Systems Biology

- History and Definitions
- Reductionism/Genetic Determination
- Holism/Emergentism/Homeostasis or Robustness
- Revolutionary and Evolutionary Systems Biology
- Networks and Computational Biology
- Basic Molecular and Cellular Components

Required Reading


Background Book References


Literature


From Systems Biology to Systems Biomedicine
Paul MA Antony, Rudi Balling and Nikos Vlassis

Systems Biology is about combining theory, technology, and targeted experiments in a way that drives not only data accumulation but knowledge as well. The challenge in Systems Biomedicine is to furthermore translate mechanistic insights in biological systems to clinical application, with the central aim of improving patients' quality of life. The challenge is to find theoretically well-chosen models for the contextually correct and intelligible representation of multi-scale biological systems. In this review, we discuss the current state of Systems Biology, highlight the emergence of Systems Biomedicine, and highlight some of the topics and views that we think are important for the efficient application of Systems Theory in Biomedicine.

Address
Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Luxembourg

Corresponding author: Antony, Paul MA (paul.antony@uni.lu)

Current Opinion in Biotechnology 2012, 23:604–608
This review comes from a themed issue on Systems biology
Edited by Jens Nielsen and Sang Yup Lee
For a complete overview see the Issue and the Editorial
Available online 24th November 2011
0958-1669/$ – see front matter, © 2011 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.copbio.2011.11.009

Introduction
Complex interactions in biological systems are an increasingly important component of translational research. Systems Biology is a unifying term that groups interdisciplinary scientific efforts that focus on the analysis of complex interactions in biological systems [1]. For the past decade, the term has been widely used in a variety of biomedical contexts. The amalgamation of two opposing concepts, namely, holism and reductionism is currently contributing to conceptual advances in Systems Biology [2,3]. The holistic approach is based on the idea that complex systems cannot be fully understood by studying isolated parts. One of the main ambitions of holistic Systems Biology is the understanding of emergent properties such as robustness [4]. The reductionists focus on modules within bigger systems and can be described as modern physiologists working on molecular explanations for biological events. There is increasing consent that it is counterproductive to engage in ongoing potentially divisive arguments comparing holism with reductionism, and it is becoming clear that both reductionists and holists, need to connect molecular and clinical scales. Systems Biomedicine can be considered as an emergent branch of Systems Biology that allows zooming through the multiple scales of life and disease by combining reductionist and holistic approaches. A widely applied research concept is a cycle composed of theoretical analysis, computational modeling, and experimental validation of model hypotheses, which drives the refinement of computational and experimental models.

Systems Biomedicine: Scope and bottlenecks
The ultimate goal in Systems Biomedicine is to apply mechanistic insights to clinical application and to improve patients’ quality of life. This aim, however, raises the need to solve some important questions on how to organise a framework supporting these interdisciplinary research efforts [5,6]. Recognition of a clinical, mechanistic, or theoretical problem is the first step. Formulating a testable hypothesis is the second. Such a hypothesis can be considered as a model that requires further validation in biological and clinical trials before clinical application. However, such interdisciplinary research can easily stagnate. To avoid stagnation, theoretical, biological, and clinical researchers need to share their current knowledge, aims, and visions for the future (Figure 1).

The human genome has now been completely sequenced, and many genetic variants and epigenetic modifications have been identified. Recent improvements in technology have facilitated the collection of unprecedented amounts of data at various biological and clinical levels. A comprehensive understanding of all this information is, however, lagging behind the accumulation of data, and we are only beginning to understand the extent to which phenotypic traits and their modes of emergence are regulated at multiple levels of biological complexity [7,8].

A classical approach for studying emergent properties from genetic or environmental factors is to perturb the factor of interest and to observe the resultant phenotypic changes. An important technology-driven development in this field is the shift from phenotypic endpoint studies to time-resolved phenotypic profiling [9]. It is, however, potentially dangerous and misleading to expect relationships of the type 'one gene – one protein – one function – one phenotype'. Frequently, phenotypic changes do not clearly reveal genetic perturbations due to extensive buffering driven by network robustness [10]. Thus, predictions of phenotypic changes need to consider both perturbation properties and network context. The effect
of certain drugs in such contexts is often not easily predicted. Such predictions would require both a sufficient understanding of molecular, cellular, and clinical scales, and a sufficient understanding of emergent properties that connect the different scales, for example, of heart function. In a pioneer study, Fink and Noble [11] demonstrated that mathematical models are beneficial for unravelling the complex interactions of pharmacodynamics in the heart. They predicted that embedding detailed cellular-scale models into anatomically correct organ-scale models would enable reconstructions of cardiac dysfunctions, providing a tool for connecting cellular-scale experimental data to clinical-scale applications.

In the context of Western medicine, molecular understanding of pathogenic events is a fundamental aim and driving force. In addition, modern Systems Biomedicine highlights the importance of a higher level of biological organisation. A much older discipline with a very similar interest in higher order biological organisation is Oriental medicine. Sasang constitutional medicine, the Korean traditional medicine, seeks patient-specific analysis and treatment. The current Western trend of recognizing the potential value of predictive, personalised, preventive, and participatory (P4) medicine [12] represents a timely convergence between Western and Oriental medical disciplines [13,14]. However, many challenges will need to be overcome before this convergence translates to actual clinical outcomes.

**Diseases and models**

Diseases can be roughly categorised into those that are due to environmental factors and those due to genetic factors. The latter can range from single causal factors to modifying risk factors and multifactorial diseases. Furthermore, the onset and progression of many diseases is influenced by a combination of both environmental and genetic factors [15,16]. In monogenic diseases, a mutation in a single gene causes the disease. Such diseases are relatively easy to trace by classical Mendelian inheritance patterns, and recent research has been directed towards exploiting such patterns in the context of whole family DNA sequencing [17]. On the contrary, multigenic or complex diseases are those in which progression is believed to be influenced by the collective action of several genes and environmental factors. Examples include epilepsy, Parkinson’s disease, and cancer. The multifactorial nature of such diseases renders making predictions relating to their onset and progression a very challenging task.

In an effort to facilitate more accurate predictions of disease onset and progression, researchers often utilise ‘models’ of the disease. The term ‘model’ is used in biology in different ways, but it often means an idealised representation of an empirical system [18]. Models can be viewed as artefacts that allow casting of a natural phenomenon in a formal language such as mathematics, physics, or computer science. Modelling, the process that combines the identification and collection of confident preliminary knowledge and the identification of potential mechanistic explanations, is a central tool in Systems Biology. As George Box said ‘all models are wrong, but some are useful’ [19].

To predict ‘simple’ diseases such as those with Mendelian inheritance patterns we probably do not need
complex models; often, we only need to determine the allelic makeup of a single gene, which usually amounts to a low-complexity computational test. The length of the polyglutamine stretch in the Huntingtin gene, for example, allows prediction of the approximate age of onset of Huntington’s disease. While additional disease-modifying factors exist, they exert only a minor influence on the age of onset [20].

On the contrary, simple models are probably inadequate for facilitating predictions relating to multifactorial complex diseases such as epilepsy. It has been shown that single-nucleotide polymorphism burden in ion channels holds little clinical predictive power for epilepsy [21], thus validating the complex multigenic inheritance of epileptic channelopathies and the principle of compensatory effects in complex systems. Individualised disease prediction appears to require an understanding of how genetic variants contribute to risk in each given individual [22,23]. In epilepsy, different ion channel variants might potentiate or suppress similar membrane currents and thereby influence neuronal firing patterns and pathological phenotypes. Furthermore, even physically non-interacting channel proteins could modulate distant channels via changes in the transmembrane potential. Appreciation of such mechanisms and modifying factors within a personalised approach would require complex models, but may also yield valuable information with far-ranging implications for future research aimed at translating genomic profiling into risk prediction [21].

