain their structures and respond to their environments. All these processes are done through regulation and control. This paper reports the study of regulation and the applicability of intelligent control to bioinformatics, particularly to biological systems. In addition to two previously-described phases of bioinformatics discipline, characterized by intelligence-free programs, and artificial intelligence-based programs, respectively, another phase is now proposed that incorporates intelligent control action and further understanding of biological regulation. All three phases can alternatively be viewed as levels corresponding to historical evolvement in our understanding of the field - of increasing degree of complexity. As the most complex of all, the intelligent control level, reported here, is dedicated to offering the necessary scientific and developmental framework for enhancing bioinformatics through the determination of optimal therapeutic strategies; tissue engineering being a far-reaching goal.

***Keywords*—** Bioinformatics, Machine learning, Intelligent control, Biological regulation, Biological control.

I. Introduction

Our aim is to integrate control theory and regulation in bioinformatics, under one unified perspective. In order to understand how normal cellular activities are altered in different disease states, the biological data must be compiled to form a comprehensive picture of these activities. Therefore, the field of bioinformatics has evolved such that the most pressing task involves the storing, retrieval, and interpretation of various types of biological data, including nucleotide and amino acid sequences, protein domains, and protein structures.

One of the major tasks of bioinformatics is the study and development of tools that enable efficient access to, and management of various types of biological data. The computational aspect of bioinformatics studies the development of new algorithms and statistical methods with which to assess relationships among members of large data sets, such as finding location of a gene within a sequence, predicting protein structure and/or function, and clustering protein sequences into families of related sequences [4].

The field of bioinformatics went through two main historical phases, during which standard heuristics-free programs were used such as database management systems (DBMSs), followed by limited artificial intelligence-based programs [25]. How can control theory enhance bioinformatics?

First of all, a control system for a physical system is an arrangement of hardware components designed to alter, to regulate, or to command, through a *control action*, that physical system so that it exhibits certain desired characteristics or behavior. Physical control systems are typically of two types: open-loop control systems, in which the control action is independent of the physical system output, and closed-loop control systems, also known as *feedback control systems*, in which the control action depends on the physical system output.

Intelligent control uses methods and techniques from artificial intelligence (AI) such as logical inference and machine learning methods (ML) like neural networks (NNs), genetic algorithms, and reinforcement learning among other methods [43]. The integration of control methodology within biology has been on the way but is still in its infancy [11]. As a complement to a previous work, concentrating on machine learning issues [16], [17] emphasis is now made on intelligent control problems and on how these theories can be integrated within a coherent framework; the aim of which is to provide future enhanced bioinformatics platforms [3].

The paper is organized as follows. Section 2 deals with the problem formulation. This section poses the fundamental question: "Why do we need a third level involving intelligent control in bioinformatics on top of the two purporting ones?" Section 3 describes some relevant bioinformatics issues. Section 4 reports a brief description of the control methodology. Section 5 describes methods from intelligent control relevant to bioinformatics. In Section 6, systems and computational issues are addressed. Biological modeling and control are discussed in Section 7. Section 8 reports the possible impacts the proposed framework is thought to induce on future bioinformatics. A conclusion sums up the main results and points towards some potential future developments.

Manuscript received March 24, 2010. This work is supported by the Deanship of Scientific Research, Qassim University, Saudi Arabia, under Contract # SR-D-007-30. Financial Support is gratefully acknowledged.

A. Hamdi-Cherif is with Computer College, Computer Science Department, Qassim University, PO Box 6688, 51452 Buraydah, Saudi Arabia. Tel. +96663800050, ext. 3301. His permanent address is with Faculty of Engineering, Computer Science Department, Université Ferhat Abbas, Setif (UFAS) 19000, Algeria. email: shrief@qu.edu.sa, elhamdi62@gmail.com

II. Problem Formulation

As far as the study of bioinformatics is concerned, we suggest using the traditional entry points available to computer and control scientists [3]. Specifically, the aim is to extend earlier works on control [14] and give directions of application to bioinformatics and especially to contribute to the enhancements of the previously-described two levels of bioinformatics [16], [17].

A. Major Intractable Natural Problems

The discovery of the structure of deoxyribonucleic acid (DNA), as a building bloc of living species, was a turning point in the history of science, culture and society. Its visible impacts on medicine, agriculture, energy production, social issues, ethics, and others, continues to create interesting challenges in all human endeavors. Such multidisciplinary efforts require scientists who are able to cross boundaries between many disciplines and who can make a valuable contribution to science and society at large. Awareness of the wholeness of this task as well as its implications, not only for science but also for humanity, requires a sense of responsibility that is equally whole [23]. In this regard, some fundamental natural problems in biology have been raised [31]:

- There are no rules without exception.
- Every phenomenon has a nonlocal component.
- Every phenomenon is intertwined with others.

The first problem poses the difficulty of using induction. Induction, also known as inductive reasoning, is a type of reasoning that involves moving logically from a set of specific facts or examples to a general conclusion. It can also be seen as a form of theory-building, in which specific facts or examples are used to create a theory that explains relationships between the facts and allows prediction and/or inference of future knowledge.

For the second and third problems, non-locality means space distribution, and intertwined phenomena imply non-separability. All these issues have been studied within the control community and can, as such, make useful contributions to the proposed framework.

B. Bioinformatics Efforts

1) Human Genome Project (HGP)

The Human Genome Project (HGP) represents one of the major bioinformatics efforts. **C**ompleted in 2003, HGP was a 13-year collaborative work coordinated by the US Department of Energy and the National Institutes of Health (NIH). During the early years of the HGP, additional contributions came from many countries such as Japan, France, Germany, China, among others. The use of computer technology for storing DNA sequence information and constructing the correct DNA sequences from fragments identified by restriction enzymes was one of the first applications arising from the different bioinformatics sequencing projects. In general, enzymes are a micro-machine that catalyses a certain reaction, like the breakdown of a food source or toxin. Specifically, restriction enzymes break up the DNA at certain points.

2) Goals of HGP

The HGP goals were to

- *identify* all the approximately 20,000-25,000 genes in human DNA,
- *determine* the sequences of the 3 billion chemical base pairs (bps) that make up human DNA,
- *store* this information in databases,
- *improve* tools for data analysis and accessibility,
- *transfer* related technologies to the private sector,
- *address* the ethical, legal, and social issues (ELSI) that may arise from the project.

Though the HGP is finished, analyses of the data will continue for many years to come.

[http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml].

