



# Bayesian Parameter Inference and Model Selection by Population Annealing in Systems Biology

Yohei Murakami\*

Department of Biophysics, Division of Biology, Graduate School of Science, Kyoto University, Kyoto, Japan

## Abstract

Parameter inference and model selection are very important for mathematical modeling in systems biology. Bayesian statistics can be used to conduct both parameter inference and model selection. Especially, the framework named approximate Bayesian computation is often used for parameter inference and model selection in systems biology. However, Monte Carlo methods need to be used to compute Bayesian posterior distributions. In addition, the posterior distributions of parameters are sometimes almost uniform or very similar to their prior distributions. In such cases, it is difficult to choose one specific value of parameter with high credibility as the representative value of the distribution. To overcome the problems, we introduced one of the population Monte Carlo algorithms, population annealing. Although population annealing is usually used in statistical mechanics, we showed that population annealing can be used to compute Bayesian posterior distributions in the approximate Bayesian computation framework. To deal with un-identifiability of the representative values of parameters, we proposed to run the simulations with the parameter ensemble sampled from the posterior distribution, named “posterior parameter ensemble”. We showed that population annealing is an efficient and convenient algorithm to generate posterior parameter ensemble. We also showed that the simulations with the posterior parameter ensemble can, not only reproduce the data used for parameter inference, but also capture and predict the data which was not used for parameter inference. Lastly, we introduced the marginal likelihood in the approximate Bayesian computation framework for Bayesian model selection. We showed that population annealing enables us to compute the marginal likelihood in the approximate Bayesian computation framework and conduct model selection depending on the Bayes factor.

**Citation:** Murakami Y (2014) Bayesian Parameter Inference and Model Selection by Population Annealing in Systems Biology. *PLoS ONE* 9(8): e104057. doi:10.1371/journal.pone.0104057

**Editor:** Maria Anisimova, Swiss Federal Institute of Technology (ETH Zurich), Switzerland

**Received:** March 24, 2014; **Accepted:** July 8, 2014; **Published:** August 4, 2014

**Copyright:** © 2014 Yohei Murakami. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability:** The author confirms that all data underlying the findings are fully available without restriction. Part of relevant data are within the paper and its Supporting Information files. Other relevant data are available from the web page of the group of Noguchi et al. (2013) *Mol. Syst. Biol. study*.

**Funding:** This work was partly supported by a Grant-in-Aid for Scientific Research on Innovative Areas from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan, by Grants for Excellent Graduate Schools, MEXT, Japan and by Grant-in-Aid for Scientific Research of the MEXT. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The author has declared that no competing interests exist.

\* Email: murakami@theory.biophys.kyoto-u.ac.jp

## Introduction

Mathematical modeling is a very useful and powerful approach in systems biology [1,2]. Mathematical models used in systems biology are often represented by ordinary or partial differential equations. These differential equations contain a number of parameters which represent the rates of biochemical reactions or amounts of components (proteins, mRNAs etc). An appropriate mathematical model together with appropriate values of parameters enables us to explain, understand and predict the target biological phenomena in system level. However, we may have a number of competing and potential mathematical models to explain the observed experimental data. In addition, concrete values of parameters in mathematical models are often not well-known in previous experimental literatures. In such cases, we need to conduct model selection and parameter inference by some sort of systematic procedures.

For parameter inference, we can use likelihood based approach as a statistical approach [3]. In addition, many optimization algorithms have already been developed and used to conduct parameter inference in systems biology [4–8]. For model selection,

AIC [9] has already used to compare a number of mathematical models in systems biology [10–12].

Alternative to those methods and approaches, Bayesian statistics enables us to conduct both model selection [11,13–19] and parameter inference [13,15,16,20,21] under the coherent philosophy. For Bayesian model selection and parameter inference in systems biology, the framework named approximate Bayesian computation (ABC) or likelihood-free computation is often used [14,15,18,22–27]. ABC is very useful when likelihood functions are analytically or computationally intractable, or it takes much cost for evaluation [14,15,23,24]. Instead of evaluating the likelihood function, ABC compares the observed data and the simulated data, and gives us the approximated posterior distribution [14,15].

However, to compute Bayesian posterior distributions, Monte Carlo methods often need to be used. As Bayesian parameter inference and model selection algorithms in the ABC framework, Markov chain Monte Carlo (MCMC) [15,22,24] and sequential Monte Carlo (SMC) [15,18,23] have already used. It is known that the efficiency of ABC-MCMC algorithm reduces when the sampler is trapped in a low probability area [23]. To overcome

the problem, ABC-SMC algorithm was developed [15,18,23]. However, even though the algorithms work successfully, the posterior distributions of parameters may be almost uniform or very similar to their prior distributions [13]. In such cases, it is difficult to choose one specific value of parameter with high credibility as the representative value of the posterior distribution.

To overcome the problems, we introduced one of the Monte Carlo algorithms, population annealing [28,29]. Population annealing is a so-called population Monte Carlo algorithm [30] as same as SMC. Although population annealing is usually used in statistical mechanics [28,31], we showed that population annealing can be used to compute Bayesian posterior distributions for parameter inference and model selection in the ABC framework. To deal with un-identifiability of the representative values of parameters, instead of choosing one specific value of parameter in the posterior distribution, we ran the simulations with the parameter ensemble sampled from the posterior distribution, named “posterior parameter ensemble”. We propose this approach is valid if our purpose of parameter inference is to reproduce or predict the system dynamics, not to estimate the correct values of parameters. We showed that population annealing is an efficient and convenient algorithm to generate the posterior parameter ensemble. In addition, we showed that the simulations with the posterior parameter ensemble can, not only reproduce the data used for parameter inference, but also capture and predict the data which was not used for parameter inference. Lastly, we introduced the marginal likelihood in the ABC framework for Bayesian model selection. We showed that population annealing enables us to compute the marginal likelihood in the ABC framework and to conduct model selection depending on the Bayes factor. The validity of our propositions was firstly judged by applying our method to the feed-forward loop network motif models [32–36], secondly to the insulin dependent AKT pathway model [47,48].

## Methods

### Bayesian parameter inference

For Bayesian parameter inference, under the given likelihood function  $f(D_{obs}|\theta, M)$  and the prior distribution of parameters  $\pi(\theta|M)$ , we try to obtain the posterior distribution of parameters  $\pi(\theta|D_{obs}, M)$ , represented as

$$\begin{aligned} \pi(\theta|D_{obs}, M) &= \frac{f(D_{obs}|\theta, M)\pi(\theta|M)}{P(D_{obs}|M)} \propto f(D_{obs}|\theta, M)\pi(\theta|M). \end{aligned}$$

$M$  is a model and fixed in parameter inference.  $\theta$  is a set of parameters (i.e. multidimensional vector) in the model  $M$ .  $D_{obs}$  is an observed experimental data. In the equation,  $P(D_{obs}|M)$  is a normalization constant, and also called as a marginal likelihood.

In the ABC framework [14,24,26], instead of the true posterior distribution, we try to obtain the augmented posterior distribution  $\pi_{ABC}(\theta, D_{sim}|D_{obs}, M)$  [24]. The augmented posterior distribution can be represented as

$$\pi_{ABC}(\theta, D_{sim}|D_{obs}, M) \propto f_w(D_{obs}|D_{sim}, \theta, M) f(D_{sim}|\theta, M) \pi(\theta|M).$$

$D_{sim}$  is the simulated data, also called as the auxiliary parameter [24].  $f_w(D_{obs}|D_{sim}, \theta, M)$  is the weighting function [24]. The weighting function is set to be large value when the observed data and the simulated data are “close”, small value when they are “distant”, and constant when they are “equal” ( $D_{obs} = D_{sim}$ ) [24].

As the weighting function, so-called indicator function is often used [14,22–24,26]. Indicator function is represented as

$$f_w(D_{obs}|D_{sim}, \theta, M) \propto I(d(D_{obs}, D_{sim}|\theta) \leq \varepsilon, M).$$

$d(D_{obs}, D_{sim}|\theta)$  is a distance measure between the observed data and the simulated data.  $\varepsilon \geq 0$  is a tolerance. The indicator function equals to 1 if the observed data and the simulated data are close ( $\leq \varepsilon$ ) and 0 if not ( $> \varepsilon$ ). If  $\varepsilon$  is small enough, the augmented posterior distribution is a good approximation of the true posterior distribution [14,24]. Thus, we can use the augmented posterior distribution alternative to the true posterior distribution for parameter inference.

### Bayesian model selection

The marginal likelihood  $P(D_{obs}|M)$  plays an important role in Bayesian model selection [11,17,37,38]. The marginal likelihood is represented as

$$P(D_{obs}|M) = \int_{\theta} f(D_{obs}|\theta, M)\pi(\theta|M)d\theta.$$

In Bayesian model selection, a model  $M$  is a variable. To compare models  $M_1$  and  $M_2$ , we can use the Bayes factor [37] represented as

$$B_{12} = \frac{P(D_{obs}|M_1)}{P(D_{obs}|M_2)}.$$

From the Bayes’ theorem, the Bayes factor is the ratio of posterior odds and prior odds as follows [37]:

$$B_{12} = \frac{P(D_{obs}|M_1)}{P(D_{obs}|M_2)} = \frac{P(M_1|D_{obs})/P(M_2|D_{obs})}{P(M_1)/P(M_2)}.$$

$P(M)$  is the prior probability of the model  $M$ .  $P(M|D_{obs})$  is the posterior probability of the model  $M$ . When the prior probabilities of the competing models are equal ( $P(M_1) = P(M_2)$ ), the Bayes factor equals to the ratio of the posterior probabilities

$$B_{12} = \frac{P(D_{obs}|M_1)}{P(D_{obs}|M_2)} = \frac{P(M_1|D_{obs})}{P(M_2|D_{obs})}.$$

Depending on the Bayes factor, we can conduct Bayesian model selection [15,17,37,38].

To conduct Bayesian model selection in the ABC framework, we need to define the marginal likelihood in the ABC framework. For definition, we start from the augmented posterior probability of a model represented as

$$\begin{aligned} P_{ABC}(M, D_{sim}|D_{obs}) &= \frac{P(M) \int_{\theta} f_w(D_{obs}|D_{sim}, \theta, M) f(D_{sim}|\theta, M) \pi(\theta|M) d\theta}{P(D_{obs})}. \end{aligned}$$

On another front, we have the relationship

$$\frac{P(M|D_{obs})}{P(M)} = \frac{P(D_{obs}|M)}{P(D_{obs})}.$$

The numerator of the right hand side of the equation corresponds to the conventional marginal likelihood. We can obtain the similar

relationship with the augmented posterior distribution as

$$\begin{aligned} & \frac{P_{ABC}(M, D_{sim}|D_{obs})}{P(M)} \\ &= \frac{\int_{\Theta} f_w(D_{obs}|D_{sim}, \theta, M) f(D_{sim}|\theta, M) \pi(\theta|M) d\theta}{P(D_{obs})}. \end{aligned}$$

Thus, we can assume the marginal likelihood in the ABC framework as

$$\int_{\Theta} f_w(D_{obs}|D_{sim}, \theta, M) f(D_{sim}|\theta, M) \pi(\theta|M) d\theta.$$

The validity of the approximation can be confirmed by setting the weighting function to the indicator function as

$$f_w(D_{obs}|D_{sim}, \theta, M) = I(D_{obs} = D_{sim}).$$

The indicator function equals to 1 if the observed data equals to the simulated data and 0 if not. By this setting, the marginal likelihood in the ABC framework is consistent with the conventional marginal likelihood as follows:

$$\begin{aligned} & \int_{\Theta} f_w(D_{obs}|D_{sim}, \theta, M) f(D_{sim}|\theta, M) \pi(\theta|M) d\theta \\ &= \int_{\Theta} I(D_{obs} = D_{sim}) f(D_{sim}|\theta, M) \pi(\theta|M) d\theta \\ &= \int_{\Theta} f(D_{obs}|\theta, M) \pi(\theta|M) d\theta \\ &= P(D_{obs}|M) \end{aligned}$$

Thus, if we set the weighting function to the indicator function as

$$f_w(D_{obs}|D_{sim}, \theta, M) \propto I(d(D_{obs}, D_{sim}|\theta) \leq \epsilon, M),$$

and  $\epsilon$  is small enough, the marginal likelihood in the ABC framework is a good approximation of the conventional marginal likelihood. This is same as the case of the augmented posterior distribution for parameter inference in the ABC framework. Therefore, the Bayes factor can be defined and calculated in the ABC framework as

$$B_{12}^{ABC} = \frac{\int_{\Theta} f_w(D_{obs}|D_{sim}, \theta, M_1) f(D_{sim}|\theta, M_1) \pi(\theta|M_1) d\theta}{\int_{\Theta} f_w(D_{obs}|D_{sim}, \theta, M_2) f(D_{sim}|\theta, M_2) \pi(\theta|M_2) d\theta}.$$

Depending on the Bayes factor  $B_{12}^{ABC}$ , we can conduct Bayesian model selection in the ABC framework.

