Support of Progress in Clinical Neurology by Models of Genetic Regulation

Paul Smolen, PhD; John H. Byrne, PhD

Many neurological disorders are based on mutations in 1 or more genes. To understand and optimally treat these disorders, it is necessary to understand the functions and regulation of the genes involved. It has become apparent that intuitive descriptions of gene regulation are often insufficient. A mathematical description adds precision and detail. Therefore, a mathematical description is important for networks of genes that underlie development, synaptic plasticity, and other complex biological processes. A mathematical description is also important to represent the combined effects of multiple genes that contribute to the phenotype of complex neurological disorders. In such gene networks, it is common for some of the gene products to regulate the expression of other network members. Also, the expression of all or some network members is commonly coregulated. Gene product proteins that regulate transcription are termed transcription factors (TFs), and many genes are activated by multiple TFs.

A mathematical model of a gene network consists of a set of equations describing the expression rates of the network members. These equations are usually differential equations, which represent the gene expression rates as functions of the amounts of regulating TFs. A computer integrates these equations over time, simulating time courses of the expression of network genes. The values of parameters, or constants in the differential equations, are adjusted until the simulated and experimental time courses are similar.

Mathematical models are important to understand how mutations, drugs, and environmental changes affect the behavior of gene networks. To simulate these effects, the equations or parameters are altered in a way that represents the mutation, environmental change, or function of a drug. The simulated time courses of gene expression with and without the change are compared. To test and improve the model, predicted responses to novel stimuli (eg, drug or neurotransmitter exposure) are compared with experimental gene expression time courses, and the model is altered to improve the comparison.

Aspects of neurology where modeling of gene regulation is likely to be useful include understanding and treating polygenic neurological disorders, using gene expression data to reconstruct the network of gene regulation in neurons and other cells, and designing vectors and transfection methods for gene therapy. These aspects are discussed below. First, however, it is of interest to review results obtained from relatively simple models that incorporate typical schemes of gene regulation.

MODELS OF SIMPLE GENE NETWORKS

Models of gene networks often exhibit steady states of fixed gene transcription rates and gene product concentrations. These states can be stable or unstable. Stable means that a small movement of the system variables away from the steady state will be followed by relaxation back to that state. Unstable means that a movement of
the variables away from the steady state will lead to further movement away from that state. Multistability refers to the coexistence of multiple stable steady states. Each state remains stable in the face of disturbances by weak stimuli, but strong stimuli can induce transitions between the states.

The type of gene regulation required for multistability is a positive-feedback loop, in which a gene activates its own expression indirectly or directly. An example is as follows: the product of gene A activates transcription of gene B, and the product of gene B activates gene A. Figure 1A illustrates multistability based on a model of a positive-feedback loop in which a single TF (TF-A) activates transcription of its own gene. The TF-A gene product is assumed to form a dimer of 2 protein molecules. The tf-a promoter incorporates an enhancer sequence or responsive element (TF-RE). Transcription of tf-a is increased when TF-A dimers bind to the TF-RE.

The system of Figure 1A can be described by a single differential equation, which expresses the rate of change of TF-A protein concentration as the synthesis rate minus the degradation rate. This model exhibits bistability (coexistence of 2 stable steady states). For a range of values of a parameter that describes how strongly TF-A activates transcription, the concentration of TF-A protein ([TF-A]) exhibits 2 steady states (Figure 1B). In 1 state, [TF-A] levels are low. Little TF-A protein is bound to the TF-RE, and therefore tf-a transcription levels are low. The positive feedback is not activated. A small basal rate of tf-a transcription is balanced by a small TF-A degradation rate proportional to [TF-A]. If [TF-A] levels are increased, regenerative positive feedback occurs, with TF-A activating tf-a transcription. At a high [TF-A] level, the positive feedback saturates as the TF-RE site becomes fully occupied with TF-A. The synthesis and degradation rates of TF-A again come into balance, creating a second steady state.

