Analysis of Temporal Propagation Property to Evaluate Gene Regulatory Networks

Tetsuya Maeshiro
School of Library and Information Science,
and Research Center for Knowledge Communities
University of Tsukuba
1-2 Kasuga, Tsukuba, Ibaraki, 305-8550 Japan

Abstract

Harmonic Pulse Analysis (HPA) is a method to analyze gene regulatory networks, which measures the plausibility of gene regulatory networks. HPA is used to predict gene regulatory networks. Generated networks were analyzed by the harmonic pulse analysis. 7.3% of systematically generated networks were within 5% margin of the harmonic pulse density of the benchmark networks.

Harmonic Pulse Analysis (HPA) (Maeshiro 2010) is a method to quantitatively evaluate biological likeliness of gene regulatory networks, and the method was evaluated using known gene regulatory networks. A gene regulatory network denotes a set of genes, where a gene may regulate and be regulated by other genes, regulatory details, and substances that function as regulators.

Dynamic behavior of gene regulatory networks is important to understand biological phenomena, and we obtain much more information than static analysis because expressions of most genes of an organism are continuously switched on and off, controlled by binding of ligands that function as activators or repressors, which are some specific substance or products of other genes' expressions. Static analysis that treats network topology detects only control relationships among genes, and temporal information is lost. Network topology cannot clarify if multiple regulations (activation or repression controls) acting on a gene are simultaneously operated or not.

HPA detects (i) Selectivity of expression control, (ii) Propagation behavior of regulatory control and pulse lifetime, and (iii) Frequency domain property, which cannot be analyzed by conventional static analysis.

Basic method of HPA is to inject single pulse to some points in the gene regulatory networks, and observe how the injected pulse travels over the network. Nodes in networks to inject pulses are denoted "pulse injection points". All substances that act as external inputs to the target gene regulatory network, and all genes that constitute the network are pulse injection points. Any substance required to regulate a gene in the target gene regulatory network, but not the product of these genes, is treated as an external input to the network.

HPA mainly analyzes two aspects of gene regulatory networks. The first one is related to the pulse propagation, which is the lifetime of the injected pulses, and the pulse spreading behavior over the network. The second one is the frequency domain behavior of the whole network. Both analysis is based on the identical quantitative measure, but the latter one has broader significance because various wave transmission behaviors are observed by sweeping the frequency of injected pulses.

The pulse propagation value of a network is calculated as follows. First, for each pulse injection, total intensity for each pulse duration is computed for predefined duration in simulation framework. Then these values are collected to calculate the average and variance values for each pulse width, generating a vector of nondiscrete values. Then for frequency domain analysis, this vector is calculated for each injected frequency (duration of injected pulses) to generate a three-dimensional graph, or a matrix, where the horizontal plane (matrix row and column) consists of pulse duration axis and frequency axis, and the vertical axis (matrix value) is the cumulative pulse intensity.

Differing from the Pulse Flux Analysis we have proposed before (Maeshiro 2009), HPA injects pulses of various lengths and analyze the global behavior of the gene regulatory network. Analysis using single pulse type focus on gene activation and repression patterns. Conversely, injection of varied pulse lengths (harmonic pulses) is equivalent to analyze the response of gene regulatory network to the expression duration. Therefore, HPA adds a new dimension to the analysis, different from temporal dimension, which is related to the pulse shape. HPA extracts global response to the duration of stimuls to the network.

One pulse is injected into one injection point at a time. Simultaneous injection to multiple points is ideal, but the number of combination increases exponentially with the number of injection points, and test of all combinations is realistically impossible. If a gene network has n injection points, 2^n combinations exist, which is not computationally feasible, since even a small gene regulatory network for the analysis has approximately 100 injection points. An advantage of HPA is that the number of pulse injection tests is proportional to the number of injection points n.

The method consists on injecting single pulse of given duration to all injection points of the target gene regulatory network. The base pulse, denoted α-pulse, has intensity 1.0 and duration T, so its frequency is $1/2T$, because single period consists of active and silent parts each with duration T. α-pulse with intensity 1.0 denotes gene expression and therefore a gene regulated by single activator binding site. The opposite is also valid, as an α-pulse also may repress a gene regulated by single repressor site. HPA rely on the use of pulses of various periods. Employed pulse types are harmonics of α-pulse, with pulse durations $2T$, $4T$, $8T$, ..., doubling the frequency. A predefined threshold intensity of the α-pulse is defined as the minimum boundary to act as a ligand on a gene with single binding site, and longer pulses have stronger regulation (activation or repression) intensity. A ligand that regulates a gene is interpreted as successful in regulating if the cumulative input quantity of the single pulse considered at that moment is greater than T, a value that is same as the area of α-pulse. Multiple ligand regulation is evaluated based on the cumulative quantity of each ligand and regulatory relationships among input ligands.

