

Bayesian experts in exploring reaction kinetics of transcription circuits

Ryo Yoshida*, Masaya M. Saito, Hiromichi Nagao and Tomoyuki Higuchi

The Institute of Statistical Mathematics, Research Organization of Information and Systems, 10-3 Midori-cho, Tachikawa, Tokyo, 190-8562, Japan

ABSTRACT

Motivation Biochemical reactions in cells are made of several types of biological circuits. In current systems biology, making differential equation (DE) models simulatable *in silico* has been an appealing, general approach to uncover a complex world of biochemical reaction dynamics. Despite of a need for simulation-aided studies, our research field has yet provided no clear answers: how to specify kinetic values in models that are difficult to measure from experimental/theoretical analyses on biochemical kinetics.

Results: We present a novel non-parametric Bayesian approach to this problem. The key idea lies in the development of a Dirichlet process (DP) prior distribution, called *Bayesian experts*, which reflects substantive knowledge on reaction mechanisms inherent in given models and experimentally observable kinetic evidences to the subsequent parameter search. The DP prior identifies significant local regions of unknown parameter space before proceeding to the posterior analyses. This article reports that a Bayesian expert-inducing stochastic search can effectively explore unknown parameters of *in silico* transcription circuits such that solutions of DEs reproduce transcriptomic time course profiles.

Availability: A sample source code is available at the URL <http://dweb.ism.ac.jp/~yoshidar/lisdas/>

Contact: yoshidar@ism.ac.jp

1 INTRODUCTION

Cells can sense many different environmental signals, and respond to extraneous stimuli by switching activation/inactivation of specific molecules, such as transcription factors, via several types of biological circuits. Kinetic biochemical modelling of such reaction circuits (often called biochemical reaction *networks*) using DEs realizes experiments *in silico* to enhance our understanding of the complex dynamic systems (Alon, 2006; Chen *et al.*, 2004; Wilkinson, 2006). It enables us to test a plausibility of current knowledge and hypothetical models on biological circuits and to create new further models when inconsistencies arise in simulation, domain knowledge and experimentally observed quantitative data (Yoshida *et al.*, 2008).

Despite of a growing need for simulation-aided studies on bio-pathways, some fundamental issues prevent us from drawing their full potentials. One critical issue addressed here is related to uncertainty in specified kinetic parameters of models: in many cases, we have no clear answer on how to specify kinetic values and initial conditions on simulations because biochemical reaction kinetics differ depending on a cell-by-cell

basis and intra-/extracellular environments such as temperatures, presence/absence of biochemical components in specific tissues.

With the aim to explore kinetic parameters in *in silico* biological circuits that make solutions of DEs fit experimental data, a range of advanced technologies have appeared in the last several years of computational systems biology, involving genetic algorithm (GA; Kikuchi *et al.*, 2003; Kimura *et al.*, 2004; Koh *et al.*, 2006), sequential Monte Carlo (SMC; Nagasaki *et al.*, 2006; Nakamura *et al.*, 2009), MCMC-driven parameter search methods (Finkenstädt *et al.*, 2008) and also gradient descent algorithms (Cao *et al.*, 2008; Fajarewicz *et al.*, 2007; Yoshida *et al.*, 2008). Quantitative data, such as transcriptomic profiles and protein expression values, are usually measured over times. In general, a Bayesian statistical method involves all the preceding methods as a unified framework: any estimation tasks are treated as evaluations of posterior distributions in which prior distributions embodying our prior belief on reaction kinetics are adapted to experimental data using likelihood functions. As demonstrated later, however, the practical relevance of the existing approaches remains far below what practical analyses require. Our view on this is that such a low performance depends highly on the lack of effective biochemical kinetic priors.

This article presents a previously unexplored class of prior distributions, called Bayesian experts on biochemical reaction kinetics. The key notion of our approach can be found in the construction of Dirichlet process (DP) prior distribution (Blei *et al.*, 2005; Escobar *et al.*, 1995; MacEachern *et al.*, 1998) whose base measure $G_0(\theta)$ is modelled by a mixture of m Gibbs distributions:

$$G_0(\theta) = \sum_{i=1}^m \frac{\alpha_i}{Z_i} \exp\left(-\frac{\phi_i(\theta)}{T}\right) \quad (1)$$

with mixing rates α_i ($i = 1, \dots, m$), temperature $T > 0$ and partition functions Z_i . The m potential functions $\phi_i(\theta)$ defined on unknown parameters θ reflect substantive knowledge on reaction kinetics for a given model such that regions of interest in the parameter space are prescribed prior to a posterior analysis. In many applications, we are able to collect several kinetic evidences on such local regions by observing temporal patterns of experimental data. For instance, rates of increasing and/or decreasing to which DE solutions have to match can be guessed in advance by observing steepness of relevant local patterns of given time course data. Alternatively, maximum and minimum levels of gene expression data provide a helpful clue to determinations of steady-state levels to which a transcription circuit model reaches in equilibrium. The Gibbsian mixture distribution, enabling Bayesian expert system in kinetic explorations, aggregates such m constraint conditions on the kinetic parameters that are induced automatically in light of experimental data and reaction speeds inherent in given models.

*To whom correspondence should be addressed.

This article describes our Bayesian experts, focusing on key examples of specific types of *in silico* transcription circuits. Transcription factor activities are modelled on Hill functions and mass action kinetics (Alon, 2006; Wilkinson, 2006) as described in Section 2. To emphasize the ability of the Bayesian expert prior, we explore unknown parameters using a simplified stochastic search algorithm without using more advanced computational techniques. The DP prior with additional technical details is introduced in Section 3. Experimental results appear in Section 4, demonstrating drastic efficiency gains from uses of the Bayesian experts and a benchmark on computational times. Section 5 concludes with more generalization of our approach. Supplementary Material and a sample program—called LiSDAS (Life Science Data Assimilation Systems)—are accessible at the Supplementary web site. LiSDAS features some basic functions to explore kinetic parameter values for given data and a model. The user can develop biological circuit models with the LiSDAS model construction file in place.