### Population annealing: algorithm

Population annealing [28,29] is one of the population Monte Carlo algorithm [30] as same as SMC. Thus, in principle, population annealing can avoid its sampler being trapped in a low probability area as same as SMC [23]. Although population annealing is usually used in statistical mechanics to sample from a canonical distribution [28,31], we show that population annealing can be used to compute Bayesian posterior distributions.

Population annealing uses particles to approximately represent a target probability distribution. Each particle has one set of concrete values of a multidimensional variable  $x$ , which corresponds to a sample from the target distribution. For population

annealing, we define the intermediate distributions  $f_{IM}(x)$  as same as annealed importance sampling [39]. The first intermediate distribution  $f_{IM}^0(x)$  is set to the probability distribution of which sampling is easy. The last intermediate distribution  $f_{IM}^N(x)$  is set to the target distribution i.e. Bayesian posterior distribution in Bayesian approach. In population annealing, the first intermediate distribution is gradually changed to the last intermediate distribution through a number of intermediate distributions  $f_{IM}^n(x)$  ( $0 < n < N$ ) between the first one ( $n = 0$ ) and the last one ( $n = N$ ). In the case of statistical mechanics, the intermediate distributions are canonical distributions with different temperatures of the system. Thus, annealing is directly corresponds to the gradual decrease of temperature of canonical distributions. Application to Bayesian statistics mimics the process [39]. In population annealing, particles are moved and their weights are changed to follow the intermediate distribution in each annealing step. Population annealing algorithm proceeds as follows [28,29]:

**PA1.** Generate  $x^k \sim f_{IM}^0(x)$  ( $1 \leq k \leq K$ ) independently and set the initial value of weight to  $w_0^k = 1/K$ , where  $K$  is the total number of particles.

**PA2.** Repeat the following procedure from  $n = 1$  to  $n = N$  for each particle independently.

(a) Update weights by following equation

$$\log w_n^k = \log w_{n-1}^k + \log \frac{f_{IM}^n(x^k)}{f_{IM}^{n-1}(x^k)} \quad \left( w_n^k = w_{n-1}^k \frac{f_{IM}^n(x^k)}{f_{IM}^{n-1}(x^k)} \right),$$

and normalize.

(b) Update  $x^k$  independently by finite number of MCMC movements as the stationary distribution is consistent with the  $n$ -th intermediate distribution  $f_{IM}^n(x)$ .

(c) Resample particles at appropriate timing by following procedure.

For each particle, set the probability as

$$P^k = \frac{w_n^k}{\sum_{k=1}^K w_n^k}.$$

Sample new particles with replacement following the probabilities and set the weight as  $w_n^k = 1/K$ .

(d) Set  $n = n+1$  and return to (a).

In the PA algorithm, we transposed the MCMC process and the resampling process in the original PA algorithm [28]. As MCMC process in PA2 (b), we use the Metropolis-Hastings algorithm (Text S1) [40,41]. To use population annealing in the ABC framework, we need to use ABC-MCMC (See Text S1) in PA2 (b) [22,24]. To decide the appropriate timing of resampling in PA2 (c), we used effective sample size (ESS) as same as Sisson et al's study [23]. ESS is defined as

$$ESS = \frac{1}{\sum_{k=1}^K (w_n^k)^2}.$$

ESS represents the bias of particles' weights. Particles are resampled when the ESS value falls below a threshold. In this study, we set the threshold to  $K/2$ . The output of the PA algorithm is the  $K$  particles which each of them has  $x^k$  associated with weight  $w_N^k$  ( $1 \leq k \leq K$ ). These weights are set to follow the target distribution  $f_{IM}^N(x)$ . By counting  $x^k$  associated with weights  $w_N^k$ , we can obtain the target distribution  $f_{IM}^N(x)$ .

## Population annealing: application to Bayesian parameter inference

Population annealing can be used to conduct Bayesian parameter inference. In the case of parameter inference, a multidimensional variable  $x$  corresponds to a set of parameters  $\theta$  and a fixed model  $M$ . The target distribution for sampling is the posterior distribution of parameters  $\pi(\theta|D_{obs},M)$  (See Text S2 as an example based on [17,39]) or the augmented posterior distribution  $\pi_{ABC}(\theta,D_{sim}|D_{obs},M)$  in the ABC framework. In the ABC framework, we firstly set the weighting function to the indicator function. Then we can define the intermediate distribution as follows:

$$f_{IM}^n(\theta) \propto I(d(D_{obs},D_{sim}|\theta) \leq \varepsilon_n, M) f(D_{sim}|\theta, M) \pi(\theta|M) \\ (\varepsilon_0 > \dots > \varepsilon_n > \dots > \varepsilon_N)$$

As the tolerance  $\varepsilon_n$  decreases, the intermediate distribution gradually changes from the first intermediate distribution ( $\varepsilon = \varepsilon_0$ ) to the last intermediate distribution ( $\varepsilon = \varepsilon_N$ ) corresponding to the augmented posterior distribution of parameters as the similar manner in ABC-SMC [15,18]. The concrete schedule of  $\varepsilon_n$  is set depending on the problem.

## Population annealing: application to Bayesian model selection

Population annealing can be used to conduct Bayesian model selection in the ABC framework. We firstly set the weighting function to the indicator function. Under this setting, with a large number  $K$ , the marginal likelihood in the ABC framework can be approximated as follows:

$$\int_{\theta} f_w(D_{obs}|D_{sim},\theta,M) f(D_{sim}|\theta,M) \pi(\theta|M) d\theta \\ = \int_{\theta} I(d(D_{obs},D_{sim}|\theta) \leq \varepsilon, M) f(D_{sim}|\theta,M) \pi(\theta|M) d\theta \\ = E_{f(D_{sim}|\theta,M)\pi(\theta|M)} [I(d(D_{obs},D_{sim}|\theta) \leq \varepsilon, M)] \\ \approx \frac{1}{K} \sum_{k=1}^K I(d(D_{obs},D_{sim}|\theta,k) \leq \varepsilon, M)$$

$E[\cdot]$  indicates an expectation value.  $D_{sim}|\theta,k$  is a sample from  $f(D_{sim}|\theta,M)\pi(\theta|M)$ . Thus, under a fixed total number of particles  $K$ , the number of particles satisfying the condition  $d(D_{obs},D_{sim}) \leq \varepsilon$  is proportional to the marginal likelihood in the ABC framework. Population annealing enables us to count the number of particles satisfying the condition. Because we use the indicator function, each particle has a common value of weight (non-zero weight) or zero weight in population annealing. Thus, starting with a fixed total number of particles  $K$ , the number of particles which have non-zero weight at the end of population annealing is proportional to the marginal likelihood in the ABC framework. This is because the particles at the end of the algorithm follow the last intermediate distribution corresponding to the target distribution

$$f_{IM}^N(\theta) \propto \pi_{ABC}(\theta,D_{sim}|D_{obs},M) \propto I(d(D_{obs},D_{sim}) \leq \varepsilon_N, M) \\ f(D_{sim}|\theta,M) \pi(\theta|M).$$

However, even though starting with a fixed total number of particles, resampling processes recover the number of non-zero weight particles to the total number of particles  $K$ . Thus, each time

of resampling, the number of non-zero weight particles before resampling  $k_{resample}$  should be memorized. By multiplying  $k_{resample}/K$  to the number of particles at the end of the algorithm, we can estimate the marginal likelihood in the ABC framework. In this study, we adopted this solution. This solution is based on the similar idea of the ‘‘ABC-SMC approximation of the marginal likelihood’’ [18]. Another solution against resampling is to eliminate the resampling process from population annealing. In this study, we did not adopt that solution. This is because the resampling process is considered to be important to avoid the fluctuation of the weight factor and conduct stable calculations [28].

The computation of the number of particles satisfying the condition  $d(D_{obs},D_{sim}) \leq \varepsilon$  can be also done by ABC rejection sampler (Text S1) [15,23,24,42]. Fixing the total number of sampling trials, the number of accepted particles in ABC rejection sampler is proportional to the marginal likelihood in the ABC framework. However, it is known that the acceptance rate often gets lower in the case that the prior distribution is very different from the posterior distribution [15]. We can avoid the problem by population annealing, because the intermediate distributions gradually changes from the prior distribution to the posterior distribution in population annealing.

## Simulations with the posterior parameter ensemble

In this study, instead of sampling a number of parameters from the posterior distribution, we ran the simulations with all the output particles of population annealing. This is fundamentally same as the simulations with the posterior parameter samples [11,13]. In population annealing, each particle at the end of the algorithm (output particles of population annealing) has concrete values of parameters  $\theta^k$  associated with weight  $w_N^k$  ( $1 \leq k \leq K$ ). These weights are set to follow the target posterior distribution. Thus, the simulations with  $\theta^k$  weighted by  $w_N^k$  for all the  $K$  particles correspond to the simulations with the parameter ensemble sampled from the posterior distribution. Practically, this can be done with use of the same particles used to obtain the posterior distribution of parameters. In this manner, population annealing is a convenient algorithm to generate the posterior parameter ensemble and conduct the ensemble simulation.

## Results

### Flow of the tests

In this section, we explain the flow of the tests which start from the next section. To test the validity of our method, we applied our method to the two test problems. In the first test, we introduced two mathematical models of the feed-forward loop network motif [32–36] in the next ‘‘Mathematical models and simulation setting of the feed-forward loop network motif’’ section. Then, we artificially generated the observed data from one of the models with the given answer values of parameters. Next, in the ‘‘Parameter inference, reproduction and prediction of the dynamics of the feed-forward loop models’’ section, we firstly tried to estimate the answer values of parameters by computing the posterior distribution of parameters. Secondly, we tested whether the simulations with posterior parameter ensemble can reproduce the data used for parameter inference. In addition, we tested whether the simulations with posterior parameter ensemble can correctly predict the newly generated data which was not used for parameter inference. Next, in the ‘‘Model selection between the feed-forward loop models’’ section, we tested whether the true model that the observed data was generated from was correctly selected or not by computing the Bayes factor. Lastly, in the

remaining sections, we showed the results of the second test. In the second test, we used the insulin dependent AKT pathway model [47,48]. In the test, we tried to reproduce the experimental time-series data of phosphorylated AKT. We also conducted Bayesian model selection between the wild type AKT pathway model and the mutant AKT pathway model. Importantly, we used the open experimental data in the second test.

