Complexity in Genetic Networks

Prasanth Sankar

1 Introduction

The most challenging problem in biology in the post-genomic era is to understand how the developing organism starts from a single cell, which divides and modifies it into many different classes of cells and many specific shapes, yet produces a complete organism with little individual variation. A large body of the emerging data indicates that the early development occurs through signaling interactions that are genetically programmed, whereas at later stages, the development of complex traits is dependent on external inputs as well[?]. A quantitative description of this entire process, if it is possible, with certainty will be the culmination of much of biology. Conventional genetic thinking in the pre-genomic era assumed the existence of as many genes as required for all the early developmental processes to take place. But the finding that the number of genes in humans is surprisingly low (approximately 31,000) compared to that of flies and worms(approximately 12,000 and 19,000 respectively)[?] eventhough humans seem to be phenotypically much more complex, is challenging the assumption of linearity in gene functioning. This suggests a picture of genome as a complex interacting system of finite number of components whose mutual interactions with others as well as the environment produces the seemingly infinite genetic control needed for the development and functioning of an organism. In addition to individual interacting genes it is found that in the genome itself there are master genes known as the regulatory genes which coordinates the action of many genes by controlling the expression of a certain combination of genes for a particular biological processes. Here, interactions between the genes and with the environment are mediated by proteins. The general study of such a system can be thought of either as a genetic network if we concentrate on the activity or inactivity of the genes for a particular process without bothering about the details of proteins involved or as a signaling network if we restrict our study to the nature, interactions and pathways of proteins involved with genetic activity taken as an initial condition or driving force. In this paper an attempt is made to review some of the recent attempts in quantifying the nature and behavior of genetic networks. Since theory or modeling and experiment are intricately connected and mutually reinforcing in such studies, some of the experimental procedures used is also mentioned.

Before proceeding into the details of the methods used, some properties unique to biological systems as compared to other physical systems should be mentioned. The biological systems cannot be thought of as machines drawn to a well defined design since the system can rebuild itself within a range of variable parameters during the course of development, functioning and evolution. To make the studies systematic one needs to quantify the boundary of this variability and this is made complicated by the fact that the boundaries are determined by as yet unknown combination of intrinsic capability and external inputs[?]. This varies from comparatively shorter time scales during the life time of an individual in response to varying external conditions to evolutionary time scales of millions of years. The other challenging features are high degree of error tolerance, near optimal performance, degeneracy(the ability of elements that are different to perform the same functions) and the seeming existence of redundancy(the presence of inactive parts or many copies of same the same parts performing the same functions). The above properties defines the complexity of biological systems in addition to the usual physical descriptions of complexity, namely the number of components and subtelity of interactions and interfaces between them. The biological systems offer some simplifying features, if it can be thought of as that, as well. The most prominent of these is the modular structure which allows the grouping of functionally similar parts.

2 Methods

The experimental studies of genetic networks can be broadly divided into two approaches. One observes the system as it is and the other makes use of designer networks which are simpler in construction and which will allow one to test theoretical model predictions[?]. Naturally occuring genetic networks are studied by means of micro-array techniques.(This makes use of DNA chips which can monitor the activity of a particular gene by analyzing the transcription process happening at that gene). In certain organisms like yeast the expression pattern of entire genome can be monitored over time[?]. to understand the influence of a particular gene, experiments are carried out with and without the presence of that gene(genes if they are temperature sensitive can be inactivate by raising the temperature, or if this is not possible mutation of a particular gene can be made to make it insensitive). These studies point to highly coordinated expression of genes which can be classified into groups of coexpressed genes that share a common temporal expression pattern and in some cases these groups can be tied to specific cellular functions such as metabolism. This modular organization is abundant in the genome. It is to be noted that these coexpressed genes are often located on different chromosomes and these genes are regulated by one or more regulatory genes. Genome wide expression analysis can some times pinpoint such regulatory genes. There are some quantitative estimates for the number of required experiments for a full reconstruction of a genetic network [?]. Simple toy model analysis points to K log(N) micro array experiments, where K is the average number of regulatory genes that affect a given gene and N is the overall number of genes. Although this kind of database is currently not available, experiments of this kind are feasible in the future. There is an unpredictability here due to the fact that the scaling theoretical estimates for simpler models for real genetic networks with noisy data is not known. Another problem associated with the experimental data analysis is the fact that a simple correlation in genetic expression is in general not enough to infer causality between genes. One needs to develop sophisticated analytical tools that closely interact with the experimental data and which can obtain causal information. Till now the experimental successes is restricted to simple systems such as bacterial-viral interactions[?]

The above mentioned experimental and data analysis difficulties has prompted studies directed at designed networks(This is done by isolating certain genes and proteins involved and doing experiments with these only). Here the hope is that by the analysis of simpler network behavior one will be able to outline the fundamental principles of genetic expression and regulation and the understanding gained can be applied to the analysis of real genetic networks. The long range goal of such work is to assemble increasingly complete models of behavior of natural systems, while maintaining at each stage the ability to test models in a tractable experimental system[?]. Some successes in this direction are the development of genetic switches (two state behavior of certain genetic regions), generation of oscillations in the concentration of cellular proteins, generation of positive and negative feedback networks and study of their effects on the stability of a particular gene expression state and generation of noise resistant oscillations using hysteresis or time delay[?]. The simpler nature of such systems allowed detailed modeling and allowed to deduce the influence of particular effects or interactions on the behavior of the system (There are two kinds of modeling in these cases: One treats genes as logic gates with two discrete states ON and OFF, and the dynamics describes how groups of genes act to change one another's state over time. This is reductionism at its extreme and is of very limited utility when one wants to modify the model by including more biological details. The other method uses the rate equation approach in which the variables of interest are the concentration of individual proteins within the cell and the dy-

namics describe the rates of production and decay of these proteins permitting the modeler to apply the techniques of nonlinear dynamics). The generation of such models allows one to suggests hypothetical experiments and out comes which might have relevance in the analysis of real networks.

