Electronic version of an article published as [Journal of Biological Systems, Volume 21, Issue 01, 1350008 (2013) [23 Pages] DOI: 10.1142/S0218339013500083]

ON MODELING OF LIVING ORGANISMS USING HIERARCHICAL COARSE-GRAINING ABSTRACTIONS OF KNOWLEDGE

YOUCEF DERBAL

TRS of Information Technology Management, Ryerson University, 350 Victoria Street Toronto, Ontario M5B 2K3, Canada
yderbal@ryerson.ca

High throughput technologies such as gene expression microarray, ChIP-chips, siRNA and protein arrays, and high throughput mass spectrometry are enabling an ever increasing amount of data becoming available about DNA, RNA, proteins, metabolites as well as biological pathways and networks. The knowledge embedded in this data deluge needs to be recast in forms that lend themselves to analysis with the expectation of developing analytical instruments to gain insight and answer questions about life and living organisms. The powers of abstraction and model building are fundamental to the quest of making sense of the biological complexity embedded in these biological and clinical datasets. The modeling of living organisms is explored with a proposed framework for model representation of biological complexity. The principal foundational assumption of the proposed modeling philosophy recognizes the symbiotic relationship between information and energy flows, required for the transformation of matter, as a fundamental organizing force underlying the observable nature of living organisms. The use of the concept of regularities to refer to complexity of structure, function and dynamics alike provides a unified approach to the reasoning about the integration of knowledge representations of varying natures and scales of granularities. The application of the proposed modeling approach is illustrated in broad qualitative terms for the human organism.

Keywords: Multi-Scale Modeling; Effective Complexity; Biological Regularities.

1. Introduction

High throughput technologies such as gene expression microarray, ChIP-chips, siRNA and protein arrays, and high throughput mass spectrometry are enabling an ever increasing amount of data becoming available about DNA, RNA, proteins, metabolites as well as biological pathways and networks. In theory, an unprecedented opportunity lies before our eyes to shed light on the structure and dynamics of biological processes through model building and analysis. Modeling and analysis will in turn contribute to the ongoing quest to elucidate the nature of life's hierarchy of complexity tying gene regulation to the states of health and disease of living organisms. The implications of this enterprise of scientific discovery span a large application space of human needs and interests which include the development of therapies and pharmaceutical drugs to treat diseases, safe food production, and the

preservation of a healthy biosphere. Computational modeling and simulation are believed to be key to twenty first century biological research and its quest to provide a satisfactory understanding of the nature of life and its processes [1]. Modeling is essentially the substitution of one knowledge representation with another. Different knowledge representations are built using different abstractions and analytical tools, and draw from different perspectives of knowledge interpretation. Models are attempts to recast knowledge in forms that lend themselves to analysis and provide as a result analytical instruments to develop insight and answer questions about the process, mechanism or system under study. Ultimately, the convergence to insight through modeling and analysis is predicated on the assumption that the models are less complex than the system under study. This assumption holds true in general, since models are inevitably confined by the abstractions and knowledge representation approaches they rely on and as a result they cannot embody the full complexity of the system being modeled. Therefore, it may be reasonable to assume that no single model view can shed light on the full complexity of a system. Model integration mechanisms are therefore necessary to yield a coherent insight about the nature and dynamics of biological processes by connecting model views that are distinct in nature, scope and scale. The proposed modeling philosophy explores the conception of such integration mechanisms starting from the assumed existence of organizing principles that support a putative thread of unity among various aspects of a complex biological system.

The powers of abstraction and model building are fundamental to problem solving. We use these capabilities as systematic intellectual tools to handle the complexity of the real world. From the planning of a family vacation to the forecasting of the weather, human activities are weaved around countless types of physical and conceptual models. We often speak of abstract models to capture a mental representation of something real or conceptual. These models enable reasoning in the midst of problem space diversity that makes up human reality from the mundane tasks of grocery store visits to the construction of skyscrapers. These models, whether explicitly formed, represented and used or implicitly incorporated in our thinking processes are the corner stones of all human decision making processes. In biological research, qualitative models are the most accessible and intuitive manifestations of abstraction. They summarize facts and knowledge within a representation framework that facilitates communication of a specific understanding. One example illustration of this modeling approach is the qualitative representation of regulatory dynamics provided by Thieffry et al. [2]. Graphical models of signaling pathways, protein interaction networks, and genetic associations embody integrated representations and bring enhanced clarity often indispensable to further analysis of the issues at hand [3,4,5]. The often limited scope of these qualitative models, such as capturing the topology of a signaling cascade without accounting for kinetics or thermodynamics, allows the separation of concerns in dealing with the overwhelming multi-factorial complexity of biological mechanisms. However, these qualitative models are limited in their ability to assist in dealing with the ever growing biolog-

ical datasets and the perplexing information complexity they seem to be carrying. Observing the unraveling of biological complexity through high throughput experiments is analogous to the opening of a black box that seems to lay bare to the eye more and more components with no accompanying hints about their mutual relationships or their contributions to biological function or disease. In this respect, visualization methods may bring a degree of clarity through graphical and video animated representations of relationships and pathways, hence facilitating the gaining of biological insight. Gehlenborg et al. [6] discusses the challenge of meaningfully visualizing Omics data in relation to protein interaction, gene expression and metabolic pathways. These methods are without a doubt valuable in the quest to discover and identify the entities and networks involved in biological processes. However, the problem of uncovering and understanding the biological basis for function and disease cannot fully be solved by the mere availability of detailed descriptions of these biological components and their associated signaling and metabolic pathways and networks. Modeling is inevitably necessary because it brings with it a framing coherency that glues back these identified components into system models where the principles and constraints responsible for the experimentally observed organization of the biological system have been substituted with organizing threads of knowledge representations. Such metamorphic gluing of biological data into model representations, which must be analytically approachable as well as computable, is essential to squeezing out the key understanding and insights necessary to advance life science and medical research in the many areas of human interests, including: (1) screening for effectiveness and side effect of drugs and therapies, (2) generation and analysis of hypotheses about the biological, biochemical and genetic underpinning of function and disease, (3) conception of biologically plausible models of organism development, growth, reproduction, and epigenetics to support human health and well being, and (4) conception of ecological models of life and the biosphere.

Modeling of biological processes and systems has been partially fueled by the qualitative models of statistical basis such as Bayesian networks [7], principal component analysis [8], and hidden Markov models [9,10]. These models are used to identify patterns and correlations in biological datasets, leading to conceptual reconstruction of signaling pathways, transcription networks and the finding of new protein coding genes. Boolean networks are qualitative logical models that have been used to study hypotheses about specific aspects of biological structure and dynamics [11]. For the modeling of gene regulation, Boolean networks and digital logic-based models rely on a binary information processing abstraction of gene expression mechanisms. In essence, such models reduce the structure and dynamics of the gene expression machinery (genes, proteins, transcription factors, RNA Polymerase complex) to a binary decision making unit not unlike a digital integrated circuit [12,13]. These logical models are instrumental in exploring the logic systems underlying gene regulation, signaling cascades and protein interactions [14,15,16,17]. However, most logical models rely on reductionist abstractions that direct their focus away from essential elements of the biological context such as thermodynamics and kinetics. As a result they are

intrinsically limited in their capacity to predict biological behaviour. Stoichiometric models of biochemical networks have a similar limitation which may be overcome with the use of constraint-based optimizations to reduce the set of possible network solutions [18]. On the other end of the modeling spectrum, quantitative models provide a bottom-up representation of the physicochemical mechanisms underlying biological processes using the mathematical methods of differential calculus, including ordinary and partial differential equations (ODEs, PDEs) and their stochastic forms [19,20,21,22,23,24]. These mechanistic models often require detailed knowledge of rates and parameters that is not always possible to obtain. Given this limitation, hybrid modeling approaches may be conceived where quantitative models are utilized to account for the mechanistic processes at some chosen level of resolution while logical abstractions are used to integrate the model. In this respect, the choice of the levels of physical and logical resolutions of modeling is essential to advancing our understanding of biological complexity [25].