## Mathematical models and simulation setting of the feed-forward loop network motif

**Mathematical models of the feed-forward loop network motif.** For the first test, we used the mathematical models of the feed-forward loop network motif [32–36]. Network motifs are building blocks of transcription network found in diverse organisms [34]. One of the significant network motifs is the feed-forward loop (FFL), which consists of three components,  $X$ ,  $Y$  and  $Z$ . There are totally eight possible structures of FFL, which can be divided into coherent FFL or incoherent FFL [33,34]. In addition, there are AND logic and OR logic for the activation of  $Z$  by  $X$  and  $Y$  [33,34]. In the current test study, we used one of the coherent FFL and one of the incoherent FFL with AND logic (Figure 1.A, B). Their dynamics can be represented by the following ordinary differential equations [33].

*Coherent FFL*

$$\frac{dY}{dt} = \beta_Y \frac{\left(\frac{X}{K_{XY}}\right)^h}{1 + \left(\frac{X}{K_{XY}}\right)^h} - \alpha_Y Y$$

$$\frac{dZ}{dt} = \beta_Z \frac{\left(\frac{X}{K_{XZ}}\right)^h \left(\frac{Y}{K_{YZ}}\right)^h}{1 + \left(\frac{X}{K_{XZ}}\right)^h + \left(\frac{Y}{K_{YZ}}\right)^h} - \alpha_Z Z$$

*Incoherent FFL*

$$\frac{dY}{dt} = \beta_Y \frac{\left(\frac{X}{K_{XY}}\right)^h}{1 + \left(\frac{X}{K_{XY}}\right)^h} - \alpha_Y Y$$

$$\frac{dZ}{dt} = \beta_Z \frac{\left(\frac{X}{K_{XZ}}\right)^h}{1 + \left(\frac{X}{K_{XZ}}\right)^h} \frac{1}{1 + \left(\frac{Y}{K_{YZ}}\right)^h} - \alpha_Z Z$$

In the equations,  $X$  is an input signal, set to a step function or a pulse in the test.  $Y$  and  $Z$  are variables. Each model contains eight parameters,  $\alpha_Y$ ,  $\alpha_Z$ ,  $\beta_Y$ ,  $\beta_Z$ ,  $K_{XY}$ ,  $K_{XZ}$ ,  $K_{YZ}$  and  $h$ . In this test study, we did not specifically define the units of these parameters and time for simplicity.

**Generation of the artificial observed data.** To use in the test, we artificially generated the observed data from the coherent FFL model (Figure 1.A). To generate the data, we firstly set the answer values of parameters  $\theta_{answer} = (\alpha_Y, \alpha_Z, \beta_Y, \beta_Z, K_{XY}, K_{XZ}, K_{YZ}) = (1, 1, 1, 1, 0.1, 0.1, 0.5)$ . In parameter inference, we tried to

infer these seven values. In model selection, these seven parameters were free parameters. Remaining Hill coefficient was set to  $h = 2$ . This value was always given and fixed in this test study. With these answer values and the fixed value of Hill coefficient, the coherent FFL and the incoherent FFL models showed distinct dynamics of  $Z$  in response to the step stimulation of  $X$  (Figure 1.A, B). Secondly, we added Gaussian noise with mean 0 and variance 0.01 as observation noise to the time-series data of  $Z$  in Figure 1.A. The generated data was represented as  $D_{obs} = \{Z^t_{obs}, t = 1 \sim 10\}$  (red points in Figure 1.C, concrete values are shown in Table S1). The observed data  $D_{obs}$  was used for all the computations of parameter inference and model selection in the first test with FFL models.

**Prior distribution and weighting function.** For Bayesian parameter inference and model selection, we set the prior distribution of each parameter to independently follow the uniform distribution on a common logarithmic scale. Logarithmic scale is often used in systems biology [13,14,43,44]. The upper-bound and the lower-bound of each uniform distribution were set to 10-folds value and 1/10-folds value of each answer value of parameter respectively i.e.  $\log_{10}\alpha_Y \sim U[0.1,10]$ ,  $\log_{10}\alpha_Z \sim U[0.1,10]$ ,  $\log_{10}\beta_Y \sim U[0.1,10]$ ,  $\log_{10}\beta_Z \sim U[0.1,10]$ ,  $\log_{10}K_{XY} \sim U[0.01,1]$ ,  $\log_{10}K_{XZ} \sim U[0.01,1]$ ,  $\log_{10}K_{YZ} \sim U[0.05,5]$ . Here,  $U$  represents the uniform distribution.

In this test study, we set the weighting function to the indicator function [14,22–24,26] represented as

$$f_w(D_{obs}|D_{sim}, \theta, M) \propto I(d(D_{obs}, D_{sim|\theta}) \leq \varepsilon, M),$$

and set the distance  $d(D_{obs}, D_{sim|\theta})$  as

$$d(D_{obs}, D_{sim|\theta}) = \sum_{t=1}^{10} (Z^t_{obs} - Z^t_{sim|\theta})^2.$$

$D_{sim|\theta} = \{Z^t_{sim|\theta}, t = 1 \sim 10\}$  is a simulated time-series data of  $Z$  with  $\theta$ .

**Numerical simulation.** The total number of particles in population annealing was set to  $K = 100000$ . As the default annealing schedule, the tolerance was gradually decreased as follows:

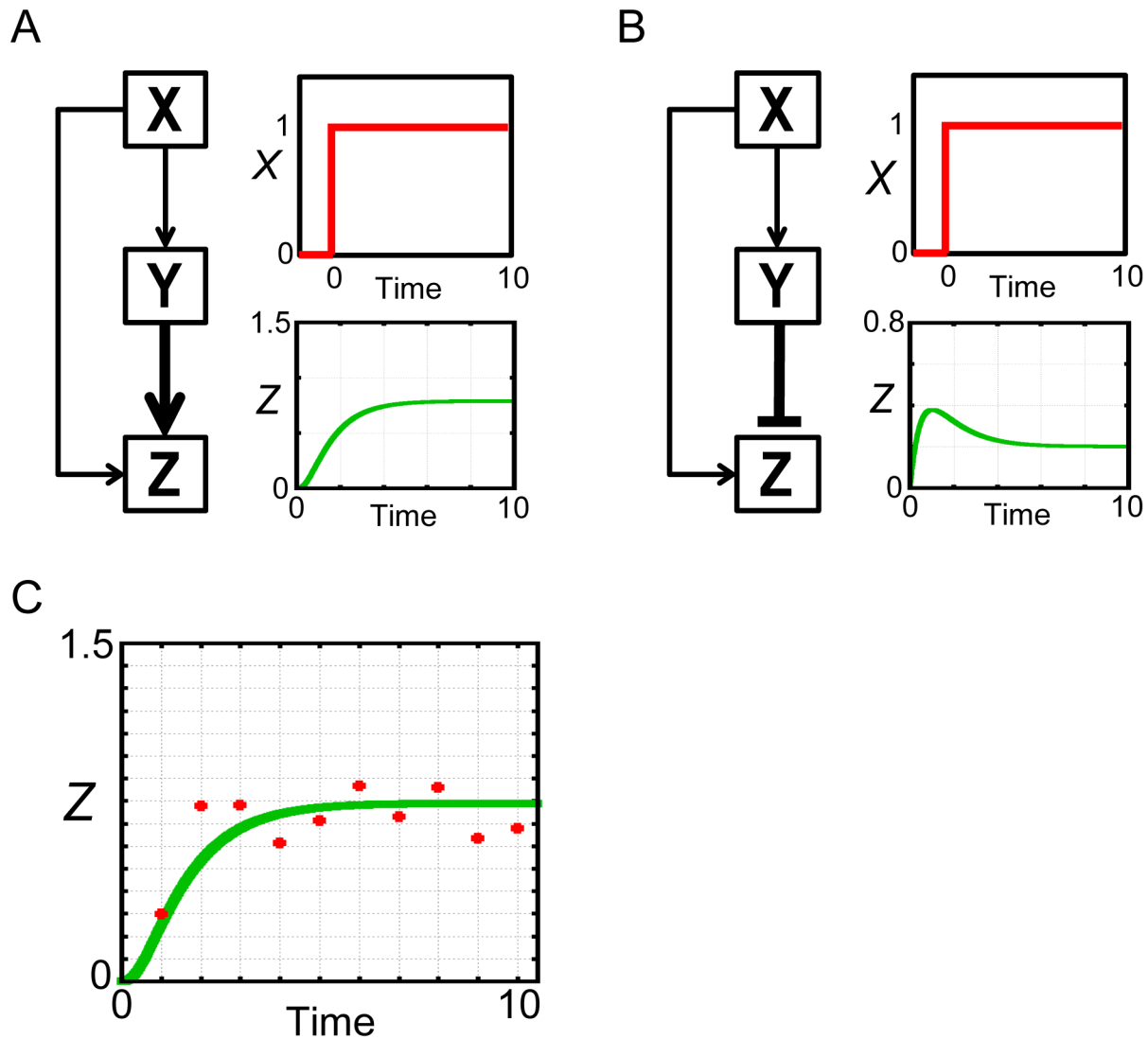
$$(\varepsilon_0, \varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5) = (\infty, 1, 0.5, 0.25, 0.2, 0.15)$$

The proposal distribution of ABC-MCMC in population annealing was set to the uniform distribution on a common logarithmic scale. In concrete terms, at each step of ABC-MCMC, one of the parameters was randomly chosen, and the uniform random number between  $-0.25$  to  $0.25$  was added on a common logarithmic scale. For each particle, ABC-MCMC movements in population annealing were set to 7 steps. This is the number of the inferred parameters.

For time-series calculations, the ordinary differential equations were numerically solved by the fourth-order Runge-Kutta method with a time step of 0.01. Both of the initial amounts of  $Y$  and  $Z$  were set to 0 in all the calculations.  $X$  was set to the step function ( $X = 1$  during the entire simulation) or the pulse ( $X = 1$  until time = 5, then  $X = 0$ ).

**Parameter inference, reproduction and prediction of the dynamics of the feed-forward loop models**

**Coherent FFL model.** Firstly, we conducted parameter inference of the coherent FFL model by population annealing.

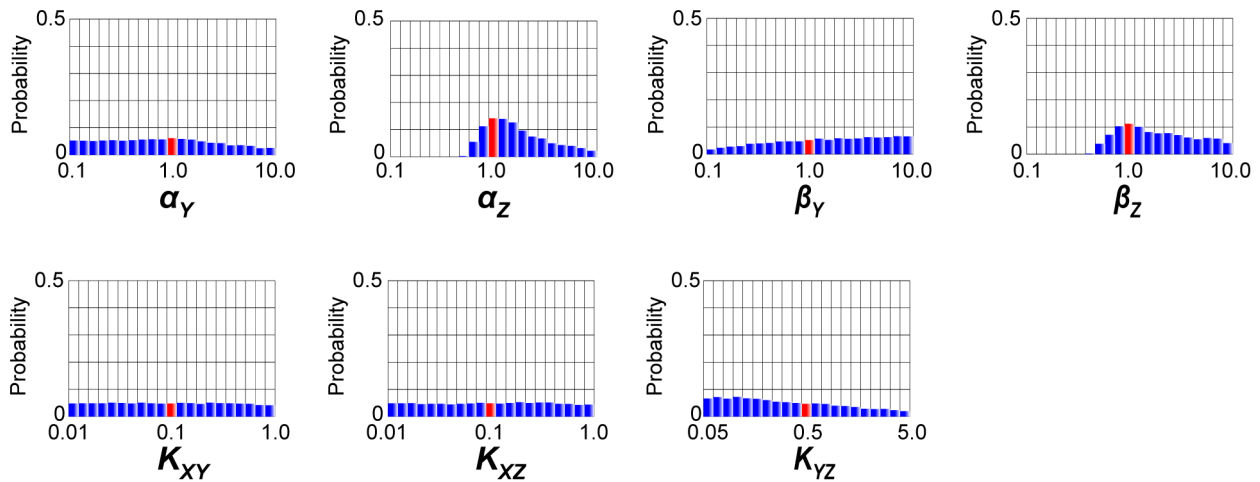


**Figure 1. Feed-forward loop network motif models.** (A) Structure and dynamics of the coherent FFL model in response to the step stimulation of  $X$ . (B) Structure and dynamics of the incoherent FFL model in response to the step stimulation of  $X$ . (C) Generated observed data. The observed data (red points) was generated from the coherent FFL model by adding observation noise into the green-colored trajectory. The green-colored trajectory is same as that of in (A).  
doi:10.1371/journal.pone.0104057.g001

The computed joint posterior distribution of the inferred parameters was marginalized and shown in Figure 2. In each marginal distribution, the red-colored class corresponds to the answer value of each parameter. In Figure 2,  $\alpha_Z$  and  $\beta_Z$  seemed to be inferred well to some extent compared to other parameters. However, marginal distributions of other parameters were almost similar to the uniform distributions, which are same as the prior distributions in this study. If our purpose of parameter inference is the estimation of the correct values of unknown parameters with high credibility, we failed in accomplishing our purpose in this case.

