Research Article

Estimating Network Kinetics of the MAPK/ERK Pathway Using Biochemical Data

Vilda Purutçuoğlu[1] and Ernst Wit[2]

[1] *Department of Statistics, Middle East Technical University, 06800 Ankara, Turkey*
[2] *Institute of Mathematics and Computing Science, Groningen University,*
  9747 AG Groningen, The Netherlands

Correspondence should be addressed to Vilda Purutçuoğlu, vpurutcu@metu.edu.tr

The MAPK/ERK pathway is a major signal transduction system which regulates many fundamental cellular processes including the growth control and the cell death. As a result of these roles, it has a crucial importance in cancer as well as normal developmental processes. Therefore, it has been intensively studied resulting in a wealth of knowledge about its activation. It is also well documented that the activation kinetics of the pathway is crucial to determine the nature of the biological response. However, while individual biochemical steps are well characterized, it is still difficult to predict or even understand how the activation kinetics works. The aim of this paper is to estimate the stochastic rate constants of the MAPK/ERK network dynamics. Accordingly, taking a Bayesian approach, we combined underlying qualitative biological knowledge in several competing dynamic models via sets of quasireactions and estimated the stochastic rate constants of these reactions. Comparing the resulting estimates via the BIC and DIC criteria, we chose a biological model which includes EGFR degradation—Raf-MEK-ERK cascade without the involvement of RKIPs.

1. Introduction

All cellular responses are regulated by various signal transduction pathways. The signal transduction starts by an external stimulus, usually a ligand binding to a receptor at the cell surface. They generate intracellular signals that are transmitted and integrated through biochemical reactions. These biochemical reactions often include changes in gene expressions and particular biological responses, which can include the cell reproduction, the motility, and others. On the other hand, any malfunction in these mechanisms has a direct influence on the expression or on the function of gene products which are components of these regulatory mechanisms, hereby resulting in alterations of the biological responses and many illnesses

such as heart disease, diabetes, and cancer [1]. Hence, knowledge about these pathways is very helpful for understanding the behaviour of biological activities and identifying drug targets which can be seen as the major aim of most of the biochemical and bioengineering studies.

The MAPK (mitogen-activated protein kinase) also known as ERK (extracellular signal-regulated kinase) pathway is one of the major signal transduction systems which are involved in the cellular growth control of all eukaryotes from the cell reproduction to death. The basic structure of the MAPK/ERK pathway includes a number of phosphorylations on the protein level. These phosphorylations are directed by core components which are Ras, Raf, MEK, and ERK proteins and regulatory components, such as ERK, RKIP, and RSK proteins. The phosphorylations by core components transduce the signal from the cell membrane to nucleus and the phosphorylations via regulatory components modulate the efficiency and the duration of the signal transduction through the pathway. The effects of these components are conducted by positive and negative feedback loops. The functionality of the MAPK pathway gives rise to nonlinear behaviour, such as ultrasensitivities, bistabilities, and periodic behaviour. Moreover, there is an influence of the effect of scaffolding proteins and subcellular compartmentalisation which makes the functionality of the system even more complex [2]. Because of these underlying features, besides the execution of the large number of interactions at the protein level, the outcomes of the signal transfer are stochastic in nature. Thus, the interaction maps of the proteins or the simple representation of the system like given in Figure 4 at Appendix are not enough to make predictive statements about this network structure, although they do provide us with a starting point for modelling the structure of the network. Hereby in order to better understand the inside of such a stochastic system, we need to find a mathematical model which can describe its behaviour and to estimate its model parameters. There are different mathematical models which enable us to represent the complex biological systems by linear and nonlinear ways [3–6] and to infer the model parameters stochastically [7]. Among modelling approaches, the nonlinear ones, which can be listed as the Langevin [8], diffusion approximation [9, 10], and inhomogeneous poisson process [10], can capture the actual randomness in the system by using the chemical master equation [11]. In these approaches, the first two models explain the system by the same expression in the sense that the former converts a deterministic model to stochastic one by adding a noise term in the equation [8]. Then it solves this extended model, that is, noise-added deterministic model, by Itô or Stratonovich integrals [11, 12]. On the other hand, the latter converts a stochastic expression to the differential equations via the Fokker-Planck equation [11]. The final approach, which is the inhomogeneous poisson process model, is based on the Gillespie technique which can exactly simulate the biological network stochastically [13]. However, since it is computationally demanding in inference of the model parameters, currently it is implemented for toy systems [10]. Therefore in this study we consider to implement the nonlinear diffusion approximation to model a realistically large MAPK/ERK pathway as it enables us to estimate the parameters via convex optimization techniques. In particular, we perform the Markov chain Monte Carlo methods for the inference of the NP hard problem.

Indeed for the purpose of estimation, there are several alternative methods. One of them is the application of the parallel computing [14]. In a Bayesian setting the parallelization technique can be effective in simple situations when the model parameters are conditionally independent on one another. Whereas it is not useful when the blocks of dependent variables are large [10, 14], like in our biochemical system. Another recent alternative is the method

discussed in [6]. They propose two approaches. The first approach is to use the reversible jump methods along with Metropolis-Hastings acceptance schemes and the second is to implement the block updating methods based on the use of Poisson process approximations and random walk schemes. From their analysis in the Lotka-Volterra model [10], they show that both methods are promising in efficiency; however, further extensions for the implementations in complex networks are needed. For the inference of the complex network dynamics, [15] suggests the application of the deterministic Michaelis-Menten model whose measurements possess independent log-normally distributed location and scale parameters. The estimates of model parameters are computed by the discretized maximum likelihood method via a piecewise constant step function in the small time interval. On the other hand, [16] uses a regression model whose regression coefficients are different for all genes for the given transcription factor activities (TFAs). The inference of the parameters is conducted by optimizing the likelihood of observations via scaled conjugate gradient algorithm. In this estimation as the number of model parameters becomes very large due to the assumption of different TFAs for each gene, a dimension reduction technique is implemented by putting a Gaussian prior on the transcription factors as the constraints in the marginalization of the likelihood function. Apart from these techniques there are some other approaches which are based mainly on the ordinary differential equations (ODEs) and hybrid methods which solve the nonlinear structure of the equilibrium state in ODE by adding noise terms [8, 17]. For instance, [18] implements ODE methods in conjunction with a Bayesian setting for the inference of a simple MAPK system. Reference [19] estimates the model parameters of the JAK2-STAT5 pathway by ODEs whose model is represented by power-law terms, resulting in noninteger kinetic orders. In that study these kinetic orders, which are the model parameters in this case, are solved via the genetic algorithm. The aim of this paper is to check the global structure of the MAPK/ERK system on western blot time-course data of a subset of proteins [20] by stochastic approach. As stated beforehand, we consider a *diffusion approximation* technique which enables to deal with the realistic complexity while maintaining the computational efficiency. In fact, the diffusion method uses continuous time and requires infinite information to describe a finite sample path, when the measurements are discrete and are only partially observed. Thus, we implement a discretized version of the diffusion approximation called the *Euler-Maruyama approximation*. This discretizes time, but keeps the continuous structure of measurements. The details of this method are described in [21]. But in this paper, different from the study in [21], we consider that the randomness in the model is also caused by the measurement error in the western-blotting dataset. Therefore as described in Section 3, we construct a more realistic model for the inference and change the updating stages of MCMC computation accordingly. Moreover in the calculation we use very limited observations for both the time points and species which enables us to check both the applicability of our proposal mathematical model and inference algorithm. Furthermore, in this study we suggest biologically plausible descriptions for the pathway and select the best-fitted model for our dataset by model-selection criteria and in the end we evaluate whether the selected model can really validate the biological findings about the system. Hereby in Section 2 we describe alternative models for the pathway structure. Section 3 outlines the formulation of the Euler-Maruyama approximation by means of the hierarchical model and gives details about the MCMC updates of the system. In Section 4.1 we describe the available western blot dataset and in Section 4.2 we present the results and compare the models via several information criteria. Finally, in Section 5, we conclude the outputs.

2. System and Methods

Due to its importance, the MAPK pathway has been studied intensively. Hence there are a number of sources that provide knowledge about this system. Most of this information is qualitative, and only a small part of it has been captured in an explicit set of reactions. Here we combine this qualitative information to present the steps of the activation of the pathway as a list of quasi reactions. A novelty of our approach is the usage of multiple state variables to deal with molecules for which different localizations in the cell are an intricate part of the dynamic process and to handle proteins that have different binding sites and various phosphorylation states. For expressing the translocation of proteins to the membrane, we use the notation m. On the other hand, the different levels of the phosphorylation are denoted by distinct abbreviations in the sense that p or p1 show the monophosphorylation, whereas p2 indicates the double-phosphorylation of a protein. For instance, in the following set of equations:

(a) $\text{Raf} + \text{PP2A} \xrightarrow{c_1} \text{Raf.I} + \text{PP2A}$;

(b) $\text{Raf.I} + \text{Ras.GTP} \xrightarrow{c_2} \text{Raf.I}_m + \text{Ras.GTP}$;

(c) $\text{MEK.p2} + \text{ERK} \xrightarrow{c_3} \text{MEK.p2} + \text{ERK.p1}$;

(d) $\text{MEK.p2} + \text{ERK.p1} \xrightarrow{c_4} \text{MEK.p2} + \text{ERK.p2}$;

we describe the interaction between PP2A and inactive Raf proteins by Reaction (a). This expression implies that the PP2A protein takes away the inhibitory phosphorylate of the inactive Raf (Raf) which enables Raf to activate. The underlying inactive and monophosphorylated Raf (Raf.I) is translocated by the recruitment of the active Ras (Ras.GTP) from the cytosol to the cell membrane (Raf.I$_m$) as denoted by Reaction (b). The activation of the ERK protein is shown by Reactions (c) and (d). Here the inactive ERK protein is double-phosphorylated via the active MEK (MEK.p2) with two steps, where ERK.p1 and ERK.p2 illustrate the mono- and double-phosphorylation of ERK proteins in the corresponding reactions. Using this representation we describe the pathway via 51 species involving 34 major proteins. Similarly, we use 94 reactions where 65 of them reflect changes in activities and translocations of species and the rest shows their degradations after dissociation. In biochemical reactions the probability per unit time of each event is stated by a stochastic rate constant. Here those constants are denoted as c_i's ($i = 1, \ldots, 94$) which are the parameters to be estimated.

For inferring c_i, we implement MCMC algorithms within a Bayesian framework. These algorithms enable us to deal with the challenges of genomic measurements, that is, missing data and a limited knowledge of the complete network structure. Moreover, they are able to include the uncertainty coming from various sources of variations [18].

We define various alternative models for the MAPK system by either combining several reactions in a single equation, hereby simplifying the system via reducing the number of total reactions, or by separating the pathway into small parts in such a way that the resulting modules will still be able to explain the biological relationships in the experiments. We suggest 5 models by using the listed totals below. But when generating these totals, we accept several assumptions by considering that the data are gathered by western blot: (i) the weight of protein complexes is heavier than the weight of single proteins, whereas (ii) single proteins have the same weight regardless whether they are mono/double-phosphorylated:
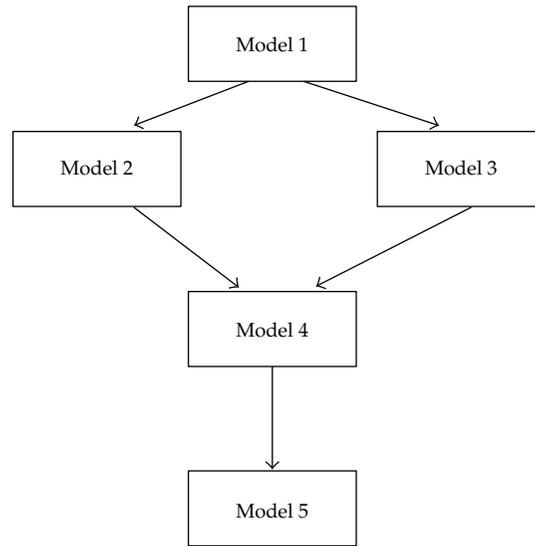
Figure 1: Simple representation of the relationship between nested structure of the five suggested MAPK/ERK pathway models.