2. In Search of an Effective Modeling Paradigm

The modeling of biological complexity can be approached through the conception of organizing principles to enable the development of integration frameworks that are operational across the multiple levels of abstractions necessary to represent the different views, scales and aspects of biological processes and their underlying mechanisms. The reliance on organizing principles that are biologically plausible may translates into models that embody a satisfactory capacity to predict biological behaviour and facilitate the reasoning about the nature of biological pathways and dynamics associated with biological function and disease. The logical conception of such organizing principles requires a closer look into the nature of biological complexity and how it may be reconstructed through the application of logical, quantitative and hybrid modeling methods to biological knowledge and data. The formidable scale of this model-based reconstruction challenge has enlisted a variety of perspectives spanning a wide spectrum of insights, all attempting to frame the issues most central to biological processes [25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41]. Irrespective of the insight driving the inquiry into the nature of life and living organisms, the sought after models of biological processes must internalize representations of biological complexity that are both accessible to analytical reasoning and computationally realizable. The effectiveness of model representations will depend on the capacity of the resulting models to yield analytical insights and offer biologically plausible simulation results that help explore the fundamental biological questions of development, growth, reproduction and disease.

Computational modeling has been applied to many areas of biological research with increasing impact on advances in biological and medical research [1]. The modeling methods and approaches being used range from strategies that are based on fine grain mechanistic algorithms with limited scope and context to coarse grain logical abstraction models. The application of different modeling strategies to divers

biological scopes yield different model views of biological complexity. These views lead to a diversity of insights about the various physiological, pathological, genetic and epigenetic dimensions of living organisms. An integrated model of these views is however needed to frame the coupling and the dynamics of interactions between the relevant biological processes. Furthermore, it may be reasonable to assert that the success of any model representation of biological complexity depends on the capacity of the modeling philosophy to provide constructs that enable the reasoning about biologically observed properties such as robustness, stability, and adaptation and their relationships to the overall physiological and pathological states of living organisms.

The notion of effective complexity put forth by Gell-Mann and Seth Lloyd [42,43] embodies a perspective about the nature of complex system that may help in the exploration of potential answers to the model integration issue raised in the previous section. Effective complexity of a system is defined as the minimum amount of information necessary to describe its regularities [42]. Regularities may be generally defined as any non-random structural or dynamic patterns. Applying this definition to biological systems, the effective complexity of a living organism would be the minimum amount of information needed to describe its structural and dynamic regularities such as the DNA transcription process, the citric acid cycle, the DNA repair process, proteins, genes, biological oscillators, oscillations, and the lipid bilayer of a cell membrane. The explicit consideration of regularities as the elemental components of biological complexity may facilitate the reasoning about knowledge representation and the degrees of representational granularity and resolution as they apply to structure, function and dynamics alike. Using the above definition of complexity, the proposed modeling approach is built on the assumption that living organisms reduce to a countable set of interacting regularities (structural, functional, dynamic and combination of thereof). Furthermore, it is also hypothesized that these regularities are dedicated to the transformation of matter for development, growth and reproduction through energy use and transfer. The operational logic of these regularities is assumed to be determined by the genetic and epigenetic DNA information content and the programs of information processing and transfer encoded therein. Ultimately, these information processing programs determine how energy is generated, transferred and used to enact and guide matter transformation. Reciprocally, energy is a necessity for all information processing activities including DNA Replication, signal transduction, and intercellular communication (see Fig. 1).

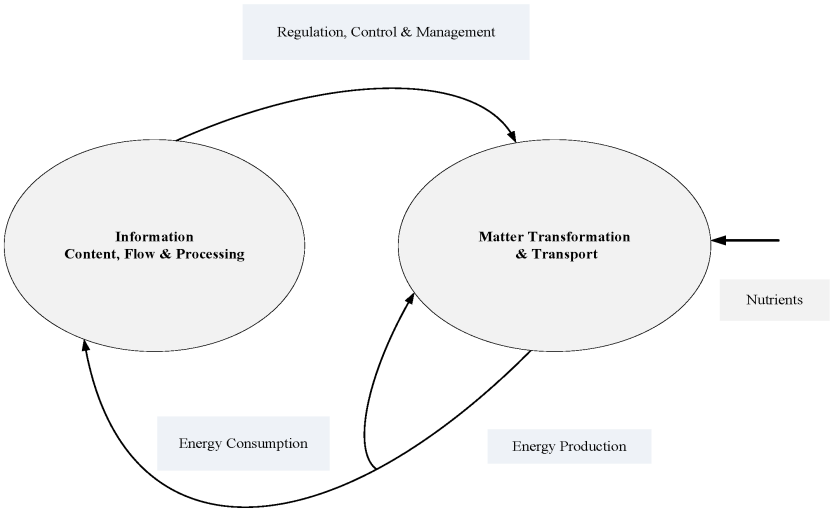

Fig. 1. Information, Energy and Matter. The symbiotic relationship between information, energy, and matter is hypothesized to be the fundamental organizing force underlying the observable nature of living organisms. Among the manifestations of such organizing force is the transformation of nutrients (matter) into proteins through the formation of peptide bonds between the constituent amino acids. This transformation (comprising the processes of RNA transcription, post-transcription modification, RNA translation, protein folding, etc.) is enacted, controlled and regulated by a biochemical machinery (involving RNA polymerases, transcription factors, spliceosome, tRNA adaptors, enzymes, ribosomes, etc.) operating under the instructional guidance of the DNA information content of the genes, the non-coding regions, and DNA imprinting. Furthermore, the translation of genes into proteins involves a transfer/flow of information (genetic sequence) from DNA to proteins where it is internalized as polypeptide sequences of amino acids. Such information is key to protein folding and ultimately to the functions of proteins as the workforce enabling development, growth and maintenance of living organisms. This transformative flow of information would be impossible without the energy generated and made available for the cell as ATP (Adenosine Triphosphate), mostly by mitochondria in eukaryotes. The electron-transport chain involved in the energy generation includes among other components the membrane proteins of mitochondria and ATP Synthase. The information content of these proteins is sourced from the DNA, closing hence the symbiotic loop between information, energy and matter.

Overall, the principal foundational assumption of the proposed modeling philosophy recognizes the symbiotic relationship between information and energy flows required for the transformation of matter as a fundamental organizing force underlying the observable nature of living organisms. Potential links may be conceivable between this assumption and a recently proposed approach to the study of organizational phenomena which conjunctures that the evolution of flow matter/energy within the physical reality is manifest as self-similar, scale invariant structures and processes emergent in all the layers of an organizational hierarchy [34]. One may postulate from the assumption being put forth that the development of biologically plausible models of living organisms is best approached through the representation of biological regularities with a deliberate consideration of information flow and the constraints on energy generation, consumption, flow and availability. The output models must as a result be architecturally equipped with mechanisms to integrate the multi-scale representations of regularities ranging from the elementary DNA information processing mechanisms such as replication and transcription to the high level functions of the Physiome. Biological plausibility of the resulting models may be entrenched by design using integration strategies that are conceived through inspiration from the experimentally observed organizing constraints on matter, information and energy flows respectively. The next sections will explore the conceptual elements of the proposed modeling philosophy and provide some practical strategies to apply them towards the development of computational models of biological processes.

3. Modeling Biological Complexity

Computational models of living organisms embody representations of knowledge about biological regularities (biological entities and processes). However, unlike descriptive knowledge, computational models incorporate a mathematical accounting for the dimensions of time and space, enabling hence the construction of virtual models of living organisms whose space-time dynamics can be simulated using a computer. The multi-scale nature of biological processes interacting within the confines of a living organism may require for their modeling a resolution hierarchy of knowledge representation. In such resolution hierarchy, ODE based models are more appropriate whenever fine grain modeling is necessary to capture the observable biological behaviour such as is the case for metabolic reactions. On the other hand coarse grain models based on Statecharts [44], for example, may be adequate to represent the state dynamics of the cell cycle. Model integration of fine and coarse grain representations of biological regularities require a multi-scale structure that echoes the natural organization of multi-cellular organisms whose constituency scaffold includes basic elements (Carbon, Oxygen, Nitrogen) , nucleotides (Adenine, Cytosine, Guanine, Thymine, Uracil), DNA/RNA, proteins, cells, organs and systems. Such multi-scale structure can be realized using a layered architecture, where the lower layers are populated with fine grain representations of mechanistic processes re-

lated to information storing, reading, replication (DNA transcription, translation, and replication), and the underlying biochemical kinetics. On the other hand, the successive higher layers would be populated with increasingly coarse grain representations of higher level biological processes associated with relevant biological entities along the hierarchy of the organisms constituency scaffold. The resulting computational model has multiple layers, each made up of interacting computational modules (CMs) to model the relevant biological regularities with the appropriate granularity in accordance to the model resolution hierarchy articulated above (see Fig. 2).