The prototypical complex model in Systems Biology is a network, whose semantics may vary depending on the particular application and degree of approximation. A network model provides a view of a biological system as a graph dynamical system, which is defined by the coordinated action of several nodes and corresponding edges, and their local dynamics [25]. This allows for the study of the global properties of dynamical biological systems via corresponding concepts in the phase space. Commonly used concepts are stable equilibrium states, for example, health and disease [26,27], and limit cycles, for example, oscillations of the cell cycle [28]. The appropriate application of these modelling concepts to biological systems can strongly enhance predictive power [29,30]. Large-scale complex networks can be difficult to analyse, and recently, a line of research has emerged that focuses on the use of mathematical theorems and results from computer science in an effort to mitigate this complexity. In particular, a variety of theoretical and experimental results demonstrate a relationship between the structural properties of a network and properties of the phase space of the corresponding dynamical system [31–35].

In general, three main types of modelling strategies can be distinguished. Modelling can be top-down, from clinical signs to molecular processes; bottom-up, from molecular processes to clinical phenotypes; or middle-out, starting from an intermediate scale [36]. Bottom-up approaches are useful when achievements from Molecular Biology allow for the formulation of hypotheses on emergent properties. Middle-out approaches typically start from the basic unit of life, the cell [37]. Since living cells coordinate the flow of spatiotemporal information between molecular and organic scales, middle-out approaches to modelling seem to be especially apt for multiple-scale models; the components incorporated into these models may include the genetic code, transcription, splicing, RNA interference, translation, post-translational modification, protein complexes, organelles and compartments, metabolism, cellular function, cell to cell interactions, tissue function, organ function, system response to environmental factors, health status, and quality of life. Examples of such multi-scale models can be found in the Virtual Cell, Physiome Project, and Virtual Physiological Human initiative [37–39].

Complex models possess the freedom to capture subtle details of the system of interest; however, some models trade off precision for simplicity and represent qualitative features of the system. For example, a Boolean network can be used for modelling gene expression in a qualitative manner, by assuming that a gene can be either ‘on’ or ‘off’, a simplified assumption that seems to work well in practice [40]. While complex models would seem a better choice in terms of predictive power, such models may lead to ‘good’ results for the wrong reasons. A ‘good fit’ between a model and observed system behaviour does not guarantee any realism of parameter values or model structure [41,42].

It can be argued that biological processes may be inherently stochastic [43,44]. This calls for models that explicitly incorporate uncertainty. The prototypical paradigms here are probability theory and Bayesian analysis. The latter, in particular, allows for the incorporation of process stochasticity and model parameter uncertainty within the same framework. For parameter estimation, the modeller places a prior probability distribution over the space of model parameters, and then uses the existing sensorial evidence to infer a posterior distribution over these parameters, thereby ‘learning’ a model from data [45]. The field of machine learning addresses questions related to the efficient learning of models based on observed data [46,47]. With regard to process stochasticity, even if certain biological processes ultimately prove to be truly deterministic (as Einstein said, “I, at any rate, am convinced that He does not throw dice”), it may still be useful to treat some of these as stochastic; stochasticity is often a convenient mathematical surrogate of our ignorance [48,49].

One important guideline however is to keep models as simple as possible while adding as much complexity as is needed in order to test hypotheses of relevance
References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

● of special interest
●● of outstanding interest


This publication constitutes an outstanding and critical view of the scope and status quo of systems biology, incorporating the concepts of both holism and reductionism.


This publication highlights the social challenges involved in big interdisciplinary research and discusses the implementation of a potential framework for successful projects.


In this paper, the authors study the spatiotemporal cell population behaviour triggered by high-throughput microfluidic perturbations, that is, analogue gene expression triggered by digital NF-κB activation.


12. Galas D, Hood L: Systems biology and emerging technologies will catalyze the transition from reactive medicine to predictive, personalized, preventive and participatory (P4) medicine. IBC 2009, 6-1-4.


This publication discusses the low predictive power of ion channel mutations with regard to epilepsy and highlights the importance of understanding the context-dependant functional downstream effects of such ion channel mutations for improving personal risk assessment.


This publication discusses how cell cycle oscillations may arise. In their tutorial, the authors elaborate on how to adapt the complexity of models to the complexity of the biological systems being studied, and comment on the extent to which the theory of non-linear systems can assist the understanding of observed events.


36. Majumder D, Mukherjee A: A passage through systems biology to systems medicine: adoption of middle-out rational


Letter to the editor: “Systems biology versus reductionism in cell physiology”

Mark A. Knepper, Viswanathan Raghuram, Davis Bradford, Chung-Lin Chou, Jason D. Hoffert, and Trairak Pisitkun

Epithelial Systems Biology Laboratory, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland

TO THE EDITOR: The following is a response to the editorial comment of Prihandoko and Tobin (15) about our recent paper in American Journal of Physiology-Cell Physiology (2), which addresses a key question in modeling of signaling networks: How to assign the protein kinases (from the entire 521-member kinome list) that are responsible for each measurable phosphorylation event in a given cell type. In our study, we used vasopressin-stimulated phosphorylation of the water channel protein, aquaporin-2, at serine-256 as an example because of its importance to the physiology of collecting duct principal cells. We thank Prihandoko and Tobin for their thorough and well thought out summary of our paper. We write now to provide additional clarification regarding the epistemological approach, which was based on a systems biological framework rather than on reductionist principles. Understanding the two ways of doing experiments is aided by a bit of history.

Attention to the problem of how to make practical scientific inferences from scientific observations peaked in 19th century with John Stuart Mill’s book “A System of Logic” (12; see chapter “Of the Four Methods of Scientific Inquiry”). Mill’s work described several approaches built from two fundamental methods, viz. the “method of difference” and the “method of agreement.” From the viewpoint of modern biology, the former method is the basis of reductionist approaches and became dominant in the 20th century. The latter method is the basis of the newly resurgent systems biology approach. We can conceptualize the method of difference as the standard hypothesis-driven experiment in which a given variable is altered and another variable is observed. This approach thrived because it has often been feasible to make the targeted measurements needed and because statistical methods were developed early in the 20th century by Fisher and others to analyze such data (14). However, reductionist approaches have drawn fire in recent years because of perceived bias in publication (7). Critics claim that positive results from reductionist experiments are publishable (often whether true-positive or false-positive), while negative results are not. In addition, the statistical approach to analysis of reductionist data draws conclusions one experiment at a time, and does not generally utilize prior information to draw conclusions (4, 14), a problem that is circumvented in systems biological approaches. The latter, roughly equivalent to Mill’s method of agreement, looks broadly for correlations in comprehensive data sets and builds models based on these correlations. Comprehensive methodologies including large-scale proteomics, DNA microarrays, and “next generation” DNA sequencing have only recently become feasible because of the availability of genome-wide sequence data needed for mapping. Thus, biological approaches based on Mill’s method of agreement (systems biology approaches), heretofore impractical, have in the 21st century become feasible. Concomitantly, statistical methodologies for analysis of comprehensive data sets have followed, e.g., the use of Bayesian statistics. Our study (2) utilized the systems approach as summarized in the next two paragraphs. The commentary (15) appeared to retell the story that we presented as a series of separately interpreted reductionist experiments, thus losing the major message of our paper, viz. that Bayes’ theorem can be used to integrate multiple imperfect data sets to provide deeper, stronger conclusions than could be expected without data integration.