3) Post-genomic Era

Although only few relatively small genomes have been completely sequenced thus far, the end of the HGP is conceptually anticipated by the community in proclaiming the start of the "post-genomic era" or the period of "functional genomics". Systematic elucidation of gene function requires to link sequence data with information about molecular mechanisms, and also with histological, anatomical and even taxonomical data. As a consequence, even "classical" branches of biological and medical research gained interest when linked to genome-based information.

[http://www.gene-regulation.com/info/cytomer.html].

The website of the US Department of Energy Genome Programs maintains useful and updated information, http://genomics.energy.gov

C. Third Level of Bioinformatics Discipline

Intelligence-free programs are the characteristics of the first level in bioinformatics development based on standard programs such as database management systems (DBMSs). Limited intelligence-based programs characterize the second level [25]. In addition to the two levels of bioinformatics described earlier [16], [17] we here provide a framework characterizing the principles that may underlie a foreseeable third level, based on the integration of intelligent control within bioinformatics. This integration allows external actions on the various elements within a given database with potential guidance in the process of drug discovery, and biological systems design such as tissue engineering, for instance.

In summary for our problem formulation, the present paper attempts to address this central issue by looking at the latest advances in intelligent control as computational problem solvers for biomedical applications.

III. Biological Relevant Issues

In this section, we concisely present the main concepts from biology relevant to our discussion. These concepts concern structure of genes, transcription, and transcription factors are detailed in [29].

A. Structure of Genes

Genes are pieces of DNA that encode for proteins through the intermediate action of messenger RNA (mRNA). Proteins

are made of amino acids arranged in a linear chain and joined together by peptide bonds. A gene and the genomic region surrounding it consists of a transcribed sequence, which is converted into an mRNA transcript, and of various *untranscribed* sequences. The mRNA is transcribed from a DNA template, and carries coding information to the sites of protein synthesis: the ribosome. The mRNA consists of a coding sequence that is translated into a protein and of several *untranslated* regions (UTRs). The untranscribed sequences and the UTRs play a major role in the regulation of expression. Notably, the *promoter region* in front of the transcribed sequence contains the *binding sites* for the *transcription factor* proteins that start up transcription. Moreover, the region upstream of the transcription start contains many binding sites for transcription factors that act as *activators* and *repressors* of gene expression, although some transcription factors can bind outside this region.

B. Transciption

Transcription means the assembly of ribonucleotides into a single strand of mRNA whose sequence is dictated by the order of the nucleotides in the transcribed part of the gene. The transcription process is initiated by the binding of several transcription factors to regulatory sites in the DNA, usually located in the promoter region of the gene. The transcription factor proteins bind each other to form a complex that associates with an enzyme called RNA polymerase. This association enables the binding of RNA polymerase to a specific site in the promoter. Together, the complex of transcription factors and the RNA polymerase loosen the DNA and separate both strands. As a result, the polymerase proceeds down on one strand while it builds up a strand of mRNA complementary to the DNA, until it reaches the terminator sequence. In this way, an mRNA is produced that is complementary to the transcribed part of the gene. Then, the mRNA transcript leaves the RNA polymerase, and the polymerase breaks its contact with the DNA. In a later stage, the mRNA is processed, transported out of the nucleus, and translated into a protein.

C. Transcription Factors

Transcription factors are proteins that bind to regulatory sequences on eukaryotic chromosomes thereby modifying the rate of transcription of a gene. Some transcription factors bind directly to specific sequences in the DNA (promoters, enhancers, and silencers), others bind to each other. Most of them bind both to the DNA as well as to other transcription factors. The transcription rate can be positively or negatively affected by the action of transcription factors. When the transcription factor significantly decreases the transcription of a gene, it is called a repressor. If, on the other hand, the expression of a gene is upregulated, biologists speak of an activator.

D. Regulatory Elements on the Web

Regulatory elements play a central role in the study of biological sequences. For this reason many databases have been developed and are available to explore known regulatory elements. Table 1 gives a list of databases of promoters and gene regulation that are accessible online. Most of these sites are also portals to specific tools for the analysis of regulatory mechanisms.

Data base	*URL*
EPD	*www.epd.isb-sib.ch*
TRANSFAC	*www.gene-regulation.de/*
PLACE	*www.dna.affrc.go.jp/htdocs/PLACE [19]*
TRRD	*www.bionet.nsc.ru*
SCPD	*cgsigma.cshl.org/jian*
HPD	*zlab.bu.edu/~mfrith/HPD.html*
COMPEL	*compel.bionet.nsc.ru/compel*

Table 1 Regulatory elements sites

IV. Control systems Contributions

As far as control theories are concerned, we only describe here the most important ones and show how these methods can be useful in bioinformatics. There are indeed many lines of research that can help in molecular biology development. Many of these are directly related to control and system theories [11], [30].

A. Dynamic Control

1) Negative Feedback control

Control is an interdisciplinary branch of engineering and mathematics, which deals with the behavior of dynamical systems. The desired output of a system is taken as a reference to be attained or maintained at a specific value. When one or more output variables of a system need to follow a certain reference over time, a controller manipulates the inputs to the system to obtain the desired effect on the output of the system. This is usually done using negative feedback, *i.e.* a procedure whereby the actual value is subtracted from the desired value to create the error signal which is used by the controller to allow correction to be undertaken at subsequent stages. This procedure is therefore done in closed-loop form.

2) Examples Physical Control Systems

A thermostat is a simple example for a closed-loop negative feedback control system. Indeed, it constantly measures the actual temperature and controls the heater's valve setting to increase or decrease the room temperature according to the user-defined setting. A simple method switches the heater either completely on, or completely off, and an overshoot or undershoot of the controlled temperature must be expected dictated by the physical inertia of the system. A more expensive method varies the amount of heat provided by the heater depending on the difference between the required temperature, or "setpoint" and the actual temperature. This minimizes over/undershoots. Other more sophisticated control schemes can also be applied [13].

An anti-lock braking system (ABS) used in car braking technology is a more complex example, consisting of multiple inputs, conditions and outputs.

3) Control Laws Construction

Whatever control strategy is used, the resulted control system must guarantee the stability of the closed-loop behavior, *i.e.* preventing that the system state or output takes unacceptable values, inducing heavy breakdowns. For linear systems, this can be obtained by directly placing the poles of the closed-loop transfer function. For multiple-input multiple-output (MIMO) systems, pole placement can be performed mathematically using a state space representation of the open-loop system and calculating a feedback matrix assigning poles in desired location of the s-plane for continuous systems or the or z-plane for discrete systems. This is usually done by computer aided control systems design (CACSD) methods tools and capabilities [14].

Whatever methods are used for linear systems, one cannot always ensure robustness, *i.e.* the ability to cope with small differences between the true system and the nominal model used for design. Furthermore, all system states are not in general measured or directly accessible and that is why estimators must be included and incorporated in pole placement design. The estimators are either observers of Luenberger type for deterministic control [13] or Kalman filters for stochastic control.