However, if our purpose is the reproduction or the prediction of the system dynamics, instead of choosing one set of representative values of parameters, we can run the simulations with the posterior parameter ensemble. In this study, we ran the simulations with use of all the 100000 output particles (parameter sets) computed by population annealing. The simulations with posterior parameter ensemble give the probability density consists of the simulated

trajectories. Reproductions of the observed data used for parameter inference were shown in Figure 3. We ran the simulations in response to the step stimulation of  $X$  (Figure 3.A) as same as the generation process of the observed data. In Figure 3.B, the area of the probability density consists of the simulated trajectories (blue-colored area) could capture the observed data used for parameter inference (red points). When the annealing schedule was changed to the smaller tolerances ( $\epsilon_0, \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_5$ ) = ( $\infty, 1, 0.5, 0.25, 0.15, 0.11$ ), the area of the probability density got narrower (Figure 3.C). When the annealing schedule was changed to the larger tolerances ( $\epsilon_0, \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_5$ ) = ( $\infty, 2, 1, 0.75, 0.5, 0.25$ ), the area of the probability density got broader and captured all the red points (Figure 3.D). As shown, the area of the probability density differs depending on the annealing schedules. This result is unsurprising because the last  $\epsilon$  values strongly restrict the acceptable trajectories to the observed data. Thus, we are recommended to try a number of annealing schedules to check the influence of the schedules on the simulated

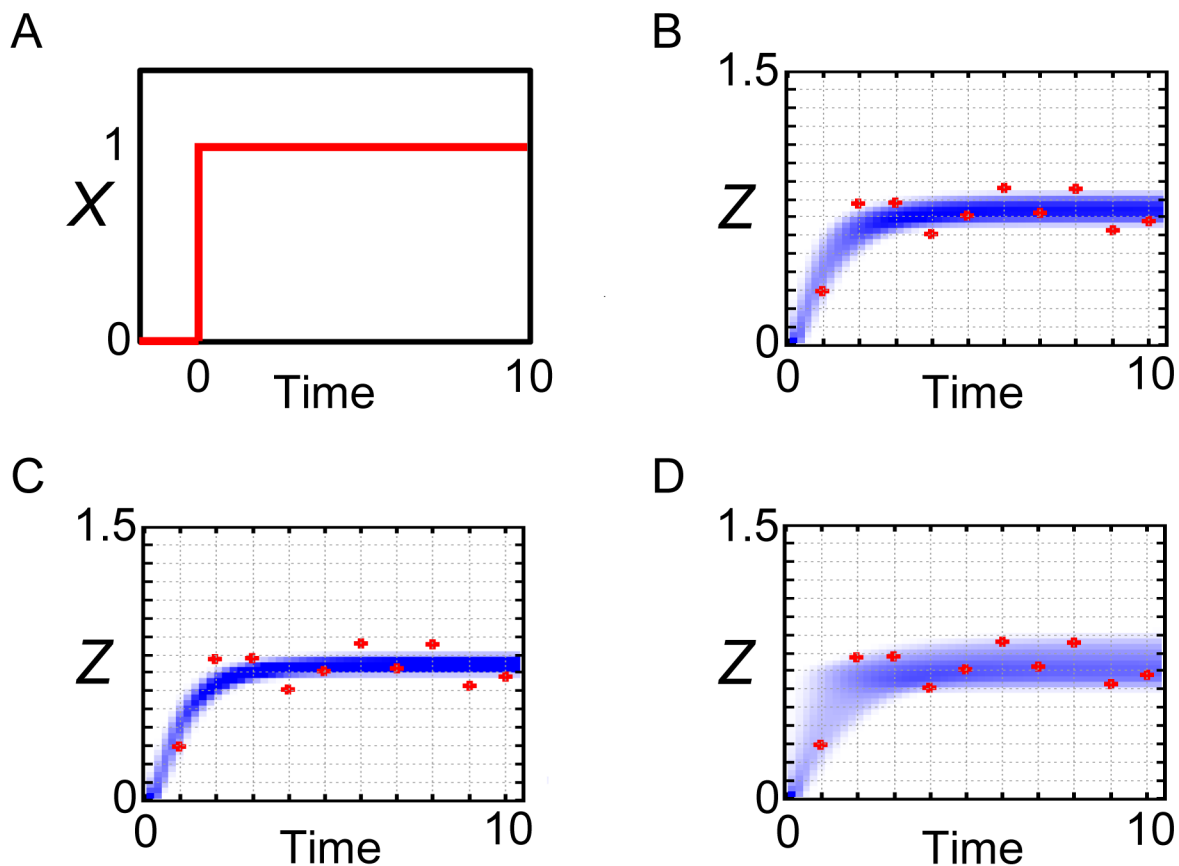


**Figure 2. Marginal probability distributions of the parameters in the coherent FFL model.** The joint probability distribution approximated by 100000 particles was marginalized for each parameter. Red-colored classes represent the “answer” value of each parameter. doi:10.1371/journal.pone.0104057.g002

data. In either case, the posterior parameter ensemble could reproduce the observed data well.

Furthermore, we examined whether the simulations with the posterior parameter ensemble can capture the data which was not

used for parameter inference. We newly generated the observed data of  $Z$  (red points in Figure 4, concrete values are shown in Table S2) in response to the pulse stimulation of  $X$  (Figure 4.A) adding the same Gaussian noise in Figure 1.C. Then we ran the



**Figure 3. Reproduction of the observed data by the coherent FFL model.** Simulations with the posterior parameter ensemble of the coherent FFL model in response to the step stimulation of  $X$ . (A) Input step stimulation of  $X$ . (B) Annealing schedule:  $(\epsilon_0, \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_5) = (\infty, 1, 0.5, 0.25, 0.2, 0.15)$ . (C) Annealing schedule:  $(\epsilon_0, \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_5) = (\infty, 1, 0.5, 0.25, 0.15, 0.11)$ . (D) Annealing schedule:  $(\epsilon_0, \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_5) = (\infty, 2, 1, 0.75, 0.5, 0.25)$ . Blue-colored area is the probability density consists of the simulated trajectories. Red points are the observed data. doi:10.1371/journal.pone.0104057.g003

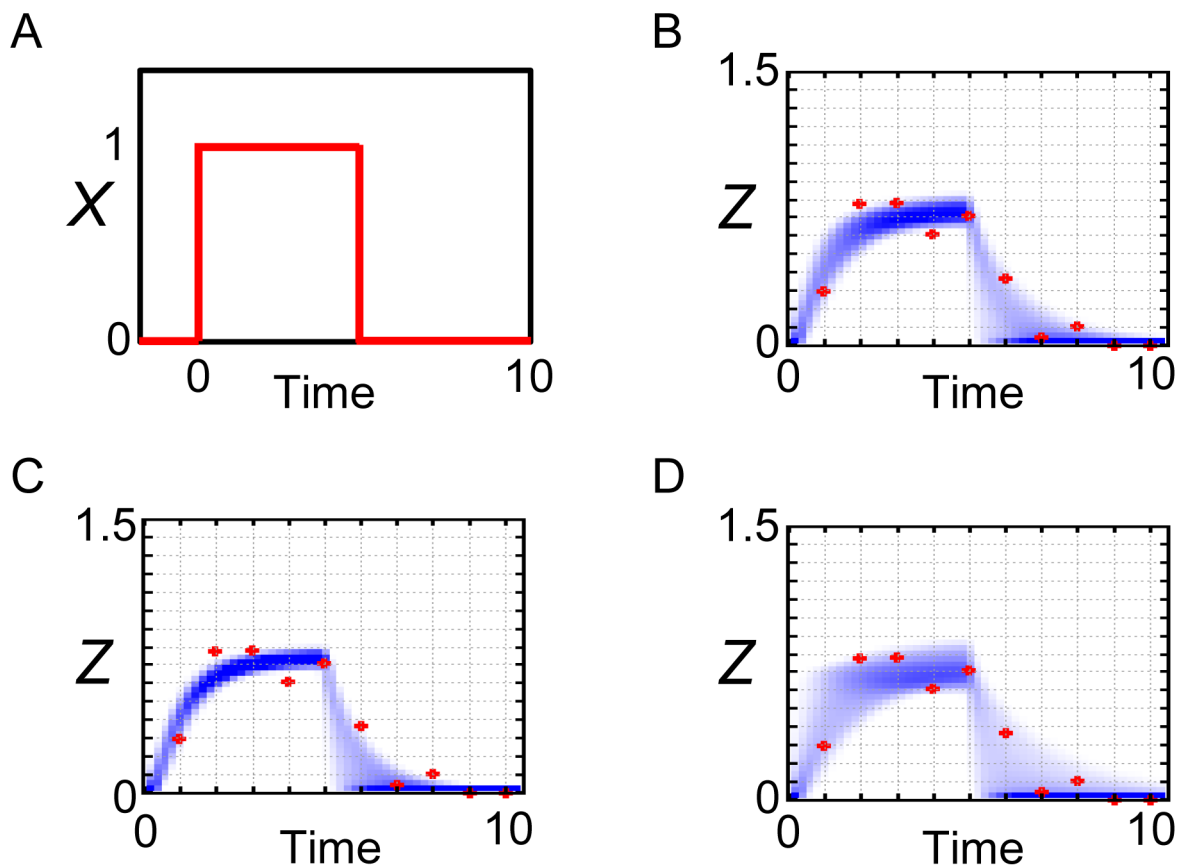
simulations in response to the pulse stimulation of  $X$  with use of all the 100000 particles which were the output of population annealing. In Figure 4.B, the probability density consists of the simulated trajectories could capture and predict the observed data which was not used for parameter inference. In addition, when the annealing schedule was changed to the smaller tolerances  $(\varepsilon_0, \varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5) = (\infty, 1, 0.5, 0.25, 0.15, 0.11)$  (Figure 4.C) or the larger tolerances  $(\varepsilon_0, \varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5) = (\infty, 2, 1, 0.75, 0.5, 0.25)$  (Figure 4.D), the areas of the probability densities could capture and predict the observed data with different rigors.

As a whole, the simulations with the posterior parameter ensemble could, not only reproduce the data used for parameter inference, but also capture and predict the data which was not used for parameter inference. Because the posterior parameter ensemble corresponds to the output particles of population annealing, population annealing is an efficient and convenient algorithm for the simulations with posterior parameter ensemble.