(a) Total Ras = Ras.GDP + Ras.GTP,

(b) Total Raf = Raf + Raf.I + Raf.I$_m$ + Raf.A$_m$,

(c) Total MEK = MEK + MEK$_S$ + MEK$_F$ + MEK.p2,

(d) Total ERK = ERK + ERK.p2 for Models 1 and 2,

(e) Total ERK = ERK + ERK.p1 + ERK.p2 for Models 3, 4, and 5,

where Ras.GDP and Ras.GTP represent inactive and active Ras substrates, respectively. MEK$_S$ and MEK$_F$ denote MEK proteins phosphorylated by PAK and ERK proteins, in the given order. Raf.I and Raf.I$_m$ designate the inactive Raf in the cytosol and near the cell membrane, respectively, and finally Raf.A$_m$ indicates the active Raf near the cell membrane. Due to the available measured totals we face a convolution problem, for example, with respect to Raf, Raf.I, Raf.I$_m$, MEK, MEK$_S$, and MEK$_F$ proteins in all models. Models 3, 4, and 5 also have this problem with ERK and ERK.p1 proteins. To use the information from those totals, we need additional constraints in inference. We assume that for observed time points the number of molecules of MEK proteins should be equal to MEK = Total MEK − (MEK$_S$ + MEK$_F$ + MEK.p2), where MEK$_S$ and MEK$_F$ are both positive. Also for most Raf measurements we observe negative values after the substraction. Therefore, we completely omit the information on the Total Raf from all models. All suggested models are listed as below and their relationships are simply drawn in Figure 1. On the other hand, the complete list of reactions for each model with the description of their associated proteins can be found in Appendices B–G.

Model 1. According to the current biological theory of the MAPK pathway [20], it is suggested that RKIP and ERK proteins are the two main components which regulate the inhibitory control of the system. We assume that this control can be directed without RKIP since the complete activity, including the inhibitory control, of the pathway is still possible without RKIP terms as long as ERK and its associated complexes exist in the system. This assumption

enables the reduction in the number of substrates from 51 to 41 without altering the main activation of the system.

Moreover, seeing that in biochemical reactions, the protein degradations are much slower than the time periods during which biochemical activation and deactivation processes take place, we consider that protein degradations are insignificant and can be canceled from the reaction list. We also assume that some reactions can be combined into a single reaction if they are executed almost simultaneously or if they introduce another variable which would have no other function than to react with proteins in the given reaction. For instance, the equations MEK.p2 + ERK $\rightarrow$ MEK.p2 + ERK.p1 and MEK.p2 + ERK.p1 $\rightarrow$ MEK.p2 + ERK.p2 are summarized as MEK.p2 + ERK $\rightarrow$ MEK.p2 + ERK.p2 resulting in the cancelation of ERK.p1 proteins. By applying this simplification, we describe the system via 39 rather than 47 reactions, and 38 rather than 41 substrates in which 20 of them are linearly independent. In our estimation, we particularly eliminate merely nonobserved substrates because of their linear dependencies on other species (see Section 3), thereby all time-course observations are used in calculations. So the estimation is conducted by 6 observed (Ras.GDP, Ras.GTP, Raf.A$_m$, MEK.p2, ERK, and ERK.p2) and 14 unobserved substrates.

Model 2. It is the extension of Model 1 by including the degradation of the EGF receptor (EGFR). The EGFR degradation is the direct result of the MAPK activation by the internalization into vesicles of this receptor, thereby can be important for the steady-state description of the system. Thus, the inference is based on 40 reactions and 38 substrates in which 21 of them are linearly independent and the same 6 substrates are used as observed.

Model 3. Model 1 is extended by describing all reactions rather than summarizing reaction groups into a single equation. Apart from those changes, the model still excludes all degradations and all reactions which have RKIP proteins or their complexes. In Model 3 the pathway is represented by 47 reactions and 41 substrates where 26 of them are linearly independent and 5 (Ras.GDP, Ras.GTP, Raf.A$_m$, MEK.p2, and ERK.p2) among 26 substrates have observed measurements. All observed substrates are included in computations and the elimination of substrates due to linear dependence is conducted among the nonobserved ones. But different from previous models, the explicit measurements of ERK proteins are lost because of the convolution between ERK and ERK.p1 proteins.

Model 4. It is an extension of Model 3 in the sense that the EGFR degradation is included into the system because of the effect of the EGFR internalization as taken in Model 2. This results in 48 reactions and 41 substrates. In the calculation we use 27 linearly independent substrates where the common 5 substrates have real time-course measurements.

Model 5. This is the largest model such that the pathway includes all types of RKIP proteins, accordingly its inhibitory control is regulated via both ERK and RKIP species. Moreover it is assumed that the EGFR degradation is essential for the best MAPK description and the system does not use any kind of summary reactions. Hence, the inference is conducted by 66 reactions and 51 substrates. But due to the elimination of linearly dependent ones, the computation is done via 29 unobserved and 5 common observed proteins.

3. Algorithm

Under the assumption that the probability distribution of the number of molecules Y in each species at time t, $P(Y, t)$, is continuous, the stochastic model can be converted to a differential

equation model. Indeed, $P(Y, t)$ can be further expanded via a Taylor series expansion and the change in states of each species at t is found by a Fokker-Planck approach, in which a correlated noise term describes the stochastic behaviour of the model over and above a deterministic drift term [11]. Under this condition a nonlinear Fokker-Planck equation, which is converted from the continuous approximation of the chemical master equation, can be written in terms of infinitesimal mean and second-moment rates of the jump in states. By applying the Itô diffusion to the underlying Fokker-Planck equation, we get a diffusion formulation of the system. In a diffusion process, a finitely observed sample path is strictly speaking intractable. But under the assumption of discrete jumps at a large set of discrete time points, we can employ the Euler-Maruyama approximation of the diffusion process [10, 22, 23], which is one of the common techniques in stochastic modelling where the optimal values of the model parameters are found from the convergence distribution of the MCMC runs. Hereby the underlying model is described as follows:

$$\Delta Y_t = \mu(Y_t, \Theta)\Delta t + \beta^{1/2}(Y_t, \Theta)\Delta W_t, \tag{3.1}$$

in which $\mu(Y, \Theta) = V'h(Y, \Theta)$ and $\beta(Y, \Theta) = V' \operatorname{diag}\{h(Y, \Theta)\}V$ are *mean*, or *drift*, and *variance*, or *diffusion*, matrices, respectively, both depending on the state $Y = (Y_1, \ldots, Y_n)$ at time t and the parameter vector $\Theta = (c_1, \ldots, c_r)'$ explicitly. n is the total number of substrates and r states the total number of reactions in the system. ΔW_t is an n-dimensional independent identically distributed Brownian random vector over time Δt. V is the net effect matrix which is the difference between the stoichiometric coefficients of products and reactants of each reaction. Finally, $h(Y, \Theta)$ indicates the hazard, also called the *rate law*, of the reaction [10]. In these expressions, the notation $(')$ denotes the transpose of the underlying matrix or vector.

In the estimation of the reaction rates of the MAPK system, we consider that the state matrix Y is composed of both observed species X and unobserved or missing species Z at given time point t, hereby $Y_t \equiv (X_t, Z_t)'$ where each Y at t consists of totally n species as previously stated. To sample the missing states given the parameters and the parameters given all states, we use Metropolis-within-Gibbs steps. This algorithm is suitable for the cases in which the Gibbs and Metropolis-Hasting algorithms can be used in a cycle [24]. In the MCMC, furthermore, we apply data augmentation for the nonobserved states in the Euler process by putting latent states within each pair of time point. In this way, the strong bias in estimators caused by the large time steps in the Euler approximation is decreased in exchange for some additional computational complexity. Hereby we augment the state matrix Y, whose observations are not evenly spaced in time, in such a way as to make the resulting full state equally spaced in time. We consider two different regimes with more and less dense states. Accordingly, we add states in every 5 minutes and 10 minutes between each pair of totally 8 observed time points (Table 1) resulting in 37 and 77 time points, respectively, in observation matrices of all suggested models.

Finally, in our model we assume that each observation at time t possesses a measurement error, thereby, the model uncertainty originates from not only the stochasticity of the protein interactions, but also the variabilities from the measurement error. The application of this assumption in financial volatility models can be found in [25]. On the other hand, [5] use this assumption for the inference of a simple model of prokaryotic autoregulation. Here this assumption is inserted in our complex MAPK model by defining every observation at t under measurement error, W_t, via $W_t = X_t + \sigma_m$, where σ_m denotes the measurement error coming from normal distribution with mean 0 and variance, also called

Table 1: Western blotting time-course dataset for the MAPK/ERK pathway [20].

Time (min)	Ras.GTP	Total Ras	Raf.A$_m$	Total Raf
0	10	175	84	231
20	67	189	307	239
40	108	278	282	243
60	60	328	441	249
120	202	252	227	221
180	157	374	321	311
240	148	348	244	266
360	156	295	289	321
Time (min)	MEK.p2	Total MEK	ERK.p2	Total ERK
0	19	217	28	694
20	186	202	133	682
40	205	210	191	682
60	152	268	151	582
120	157	250	82	673
180	137	238	167	710
240	141	289	83	572
360	126	240	115	460

the *tuning parameter*, φ^2 for all time and observed substrates, that is, $\sigma_m \sim N(0, \varphi^2)$. In the sampling, since the tuning parameter φ^2 has an effect on the acceptance rate of σ_m, φ^2 is adjusted adaptively during the first 10,000 iterations of the burn-in period in the MCMC algorithm. More details about this sampling and the augmentation of Y can be found in [21].

In inference of the parameters $\Theta = (c_1, \ldots, c_r, \sigma_m)$, the positivity of rate constants c_i ($i = 1, \ldots, r$) is maintained by independent exponential priors with rate 1. For the prior of the measurement error σ_m, we select the inverse gamma with scale and location parameters 1 due to its conjugality of the normal distribution. Both densities produce flat and heavily tailed prior distributions for the elements of Θ. In the update of the system, we use block updates to sample the reaction rates c_i. In our block scheme we divide the r-dimensional vector of c into small and equally sized groups with a dimension d and then simulate each d-dimensional group sequentially. The size of the block update dimension d in each model is varied. We set d to 3, 4, and 5 for Models 1 and 2, Models 3 and 4, and Model 5, respectively. After the update of rate constants, we update σ_m.

When we renew c_i ($i = 1, \ldots, r$), we begin to update Y column by column whose entries are composed of partially observed and completely augmented data. The candidate values of each Y column are generated from distinct transition kernels via the Metropolis-within-Gibbs algorithm. We include the measurement error in the update of the partially observed states W_t from a normal distribution with mean X_t and variance σ_m via $W_t \sim N(X_t, \sigma_m) \cdot X_t$ denotes the observed values of the substrates at the given time t as described in advance. On the other side, the proposal of the completely augmented states is produced as stated in [9]. In Appendix A we present the derivations of candidate generators for partially observed states as an example of transition probabilities. At the end of the update Y, the convergence of the chain is controlled and the algorithm is repeated until the convergence is attained.