2 MODEL AND INVERSION ANALYSIS

2.1 Transcription regulatory circuits

The proposed Bayesian method will be derived and described, in later sections, with examples on analyses of transcriptional reaction kinetics. Before that, we briefly summarize specific forms of the DEs making up elements of the transcription circuit models.

2.1.1 Transcription Control of transcription factor-inducing gene expressions is a key regulatory mechanism in cells. Transcription factors can sense external stimuli of cells by mediations of signal transduction pathways. A transcription factor—activated by the transmitted signals—acts as either activator or repressor in transcription processes of specific target genes with many recruited enzymatic activities at the promoter regions. One commonly used DE for describing the transcriptional process is Hill function which is in theory derived by considering an equilibrium binding of a transcription factor to its target site on the promoter (see Appendix of Alon, 2006). The time evolution of mRNA concentration $[X]$ is modelled as

$$\frac{d[X]}{dt} = -\psi_{rd}[X] + \psi_{tc} \prod_{i \in \mathcal{A}} h_a([P_i]; K_i, \sigma_i) \prod_{i \in \mathcal{R}} h_r([P_i]; K_i, \sigma_i)$$

where the mRNA decay rate and the maximal rate of transcription are denoted by ψ_{rd} and ψ_{tc} . In addition to this, some studies commonly include a baseline ψ_b (e.g. Cantone *et al.*, 2009). The Hill functions involving the transcriptional activators \mathcal{A} and repressors \mathcal{R} are

$$h_a([P_i]; K_i, \sigma_i) = \frac{1}{1 + (K_i/[P_i])^{\sigma_i}} \rightarrow I([P_i] > K_i), \quad (2)$$

$$h_r([P_i]; K_i, \sigma_i) = \frac{1}{1 + ([P_i]/K_i)^{\sigma_i}} \rightarrow I([P_i] < K_i). \quad (3)$$

The unknown kinetic parameters σ_i and K_i control, respectively, location and steepness of each Hill function. As σ_i gets larger, the Hill function becomes steeper and looks more like the indicator function $I(\cdot)$, taking the value one if argument is true, otherwise zero. The multiple set of the Hill functions totally defines the AND gate that describes an abrupt increasing/decreasing of mRNA transcript level in response to excess concentrations of the regulatory proteins in \mathcal{A} and \mathcal{R} .

2.1.2 Mass action law Multimeric molecules can make up a transcription factor *complex* in any of transcription regulatory circuits. In modelling, we need to build a DE describing molecular binding reactions, such as protein–protein binding or one or more proteins interacting with small molecules. Instantaneous rate of a binding reaction is described here by mass action kinetics

$$\frac{d[C]}{dt} = -\psi_{cd}[C] + \psi_{cb} \prod_{i=1}^d [P_i]^{\tau_i}. \quad (4)$$

In the formation of a complex C , the binding efficiency among d molecules, P_i ($i = 1, \dots, d$), is determined on the affinity proportional to the active masses with the binding rate constant ψ_{cb} and powers τ_i .

2.1.3 Translation An mRNA transcribed at level $[X]$ proceeds to its translation process. The model that we employ describes a production of protein P simply as

$$\frac{d[P]}{dt} = -\psi_{pd}[P] + \psi_{tl}[X] - \psi_{cb} \prod_{i=1}^d [P_i]^{\tau_i}. \quad (5)$$

Instantaneous rate of translation is represented by the linear functions in the first two terms with the rates of degradation ψ_{pd} and translation ψ_{tl} , respectively. The last term involving (4) appears in (5) when the protein P binds to the other $d - 1$ molecules $\{P_i | P_i \neq P, i = 1, \dots, d\}$ in system. Note that we omit dissociation processes of the protein–protein complex C , relying on the assumption that molecular stability of C is much higher and reaction kinetics involved in the dissociation is negligible in the entire system.

2.2 State space model

A biological circuit is modelled as a set of DEs $dx_i(t)/dt = f_i(\text{pa}_i(x), \psi)$ defining rates of change in evolving concentrations of the p biological entities, $x(t) = (x_i(t))_{1 \leq i \leq p}$, over continuous times $t \in \mathcal{T}$. In the above context, each endogenous variable involved in a bio-pathway makes up a state variable $x_i(t)$. The i -th state variable is regulated by the parent variables $\text{pa}_i(x)$ —a subset of the state variables appearing in the right-hand side of each DE—with the rate equation f_i defined by the kinetic parameters ψ . Conduction of quantitative experiments along times enables us to measure changes of the *unknown* state variables, $x(t)$, directly or indirectly via observed data $y(t) = (y_i(t))_{1 \leq i \leq p} \in \mathbb{R}_+^p$ at discrete time points $t \in \mathcal{N} \subset \mathcal{T}$. Here, to cope with partially observed systems where data of the state variables are in part missing, it is helpful to define a set of observable variables, $i \in \mathcal{S} \subset \{1, \dots, p\}$. For instance, in transcriptomic experiments using DNA microarrays, it is the case that amounts of protein expression levels are unobservable.

To proceed with an inversion analysis, the DEs are related to the experimental data using a state space model:

$$y_i(t) = x_i(t) + w_i(t) \quad \text{for } t \in \mathcal{N} \text{ and } i \in \mathcal{S} \quad (6)$$

$$\frac{dx_i(t)}{dt} = f_i(\text{pa}_i(x), \psi) \quad (7)$$

where $w_i(t)$ denotes measurement error independently and identically distributed.