**Incoherent FFL model.** Next, we conducted parameter inference of the incoherent FFL model by population annealing. Note that parameter inference was conducted with the observed data generated from the coherent FFL model in Figure 1.C, not from the incoherent FFL model. The joint posterior distribution of the inferred parameters was marginalized and shown in Figure 5. We can see the distinct tendencies of marginal distributions between the coherent FFL model and the incoherent FFL model. For example, the marginal distribution of  $\alpha_Z$  had a tail in smaller

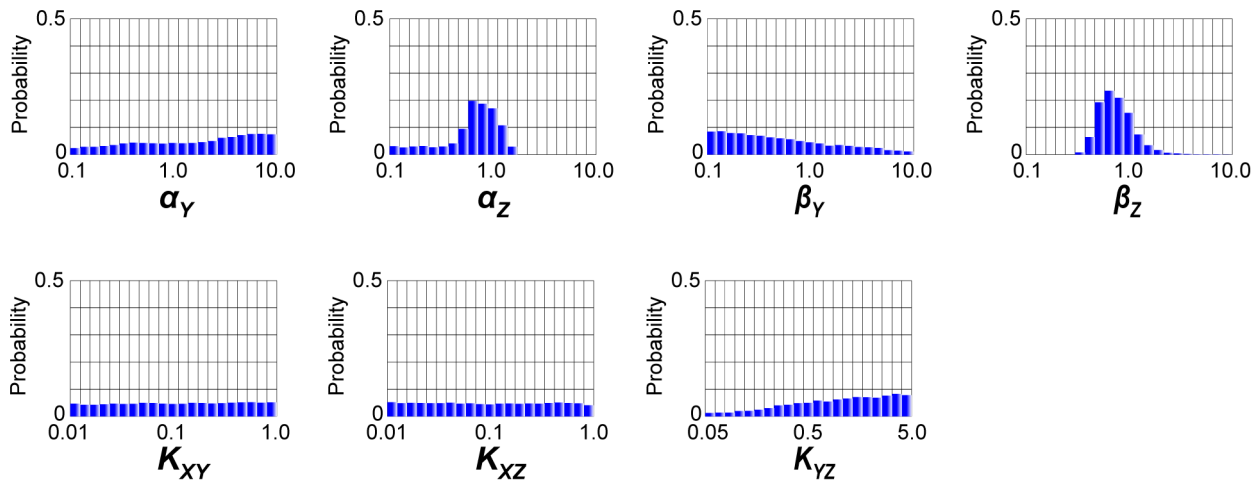
values. This is opposite to that of the coherent FFL (Figure 2).  $\beta_Y$  and  $K_{YZ}$  also showed the opposite tendency between the incoherent FFL model and the coherent FFL model (Figure 2). These differences were considered to come from the difference of the interaction from  $Y$  to  $Z$  (Figure 1.A, B). Although there were these kinds of small differences, most of the marginal posterior distributions of the incoherent FFL model were similar to the uniform distributions. This result is similar to the result of the coherent FFL model.

In addition, as shown in Figure 6.B, C and D, the simulations with the posterior parameter ensemble of the incoherent FFL model in response to the step stimulation of  $X$  (Figure 6.A) showed similar results to those of the coherent FFL model (Figure 3). We again note that the observed data used for parameter inference was generated from the coherent FFL model, not from the incoherent FFL model. This result indicates that, by setting the values of parameters suitably, even the false model (incoherent FFL model) can reproduce the observed data with comparable level to the true model (coherent FFL model). This result also emphasizes the importance of the concrete values of kinetic parameters for the dynamics of the system [45], not only the network structures. One attention point is the case of the annealing schedule  $(\varepsilon_0, \varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5) = (\infty, 1, 0.5, 0.25, 0.15, 0.11)$  (Figure 6.C). In this case, the probability density became slightly parabolic. This might be a kind of over-fitting to the data.



**Figure 4. Prediction of the observed data by the coherent FFL model.** Simulations with the posterior parameter ensemble of the coherent FFL model in response to the pulse stimulation of  $X$ . (A) Input pulse stimulation of  $X$ . (B) Annealing schedule:  $(\varepsilon_0, \varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5) = (\infty, 1, 0.5, 0.25, 0.2, 0.15)$ . (C) Annealing schedule:  $(\varepsilon_0, \varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5) = (\infty, 1, 0.5, 0.25, 0.15, 0.11)$ . (D) Annealing schedule:  $(\varepsilon_0, \varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5) = (\infty, 2, 1, 0.75, 0.5, 0.25)$ . Blue-colored area is the probability density consists of the simulated trajectories. Red points are the observed data. doi:10.1371/journal.pone.0104057.g004

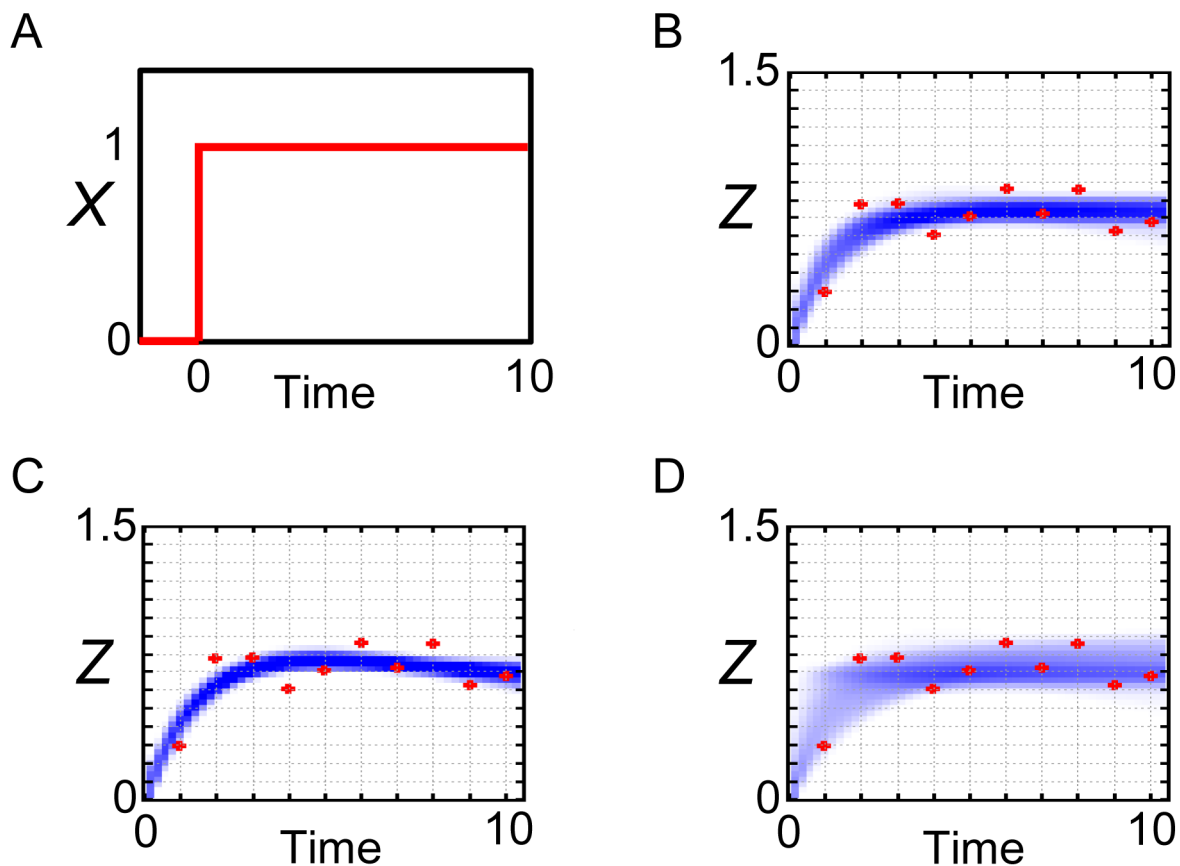




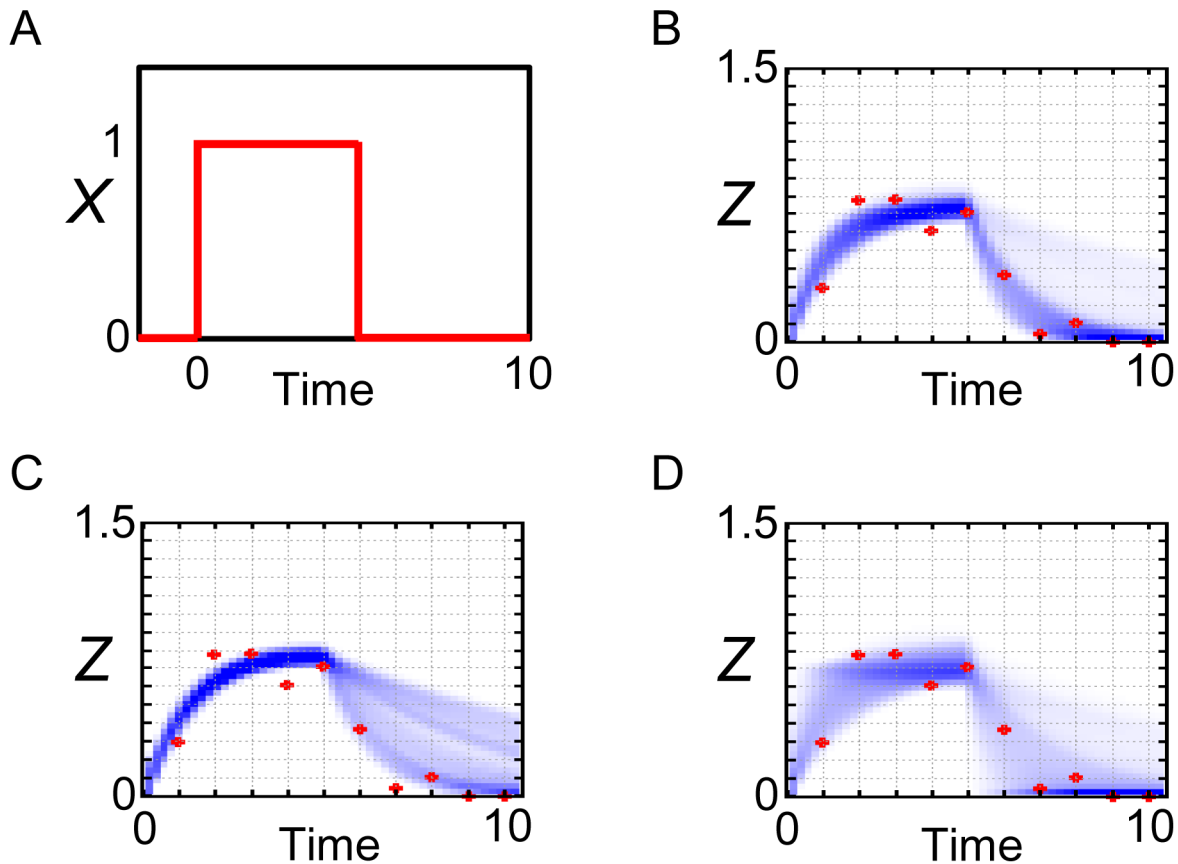
**Figure 5. Marginal probability distributions of the parameters in the incoherent FFL model.** The joint probability distribution approximated by 100000 particles was marginalized for each parameter. doi:10.1371/journal.pone.0104057.g005

However, the prediction of  $Z$  dynamics in response to the pulse stimulation of  $X$  (Figure 7.A) did not succeed well (Figure 7.B, C and D). After the decrease of  $X$  at time = 5, part of the trajectories of  $Z$  could not capture the observed data well. Thus, if we have the

observed data in response to the pulse stimulation of  $X$ , we might be able to naively select the true model (coherent FFL model) in a visual way. However, it is not always possible to obtain the



**Figure 6. Reproduction of the observed data by the incoherent FFL model.** Simulations with the posterior parameter ensemble of the incoherent FFL model in response to the step stimulation of  $X$ . (A) Input step stimulation of  $X$ . (B) Annealing schedule:  $(\epsilon_0, \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_5) = (\infty, 1, 0.5, 0.25, 0.2, 0.15)$ . (C) Annealing schedule:  $(\epsilon_0, \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_5) = (\infty, 1, 0.5, 0.25, 0.15, 0.11)$ . (D) Annealing schedule:  $(\epsilon_0, \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_5) = (\infty, 2, 1, 0.75, 0.5, 0.25)$ . Blue-colored area is the probability density consists of the simulated trajectories. Red points are the observed data. doi:10.1371/journal.pone.0104057.g006



**Figure 7. Prediction of the observed data by the incoherent FFL model.** Simulations with the posterior parameter ensemble of the incoherent FFL model in response to the pulse stimulation of  $X$ . (A) Input pulse stimulation of  $X$ . (B) Annealing schedule:  $(\varepsilon_0, \varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5) = (\infty, 1, 0.5, 0.25, 0.2, 0.15)$ . (C) Annealing schedule:  $(\varepsilon_0, \varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5) = (\infty, 1, 0.5, 0.25, 0.15, 0.11)$ . (D) Annealing schedule:  $(\varepsilon_0, \varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5) = (\infty, 2, 1, 0.75, 0.5, 0.25)$ . Blue-colored area is the probability density consists of the simulated trajectories. Red points are the observed data. doi:10.1371/journal.pone.0104057.g007

convenient data for model selection. To deal with this kind of problem, we can conduct Bayesian model selection.