Several authors have addressed possible functional consequences of multistability in gene expression rates. Brief exposure of cells to a neurotransmitter or hormone might trigger the switch and lead to long-lasting changes in gene expression. Switching of progenitor cells among gene expression states has been hypothesized as a mechanism for cell differentiation. In neuronal circuits, induced steady states of gene expression could lead to the synthesis of proteins required for synaptic growth and the formation of long-term memory.

Oscillatory biological phenomena can depend on periodic variation of transcription. Mathematical modeling of biochemical reaction pathways and gene regulation has suggested that negative-feedback loops are generally important for sustaining oscillations of the rates of biochemical reactions, including gene transcription. In a negative-feedback loop, a gene represses its own expression, often indirectly. An example is as follows: the product of gene A activates transcription of gene B, and the product of gene B represses gene A. Circadian rhythms, generated by neurons in the suprachiasmatic nucleus, appear to be based on oscillations in gene expression. Experimental data and modeling have suggested that the central negative-feedback loop is based on modification of 1 or more proteins by a series of phosphorylations, with the phosphorylated proteins acting to repress transcription of their own genes.4,5

One gene network often implicated in the control of synaptic plasticity and memory formation is the network based on transcriptional regulation by the family of TFs that includes the protein termed the calcium ion/cyclic adenosine monophosphate (Ca2+/cAMP)–response element binding protein (CREB).6 Aspects of this network are illustrated in Figure 2. Neurotransmitters bind to receptors and activate G proteins. Production of second messenger molecules, such as cAMP, is enhanced. Kinases such as protein kinase A or mitogen-activated pro-
tein kinase are activated after elevation of cAMP levels. These kinases phosphorylate CREB and related TFs. Phosphorylated CREB, in turn, activates genes required for synaptic plasticity (synapse formation or growth) and the consequent formation and stabilization of long-term memories.

A positive-feedback loop may exist in this network (Figure 2). In this loop, phosphorylated CREB binds to DNA sequences termed Ca^2+/cAMP response elements (CREs). The CREB gene itself has a CRE near the CREB promoter, which can therefore activate CREB transcription. There may also be a negative-feedback loop, in which the gene for a TF termed inducible Ca^2+/cAMP-responsive early repressor (ICER) is induced by phosphorylated CREB. Upon binding to CREs, ICER closes the negative-feedback loop by repressing transcription of CREB. The positive- and negative-feedback loops could allow complex behaviors such as oscillations of gene expression rates. Modeling in our laboratory (P.S. and J.H.B., unpublished data, April 2003) suggests that positive feedback in this network can generate a form of bistability. A critical number of spaced stimuli (neurotransmitter exposures) may often be required to switch the network into a state corresponding to a long-lasting increase in the level of phosphorylated CREB, a long-lasting induction of genes regulated by CREB, and the consequent formation of stable long-term memory.

**IMPORTANCE OF MATHEMATICAL MODELS FOR UNDERSTANDING AND TREATING NEUROLOGICAL DISORDERS**

Models of the regulation of multiple genes will be important to accurately describe the etiology and progression of polygenetic neurological disorders. The phenotype of such disorders depends on a combination of variations (alleles) in a number of genes. Neurological conditions known or thought to be polygenic include many cases of Alzheimer disease (AD), autism, schizophrenia, reading disability, and attention-deficit/hyperactivity disorder. Extensive linkage studies are required to identify those genes with variant alleles that contribute to these disorders. Contributing genes have been identified for AD^7 and schizophrenia. Within the next few years, it is likely that contributing genes will be found for more polygenic diseases.

Models will be needed to predict and understand the effect of therapies for polygenetic disorders, because multiple modes of regulation of the genes involved are likely to make treatment responses too complex to understand solely by intuition. For example, proposed therapies may involve multiple drugs to compensate for altered functions or expression levels of the protein products of several genes. A model is necessary to predict the effects of different dosages of such drug combinations.

