Advanced Drug Delivery Reviews 65 (2013) 918–928

Contents lists available at SciVerse ScienceDirect

Advanced Drug Delivery Reviews

journal homepage: www.elsevier.com/locate/addr

Systems mapping of metabolic genes through control theory ☆

Guodong Liu [a], Lan Kong [a], Zhong Wang [a], Chenguang Wang [b,*], Rongling Wu [a,**]

[a] Division of Biostatistics and Bioinformatics, Pennsylvania State University, Hershey, PA 17033, USA
[b] The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, MD 21287, USA

ARTICLE INFO

Article history:
Accepted 10 April 2013
Available online 17 April 2013

Keywords:
Systems mapping
QTL
Metabolic system
Differential equations
Mixture model

ABSTRACT

The formation of any complex phenotype involves a web of metabolic pathways in which one chemical is transformed through the catalysis of enzymes into another. Traditional approaches for mapping quantitative trait loci (QTLs) are based on a direct association analysis between DNA marker genotypes and end-point phenotypes, neglecting the mechanistic processes of how a phenotype is formed biochemically. Here, we propose a new dynamic framework for mapping metabolic QTLs (*m*QTLs) responsible for phenotypic formation. By treating metabolic pathways as a biological system, robust differential equations have proven to be a powerful means of studying and predicting the dynamic behavior of biochemical reactions that cause a high-order phenotype. The new framework integrates these differential equations into a statistical mixture model for QTL mapping. Since the mathematical parameters that define the emergent properties of the metabolic system can be estimated and tested for different *m*QTL genotypes, the framework allows the dynamic pattern of genetic effects to be quantified on metabolic capacity and efficacy across a time-space scale. Based on a recent study of glycolysis in *Saccharomyces cerevisiae*, we design and perform a series of simulation studies to investigate the statistical properties of the framework and validate its usefulness and utilization in practice. This framework can be generalized to mapping QTLs for any other dynamic systems and may stimulate pharmacogenetic research toward personalized drug and treatment intervention.

Contents

1. Introduction

The formation of any complex traits or diseases occurs in a series of building blocks of cellular processes, involving metabolic pathways or protein–protein interaction networks [1]. Metabolomics, an emerging discipline of studying small-molecule metabolites, may be particularly useful for understanding the physiology of dynamic cellular processes and for diagnosis of disease [2,3]. For example, by studying granular

☆ This review is part of the *Advanced Drug Delivery Reviews* theme issue on "Mathematical modeling of systems pharmacogenomics towards personalized drug delivery".

* Corresponding author. Tel.: +1 410 614 4961; fax: + 1 410 955 0859.
** Corresponding author. Tel.: +1 717 531 2037; fax: + 1 717 531 0480.
E-mail addresses: cwang68@jhmi.edu (C. Wang), rwu@phs.psu.edu (R. Wu).

0169-409X/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.addr.2013.04.007

metabolic pathways, which reflect biological responses to exogenous and endogenous factors, the molecular mechanisms of cardiovascular diseases can be better uncovered [4]. It has been recognized that metabolic profiles of the neonate, in terms of levels of amino acid, organic acid and fatty acid oxidation metabolites, are associated with obesity, diabetes and cardiovascular disease in adulthood through metabolic programming [2].

The phenotypes of metabolites and their associations with disease phenotypes may be controlled by genes [2,5,6]. Therefore, the elucidation of the underlying genetic architecture of metabolic traits predisposing to diseases can enhance our understanding of the genetic regulation of complex metabolic networks. Genetic mapping by linking molecular markers and trait phenotypes in segregating populations has been instrumental for identifying the genetic determinants of metabolic profiles, known as metabolic quantitative trait loci (*m*QTL), in a relation to particular traits or diseases [7–9]. Traditional strategies for *m*QTL mapping are based on a direct association analysis between marker genotypes and end-point phenotypes. It is likely that these strategies affect our statistical inference about the dynamic regulation of *m*QTL, given that metabolic responses and fluxes respond to various environmental or physiological stimuli in a dynamic matter [10,11].

More recently, a more powerful mapping strategy has been developed for genetic mapping of complex traits by dissolving the phenotype of a trait into its developmental, physiological, anatomical or biochemical components and further mapping QTLs involved in coordination and organization among a web of interactive components [12]. This strategy, called systems mapping, capitalizes on a system of differential equations to specify the dynamic behavior of phenotypic formation as a biological system. Compared to traditional static mapping for a steady state, systems mapping possesses several unique advantages: first, it maps and studies the genetic architecture of complex traits from their underlying mechanistic processes and pathways, with results being biologically more meaningful; second, it makes full use of mathematical equations to simplify the complexity of trait formation into key tractable elements, enhancing the statistical power of QTL mapping; and third, it allows various hypotheses about the interplay between genes and development to be formulated and tested in a quantitative way. Systems mapping has been employed for genetic mapping in a variety of biological phenomena from biomass allocation [13] to circadian rhythm [14,15] to viral dynamics [16].