Our previous study in AJP-Cell (5) showed, using mass spectrometry, that protein kinases are low fidelity enzymes and when combined with prior observations (11) suggested that protein kinases gain specificity in the cell chiefly through factors that cause them to colocalize with specific substrates. From this and other studies, it was already clear that we can rely only on very general specificity constraints, basically whether they phosphorylate tyrosines or serines/threonines, and whether the latter are basophilic, acidophilic, or proline-directed. Thus, the question of what protein kinase(s) phosphorylate serine-256 of aquaporin-2 was not answerable simply by looking at the amino acid sequence surrounding it. More information was needed. To address the question, we integrated prior information from several sources using Bayes’ theorem to rank all 521 kinases in the rat genome with regard to the probability that they phosphorylate serine-256 of aquaporin-2 in the rat inner medullary collecting duct (IMCD). This included data gleaned from prior large-scale (proteomic or transcriptomic) experiments in the IMCD. This Bayes’ approach allowed us to utilize data, which in isolation did not answer the question, but narrowed the choices. For example, transcriptomic experiments divided the 521 protein kinase genes into those that were expressed in IMCD and those that were not detectable, and thus were unlikely to play a regulatory role regardless of kinase specificity. Use of Bayes’ theorem to integrate information from many sources is not new; it was used for example to establish the conclusion that smoking is harmful to health in the 1950s (3). However, as far as we can tell, the use of Bayes’ theorem to integrate multiple data sets in cell physiology is novel and it is therefore surprising that it was not explicitly discussed in the Prihandoko and Tobin commentary.

Using the Bayesian integration of prior data as a launching point, our study (2) addressed whether addition of inhibitor data could sharpen the Bayesian estimates. Protein kinase inhibitors have been used in physiology for many decades, always with tacit recognition that they inhibit multiple kinases in addition to the nominal target kinase. Now, the International Centre for Kinase Profiling (ICKP, http://www.kinase-screen.mrc.ac.uk/kinase-inhibitors) has provided profiling data for many commonly used protein kinase inhibitors. This comprehensive
data set identifies which kinases are and which kinases are not inhibited by a given small-molecule kinase inhibitor, and estimates the percentage of kinase activity remaining for relevant inhibitor concentrations. The ICKP data give new life to the use of inhibitors in physiological experiments by its comprehensive nature. It allowed phosphorylation data from immunoblotting of IMCD suspensions to be integrated with prior data using Bayes’ theorem, thereby significantly improving discrimination among candidate kinases involved in aquaporin-2 phosphorylation at serine-256. The overall Bayes’ analysis shows that the conventional wisdom, that protein kinase A phosphorylates this site in the collecting duct cell, is not any better supported by the data than roles for several other basophilic protein kinases including calcium/calmodulin-dependent protein kinase 2δ (Camk2d) and protein kinase B-α (Akt1). In fact, the top ranked protein kinase in the Bayes’ analysis, calcium/calmodulin-dependent protein kinase 2δ, was shown in mass spectrometry experiments to be as potent in phosphorylating aquaporin-2 in vitro as was protein kinase A, or more so.

In summary, our paper used a systems biological approach involving application of Bayes’ theorem to integrate multiple data sets. Such an approach appears to be new to cell physiology and appears to provide significant advantages for certain physiological problems such as the assignment of kinases to phosphorylation sites. We as authors recognize that the onus is on us to provide a persuasive argument for the systems approach. It may indeed be difficult for many biologists to embrace systems biology after a 100 years of reductionism. Toward that end, we invite the interested reader to view our previous writings about systems biology in AJP-Cell (8, 9) as well as recent articles by others in this journal (1, 6, 10, 13, 16).

ACKNOWLEDGMENTS

Present address of T. Pisitkun: Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand. Supported by NHLBI Division of Intramural Research (Project Z01-HL-001285).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS


REFERENCES

Instruction Format:
- One 1.5 hour overview/lecture per week
- One 1.5 hour literature review/discussion session per week

Course Requirements:
1. Attendance
2. Participation in literature and discussion sessions

Graduate Students:
3. Grant proposal (6 page limit) due week of April 23
4. Student Grant Review session on week of April 30

Undergraduate Students:
3. Two exams

Grading and Evaluation:
- Both in class attendance (10%) and discussion participation (25%) and (graduate students) the proposal (65%) or (undergraduate students) exams (65%) will be factors considered.
- Grading scale A(90%), B(80%), C(70%), D(60%), F(<60%)

References and Textbook:
- Reading literature and references provided one week prior to session
- No required textbook (suggested reading provided from selected textbooks such as The Cell or Genes and review articles).
Systems Biology
• History and Definitions
• Reductionism/Genetic Determination
• Holism/Emergentism/Robustness
• Revolutionary and Evolutionary Systems Biology
• Networks and Computational Biology
• Basic Molecular and Cellular Components

Required Reading

Background Book References

Spring 2019 – Epigenetics and Systems Biology
Discussion Outline – Systems Biology
Michael K. Skinner – Biol 476/576
Weeks 1 and 2 (January 8-10, 15 & 17, 2019)

Systems Biology

Discussion

Student 1 – Ref #1 above
- What are simulation and in silico experiments?
- How can this computational approach help medicine?

Student 2 – Ref #2 above
- What is mechanical robustness?

Student 3 – Ref #3 above
- What is the relationships with various organisms?
Systems Biology

Definition

History

Theory

Paradigm Shift

Parameters

Systems biology is a comprehensive quantitative analysis of the manner in which all the components of a biological system interact functionally over time. Such an analysis is executed by an interdisciplinary team of investigators that is also capable of developing required technologies and computational tools. In this model, biology dictates what new technology and computational tools should be developed, and, once developed, these tools open new frontiers in biology for exploration. Thus, biology drives technology and computation, and, in turn, technology and computation revolutionize biology.

"systems biology is the study of an organism, viewed as an integrated and interacting network of genes, proteins and biochemical reactions which give rise to life " (Hood 2005).

Systems Biology Theory

Evolutionary Systems Biology- Extension of classical biology paradigm with new technology

Revolutionary Systems Biology- New paradigm shift in biology with altered perspective on causal relationships and systems
Evolutionary Systems Biology History

Systems biology extension current paradigm and history of biology with new technology

300BC Aristotle, System has 4 properties or causes: Material, Formal, Efficient, Teleological
200AD Galen (Roman Physician), Teleological important role in organism function
1500s Fernel, Systematic approach Anatomy
1600s Harvey, Physiology, Cell Biology, Circulation
1700s Newton, Physics leads to mechanistic determinism to explain systems
   La Mettrie, Define Biological Machine (eg Clock)
1800s Bernard, Father physiology and integration biological systems (milieu interieur)
   -Discovery DNA/Structure/Genes (Molecular Biology)
   -Computational Biology (non-equilibrium thermodynamics and kinetics metabolism)
2000s -Genome Sequence
   -Oomics Technology

Evolutionary System Biology Definitions

Extension of traditional biological paradigm

Marc Kirchner 2005

"Systems biology is the study of the behavior of complex biological organization and processes in terms of the molecular constituents"

Westerhoff and Alberghina 2005

Systems biology is “nothing but good old physiology” or that is “molecular biology claiming additional money”

Sorger 2005

"System biology aim is to build numerical models of biological processes and test the models experimentally"
Revolutionary Systems Biology History

Jan Smuts (1870-1950), South Africa, Defined Holism (Tendency in nature to form wholes that are greater than the sum of the parts through creative evolution)