B. Optimal Control

1) Optimal Control Application

Loosely defined, an optimal control problem is stated as follows. Given a system with known dynamics and output and some initial state, find a control law that minimizes a given cost functional subject to some prescribed constraints. As a control methodology, optimal control, rooted in the calculus of variations, is not considered within the framework of intelligent control, discussed in next section. The main link that exists between optimal control and intelligent control is perhaps the parameter identification procedure, when needed, whereby some sort of learning is used.

2) General methodology

2.1 Principle of Optimality

Optimal control problems can be solved using the technique of dynamic programming. This technique, pioneered by Bellman in the 1960s, is based on the so-called principle of optimality. When an optimal strategy exists, the principle of optimality asserts: if one searches for an optimal strategy over a subset of the original number of steps, then this new optimal strategy will be given by the overall optimal strategy, restricted to the steps being considered. It can be used to arrive at the dynamic programming solution of the basic problem expressed above [5].

2.2 Maximum Principle / Dynamic Programming

Whenever it is possible to use a state-variable to represent the system to be controlled, then optimal control theories can be applied such as the Pontryagin's maximum principle, *via* the optimization of the Hamiltonian, or dynamic programming, based on the principle of optimality. Optimal control methods have been applied in regulation of transcription. Using complete genome sequences available for many eukaryotic organisms, the genome-based analysis will become more and more important in interpreting gene regulation. In this respect, the suppression of responses to mechanical stimuli in human joint tissue (synovial cells) by solving a Ricatti equation has been applied. Since the derived control law can be implemented by a DNA transfer technique such as a promoter competition assay, the novel genome-wide model-based approach would be useful in developing a strategy for gene therapies and tissue engineering [27].

2.3 Steps for optimal control bioinformatics application

Step 1: *Initialization*. Select genes involved in the responses to mechanical stimuli.

Step 2: *Mathematical model*. Find a matrix linear system of equations and the measurement equation in a general state-space form. Define a state vector, a control vector, and a measurement vector, with three matrices which are the system matrix **A**, the control or input matrix **B**, and the measurement matrix **C** for the genes associated with the mechanical responses,

Step 3: *Identification*. Practical determination of the matrices **A**, **B**, and **C.** Here identification is necessary. The system matrix **A** can be identified from the temporal profile of the estimated state, **x**, assuming that the dynamics for each state as a linear second order differential equation.

Step 4: Eigenanalysis. Single value decomposition (SVD) can be used to factorize the unprocessed matrix, **C.**

Step 5: *Control law construction*. LQ regulator solution. Solve the linear quadratic (LQ) problem and find the closed-loop control. This requires solving an algebraic Ricatti equation (ARE).

Step 6: Numerical integration and Monte Carlo simulation. **N**umerically integrate the dynamical responses. Establish statistical significance of the modeled mRNA expression profiles using Monte Carlo simulation [27].

V. Intelligent Control Paradigm

Roughly speaking, intelligent control lies at the intersection or artificial intelligence (AI) and control. It uses various AI computing approaches like neural networks (NNs), Bayesian probability, fuzzy logic (FL), machine learning (ML), evolutionary computation, genetic algorithms (GAs), expert systems and consciousness / cognition to control a given dynamic system [24].

A. Soft Computing Methods

Soft computing is not a closed and clear-cut discipline. It incorporates an emerging family of problem-stating and problem-solving methods that attempt to mimic natural intelligence; this latter based on approximate reasoning, heuristics and the power of easy generalization. Basically, there are two important components *i.e.* fuzzy logic-based models (FLMs) and experimental data learning methods such as neural networks (NNs) and support vector machines (SVMs). In addition, there are methods based on genetic

algorithms (GAs), evolutionary algorithms (EAs), probabilistic reasoning, belief networks, rough sets, wavelets, fractal and chaos theories. Soft computing methods are used whenever it is not possible to devise a mathematical model from first principles. The aim is to:

- Learn from experimental data (examples, samples, measurements, records, patterns, observations,…) by NNs or SVMs.
- Embed existing structural human knowledge such as experience, expertise, and heuristics, into efficient mathematical framework such as IF-THEN rules.

B. Neural Networks (NNs)

1) NNs in Bioinformatics

Neural networks (NNs) represent a body of machine learning knowledge from the field of AI with proven pattern recognition capabilities and have been utilized in many areas of science and technology including bioinformatics. NNs learning methods are robust approximators of real-valued, discrete-valued, and vector-valued functions. For certain types of problems, such as learning to interpret complex real-world sensor data, NNs are among the most effective learning methods currently known.

For example, in spite of its simplicity of implementation, the Backpropagation algorithm has proven surprisingly successful in many practical problems such as learning to recognize handwritten characters, learning to recognize spoken words, and learning to recognize faces [24]. NNs have been applied to biomedical problems such as disease classification and identification of biomarkers. This is due to their ability to cope with highly dimensional complex datasets such as those developed by protein mass spectrometry and DNA microarray experiments [26]. NNs play a central role in areas as diverse as protein structure and function prediction. A critical overview of recent advances in bioinformatics which have used NNs methods have been reported [39].

2) Advantages of NNs

- *Learning*. NNs have an ability of learning from data, mimicking human ability.
- *Approximators*. NNs can approximate any multivariate nonlinear function.
- *Simplicity*. NNs do not need deep understanding of the process or the problem being studied.
- *Parallelism*. Have parallel structure and can easily be implemented on hardware.
- *Ubiquity*. Some NNs can cover broad and different classes of tasks.

3) Disadvantages of NNs

- *Long training*. NNs need long time in training and learning especially for problems with local minima or multiple solutions. Impedes many real-time applications.
- *No additional knowledge*. NNs do not uncover basic internal relations of physical variables and do not increase our knowledge about the process.
- *Bad generalizations*. NNs are prone to bad generalization, with large number of weights, tendency to overfit the data, poor performance on previously unseen data during training phase.
- *Difficult choice*. Little or no guidance is offered about NNs structure or optimization procedure, or type of NNs to use for a particular problem.

C. Fuzzy Logic Paradigm

Classical, also known as crisp or Boolean logic, is a mathematical system that operates on discrete values of either 0 or 1 (true or false). Fuzzy logic considers real input values in terms of logical variables that take on continuous values between 0 and 1, in contrast with crisp logic.