The processes $Y = \{y_i(t) | t \in \mathcal{N}, i \in \mathcal{S}\}$ and $X = \{x(t) | t \in \mathcal{T}\}$ follow (6) and (7) with initial state variables $x(0)$ having a certain prior distribution $x(0) \sim p(x(0))$. Bayesian inversion

analysis explores all the unknown parameters in the model—initial states $x(0)$ and kinetic parameters ψ —based on the posterior distribution $P(x(0), \psi | Y, \text{Pa})$ under which a circuit structure $\text{Pa} = \{\text{pa}_i(x)\}_{1 \leq i \leq p}$ —a collection of parent variable sets—is specified and a priori knowledge on reaction kinetics is expressed via a prior distribution $P(\psi)$. Rather than $P(x(0), \psi | Y, \text{Pa})$, in many cases, it is of more interest to evaluate the *augmented* posterior distribution $P(x(0), \psi, \text{Pa} | Y)$ where a circuit structure is unknown or unreliable. In what follows, we focus on the estimation of ψ with a given circuit structure. LiSDAS, however, has a pre-mature function for structural search of reaction circuits.

2.3 Learning reaction kinetics and relevant methods

All the unknown parameters relevant to an analysis are expressed by θ , i.e. $\theta = \{\psi, x(0)\}$ or augmented parameters involving circuit structures Pa . In usual, an analytical form of the posterior distribution $P(\theta | Y) \propto P(Y | \theta)P(\theta)$ is unavailable due to non-linearity of DEs, thereby driving a need for efficient posterior approximation techniques in solving the inversion problems—search for posterior means or modes. The recent systems biology has explored parallel developments of SMC, MCMC-driven posterior explorations and gradient descent methods. However, with a view to scaling to larger numbers of unknown parameters, efficient and effective computation remains a challenge.

Our study started from Nagasaki *et al.* (2006) that applied a simplest SMC to the stochastic search for transcriptional reaction kinetics. The use of SMC algorithm had been suggested also by Quach *et al.* (2007) whereas they actually applied the unscented Kalman filter as a variant of conventional SMC. A key notion of the SMC and MCMC-driven search—a more efficient counterpart of SMC—is that the posterior distribution is approximately evaluated on a finite number of Monte Carlo samples called *particles* such that the empirical distributions are ensured to reach the true posterior distribution as the number of particles approaches to infinity. Whenever following such an approach, however, a quite huge number of particles are needed to draw as raising the number of unknown parameters. Nagasaki *et al.* (2006) and Yoshida *et al.* (2008) reported that the practical relevance of the SMC was limited by estimations of only a few parameters.

Some early studies focused on GA (Kikuchi *et al.*, 2003; Kimura *et al.*, 2004; Koh *et al.*, 2006) that explores values of parameters by alternating crossover and mutation of initially generated seeds for the parameters. With no loss of generality, any optimization problems for which GA aims to solve can be converted to posterior mode searches in Bayesian context. The key for success in GA is whether or not an initial set of candidate parameters lies in the region close to the posterior mode. According to Higuchi (1997), the intrinsic nature of GA is essentially the same to the SMC. Hence, the low practical relevance of the SMC (Nagasaki *et al.*, 2006; Yoshida *et al.*, 2008) can be a rule of thumb for a performance of GA.

The substantial merit of Bayesian kinetic inference is in the use of prior distributions, $P(\theta)$, for conducting effective regularization in the inversion analysis. To the best of our knowledge, however, no one has addressed this fundamental issue while the range of cutting-edge *computational* technologies have appeared over the last several years. The Bayesian expert system that we present is the first attempt

to make a class of prior distributions aiming to enhance the power of previously explored class of kinetic search methods.

3 BAYESIAN EXPERTS ON KINETIC LEARNING

3.1 Dirichlet process

A prior distribution $P(\theta)$ on the unknown parameters is modelled by a DP written as

$$G \sim \text{DP}(G | \gamma, G_0) \text{ and } \theta | G \sim G$$

The distribution G that θ follows is treated as an infinite dimensional random variable distributed according to the DP— $\text{DP}(G | \gamma, G_0)$. The DP is defined by the two parameter: base measure G_0 and concentration parameter $\gamma \geq 0$. The G_0 is a distribution function defined on the support same as G . The smaller γ is, the more G looks like the G_0 , and vice versa as larger γ . A key notion of DP is in a discrete representation of probability measure; G is decomposable via the sum of the Dirac measures $\delta_{\theta_i}(\cdot)$ at the infinite set of atoms $\theta_i \sim G_0$ ($i = 1, \dots, \infty$). The explicit form is hence given as the infinite mixture $G(\theta) = \sum_{i=1}^{\infty} \pi_i \delta_{\theta_i}(\theta)$ with the mixing rates, $\pi_i \geq 0$ and $\sum_{i=1}^{\infty} \pi_i = 1$. The infinite sequence of θ_i can be made by successively conducting stick breaking construction or Chinese restaurant process (Blei *et al.*, 2005; Escobar *et al.*, 1995).

3.2 Base measure construction by a mixture of experts

Of the base measure with the Gibbsian mixture (1), the m potential functions represent a set of implicit functions $\phi_i(\theta) \approx 0$ ($i = 1, \dots, m$) that specifies regions of interest in the kinetic parameters. We here describe a simple idea on how to make such potentials with an example of the transcription regulatory models in Section 2.1.

3.2.1 Experts on maximal/minimal production rates In many applications, maximal or minimal level of each observable variable can be a helpful kinetic evidence to prescribe steady-state conditions on which each variable should reach in simulation. We illustrate some examples on how to make such prior distributions.