Alfred Whitehead (1861-1947), USA, Defined Organisms (Philosophy of organism to explain the complexity of natural processes- including biological organisms)

Ludwig von Bertalanffy (1901-1972), Austria, Defined Disequilibrium (Biological organisms are open systems, which respond to changes in environment, such that disequilibrium is state of living organism and equilibrium is death)

Norbert Wiener (1894-1964), USA, Defined Cybernetics (Application mathematics to explain biological mechanisms)

Joseph Woodger (1894-1981), UK, Defined Bauplan (Bauplan as the essential structural plan or morphology of an organism body plan, eg vertebrates)

Conrad Waddington (1905-1975), Scotland, Defined Epigenetics (Discuss later)

Walter Elsasser (1904-1991), Hungarian, Defined Biotic (Laws not reducible to physical or chemical laws)

1980s Theoretical Biology Holism (Elsasser and Laszlo) (Butterfly Effect)

1990s High throughput sequencing and expansion epigenetic area

2000s Sequence genome and transcriptome (Omnics technologies)

Revolutionary Definitions for Systems Biology

Leroy Hood (2005)

“The inter-relationships of all the elements in a system rather than studying them one at a time”

Methodological Approach-

1) Develop simple descriptive, graphical, or mathematical model of how system functions

2) Identify and define the various components of the system and their state (eg omics)

3) Disturb the system with external perturbation and document changes in the components

4) Integration of the two data sets from step 3 and comparison to model in step 1

5) Adjust model until harmony or conjunction exists between data and model

Hiroaki Kitano (2002)

Four factors for comprehensive systems biology definition

1) System Structure, organization of components (macromolecules, genes, cells, tissues etc

2) System Dynamics, interactions between or relationships of the various hierarchical levels over time

3) Systems Control Method, regulatory mechanisms involved in the maintenance of the organizational hierarchy

4) Systems Design Method, hierarchical organization with specific properties and manipulate

Table 1. Comparison of features for revolutionary and evolutionary systems biology

<table>
<thead>
<tr>
<th>Revolutionary systems biology</th>
<th>Evolutionary systems biology</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Holism</td>
<td>Reductionism</td>
</tr>
<tr>
<td>2. Top-down causation</td>
<td>Bottom-up causation</td>
</tr>
<tr>
<td>3. Epigenetics</td>
<td>Genetic determinism</td>
</tr>
<tr>
<td>4. Emergentism</td>
<td>Mechanism</td>
</tr>
<tr>
<td>5. Synergy</td>
<td>Synthesis</td>
</tr>
<tr>
<td>6. Robustness</td>
<td>Homoestasis</td>
</tr>
<tr>
<td>7. Nonlinear dynamics</td>
<td>Linear stasis</td>
</tr>
</tbody>
</table>
Reductionism
The view that the ultimate scientific understanding of a range of phenomena is to be gained exclusively from looking at the constituents of these phenomena and their properties

Ontological Reductionism
That complex phenomena are reducible to or determinable by simpler entities and forces that compose them (eg genetic determinism) and (bottom-up or upward causation)

Methodological Reductionism
Reducing wholes to parts and explaining the higher levels in terms of lower ones as the ultimate direction for all scientific research (eg physics)

Epistemological Reductionism
Reduction of scientific knowledge, whether in terms of theories, laws, or explanations, from a higher level of organization to that of a lower or more basic one

Neuropharmacology beyond reductionism - A likely prospect.
Margineanu DG.

Abstract
Neuropharmacology had several major past successes, but the last few decades did not witness any leap forward in the drug treatment of brain disorders. Moreover, current drugs used in neurology and psychiatry alleviate the symptoms, while hardly curing any cause of disease, basically because the etiology of most neuro-psychic syndromes is but poorly known. This review argues that this largely derives from the unbalanced prevalence in neuroscience of the analytic reductionist approach, focused on the cellular and molecular level, while the understanding of integrated brain activities remains flimsier. The decline of drug discovery output in the last decades, quite obvious in neuropharmacology, coincided with the advent of the single target-focused search of potent ligands selective for a well-defined protein, deemed critical in a given pathology. However, all the widespread neuro-psychic troubles are multi-mechanistic and polygenic, their complex etiology making unsuited the single-target drug discovery. An evolving approach, based on systems biology considers that a disease expresses a disturbance of the network of interactions underlying organismic functions, rather than alteration of single molecular components. Accordingly, systems pharmacology seeks to restore a disturbed network via multi-targeted drugs. This review notices that neuropharmacology in fact relies on drugs which are multi-target, this feature having occurred just because those drugs were selected by phenotypic screening in vivo, or emerged from serendipitous clinical observations. The novel systems pharmacology aims, however, to devise ab initio multi-target drugs that will appropriately act on multiple molecular entities. Though this is a task much more complex than the single-target strategy, major informatics resources and computational tools for the systemic approach of drug discovery are already set forth and their rapid progress forecasts promising outcomes for neuropharmacology.

Overcoming the Newtonian paradigm: the unfinished project of theoretical biology from a Schellingian perspective.
Gare A.
Prog Biophys Mol Biol. 2013 Sep;113(1):5-24

Abstract
Defending Robert Rosen's claim that in every confrontation between physics and biology it is physics that has always had to give ground, it is shown that many of the most important advances in mathematics and physics over the last two centuries have followed from Schelling's demand for a new physics that could make the emergence of life intelligible. Consequently, we argue that reductionism is not a reasonable anti-reductionist tradition. The history is used to identify and defend a fragmented but progressive tradition of anti-reductionist biomathematics. It is shown that the mathematically-physico-chemical morphology research program, the biosemiotics movement, and the relational biology of Rosen, although they have developed independently of each other, are built on and advance this anti-reductionist tradition of thought. It is suggested that understanding this history and its relationship to the broader history of post-Newtonian science could provide guidance for and justify both the integration of these strands and radically new work in post-reductionist biomathematics.
Holism (Revolutionary Systems Biology)
The living world consists in a reality that can be understood only in its global and inseparable unity. The whole is fundamental, not any one level. The whole is greater than the sum of its parts or of its levels.

Ontological Holism
Putting together the parts will not produce the wholes (such as living systems) or account for their properties and behaviors. Downward causation claims that higher order entities determine causally the properties or behavior of lower-level entities.

Methodological Holism
That life can only be understood by studying it as a whole. The world is disordered and it recognized that each hierarchical level requires its own research strategy not reducible to the methodological strategy below it.

Epistemological Holism
Complex wholes are considered not to be understandable from the mere knowledge of the behavior of the parts in isolation; only properties of the system as a whole may offer understanding.

The new holism: P4 systems medicine and the medicalization of health and life itself.

Abstract
The emerging concept of systems medicine (or ‘P4 medicine’-predictive, preventive, personalized and participatory) is at the vanguard of the post-genomic movement towards ‘precision medicine’. It is the medical application of systems biology, the biological study of wholes. Of particular interest, P4 systems medicine is currently promised as a revolutionary new biomedical approach that is holistic rather than reductionist. This article analyzes its concept of holism, both with regard to methods and conceptualization of health and disease. Rather than representing a medical holism associated with basic humanistic ideas, we find a technoscientific holism resulting from altered technological and theoretical circumstances in biology. We argue that this holism, which is aimed at disease prevention and health optimization, points towards an expanded form of medicalization, which we call ‘holistic medicalization’. Each person’s whole life process is defined in biomedical, technoscientific terms as quantifiable and controllable and underlain a regime of medical control that is holistic in that it is all-encompassing. It is directed at all levels of functioning, from the molecular to the social, continual throughout life and aimed at managing the whole continuum from cure of disease to optimization of health. We argue that this medicalization is a very concrete materialization of a broader trend in medicine and society, which we call ‘the medicalization of health and life itself’. We explicate this holistic medicalization, discuss potential harms and conclude by calling for preventive measures aimed at avoiding eventual harmful effects of overmedicalization in systems medicine (quaternary prevention).