1) Advantages of Fuzzy Logic Models (FLMs)

- *Human knowledge embedding*. Fuzzy logic models (FLMs) are efficient tools for embedding human knowledge into useful algorithms.
- *Approximators*. FLMs are good approximators of any multivariate nonlinear function.
- *No need for explicit modeling*. FLMs are useful when no mathematical model is available or when it is impossible to obtain it.
- *Robustness*. FLMs operate successfully under a lack of precise sensor information.
- *Genericity*. FLMs are appropriate tool in generic decision-making.

2) Disadvantages of FLMs

- *Structuring knowledge*. Human experts may have problems in structuring their knowledge.
- *Inconsistencies and human subjectivity*. Human expert may sway between extreme decisions or tend to hide their knowledge.
- *Exponential explosion*. The number of rules increases exponentially with increase in the number of fuzzy subsets per input variable.
- *High constraints*. Learning, *i.e.* changing membership functions, shapes is highly constrained, typically more complex than NNs.

3) Fuzzy Control Systems (FCS) Design

A fuzzy control system (FCS) is a control system based on fuzzy logic. From a control theoretical point of view, fuzzy logic has been intermixed with all the important aspects of systems theory: modeling, identification, analysis, stability, synthesis, filtering, and estimation.

A number of assumptions are implicit in any FCS design. Six basic assumptions are commonly made whenever a fuzzy rule-based control policy is selected.

(i) A solution exists.

(ii) The plant is observable and controllable: state, input, and output variables are usually available for observation and measurement or computation.

(iii) There exists a body of knowledge comprised of a set of linguistic rules of type IF-THEN, engineering common sense, intuition, or a set of input–output measurements data from which rules can be extracted.

(iv) The control engineer is looking for a ‘‘good enough’’ solution, not necessarily the optimum one.

(v) The controller will be designed within an acceptable range of precision.

(vi) The problems of stability and optimality are not addressed explicitly; such issues are still open problems in FCS design. However, interest in stability criteria for FCSs has grown in recent years.

4) FCS Stability Issues

The basic idea behind any FCS design is obtaining the control surface from approximations based on a collection of fuzzy IF–THEN rules that describe the dynamics of the controller. One of the most important difficulties with the creation of new stability criteria for any FCS has been the analytical interpretation of the linguistic part of fuzzy controller IF-THEN rules. Often FCSs are designed with very modest or no prior knowledge of a solid mathematical model, which, in turn, makes it relatively difficult to tap into many tools for the stability of conventional control systems. With the help of Takagi-Sugeno fuzzy IF-THEN rules in which the consequences are analytically derived, sufficient conditions to check the stability of fuzzy control systems are now available. These schemes are based on the stability theory of interval matrices and those of the Lyapunov approach. Frequency-domain methods such as describing functions are also employed for this purpose [10].

D. Hybrid Soft Computing Methods

Hybrid soft computing methods incorporate a combination of two or more methods from previously-described soft computing methods. We concentrate our talk on one of the most popular hybrid methods or the so-called adaptive neuro-fuzzy inference system (ANFIS).

1) ANFIS Control

An adaptive neuro-fuzzy inference system (ANFIS) is a cross between neural network and a fuzzy inference system (FIS). An adaptive network is a multi-layer feed-forward network in which each node (neuron) performs a particular function on incoming signals. The form of the node functions may vary from node to node. In an adaptive network, there are two types of nodes: adaptive and fixed. The function and the grouping of the neurons are dependent on the overall function of the network. Based on the ability of an ANFIS to learn from training data, it is possible to create an ANFIS structure from an extremely limited mathematical representation of the system. The ANFIS architecture can identify the near-optimal membership functions of fuzzy logic controller (FLC) for achieving desired input-output mappings. The network applies a combination of the least squares method and the back propagation gradient descent method for training fuzzy inference system (FIS) membership function parameters to emulate a given training data set. The system converges when the training and checking errors are within an acceptable bound. For example, the ANFIS generated by the fuzzy toolbox available in MATLAB™ allows for the generation of a standard Sugeno style FIS or a FIS based on sub-clustering of the data [21].

2) Hybrid Soft Computing Application to Cell Culture

As an example of hybrid soft computing methods application to bioinformatics, we describe a system a neuro-fuzzy control system for recombinant cell culture [42]. The introduced system has learnt the dynamics of the bioprocess in the form of a FIS and also estimated major parameters of the controlled process.

To produce a recombinant protein, it is critically important to optimize and control bioprocesses based on knowledge of a cell's genetic, metabolic, and kinetic behavior. It is, however, not straightforward due to the fact that the biosystem is highly nonlinear, time variant, and complex. Some intelligent control systems have been implemented for control of fed-batch cultivation of recombinant *Escherichia coli* and yeast, namely, fuzzy pH-stat, fuzzy neural network, and fuzzy control coupled with a neural network estimator. In a fuzzy pH-stat control system, the relationship between pH change in the medium and glucose consumption rate is modeled by a fuzzy set and subsequently used to control the feed rate of glucose to obtain cell density as high as 72 g/L. In a fuzzy neural network control system (FNN-CS), a FNN was constructed to learn fuzzy control inference and then was applied to fed-batch cultivation of recombinant *Escherichia coli* to attain a high expression of recombinant protein. In addition, a FCS was developed and coupled with NN estimators that can on-line estimate residual glucose and galactose concentrations, which were utilized to control the feed rate of glucose (during the cell growth phase) and the feed rate of galactose (during the expression phase). Such results from the application of these control strategies demonstrate usefulness in the fed-batch cultivation of recombinant strains. The idea behind these studies is to utilize predetermined experimental data to develop repetitive learning control using intelligent techniques.

VI. Systems and Computational Issues

A. The Systems Approach Heritage

Control theory is deeply rooted in systems theory - an interdisciplinary theory about the nature of complex systems as they appear in nature, society, and science. System theory is a framework through which one can study any group of objects that work together to produce some result. Systems theory originated first in biology, in the 1920s, sprang by the need to explain the correlation between organisms and ecosystems. As a technical and general academic area of study, system theory encompasses the science of systems that resulted from Bertalanffy's general system theory (GST), among others, in initiating what became a project of systems research and practice. On the other hand, systems dynamics in the sense of Forrester is an offshoot of system theory and has had application in as diverse field as urban dynamics, world dynamics, and defense systems through the so-called WHIRLPOOL and SAGE Projects and K-12 learning methods [http://www.systemdynamics.org/DL-IntroSysDyn/].

As far as control is concerned, if we take the example of parameter estimation, needed for most control applications, especially adaptive control, we can easily discern the heavy

heritage of intelligent control *vis-a-vis* system theory [43]. Thus, the systems approach is particularly useful for our proposed framework since it helps in the integration of intelligent control methods and bioinformatics in a unifying manner. The proposed framework further offers the possibility to use more specialized machine learning-oriented methods such as grammatical inference with its applications to self-assembly [15].