In this article, we develop a new dynamic framework for mapping *m*QTLs that control metabolic pathways within the context of systems mapping. For a particular metabolic pathway, one chemical is transformed through the catalysis of enzymes into the other which may further form a new chemical. We use a system of ordinary differential equations (ODEs) to describe and quantify this process, aimed to investigate how different metabolites function synergistically to produce a desirable final phenotype. We implement ODE aspects of metabolic pathways into a statistical mixture model in which a mixture component is represented by an *m*QTL genotype. Since the mathematical parameters that define the emergent properties of the metabolic system can be estimated and tested for different *m*QTL genotypes, the framework allows the dynamic pattern of genetic effects to be quantified on metabolic capacity and efficacy across a time-space scale.

The way of how genotype-specific ODE parameters for systems mapping are estimated represents a statistical challenge. The last five years have witnessed the development of various statistical methods for estimating ODE parameters based on state variables measured at multiple time points [17–20]. The estimation of ODE parameters within systems mapping is even more complex because the QTL genotypes that define the dynamic system are unobserved. Although conventional mathematical approaches for ODE solution have been successfully integrated into systems mapping [12,14,15], the performance may deteriorate quickly when the number of parameters increases. In order to effectively and efficiently solve ODEs for *m*QTL mapping, we propose to adopt the well-established results in control theory by representing the system of ODEs as a feedback control system. Control theory has proven to be useful in the study of biological and biochemical systems where the issues of regulation and control are central [21]. Characterization of general control rules that underpin metabolic dynamics is an important part of systems analysis in biology. It has been long recognized that many biological regulatory mechanisms have evolved to optimize cellular adaptation in response to external stimuli, concordant with the design principle of control theory. We have performed extensive simulation studies to investigate the statistical properties of systems mapping equipped with control theory. The results from computer simulation have validated the utility and usefulness of the new mapping model for *m*QTL identification in practice.

2. Modeling metabolic reactions as a system

We consider metabolic dynamics as a feedback control system and apply the concept of control theory to facilitate the optimization problem in *m*QTL mapping. To describe our mapping model, we use glycolytic oscillations, a well-studied metabolic reaction, as a case study. Glycolysis is a metabolic process that provides central energy in the form of ATP (adenosine triphosphate) for a living cell. As the most studied control system, many mathematical models have been developed to describe glycolytic oscillations, in which the concentrations of metabolites fluctuate. To identify the function and underlying mechanisms of glycolytic oscillations, Chandra et al. [22] have recently proposed a minimal three-reaction system with specific mechanisms both necessary and sufficient for oscillations in *Saccharomyces cerevisiae*, which is specified by a system of ODEs, expressed as

$$\begin{aligned}\dot{x} &= \frac{2y^a}{1+y^{2h}} - \frac{2kx}{1+y^{2g}} \\ \dot{y} &= -q\frac{2y^a}{1+y^{2h}} + (q+1)\frac{2kx}{1+y^{2g}} - (1+\delta) \\ \begin{bmatrix}\dot{x}\\ \dot{y}\end{bmatrix} &= \underbrace{\begin{bmatrix}1\\ -q\end{bmatrix}\frac{2y^a}{1+y^{2h}}}_{PFK} + \underbrace{\begin{bmatrix}-1\\ (q+1)\end{bmatrix}\frac{2kx}{1+y^{2g}}}_{PK} + \underbrace{\begin{bmatrix}0\\ -(1+\delta)\end{bmatrix}}_{Consumption}\end{aligned} \quad (1)$$

where functions $x = x(t)$ and $y = y(t)$ represent the time-varying concentrations of the lumped variable of intermediate metabolites and output ATP, respectively. Parameter a is the cooperativity of ATP binding to phosphofructokinase (PFK); h is the feedback strength of ATP on PFK; k is the intermediate reaction rate; g is the feedback strength of ATP on PK; q is the autocatalytic stoichiometry; and δ indicates the external perturbation in ATP consumption. Five parameters (a, q, g, h, k), each with a biological interpretation, determine the properties of glycolysis. As being modeled by system of ODEs (1), q ATP molecules are used up by PFK to produce a pool of intermediate metabolites, which are then fed into PK to generate $q + 1$ ATP molecules. So during each round of reactions, a net one ATP molecule is produced by PFK + PK reactions and then consumed at the third reaction that models the cell's use of ATP. ATP inhibits both PFK and PK reactions.

Chandra et al. [22] used control theory to illustrate how the trade-offs between efficiency and robustness arise from individual parameters in this model, including the interplay of feedback control with autocatalysis of network products necessary to power and catalyze intermediate reactions. In addition, they explicitly derived analytic equations for hard trade-offs with oscillations as an inevitable side effect. We integrate ODEs (1) into a systems mapping framework to devise a computational model for identifying specific *m*QTLs that control glycolytic oscillations and further explore how the change of QTL