Genetic Determinism
The view that genes (genotype) cause traits (phenotype)

Genetic determinism also referred to as- Geneticism, Genetic Essentialism and Genetic Fatalism

Strong Genetic Determinism- genotype “always” dictates phenotype

Weak Genetic Determinism- genotype “sometimes” dictates phenotype, also potentials or predispositions

Classical Genetics (Mendel) to Molecular Genetics (DNA) to Molecular Biology
Two applications of network-based analyses of GWAS. (a) GWAS analysis computes the association between a SNP and case/control, reporting a P-value for each SNP. (b) Casual gene identification is the problem of identifying a single causal gene (circled in red) for the phenotype from a larger locus of candidate genes that is significantly associated with the phenotype. (c) Causal network identification is the problem of finding a group of interacting genes (e.g. a signaling pathway or protein complex) containing SNPs that distinguish cases and controls.


Network-specific pathology and functional characterization of CN Thistle2 module. (A,B) Differential connectivity analysis reveals network-level alterations (light purple) that were not observed by previous differential expression analysis in the same cohort (dark purple). (B) Venn diagrams depict the number of genes identified by differential connectivity (light purple) and differential expression analyses (dark purple), as well as their overlap. (C) CN modules showing enrichment for previously published cell-type specific gene signatures identified by FACS (F) and in situ hybridization (I) experiments. Fisher’s exact test odds ratios are plotted only for modules with P < 0.05, two-sided, Bonferroni corrected. (D) Circos plot depicting FOXO3 as the top TF associated with Thistle2 in CN; rings are numbered 1 (outermost) to 5 (innermost). TF binding site enrichment scores are depicted in rings 2, 3, and 4 (Z score, Fisher’s score, and CompositeRank, respectively). Ring 5 depicts the differential expression profile of each TF in HD (log2(FP)). Blue histogram height (ring 1) reflects the cumulative scores of each TF based upon rings 2–5, with taller heights depicting greater relevance to Thistle2.


Abstract
One major approach to weight stigma reduction consists of decreasing beliefs about the personal controllability of-and responsibility for-obesity by educating about its biogenetic causes. Evidence on the efficacy of this approach is mixed, and it remains unclear whether this would create a deterministic view, potentially leading to detrimental side-effects. Two independent studies from Germany using randomized designs with delayed-intervention control groups served to (1) develop and pilot a brief, interactive stigma reduction intervention to educate N = 128 university students on gene × environment interactions in the etiology of obesity; and to (2) evaluate this intervention in the general population (N = 128) and determine mechanisms of change. The results showed (1) decreased weight stigma and controllability beliefs two weeks post-intervention in a student sample; and (2) decreased internal attributions and increased genetic attributions, knowledge, and deterministic beliefs four weeks post-intervention in a population sample. Lower weight stigma was longitudinally predicted by a decrease in controllability beliefs and an increase in the belief in genetic determinism, especially in women. The results underline the usefulness of a brief, interactive intervention promoting an interactionist view of obesity to reduce weight stigma, at least in the short term, lending support to the mechanisms of change derived from attribution theory. The increase in genetic determinism that occurred despite the intervention’s gene × environment focus had no detrimental side-effect on weight stigma, but instead contributed to its reduction. Further research is warranted on the effects of how biogenetic causal information influences weight management behavior of individuals with obesity.
Abstract

The concept of geneticization belongs to a style of thinking within the social sciences that refers to wide-ranging processes and consequences of genetic knowledge. Lippman's original use of the term was political, anticipating the onerous consequences of genetic reductionism and determinism, while more recent engagements emphasise the productivity and heterogeneity of genetic concepts, practices and technologies. This paper reconstructs the geneticization concept, tracing it back to early political critiques of medicine. The argument is made that geneticization belongs to a style of constructionist thinking that obscures and exaggerates the essentializing effects of genetic knowledge. Following Hacking's advice, we need a more literal sense of construction in terms of 'assembly' to give a clearer account of the relationship between processes and products. Using the 'assemblage' concept to explore the social ontology of genetics, the paper reviews three areas of the empirical literature on geneticization - disease classification, clinical practice and biosociality - to show that a new style of thinking has appeared within the social sciences. In the final assessment, the conditions that gave rise to geneticization are now obsolete. While it may serve as a useful ritual of debate, conceptually geneticization offers a limited account of the heterogeneity of socio-technical change.

Epigenetics

Waddington (1940s) coined term to describe environment-gene interactions that promote phenotype.

Non-genetic factors in the control of developmental processes and phenotype (? anti-genetic determinism)

Art Riggs (1996), defined as “mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence”

Epigenetics represents for many systems biologists a promise for control of biological phenomena unfulfilled by genetic determinism (Silverman 2004)
Epigenetic Mechanisms of Gene Regulation

- DNA Methylation
- Histone Modification
- Chromatin Structure
- DNA Organization into Domains (eg Loops)
- Nuclear Compartmentalization (eg nuclear matrix)
- Noncoding functional RNAs

Mechanism and Emergence

Mechanism:

Glennan 2002- “is a complex system that produces that behavior by the interaction of a number of parts, where the interactions between parts can be characterized by direct, invariant, change relating generalizations”

Machamer, Darden, Craver 2000- “are intitites and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions” (A to B to C)

Mechanisms are especially open to investigation particularly through experimentation

Table 1. Comparison of features for revolutionary and evolutionary systems biology

<table>
<thead>
<tr>
<th>Revolutionary systems biology</th>
<th>Evolutionary systems biology</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Holism</td>
<td>Reductionism</td>
</tr>
<tr>
<td>2. Top-down causation</td>
<td>Bottom-up causation</td>
</tr>
<tr>
<td>3. Epigenetics</td>
<td>Genetic determinism</td>
</tr>
<tr>
<td>4. Emergentism</td>
<td>Mechanism</td>
</tr>
<tr>
<td>5. Synergism</td>
<td>Synthesis</td>
</tr>
<tr>
<td>6. Robustness</td>
<td>Homeostasis</td>
</tr>
<tr>
<td>7. Nonlinear dynamics</td>
<td>Linear stasis</td>
</tr>
</tbody>
</table>

Emergence. Complex systems display properties, often called “emergent properties,” that are not demonstrated by their individual parts and cannot be predicted even with full understanding of the parts alone. For example, understanding the properties of hydrogen and oxygen does not allow us to predict the properties of water. Life is an example of an emergent property. It is not inherent in DNA, RNA, proteins, carbohydrates, or lipids but is a consequence of their actions and interactions. A comprehensive understanding of such emergent properties requires systems-level perspectives and cannot be gleaned from simple reductionist approaches.
“What is the difference between a live cat and a dead one? One scientific answer is systems biology. A live cat is the emergent behavior of the system incorporating those parts.”