The theoretical models which tries to predicts the behavior of the genetic networks independently of experimental data uses the modified versions of complex dynamical networks[?]. Here the nodes of the networks are particular genes and the edges represent the interactions mediated by proteins. A lot of complexity can be built into such networks. The nodes of the network and the interaction links between the nodes may be different and diverse. The complex wiring pattern can be used to incooperate structural complexity. Dynamical complexity can be incooperated by introducing complicated interaction between the nodes. The network may be allowed to evolve over time by the rearrangement of the wiring coupled to the network dynamics itself and this can be thought of as simulating biological evolution. Tools for treating such complex dynamical networks have been developed in the fields of graph theory and statistical physics while attempting the study of neural networks. Percolation theory can be used to study how information spreads over such a network. Attempts are made to predict certain experimental observations such as the presence of modular structure, ability of mutations to produce major structural innovation, the selective pressures that promote modularity and influence of genetic interactions on macroevolution as well as the boundary conditions on evolution and speciation. The next paragraph lists some such network models[?]
.

One approach uses a subclass of Boolean networks known as threshold networks. Here each node is allowed to take two discrete values that at some time is a function of the value of some fixed set of other nodes. Here, an updating rule is used based on the sum of states of neighbor nodes exceeding a certain threshold. Study of these networks show that these can to a certain approximation represent the behavior of transcriptional regulation of genes. Depending on the connectivity the network takes on different topologies. Below a critical connectivity the network breaks into modular regions and above it long range connections exists making the network vulnerable to local faults which can propagate throughout the network. If the effect or random mutations is taken into account by allowing random rewiring of the network, followed by an optimization check for acceptability of such a rewiring. Choice of this optimization check is arbitrary, and can be chosen as dynamics preserving or as having maximal overlap with idealized network or as something that preserves the expression pattern. Such networks exhibits certain scaling behaviors as well as some properties of real gene networks such as the switching of genes from inactive state to active state and vice versa. The optimization step can be chosen as a global optimization or as local optimization. If a local optimization condition is used the network is shown to evolve to a statistically stationary state with a given connectivity and the evolved network is shown to have considerable robustness against noise.

While comparing the above networks to experimental data (for this one can use the direct probing of genetic networks as well as evolution experiments of fast evolving organisms like E-coli) it is found that such network models can capture some of the features of the actual network like approximate connectivity, cooperative nature, divergence of the activity of redundant genes, and simplicity of biological expression pattern(This implies activity of a gene is low turning ON and OFF only a few times during the expression cycle). But it is also to be noted that in all cases the distinction between the predictions of a random network and the modeled work based on some optimization principle is not always clear. The main differences of genetic networks that cannot be accounted by random networks are the robustness or error tolerance, and low activity and these

features are captured by optimized networks.

3 Results and Discussion

Identification of the genetic network is one of the most challenging problems of post-genomic era. Traditional biological approaches of qualitative descriptions seems to fail in this case due to the fact that the amount of information to be processed and analyzed to get definite patterns is enormous. This calls for a combination of theoretical modeling firmly based on the experimental data. There are two major developments along these lines. The first theoretical approach goes hand in hand with experiments in the sense that simple systems are analyzed and modeled in detail and these systems are designed in the lab to verify the model predictions as well as to point out the fundamentals. The output from the experiment is used to refine the modeling and the hypothetical cases that can be considered by the model is further used in the design of experiments. The long term goal of such approaches is the incremental increase in complexity captured in experiments and modeling leading to full understanding of the entire genetic network. The second theoretical approach is based on first identifying certain abstract characterization of biological systems such as robustness or optimized behavior and basing the model construction on these principles and then comparing the model predictions with experimental data. Eventhough this approach is abstractly concrete, the predictions of the model are also along abstract lines in the sense that it cannot generate further information than that has already been put into its construction except for identifying certain generalities which when compared to the sheer diversity of biological systems seem to be not that important. Nor is it clear whether some other predictions of such models such as self organization or dynamical phase transitions have any relevance to biological systems. One step that will make this kind of approaches more useful seems to be analysis of real genetic networks for detailed global information. For this presently available experimental data is not sufficient but in the future such attempts will be tried.

The progress in this field at the present moment is restricted only by the availability of experimental data and theoretical tools to analyze and identify certain patterns from the data. Modeling of the data is necessitated by the fact that it corresponds to functions of numerous genes in parallel and linear and quantitative attribution of data to a particular event or component is not possible. A recent progress in this field is functional genomics[**?**] with the underlying logic that, to understand complex biological systems it might be necessary to accumulate enormous amounts of data on numerous genes and then use data-mining algorithms to find the underlying patterns in the data. This requires the setting up of web based databases and development of functional data analysis tools.

Another approach to these problems which can be undertaken be theorists, is to consider the different aspects of complexity of biological systems such as degeneracy, redundancy, error tolerance or efficiency and trying to give quantitative estimates of these. This will further involve the incooperation of evolutionary effects into conventional physical reasoning.

References

[1] Weng G., Bhalla U. S., and Iyengar R. *Science 284 92 1999*

[2] Hasty et al. , *Nature Reviews Genetics, 2 268 2001*

[3] Bornholdt S., *Biol. Chem., 382 1289 2001*

[4] Kim S. K., *Nature Reviews Genetics, 2 681 2001*