#### Model selection between the feed-forward loop models

Next, we conducted Bayesian model selection, comparing the coherent FFL model with the incoherent FFL model. To conduct Bayesian model selection, we computed the Bayes factor in the ABC framework represented as

$$B_{CI}^{ABC} = \frac{\int_{\theta} I(d(D_{obs}, D_{sim}|\theta) \leq \varepsilon, M_{coherent}) f(D_{sim}|\theta, M_{coherent}) \pi(\theta|M_{coherent}) d\theta}{\int_{\theta} I(d(D_{obs}, D_{sim}|\theta) \leq \varepsilon, M_{incoherent}) f(D_{sim}|\theta, M_{incoherent}) \pi(\theta|M_{incoherent}) d\theta}.$$

Larger  $B_{CI}^{ABC}$  indicates that the coherent FFL model is selected with stronger evidence against the incoherent FFL model. In the current test study, model selection was conducted with the observed data generated from the coherent FFL model (Figure 1.C). Thus,  $B_{CI}^{ABC}$  should be a large value, at least larger than 1 to propose the validity of Bayesian model selection by population annealing.

Computation of  $B_{CI}^{ABC}$  was done under different annealing schedules in population annealing. Annealing schedules were set to same as those in parameter inference. For each annealing schedule, the mean and the standard deviation of 10 independent computations of  $B_{CI}^{ABC}$  were shown in Table 1. In Table 1, as the

value of the last  $\varepsilon$  got smaller,  $B_{CI}^{ABC}$  got larger. This result indicates that, as the models need to reproduce the observed data more rigorously, the coherent FFL model is selected with stronger evidence against the incoherent FFL model. In addition,  $B_{CI}^{ABC}$  was always larger than 1. This result is consistent with the fact that the observed data used for model selection was generated from the coherent FFL model, and the prediction that the coherent FFL model must be selected in this case. Thus, Bayesian model selection by population annealing is valid.

We also conducted Bayesian model selection by ABC rejection sampler (Text S1). For comparison, the value of  $\varepsilon$  in ABC rejection sampler was set to the same value of the last  $\varepsilon$  of population annealing. The total number of sampling trials were set to 100000, same as the number of particles in population annealing. As shown in Table 1, the means of Bayes factors computed by ABC rejection sampler were similar to those computed by population annealing. In the case of large or middle values of  $\varepsilon$ , the standard deviations of Bayes factors computed by ABC rejection sampler were also similar to those of computed by population annealing. However, in the case of small value of  $\varepsilon$ , the standard deviation of the Bayes factor computed by ABC rejection sampler was larger than that of computed by population annealing. Small standard deviation indicates a stable computational result. Thus, these results demonstrate the efficiency of population annealing for computation of the Bayes factors.

**Table 1.** Bayes factor  $B_{CI}^{ABC}$  computed with different annealing schedules in population annealing and different last epsilons in ABC rejection sampler.

annealing schedule/last $\varepsilon$	$B_{CI}^{ABC}$ (PA)	$B_{CI}^{ABC}$ (ARS)
AS1/0.25	1.367±0.033	1.391±0.031
AS2/0.15	1.896±0.089	1.935±0.096
AS3/0.11	9.618±1.175	10.079±2.830

Abbreviations are as follows: PA: population annealing, ARS: ABC rejection sampler. AS1: annealing schedule 1 = ( $\infty$ , 2, 1, 0.75, 0.5, 0.25), AS2: annealing schedule 2 = ( $\infty$ , 1, 0.5, 0.25, 0.2, 0.15), AS3: annealing schedule 3 = ( $\infty$ , 1, 0.5, 0.25, 0.15, 0.11). doi:10.1371/journal.pone.0104057.t001

## Mathematical model and simulation setting of the AKT pathway model

**Mathematical model of the AKT pathway model.** For the second test, we focused on the insulin dependent AKT pathway model [47,48]. This is because the model is more complicated than the FFL models, and the test with the model seems to be difficult. In addition, the experimental data about the pathway are open in Noguchi et al.'s web page [48].

Insulin is an important hormone which regulates various metabolic processes [49]. Especially, regulation of sugar metabolism is a very important role of insulin because defect in insulin action is related to type 2 diabetes mellitus [49]. As an intracellular signal transduction pathway, the AKT pathway plays an important role for the action of insulin [49,50]. Using a combination of experiments and mathematical modeling, Kubota et al. and Noguchi et al. demonstrated that temporal patterns of insulin selectively control glucose metabolism through the AKT pathway [47,48].

For the second test, we decided to use a part of the original insulin dependent AKT pathway models. The original models incorporates additional metabolic pathways downstream of AKT [47,48]. However, for simplicity, and because the downstream pathways differs between Kubota et al.'s model [47] and Noguchi et al.'s model [48], we employed the insulin-AKT module which is common in the original models. In the employed AKT model (Figure 8.A), the input signal is insulin and the output signal is phosphorylated AKT (pAKT in Figure 8.A). The dynamics of the model is represented by 6 differential equations with 16 rate constants and 9 initial amounts of the components. In this study, we used the same differential equations of the original models [47,48]. In addition, we used the same values of the initial amounts of components shown in Noguchi et al.'s paper [48]. Remaining 16 rate constants were free parameters in the test.

**Experimental data of AKT dynamics.** A part of the experimental results are open in Noguchi et al.'s web page [48]. In the experimental results, we used the time-series data of pAKT in response to the step stimulation of 1 nM insulin (red points in Figure 8.B) [48]. In their experiments, pAKT level is measured at eight time points (time = 0, 10, 30, 60, 120, 240, 360, 480 minutes) [48]. When the rate constants in the model are set to their default values shown in Noguchi et al.'s paper [48], the simulated trajectory could capture the experimental time-series data of pAKT except for the point at 480 minutes (green trajectory in Figure 8.B). Thus, we decided to use the experimental time-series data of pAKT except for the point at 480 minutes in the test. Those time-series data of pAKT were used for all the computations in the remaining sections.

**Prior distribution and weighting function.** As same as the case of the feed-forward loop network motif models, we set the

prior distribution of each parameter to independently follow the uniform distribution on a common logarithmic scale. The upper-bound and the lower-bound of each uniform distribution were set to  $10^4$  and  $10^{-6}$  respectively for all the 16 rate constants (parameters) in the model.

In this test study, we set the weighting function to the indicator function [14,22–24,26] represented as

$$f_w(D_{obs}|D_{sim}, \theta, M) \propto I(d(D_{obs}, D_{sim}|\theta) \leq \varepsilon, M),$$

and set the distance  $d(D_{obs}, D_{sim}|\theta)$  as

$$d(D_{obs}, D_{sim}|\theta) = \sum_{t=0,10,30,60,120,240,360} (pAKT_{obs}^t - pAKT_{sim}^t|\theta)^2.$$

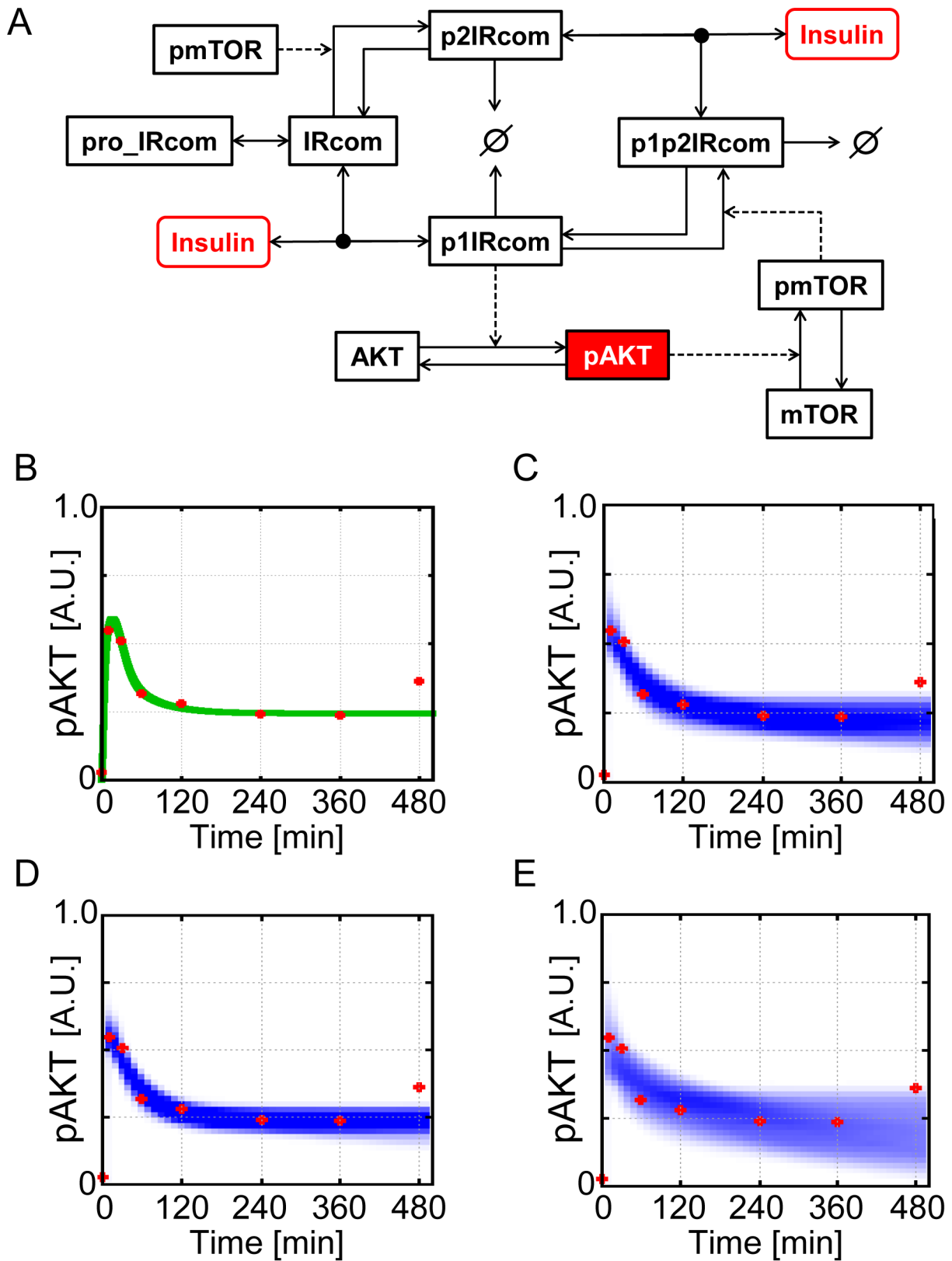
$\theta$  represents the 16 rate constants in the model.  $D_{obs} = \{pAKT_{obs}^t, t=0, 10, 30, 60, 120, 240, 360\}$  is the experimental time-series data of pAKT except for the point at 480 minutes in Noguchi et al.'s study [48].  $D_{sim}|\theta = \{pAKT_{sim}^t|\theta, t=0, 10, 30, 60, 120, 240, 360\}$  is a simulated time-series data of pAKT with  $\theta$ .