The transcriptional process shown in Section 2.1 with each Hill curve varying sigmoidally between 0 and 1 defines the maximal production rate as

$$\frac{d[X]}{dt} = -\psi_{rd}[X] + \psi_{tc} + \psi_b$$

where the AND gate opens. The equilibrium condition $d[X]/dt = 0$ induces the steady-state level of $[X]$ —the maximal level of the ODE solution—as the ratio of the production and dilution rates $[X]_{\max} = (\psi_{tc} + \psi_b)/\psi_{rd}$. If time course profiles are available on the mRNA, we are able to evaluate, in some ways, the maximal expression level of each mRNA, y_{\max} . Then, all possible configurations of $\{\psi_{tc}, \psi_{rd}, \psi_b\}$ such that $y_{\max} \gg (\psi_{tc} + \psi_b)/\psi_{rd}$ can be eliminated in advance by making an arbitral potential function. For instance, we will use, in later, $\phi_{\max} = (\lambda y_{\max} - (\psi_{tc} + \psi_b)/\psi_{rd})^2$ with a given discount factor $\lambda \in [1, \infty)$. Solutions with the parameters satisfying $\phi_{\max} \approx 0$ yield trajectories saturated around at y_{\max} when $\lambda = 1$. The specified $\lambda = 1$ is to express the prior belief that the observed maximal level results in an equilibrium state to which the underlying dynamic system reaches. The restriction can be weaken/strengthen by the control of the temperature parameter T . By analogy, the minimal expression level y_{\min} can also be associated

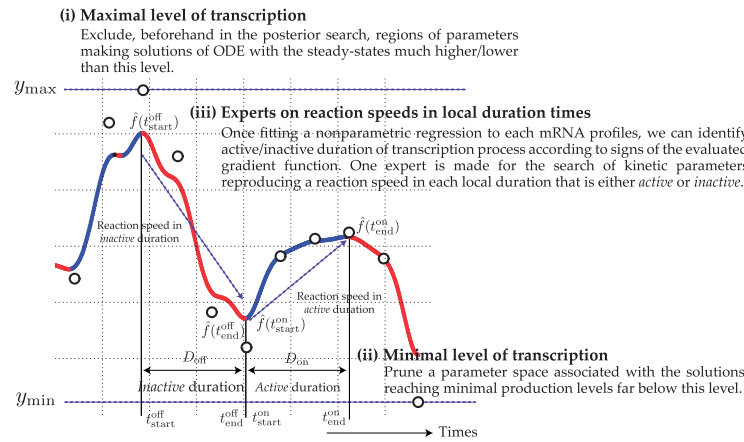


Fig. 1. Schematic expression on making Bayesian experts for (i)–(ii) maximal/minimal production levels, and (iii) local reaction speeds.

with the minimal production rate on the model with the closed AND gate, $-\psi_{rd}[X] + \psi_b = 0$, which derives a potential function $\phi_{\min} = (\lambda y_{\min} - \psi_b / \psi_{rd})^2$.

Relevant to the mass action kinetics (4) and the translation process (5) that involve protein-level reactions, we can often specify the maximal/minimal expression level of protein in data-driven ways or empirical bases. Conventional DNA microarray experiments measure transcript levels of mRNA molecules in arbitrary units, and hence any kinetic parameters in a fitted model result in unit free as well as quantities of regulatory proteins. A natural way to cope with such cases is to put an arbitrary chosen maximal/minimal value to the solutions for (4) and (5). The DE (5) attains the maximal speed of translation when the maximal level of mRNA, $[X]_{\max} = (\psi_{tc} + \psi_b) / \psi_{rd}$, is specified:

$$\frac{d[P]}{dt} = -\psi_{pd}[P] + \psi_{tl}[X]_{\max} - \psi_{cb}[P] \prod_{P_i \neq P} [P_i]^{\tau_i}_{\min}.$$

where the minimal protein concentrations $[P_i]_{\min}$ are specified arbitrary in the rest of $d-1$ reactants. Given this, a maximal level p_{\max} is then related to the equilibrium state $\psi_{tl}[X]_{\max} / (\psi_{pd} + \psi_{cb} \prod_{P_i \neq P} [P_i]_{\min})$, which makes a potential function $\phi_{\max} = (\lambda p_{\max} - \psi_{tl}[X]_{\max} / (\psi_{pd} + \psi_{cb} \prod_{P_i \neq P} [P_i]_{\min}))^2$ with a discount factor λ . By analogy, the solution of (4) describing a complex formation attains at the maximal rate of production with $[P_i]_{\max}$ ($i=1, \dots, d$). The steady-state condition induces an expert $\phi_{\max} = (\lambda c_{\max} - \psi_{cb} \prod_{i=1}^d [P_i]_{\max} / \psi_{cd})^2$ with an arbitrary specified maximum value of the complex, c_{\max} .

3.2.2 Experts on reaction speeds As well as the maximal and minimal levels of production, experimental data provide us a further clue to what temporal patterns—reaction speeds—DEs must exhibit in simulation. Suppose that a temporal profile of mRNA is observed as shown in Figure 1, showing increasing and decreasing trends alternatively. Gradients of any fitted DE models should reproduce the observed steepness during each trend.

Of many possible ways to develop such experts, we present a practically useful, simple procedure that is implemented on LiSDAS. The first task to be addressed is a decomposition of an entire time interval into *active* and *inactive* durations according to observed gene expression profiles, and values of gradients or steepness in

(A) Experts on maximal/minimal transcription levels

$$\phi_{\max} = \left(\lambda y_{\max} - \frac{\psi_{tc} + \psi_b}{\psi_{rd}} \right)^2$$

$$\phi_{\min} = \left(\lambda y_{\min} - \frac{\psi_b}{\psi_{rd}} \right)^2$$

(B) Experts on maximal/minimal speeds in transcriptional reactions

$$(\text{active duration}) \quad \phi_{\text{on}} = \left\{ \hat{f}_i(t_{\text{end}}^{\text{on}}) - \frac{\psi_{tc} + \psi_b}{\psi_{rd}} - \left(\hat{f}_i(t_{\text{start}}^{\text{on}}) - \frac{\psi_{tc} + \psi_b}{\psi_{rd}} \right) e^{-\psi_{rd} t_{\text{on}}} \right\}^2$$

$$(\text{inactive duration}) \quad \phi_{\text{off}} = \left\{ \hat{f}_i(t_{\text{end}}^{\text{off}}) - \frac{\psi_b}{\psi_{rd}} - \left(\hat{f}_i(t_{\text{start}}^{\text{off}}) - \frac{\psi_b}{\psi_{rd}} \right) e^{-\psi_{rd} t_{\text{off}}} \right\}^2$$