De novo protein-coding genes originating from lncRNAs. (A) Computational pipeline for ab initio identification and meta-analysis of de novo genes in the hominoid lineage. (B) Number of de novo genes on the phylogenetic tree, with the branch length proportional to the divergence time. (C) Stacked histograms showing the percentage of de novo gene orthologs that also show expression in chimpanzee or rhesus macaque. (D) Bipartite graph showing relative expression levels of the transcripts and their nearby regions corresponding to de novo genes (orthologs) in human (chimpanzee or rhesus macaque). The nearby regions are defined as upstream and downstream regions with equal length to the corresponding gene. For each region, the relative expression was calculated by normalizing the expression level of the region with the sum of the expression levels of the genic region and the nearby regions. (E) Percentage of splicing junctions with supporting RNA-Seq reads in human, chimpanzee, and rhesus macaque. (F) For each pair of tissues, Spearman correlation coefficients were computed separately, and the extent of tissue-specific differences in de novo gene expressions are shown (based on the color scale). Dotted lines highlight parallel comparisons between two different species.
Contextual organismality: Beyond pattern to process in the emergence of organisms.
Díaz-Muñoz SL, Boddy AM, Dantas G, Waters CM, Bronstein JL.
Evolution. 2016 Dec;70(12):2669-2677.

The cooperation-conflict space is useful to visualize and evaluate potentially organismal interactions. Panel (A) illustrates organismality space (after Queller and Strassmann 2009) and some of the potential paths (numbered 1–4) organisms can move through under changing ecological contexts, such as development, resource availability, population size, and species interactions. In Panel (B), we provide examples of movement across organismal space in honey bee colonies (blue) and groups of microbial cells (red). In both examples, the cloud plot depicts the movement over “organismality space” and the labels represent the context that facilitates this change. The shading around the points is meant to convey the possibility of small changes in cooperation-conflict in any context.

Homeostasis vs Robustness

Homeostasis-

Claude Bernard (1800s)- “internal milieu’s constancy”

Cannon (1939)- “steady states in the body... a condition that may vary, but is relatively constant”

Migliani (2006)- “a mechanism for promoting the stability of phenotypic expression of a genotype when grown over a wide range of environments”

Illustration of environmental influences and the effect of perturbations on inner dynamics. In (A), two environments are shown (rich and minimal media). Plots adapted from (Freilich et al., 2010). In (B), a current state of an internal control can be modified by small or large perturbations (thick black arrows) pushing the agent-internal dynamics within the current boundary of attraction or far from it. NN, neural network. See main text for further details.

Robustness. Biological systems maintain phenotypic stability in the face of diverse perturbations imposed by the environment, stochastic events, and genetic variation. Robustness often arises through positive and negative feedback loops and other forms of control that constrain a gene’s output. This feedback insulates the system from fluctuations imposed on it by the environment. Positive feedback, in general, enhances sensitivity, whereas negative feedback can dampen noise and reject perturbations. Robustness is an inherent property of all biological systems and is strongly favored by evolution.
Robustness as an organizational principle

Robustness enables the system to maintain its functionalities against external and internal perturbations. This property has been widely observed across many species, from the level of gene transcription to the level of systemic homeostasis.
Illustration of redundancy (A) and distributed robustness (B). Plots show a hypothetical organization in which an upstream signal from the upper white circles is processed by a number of intermediate components (dark circles) to a downstream effector (lower white circles).
Developmental Plasticity and Robustness of a Nematode Mouth-Form Polyphenism.
Sieriebriennikov B, Sommer RJ.
Modularity. A further characteristic of complex systems is their modularity. Multiple useful definitions of a module exist. To an engineer, a module is a functional unit, a collection of parts that interact together to perform a distinct function. Such a module would have distinct inputs, things it is sensitive to, and outputs, things it controls. To a biologist, a module in a network is a set of nodes that have strong interactions and a common function. Modularity can contribute to both robustness of the entire system, by confining damage to separable parts, and to evolution, by simply rewiring modules. Furthermore, modularity decreases the risk of failure of the system by preventing the spread of damage in one part of the network throughout the entire network.
Table 1. Comparison of features for revolutionary and evolutionary systems biology

<table>
<thead>
<tr>
<th>Revolutionary systems biology</th>
<th>Evolutionary systems biology</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Holism</td>
<td>Reductionism</td>
</tr>
<tr>
<td>2. Top-down causation</td>
<td>Bottom-up causation</td>
</tr>
<tr>
<td>3. Epigenetics</td>
<td>Genetic determinism</td>
</tr>
<tr>
<td>4. Emergentism</td>
<td>Mechanism</td>
</tr>
<tr>
<td>5. Synergism</td>
<td>Synthesis</td>
</tr>
<tr>
<td>6. Robustness</td>
<td>Homeostasis</td>
</tr>
<tr>
<td>7. Nonlinear dynamics</td>
<td>Linear stasis</td>
</tr>
</tbody>
</table>

Table 1. Categorizations of systems biology

<table>
<thead>
<tr>
<th>Type One</th>
<th>Type Two</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hackett et al., 2017</td>
<td>Systems-oriented biology</td>
</tr>
<tr>
<td>Pelc et al.</td>
<td>Reductionist systems biology</td>
</tr>
<tr>
<td>Focus</td>
<td>Genetic determinism</td>
</tr>
<tr>
<td>HPPL</td>
<td>Synteny</td>
</tr>
<tr>
<td>1. Biological systems biology</td>
<td>Systems-oriented biology</td>
</tr>
<tr>
<td>2. Reductionist systems biology</td>
<td>Reductionist systems biology</td>
</tr>
<tr>
<td>3. Genetic determinism</td>
<td>Genetic determinism</td>
</tr>
<tr>
<td>4. Synteny</td>
<td>Synteny</td>
</tr>
<tr>
<td>5. Evolutionary systems biology</td>
<td>Evolutionary systems biology</td>
</tr>
<tr>
<td>6. Homeostasis</td>
<td>Homeostasis</td>
</tr>
<tr>
<td>7. Linear stasis</td>
<td>Linear stasis</td>
</tr>
</tbody>
</table>

Table 1. Comparison of features for revolutionary and evolutionary systems biology

<table>
<thead>
<tr>
<th>Revolutionary systems biology</th>
<th>Evolutionary systems biology</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Holism</td>
<td>Reductionism</td>
</tr>
<tr>
<td>2. Top-down causation</td>
<td>Bottom-up causation</td>
</tr>
<tr>
<td>3. Epigenetics</td>
<td>Genetic determinism</td>
</tr>
<tr>
<td>4. Emergentism</td>
<td>Mechanism</td>
</tr>
<tr>
<td>5. Synergism</td>
<td>Synthesis</td>
</tr>
<tr>
<td>6. Robustness</td>
<td>Homeostasis</td>
</tr>
<tr>
<td>7. Nonlinear dynamics</td>
<td>Linear stasis</td>
</tr>
</tbody>
</table>
Top-down models in biology: explanation and control of complex living systems above the molecular level.