In the proposed modeling strategy, the effect of the organizing constraints of energy flow, matter transformation and information flow in living organisms is realized through hierarchical state feedback between the layers of the computational model. In particular, the computational modules representing the genetic and epigenetic programs of development, growth and reproduction drive the model's state dynamics in accordance to the DNA information content (including genomic imprinting) of the living organism. The dynamics of the model are also constrained, through hierarchical state feedback, by the output of the CMs that represent the processes of adaptation to environmental pressures and constraints to which living organisms are subjected, including the varying supply of food, energy, oxygen as well as the exposure to toxins, pathogens and trauma. Furthermore, hierarchical state feedback would provide a natural mechanism of structural integration between CMs such as would be the case in linking CMs that model metabolic pathways and lower level CMs that model gene networks that regulate the expression of the relevant enzymes.

The model architecture with its embedded resolution hierarchy of knowledge representation is intended to reflect the assumption that a living organism may be viewed as a system layered in relation to the biological processes relevant to the hierarchy of the organism constituency scaffold (Basic Chemical Element, Nucleotide, DNA/RNA, Protein, Cell, Organ, and System). Furthermore, each layer of such system is also assumed to embody a complexity of a different nature in the sense that it requires for its modeling a set of computational modules that are layer-specific in their resolution of modeling and underlying abstractions.

The hierarchy of dynamics (regularities) represented with different degrees of granularities is one of the defining elements of hierarchical control systems which are often qualified as intelligent [45,46,47,48]. One of the design principles of these human engineered systems is to employ high precision control models for the lower levels of the hierarchy, and rely on intelligent models of decision making for the higher levels of the hierarchy. In particular, differential calculus is used to formulate control strategies with a desired control precision at the lowest levels of the system, often associated with actuation. Models that utilize coarse grain reasoning, such as fuzzy logic, are used for high level supervisory and management control, which take instructions and commands from within the context of human mundane reality. This design philosophy is often summarized as entailing an inverse relationship between

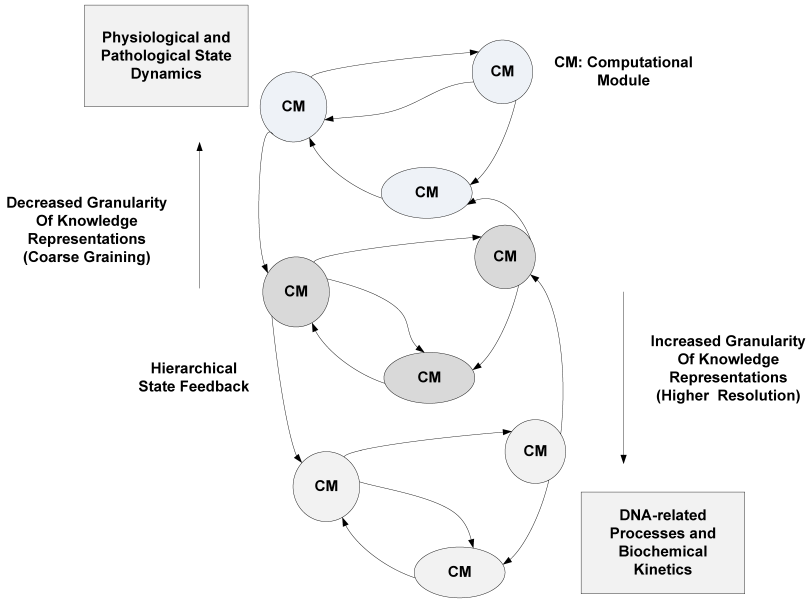

Fig. 2. Hierarchical Model of Living Organisms. Living organisms are modeled as a hierarchy of layers comprising representations of biological processes distinguishable by the resolution, interpretation, and processing strategy of the relevant biological information. The layers are interdependent through state feedback but they each embody a complexity of a different nature. For example, the proliferation of cancer and the morphological state of the associated lesion may be represented in the highest model layer using an FSM (finite state machine). In a lower layer of the model, the dynamics of production and diffusion of chemokines and cytokines in the tumor's environment may be modeled using ODEs (Ordinary Differential Equations) and PDEs (Partial Differential Equations). The two representations (ODE/PDE and FSM) are addressing two distinct types of complexity. Although the morphological state of the lesion may depend in its emergence on the circulation of chemokines and cytokines in the tumor's environment, the complexity of the lesion state dynamics, embodied by the FSM, is not reducible to the complexity, embodied by the ODEs/PDEs, of the dynamics of production and diffusion of chemokines and cytokines. Put differently, one may be able to reason about the health of the affected individual based on the state of the lesion. This involves a complexity associated with predicting the cancer's course of progression and the most appropriate therapies based on clinical knowledge. On the other hand attempting the same assessment of the individual's health by inspecting the distribution of chemokines and cytokines instantiates a complexity of a different nature in the sense that it may involve the need to consider inter-cellular communication mechanisms mediated by these cell signaling proteins and their role in cancer progression.

intelligence and precision. In other words, the lower layers of the hierarchy realize high precision control with a reduced capability for intelligent decision making, whereas the higher layers of the hierarchy exercise intelligent decision making with reduced precision of control. The applications of this design philosophy to human engineered systems such as trains, planes and power plants provide an instructive insight about the effectiveness of hierarchical control models and the relationships

between the distinct representations of regularities that must be integrated in the computational models of these systems.

The proposed multi-scale model of biological systems carries a significant mark of analogy to successfully designed hierarchical control systems. The highlighted parallels between engineering and biological systems provide the motivation for the applications of mathematical and computational engineering tools to the development of models of living organisms. However, for these models to be biologically plausible they must have an adequate accounting for the biological complexity emergent from the working of the DNA-information driven metabolic reactions that transform matter and energy, and the gene-regulated biological mechanisms of adaptation and robustness that maintain the organism's functions under the constraints and pressures of the environment. Ultimately, useful models of living organisms must be biologically plausible both in structure and behaviour. Specifically, such models must internalize the causal chains of biological mechanisms that lead to the observable physiological and pathological behaviours of interest. In addition, computer simulations of these models must exhibit the expected physiologic and pathologic behavioral scenarios of the living organism for specific environmental and internal state conditions. The physiologic and pathologic behavioral scenarios may be defined based on the time profile of the measurable or observable biological variables that convey physiological or pathological information about living organisms such as metabolite and protein concentrations, blood pressure, tissue lesion size, inflammation status, antibody count, and viral count. Let $\mathcal{V}$ be the set of such physiological and pathological variables, then $\forall v \in \mathcal{V}$ there is a set P_v^T of time traces, referred to as profiles, of the values of the variable v during the time period T. Different profiles $p_v(T) \in P_v^T$ for a given physiological/pathological variable result from different environmental and internal conditions of the living organism. An organism's behavioral scenario $s(T)$ is defined as the collection of the concurrent profiles of all $v \in V$ for the same time period T under the assumed environmental conditions. Note that for any given time period T a physiological/pathological variable v has a single profile $p_v(T)$ for the specific environmental conditions being assumed.

Integral to the modeling approach, all asserted or putative knowledge about the organism's biological mechanisms will be explicitly internalized in the computational modules. The causal chains linking the biological processes are modeled by the structural associations between the computational modules. These associations represent the channels of interactions between the biological mechanisms within the hierarchical model architecture articulated above. The behavioral (physiologic/pathologic) scenarios of the model collectively provide a simulated instance of the dynamic behaviour of the living organism. In essence, the simulated behavioral scenarios represent the sum model integration of known or hypothesized regularities assumed to yield the observable or hypothesized physiological and pathological behaviours of the living organism.