(C) Experts on maximal translation and mass action kinetics

$$(\text{translation}) \quad \phi_{\max} = \left(\lambda p_{\max} - \frac{\psi_{tl}[X]_{\max}}{\psi_{pd} + \psi_{cb} \prod_{P_i \neq P} [P_i]^{\tau_i}_{\min}} \right)^2$$

$$(\text{molecular binding}) \quad \phi_{\max} = \left(\lambda c_{\max} - \frac{\psi_{cb} \prod_{i=1}^d [P_i]_{\max}}{\psi_{cd}} \right)^2$$

each of the identified phases. The proposed procedure starts with fitting a non-parametric regressor, $\hat{f}_i(t) = \mu_i + w_i^T g(t) + v_i(t)$, to the i -th mRNA profile, $y_i(t)$ ($t \in \mathcal{N}$), where μ_i and $v_i(t)$ are intercept and residual, respectively. The J basis functions placed at equally-spaced grids in \mathcal{T} are aggregated to $g(t) \in \mathbb{R}^J$, and $w_i \in \mathbb{R}^J$ is a vector of coefficients. LiSDAS fits the Gaussian radial functions by using ℓ_2 regression scheme. A value of bandwidth common to all basis functions and a shrinkage parameter in the ℓ_2 regression are prescribed by users.

The fitted curve $\hat{f}_i(t) = \hat{\mu}_i + \hat{w}_i^T g(t)$ offers a gradient function over time, $\hat{w}_i^T (dg(t)/dt)$. Let D_{on} and D_{off} be *duration* times of increasing and decreasing trends that are identified as subsets of the entire time interval. Each duration keeps a same sign of the evaluated gradients during the consecutive times. The key idea here is to derive a set of experts such that each potential function reflects a reaction speed of increasing or decreasing during each of the identified local times.

The Hill curve transcription dynamics in Section 2.1 are locally linear where the concentration levels of all the transcription factors rise much above or fall much below the relevant K_i s. At the reaction speeds of transcriptional process with open/closed AND gates, we have the solutions:

$$[X] = \begin{cases} \frac{\psi_{tc} + \psi_b}{\psi_{rd}} + \left([X]_0 - \frac{\psi_{tc} + \psi_b}{\psi_{rd}} \right) e^{-\psi_{rd} t} & (\text{open gate}) \\ \frac{\psi_b}{\psi_{rd}} + \left([X]_0 - \frac{\psi_b}{\psi_{rd}} \right) e^{-\psi_{rd} t} & (\text{closed gate}) \end{cases}$$

where $[X]_0$ is an initial condition.

Denote the start and end times of each identified duration by $(t_{\text{start}}^{\text{on}}, t_{\text{end}}^{\text{on}})$ for D_{on} and $(t_{\text{start}}^{\text{off}}, t_{\text{end}}^{\text{off}})$ for D_{off} , respectively. Of interest here is to identify regions of the parameter spaces such that the simulations instantaneously reproduce the observed reaction speeds— $\hat{f}_i(t)$, respectively, for $t \in [t_{\text{start}}^{\text{on}}, t_{\text{end}}^{\text{on}}]$ or $t \in [t_{\text{start}}^{\text{off}}, t_{\text{end}}^{\text{off}}]$. Specify $\hat{f}_i(t_{\text{start}}^{\text{on}})$ and $\hat{f}_i(t_{\text{end}}^{\text{on}})$ to $[X]$ and $[X]_0$ in the former ODE solution corresponding to the active duration where t is set to the length of the duration, $t_{\text{on}} = t_{\text{end}}^{\text{on}} - t_{\text{start}}^{\text{on}}$. An expert on reaction speed is then made of the potential function:

$$\phi_{\text{on}} = \left\{ \hat{f}_i(t_{\text{end}}^{\text{on}}) - \frac{\psi_{tc} + \psi_b}{\psi_{rd}} - \left(\hat{f}_i(t_{\text{start}}^{\text{on}}) - \frac{\psi_{tc} + \psi_b}{\psi_{rd}} \right) e^{-\psi_{rd} t_{\text{on}}} \right\}^2.$$

By analogy, the implicit function relevant to the latter solution induces the expert on speed in the inactive duration with $t_{\text{off}} =$

$$t_{\text{end}}^{\text{off}} - t_{\text{start}}^{\text{off}} :$$

$$\phi_{\text{off}} = \left\{ \hat{f}_i(t_{\text{end}}^{\text{off}}) - \frac{\psi_b}{\psi_{rd}} - \left(\hat{f}_i(t_{\text{start}}^{\text{on}}) - \frac{\psi_b}{\psi_{rd}} \right) e^{-\psi_{rd} t_{\text{off}}} \right\}^2.$$

These two potential functions form elements of a Gibbsian distribution in various ways, for instance $\propto \exp(-\phi_{\text{off}} \times \phi_{\text{on}}/T)$. Alternatively, we would develop an expert based on either $\propto \exp(-\phi_{\text{off}}/T)$ or $\propto \exp(-\phi_{\text{on}}/T)$. In advance to the posterior exploration, LiSDAS generates many experts for all the identified duration times by joining/dissociating ϕ_{on} and ϕ_{off} at random.

3.2.3 Summary Figure 1 summarizes all the potential functions that we use in later and a schematic view on the idea. It is important to see that the preceding potential functions constitute, with their combination, a single potential function in (1). For instance, the potential functions relevant to each mRNA— ϕ_{max} , ϕ_{min} , ϕ_{on} and ϕ_{off} —are defined on the transcriptional parameters (ψ_{rd} , ψ_{tc} , ψ_b), while the other potentials involve only the kinetics relevant to protein regulations. LiSDAS draws at random potential functions and conventional Gaussian priors (the user can specify in the model construction file) for each set of kinetic parameters. Each ϕ_i in (1) is then constructed by aggregating the extracted component prior distribution corresponding to different types of kinetic parameters.