**Lecture Outline – Systems Biology**

Michael K. Skinner – Biol 476/576
CUE 418, 10:35-11:50 am, Tuesdays & Thursdays
January 8, 10, 15 & 17, 2019
Weeks 1 and 2

**Systems Biology**
- History and Definitions
- Reductionism/Genetic Determination
- Holism/Emergentism/Homeostasis or Robustness
- Revolutionary and Evolutionary Systems Biology
- Networks and Computational Biology
- Basic Molecular and Cellular Components

**Required Reading**

**Background/Book References**

---

**Evolutionary Systems Biology**
- Ecosystem
- Populations
- Organisms
- Organisms

**Actual Systems Biology**
- Ecosystem
- Populations
- Organisms

**Revolutionary Systems Biology**
- Nonlinear
- Robustness
- Synergy
- Emergence
- Epigenetics
- Holism

---

**Linear Stasis**
- Ecosystem
- Populations
- Organisms
- Organisms

**Homeostasis**
- Physiology
- Organ Systems
- Organs
- Organs

**Mechanism**
- Cells
- Tissues
- Organelles
- DNA

**Genetic Determinism**
- Cells
- Tissues
- Organelles
- DNA

**Reductionism**
- Macromolecules
- DNA

**Holism**
- Ecosystem
- Populations
- Organisms
- Organisms

---

**Ecosystem**
- Populations
- Organisms
- Physiological
- Organ Systems
- Organs
- Tissues
- Cells
- Organelles
- Macromolecules
- DNA
Computational Biology

- Mathematical modeling
- Data set analysis to develop models

### Computational Models

- Model Scope (mathematical elements)
- Model Statement (equations)
- System State (dynamic, snapshot)
- Variables, Parameters and Constants
- Model Behavior (environmental and internal processes)
- Model Assignment (biology described mathematical)
- Data Integration (omics data)
Figure 2.1 Change of free energy along the course of a reaction. The substrate and the product are situated in local minima of the free energy; the active complex is assigned to the local maximum. The enzyme may change the reaction path and thereby lower the barrier of free energy.

Figure 2.2 Dependence of reaction rate υ on substrate concentration $S$ in Michaelis-Menten kinetics. $V_{max}$ denotes the maximal reaction rate that can be reached for large substrate concentration. $K_m$ is the substrate concentration that leads to half-maximal reaction rate. For low substrate concentration, υ increases almost linearly with $S$, while for high substrate concentrations υ is almost independent of $S$.

Table 2.3 Different approaches for the description of Michaelis-Menten enzyme kinetics.

Figure 2.3 General scheme of inhibition in Michaelis-Menten kinetics. Reactions 1 and 2 belong to the standard scheme of Michaelis-Menten kinetics. Competitive inhibition is given, if in addition reaction 3 (and not reactions 4, 5, or 6) occurs. Uncompetitive inhibition involves reactions 1, 2, and 4, and non-competitive inhibition comprises reactions 1, 2, 3, 4, and 5. Occurrence of reaction 6 indicates partial inhibition.
An unbranched metabolic pathway under hierarchical regulation. (a) The first enzyme is regulated through both transcriptional repression and allosteric activity inhibition by the end product. Enzymes in other steps might also be regulated through gene expression (in dashed arrows), but this is not explicitly considered here. TF denotes transcription factor. (b) The hierarchical supply–demand representation of the pathway (a). The lower part represents the classical metabolic supply-demand system, in which only the metabolic regulation (in this case allosteric inhibition) is considered. The letter 'X' denotes the penultimate product in the pathway. The supply is catalysed by enzyme E1 (i.e., E1 here or enzymes stemming from an entire operon). (c) Illustration of the steady-state properties of a supply–demand system in terms of changes in the flux, intermediate concentration and elasticity coefficients.
Parameter Estimations

- Regression (minimum of the function)
- Estimators (distance measure)
- Maximum likelihood estimation (Gaussian noise)
- Identifiability (landscape in parameter space)
- Bootstrapping (sampling and noisy data)
- Cross Validation (model fitting and prediction)
- Bayesian Parameter Estimation (parameter not fixed, random variables)
- Local and Global Optimization
- Machine Learning Algorithms (simulations)

(Mathematica / Matlab / Systems Biology Markup Language, SBML)
Patterning with activator-inhibitor systems. (A) Local activation and lateral inhibition generates spatially heterogeneous patterns. (B) Interactions between black and yellow pigment cells produce Turing patterns in zebrafish skin. Mutual inhibition between them functions as self-activation for the yellow cells. Each yellow cell activates distant black cells. Therefore, inhibition of the yellow cell by the black cell works as a lateral inhibition. (C) Different modeling approaches to spontaneous pattern formation.

Parameter Estimations

- Regression (minimum of the function)
- Estimators (distance measure)
- Maximum likelihood estimation (Gaussian noise)
- Identifiability (landscape in parameter space)
- Bootstrapping (sampling and noisy data)
- Cross Validation (model fitting and prediction)
- Bayesian Parameter Estimation (parameter not fixed, random variables)
- Local and Global Optimization
- Genetic Algorithms (simulations)

(Mathematica / Matlab / Systems Biology Markup Language, SBML)
Parameter Estimations

- Regression (minimum of the function)
- Estimators (distance measure)
- Maximum likelihood estimation (Gaussian noise)
- Identifiability (landscape in parameter space)
- Bootstrapping (sampling and noisy data)
- Cross Validation (model fitting and prediction)
- Bayesian Parameter Estimation (parameter not fixed, random variables)
- Local and Global Optimization
- Genetic Algorithms (simulations)

(Mathematica / Matlab / Systems Biology Markup Language, SBML)
Machine Learning Modeling

- Large data set with manipulations
- Test data set with known outcomes parameters (learning data set)
- Mathematical Algorithm development from training set
- Refine Algorithm development with large data set
- Final Algorithm should be correct with training set and reveal new biology insight

Methods of information theory and algorithmic complexity for network biology.

Abstract

We survey and introduce concepts and tools located at the intersection of information theory and network biology. We show that Shannon’s information entropy, compressibility and algorithmic complexity quantify different local and global aspects of synthetic and biological data. We show examples such as the emergence of giant components in Erdös-Rényi random graphs, and the recovery of topological properties from numerical kinetic properties simulating gene expression data. We provide exact theoretical calculations, numerical approximations and error estimations of entropy, algorithmic probability and Kolmogorov complexity for different types of graphs, characterizing their variant and invariant properties. We introduce formal definitions of complexity for both labeled and unlabeled graphs and prove that the Kolmogorov complexity of a labeled graph is a good approximation of its unlabeled Kolmogorov complexity and thus a robust definition of graph complexity.
Parameter Estimations

- Regression (minimum of the function)
- Estimators (distance measure)
- Maximum likelihood estimation (Gaussian noise)
- Identifiability (landscape in parameter space)
- Bootstrapping (sampling and noisy data)
- Cross Validation (model fitting and prediction)
- Baysian Parameter Estimation (parameter not fixed, random variables)
- Local and Global Optimization
- Genetic Algorithms (simulations)

(Mathematica / Matlab / Systems Biology Markup Language, SBML)
Networks

- Modules
- Nodes
- Clusters
- Interactomes

### 4.1 Structure of Biochemical Networks

**Summary**

The structure of complex biochemical systems — e.g., metabolism or transcriptional regulation — can be represented by networks. Nodes typically correspond to molecule types or genes, while edges represent, for instance, molecular interactions, causal influences, or correlations in high-throughput data. To detect significant structures that deserve further explanation, networks can be compared to random graphs with defined statistical properties. Various characteristic structures have been found in biological networks, including scale-free degree distributions, small average path lengths, modules and clustering, as well as network motifs.
Figure 7.1. A section of the metabolic network of a "simple" bacterium. Note that each point (each chemical compound) is linked to any other point via the complexity of the network.
Figure 3.2 Small graphs. (a) Directed graph with 6 nodes and 9 edges. (b) An undirected graph with similar topology. (c) By rewiring, we can obtain a new graph without changing the degrees \( k_i \).
(a) A network of 750 nodes was generated by means of the PS model, with target average node degree $2m = 10$, scaling exponent $\gamma = 2.75$ and network temperature $T = 0$. The network is embedded to the hyperbolic plane $\mathbb{H}^2$. An external file that holds a picture, illustration, etc. Object name is srep30108-m31.jpg with LaBNE to reveal the angular position of the nodes in the hyperbolic circle containing the network. (b) Finally, the radial coordinates of the nodes are assigned, so that they resemble the rank of each node according to its degree. By the colour of the nodes, which highlights their angular coordinates, one can note that the embedding by LaBNE is rotated by some degrees with respect to the actual node angular coordinates obtained with the PS model. This does not impact the hyperbolic, distance-dependent connection probabilities, because distances are invariant under rotations. Edges in the raw embedding by LaBNE are not shown for clarity.