The bewildering complexity of living organisms is associated with an infinite

set $\mathcal{S}$ of possible physiologic and pathologic behavioral scenarios. However, the modeling of a living organism is driven by a finite subset $\mathcal{S}_0 \subset \mathcal{S}$ of observable behavioral scenarios. Let $g : \mathcal{M} \to \mathcal{S}$ be the mapping from the set $\mathcal{M}$ of possible computational models of a living organism to the set $\mathcal{S}$ of behavioral scenarios. Developing a computational model $m_o \in \mathcal{M}$ that exhibits the set $\mathcal{S}_0$ of behavioral scenarios is essentially a process of estimating the inverse mapping $g^{-1} : \mathcal{S} \to \mathcal{M}$. It is clear that there are more than one model m_o that satisfies $g(m_0) = \mathcal{S}$. This is despite the constraints imposed by the hierarchical model architecture $\mathcal{A}$ articulated earlier, and the structural knowledge and known biological mechanisms embedded in the computational modules. In practical terms, the non-uniqueness of m_o means that a large space of modeling solutions may need to be explored before converging to a model that is satisfactory in some defined sense. Furthermore, given that $\mathcal{S}_0$ is a finite subset of $\mathcal{S}$, the prediction power of $m_o \in \mathcal{M}$ will be influenced by the limitation of approximating the inverse mapping using a subset of the organism's behavioral scenarios. Moreover, irrespective of the quality of the approximation, the prediction power of $m_o \in \mathcal{M}$ would still be limited in relation to the classical problem of generalization of discovery from a finite data set. Overall, it is worth noting that irrespective of the modeling approach being used, true-to-knowledge computational models would still not recover the entire complexity of a living organism and as a result would have intrinsic limitations with respect to their prediction power. However, it is reasonable to assume that the development of such models is the most logical way forward in dealing with the challenges of 21st century biological research. The conception of quantitative models of these limitations would be instrumental in developing more effective models of biological processes and living organisms.

The set $\mathcal{S}_0$ of behavioral scenarios cover all data collected through observation or experimentation about the physiology and pathology of a living organism. Such data also include experimental results or observation data about similar or comparable biological, physiological and pathological processes of other living organisms. Most knowledge about living organisms has been acquired through analytical methods of induction, deduction and inference applied to experimental data and observations. However, significant portion of such knowledge which concerns the structure and internal biological mechanisms has been sufficiently validated that it would be reasonable to take as a starting point for any modeling project. In this respect, structural information such as the organism's genome and proteome as well as known metabolic reactions and genetic mechanisms of transcription and translation would constitute the initial structural skeleton of the model. In particular, gene expression and metabolic processes would be internalized as the algorithms of the computational modules acting on the genes and protein entities. Information about the metabolic and gene regulatory networks will translate into structural relationships between the computational modules serving as channels for information flow.

Given the chosen model architecture $\mathcal{A}$, and the starting model fabric provided

by the known structural and mechanistic information about the living organism, the modeling project consists in the development of a computational model that exhibits the subset $\mathcal{S}_0$ of the organism's behavioral scenarios. This is clearly an engineering design problem aimed at the full specification of the algorithms of the computational modules and the protocols of information flows that control their interactions. The performance or effectiveness of the resulting model may be defined as $E = \sum_T \sum_v \parallel p_v^T - q_v^T \parallel$, where the inner sum is over all $v \in V$, and the outer sum is over all possible finite intervals of time T. q_v^T is the model's generated profile of the living organism for the variable v during the time interval T, and $\parallel . \parallel$ is a suitable norm. Although the defined model effectiveness E can only be approximated, it provides a theoretical criterion to guide the development of modeling strategies that would enable the development of the sought after computational model solution m_0. As stated earlier, m_0 is not unique implying as a result the need to explore a potentially large solution space through the use of different modeling strategies, techniques and algorithms. The convergence of such exploration may not be possible without the reliance on additional system-wide and biologically plausible constraints to specify the algorithms of the computational modules and the inter-module protocols of information flow. This question brings us full circle back to the assumption that the fundamental organizing force of living organisms emerges from the symbiotic relationship between the information and energy flows necessary for the transformation of matter in accordance to the DNA stored programs with their genetic and epigenetic dimensions. This relationship is hypothesized to be at the source of system-wide constraints that translate into observable system properties such as adaptation and robustness. The pertinent question at this point is how to reflect the need for the desired properties of adaptation and robustness in the detail design of the algorithms, parameters, and structure of the computational modules and their protocols of information flows. Control theory offers a wealth of adaptive and robust control design strategies that are readily applicable to the modeling of biological dynamics with quantitative assurance models of stability, robustness, and sensitivity. The exploration of specific control theoretic strategies for the development of models of biological processes is beyond the scope of this work. It may suffice to note here that the literature on biological modeling already includes a significant number of works dedicated to the application of control theoretic approaches and concepts to the development of models of biological processes [21,49,50,51,52,53,54]. These works include a theory of approximation for stochastic biochemical processes which can be instrumental in the development of the algorithmic details of the computational modules [55]. The use of Wasserstein pseudo-metrics as local objective functions for the estimation of model parameters within a computational module may be coupled to an approximation of the model effectiveness metric defined earlier. System-theoretic approaches to the biological network reconstructions constitute another example of research results which provide control-theoretic tools and strategies to discover system structure [56]. These results include analyses about the fundamental limitations of structure discovery from input-output data.

The suggested approach of structural perturbation experiments and the associated design are particularly relevant to the practical realization of the proposed modeling philosophy given the framing of this last as a reverse engineering process taking as input the observed behavioral scenarios of the living organism.

4. Computational Modeling of Living Organisms

The previous section outlined a proposed philosophy for the modeling of biological complexity. The ideas and concepts put forth will be further developed in future works for specific biological processes and living organisms. Meanwhile, the human organism is taken as an example to broadly illustrate the possible application of the discussed modeling perspective.

Modeling involves incremental steps where the scope, level of details and accuracy of the model are evolved as more knowledge is acquired about the living organism. For the human organism one may start with a graphical rendering of a very coarse-grain representation of some key biological regularities (see Fig. 3). The identified regularities and the abstractions used in the elaboration of their graphical rendering are only intended to illustrate the modeling approach rather than provide a precise and accurate representation of the relevant biological knowledge about cellular microbiology, human physiology and pathology, and genetics. For a detail account of such knowledge, the reader is referred to the elaborate and excellent treatments provided in the relevant texts [57,58,59,60,61].

The graphical illustration of Fig. 3 provides a sketchy knowledge model about a selection of major biological regularities of the human organism. The application of the proposed modeling philosophy requires a multi-steps process that starts with the allocation of biological regularities to their appropriate layers of the model's architecture. The modeling abstractions most appropriate for knowledge representation of the identified biological regularities are subsequently chosen to establish a resolution hierarchy across the layers of the model. Such resolution hierarchy emphasizes the representation of biological regularities with varying degrees of granularities. Fine grain mechanistic models are necessarily needed for the layers that are closest to the biochemical processes of metabolism, DNA/RNA information processing, energy production and matter transformation. Coarse grain models may provide on the other hand more appropriate abstractions for the upper layers of the model where physiological and pathological behaviour are described using macro-states such as *healthy, sick,* and *stable*.