1) Systems Biology

The proposed three-level bioinformatics framework is to be addressed through general systems theory in the sense of Bertalanffy. The passage from DNA to the cell to the organ to the organism to community of organisms to ecosystems represents different living levels, usually addressed by systems theory.

All these levels express different behaviors and cannot be reduced to, nor understood only from, lower levels. Systems biology is based on this holistic view of biology. Going back to antiquity, holism is based on the idea that the whole is more than the sum of its constituent parts. For example, the different parts of a living organism taken separately do not tell us about what that organism might be. The functions of organisms are based not only on its constituent parts but also on the relation between them. Because one of the objectives of systems biology is the modeling of biological processes *via* mathematical models and computer simulation, it can therefore be a good candidate for integration in our proposed framework and can represent a field of predilection of intelligent control application [23]. Indeed, for many years, system biology has been part of the interests of control systems community [37].

2) Ontologies Construction

Complex problems, like the one addressed in the present work, require ontologies, *i.e.* the ways in which various entities can be grouped, related within a hierarchy, and subdivided according to similarities and clear discriminatory features. Vast amounts of biological data have been made available thanks to advances in biotechnology and experimental techniques. Data mining and *ad hoc* mathematical models provide a method of analyzing this data. However, there remain some issues that need to be addressed such as:

(i) the need for standards for defining cell models so they can, for example, be exchanged across the World Wide Web, and also read into simulation software in a consistent format, [9].

(ii) the elimination of the errors which arise with the current method of model publication. In order to address these stringent issues, markup languages have been developed. Examples of these are: *SBML* [22] and *CellML* has evolved to meet these needs of the modeling community. *CellML* is a free, open-source, eXtensible markup language-based standard for defining mathematical models of cellular function. The structure of *CellML*, its current applications, including biological pathway and electrophysiological models, and its future development—in particular, the development of toolsets and the integration of ontologies have been investigated and detailed [28].

B. Computational Aspects

1) Computational Biology vs. Biology

An area called computational biology preceded what is now called bioinformatics. Computational biologists also gathered their inspiration from biology and developed some very important algorithms that are now used by biologists.

Computational biologists take pride in the formal aspects of their work which often involves proofs of algorithmic correctness, complexity estimates, and other themes that are central to theoretical computer science. Nevertheless, the biologists' needs are so pressing and broad that many other aspects related to computer science have to be explored. For example, biologists need software that is reliable enough and can deal with huge amounts of data, as well as interfaces that facilitate the human-computer interactions (HCI) with high-resolution graphics systems and intelligent search and retrieval processes [4].

2) DNA Computing Contributions

One of the breakthroughs of computational science is that DNA can be used as a computational element. An assembly of DNA strands can process data in a similar way as an electronic computer, and has the potential to solve far more complex problems and store a greater amount of information, for substantially less energy costs than do conventional microprocessors. Thus DNA computation [6].

DNA computation was used to solve the 'traveling salesman problem (TSP)' by mixing together the strands, joining the cities connected by roads, weeding out any 'wrong answers', and finally showing that the strands could self-assemble to solve the problem [40].

A natural extension of DNA computing lies in its relation with nanotechnology. A first link between DNA computation and DNA nanotechnology was established, suggesting that short branched DNA molecules could be 'programmed' to undergo algorithmic self-assembly and serve for computation [38].

C. Available Software for Systems Study

The software used in computational biology is diversified. On just one Website, there are 31 different systems and tools *e.g.* [http://www.scfbio-iitd.res.in/bioinformatics/bioinformaticssoftware.htm].
Another example is depicted in the Web site [http://www.netsci.org/Resources/Software/Bioinform/] where various tools and solutions are made available on line such as sequence databases, pathway analysis, structure prediction and analysis, sequence analysis, sequence management, and visualization. Other Web sites offer free resources *e.g.* [http://www.clcbio.com/]. In addition to bioinformatics software reported in [16], [17] emphasis is now made on general-purpose systems, since we are concerned with biological processes as dynamical systems to be modeled and eventually controlled. The main relevant tools are depicted in Table 2 below.

Software Name / Toolbox	url
Matlab™/ *Bioinformatics Toolbox* *Control Systems Toolbox*	www.mathworks.com
Mathematica™/ *Bioinformatics* *Data Analysis and Mining*	www.wolfram.com
Maple™/ *Control Systems*	www.maplesoft.com
Stella™/ *iThink* *isee NetSim*	www.iseesystems.com

Table 2 Software for Systems Study

VII. Biological Modeling and Control

Beside *in vivo* experimentation that characterizes biology, and as far as bioinformatics is concerned, we have to inevitably use computational models for the understanding of the relevant biological phenomena [25].

Modeling is usually unavoidable because the production of data from techniques of genomic analysis is not always amenable to interpretation mainly due to the complexity of the data and the large amount of data points. Modeling can handle the data and allow the testing of a given hypothesis; for instance, whether gene A is regulated by protein B that can be verified experimentally. Hence, modeling and simulation of genetic regulatory systems [7].

A. Cell Modeling and Simulation

1) The "E-Cell "

One of the central questions in computational biology is: "what can be determined or measured to infer cell behavior"? Many attempts have been carried out to address this issue [9], [18]. Perhaps one of the fundamental results obtained so far in bridging the gap between mathematical modeling and cell behavior is the so-called *E-CELL* [34]. Using *E-CELL* Simulation Environment often requires kinetic data of biochemical reactions. However, it is generally difficult to obtain these data from literature alone. One solution is to measure values directly using wet experiments. Though the "wet-approach" is currently being done, the so-called "dry-approach" is also devised in order to estimate parameters computationally from limited data. Novel parameter estimation methods need to be developed with heavy mathematical bend. The e-cell group is currently developing methods for metabolic control analysis, mebolic flux analysis and flux balance analysis, and also the application of control theory to cell simulation [35], [http://www.e-cell.org].

2) Example of Human Disease Modeling

If we can model a given disease, we can therefore make scenario studies concerning it and eliminate it. As an example of disease modeling tool, we find *PathoSign Public.* It is a database which collects information about defective cell signaling molecules causing human diseases. While constituting a useful data repository in itself, *PathoSign* is also aimed at being a foundational part of a platform for modeling human disease processes, which is the ultimate goal of these efforts. Unfortunately, it is a descriptive tool not a prescriptive one. Indeed, it describes what happens not what we want to happen.

[http://www.gene-regulation.com/info/cytomer.html]

B. Metabolism as a Control System

Most of the structures that make up animals, plants and microbes are made from three basic classes of molecule: amino acids, carbohydrates and lipids, often called fats. As these molecules are vital for life, metabolic reactions focus on making these molecules during the construction of cells and tissues, or breaking them down and using them as a source of energy, in the digestion and use of food. Many important biochemicals can be joined together to make polymers such as DNA and proteins [12].