During the application of these MCMC methods to the five MAPK pathway models, we suggest an updating scheme which is able to deal with any kind of dependency coming

from the complexity of our system in the sampling process. We mainly divide the cause of dependency into two parts, namely *structural* and *incidental* dependence. The former is originated from the linear dependence of the substrates in the net effect matrix $V'V$. We eliminate these problematic species when starting the algorithm so that there is no singularity problem in the calculation of the likelihood. The latter, on the other hand, is caused by the singularity of $\beta(Y,\Theta) = V' \operatorname{diag}\{h(Y,\Theta)\}V$ or any numeric problems resulting from $h(Y,\Theta)$ where $h(Y,\Theta)$ is the n-dimensional hazard vector of the system for a given Y and Θ, and $\operatorname{diag}\{h(Y,\Theta)\}$ represents the $n \times n$ diagonal matrix of $h(Y,\Theta)$. To work with the incidental dependency we check whether a candidate value for either missing states of Y or Θ causes a singularity in the diffusion if it were accepted. If it preserves the nonsingularity, the acceptance probability is computed and the system is updated in the usual way. Otherwise, the candidate value is rejected even before calculating the acceptance probability. The details about this MCMC scheme can be found in [21].

4. Implementation

4.1. Description of the Experimental Data

We use the experimental measurements of the protein levels and activities of Ras, Raf, MEK, and ERK proteins from mitogen stimulated COS1 green monkey kidney cells. These cells can easily be propagated in the cell culture and are commonly used for biochemical experiments. The cells and their crude protein extracts contain many ten thousands of different proteins such that the proteins of interest constitute only a small proportion of the mixture. Hence the cell extracts usually need to be separated in order to be able to visualize the protein of interest. In collecting the experimental protein data from the underlying cell extracts, the technique of the *western blotting*, the *immunoprecipitation*, and the *pulldown* are widely utilized. If the proteins in a sample of the dissolved cell are detected by separating proteins according to their weights, the method is called the *western blotting* [26]. This method is used here to measure MEK and ERK activities. If a further purification is required, in order to measure enzymatic activities, the *immunoprecipitation* technique is used. So the protein of interest from the crude is extracted and its enzymatic activities are assayed by the incubation with relevant substrates *in vitro*. This is the way how the Raf activation is measured. A *pulldown* assay is a variation of the immunoprecipitation in the sense that the purification of pulldown is done by means of an assay specific to the binding side rather than a substance. In our data the measurements of Ras proteins are collected by this technique [27].

In our dataset initially the expression levels of these proteins which do not change over the time of the mitogen stimulation are examined as the control expressions and then the measured activities are reflected as the relative changes between untreated and mitogen-stimulated cells. The ODE method is successful in working with such changes in concentrations but they are insufficient to present the stochastic manner of biological activations [28, 29]. On the contrary, the stochastic methods enable to take these effects into account by working with the number of molecules. By assuming that the measurements are proportional to the true underlying number of protein molecules, we arbitrarily set the lowest intensity in the order to 10 molecules and extrapolate all other molecules from that proportion as shown in Table 1. On the other hand, the actual time-course data without extrapolation are represented in Table 5 and are plotted in Figure 5 in Appendix.

Table 2: Stability of the PAM clustering of alternative MAPK/ERK pathway models via the posterior means of model parameters for $\Delta t = 5$ and $\Delta t = 10$ with 5 clusters. The estimates are based on 15,000 MCMC runs after 85,000 burn-in runs.

Model	Number of parameters	Number of changes from $\Delta t = 5$ to $\Delta t = 10$
Model 1	39	1
Model 2	40	1
Model 3	47	5
Model 4	48	0
Model 5	66	1

Table 3: Computational time (where h, min, and sec stand for hour, minute, and second, resp.) for estimating model parameters via 100,000 MCMC runs of all models when $\Delta t = 5$ and $\Delta t = 10$ in R on the 3.00 GHz Dual Core Xeon processor with a single-trade application.

Model	$\Delta t = 5$	$\Delta t = 10$
	Real time	Real time
1	42 h 35 min 36 sec	21 h
2	45 h 23 min 54 sec	22 h 53 min 40 sec
3	52 h 26 min 43 sec	27 h 35 min 17 sec
4	56 h 01 min 41 sec	28 h 52 min 05 sec
5	91 h 46 min 42 sec	26 h 57 min 21 sec

4.2. Application to the Experimental Data

In inference of all models, the total number of iterations is chosen as 100,000 in which the first 85,000 are taken as the burn-in. The full list of reactions and concerning estimations are presented in Appendices B–F. From these results, first we check how strong the prior distributions of parameters affect their probability distributions. For the practical purpose we choose the Kolmogorov-Smirnov test in calculations, thereby compare MCMC outputs of rate constants and measurement error to random variables generated from exponential distribution with rate 1 and inverse gamma with parameters 1, respectively. The test statistics indicate significant differences for all models, empirically implying that the selected prior densities in fact cause no effect on the corresponding posterior densities. Then we compare the estimates found at $\Delta t = 5$ and $\Delta t = 10$. Using PAM clustering [30] with $k = 5$ classes (referring to very slow, slow, moderately slow, moderately fast, very fast) on the parameter estimates, we evaluate their stabilities. Table 2 shows that the estimated values are very stable between the two approximations. Moreover, from the analysis we see that although the majority of model parameters possess good mixing in the sense that the acceptance ratios are not less than our lower bound of 0.05, some of them own low mixing property such as 0.02 or 0.03. We consider that the sparsity of our dataset, high dependencies between species, and the existence of latent states cause less-likely candidate values in MCMC runs although the adaptive sampling plan during the first 10,000 MCMC runs enables us to get good tuning parameters. Furthermore, from the comparison of the computational demands under both Δt's, it is observed that the increase of the augmented states by going from $\Delta t = 10$ to $\Delta t = 5$ significantly raises the computational time (Table 3). On the other hand, from the comparison of estimates, it is found that the correlations between MCMC runs are slightly higher in $\Delta t = 5$ than in $\Delta t = 10$. Moreover we get better mixing in some of the estimates under $\Delta t = 5$ which is consistent with [9] since the augmentation in latent states leads to

Table 4: Results of BIC and DIC criteria for the alternative MAPK/ERK pathway models based on 500 MCMC samples when $\Delta t = 5$ by data in Table 1.

	Model 1	Model 2	Model 3	Model 4	Model 5
BIC	183622.8	**28450.14**	76446.05	41631.55	136870.2
DIC	325701	**46548.61**	128999.5	66110.5	186086.0

higher dependencies between states. But we do not observe a complete improvement for all rates under each model. We consider that the reason is relevant to the structure of our observed time points. Because for our western blotting data, the number of augmented states rises after the observed time point $t = 60$ min (12 latent states from $t = 60$ to $t = 180$ min and 24 latent states from $t = 240$ to $t = 360$ min), hereby, the dependency between the parameters and those states increases. On the other side, since the number of augmented states is not large from $t = 0$ to $t = 60$ min, the underlying augmentation does not very much results in worsening the dependency among the updates of these particular states and estimates (4 latent states from $t = 0$ to $t = 60$ min). Finally, it is seen that the estimated measurement errors are very big with respect to the observations in Table 1. We consider that similar to the estimates of reaction rates, the high correlation between the species as well as the sparsity of the observed time points and observed species lead to low mixing property, thereby, high measurement errors.

As a result of all these outcomes we think that though the current data are not very reliable for finding the possible reaction rates of the MAPK system, the estimates can be still used for the procedure of the model selection since the inference of all models is conducted under the same challenges and the main interest from the estimation results is to investigate whether any proposal model for the MAPK pathway (Section 2) suggests a better fit for the available western blotting dataset (Table 1). Although there are a number of model selection methods in literature, we choose the *Bayesian information criteria* (BIC) and *deviance information criteria* (DIC) as the model selection methods since they give a distance of the data to each of the alternative models which are not generalized from others, and enable us to compare all models even if all of which are incorrect in a biological sense [31].

In inference, the number of associated substrates having the structural dependence is different for each model. But considering our sparse western blotting measurements, we particularly order the species for every model in such a way that none of the observed substrate presented in Table 1 is discarded due to the structural dependency. In this way, we can execute the complete measurements in Table 1 in our estimation, thereby implement the comparison of both BIC and DIC via all data given in Table 1 in each alternative model. In our analysis although we observe slightly high autocorrelation under $\Delta t = 5$ than the estimates under $\Delta t = 10$, we still choose the results from $\Delta t = 5$ for the model selection seeing that the estimates of the associated measurement errors are lower that the ones under $\Delta t = 10$. Table 4 summarizes the test results based on MCMC outputs sampled from the convergent parts of the chains which are taken as the last 15,000 runs for each model. In order to select samples for the comparison, we check the autocorrelation functions (ACF) of each substrate per model and take 500 thinned samples whose MCMC outputs have relatively less correlations within each other. From the tabulated BIC and DIC values it is clear that Model 2 is the best fitting model for our western blot data. This model suggests that the system does not include any kinds of RKIP proteins and its components, and can be described by summarizing simultaneous reactions into a single equation. Moreover, it indicates that the degradation of EGFR is essential in the description of this system.

Table 5: Western blotting time-course dataset for the MAPK pathway collected from the School for Cancer Studies in Glasgow University.

Time (min)	Ras.GTP	Total Ras	Raf.A$_m$	Total Raf
0	35618	623520	300095	824192
20	240222	672847	1092111	850161
40	385132	989878	1005269	865119
60	212275	1169001	1570648	887820
120	718095	898817	810234	785545
180	559659	1332240	1142328	1105969
240	525533	1240655	867336	948456
360	555147	1049890	1028401	1141482
Time (min)	MEK.p2	Total MEK	ERK.p2	Total ERK
0	67914	772901	98683	2470824
20	662655	718812	472934	2430305
40	729792	748227	678992	2428199
60	539726	954104	536933	2071863
120	559206	889207	290295	2396161
180	487896	846376	594756	2529436
240	501463	1030196	295292	2038472
360	449943	856206	408216	1636536

In a biological sense Model 2 implies that the inhibited control of the Raf-MEK pathway is directed via ERK, rather than ERK and RKIP, proteins by phosphorylating MEK at T292 binding site or by phosphorylating SOS, as long as the EGR receptor triggers the activation of the pathway. Moreover, the chains of reactions which produce, particularly, Shc-Grb2-SOS$_m$, Grb2-SOS, ERK.p2, and Ras.GTP are very quick in the sense that Raf.A-Ras.GTP$_m$, ERK.p1, and Shc$_m$ proteins, which are used in the production of those proteins, cannot stay longer in the cell to count them as separate species. On the other hand although the knowledge about the MAPK system claims the existence of all kinds of RKIP proteins in the system [32–34], Model 2 describes the system without any RKIP species.

Once the selection of the best-fitted model we compare it with the proposal models of the system in [35] which are suggested by [36–38]. The main differences among these models are the function of the EGF receptor during the activation, the existence of Shc substrates and negative-feedback loops. Among these alternative models, Model 2 supports the model of [36]. The main feature of that model is the existence of the Ras activation with Shc and without Shc species, the degradation of EGRF, and core cascade of ERK via Raf and MEK. Moreover, different from this model, Model 2 includes the negative feedback phosphorylation of SOS by active ERK and defines the system with less species and reactions. Schoeberl et al. [36] model describes the pathway with 94 reactions and 125 reactions. On the other hand to the best of our knowledge since none of this model is validated by a real time-course dataset within a stochastic modelling, there have not yet any wet-data/approaches which can verify Model 2 or other suggested models.