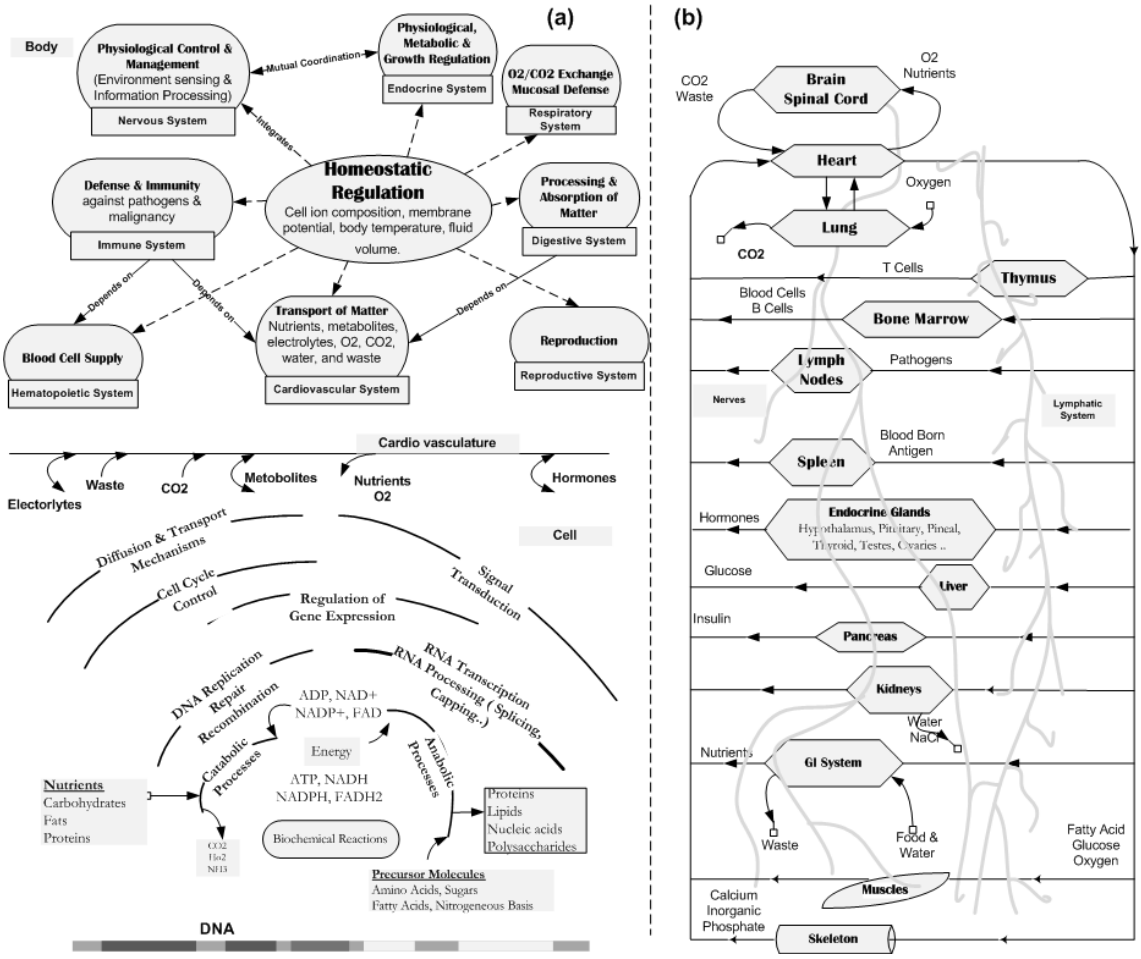

Fig. 3. Biological Regularities of the Human Organism. (a) key regularities at the cellular and multi-cellular levels respectively. The concentric arcs represent the direct dependence between the outer and inner layers. The biochemical processes of metabolism, DNA/RNA information processing, energy production and matter transformation are placed schematically close to the DNA as the primary information store. The concentric visual arrangement of the biochemical processes is also intended to reflect a logical axis of information flow and molecule transport stretched from the DNA to the lipid bilayer of the cell membrane and beyond. Biological processes are ultimately integrated at the organism level by the flow of information and matter /energy transported through the cardio-vasculature, the nerves, and the lymphatic system. The physiological systems are functional regularities with a system-wide coordination scope over biological processes operating at the cell, tissue and organ levels. Reciprocally, the cellular and inter-cellular processes provide the life line for these system-wide functional regularities. Nomenclature: ADP-adenosine diphosphate, NAD+ - nicotinamide adenine dinucleotide, NADP- nicotinamide adenine dinucleotide phosphate, FAD - flavin adenine dinucleotide, NH3 - ammonia. (b) body-wide structural regularities provide the channels for the flow of information, energy and matter. They are also involved in the processing of information and the transformation of matter/energy as illustrated with the molecular species exchanged at their interfaces. For example pathogens are shown flowing inward of the lymph nodes where they are confronted by the immune system, whereas Glucose is flowing outwards of the liver which regulates blood sugar level.

The result of the first two steps of the modeling process is a model skeleton with distinct layers and a structure binding the computational modules dedicated to the realization of the models of identified biological regularities (see Fig. 4). In line with the assumption that a living organism is an integrated collection of biological regularities, the modeling strategy treats physiologic/pathologic behavioral scenarios as a net manifestation of integrated systems of biological regularities rather than behavioral regularities emerging from the hierarchical compositions of constituents. For example, the dynamic regulation of fluid volume is not necessarily possible by mere assemblage of the kidneys, the cardio vasculature and the Gastrointestinal (GI) and urinary tracks. While these organs are necessary for the realization of the regulation function, however, such function is not possible without the integration of other physiological functions, including circulation, filtration and the excretion of water and electrolytes. Indeed, the flow of matter (blood, metabolites, electrolytes and water), the consumption of energy, and the flow of information (endocrinal signaling, nervous system signaling) through the spatial contexts of the regularities in question lead to their integration and the subsequent realization of the fluid volume control function. The implication of this assumption is that each model layer and associated computational modules must capture the complexity of the relevant biological regularities independently of the structure and processes of the lower layers of the model. However, the workings of the model layers are integrated by the information flows which lead as a result to the coupling between the different abstractions and knowledge representations of the computational modules. Such coupling does however require information processing such as quantization and filtering to bridge the different types of data being generated or consumed by the computational modules. For example, the Cell Cycle Control (C3) module responsible for orchestrating the transition between the phases of the cell cycle may be realized using a Statechart-based reactive model [29,44]. Biologically plausible states of the cell cycle may be defined based on the instantaneous concentrations of the relevant proteins being within specific value ranges. Such coarse-grain representation of protein concentration, supplied by the lower layer of the model filters out the complexity associated with the consideration of specific values of protein concentration. Fine grain knowledge details may not need to be accounted for at the level of cell cycle control where the focus is on the overall dynamical behaviour of the cell cycle. The nature of information processing applied as part of the resolution hierarchy of knowledge representation depends on the nature of the specific processes being modeled. Indeed, to model the mechanisms of diffusion and transport across the cell membrane, quantization of the temporal and spatial distribution of transported ions may be more appropriate. Such quantized ion distributions could then be used to define model states of Intracellular fluid / extracellular fluid (ICF/ECF) volume and composition balance.

The Signal Transduction (STR) module is essentially driven by external stimuli especially for the case of endocrine signaling. The challenge here is to abstract away the mechanistic implementation of the signaling pathways, such as the JAK-STAT

(Janus Kinase /Signal Transducer and Activator of Transcription), while safeguarding the information content and the temporal dynamics of signal reception, relay and decoding. One possible approach to this issue is to define a list of biologically plausible basis signals for all potential outputs of the STR module and leverage communication system theory to develop a signal processing model of the cell signal transduction process. Nevertheless, the complexity of the modeling task would still be a difficult challenge given the large number of combinations of protein interaction domains being used in the intracellular signaling cascades [62]. An example basis signal could be a discrete-time binary sequence generated by the STR module to overcome the G0 blockade of the cell cycle, in response to the external stimuli of hormones and growth factors. Such signal would trigger the DNA eXpression Regulator (DXR) and Metabolic Control (MC) to initiate the timely degradation, inhibition, activation or synthesis of protein products such as P27, and Cyclins C and D implicated in keeping the cell in G0 and also known to be responsible for its G1 re-entry respectively. The DXR and MC modules may be realized using a variety of mathematical and control theoretic models, some of which have been recently reported in the literature [63,64]. These models could be readily integrated into larger discontinuous control structures using bang-bang and sliding mode control [65,66]. Such variable structure control models would then be driven by the discrete-time binary signal output of the STR module. Such binary signal may on the other hand require filtering or mathematical integration over time before being feed to the cell cycle control and the Cell State Dynamics (CSD) module. The computational realization of the C3 and CSD modules may be achieved using Statechart based reactive models, where state transitions are guarded by conditions dependant on threshold crossings by the signal inputs to the modules. The cell states of the CSD module may be defined based on a potentially long list of conditions and factors that involve among other contributors, the concentrations of cellular proteins and molecular species, cell morphology as deduced from the presence of some biological markers, the state of the DNA chromatin, and the state of the cell receptors. The resulting model states may include *normal, infected, atypical,* and *abnormal*. It should be noted that this list is only an indication of the possible states that must be defined to capture a biologically and clinically plausible model characterization of the physiological/pathological states of a cell.