A further layer of complexity consists of modeling normal and aberrant neural circuitry involved in neurological disorders. Alzheimer disease, Parkinson disease, schizophrenia, and other conditions involve degradation or disruption of synaptic links among the hippocampus, cortex, and other brain regions. Mathematical models have been formulated to describe the synaptic disruptions in these disorders. Pharmacological or genetic therapies that affect synaptic formation or maintenance might be beneficial for these diseases. However, to design and modify such therapies, it will be essential to develop models that describe the relevant synaptic pathways and their dysfunction and how altering gene regulation affects synaptic plasticity and maintenance within these pathways.

Consideration of AD further illustrates how combining models of gene regulation with models of other elements of neuronal biochemistry is important. Such a combination provides a framework within which the effects of new therapies can be predicted. At least 4 genes have alleles that cause or increase the risk for development of AD (the β-amyloid precursor protein [APP] gene, presenilin-1, presenilin-2, and apolipoprotein E), and there are probably a number of other significant genes. These 4 genes all appear to influence the biochemical pathway that begins with the synthesis of APP and ends with brain deposition of plaques consisting largely of an APP product termed amyloid-β protein. It has been hypothesized that AD is due to neurotoxicity of aggregates of amyloid-β protein.

Mathematical modeling has been used to gain understanding of the dynamics of aggregation of amyloid-β protein and to predict the effects of treatments that interfere with aggregation. However, modeling has not yet predicted how agents that influence expression of the genes involved in the production of amyloid-β protein will affect its deposition. It would be useful to combine a model of aggregation of amyloid-β protein with a model of the biochemical reactions and gene regulation in the pathways for production and degradation of amyloid-β protein. The combined model would create a more complete framework to predict the efficacy of proposed

![Figure 2. Signaling pathway involving transcriptional regulation by calcium ion/cyclic adenosine monophosphate (Ca^2+/cAMP)–response element binding protein (CREB). Neurotransmitters act through G proteins to elevate levels of intracellular messengers (eg, cAMP, Ca^2+). Kinases such as protein kinase A are activated, resulting in phosphorylation (P) of CREB and related transcription factors. Phosphorylated CREB stimulates transcription by binding to enhancer sequences termed Ca^2+/cAMP-response elements (CREs). A positive-feedback loop (blue arrows) may regulate CREB synthesis. In this loop, CREB binds to CREs near CREB and activates CREB transcription. Inducible Ca^2+/cAMP-responsive early repressor (ICER) is an element of a negative-feedback loop (brown arrows). Transcription of ICER is increased when the level of CREB increases, and ICER in turn can bind to CREs near the CREB gene, repressing CREB transcription.](image-url)
therapies. Comparison of model simulations with experimental estimates of biochemical reaction rates, or gene expression rates, could identify the rate-limiting steps in the production and degradation of amyloid-ß protein. Drugs designed to alter the rates of these steps would have the greatest effect on the levels of amyloid-ß protein.

MODELS ARE NECESSARY TO EXTRACT MEANINGFUL INFORMATION FROM GENE EXPRESSION DATA

The advent of DNA microarrays spotted with large numbers of oligonucleotides allows the simultaneous measurement of expression time courses for up to approximately 30,000 genes. In animal studies, DNA microarrays have been used to monitor expression of groups of genes during development of the nervous system. Microarrays have also been used to compare normal gene expression with expression in AD and multiple sclerosis and after brain injury. In these studies, alterations in the expression of many genes have been reported, although further experiments are necessary to determine functional relevance of these differences.

Mathematical analysis and modeling are required if a picture of neuronal genetic regulation is to be assembled from such large, complex data sets. Computer software is being developed to construct models using gene expression time courses and data for binding of transcription factors to DNA sites. Such software can suggest which combinations of TFs regulate individual genes in the data set. When aspects of gene regulation in healthy individuals are well understood, then comparison with models of gene regulation in individuals with disorders is likely to generate hypotheses of disease causation and to suggest pharmacological interventions.