When multiple ligands act simultaneously, the gene expression, or the output, is calculated separately for each ligand and then integrated. For the simplest case of two ligands A and B, the total quantities of A and B are calculated separately for time instances where both A and B are present (intersection time). The gene is regulated if each area of ligands A and B, that corresponds to their quantity, in intersection time is greater than T, which is equal to the area of α-pulse. Intersection is used because the AND relationship requires simultaneous presence of both ligands A and B. Constraints derived from α-pulse is the basic assumption of HPA to define the valid regulatory threshold. Evidently the gene is expressed or repressed depending on regulation type. Separate calculation for each ligand is necessary because the intensity of pulses may be different and not 1.0.

Some genes, for instance λ repressor, have multiple binding sites for the same ligand (Ptashne 2004), and the gene expression rate is proportional to the number of bound ligands. This regulatory mechanism is simplified in our model, and a single arrow represents for all cases of multiplicity by a single ligand. The purpose of multiple binding sites to recruite identical ligand is the quantitative regulation of gene expression, and fulfillment of one of binding sites is enought to regulate, mainly for gene activation. Therefore, the reduction from multiple to single ligand regulation still represents the control (activation or repression) type, and ignores only the quantitative aspect of the regulation result. This simplification is due to the fact that we focus on the pulse propagation behavior over the network, and not on the detailed quantitative analysis of individual gene expression. In our model, output pulse that represents gene expression quantity is fixed to a constant value, which justify the reduction of number of controls. Such diminution is excluded when binding sites are for different ligands. Moreover, precise model requires the incorporation of binding strength of ligands to binding sites, but it is unrealistic because most values are unknown.

Each pulse injection is analyzed based on ativation pattern of each gene in the network. Analyzed gene regulatory networks were selected from KEGG pathway database [1]. External ligands that act on genes also exist, which consist of outputs from external pathways, identified substances, unknown activators and repressors. These networks are used as benchmark networks. Artificial gene regulatory networks were generated to evaluate the estimation performance of HPA. Between 1 and 5 random modifications were introduced to the benchmark networks, generating 200 networks for each number of modifications. Introduced changes in networks are addition, deletion and change of regulation type and target genes. 2.4% of systematically generated networks were within 1% margin of the harmonic pulse density of the Sea urchin early embryo network. 7.3% of networks were within 5% margin, and 28.6% within 10% margin.

HPA evaluates more accurately the gene regulatory network architectures than methods that consider only the static topological network properties, since conventional methods ignore many time related aspects of gene network behavior. Inclusion of biological information such as protein function, specie, cell type (lineage) and developmental stage, should increase the accuracy of network evaluation. However, such data is incomplete and there is no definite approach to integrate these data and employ for evaluation. Applications to other biological networks such as metabolic reaction networks and signal transduction networks are also possible.

References

T. Maeshiro, S. Nakayama, "Harmonic Pulse Analysis to detect biologically plausible gene regulatory networks", SICE Annual Conference 2010, 3233-3239, 2010.

M. Ptashne, A. Gann, *Genes & Signals*, Cold Spring Harbor Laboratory Press, 2002.

T. Maeshiro, H. Hemmi, K. Shimohara, "Ultra-fast Genome wide Simulation of Biological Signal Transduction Networks - Starpack", *Frontiers of Computational Science*, 177-180, Springer, 2007.

T. Maeshiro, K. Shimohara, S. Nakayama, "Fitness function for evolutionary system to predict unknown gene regulatory networks", *Proceesings of the International Conference on Instrumentation, Control and Information Technology*, 2722-2727, 2009.

M. Levine, E. H. Davidson, "Gene regulatory networks for development", *Proc. Natl. Acad. Sci. USA*, 102, 4936-4942, 2005.

M. Ptashne, *Genetic Switch: Phage Lambda Revisited*, Cold Spring Harbor Lab. Press, 2004.

[1] http://www.genome.jp/kegg/pathway.html