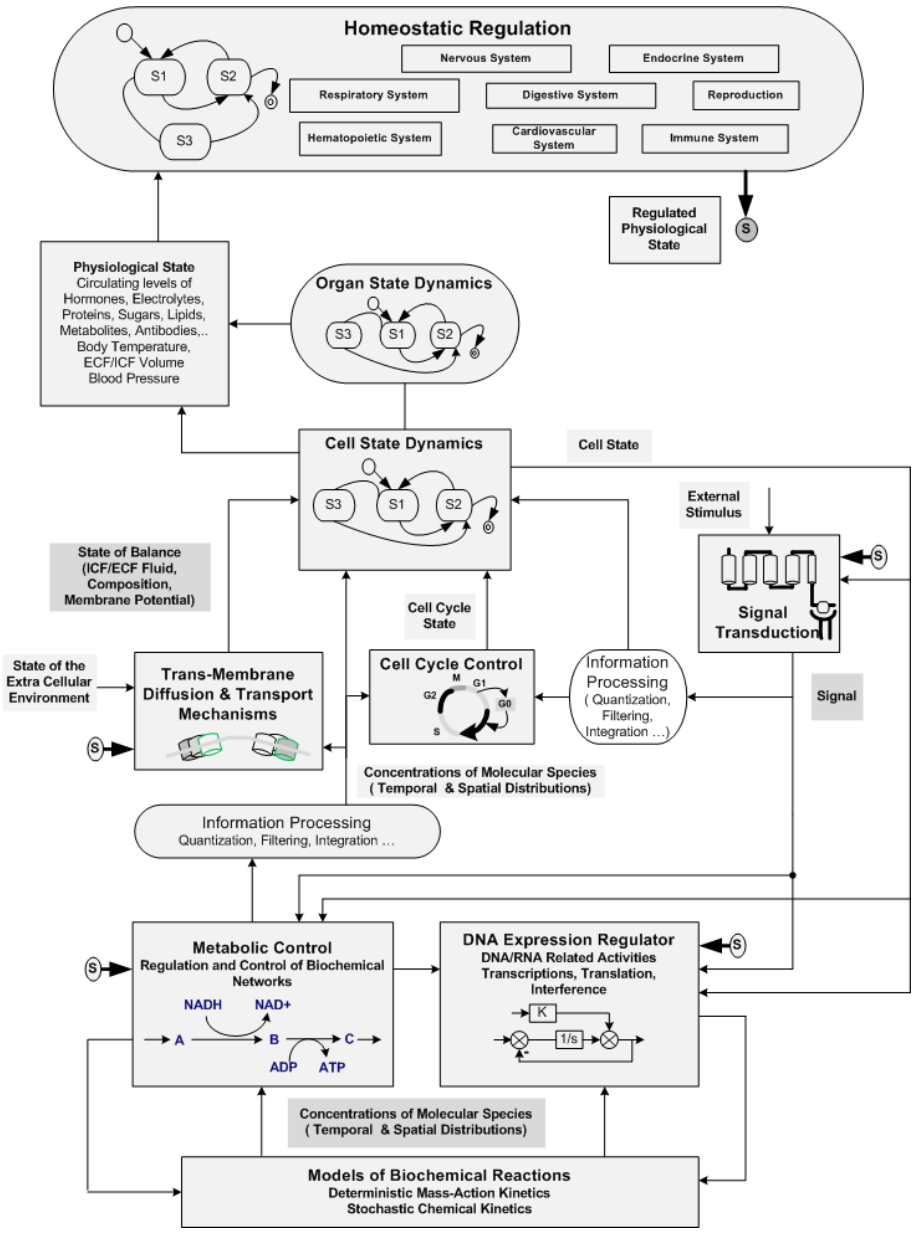

Fig. 4. Multi-Scale Computational Model of the Human Organism. The biological processes of a human organism are categorized according to their scope (cell, tissue/organ, and body) and their involvement in: (1) matter transformation and transport, (2) information processing and flow, and (3) control, regulation and management. The model is conceived as a collection of computational modules each dedicated to the modeling of biological processes that belong to the same realm of functional category (metabolism, signal transduction, gene expression regulation, etc.). For example the metabolic control module represents the algorithmic model of regulation and control of biochemical networks. The modules are loosely coupled and equipped with well defined information interfaces made up of observable or experimentally measurable variables such as the concentrations of molecular species and the cell states. The model integration is enabled by the information flowing between the modules including signals, protein concentration values, cell state, physiological state, etc. The modules interact in a context defined by the hierarchical coarse-grain knowledge abstraction which gradually links biochemical reactions and genetic information processing populating the inner layers of the model all the way to the pathologic and physiologic behaviour exhibited through the higher layers of the model.

The convergence to a coarse grain model representation of cell state dynamics from the basic models of mass-action kinetics would, if realizable as hypothesized so far, provide an effective framework for the integration of multi-scale models of biological processes. Indeed, such pattern of multi-scale integration can be replicated by linking the states of the cells to the states of their home organs. Hierarchical Statechart models may prove to be an adequate approach to the realization of such integration [29]. This model integration would require the definition of the possible physiological and pathological states of organs (*normal, inflamed, infected*, etc.), and a model of how the dynamics of cell ensembles lead to specific organ states. The organs' states and the information about what may be called the body's physiological signature (the levels of circulating hormones, body temperature, blood pressure, sugar blood level, level of electrolytes, etc.) may provide the input to a homeostatic regulation model representing the coordinated interaction of the physiological systems (nervous, cardiovascular, endocrine, hematopoietic, digestive, reproductive, immune, and respiratory) to maintain the healthy state of the organism. The homeostatic regulation may also be realized using a Statechart reactive model with outputs acting as feedback control inputs to the modules of the cell mechanisms as shown in Fig. 4. Deciphering and modeling the interactions between the various physiological systems and their coordinated actions on DNA expression and metabolic networks is another item in the long list of biological research challenges. Collaborative research efforts such as the Physiome project will inevitably be instrumental in leading to scientific advances on this front [67]. The quantitative models resulting from the Physiome projects [68] for various physiological systems and diseases would directly contribute to the structural fabric and mechanistic processes targeted for realization by the modules of the proposed computational model. One potential application example of the proposed framework is the study of the antagonistic dynamics between cell proliferation and apoptosis in cancer, and the exploration of drug-mediated control of their ultimate outcome. For a highly simplified apoptosis/proliferation model (drawn principally from Green [69]), cell proliferation is induced by Myc, a transcription factor which can also trigger apoptosis unless blocked by factors, such as interleukin-2, epidermal growth factor, and insulin, supplied by neighboring cells. MCL-1 (Myeloid Cell Leukemia sequence 1) is one such anti-apoptotic protein whose degradation by MULE-E3-ubiquitin is prevented by TCTP (Translationally Controlled Tumor Protein). On the other hand TCTP is repressed by p53, itself locked with MDM-2 (Mouse Double Minute - 2) in a feedback loop of expression and inhibition that prevents any significant accumulation of neither in normal cells. Furthermore, the tumor suppressor ARF (inducible by oncogenes) exert another layer of control on apoptosis through MULE-E3 that targets p53 and MCL-1 for degradation, seemingly inhibiting apoptosis on one hand and promoting apoptosis on the other! ARF can disrupt the p53-MDM-2 interaction of p53 leading to the accumulation of this later and the subsequent repression of TCTP, allowing hence the degradation of MCL-1 and the promotion of apoptosis. The complexity of interactions involved in the control of apoptosis