3.3 Computational algorithm

DP is discrete with probability one, and represented as an infinite mixture $P(\theta|\Theta_\infty) = \sum_{i=1}^{\infty} \pi_i \delta_{\theta_i}(\theta)$ conditional on infinite number of the atoms, $\Theta_\infty = \{\theta_i | i=1, \dots, \infty\}$ following G_0 . Hence, the time course data also follow the infinite mixture

$$P(Y|\Theta_\infty) = \sum_{i=1}^{\infty} \pi_i P(Y|\theta_i).$$

Of interest in the subsequent posterior analysis is to realize random sample from the posterior distribution of Θ_∞ , *a posteriori*, and also the mixing rates π_i , conditional on Y . The Bayesian model in total forms a DP mixture (DPM) with the key hierarchical structure: (i) $G \sim \text{DP}(G|\gamma, G_0)$, (ii) $\theta|G \sim G$ and (iii) $Y|\theta \sim p(Y|\theta)$. In statistical science, there have been several excellent reviews, providing technical details on DPM (Blei *et al.*, 2005; Escobar *et al.*, 1995; MacEachern *et al.*, 1998). We therefore present only a specific computational algorithm.

The procedure is computationally very straightforward. Sample from the posterior is constructed in a recursive way as a conventional sampling algorithm does (e.g. Escobar *et al.*, 1995). Suppose that, in a current *distinct* parameter set, $\Theta_k = \{\theta_i | i=1, \dots, k\}$ with the realized values being all distinct, there are n_i occurrences of each θ_i . The total sample size is $N_k = \sum_{i=1}^k n_i$. Then, the $(k+1)$ -th case is newly generated by the conditional posterior $(\theta_{k+1}|\Theta_k, Y)$ given as the mixture

$$P(\theta_{k+1}|\Theta_k, Y) \propto \sum_{j=1}^k \omega_j \delta_{\theta_j}(\theta_{k+1}) + \omega_{k+1} P(\theta_{k+1}|G_0, Y),$$

with the mixing rates given by

$$\omega_j \propto \begin{cases} \frac{n_j}{\gamma + N_k} P(Y|\theta_j) & \text{for } j=1, \dots, k, \\ \frac{\gamma}{\gamma + N_k} \int P(Y|\theta) dG_0(\theta) & \text{for } j=k+1 \end{cases}.$$

The first k component distributions consist of the Dirac measures $\delta_{\theta_j}(\theta_{k+1})$ placed at the preceding sample points, with the mixing rates involving the frequency of sampling n_j and the likelihood $P(Y|\theta_j)$. The last component represents the conditional posterior $P(\theta_{k+1}|G_0, Y) \propto P(Y|\theta_{k+1})G_0(\theta_{k+1})$, and the normalizing constant cannot be evaluated analytically. If a sample θ_{k+1} is drawn from one of the preceding samples, the occurrence rate is updated, $n_j \rightarrow n_j + 1$. Given a new θ^* coming from the $(k+1)$ -th distribution, we increment the current set such that $\Theta_{k+1} = \Theta_k \cup \{\theta^*\}$ and $n_{k+1} = 1$. Random draws having the higher likelihood $P(Y|\theta)$ appear more frequently and result in an approximation of the posterior distribution $P(\theta|Y)$ as the sampling process proceeds iteratively until convergence.

The remaining task to be addressed is to make a sample from $P(\theta_{k+1}|G_0, Y)$, and evaluation of its normalizing constant. We here provide a simple Monte Carlo technique using importance sampling (IS). One difficulty is relevant to making a sample from the Gibbsian mixture $G_0 \propto \sum_{i=1}^m \alpha_i Z_i^{-1} \exp(-T^{-1} \phi_i(\theta))$. Our approach is very straightforward: (i) conduct the second-order approximation of each potential function $\phi_i(\theta) \approx \phi_i(\bar{\theta}_i) + (\theta - \bar{\theta}_i)^T \Phi_i(\bar{\theta}_i)(\theta - \bar{\theta}_i)$ around $\bar{\theta}_i = \arg \max \phi_i(\theta)$ with $\Phi_i(\theta)$ the Hessian matrix of ϕ_i , and (ii) the proposal distribution of IS is then defined as the Gaussian distribution $Q_i(\theta) = N(\theta|\bar{\theta}_i, T\Phi_i(\bar{\theta}_i)^{-1})$ with the covariance matrix given by the inverse of Hessian matrix times T . Let $Q(\theta) = \sum_{i=1}^m \alpha_i Q_i(\theta)$ be the Gaussian mixture proposal distribution. Assuming that we have a realization of k samples from the Q and the importance weights $\beta_i = G_0(\theta_i)/Q(\theta_i)$ ($i=1, \dots, k$), then the normalization constant appearing in the mixing rate of the $(k+1)$ -th component can be approximated by $(\sum_{j=1}^k \beta_j)^{-1} \sum_{j=1}^k \beta_j P(Y|\theta_j)$. The approximated mixing rates are then proportional to

$$\hat{\omega}_j \propto \begin{cases} \frac{n_j}{(\gamma + N_k) \sum_{i=1}^k \beta_i} \beta_j P(Y|\theta_j) & \text{for } j=1, \dots, k, \\ \frac{\gamma}{(\gamma + N_k) \sum_{i=1}^k \beta_i} \sum_{i=1}^k \beta_i P(Y|\theta_i) & \text{for } j=k+1 \end{cases}.$$

One of the previously obtained k samples is replicated if a sampling process draws a component coincident to one of $\{1, \dots, k\}$. If not, we add a new case $\theta_{k+1} = \theta^*$ from Q , and then assign the likelihood $P(Y|\theta_{k+1})$ and the IS weight β_{k+1} . All the cases, whether new or old, are independently distributed according to the Q , thereby ensuring the statistical consistency of the Monte Carlo approximation of $\hat{\omega}$.