One important technique used in reconstructing gene regulatory networks is cluster analysis, which groups genes on the basis of similarly shaped time courses of expression. Cluster analysis can reveal similarities between the expression of genes with known and unknown functions. For the genes of unknown function, it is reasonable to hypothesize functions analogous to those of known genes with similar expression time courses. These hypotheses can then be tested with more specific experiments.

In the complementary field of proteomics, high-throughput methods are developed to characterize the amounts and interactions of all or most proteins expressed by neurons or other cells under a variety of conditions. Proteomics data can complement gene expression data by providing protein time courses. Including these data in models of gene networks allows more accurate modeling of gene expression regulation by TF proteins.

ROLE OF MODELS IN DESIGNING OPTIMAL VECTORS AND DELIVERY PROTOCOLS FOR GENE THERAPY

Several clinical trials of neuronal gene therapy have been reported or are under way. Transgenes are usually delivered by injection of a viral vector, although injection into cerebral ventricles of neural stem cells carrying a transgene may provide an alternative. Disorders that might be alleviated by neuronal transgenes include Canavan disease, various lysosomal storage diseases, and Parkinson disease.

Mathematical models have been used to design vectors with maximal transfection efficiency and to design procedures that maximize transfection for a given vector. However, detailed models that represent vector behavior, integration of transgenes into chromosomes, and regulation of transgene expression will be necessary to predict the effects of vectors or transgenes in vivo. Correction of polygenetic diseases is likely to require coordinated and also independently adjustable expression of several transgenes. Designing a transgenic system subject to appropriate endogenous regulation will be a challenge requiring collaborative experimentation and modeling.

Designing vectors to transfect neurons with multiple genes has become more feasible after advances in engineering of gene networks. Collaboration between modeling and experiments has aided construction of artificial networks of a few genes that exhibit desired behaviors, such as regular oscillations of gene expression rates. In one example, a 2-gene network was constructed that contained a positive-feedback loop, in which gene 1 inhibited gene 2 and vice versa. This system had 2 stable states. Gene 1 was highly expressed and gene 2 was repressed, or vice versa. This system behaved as an epigenetic switch, because brief changes in the environment (temperature changes or chemical exposure) would switch the system back and forth between the stable states. Packaging such gene networks into vectors that can reliably transfected neurons or neuronal stem cells will present significant technical challenges.

CONCLUSIONS

Mathematical modeling of genetic regulatory networks is likely to be an essential component of research to improve therapies for polygenetic neurological disorders, to understand neuronal gene expression in normal and diseased states, and to develop gene therapies for neurological disorders.

Accepted for publication March 6, 2003.

Author contributions: Study concept and design (Drs Smolen and Byrne); drafting of the manuscript (Dr Smolen); critical revision of the manuscript for important intellectual content (Drs Smolen and Byrne); administrative, technical, and material support (Dr Byrne); study supervision (Dr Byrne).

This study was supported by grant P01 NS38310 from the National Institutes of Health, Bethesda, Md, and grant N00014-01-1-1031 from the Defense Advanced Research Projects Agency, Washington, DC.

Corresponding author and reprints: John H. Byrne, PhD, Department of Neurobiology and Anatomy, W. M. Keck Center for the Neurobiology of Learning and Memory, The University of Texas–Houston Medical School, PO Box 20708, Houston, TX 77225 (e-mail: John.H.Byrne@uth.tmc.edu).
REFERENCES


Call for Papers

ARCHIVES Express

The ARCHIVES launched a new ARCHIVES Express section in the September 2000 issue. This section will enable the editors to publish highly selected papers within approximately 2 months of acceptance. We will consider only the most significant research, the top 1% of accepted papers, on new important insights into the pathogenesis of disease, brain function, and therapy. We encourage authors to send their most exceptional clinical or basic research, designating in the cover letter a request for expedited ARCHIVES Express review. We look forward to publishing your important new research in this accelerated manner.

Roger N. Rosenberg, MD
Editor