\[ \text{Input: } A, \text{ the } N \times N \text{ adjacency matrix representing network } G = (V, E) \]
\[ \text{Output: } Y_{\mathbb{H}^2} \text{, the hyperbolic coordinates for the set of nodes } V \]

1. Compute the average node degree of the network $2m$
2. Determine the network's scaling exponent $\gamma$:
   \[ \beta = 1/(\gamma - 1) \]
3. Compute the degree matrix $D$
   \[ L \leftarrow \tilde{D} - A \]
4. Embed $G$ to $\mathbb{H}^2$ via $L Y_{\mathbb{H}^2} \approx \lambda_{k+1} D V_{\mathbb{H}^2}$ with $k = 2$
5. Since the smallest eigenvalue is 0, $Y_{\mathbb{H}^2} = [y_1 = v_1, y_2 = v_2]$
6. Sort nodes decreasingly by degree and label them $i = [1, 2, \ldots, N]$
7. Assign each node with radial coordinates $r(i) = 2\beta \ln(i) + 2(1 - \beta) \ln(N)$
8. \[ \theta \leftarrow \arctan(y_2/y_1) \]
9. Finally, $Y_{\mathbb{H}^2} \leftarrow [r, \theta]$

Note that to embed a network $G$ to $\mathbb{H}^2$, the truncated spectral decomposition of $L$ is used. This gives the closest approximation to the eigen-decomposition by a matrix $\lambda \times 1$ of rank $\lambda_{k+1}$, and ensures that the computational complexity of LaBNE is $O(N^2)$. 

Note: To embed a network $G$ to $\mathbb{H}^2$, the truncated spectral decomposition of $L$ is used. This gives the closest approximation to the eigen-decomposition by a matrix $\lambda \times 1$ of rank $\lambda_{k+1}$, and ensures that the computational complexity of LaBNE is $O(N^2)$. 

31
Efficient embedding of complex networks to hyperbolic space via their Laplacian.
Alanis-Lobato G, Mier P, Andrade-Navarro MA.

1. Feedback is an essential part of molecular networks. It allows the cell to adjust the repertoire of functional proteins to current needs.
2. A FL is primarily characterized by its sign: negative feedback for maintaining homeostasis, positive feedback for obtaining ultrasensitivity or multiple stable states of the cellular composition.
3. Negative feedback can cause oscillations if signal propagation around the FL is sufficiently slow. High Hill coefficients, additional positive FLs, or saturated degradation facilitates oscillations in a negative FL.
4. Positive feedback can come from strong self-activation of a gene, from mutual repression between proteins, or by autocatalytic processes. In all cases one can obtain bistability if reactions involve some sort of cooperativity.
5. Metabolism of small molecules is characterized by a separation of scales. Typically, the intracellular pool of available small molecules is much smaller than the total amount of small molecules consumed during one cell generation.
6. Combinations of FLs in small-molecule uptake and metabolism can result in new behavioral features that are significantly different from a simple sum of the behaviors of single loops.

Summary Points
1. Feedback is an essential part of molecular networks. It allows the cell to adjust the repertoire of functional proteins to current needs.
2. A FL is primarily characterized by its sign: negative feedback for maintaining homeostasis, positive feedback for obtaining ultrasensitivity or multiple stable states of the cellular composition.
3. Negative feedback can cause oscillations if signal propagation around the FL is sufficiently slow. High Hill coefficients, additional positive FLs, or saturated degradation facilitates oscillations in a negative FL.
4. Positive feedback can come from strong self-activation of a gene, from mutual repression between proteins, or by autocatalytic processes. In all cases one can obtain bistability if reactions involve some sort of cooperativity.
5. Metabolism of small molecules is characterized by a separation of scales. Typically, the intracellular pool of available small molecules is much smaller than the total amount of small molecules consumed during one cell generation.
6. Combinations of FLs in small-molecule uptake and metabolism can result in new behavioral features that are significantly different from a simple sum of the behaviors of single loops.
8.2 Network Motifs

Summary

Signal transduction pathways and transcription networks process biochemical signals, which are coded by the concentrations, modifications, and localization of molecules. Regulatory networks contain characteristic motifs, which may reveal small subnetworks with typical dynamic behavior and specific regulatory functions. The adaptation motif, for instance, translates jumps of its input signal into a transient response, but in steady-state situations, its response is completely independent of the magnitude of the input. Other typical motifs comprise negative feedback loops, which speed up response times and contribute to stability, but also to oscillations, and the feed-forward loops, which can act as filters, sign-sensitive delays, or pulse generators.

Figure 8.7 Network motifs in the transcription network of the yeast S. cerevisiae. Gene names refer to specific examples in the network. Redrawn from Lee et al. [10].

Transcription network in bacteria

Figure 8.13 Epistatic effects reflect the shared functions of genes. Circles show the abundance of epistatic interactions between genes belonging to functional groups (rows and columns). Circle radii represent numbers of epistatic interactions. Aggravating and buffering interactions are shown as red and green pie slices, respectively. From Segre et al. [86].
Network Application Examples

(a) Gene regulatory network for Drosophila gap genes, showing relationship between input genes (Bcd, Cad, Hb, Tl) and output genes (Kni, Hb, Kr, Gt). (After figure 1 of Papatsenko and Levine (2011)).

(b) Concentration of Gap genes along the anterior posterior axis of the embryo. Model was fitted to this data. Hb, hunchback; Gt, giant; Kr, Kruppel; Kni, Knirps.

The (r)evolution of gene regulatory networks controlling Arabidopsis plant reproduction: a two-decade history.

A schematic of the network perturbations of one neural degenerative network over the 20 weeks of the progression of this disease in a mouse model. The red nodes indicate mRNAs that have become disease perturbed as compared with the brain transcripts of normal mice. The spreading of the disease-perturbed networks at the three different times points is striking – indicating the progressive disease perturbation of this neurodegenerative network.
Use of comparative genomics approaches to characterize interspecies differences in response to environmental chemicals: challenges, opportunities, and research needs.
OpenWorm: overview and recent advances in integrative biological simulation of Caenorhabditis elegans.

Table 5. Summary of the genes used in systems biology-based drug discovery paradigms and the names of the hypotheses that can be inferred from these analyses.

<table>
<thead>
<tr>
<th>Role type</th>
<th>Data type</th>
<th>Hypotheses from a computational analysis of interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemistry</td>
<td>Small molecule interactions</td>
<td>1. Identify common network and sequence conserved interacting proteins</td>
</tr>
<tr>
<td>Pathology</td>
<td>Drug-Protein interactions</td>
<td>2. Identify drug targets in human tissues</td>
</tr>
<tr>
<td>Genomics</td>
<td>Drug-RNA interactions</td>
<td>3. Identify drug targets in human tissues</td>
</tr>
<tr>
<td>Bioinformatics</td>
<td>Drug-Target interactions</td>
<td>4. Identify drug targets in human tissues</td>
</tr>
</tbody>
</table>

Network modeling by iterative refinement

Data Acquisition
High-throughput "omics" data

Global Databases

Analysis Modules
Network Refinement
Hypothesis Generation

Network Visualization/Modeling