4 APPLICATION

An example in kinetic search of a circadian clock transcriptional circuit illustrates the practical relevance of the Bayesian expert systems, and we make a comparison to the non-expert approach using a conventional SMC (Nagasaki *et al.*, 2006; Nakamura *et al.*, 2009). Our model considered here is an extended version of the gene regulatory model developed in Matsuno *et al.* (2000) where among several studies used in evaluating relevance of various parameter search algorithms (Nagasaki *et al.*, 2006; Nakamura *et al.*, 2009; Yoshida *et al.*, 2008). The original model describes a reaction circuit of 12 endogenous variables, involving five mRNAs, five translated proteins and two protein complexes, which constitute an interlock of several types of positive and negative feedback loops. We here added, to the previous model, seven clock-related variables and their regulatory mechanisms, reflecting in part the recent discoveries (Baggs *et al.*, 2009; Ueda *et al.*, 2005). The newly constructed DE

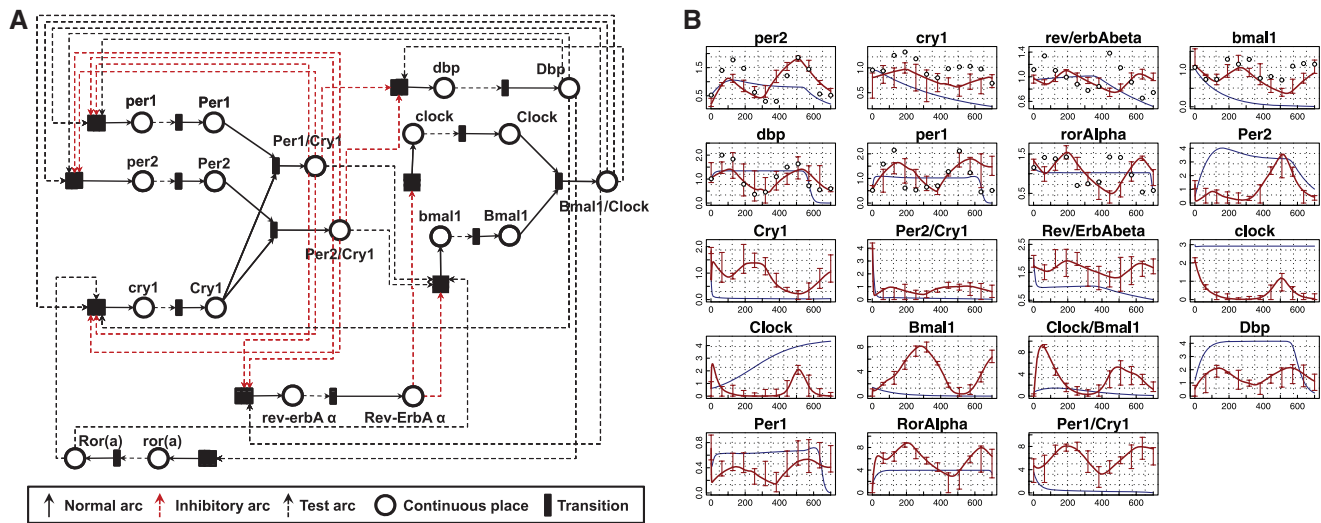


Fig. 2. Results of analysis of reaction kinetics. (A) Petri net diagram for *in silico* circadian clock transcriptional circuit involving 19 endogenous variables and 116 unknown parameters. (B) Simulation trajectories of the 19 variables with the estimated parameters (red; Bayesian expert, blue; non-expert SMC). The black circles in the panels corresponding to the seven mRNAs denote the experimental data at the 12 time points. The error bars (vertical red lines) indicate the maximum and minimum values among 20 simulated trajectories with different sets of the parameters that were estimated by repeating the Bayesian expert search independently.

model—its mathematical form is available in the Supplementary Material—contains 116 unknown parameters in total, consisting of 97 kinetic parameters and 19 initial conditions on $x_{(0)}$.

The search methods were applied to time course gene expression indices from GeneChip mouse genome microarrays (Ueda *et al.*, 2002). Of the 19 variables, temporal expression changes of the seven mRNAs were measured along 12 time points equally spaced on 44 h ([http://sirius.cdb.riken.jp/MouseSCN/MouseSCNCCG\(020707\).html](http://sirius.cdb.riken.jp/MouseSCN/MouseSCNCCG(020707).html)). A much detailed information on the specified prior distributions—values for γ , temperature and conventional Gaussian priors used in combinations—is accessible from the Supplementary web site, as well as the LiSDAS code that can reproduce the subsequent numerical analyses with a default configuration file.

Figure 2 displays simulation trajectories of the 19 endogenous variables, where the identified kinetic parameters and initial state variables were defined as the θ_i exhibiting the best fit in terms of $P(Y|\theta_i)$. The DEs were solved using the Euler discretization with 700 time steps equally spaced on the entire time interval. CPU time required for the execution of LiSDAS was ~ 40 min (Intel Core2 Quad processor, 2.66 GHz) for the particles of size $k=5 \times 10^5$. Compared with the very poor results of the non-expert SMC method using only the conventional Gaussian priors, the LiSDAS could capture the underlying circadian rhythm inherent in the observed data with much better degree of fitness. Note that the applied search procedure is almost the same as a random search. The particles were all drawn from G_0 , and then the best fit value was merely chosen. Therefore, the good performance arose from only the added Bayesian expert system. To the best of our knowledge, the methods being able to estimate such many parameters have never appeared in the related research field, while so far, our previous study (Yoshida *et al.*, 2008) and the GA of Koh *et al.* (2006) could draw reasonably good performances in analyses of a transcription model and *in*

silico signal transduction pathway, containing, 44 and 84 unknown parameters respectively.

5 CONCLUDING REMARKS

This brief article has described the new idea on data-driven parameter search for bio-pathway DE models. The key contributions lie in the novel Bayesian expert systems induced by the base measure of DP prior. The expert systems rely on the Gibbsian potential functions that reflect automatically recognized model-specific contexts—reaction speeds, equilibrium states—to the subsequent search process so as to prune insignificant parameter spaces. The ability to aid in improving efficiency has been demonstrated in the example on transcriptional regulatory models. The practical benefit was surprisingly significant; the applied simple search algorithm could draw fairly good performances only by a plug-in of the expert systems.