is clearly prohibitive, suggesting hence the existence of an intrinsic difficulty in making conclusions about potential targets to either promote apoptosis or inhibit anti-apoptotic pathways in cancerous cells. However, the consideration of the net effect of these interactions on the balance between apoptosis and proliferation may make the problem more tractable [69]. The proposed framework may provide the context to enact this insight. The lower levels of the model would be responsible for the modeling and simulation of the relevant biochemical reactions, gene regulation and metabolic control using state of the art fine-grain mathematical models, many of which were mentioned earlier. The information carried by the quantized protein concentrations is interpreted within the context of cell state dynamics. The cell state space can be defined in relation to the relative abundance of the involved tumor suppressor proteins such as p53 and ARF and other determinant proteins such as the anti-apoptotic protein MCL-1 and the ligase MULE E3. Such definition needs to be informative in relation to the ultimate outcome (apoptosis or proliferation). In addition, the defined states must have a reasonable likelihood of being discernable through experimentation (using imaging technology, markers, etc). One approach that may cover all of these criteria is to inspect the protein interaction network described above and define a state for each pathway node that could be implicated in tipping the balance between apoptosis and proliferation. For example, the competitive action of MULE and MDM2 on p53 would inspire the definition of a state where a threshold associated with the collective abundance of these proteins has been reached taking the cell to this state and maintaining it there until the time at which the contest is resolved. Such resolution could be the abundance of p53 beyond a certain threshold taking the cell to another state where TCTP compete with MULE for the stability of MCL-1 and so on. The states of the individual cells inform the organ/tissue state dynamics which would be focused on a tissue/organ centric interpretation of the physiology/pathology. Here again experimental as well as clinical data and observations would be the arbiter as to what might be the most pertinent and discernible states of the tissue vis-a-vis a progression towards cancer. The algorithmic core of the organ/tissue state transition could simply be based on the distribution of state occupancy by the individual cells. Other strategies may be conceived to define the behaviour of an ensemble in relation to the state of individual cells. However, analytical accessibility and biological plausibility of the chosen approach should be given prime consideration. The sketchy details raised in this example are only meant to provide an illustrative highlight of the potential utility of the proposed framework. Some broad ideas are discussed as to how such utility might begin to be enacted in tackling complex diseases such as cancer. In future works, the underlying strategies and concepts of the proposed framework will be unpacked towards full application. Experimental data is expected to provide the necessary support to shape the framing of the specific questions such as the dynamics of apoptosis/proliferation in cancer.

5. Conclusions and Future Works

Living organisms are viewed as collections of integrated biological regularities that may be modeled along a resolution hierarchy of knowledge representation to reflect the nature of their observed complexity. The constraints resulting from the flow of information and energy involved in the transformation of matter necessary for development, reproduction and growth are assumed to make up the organizing force of living organisms. This assumption inspired the hierarchical model architecture where metabolic reactions and DNA/RNA information processing mechanisms are assigned to the lower layers of the model acting as the engine for the information and energy flow. Hierarchical state feedback and the progressive coarse grain representation of information flowing to the upper layers of the model link the internal model dynamics to its output, which is intended to mirror the living organism's observed physiological and pathological behavior. The layered model encapsulation of biological complexity along with the enforcement of a resolution hierarchy of knowledge representation may lead to a computational realization that is more accessible to analysis of the causes underlying the emergence of pathological patterns of dynamics and the information pathways critical to robust maintenance of physiological behaviour under environmental perturbations. Computational model realizations and the explicit consideration of energy are planned for future works. In particular, an explicit modeling of the links between gene regulation and the availability of energy may be explored using control-theoretic approaches. Such exploration may benefit from control system design methods especially those focused on the synthesis of control laws under constraints on the amount of energy available for actuation. In particular, hybrid and multi-scale computational model realizations (combining strategies such as differential calculus-based methods, state machines, and Markov chains) will be explored to capture the different modeling resolutions most adequate for the layers and modules in questions. Attention will be given to leveraging the modularity emphasized in the framework by opting for algorithmic realizations targeting a multi-threaded multi-core computing environment and strengthening as a result the capacity of the model realization to absorb new research discoveries. A particular consideration will be given to the application of control-theoretic concepts such as robustness, stability, and adaptation to gauge the model's biological plausibility and provide insight about the relationship between these properties and pathogenesis. One possible starting point in this direction would be the definition and subsequent analysis of sensitivity, robustness and adaptation measures of the time-varying dynamics of the physiological state. In this respect, Omics profiling combined with physiological state monitoring could be instrumental in model driven reconstruction of the genotype-phenotype chain as recently reported [70]. In particular the dataset resulting from Omics profiling may be mined to construct physiologic and pathologic behavioral scenarios that would serve as a system-wide integrated reference behaviour, which may consequently be used to establish the biological plausibility of model realizations of the proposed

framework.

6. References

[1] John C. Wooley and Herbert S. Lin. *Catalyzing Inquiry at the Interface of Computing and Biology*. The National Academies (US), 2005.

[2] D. Thieffry and R. Thomas. Qualitative analysis of gene networks. In *Pac Symp Biocomput*, pages 77–88, 1998.

[3] R. Mourad, C. Sinoquet, and P. Leray. Probabilistic graphical models for genetic association studies. *Brief Bioinform*, 13(1):20–33, 2012.

[4] J. F. Yu and X. Sun. Reannotation of protein-coding genes based on an improved graphical representation of dna sequence. *J Comput Chem*, 31(11):2126–35, 2010.

[5] Y. Moreau, P. Antal, G. Fannes, and B. De Moor. Probabilistic graphical models for computational biomedicine. *Methods Inf Med*, 42(2):161–8, 2003.

[6] N. Gehlenborg, S. I. O'Donoghue, N. S. Baliga, A. Goesmann, M. A. Hibbs, H. Kitano, O. Kohlbacher, H. Neuweger, R. Schneider, D. Tenenbaum, and A. C. Gavin. Visualization of omics data for systems biology. *Nat Methods*, 7(3 Suppl):S56–68, 2010.

[7] N. Friedman, M. Linial, I. Nachman, and D. Pe'er. Using bayesian networks to analyze expression data. *J Comput Biol*, 7(3-4):601–20, 2000.

[8] F. Reverter, E. Vegas, and P. Sanchez. Mining gene expression profiles: an integrated implementation of kernel principal component analysis and singular value decomposition. *Genomics Proteomics Bioinformatics*, 8(3):200–10, 2010.

[9] T. Hancock and H. Mamitsuka. A markov classification model for metabolic pathways. *Algorithms Mol Biol*, 5:10, 2010.

[10] A. Senf and X. W. Chen. Identification of genes involved in the same pathways using a hidden markov model-based approach. *Bioinformatics*, 25(22):2945–54, 2009.

[11] S. A. Kauffman and S. Johnsen. Coevolution to the edge of chaos: coupled fitness landscapes, poised states, and coevolutionary avalanches. *J Theor Biol*, 149(4):467–505, 1991.

[12] R. Silva-Rocha and V. de Lorenzo. Implementing an or-not (orn) logic gate with components of the sos regulatory network of escherichia coli. *Mol Biosyst*, 7(8):2389–96, 2011.

[13] C. Lou, X. Liu, M. Ni, Y. Huang, Q. Huang, L. Huang, L. Jiang, D. Lu, M. Wang, C. Liu, D. Chen, C. Chen, X. Chen, L. Yang, H. Ma, J. Chen, and Q. Ouyang. Synthesizing a novel genetic sequential logic circuit: a push-on push-off switch. *Mol Syst Biol*, 6:350, 2010.

[14] G. Karlebach and R. Shamir. Modelling and analysis of gene regulatory networks. *Nat Rev Mol Cell Biol*, 9(10):770–80, 2008.

[15] J. Saez-Rodriguez, L. G. Alexopoulos, J. Epperlein, R. Samaga, D. A. Lauffenburger, S. Klamt, and P. K. Sorger. Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. *Mol Syst Biol*, 5:331, 2009.

[16] S. Raza, K. A. Robertson, P. A. Lacaze, D. Page, A. J. Enright, P. Ghazal, and T. C. Freeman. A logic-based diagram of signalling pathways central to macrophage activation. *BMC Syst Biol*, 2:36, 2008.

[17] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon. Network motifs in the transcriptional regulation network of escherichia coli. *Nature Genetics*, 31(1):64–8, 2002.

[18] M. A. Kenyon and B. Palsson. *Genome-Scale Reconstruction, Modeling, and Simulation*, chapter 12, pages 237–255. Springer, 2009.