On the other hand, if we only evaluate the estimated values of reaction rates of Model 2, we see that the dissociation type of reactions is the fast reactions and among other dissociations, the ones within the inactive Raf and active Ras complex (reaction 15 in Model 2) as well as within the active ERK, RSK, and transcription factor (reaction 39 in Model 2) have the highest speed. Whereas the reactions which regulate the Raf-MEK cascade by the
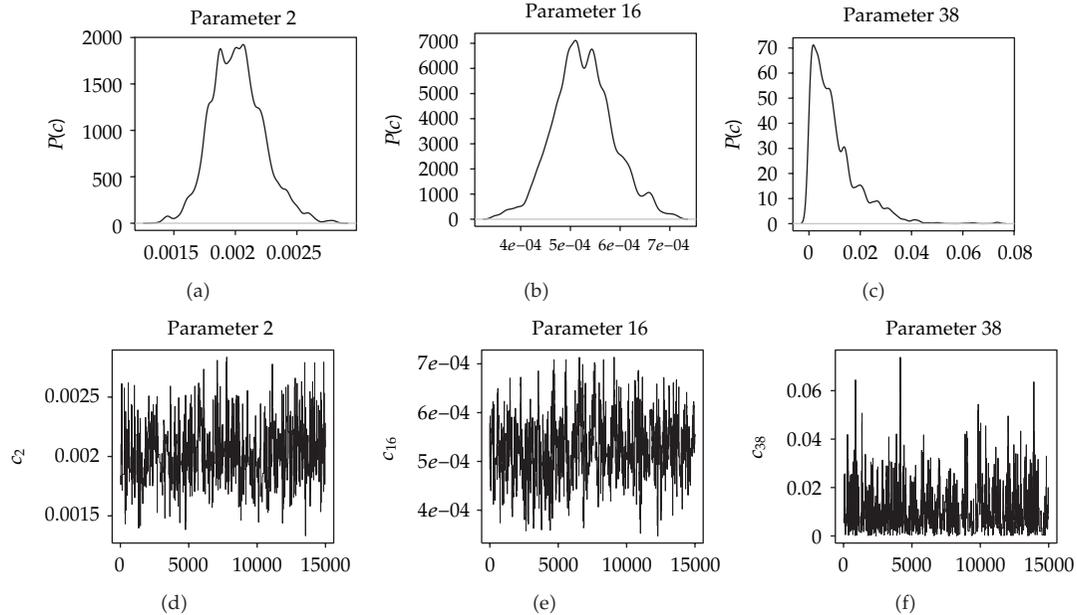
Figure 2: Examples of the probability distribution (a, b, and c) and trace plots (d, e, and f) of reaction rates 2 (which indicates the simultaneous recruitment of Shc and Grb2-SOS complex from the cytosol to the cell membrane by the recruitment of EGFR (a and d)), 16 (which shows the activation of MEK proteins by active Raf (b and e)), and 38 (which refers to the dissociation of active ERK and RSK complex (c and f)) of Model 2 under $\Delta t = 5$ after 85,000 burn-in runs.

inhibition of ERK (reactions 20 and 23 in Model 2) occur slowly. Figure 2 is shown as example from the posterior distributions of some reaction rates and the trace plots of their MCMC outputs. Figure 3, on the one hand, briefly describes Model 2. In this figure the thickness of the interaction indicates the speed of the associated reactions.

On the other hand, [39] shows that the single element update of states causes poor mixing due to the high correlation within the latent data. Thereby as an extension of this study, we can investigate the possibility of block updates of missing states [40, 41] to get better mixing. We consider that this suggested approach improves the current approximation technique in the class of stochastic methods where the optimal solution of the system cannot be found globally. Moreover, as shown in Figure 2, the curves may also imply fractal time series [42, 43] in which the system can be also modelled via the fractional Brownian motion (FBM). In FBM, the density of the time series has heavy-tailed distribution and indicates slowly decreasing autocorrelation functions with power spectrum of power-law type that can be observed as different cell signalling processes [44] as well as earthquake modelling and financial time series analysis. The power-law spectrum is seen if the process shows the self-similarity, that is, a time segment in the process presents the same behaviour as any segments of other time scales, and stationarity of its increments in diffusion model [45, 46]. If the self-similarity is found locally, rather than globally as in FBM, the system can be modelled by a more general modelling approach, called the multifractional Brownian motion [45]. On the other side if we deal with the change in increment of the Gaussian process, the system can be described by the fractional Gaussian model (FGM) [47–50]. Hereby as the extension of the diffusion modelling, the high correlation in the system can be solved by both FBM and FGM models that are specifically designed for such short- or long-correlated time series datasets.
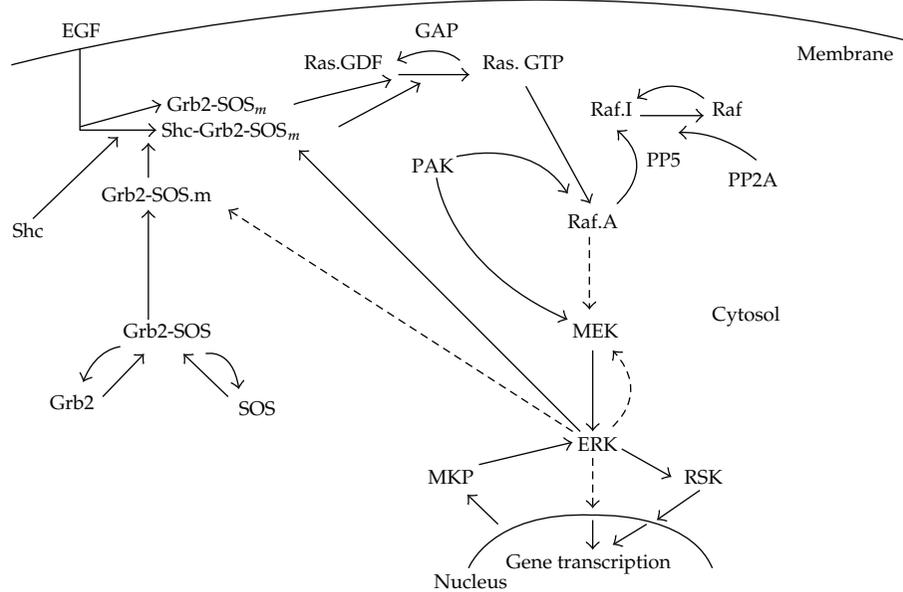
Figure 3: Simple representation of the MAPK/ERK pathway via Model 2 under $\Delta t = 5$. The thickness of the arrow indicates the speed of the interaction. The thick arrows represent fast interactions (estimated reaction rates have power 10^{-1}), only simple arrows describe moderately slow reactions (estimated reaction rates have power 10^{-2} or 10^{-3}), and finally dotted lines show very slow reactions among substrates (estimated reaction rates have power 10^{-4} or 10^{-5}).

5. Conclusion

We have described the MAPK pathway by quasireactions and applied multiple parameterizations for the same protein in different localizations and different phosphorylations. This reaction set has been used for estimating reaction rates of the real time-course data under the assumption that the observed substrates possess the measurement errors. In inference of the model parameters we have used the nonlinear Euler approximation combined with a data augmentation step and implemented it within a Metropolis-within-Gibbs algorithm by taking into account the singularity of the system at different stages of the estimation. This method is one of the best-known mathematical methods in system and computational biology. The alternative models for the western blot MAPK data were compared via BIC and DIC techniques. Both results have shown Model 2 as the best-selected model which fits the available experimental data. In this study even though we have been interested in the MAPK pathway as a result of its leading role in the cellular life cycle, the methods which we have developed for this system can be applied in the analysis of any large biochemical systems with many genes used in the biochemical and bioengineering researches.

Appendices

A. Derivation of Candidate Generators for Partially Observed States

In order to estimate the stochastic reaction rate constants of the MAPK/ERK pathway (Figure 4) when the states of the system are partially observed, we use the candidate generators whose details of the derivations are presented as follows.
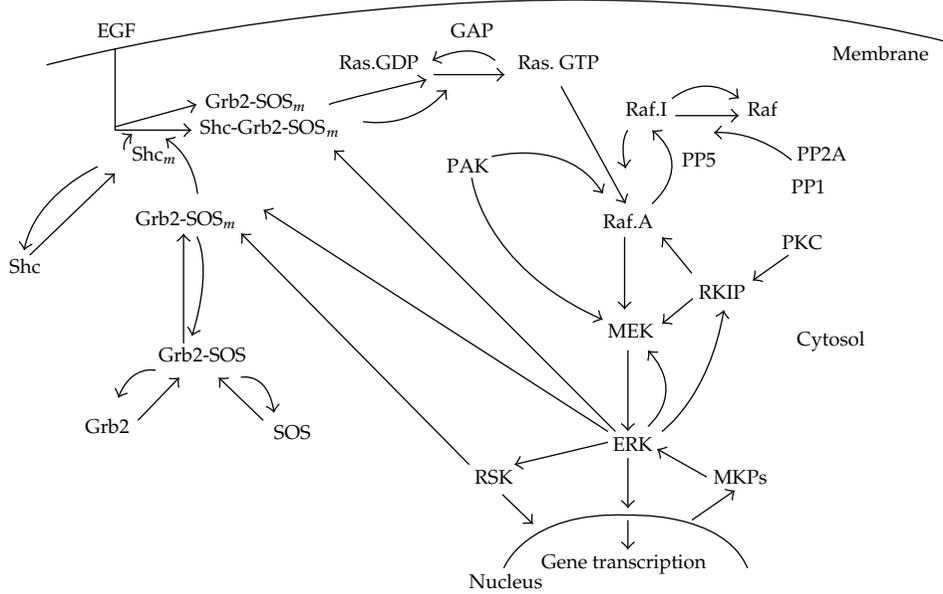
Figure 4: Simple representation of the structure of the MAPK/ERK pathway taken from Figure 2 in the study of [21].
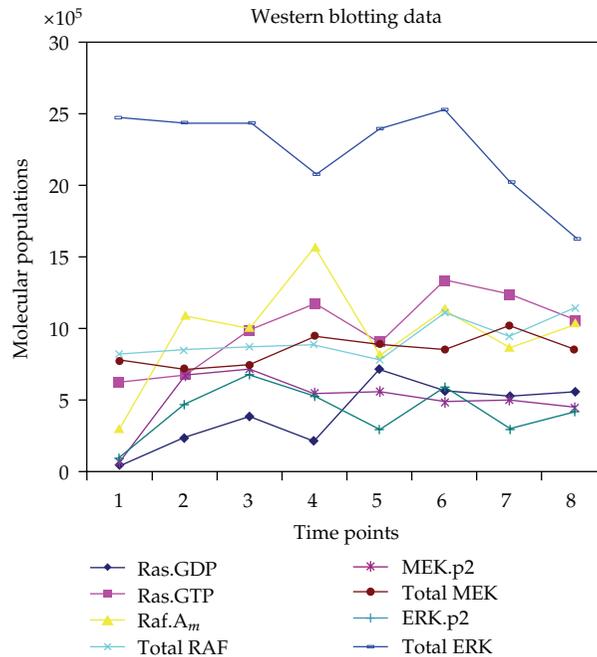
Figure 5: Plot of the western blotting data of the MAPK pathway observed at 8 time points according to the given molecular populations in Table 5.