**Numerical simulation.** The total number of particles in population annealing was set to  $K=10000$ . As the default annealing schedule, the tolerance was gradually decreased as follows:

$$(\varepsilon_0, \varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5, \varepsilon_6, \varepsilon_7, \varepsilon_8, \varepsilon_9, \varepsilon_{10}, \varepsilon_{11}, \varepsilon_{12}, \varepsilon_{13}, \varepsilon_{14}) \\ = (\infty, 1.5, 1.0, 0.75, 0.5, 0.25, 0.1, 0.09, 0.08, 0.07, 0.06, 0.05, 0.04, 0.03, 0.02)$$

The proposal distribution of ABC-MCMC in population annealing was set to the uniform distribution on a common logarithmic scale. In concrete terms, at each step of ABC-MCMC, one of the parameters was randomly chosen, and the uniform random number between  $-0.25$  to  $0.25$  was added on a common logarithmic scale. For each particle, ABC-MCMC movements in population annealing were set to 16 steps. This is the number of the free parameters.

For time-series calculations, the ordinary differential equations were numerically solved by the fourth-order Runge-Kutta method with a time step of 0.001 minutes. Initial amounts of the components in the model were set to the values shown in Noguchi et al.'s paper [48]. As the input signal of the model, insulin is set to the step function (Insulin = 1 nM during the entire simulation) in all the computations.



**Figure 8. AKT pathway model.** (A) Structure of the AKT pathway model. The input signal is insulin. The output signal is pAKT. Solid arrows represent mass flows. Solid arrows directional to a lined circle represent degradation processes. Solid arrows with black circles represent association/dissociation processes. Dotted arrows represent enhancement of the processes. (B) Experimental data (red points) and simulated trajectory (green trajectory) of pAKT in response to the step stimulation of 1 nM insulin. (C), (D), (E) Reproduction of the experimental data. Simulations were run with the posterior parameter ensemble in response to the step stimulation of 1 nM insulin. (C) Annealing schedule:  $(\epsilon_0, \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_5, \epsilon_6, \epsilon_7, \epsilon_8, \epsilon_9, \epsilon_{10}, \epsilon_{11})$ .

$\varepsilon_{12}, \varepsilon_{13}, \varepsilon_{14}) = (\infty, 1.5, 1.0, 0.75, 0.5, 0.25, 0.1, 0.09, 0.08, 0.07, 0.06, 0.05, 0.04, 0.03, 0.02)$ . (D) Annealing schedule:  $(\varepsilon_0, \varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5, \varepsilon_6, \varepsilon_7, \varepsilon_8, \varepsilon_9, \varepsilon_{10}, \varepsilon_{11}, \varepsilon_{12}, \varepsilon_{13}, \varepsilon_{14}, \varepsilon_{15}) = (\infty, 1.5, 1.0, 0.75, 0.5, 0.25, 0.1, 0.09, 0.08, 0.07, 0.06, 0.05, 0.04, 0.03, 0.02, 0.01)$ . (E) Annealing schedule:  $(\varepsilon_0, \varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5, \varepsilon_6, \varepsilon_7, \varepsilon_8, \varepsilon_9, \varepsilon_{10}, \varepsilon_{11}) = (\infty, 1.5, 1.0, 0.75, 0.5, 0.25, 0.1, 0.09, 0.08, 0.07, 0.06, 0.05)$ . Blue-colored area is the probability density consists of the simulated trajectories. Red points are the experimental data.  
doi:10.1371/journal.pone.0104057.g008

### Reproduction of the experimental time-series data of insulin dependent AKT dynamics

We tested whether the posterior parameter ensemble can capture the experimental time-series data of insulin dependent AKT dynamics. We ran the simulations with use of all the 10000 output particles (parameter sets) of population annealing. Reproductions of the experimental data were shown in Figure 8.C, D, E. In Figure 8.C, the area of the probability density consists of the simulated trajectories (blue-colored area) could capture the experimental time-series data of pAKT (red points). When the annealing schedule was changed to the smaller tolerances  $(\varepsilon_0, \varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5, \varepsilon_6, \varepsilon_7, \varepsilon_8, \varepsilon_9, \varepsilon_{10}, \varepsilon_{11}, \varepsilon_{12}, \varepsilon_{13}, \varepsilon_{14}, \varepsilon_{15}) = (\infty, 1.5, 1.0, 0.75, 0.5, 0.25, 0.1, 0.09, 0.08, 0.07, 0.06, 0.05, 0.04, 0.03, 0.02, 0.01)$ , the area of the probability density got narrower (Figure 8.D). When the annealing schedule was changed to the larger tolerances  $(\varepsilon_0, \varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5, \varepsilon_6, \varepsilon_7, \varepsilon_8, \varepsilon_9, \varepsilon_{10}, \varepsilon_{11}) = (\infty, 1.5, 1.0, 0.75, 0.5, 0.25, 0.1, 0.09, 0.08, 0.07, 0.06, 0.05)$ , the area of the probability density got broader (Figure 8.E). In any of these cases, the simulations with the posterior parameter ensemble could reproduce the real experimental data of AKT dynamics, demonstrating the efficiency of our method.

### Model selection of the AKT pathway model

Lastly, we conducted Bayesian model selection of the AKT pathway model. As shown in Figure 8.B, pAKT shows strong transient and weak sustained response, so-called a partial adaptive response. From the network structure of the model, we can predict the mTOR-related negative feedback plays an important role for a partial adaptive response. Thus, for comparison with the wild type model shown in Figure 8.A, we prepared the mutant model which lacks the mTOR-related negative feedback. Lack of the negative feedback was realized by setting the values of related rate constants to zero.

To conduct Bayesian model selection, we computed the Bayes factor in the ABC framework represented as

$$B_{WM}^{ABC} = \frac{\int_{\theta} I(d(D_{obs}, D_{sim}(\theta)) \leq \varepsilon, M_{wild}) f(D_{sim}(\theta), M_{wild}) \pi(\theta | M_{wild}) d\theta}{\int_{\theta} I(d(D_{obs}, D_{sim}(\theta)) \leq \varepsilon, M_{mutant}) f(D_{sim}(\theta), M_{mutant}) \pi(\theta | M_{mutant}) d\theta}$$

Larger  $B_{WM}^{ABC}$  indicates that the wild type model is selected with stronger evidence against the mutant model. If the mTOR-related negative feedback is destroyed, the model will not be able to reproduce a partial adaptive response. Thus,  $B_{WM}^{ABC}$  should be a large value, at least larger than 1 to propose the validity of Bayesian model selection by population annealing.

Computation of  $B_{WM}^{ABC}$  was done under different annealing schedules in population annealing. Annealing schedules were set to same as those in parameter inference. For each annealing schedule, the mean and the standard deviation of 10 independent computations of  $B_{WM}^{ABC}$  were shown in Table 2. In Table 2, the means of  $B_{WM}^{ABC}$  were always larger than 1 independent of the annealing schedules. This result indicates that the wild type model was selected with stronger evidence against the mutant model independent of the annealing schedules. This is consistent with the predictions that the mTOR-related negative feedback is important for a partial adaptive response of pAKT, and the wild type model

must be selected in this case. These results demonstrate the validity of Bayesian model selection by population annealing.

We also conducted Bayesian model selection by ABC rejection sampler (Text S1). For comparison, the value of  $\varepsilon$  in ABC rejection sampler was set to the same value of the last  $\varepsilon$  of population annealing. The total number of sampling trials were set to 10000, same as the number of particles in population annealing. As shown in Table 2, ABC rejection sampler could compute the Bayes factor in the case of large value of  $\varepsilon$ . However, the standard deviation was larger than that of computed by population annealing. In addition, in the cases of middle or small value of  $\varepsilon$ , the number of acceptable particle in each independent run was mostly zero. In these cases, we could not calculate the Bayes factors (“-” in Table 2). These results demonstrate the efficiency of population annealing for computation of the Bayes factors.

### Discussion

In this paper, we introduced Bayesian model selection and parameter inference by population annealing. Firstly, we showed that population annealing can be used to compute the Bayesian posterior distributions. Secondly, we showed that the simulations with the posterior parameter ensemble could reproduce the artificial observed data and the experimental data used for parameter inference. In addition, the simulations with the posterior parameter ensemble could capture and predict the observed data which was not used for parameter inference. For both reproduction and prediction, population annealing enables us to run the simulations with the posterior parameter ensemble. These results also emphasize the importance to consider the ensemble or samples of parameters from the posterior distributions for parameter inference and subsequent simulations [11,13]. Lastly, we showed that the true model was correctly selected by computing the Bayes factor in the test with the FFL models. In the test with the AKT pathway model, the wild type model was correctly selected as expected. In addition, compared to ABC rejection sampler, population annealing showed smaller standard deviations of the Bayes factors. These results indicate that population annealing provides more stable computational result of the Bayes factor than ABC rejection sampler. All of these results support the efficiency of population annealing for Bayesian model selection, parameter inference and subsequent simulations with the parameter ensemble.

In the first test with the FFL models, we used the time-series data consists of 10 time points of  $Z$ . In the second test with the AKT pathway model, we used the time-series data consists of 7 time points of pAKT. These data scales may seem to be small. Our results may seem to be influenced by the smallness of the data scale. However, the results with the small data were not so largely different from the results obtained with the large data in this study (Text S3, Figure S1). These results indicate that, even though the available data scale is small, our approach can give us reasonable computational results. This is very important for real data analysis because it is not always possible to obtain the large experimental data.

In parameter inference of the FFL models, most of the marginal distributions of parameters were almost similar to the uniform distributions, which were same as the prior distributions in this

**Table 2.** Bayes factor  $B_{WM}^{ABC}$  computed with different annealing schedules in population annealing and different last epsilons in ABC rejection sampler.

annealing schedule/last $\epsilon$	$B_{WM}^{ABC}$ (PA)	$B_{WM}^{ABC}$ (ARS)
AS1/0.05	1.567 ± 0.496	2.600 ± 1.750
AS2/0.02	2.650 ± 0.959	-
AS3/0.01	1.895 ± 0.933	-

Abbreviations are as follows: PA: population annealing, RS: ABC rejection sampler. AS1: annealing schedule 1 = ( $\infty$ , 1.5, 1.0, 0.75, 0.5, 0.25, 0.1, 0.09, 0.08, 0.07, 0.06, 0.05), AS2: annealing schedule 2 = ( $\infty$ , 1.5, 1.0, 0.75, 0.5, 0.25, 0.1, 0.09, 0.08, 0.07, 0.06, 0.05, 0.04, 0.03, 0.02), AS3: annealing schedule 3 = ( $\infty$ , 1.5, 1.0, 0.75, 0.5, 0.25, 0.1, 0.09, 0.08, 0.07, 0.06, 0.05, 0.04, 0.03, 0.02, 0.01).

doi:10.1371/journal.pone.0104057.t002

study. If our purpose of parameter inference is the estimation of the correct values of parameters with high credibility, we should conclude the parameter inference was failed in this case. However, if our purpose is the reproduction or the prediction of the system dynamics, instead of choosing one set of representative values of parameters, we can run simulations with the posterior parameter ensemble. In this study, we want to propose the validity of this approach.