Many proteins are the enzymes that catalyze the chemical reactions in metabolism *i.e.* the set of chemical reactions that happen in living organisms to maintain life. Other proteins have structural or mechanical functions, such as the proteins that form the cytoskeleton, a system of scaffolding that maintains the cell shape. Proteins are also important in cell signaling, immune responses, cell adhesion, active transport across membranes, and the cell cycle [30].

1) Catabolism and Anabolism

The processes described above allow organisms to grow and reproduce, maintain their structures, and respond to their environments. There are two categories of metabolism. *Catabolism* breaks down organic matter, for example to use energy in cellular respiration. *Anabolism*, uses energy to construct components of cells such as proteins and nucleic acids. Catabolism is consumption of energy in living matter. Anabolism is responsible for production of living matter.

The chemical reactions of metabolism are organized into metabolic pathways, in which one chemical is transformed through a series of steps into another chemical, by a sequence of enzymes. Enzymes are crucial to metabolism because they allow organisms to drive desirable reactions that require energy and will not occur by themselves, by coupling them to spontaneous reactions that release energy. As enzymes act as catalysts they allow these reactions to proceed quickly and efficiently. Enzymes also allow the regulation of metabolic pathways in response to changes in the cell's environment or signals from other cells.

The metabolism of an organism determines which substances it will find nutritious and which it will find poisonous. For example, some prokaryotes use hydrogen sulfide as a nutrient, yet this gas is poisonous to animals. The speed of metabolism or metabolic rate also influences how much food an organism will require.

A striking feature of metabolism is the similarity of the basic metabolic pathways and components between even vastly different species. For example, the set of carboxylic acids that are best known as the intermediates in the citric acid cycle are present in all organisms, being found in species as diverse as the unicellular bacteria *Escherichia coli* and huge multicellular like elephants [30].

2) Metabolic Control Analysis

Metabolic control analysis (MCA) is a useful mathematical framework for describing metabolic, signaling and genetic pathways [11]. MCA quantifies how variables, such as fluxes and species concentrations, depend on network parameters. In particular, it is able to describe how network dependent properties, called *control coefficients*, depend on local properties called *elasticities*. MCA was originally developed to describe the control in metabolic pathways but was subsequently extended to describe signaling and genetic networks. More recent work has shown that MCA can be mapped directly on to classical control theory and are as such equivalent

[http://dbkgroup.org/mca_home.htm], [20]

Concerning MCA, a useful set of frequently-asked questions (FAQs) is available at the site [http://bip.cnrs-mrs.fr/bip10/mcafaq.htm]. Biochemical systems theory is a similar formalism, though with a rather different objectives. Both are evolutions of an earlier theoretical analysis of sequential reactions dating back to the early sixties.

3) Gene Regulation

The genome of a given organism contains thousands of genes, but not all these genes need to be active at any given moment. A gene is expressed when it is being transcribed into mRNA, and translated into protein, and there exist many cellular methods of controlling the expression of genes such that proteins are produced only when needed by the cell. Gene regulation gives the cell control over structure and function, and is the basis for cellular differentiation and morphogenesis. It is also responsible for the versatility and adaptability of any organism. Gene regulation may also serve as a substrate for change, since control of the timing, location, and amount of gene expression can have a profound effect on the functions or actions of the gene in a cell or in a multicellular organism [41].

[http://www.news-medical.net/health/Genetics-and-Gene-Expression.aspx]

4) From Molecular to Gene Regulatory Networks

4.1 Molecular Regulatory Networks

The cells physiological responses to external and internal stimuli are governed by genes and proteins interacting in complex networks whose dynamical characteristics are impossible to understand by intuitive reasoning alone. Recent advances in theoretical biology have demonstrated that molecular regulatory networks and particularly gene regulatory networks can be accurately modeled in mathematical terms. These models give insight to the design principles of biological control systems and make predictions that have been verified experimentally [36].

4.2 Gene Regulatory Networks

Owing to the multivariate ways and means in which genes manage cellular function, including their regulatory effects on each other, the modeling of gene regulatory networks is a prominent issue in systems biology. Appropriate network modeling is critical to understanding the manner in which cells execute and control the huge number of operations required for normal function. Precise network modeling can detect the failure in cellular systems that occurs in disease.

Many approaches to modeling gene regulatory networks have been proposed, each with its own assumptions, data requirements, and goals, including linear models, Bayesian networks, neural networks, nonlinear ordinary differential equations, stochastic logical networks, and graph-based models [5].

4.2 Epistemological Translation

Epistemology can act as an abstract framework for translation of a given science. Epistemology, or theory of knowledge, is a branch of philosophy concerned with what constitutes knowledge, its limits, and how it is acquired by people. For something to count as knowledge, it must be true. Beliefs are not necessarily knowledge. An attempt of translating one science has recently been made in bioinformatics; the science to be translated is genomics, the framework is genomic signal processing, and the goal of the translation is to provide therapeutic strategies based on controlling gene regulation. The problem considered here is different in two ways. First, the scientific model is not a stochastic time series or random set, but a dynamical network whose probabilistic description is characterized *via* a Markov chain and, second, the translational problem will be to alter the steady-state distribution of the network [8].

VIII. Impacts of Proposed Framework

The three-level framework is shown in Figure 3. We believe that the study and integration of previously-described theories will advance our knowledge of biological processes based on the most powerful theoretical and technological tools available to computer and control scientists, entailing a better understanding of molecular biology. The impacts on many fields of research are expected to be important, not only on computer science and control theory *per se* but also on medicine, pharmacy and technology at large. We expect impacts of our framework on the following fields of research and technology.

A. Impacts on Bioinformatics

1. To further formalize bioinformatics problems and solutions.
2. To speed up the process of modeling and simulation of biological processes by the use of *ad hoc* intelligent control methods. Singular value decomposition (SVD) can used to reduce data [1], [32]
3. To contribute to the identification of unknown sequence patterns across single or multiple DNA and protein sequences through grammatical inference models. If the analysis is performed on several sequences at a time the

method has to search for patterns which are common between all the sequences.
4. To reduce the time of drug production, inducing a considerable impact on biochemistry and pharmacy.
5. To contribute to the design of drugs for therapy and for existing and actually incurable diseases
6. To ultimately contribute to the design of useful ethically-monitored biological systems, as a far-reaching goal.