Under the assumption that the observed values $W_i = w_i$ involves the measurement error and generated by $w_i \sim N(X_i, \sigma_m)$ where $X_i = (W_{i-1} + W_{i+1})/2$, the mean of Z_i^* conditional on $W_i = w_i$, denoted as $\eta_{Z_i^*}$, is derived via

$$\eta_{Z_i^*} = \mu_i^z + \delta \frac{(\beta_i^{zz})^{1/2}}{(\beta_i^{ww})^{1/2}} (w_i - X_i) \tag{A.1}$$

from the multivariate normal theories. Here $\mu_i^z = \mu(Z_i, \Theta)$, $\beta_i^{zz} = \beta(Y_i^{zz}, \Theta)$, and $\beta_i^{ww} = \beta(Y_i^{ww}, \Theta)$. Moreover, $Y_i^{zz} = Z_i{}'$ and $Y_i^{ww} = W_i{}'$. In that equation, δ refers to the correlation between Z_i and w_i and equals to $\delta = \beta_i^{zw} / [(\beta_i^{zz})^{1/2}(\beta_i^{ww})^{1/2}]$ in which $\beta_i^{zw} = \beta(Y_i^{zw}, \Theta)$ and $\beta_i^{wz} = \beta(Y_i^{wz}, \Theta)$. Furthermore, $Y_i^{zw} = (Z_i, W_i)'$ and $Y_i^{wz} = (W_i, Z_i)'$, while $(')$ describes the transpose of the given vector. By substituting δ into (A.1), we get

$$\begin{aligned}
\eta_{Z_i^*} &= \mu_i^z + \frac{\beta_i^{zw}}{(\beta_i^{zz})^{1/2}(\beta_i^{ww})^{1/2}} \frac{(\beta_i^{zz})^{1/2}}{(\beta_i^{ww})^{1/2}} (w_i - X_i) \\
&= \mu_i^z + \beta_i^{zw}(\beta_i^{ww})^{-1}(w_i - X_i),
\end{aligned} \tag{A.2}$$

where β_i^{ww} has full rank [51].

Since in our estimation $\mu_i^z = (Z_{i-1} + Z_{i+1})/2$ is similar to the idea of X_i, $\mu_{i-1}^z = \mu(Z_{i-1}, \Theta)$, and $\mu_{i+1}^z = \mu(Z_{i+1}, \Theta)$, hereby (A.2) can be written as

$$\eta_{Z_i^*} = \frac{1}{2}(Z_{i-1} + Z_{i+1}) + \beta_{i-1}^{zw}(\beta_{i-1}^{ww})^{-1}\left(w_i - \frac{1}{2}[W_{i-1} + W_{i+1}]\right). \tag{A.3}$$

On the other hand, in order to derive the conditional variance of Z_i^*, $\Sigma_{Z_i^*}$, under $W_i = w_i$, we use

$$\begin{aligned}
\Sigma_{Z_i^*} &= \beta_i^{zz}\left(1 - \delta^2\right) \\
&= \beta_i^{zz}\left(1 - \frac{\beta_i^{zw}\beta_i^{wz}}{\beta_i^{ww}\beta_i^{zz}}\right) = \beta_i^{zz} - \beta_i^{zz}\frac{\beta_i^{zw}\beta_i^{wz}}{\beta_i^{ww}\beta_i^{zz}} \\
&= \beta_i^{zz} - \beta_i^{zw}(\beta_i^{ww})^{-1}\beta_i^{wz},
\end{aligned} \tag{A.4}$$

from the consequence of the multivariate normal theory. In (A.4) seeing that $\beta_i^{zz} = \Delta t \beta_{i-1}^{zz}$, $\beta_i^{ww} = \Delta t \beta_{i-1}^{ww}$, and $\beta_i^{zw} = \Delta t \beta_{i-1}^{zw}$, the final $\Sigma_{Z_i^*}$ is found via

$$\begin{aligned}
\Sigma_{Z_i^*} &= \Delta t \beta_{i-1}^{zz} - \Delta t \beta_{i-1}^{zw}(\Delta t \beta_{i-1}^{ww})^{-1}\Delta t \beta_{i-1}^{wz} \\
&= \Delta t \beta_{i-1}^{zz} - \Delta t \beta_{i-1}^{zw}(\beta_{i-1}^{ww})^{-1}\beta_{i-1}^{wz},
\end{aligned} \tag{A.5}$$

in which $\beta_{i-1}^{zw} = \beta(Y_{i-1}^{zw}, \Theta)$ and $\beta_{i-1}^{wz} = \beta(Y_{i-1}^{wz}, \Theta)$. β_{i-1}^{ww}, similar to β_i^{ww}, has full rank [51].

Table 6: List of proteins used in inference of the reaction rates of Model 1.

Independent	Ras.GDP, Ras.GTP, Raf, Raf.I, Raf.A$_m$, MEK, MEK$_F$, MEK.p2 ERK, ERK.p2, ERK.p2-TF.p2, ERK.p2-RSK.A, Grb2, Shc, SOS, Grb2-SOS, Grb2-SOS$_m$, c-Fos, c-Fos.RNA, MKP.
Dependent	Raf.I$_m$, Raf.I-Ras.GTP$_m$, MEK$_S$, ERK.p2-RSK.A-TF.p2, Grb2$_m$, Shc-Grb2-SOS$_m$, Shc-Grb2$_m$, c-Fos.p, c-Fos.DNA, MKP.DNA, MKP.RNA, EGFR, TF, GAP, PP2A, PAK, PP5, RSK.

Therefore, from $Y_i = (w_i, Z_i)$, the candidate generator of Z_i conditional on $W_i = w_i$, $q(\cdot \mid w_i, Y_{i-1}, Y_{i+1}, \Theta)$, which converges pointwise to $\pi(\cdot \mid Y_{i-1}, Y_{i+1}, \Theta)$ when $\Delta t \to 0$ is sampled from

$$Z_i^* \sim N\left(\eta_{Z_i^*}, \Sigma_{Z_i^*}\right) \tag{A.6}$$

with mean $\eta_{Z_i^*} = (1/2)(Z_{i-1} + Z_{i+1}) + \beta_{i-1}^{zw}(\beta_{i-1}^{ww})^{-1}(w_i - (1/2)[W_{i-1} + W_{i+1}])$ and variance $\Sigma_{Z_i^*} = (1/2)\Delta t(\beta_{i-1}^{zz} - \beta_{i-1}^{zw}(\beta_{i-1}^{ww})^{-1}\beta_{i-1}^{wz})$. The candidate Y_i^* which is combination of both proposal w_i and Z^*, sequentially, that is, $Y^* \equiv (w^*, Z^*)'$, is accepted according to the acceptance probability

$$\alpha(Y_i^* \mid Y_i) = \min\left\{1, \frac{p(Y_i^* \mid Y_{i-1}, Y_{i+1}, \Theta)q(Y_i \mid Y_{i-1}, Y_{i+1}, \Theta)}{p(Y_i \mid Y_{i-1}, Y_{i+1}, \Theta)q(Y_i^* \mid Y_{i-1}, Y_{i+1}, \Theta)}\right\}. \tag{A.7}$$

B. Description of Model 1 and Related Estimations

For the estimation of the stochastic reaction rate constants, we use the western blotting data presented in Table 5 and plotted in Figure 5. The data are gathered from the School for Cancer Studies in the University of Glasgow and are already published in the study of [20]. In inference of the parameters in all suggested MAPK/ERK models, we arbitrarily equate the lowest intensity to 10 molecules and extrapolate the remaining proportionally.

We use the following list of reactions for estimating the reaction rates of Model 1. The estimated rates and associated statistics are summarized in Table 7. The list of substrates, on the other hand, is given in Table 6.

(1) Grb2 + SOS → Grb2-SOS

(2) EGFR + Shc + Grb2-SOS → EGFR + Shc-Grb2-SOS$_m$

(3) EGFR + Grb2-SOS → EGFR + Grb2-SOS$_m$

(4) Shc-Grb2-SOS$_m$ → Shc + Grb2-SOS

(5) Grb2-SOS → Grb2 + SOS

(6) Shc-Grb2-SOS$_m$ + Ras.GDP → Shc-Grb2-SOS$_m$ + Ras.GTP

(7) Grb2-SOS$_m$ + Ras.GDP → Grb2-SOS$_m$ + Ras.GTP

(8) Ras.GTP → Ras.GDP

(9) GAP + Ras.GTP → GAP + Ras.GDP

(10) Raf + PP2A $\rightarrow$ Raf.I + PP2A

(11) Raf.I + Ras.GTP $\rightarrow$ Raf.I$_m$ + Ras.GTP

(12) Raf.I$_m$ + Ras.GTP + PAK $\rightarrow$ Raf.A$_m$ + Ras.GTP + PAK

(13) PP5 + Raf.A$_m$ $\rightarrow$ PP5 + Raf.I$_m$

(14) Raf.I$_m$ $\rightarrow$ Raf

(15) Raf.I-Ras.GTP$_m$ $\rightarrow$ Raf.I$_m$ + Ras.GTP

(16) Raf.A$_m$ + MEK $\rightarrow$ Raf.A$_m$ + MEK.p2

(17) PAK + MEK $\rightarrow$ PAK +MEK$_F$

(18) MEK$_F$ + Raf.A$_m$ $\rightarrow$ MEK.p2 + Raf.A$_m$

(19) MEK.p2 + ERK $\rightarrow$ MEK.p2 + ERK.p2

(20) ERK.p2 + MEK $\rightarrow$ ERK.p2 +MEK$_S$

(21) MEK$_S$ + Raf.A$_m$ $\rightarrow$ MEK.p2 + Raf.A$_m$

(22) ERK.p2 + Shc-Grb2-SOS$_m$ $\rightarrow$ ERK.p2 + Shc-Grb2$_m$ + SOS

(23) ERK.p2 + Grb2-SOS$_m$ $\rightarrow$ ERK.p2 +Grb2$_m$ + SOS

(24) Shc-Grb2$_m$ $\rightarrow$ Shc + Grb2

(25) Grb2$_m$ $\rightarrow$ Grb2

(26) ERK.p2 + TF $\rightarrow$ ERK.p2-TF.p2

(27) ERK.p2-TF.p2 + c-Fos.DNA $\rightarrow$ ERK.p2-TF.p2 + c-Fos.DNA + c-Fos.RNA + c-Fos

(28) ERK.p2 + c-Fos $\rightarrow$ ERK.p2 + c-Fos.p

(29) ERK.p2-TF.p2 + MKP.DNA $\rightarrow$ ERK.p2-TF.p2 + MKP.DNA + MKP.RNA + MKP

(30) MKP + ERK.p2 $\rightarrow$ MKP + ERK

(31) ERK.p2 + RSK $\rightarrow$ ERK.p2-RSK.A

(32) ERK.p2-RSK.A + TF $\rightarrow$ ERK.p2-RSK.A-TF.p2

(33) ERK.p2-RSK.A-TF.p2 + c-Fos.DNA $\rightarrow$ ERK.p2-RSK.A-TF.p2 + c-Fos + c-Fos.DNA + c-Fos.RNA

(34) ERK.p2-RSK.A + c-Fos $\rightarrow$ ERK.p2-RSK.A + c-Fos.p

(35) ERK.p2-RSK.A-TF.p2 + MKP.DNA $\rightarrow$ ERK.p2-RSK.A-TF.p2 + MKP + MKP.DNA + MKP.RNA

(36) MKP + ERK.p2-RSK.A $\rightarrow$ MKP + ERK + RSK

(37) ERK.p2-TF.p2 $\rightarrow$ ERK.p2 + TF

(38) ERK.p2-RSK.A $\rightarrow$ ERK.p2 + RSK

(39) ERK.p2-RSK.A-TF.p2 $\rightarrow$ ERK.p2-RSK.A + TF

Table 7: Posterior means (μ), standard deviations (σ), and acceptance ratios (p) for the estimated rate constants c_i for Model 1 from the western blotting data when $\Delta t = 5$ and $\Delta t = 10$ time unit, respectively. The estimates are based on 15,000 MCMC runs after 85,000 burn-in runs.