With a view to general versatility of the current approach, developments of more practically relevant expert systems remain a challenge yet. A biological circuit is made of several types of biochemical reactions—phosphorylation cascade-mediated signal transductions and metabolic reactions—other than the transcriptional processes we studied. In many cases, these reactions are modelled by more complex equations, being either stochastic or non-stochastic, than the Hill functions used in this article. Though it would be straightforward, even for more complex DE systems, to relate observed maximal/minimal production levels to steady-state conditions, some difficulties arise in evaluations of experts on reaction speeds with a given more complex model due to unavailability of the analytical expression for the reaction speeds. We then need to explore some numerical techniques for making expert system priors.

Making further variants in expert construction is possible, and will then be inherently context specific. In view of biochemical kinetics, it would be more helpful to incorporate, to further prior modelling, differences of time scales on each reaction speed, such as difference in molecular stability of different proteins and mRNAs. Signal transduction pathways usually change in transcription factor activities on subsecond time scales. Binding of transcription factor to its target promoter reaches equilibrium in seconds. Transcription and translation of genes take many minutes in reaching steady state. We need to explore ways of reflecting these time scales to subjective prior distributions expressing a biologically significant subspace of kinetic parameters.

Another important issue is unreliability of model. *In silico* circuits embody currently obtained knowledge on many pairs of interacting molecules. In many applications, such a circuit structure is totally unreliable because of environmental dependency and diversity of cells. Very often, simulations fail to reproduce experimental data irrelevant to which kinetic parameters are specified. The inconsistency indicates presence or absence of some regulatory mechanisms to be reflected into model reconstruction. Recent systems biology therefore has explored a structural learning of reactions circuits in data-driven ways (Cantone *et al.*, 2009; Cao *et al.*, 2008; Porreca *et al.*, 2010). LiSDAS with the procedure specification file in place can explore the most plausible circuit with the highest degree of fit as well as kinetic values by adding/deleting edges completely at random. However, with a view to larger scale models, the ability of such a low-level search method reaches a performance limit due to a huge space of possible edge configurations. Our next challenge is the development of Bayesian experts in exploring circuit structure.

Conflict of Interest: none declared.

REFERENCES

- Alon, U. (2006) *An Introduction to Systems Biology*. Chapman and Hall, London, UK.
- Baggs, J.E. *et al.* (2009) Network features of the mammalian circadian clock. *PLoS Biol.*, **7**, e1000052.
- Blei, D.M. *et al.* (2005) Variational inference for Dirichlet process mixtures. *Bayesian Anal.*, **16**, 121–144.
- Cantone, I. *et al.* (2009) A yeast synthetic network for *in vivo* assessment of reverse-engineering and modeling approaches. *Cell*, **137**, 172–181.
- Cao, J. *et al.* (2008) Estimating dynamic models for gene regulation networks. *Bioinformatics*, **24**, 1619–1624.
- Chen, K.C. *et al.* (2004) Integrative analysis of cell cycle control in budding yeast. *Mol. Biol. Cell.*, **15**, 3841–3862.
- Escobar, M.D. *et al.* (1995) Bayesian density estimation and inference using mixtures. *J. Am. Stat. Assoc.*, **90**, 577–588.
- Koh, G. *et al.* (2006) A decomposition approach to parameter estimation in pathway modeling: a case study of the Akt and MAPK pathways and their crosstalk. *Bioinformatics*, **22**, e271–e280.
- Finkensstädt, B. *et al.* (2008) Reconstruction of transcriptional dynamics from gene reporter data using differential equations. *Bioinformatics*, **24**, 2901–2907.
- Fujarewicz, K. *et al.* (2007) Adjoint systems for models of cell signaling pathways and their applications to parameter fitting. *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, **4**, 322–335.
- Higuchi, T. (1997) Monte Carlo filter using the genetic algorithm operators. *J. Stat. Comput. Simul.*, **59**, 1–23.
- Kikuchi, S. *et al.* (2003) Dynamic modeling of genetic networks using genetic algorithm and S-system. *Bioinformatics*, **19**, 643–650.
- Kimura, S. *et al.* (2004) Inference of S-system models of genetic networks using a cooperative coevolutionary algorithm. *Bioinformatics*, **21**, 1154–1163.
- MacEachern, S.N. *et al.* (1998) Estimating mixture of Dirichlet process models. *J. Comput. Graph. Stat.*, **7**, 223–238.
- Matsuno, H. *et al.* (2000) Hybrid Petri net representation of gene regulatory network. *Pac. Symp. Biocomput.*, **5**, 341–352.
- Nagasaki, M. *et al.* (2006) Genomic data assimilation for estimating hybrid functional petri net from time-course gene expression data. *Genome Inform.*, **17**, 46–61.
- Nakamura, K. *et al.* (2009) Parameter estimation of *in silico* biological pathways with particle filtering towards a petascale computing. *Pac. Symp. Biocomput.*, **14**, 227–238.
- Perrin, B.E. *et al.* (2003) Inference of gene regulatory network using dynamic Bayesian networks. *Bioinformatics*, **19**, ii138–ii148.
- Porreca, R. *et al.* (2010) Identification of genetic network dynamics with unate structure. *Bioinformatics*, **26**, 1239–1245.
- Quach, M. *et al.* (2007) Estimating parameters and hidden variables in non-linear state-space models based on ODEs for biological networks inference. *Bioinformatics*, **23**, 3209–3216.
- Ueda, H.R. *et al.* (2002) A transcription factor response element for gene expression during circadian night. *Nature*, **418**, 534–539.
- Ueda, H.R. *et al.* (2005) System-level identification of transcriptional circuits underlying mammalian circadian clocks. *Nat. Genet.*, **37**, 187–192.
- Wilkinson, D.J. (2006) *Stochastic Modelling for Systems Biology*. Chapman and Hall, London, UK.
- Yoshida, R. *et al.* (2008) Bayesian learning of biological pathways on genomic data assimilation. *Bioinformatics*, **24**, 2592–2601.