B. Impacts on Biological Regulation and Control

1. To integrate stability and control for biological processes considered as complex dynamical systems by relying on catastrophe theory for the prediction and formation of biological sequences [2], [33].
2. To contribute to the understanding of the regulome, *i.e.* the set of regulation components in a cell. This will allow the modeling of cell behavior with greater precision [41].
3. To contribute to the understanding of metabolic regulation and control of cell differentiation and specialization (the so-called stem cells), hopefully leading to the prediction and reduction of major diseases [7].
4. To contribute to functional genomics; a major goal of functional genomics being to identify genes that determine specific cellular malfunctioning and model their activity in such a way as to distinguish between normal and abnormal behaviors. Once modeling is done, then external corrective intervention *via* control becomes possible.
5. To ultimately act as "genomics-oriented translator" by finding therapeutic strategies based on controlling gene regulation using genomic signal processing (GSP).

	AI-free methods	*AI-based methods*	*Intelligent control methods*
Bioinf. Level 1	Yes *e.g.* Sequence Alignment and Analysis	No use	No use
Bioinf. Level 2	Yes *e.g.* 1/IT in Biol. and Health 2/DBMS	Yes *e.g.* 1/Data Mining 2/Biomedical Eng. 3/General Biological Properties 3/ Tools 4/Protein Structure 5/Function Proteomics	No use
Bioinf. Level 3	Yes *e.g.* 1/Microarray Analysis 2/Biological Modeling & Classification	Yes *e.g.* 1/Gene Expression 2/Regulatory Network 3/Genomics and Systems Biology 4/Personalized medicine 5/Health Informatics	Yes *e.g.* 1/Cancer Informatics 2/Computat. Biology 3/Drug Design 4/ Functional Informatics 5/Tissue Design 6/Genomics Translation using GSP

Table 3 Three-Level Bioinformatics

IX. Conclusion

We have presented a framework for bioinformatics, incorporating intelligent control. It is highly expected that intelligent control coupled with machine learning will uncover more useful structures hidden in biological sequences. On top of actual query search methods now available, however intelligent these might be, future public bioinformatics platforms have to include an array of "what-if" simulation scenarios capable of producing intelligently-controlled and / or intelligently-produced novel elements right from the Web. It is hoped that *via* intelligent control, specialists can use external actions to control and / or produce novel elements, *e.g.* new drugs for existing and actually incurable diseases or novel useful biological systems. The proposed framework represents an early contribution to this far-reaching goal. Further integration of diverse theories from machine learning, control theory and bioinformatics will remain indeed a challenging task for a long time.

Acknowledgment

This work has been gratefully supported by the Deanship of Scientific Research, Qassim University, Saudi Arabia, under Contract # SR-D-007-30.

References

[1] Alter O., P.O. Brown, D. Botstein. "Singular value decomposition for genome-wide expression data processing and modeling," *Proc Natl. Acad. Sci. USA*, vol. 97, pp. 101-6, 2000.

[2] Arnold V.I. (Ed.). *"Dynamical Systems V: Bifurcation Theory and Catastrophe Theory"*. Vol. 5 of *Encyclopedia of Mathematical Sciences*, Springer-Verlag, Berlin, 1994.

[3] Araujo R.P, L.A. Liotta. "A control theoretic paradigm for cell signaling networks: a simple complexity for a sensitive robustness", *Current Opinion in Chem. Biol.,* 10(1):81-87, 2006.

[4] Cohen J. "Bioinformatics—An introduction for computer scientists", *ACM Computing Surveys*, 36(2):122-158, June 2004.

[5] Datta A., R. Pal, A. Choudhary, E.R. Dougherty. Control Approaches for Probabilistic Gene Regulatory Networks", *IEEE Sign. Proc. Mag.*, p.54-63, Jan. 2007.

[6] Deaton, R., J.-W. Kim, J. Chen. "Design and test of noncrosshybridizing oligonucleotide building blocks for DNA computers and nanostructures", *App. Physics. Lt.,* 82(8):1305-1307, 2003.

[7] DeJong, H. "Modeling and simulation of genetic regulatory systems: A literature review, *J. Comput. Biol.* 9(1):67–103, 2002.

[8] Dougherty, E.R., "Epistemology and the Role of Mathematics in Translational Science" In Tabus, I., K. Egiazarian, M. Gabbouj, (Eds.) *Festschrift in Honor of Jaakko Astola on the Occasion of his 60th Birthday,* Tampere International Center for Signal Processing, TICSP Series #47, 60-78, 2009.

[9] Endy D., R. Brent, "Modeling cellular behaviour," *Nature*, 409:391-395, 2001.

[10] Farrell J.A., M.M. Polycarpou. *"Adaptive Approximation Based Control: Unifying Neural, Fuzzy and Traditional Adaptive Approximation Approaches"*, John Wiley & Sons Ltd., 2006.

[11] Fell D., "*Understanding the Control of Metabolism"*, Portland Press, 1997.

[12] Goutsias J. "A hidden Markov model for transcriptional regulation in single cells", *IEEE/ACM Trans. on Comp. Biol. And Bioinf.* 3(1):57-71, 2006

[13] Hamdi-Cherif A., N. Golea, M.-L. Hadjili. "Modal control by microcomputer : a sample program from the *CASCIDA* Project", *Proc. of the Eur. Simul. Multiconf. (ESM'94)*, Barcelona, Spain, pp. 748-752, 1-3 June 1994.

[14] Hamdi-Cherif A. "The *CASCIDA* Project - A computer-aided system control for interactive design and analysis". *Proc. of IEEE/IFAC Joint Symp. on CACSD (CASCD'94),* Tucson, AZ, USA, p. 247-251, 1994.

[15] Hamdi-Cherif A. "Grammatical Inference Methodology for Robotic Self-Assembly", *WSEAS Trans. on Computers,* 8(4): 610-619, April 2009.

[16] Hamdi-Cherif A. "Machine learning for intelligent bioinformatics – Part 1 machine learning integration", Invited Paper In *9th WSEAS Int. Conf. on Artif. Intell. and Knowl. Eng. and Data Bases (AIKED'10)*, Cambridge, UK, 20-25 February 2010.

[17] Hamdi-Cherif A. "Integration of machine learning and intelligent bioinformatics", *WSEAS Trans. On Computers*. 4(9):406-417, April 2010.

[18] Heinrich R., S. Schuster, *"The Regulation of Cellular Systems"*. Chapman & Hall, 1996.

[19] Higo K., Y. Ugawa, M. Iwamoto, T. Korenaga. "Plant cis-acting regulatory DNA elements (PLACE) database", *Nucl. Acids Res.*, 27(1):297-300, 1999.

[20] Ingalls B.P. "A Frequency domain approach to sensitivity analysis of biochemical systems", *J. of Phys. Chem.* ***B*****,** 108:1143-1152, 2004.