Reaction	$\Delta t = 5$ unit			$\Delta t = 10$ unit		
	μ	σ	p	μ	σ	p
c_1	1.008×10^{-2}	1.219×10^{-3}	0.133	1.030×10^{-2}	1.291×10^{-3}	0.070
c_2	7.707×10^{-3}	1.148×10^{-3}	0.134	5.344×10^{-3}	8.487×10^{-4}	0.070
c_3	8.061×10^{-3}	1.441×10^{-3}	0.134	2.808×10^{-2}	4.336×10^{-3}	0.070
c_4	1.583×10^{-2}	1.280×10^{-2}	0.170	1.115×10^{-2}	9.342×10^{-3}	0.110
c_5	3.920×10^{-3}	3.551×10^{-3}	0.170	5.615×10^{-3}	4.142×10^{-3}	0.110
c_6	1.936×10^{-3}	1.729×10^{-3}	0.170	4.269×10^{-2}	2.200×10^{-3}	0.110
c_7	1.495×10^{-2}	9.222×10^{-4}	0.116	8.626×10^{-5}	8.144×10^{-5}	0.152
c_8	4.863×10^{-2}	3.140×10^{-2}	0.116	1.170×10^{-2}	1.055×10^{-2}	0.152
c_9	8.207×10^{-2}	2.955×10^{-2}	0.116	7.175×10^{-2}	1.221×10^{-2}	0.152
c_{10}	1.323×10^{-2}	3.806×10^{-3}	0.067	2.417×10^{-2}	2.847×10^{-3}	0.069
c_{11}	6.802×10^{-5}	1.906×10^{-5}	0.067	1.466×10^{-4}	2.753×10^{-5}	0.069
c_{12}	1.911×10^{-1}	4.272×10^{-3}	0.067	1.143×10^{-1}	4.948×10^{-3}	0.069
c_{13}	9.018×10^{-2}	2.693×10^{-3}	0.063	5.981×10^{-2}	2.650×10^{-3}	0.082
c_{14}	3.771×10^{-2}	1.019×10^{-2}	0.063	1.268×10^{-1}	1.976×10^{-2}	0.082
c_{15}	3.075	4.708×10^{-2}	0.063	5.404	1.265×10^{-1}	0.081
c_{16}	5.820×10^{-4}	4.940×10^{-5}	0.095	7.620×10^{-4}	5.648×10^{-5}	0.087
c_{17}	2.593×10^{-1}	1.173×10^{-3}	0.093	7.053×10^{-1}	6.147×10^{-4}	0.084
c_{18}	1.126×10^{-4}	1.629×10^{-5}	0.095	1.220×10^{-4}	6.378×10^{-6}	0.087
c_{19}	4.486×10^{-3}	2.599×10^{-4}	0.080	2.544×10^{-4}	9.265×10^{-6}	0.015
c_{20}	1.053×10^{-3}	1.274×10^{-4}	0.080	3.734×10^{-4}	6.867×10^{-5}	0.015
c_{21}	7.311×10^{-3}	5.164×10^{-4}	0.080	1.140×10^{-2}	2.944×10^{-4}	0.015
c_{22}	2.370×10^{-3}	2.846×10^{-4}	0.081	1.008×10^{-3}	1.792×10^{-4}	0.062
c_{23}	3.346×10^{-5}	9.889×10^{-6}	0.081	1.804×10^{-4}	2.893×10^{-5}	0.063
c_{24}	4.694×10^{-1}	4.176×10^{-3}	0.081	2.526×10^{-1}	2.539×10^{-3}	0.062
c_{25}	3.046×10^{-2}	4.346×10^{-3}	0.020	2.373×10^{-1}	3.101×10^{-2}	0.116
c_{26}	5.659×10^{-5}	3.625×10^{-5}	0.020	6.631×10^{-4}	1.615×10^{-4}	0.116
c_{27}	1.614×10^{-3}	6.033×10^{-4}	0.020	3.801×10^{-3}	1.908×10^{-3}	0.116
c_{28}	4.537×10^{-5}	1.401×10^{-5}	0.054	2.085×10^{-5}	1.327×10^{-5}	0.061
c_{29}	1.878×10^{-1}	5.200×10^{-3}	0.054	3.369×10^{-2}	2.679×10^{-3}	0.061
c_{30}	5.698×10^{-1}	3.673×10^{-2}	0.054	1.398×10^{-2}	8.928×10^{-4}	0.061
c_{31}	5.442×10^{-2}	2.554×10^{-3}	0.050	4.812×10^{-2}	2.139×10^{-3}	0.137
c_{32}	4.123×10^{-1}	2.216×10^{-2}	0.050	4.021×10^{-1}	2.590×10^{-2}	0.137
c_{33}	1.584×10^{-3}	1.401×10^{-3}	0.050	9.498×10^{-3}	5.594×10^{-3}	0.137
c_{34}	1.787×10^{-4}	1.531×10^{-4}	0.075	5.371×10^{-4}	2.169×10^{-4}	0.103
c_{35}	5.809×10^{-3}	5.354×10^{-3}	0.075	5.067×10^{-2}	1.478×10^{-2}	0.104
c_{36}	1.825×10^{-1}	1.468×10^{-2}	0.075	1.458×10^{-1}	6.914×10^{-3}	0.103
c_{37}	8.745×10^{-3}	2.396×10^{-3}	0.034	3.347×10^{-2}	7.648×10^{-3}	0.118
c_{38}	9.191×10^{-3}	8.221×10^{-3}	0.034	1.433×10^{-2}	1.313×10^{-2}	0.119
c_{39}	3.930	1.817×10^{-1}	0.034	4.362	2.063×10^{-1}	0.118

C. Description of Model 2 and Related Estimations

In Model 2, basically, we use the same reaction list given for Model 1. But different from that model, we also add the reaction of the EGFR degradation which is denoted by EGFR $\rightarrow$ $\emptyset$ as the 40th reaction. The estimated rate constants concerning Model 2 and other statistics are summarized in Table 8. Apart from the EGF receptor, the list of substrates and their classes are the same as given in Table 6. Under this model the EGF receptor is counted as independent, rather than dependent, term.

D. Description of Model 3 and Related Estimations

We apply the following list of reactions to infer the reaction rates of Model 3. The estimated reaction rates and calculated statistics are presented in Table 9. Table 10 gives the list of substrates used in inference.

(1) Grb2 + SOS $\rightarrow$ Grb2-SOS

(2) EGFR + Shc $\rightarrow$ EGFR + Shc$_m$

(3) EGFR + Grb2-SOS $\rightarrow$ EGFR + Grb2-SOS$_m$

(4) Shc$_m$ + Grb2-SOS$_m$ $\rightarrow$ Shc-Grb2-SOS$_m$

(5) Shc-Grb2-SOS$_m$ $\rightarrow$ Shc$_m$ + Grb2-SOS$_m$

(6) Shc$_m$ $\rightarrow$ Shc

(7) Grb2-SOS$_m$ $\rightarrow$ Grb2-SOS

(8) Grb2-SOS $\rightarrow$ Grb2 + SOS

(9) Shc-Grb2-SOS$_m$ + Ras.GDP $\rightarrow$ Shc-Grb2-SOS$_m$ + Ras.GTP

(10) Grb2-SOS$_m$ + Ras.GDP $\rightarrow$ Grb2-SOS$_m$ + Ras.GTP

(11) Ras.GTP $\rightarrow$ Ras.GDP

(12) GAP + Ras.GTP $\rightarrow$ GAP + Ras.GDP

(13) Raf + PP2A $\rightarrow$ Raf.I + PP2A

(14) Raf.I + Ras.GTP $\rightarrow$ Raf.I$_m$ + Ras.GTP

(15) Raf.I$_m$ + Ras.GTP $\rightarrow$ Raf.I-Ras.GTP$_m$

(16) Raf.I-Ras.GTP$_m$ + PAK $\rightarrow$ Raf.A-Ras.GTP$_m$ + PAK

(17) Raf.A-Ras.GTP$_m$ $\rightarrow$ Raf.A$_m$ + Ras.GTP

(18) PP5 + Raf.A$_m$ $\rightarrow$ PP5 + Raf.I$_m$

(19) Raf.I$_m$ $\rightarrow$ Raf

(20) Raf.I-Ras.GTP$_m$ $\rightarrow$ Raf.I$_m$ + Ras.GTP

(21) Raf.A$_m$ + MEK $\rightarrow$ Raf.A$_m$ + MEK.p2

(22) PAK + MEK $\rightarrow$ PAK + MEK$_F$

(23) MEK$_F$ + Raf.A$_m$ $\rightarrow$ MEK.p2 + Raf.A$_m$

(24) MEK.p2 + ERK $\rightarrow$ MEK.p2 + ERK.p1

(25) MEK.p2 + ERK.p1 $\rightarrow$ MEK.p2 + ERK.p2

Table 8: Posterior means (μ), standard deviations (σ), and acceptance ratios (p) for the estimated rate constants c_i for Model 2 from the western blotting data when $\Delta t = 5$ and $\Delta t = 10$ are time units, respectively. The estimates are based on $15,000$ MCMC runs after $85,000$ burn-in runs.