[19] D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81(25):2340–2361, 1977.
[20] I. W. Jolma, X. Y. Ni, L. Rensing, and P. Ruoff. Harmonic oscillations in homeostatic controllers: Dynamics of the p53 regulatory system. *Biophys J*, 98(5):743–52, 2010.
[21] T. M. Yi, Y. Huang, M. I. Simon, and J. Doyle. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc Natl Acad Sci U S A*, 97(9):4649–53, 2000.
[22] A. P Iglesias. Spatial modeling. In a. p Iglesias and B. ingalls, editors, *Control Theory and System Biology*, chapter 3, pages 45–68. MIT Press, 2010.
[23] S. Frey, O. Wolkenhauer, and T. Millat. *Quantifying Properties of Cell Signaling Cascades*, chapter 4, pages 69–84. 2010.
[24] C. Trane and E. W. Jacobsen. *Structural Robustness of Biochemical Networks: Quantifying Robustness and Identifying Fragilities*, chapter 10, pages 189–224. 2010.
[25] Drew Endy and Roger Brent. Modelling cellular behaviour. *Nature*, 409(6818):391–395, 2001.
[26] P. Nurse. Life, logic and information. *Nature*, 454(7203):424–6, 2008.
[27] H. Kitano. Systems biology: a brief overview. *Science*, 295(5560):1662–4, 2002.
[28] T. Ideker, T. Galitski, and L. Hood. A new approach to decoding life: systems biology. *Annu Rev Genomics Hum Genet*, 2:343–72, 2001.
[29] Sol Efroni, David Harel, and Irun R. Cohen. Toward rigorous comprehension of biological complexity: modeling, execution, and visualization of thymic t-cell maturation. *Genome research*, 13(11):2485–2497, 2003.
[30] Y. Lazebnik. Can a biologist fix a radio? – or, what i learned while studying apoptosis, (cancer cell. 2002 sep;2(3):179-82). *Biochemistry (Mosc)*, 69(12):1403–6, 2004.
[31] S. A. Kauffman. Antichaos and adaptation. *Sci Am*, 265(2):78–84, 1991.
[32] U. Alon. Biological networks: The tinkerer as an engineer. *Science*, 301(5641):1866–1867, 2003.
[33] P. W. Anderson. More is different. *Science*, 177(4047):393–6, 1972.
[34] Alexei Kurakin. The self-organizing fractal theory as a universal discovery method: the phenomenon of life. *Theoretical Biology and Medical Modelling*, 8(1):4, 2011.
[35] P. Nurse and J. Hayles. The cell in an era of systems biology. *Cell*, 144(6):850–4, 2011.
[36] Arthur D. Lander. A calculus of purpose. *PLoS biology*, 2(6):e164–e164, 2004.
[37] Aviv Regev and Ehud Shapiro. Cellular abstractions: Cells as computation. *Nature*, 419(6905):343–343, 2002.
[38] L. A. Segel. Computing an organism. *Proceedings of the National Academy of Sciences of the United States of America*, 98(7):3639–3640, 2001.
[39] Douglas W. Selinger, Matthew A. Wright, and George M. Church. On the complete determination of biological systems. *Trends in Biotechnology*, 21(6):251–254, 2003.
[40] O. Wolkenhauer and M. Mesarovic. Feedback dynamics and cell function: Why systems biology is called systems biology. *Mol Biosyst*, 1(1):14–6, 2005.
[41] O. Wolkenhauer, C. Auffray, S. Baltrusch, N. Bluthgen, H. Byrne, M. Cascante, A. Ciliberto, T. Dale, D. Drasdo, D. Fell, Jr. Ferrell, J. E., D. Gallahan, R. Gatenby, U. Gunther, B. D. Harms, H. Herzel, C. Junghanss, M. Kunz, I. van Leeuwen, P. Lenormand, F. Levi, M. Linnebacher, J. Lowengrub, P. K. Maini, A. Malik, K. Rateitschak, O. Sansom, R. Schafer, K. Schurrle, C. Sers, S. Schnell, D. Shibata, J. Tyson, J. Vera, M. White, B. Zhivotovsky, and R. Jaster. Systems biologists seek fuller integration of systems biology approaches in new cancer research programs. *Cancer Res*, 70(1):12–3, 2010.
[42] Murray Gell-Mann and Seth Lloyd. *Effective Complexity*. Oxford University Press,

2003.

[43] Murray Gell-Mann and Seth Lloyd. Information measures, effective complexity, and total information. *Complexity*, 2(1):44–52, 1996.

[44] S. Efroni, D. Harel, and I. R. Cohen. Reactive animation: realistic modeling of complex dynamic systems. *Computer*, 38(1):38–47, 2005.

[45] George N. Saridis and Kimon P. Valavanis. Analytical design of intelligent machines. *Automatica*, 24(2):123–133, 1988.

[46] G. N. Saridis. Toward the realization of intelligent controls. *Proceedings of the IEEE*, 67(8):1115–1133, 1979.

[47] George N. Saridis. Analytic formulation of the principle of increasing precision with decreasing intelligence for intelligent machines. *Automatica*, 25(3):461–467, 1989.

[48] J. S. Albus. Rcs: a reference model architecture for intelligent control. *Computer*, 25(5):56–59, 1992.

[49] A. P Iglesias and B. P Ingalls. *Control Theory and Systems Biology*. 2010.

[50] M. E. Csete and J. C. Doyle. Reverse engineering of biological complexity. *Science*, 295(5560):1664–9, 2002.

[51] J. M. Carlson and J. Doyle. Complexity and robustness. *Proc Natl Acad Sci U S A*, 99 Suppl 1:2538–45, 2002.

[52] S. Bhartiya, N. Chaudhary, K. V. Venkatesh, and 3rd Doyle, F. J. Multiple feedback loop design in the tryptophan regulatory network of escherichia coli suggests a paradigm for robust regulation of processes in series. *J R Soc Interface*, 3(8):383–91, 2006.

[53] H. Kitano. Towards a theory of biological robustness. *Mol Syst Biol*, 3:137, 2007.

[54] Y. Derbal. Control theoretic modeling of a genetic switch. In *24th Canadian Conference on Electrical and Computer Engineering*, pages 000603–000606. IEEE.

[55] D. Thorsley and E. Klavins. Approximating stochastic biochemical processes with wasserstein pseudometrics. *IET Syst Biol*, 4(3):193–211, 2010.

[56] J. Goncalves and S. Warnick. System-theoretic approaches to network reconstruction. In a. p Iglesias and B. ingalls, editors, *Control Theory and System Biology*, chapter 3, pages 265–295. 2010.

[57] Bruce Alberts, John Wilson, and Tim Hunt. *Molecular biology of the cell*. Garland Science, New York, 5th edition, 2008.

[58] Bruce M. Koeppen and Bruce A. Stanton. *Berne and Levy Physiology, Updated Edition, 6th Edition*. Mosby, 2010.

[59] Vinay Kumar, Abul K. Abbas, Nelson Fausto, and Richard Mitchell. *Robins Basic Pathology*. Saunders, 2007.

[60] Michael Cox and David L. Nelson. *Lehninger Principles of Biochemistry*. W. H. Freeman, 2009.

[61] James D. Watson. *Molecular biology of the gene*. Pearson/Benjamin Cummings ; Cold Spring Harbor Laboratory Press, San Francisco, Cold Spring Harbor, N.Y., 6th edition, 2008.

[62] Friedrich Marks, Ursula Klingmlle, and Karin Mller-Decker. *Cellular signal processing: an introduction to the molecular mechanisms of signal transduction*. Garland Science, 2009.

[63] B. P Ingalls. A control theoretic interpretation of metabolic control analysis. In a. p Iglesias and B. ingalls, editors, *Control Theory and System Biology*, chapter 8, pages 145–168. 2010.

[64] H. El-Samad. Control strategies in times of adversity : How organisms survive stressful conditions. In A. P Iglesias and B. Ingalls, editors, *Control Theory and System Biology*, chapter 5, pages 85–99. 2010.

[65] Donald R. Snow. Functional analysis and time optimal control (henry hermes and joseph p. lasalle). *SIAM Review*, 18(2):307, 1976.
[66] U. Itkis. *Control Systems of Variable Structures*. 1976.
[67] J. B. Bassingthwaighte and H. J. Chizeck. The physiome projects and multiscale modeling. *IEEE Signal Process Mag*, 25(2):121–144, 2008.
[68] A. S. Popel and P. J. Hunter. Systems biology and physiome projects. *Wiley Interdiscip Rev Syst Biol Med*, 1(2):153–8, 2009.
[69] D. R. Green. *Means to an end*. CSH, 2011.
[70] M. Snyder. Adventures in personal genomics and whole omics profiling. In *The 13th International Conference on Systems Biology*, page 76, 2012.