[21] Jang, J.S., Sun, C., and Mizutani, E., *Neuro Fuzzy and Soft Computing*, Prentice Hall, Upper Saddle River, NJ, 1997.

[22] Hucka M., A. Finney, H.M. Sauro *et al.* "The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models". *Bioinf.* **19** (4):524–31, 2003.

[23] Jardon M. "Systems biology: an overview", *The Science Creative Quaterly*, Issue 4, July 2009, http://www.scq.ubc.ca/systems-biology-an-overview/

[24] Kecman V. *"Learning and Soft Computing : Support Vector Machines, Neural Networks and Fuzzy Logic Models"*, MIT Press, 2001.

[25] Keedwell E., A. Narayanan *"Intelligent Bioinformatics - The Application of Artificial Intelligence Techniques to Bioinformatics Problems"*, John Wiley & Sons Ltd., 2005.

[26] Lancashire, L.J., Ch. Lemetre and G.R. Ball. "An introduction to artificial neural networks in bioinformatics—application to complex microarray and mass spectrometry datasets in cancer studies", *Briefings in Bioinf.,* 10(3):315-329, 2009

[27] Liu Y., H.B. Sun, H. Yokota. "Regulating gene expression using optimal control theory. *Proceedings of the Third IEEE Symposium on BioInf. & BioEng. (BIBE'03)*, p. 1-6, 2003.

[28] Lloyd C.M., M.D. Halstead, P.F. Nielsen. "CellML: its future, present and past". *Prog. Biophys. Mol. Biol.* **85** (2-3): 433–50, 2004.

[29] Moreau Y., F. De Smet, G. Thijs, K. Marchal, B. De Moor. "Functional bioinformatics of microarray data: from expression to regulation", *Proc. Of the IEEE,* 90(11):1722-1743, Nov. 2002.

[30] Salter M, Knowles R, Pogson C. "Metabolic control". *Essays Biochem* **28**:1–12, 1994.

[31] Searls D.B. "Grand challenges in computational biology", In S.L. Salzberg, D.B. Searls, S. Kasif, (Eds.), *Computat. Meth. in Molec. Biol.*, Elsevier, Amsterdam, 1998.

[32] Simek K., K. Fujarewicz, A. Świernia, M. Kimmel, B. Jarząb, M. Wiench, J. Rzeszowska. "Using SVD and SVM methods for selection, classification, clustering and modeling of DNA microarray data", *Eng. App. of Artif. Intel.,* 17(4):417-427, June 2004.

[33] Thom R. *"Structural Stability and Morphogenesis",* Benjamin-Addison Wesley, 1975.

[34] Tomita M., K. Hashimoto, K. Takahashi, T. Shimizu, Y. Matsuzaki, F. Miyoshi, K. Saito, S. Tanida, K. Yugi, J. C. Venter, C. A. Hutchison. "E-CELL: software environment for whole cell simulation". *Genome Inform. Ser. Worksh. Genome Inform.* 1997;8:147-155. http://web.sfc.keio.ac.jp/~mt/mt-lab/publications/Paper/ecell/bioinfo99/btc007_gml.html

[35] Tomita M., "Whole-cell simulation: a grand challenge of the 21st century", *TRENDS in Biotech.* 19(6):205-210, June 2001. http://tibtech.trends.com

[36] Tyson J.J., K.C. Chen and B. Novak. "Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell", *Curr Opin Cell Biol* **15**:221–231, 2003.

[37] Wellstead, P. "Systems Biology and the Spirit of Tustin", *IEEE Control Systems Magazine*, pp. 57-71, Feb. 2010.

[38] Winfree E. "Algorithmic self-assembly of DNA: theoretical motivations and 2D assembly experiments", *J. Biol. Mol. Struct. Dynamics Conversat*. 11(2):263–270, 2000.

[39] Wood, M.J., Jonathan D. Hirst. "Recent Applications of Neural Networks in Bioinformatics". In *Biol. and AI Env., 15th Italian Worksh. on NNs, WIRN VIETRI*, pp. 91-97, Springer 2004.

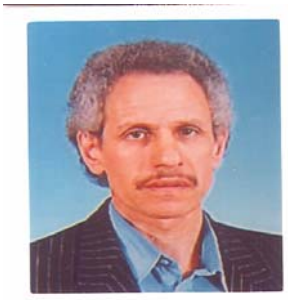
[40] Xiong F., D. Spetzler, W. D. Frasch. "Solving the fully-connected 15-city TSP using probabilistic DNA computing", *Integrative Biol.*, 1:275-280, 2009.

[41] Wolkenhauer O., M. Ullah, P. Wellstead, K.-H. Cho. "The dynamic systems approach to control and regulation of intracellular networks", *Fed. of Eur. Bioch. Soc. (FEBS) Letters* 579:1846–1853, 2005.

[42] Ye K., S. Jin, K. Shimizu "On the development of an intelligent control system for recombinant cell culture", *Int. J. of Intel. Syst.*, 13(6):539 – 560, 1998.

[41] Zilouchian A., M. Jamshidi (Eds.). *"Intelligent Control Systems Using Soft Computing Methodologies"*, CRC Press LLC, 2001.

Aboubekeur Hamdi-Cherif was born in Setif, Algeria. He received BSc (Honors) in Electrical Engineering and MSc in Electronic Control Engineering, both from Salford University, Manchester, England, and PhD degree in Computer Science (AI in Robotics) from Université Pierre et Marie Curie Paris 6, (UPMC), Paris, France. He is member of IEEE and ACM.



He worked as Control Engineer with Algerian Petroleum Company SONATRACH. He taught at Ecole Supérieure des Transmissions, Algiers, Algeria, at Université de Bretagne Occidentale, (UBO) Brittany, France, at Ecole Supérieure Libre des Sciences Commerciales Appliquées (ESLSCA), Paris, France, and at Ferhat Abbas University, Setif, (UFAS), Algeria. While at UFAS, he was Director of Postgraduate Studies and Scientific Research, reporting to the Vice-Rector. In 2001, he joined Computer College, Qassim University, Saudi Arabia, where he is Associate Professor and Deputy Head of Computer Science Department. He supervised about 80 BSc student projects, and 10 master and doctoral students. He is currently interested in bioinformatics, machine learning as applied to control and grammatical inference, and Arabic language processing.

In 1998, he was awarded the best technological thesis prize by UFAS. Dr. Aboubekeur received financial support for his research projects from Université Pierre et Marie Curie Paris 6, Paris, France, from UFAS, from Qassim University, and from King Abdulaziz City for Science and Technology (KACST), Riyadh.