Reaction	$\Delta t = 5$ unit			$\Delta t = 10$ unit		
	μ	σ	p	μ	σ	p
c_1	1.228×10^{-2}	1.246×10^{-3}	0.064	9.575×10^{-3}	1.589×10^{-3}	0.028
c_2	2.015×10^{-3}	2.131×10^{-4}	0.064	6.589×10^{-4}	1.036×10^{-4}	0.028
c_3	5.250×10^{-3}	7.264×10^{-4}	0.064	3.804×10^{-3}	6.289×10^{-4}	0.028
c_4	1.251×10^{-2}	1.151×10^{-2}	0.159	4.674×10^{-2}	2.835×10^{-2}	0.125
c_5	4.588×10^{-3}	4.039×10^{-3}	0.159	2.363×10^{-3}	2.087×10^{-3}	0.125
c_6	1.024×10^{-2}	3.350×10^{-3}	0.159	2.825×10^{-3}	2.120×10^{-3}	0.125
c_7	9.500×10^{-3}	6.363×10^{-4}	0.119	6.488×10^{-3}	4.798×10^{-4}	0.136
c_8	2.775×10^{-2}	1.612×10^{-2}	0.119	3.552×10^{-2}	9.800×10^{-3}	0.136
c_9	8.103×10^{-2}	1.749×10^{-2}	0.119	1.430×10^{-2}	9.956×10^{-3}	0.136
c_{10}	1.814×10^{-2}	3.106×10^{-3}	0.086	4.775×10^{-2}	4.988×10^{-3}	0.075
c_{11}	1.756×10^{-4}	2.908×10^{-5}	0.086	1.837×10^{-4}	2.674×10^{-5}	0.075
c_{12}	1.951×10^{-1}	3.997×10^{-3}	0.085	1.192×10^{-1}	4.119×10^{-3}	0.075
c_{13}	8.965×10^{-2}	2.181×10^{-3}	0.083	8.328×10^{-2}	2.475×10^{-3}	0.110
c_{14}	8.266×10^{-2}	1.329×10^{-2}	0.083	1.696×10^{-1}	2.793×10^{-2}	0.110
c_{15}	3.546	1.096×10^{-1}	0.082	2.837	1.250×10^{-1}	0.110
c_{16}	5.263×10^{-4}	6.079×10^{-5}	0.098	1.197×10^{-3}	6.928×10^{-5}	0.071
c_{17}	4.428×10^{-1}	1.440×10^{-3}	0.097	6.341×10^{-1}	6.820×10^{-4}	0.070
c_{18}	1.536×10^{-4}	7.442×10^{-6}	0.098	8.151×10^{-5}	4.562×10^{-6}	0.071
c_{19}	2.526×10^{-3}	1.375×10^{-4}	0.056	2.330×10^{-4}	9.969×10^{-6}	0.033
c_{20}	7.068×10^{-4}	8.094×10^{-5}	0.056	8.367×10^{-4}	1.139×10^{-4}	0.033
c_{21}	7.988×10^{-3}	3.053×10^{-4}	0.056	6.698×10^{-3}	6.198×10^{-4}	0.033
c_{22}	2.948×10^{-3}	3.035×10^{-4}	0.075	1.318×10^{-3}	1.973×10^{-4}	0.067
c_{23}	2.803×10^{-4}	4.000×10^{-5}	0.075	2.004×10^{-4}	3.455×10^{-5}	0.067
c_{24}	5.324×10^{-1}	5.017×10^{-3}	0.074	2.570×10^{-1}	4.554×10^{-3}	0.067
c_{25}	1.632×10^{-1}	9.574×10^{-3}	0.058	2.024×10^{-1}	1.598×10^{-2}	0.122
c_{26}	3.513×10^{-4}	7.545×10^{-5}	0.058	2.492×10^{-4}	9.149×10^{-5}	0.122
c_{27}	5.338×10^{-3}	2.018×10^{-3}	0.058	3.546×10^{-3}	2.618×10^{-3}	0.122
c_{28}	3.300×10^{-5}	2.190×10^{-5}	0.044	4.668×10^{-5}	1.335×10^{-5}	0.080
c_{29}	5.371×10^{-2}	3.414×10^{-3}	0.044	4.370×10^{-2}	4.481×10^{-3}	0.080
c_{30}	3.244×10^{-1}	1.549×10^{-2}	0.044	2.038×10^{-2}	1.001×10^{-3}	0.080
c_{31}	7.509×10^{-2}	3.647×10^{-3}	0.055	4.297×10^{-2}	2.031×10^{-3}	0.110
c_{32}	5.220×10^{-1}	2.315×10^{-2}	0.055	2.528×10^{-1}	1.250×10^{-2}	0.110
c_{33}	7.916×10^{-3}	5.697×10^{-3}	0.055	1.641×10^{-2}	7.164×10^{-3}	0.111
c_{34}	7.036×10^{-4}	3.411×10^{-4}	0.108	1.560×10^{-4}	1.302×10^{-4}	0.073
c_{35}	1.228×10^{-2}	9.204×10^{-3}	0.108	1.454×10^{-2}	9.950×10^{-3}	0.073
c_{36}	2.565×10^{-1}	1.107×10^{-2}	0.108	1.111×10^{-1}	7.249×10^{-3}	0.073
c_{37}	1.090×10^{-2}	3.086×10^{-3}	0.086	1.454×10^{-2}	4.631×10^{-3}	0.105
c_{38}	9.962×10^{-3}	9.301×10^{-3}	0.086	1.094×10^{-2}	9.427×10^{-3}	0.105
c_{39}	3.960	1.188×10^{-1}	0.085	3.215	1.444×10^{-1}	0.105
c_{40}	4.201×10^{-3}	7.856×10^{-4}	0.077	4.977×10^{-3}	1.099×10^{-3}	0.094

Table 9: Posterior means (μ), standard deviations (σ), and acceptance ratios (p) for the estimated rate constants c_i for Model 3 from the western blotting data when $\Delta t = 5$ and $\Delta t = 10$ are time units, respectively. The estimates are based on 15,000 MCMC runs after 85,000 burn-in runs.

Reaction	$\Delta t = 5$ unit			$\Delta t = 10$ unit		
	μ	σ	p	μ	σ	p
c_1	6.480×10^{-3}	9.529×10^{-4}	0.079	7.552×10^{-3}	1.218×10^{-3}	0.072
c_2	1.573×10^{-2}	2.645×10^{-3}	0.079	3.697×10^{-2}	5.668×10^{-3}	0.072
c_3	1.933×10^{-2}	3.191×10^{-3}	0.079	5.075×10^{-2}	7.618×10^{-3}	0.072
c_4	2.432×10^{-3}	4.300×10^{-4}	0.055	7.508×10^{-3}	1.107×10^{-3}	0.136
c_5	9.854×10^{-3}	6.114×10^{-3}	0.055	1.807×10^{-2}	1.304×10^{-2}	0.136
c_6	1.818×10^{-3}	1.349×10^{-3}	0.055	5.590×10^{-3}	4.128×10^{-3}	0.136
c_7	1.026×10^{-3}	9.363×10^{-4}	0.089	8.907×10^{-3}	7.014×10^{-3}	0.109
c_8	2.913×10^{-3}	2.128×10^{-3}	0.089	2.330×10^{-3}	2.240×10^{-3}	0.109
c_9	1.076×10^{-1}	6.713×10^{-3}	0.089	3.215×10^{-2}	4.551×10^{-3}	0.109
c_{10}	7.288×10^{-4}	5.691×10^{-4}	0.140	4.344×10^{-3}	7.679×10^{-4}	0.077
c_{11}	1.403×10^{-1}	1.724×10^{-2}	0.140	1.665×10^{-2}	1.008×10^{-2}	0.077
c_{12}	3.147×10^{-2}	1.805×10^{-2}	0.140	4.509×10^{-2}	1.067×10^{-2}	0.077
c_{13}	1.479×10^{-2}	2.592×10^{-3}	0.061	3.214×10^{-2}	4.592×10^{-3}	0.047
c_{14}	1.508×10^{-4}	2.329×10^{-5}	0.061	2.011×10^{-4}	2.734×10^{-5}	0.047
c_{15}	9.137×10^{-2}	1.831×10^{-3}	0.061	7.562×10^{-2}	1.470×10^{-3}	0.047
c_{16}	10.575	1.435×10^{-1}	0.094	10.816	1.745×10^{-1}	0.128
c_{17}	1.659	3.287×10^{-2}	0.095	2.395	5.379×10^{-2}	0.130
c_{18}	6.626×10^{-2}	1.319×10^{-3}	0.095	5.601×10^{-2}	1.078×10^{-3}	0.131
c_{19}	9.251×10^{-2}	1.660×10^{-2}	0.085	1.731×10^{-1}	1.378×10^{-2}	0.140
c_{20}	5.994×10^{-2}	3.102×10^{-2}	0.085	6.973×10^{-1}	1.754×10^{-2}	0.140
c_{21}	1.380×10^{-3}	1.449×10^{-4}	0.085	1.317×10^{-3}	1.326×10^{-4}	0.140
c_{22}	6.498×10^{-1}	3.442×10^{-3}	0.055	4.784×10^{-1}	6.560×10^{-3}	0.038
c_{23}	3.882×10^{-4}	1.288×10^{-5}	0.055	2.441×10^{-4}	9.155×10^{-6}	0.038
c_{24}	4.990×10^{-2}	6.690×10^{-4}	0.055	3.156×10^{-2}	3.652×10^{-4}	0.038
c_{25}	1.931×10^{-3}	2.731×10^{-5}	0.077	9.613×10^{-4}	1.230×10^{-5}	0.039
c_{26}	4.488×10^{-3}	2.598×10^{-4}	0.077	5.208×10^{-3}	2.491×10^{-4}	0.039
c_{27}	4.509×10^{-2}	9.108×10^{-4}	0.076	2.618×10^{-2}	6.999×10^{-4}	0.039
c_{28}	1.512×10^{-3}	1.737×10^{-4}	0.103	1.345×10^{-3}	1.984×10^{-4}	0.082
c_{29}	1.550×10^{-4}	2.810×10^{-5}	0.103	1.673×10^{-4}	3.721×10^{-5}	0.082
c_{30}	7.065×10^{-2}	4.044×10^{-3}	0.103	2.363×10^{-1}	1.072×10^{-2}	0.082
c_{31}	6.694×10^{-2}	1.036×10^{-2}	0.099	9.402×10^{-2}	2.632×10^{-2}	0.108
c_{32}	5.092×10^{-4}	1.303×10^{-4}	0.099	1.150×10^{-3}	3.009×10^{-4}	0.108
c_{33}	2.767×10^{-2}	4.969×10^{-3}	0.099	6.271×10^{-3}	4.227×10^{-3}	0.108
c_{34}	3.111×10^{-2}	2.389×10^{-3}	0.087	4.709×10^{-2}	7.590×10^{-3}	0.061
c_{35}	1.008×10^{-4}	2.577×10^{-5}	0.087	3.904×10^{-5}	9.106×10^{-6}	0.061
c_{36}	2.658×10^{-1}	1.694×10^{-2}	0.087	7.654×10^{-3}	3.130×10^{-3}	0.061
c_{37}	1.177×10^{-1}	1.146×10^{-2}	0.167	4.725×10^{-2}	7.708×10^{-3}	0.107
c_{38}	6.363×10^{-2}	2.699×10^{-3}	0.167	3.570×10^{-2}	1.259×10^{-3}	0.106
c_{39}	6.107×10^{-1}	1.369×10^{-2}	0.167	3.344×10^{-1}	6.586×10^{-3}	0.106
c_{40}	1.659	1.671×10^{-2}	0.048	2.942	4.605×10^{-2}	0.016
c_{41}	1.266×10^{-2}	1.137×10^{-2}	0.049	4.349×10^{-2}	2.590×10^{-2}	0.016
c_{42}	1.375×10^{-4}	1.298×10^{-4}	0.049	6.450×10^{-5}	5.661×10^{-5}	0.016
c_{43}	5.018×10^{-3}	4.666×10^{-3}	0.127	2.916×10^{-2}	1.795×10^{-2}	0.081
c_{44}	6.514×10^{-1}	1.299×10^{-2}	0.126	4.693×10^{-1}	1.356×10^{-2}	0.081

Table 9: Continued.

Reaction	$\Delta t = 5$ unit			$\Delta t = 10$ unit		
	μ	σ	p	μ	σ	p
c_{45}	1.776×10^{-2}	3.850×10^{-3}	0.127	2.496×10^{-2}	6.222×10^{-3}	0.081
c_{46}	9.937×10^{-1}	6.743×10^{-2}	0.235	1.927	6.363×10^{-2}	0.231
c_{47}	27.171	3.838×10^{-1}	0.234	33.065	4.008×10^{-1}	0.230

Table 10: List of proteins used in inference of the reaction rates of Model 3.