In the simulations with the posterior parameter ensemble, the area of the probability density consists of the simulated trajectories differs among the annealing schedules. This is because the last  $\epsilon$  values restrict the acceptable trajectories to the observed data. In both of the tests, we set the distance between the observed data and the simulated data as the sum of squared errors. In this case, the computation in the ABC framework is comparable to the maximum-likelihood estimation that errors are assumed to follow Gaussian distribution. This is because minimizing the distance (the smallest  $\epsilon$  wherever possible) is equivalent to maximizing the Gaussian likelihood function, as pointed out by Toni et al [15]. However in the ABC framework, we can compute, not the maximum-likelihood estimate as a point, but the posterior distribution as a distribution. This allows us to examine the interactions among parameters [21] and to run the simulations with the posterior parameter ensemble. This seems to be an advantage of ABC. In addition, we can change the annealing schedules in population annealing. This allows us to control the rigor of reproduction of the observed data or experimental data flexibly. Although we need to check the influence of the annealing schedules on the simulated data, flexibility of annealing schedule is one of the advantages of population annealing.

In model selection, ABC rejection sampler showed larger standard deviations than those computed by population annealing (Table 1, 2). In addition, ABC rejection sampler could not compute the Bayes factors in some cases (Table 2). As is known, the acceptance rate of ABC rejection sampler often gets lower in the case that the prior distribution is very different from the posterior distribution [15]. This was reconfirmed in this study, because larger standard deviations were obtained when  $\epsilon$  was smaller (Table 1), which strongly restricts the distributions of parameters. On the other hand, population annealing can avoid the problem and provide the stable computational result of the Bayes factor. This is because the intermediate distributions gradually changes from the prior distribution to the posterior distribution in population annealing.

In this study, with use of the indicator function, we could estimate the marginal likelihood in the ABC framework by population annealing. However, this is valid when the number of non-zero weight particles is proportional to the marginal likelihood. Thus, when the premise is not satisfied, this method

cannot be applied. For example, this is the case that the likelihood function is assumed as Gaussian distribution, not in the ABC framework. However, as the solution of this problem, thermodynamic integration by various kinds of Monte Carlo methods have already been developed and used to compute the Bayes factor [13,17,38]. Thus, those methods and population annealing can support each other to compute the Bayes factor.

For model selection, we needed to compute the marginal likelihoods of the competing models one-by-one. However, SMC can perform model selection among a number of competing models at one time [15,18]. One solution to overcome this weak point in population annealing is to use reversible jump MCMC [46] which sampler can jump among parameter subspaces of different dimensions. By jumping among a number of competing models with different parameter dimensions, we can conduct Bayesian model selection by population annealing at one time. This will be a future expansion of population annealing.

Although population annealing still has a room for improvement, population annealing will help us to conduct Bayesian model selection, parameter inference and subsequent simulations with the posterior parameter ensemble for better understanding and prediction of various biological phenomena in system level as shown in this study.

## Supporting Information

### Figure S1 Reproduction and prediction of the large scale observed data.

(A) Simulations with the posterior parameter ensemble of the coherent FFL model in response to the step stimulation of  $X$ . (B) Simulations with the posterior parameter ensemble of the incoherent FFL model in response to the step stimulation of  $X$ . (C) Simulations with the posterior parameter ensemble of the coherent FFL model in response to the pulse stimulation of  $X$ . (D) Simulations with the posterior parameter ensemble of the incoherent FFL model in response to the pulse stimulation of  $X$ . Blue-colored area is the probability density consists of the simulated trajectories. Red points are the observed data.

(PDF)

### Table S1 Observed data in response to the step stimulation of $X$ .

(PDF)

### Table S2 Observed data in response to the pulse stimulation of $X$ .

(PDF)

### Text S1 Supporting algorithms.

(PDF)

**Text S2 Population annealing in the conventional framework.**

(PDF)

**Text S3 Test with large data.**

(PDF)

**Acknowledgments**

We thank Prof. Shoji Takada for helpful discussions and comments.

**References**

- Kitano H (2002) Systems biology: a brief overview. *Science* 295: 1662–1664.
- Kitano H (2002) Computational systems biology. *Nature* 420: 206–210.
- Kirk P, Toni T, Stumpf MP (2008) Parameter inference for biochemical systems that undergo a Hopf bifurcation. *Biophys J* 95: 540–549.
- Ashyraliyev M, Fomekong-Nanfack Y, Kaandorp JA, Blom JG (2009) Systems biology: parameter estimation for biochemical models. *FEBS J* 276: 886–902.
- Mirsky HP, Liu AC, Welsh DK, Kay SA, Doyle FJ 3rd (2009) A model of the cell-autonomous mammalian circadian clock. *Proc Natl Acad Sci U S A* 106: 11107–11112.
- Moles CG, Mendes P, Banga JR (2003) Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res* 13: 2467–2474.
- Zeilinger MN, Farre EM, Taylor SR, Kay SA, Doyle FJ 3rd (2006) A novel computational model of the circadian clock in Arabidopsis that incorporates PRR7 and PRR9. *Mol Syst Biol* 2: 58.
- Mendes P, Kell D (1998) Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics* 14: 869–883.
- Akaike H (1974) A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control* 19: 716–723.
- Aoki K, Yamada M, Kunida K, Yasuda S, Matsuda M (2011) Processive phosphorylation of ERK MAP kinase in mammalian cells. *Proc Natl Acad Sci U S A* 108: 12675–12680.
- Kirk P, Thorne T, Stumpf MP (2013) Model selection in systems and synthetic biology. *Curr Opin Biotechnol* 24: 767–774.
- Kondo Y, Kaneko K, Ishihara S (2013) Identifying dynamical systems with bifurcations from noisy partial observation. *Phys Rev E* 87: 042716.
- Eydgahi H, Chen WW, Muhlich JL, Vitkup D, Tsitsiklis JN, et al. (2013) Properties of cell death models calibrated and compared using Bayesian approaches. *Mol Syst Biol* 9: 644.
- Barnes CP, Silk D, Sheng X, Stumpf MP (2011) Bayesian design of synthetic biological systems. *Proc Natl Acad Sci U S A* 108: 15190–15195.
- Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MP (2009) Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J R Soc Interface* 6: 187–202.
- Pullen N, Morris RJ (2014) Bayesian model comparison and parameter inference in systems biology using nested sampling. *PLoS One* 9: e88419.
- Vyshemirsky V, Girolami MA (2008) Bayesian ranking of biochemical system models. *Bioinformatics* 24: 833–839.
- Toni T, Stumpf MP (2010) Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics* 26: 104–110.
- Toni T, Ozaki Y, Kirk P, Kuroda S, Stumpf MP (2012) Elucidating the in vivo phosphorylation dynamics of the ERK MAP kinase using quantitative proteomics data and Bayesian model selection. *Mol Biosyst* 8: 1921–1929.
- Murakami Y, Takada S (2013) Bayesian parameter inference by Markov chain Monte Carlo with hybrid fitness measures: theory and test in apoptosis signal transduction network. *PLoS One* 8: e74178.
- Secrier M, Toni T, Stumpf MP (2009) The ABC of reverse engineering biological signalling systems. *Mol Biosyst* 5: 1925–1935.
- Marjoram P, Molitor J, Plagnol V, Tavaré S (2003) Markov chain Monte Carlo without likelihoods. *Proc Natl Acad Sci U S A* 100: 15324–15328.
- Sisson SA, Fan Y, Tanaka MM (2007) Sequential Monte Carlo without likelihoods. *Proc Natl Acad Sci U S A* 104: 1760–1765.
- Sisson SA, Fan Y (2011) Likelihood-Free Markov chain Monte Carlo; Brooks S, Gelman A, Jones G, Meng XL, editors: *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC Press. pp.313–335.
- Sunnaker M, Busetto AG, Numminen E, Corander J, Foll M, et al. (2013) Approximate Bayesian computation. *PLoS Comput Biol* 9: e1002803.
- Marin JM, Pudlo P, Robert CP, Ryder RJ (2012) Approximate Bayesian computational methods. *Statistics and Computing* 22: 1167–1180.
- Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics* 162: 2025–2035.
- Hukushima K, Iba Y (2003) Population Annealing and Its Application to a Spin Glass. *AIP Conference Proceedings* 690: 200–206.
- Iba Y (2005) An introduction to sequential Monte Carlo (Chikuiji monte karuro hou nyuumon) (in Japanese); Amari S, Takeuchi K, Takemura A, Iba Y, editors: *Keisan toukei 2*. Iwanami Shoten. pp.293–326.
- Iba Y (2001) Population Monte Carlo algorithms. *Transactions of the Japanese Society for Artificial Intelligence* 16: 279–286.
- Machta J (2010) Population annealing with weighted averages: a Monte Carlo method for rough free-energy landscapes. *Phys Rev E* 82: 026704.
- Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, et al. (2002) Network motifs: simple building blocks of complex networks. *Science* 298: 824–827.
- Mangan S, Alon U (2003) Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci U S A* 100: 11980–11985.
- Alon U (2007) Network motifs: theory and experimental approaches. *Nat Rev Genet* 8: 450–461.
- Shen-Orr S, Milo R, Mangan S, Alon U (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genetics* 31: 64–68.
- Hayot F, Jayaprakash C (2005) A feedforward loop motif in transcriptional regulation: induction and repression. *Journal of Theoretical Biology* 234: 133–143.
- Kass RE, Raftery AE (1995) Bayes Factors. *Journal of the American Statistical Association* 90: 773–795.
- Lartillot N, Philippe H (2006) Computing Bayes Factors Using Thermodynamic Integration. *Systematic Biology* 55: 195–207.
- Neal R (2001) Annealed importance sampling. *Statistics and Computing* 11: 125–139.
- Metropolis N, Rosenbluth AR, Rosenbluth MN, Teller AH, Teller E (1953) Equation of State Calculations by Fast Computing Machines. *J Chem Phys* 21: 1087–1092.
- Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57: 97–109.
- Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW (1999) Population Growth of Human Y Chromosomes: A Study of Y Chromosome Microsatellites. *Molecular Biology and Evolution* 16: 1791–1798.
- Zi Z, Sun Z (2005) Robustness Analysis of the IFN- $\gamma$  Induced JAK-STAT Signaling Pathway. *Journal of Computer Science and Technology* 20: 491–495.
- Eißing T, Allgöwer F, Bullinger E (2005) Robustness properties of apoptosis models with respect to parameter variations and intrinsic noise. *IEE Proceedings - Systems Biology* 152: 221–228.
- Ingram PJ, Stumpf MP, Stark J (2006) Network motifs: structure does not determine function. *BMC Genomics* 7: 108.
- Green PJ (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82: 64–68.
- Kubota H, Noguchi R, Toyoshima Y, Ozaki Y, Uda S, Watanabe K, Ogawa W, Kuroda S (2012) Temporal Coding of Insulin Action through Multiplexing of the AKT Pathway. *Mol Cell* 46: 820–832.
- Noguchi R, Kubota H, Yugi K, Toyoshima Y, Komori Y, Soga T, Kuroda S (2013) The selective control of glycolysis, gluconeogenesis and glycogenesis by temporal insulin patterns. *Mol Syst Biol* 9: 664.
- Whiteman EL, Cho H, Birnbaum MJ (2002) Role of Akt/protein kinase B in metabolism (2002). *Trends Endocrinol Metab* 13:444–451.
- Manning BD, Cantley LC (2007) AKT/PKB Signaling: Navigating Downstream (2007). *Cell* 129:1261–1274.

**Author Contributions**

Conceived and designed the experiments: YM. Performed the experiments: YM. Analyzed the data: YM. Contributed reagents/materials/analysis tools: YM. Contributed to the writing of the manuscript: YM.