Independent	Ras.GDP, Ras.GTP, Raf, Raf.I, Raf.I$_m$, Raf.A$_m$, MEK, MEK.p2, MEK$_F$, ERK, ERK.p2, ERK.p2-TF.p2, ERK.p2-RSK.A, Grb2, Shc, SOS, Grb2-SOS, Grb2-SOS$_m$, c-Fos, c-Fos.RNA, c-Fos.p, MKP, MKP.RNA
Dependent	Raf.I-Ras.GTP$_m$, MEK$_S$, ERK.p2-RSK.A-TF.p2, Shc-Grb2-SOS$_m$, Grb2$_m$, Shc-Grb2$_m$, c-Fos.DNA, MKP.DNA, EGFR, TF, GAP, PP2A, PAK, PP5, RSK

(26) ERK.p2 + MEK → ERK.p2 + MEK$_S$

(27) MEK$_S$ + Raf.A$_m$ → MEK.p2 + Raf.A$_m$

(28) ERK.p2 + Shc-Grb2-SOS$_m$ → ERK.p2 + Shc-Grb2$_m$ + SOS

(29) ERK.p2 + Grb2-SOS$_m$ → ERK.p2 + Grb2$_m$ + SOS

(30) Shc-Grb2$_m$ → Shc + Grb2

(31) Grb2$_m$ → Grb2

(32) ERK.p2 + TF → ERK.p2-TF.p2

(33) ERK.p2-TF.p2 + c-Fos.DNA → ERK.p2-TF.p2 + c-Fos.DNA + c-Fos.RNA

(34) c-Fos.RNA → c-Fos

(35) ERK.p2 + c-Fos → ERK.p2 + c-Fos.p

(36) ERK.p2-TF.p2 + MKP.DNA → ERK.p2-TF.p2 + MKP.DNA + MKP.RNA

(37) MKP.DNA → MKP

(38) MKP + ERK.p2 → MKP + ERK

(39) ERK.p2 + RSK → ERK.p2-RSK.A

(40) ERK.p2-RSK.A + TF → ERK.p2-RSK.A-TF.p2

(41) ERK.p2-RSK.A-TF.p2 + c-Fos.DNA → ERK.p2-RSK.A-TF.p2 + c-Fos.DNA + c-Fos.RNA

(42) ERK.p2-RSK.A + c-Fos → ERK.p2-RSK.A + c-Fos.p

(43) ERK.p2-RSK.A-TF.p2 + MKP.DNA → ERK.p2-RSK.A-TF.p2 + MKP.DNA + MKP.RNA

(44) MKP + ERK.p2-RSK.A → MKP + ERK + RSK

(45) ERK.p2-TF.p2 → ERK.p2 + TF

(46) ERK.p2-RSK.A → ERK.p2 + RSK

(47) ERK.p2-RSK.A-TF.p2 → ERK.p2-RSK.A + TF

Table 11: Posterior means (μ), standard deviations (σ), and acceptance ratios (p) for the estimated rate constants c_i for Model 4 from the western blotting data when $\Delta t = 5$ and $\Delta t = 10$ are time units, respectively. The estimates are based on 15,000 MCMC runs after 85,000 burn-in runs.

Reaction	$\Delta t = 5$ unit			$\Delta t = 10$ unit		
	μ	σ	p	μ	σ	p
c_1	6.275×10^{-3}	7.015×10^{-4}	0.035	5.172×10^{-3}	7.401×10^{-4}	0.082
c_2	8.505×10^{-3}	1.047×10^{-3}	0.035	5.261×10^{-3}	8.486×10^{-4}	0.082
c_3	1.076×10^{-2}	1.238×10^{-3}	0.035	1.069×10^{-2}	1.605×10^{-3}	0.082
c_4	1.034×10^{-2}	1.391×10^{-3}	0.131	6.510×10^{-3}	1.077×10^{-3}	0.101
c_5	6.702×10^{-3}	5.856×10^{-3}	0.131	1.161×10^{-2}	8.984×10^{-3}	0.101
c_6	4.523×10^{-3}	3.628×10^{-3}	0.131	3.256×10^{-3}	2.676×10^{-3}	0.101
c_7	2.405×10^{-3}	2.375×10^{-3}	0.109	5.338×10^{-3}	3.998×10^{-3}	0.166
c_8	4.204×10^{-3}	3.283×10^{-3}	0.109	6.539×10^{-3}	4.417×10^{-3}	0.166
c_9	7.614×10^{-3}	3.965×10^{-3}	0.109	2.167×10^{-2}	5.495×10^{-3}	0.166
c_{10}	1.763×10^{-2}	1.320×10^{-3}	0.129	2.370×10^{-3}	6.406×10^{-4}	0.096
c_{11}	4.517×10^{-2}	2.905×10^{-2}	0.129	3.211×10^{-2}	6.630×10^{-3}	0.096
c_{12}	8.207×10^{-2}	2.937×10^{-2}	0.129	1.061×10^{-2}	8.291×10^{-3}	0.096
c_{13}	1.918×10^{-2}	2.491×10^{-3}	0.048	2.476×10^{-2}	3.477×10^{-3}	0.081
c_{14}	1.158×10^{-4}	1.811×10^{-5}	0.048	2.473×10^{-4}	3.824×10^{-5}	0.081
c_{15}	6.026×10^{-2}	1.525×10^{-3}	0.048	5.419×10^{-2}	1.825×10^{-3}	0.081
c_{16}	6.167	1.020×10^{-1}	0.066	6.075	7.451×10^{-2}	0.106
c_{17}	5.960×10^{-1}	1.728×10^{-2}	0.067	8.225×10^{-1}	2.239×10^{-2}	0.106
c_{18}	3.473×10^{-2}	8.592×10^{-4}	0.066	2.956×10^{-2}	7.683×10^{-4}	0.106
c_{19}	8.612×10^{-2}	1.231×10^{-2}	0.123	2.013×10^{-1}	1.239×10^{-2}	0.099
c_{20}	2.490×10^{-1}	2.893×10^{-2}	0.122	1.264	1.641×10^{-2}	0.098
c_{21}	1.187×10^{-3}	1.356×10^{-4}	0.122	8.140×10^{-4}	8.502×10^{-5}	0.099
c_{22}	9.444×10^{-1}	2.770×10^{-3}	0.015	7.763×10^{-1}	8.637×10^{-4}	0.070
c_{23}	2.168×10^{-4}	9.382×10^{-6}	0.015	1.207×10^{-4}	8.157×10^{-6}	0.071
c_{24}	6.633×10^{-3}	2.008×10^{-4}	0.015	8.142×10^{-3}	8.785×10^{-4}	0.071
c_{25}	4.634×10^{-4}	1.107×10^{-5}	0.030	3.176×10^{-4}	1.465×10^{-5}	0.064
c_{26}	4.655×10^{-3}	2.715×10^{-4}	0.030	2.646×10^{-3}	1.518×10^{-4}	0.064
c_{27}	2.156×10^{-2}	6.485×10^{-4}	0.030	2.119×10^{-2}	5.370×10^{-4}	0.064
c_{28}	2.741×10^{-3}	3.529×10^{-4}	0.077	1.479×10^{-3}	2.521×10^{-4}	0.083
c_{29}	1.245×10^{-4}	3.052×10^{-5}	0.077	6.407×10^{-5}	1.792×10^{-5}	0.083
c_{30}	5.922×10^{-1}	1.230×10^{-2}	0.076	6.416×10^{-1}	5.345×10^{-3}	0.082
c_{31}	4.891×10^{-2}	1.112×10^{-2}	0.081	4.869×10^{-2}	1.459×10^{-2}	0.058
c_{32}	5.505×10^{-4}	1.569×10^{-4}	0.081	1.088×10^{-4}	6.024×10^{-5}	0.058
c_{33}	3.075×10^{-3}	2.056×10^{-3}	0.081	5.887×10^{-3}	1.378×10^{-3}	0.058
c_{34}	1.558×10^{-2}	1.467×10^{-3}	0.042	4.978×10^{-2}	6.273×10^{-3}	0.093
c_{35}	2.073×10^{-5}	1.425×10^{-5}	0.042	3.354×10^{-5}	1.084×10^{-5}	0.093
c_{36}	1.284×10^{-3}	7.211×10^{-4}	0.042	5.664×10^{-3}	1.168×10^{-3}	0.093
c_{37}	1.236×10^{-1}	1.241×10^{-2}	0.134	9.668×10^{-2}	1.080×10^{-2}	0.064
c_{38}	1.109×10^{-2}	1.147×10^{-3}	0.133	2.377×10^{-2}	1.442×10^{-3}	0.064
c_{39}	2.599×10^{-1}	6.477×10^{-3}	0.133	1.124×10^{-1}	5.128×10^{-3}	0.064
c_{40}	5.226×10^{-1}	9.351×10^{-3}	0.005	1.008	5.664×10^{-2}	0.083
c_{41}	2.507×10^{-2}	1.084×10^{-2}	0.005	7.539×10^{-3}	5.025×10^{-3}	0.083
c_{42}	1.118×10^{-4}	4.154×10^{-5}	0.005	1.158×10^{-4}	8.790×10^{-5}	0.083
c_{43}	7.870×10^{-3}	4.093×10^{-3}	0.012	5.404×10^{-3}	4.543×10^{-3}	0.051
c_{44}	1.206×10^{-1}	3.997×10^{-3}	0.012	1.013×10^{-1}	1.766×10^{-2}	0.051

Reaction	$\Delta t = 5$ unit			$\Delta t = 10$ unit		
	μ	σ	p	μ	σ	p
c_{45}	1.165×10^{-2}	2.332×10^{-3}	0.012	6.044×10^{-3}	1.799×10^{-3}	0.051
c_{46}	2.027×10^{-1}	1.903×10^{-2}	0.011	4.904×10^{-1}	4.001×10^{-2}	0.066
c_{47}	16.171	3.979×10^{-1}	0.011	15.218	6.651×10^{-1}	0.065
c_{48}	3.350×10^{-3}	6.139×10^{-4}	0.011	3.205×10^{-3}	7.732×10^{-4}	0.066

E. Description of Model 4 and Related Estimations

In Model 4, apart from the additional reaction of the EGFR degradation as the 48th reaction, we apply the same list of reactions presented for Model 3. Similarly we use the same list of substrates stated for Model 3. But in this case the receptor of the EGF protein (EGFR) is grouped with the independent, rather than dependent, species as found in Model 2. The estimates of rate constants and corresponding statistics are given in Table 11.

F. Description of Model 5 and Related Estimations

We take the following list of reactions to infer the rate constants of Model 5. Apart from those reactions of activations, we list all associated degradations of the proteins after their dissociations in Table 12. Table 14 summarizes the estimated model parameters merely via the following reactions list. On the other hand, the substrates used in inference are listed in Table 13.

(1) Grb2 + SOS $\rightarrow$ Grb2-SOS

(2) EGFR + Shc $\rightarrow$ EGFR + Shc$_m$

(3) EGFR + Grb2-SOS $\rightarrow$ EGFR + Grb2-SOS$_m$

(4) Shc$_m$ + Grb2-SOS$_m$ $\rightarrow$ Shc-Grb2-SOS$_m$

(5) Shc-Grb2-SOS$_m$ $\rightarrow$ Shc$_m$ + Grb2-SOS$_m$

(6) Shc$_m$ $\rightarrow$ Shc

(7) Grb2-SOS$_m$ $\rightarrow$ Grb2-SOS

(8) Grb2-SOS $\rightarrow$ Grb2 + SOS

(9) Shc-Grb2-SOS$_m$ + Ras.GDP $\rightarrow$ Shc-Grb2-SOS$_m$ + Ras.GTP

(10) Grb2-SOS$_m$ + Ras.GDP $\rightarrow$ Grb2-SOS$_m$ + Ras.GTP

(11) Ras.GTP $\rightarrow$ Ras.GDP

(12) GAP + Ras.GTP $\rightarrow$ GAP + Ras.GDP

(13) Raf + PP2A $\rightarrow$ Raf.I + PP2A

(14) Raf.I + Ras.GTP $\rightarrow$ Raf.I$_m$ + Ras.GTP

(15) Raf.I$_m$ + Ras.GTP $\rightarrow$ Raf.I-Ras.GTP$_m$

(16) Raf.I-Ras.GTP$_m$ + PAK $\rightarrow$ Raf.A-Ras.GTP$_m$+ PAK

(17) Raf.A-Ras.GTP$_m$ $\rightarrow$ Raf.A$_m$ + Ras.GTP

(18) PP5 + Raf.A$_m$ $\rightarrow$ PP5 + Raf.I$_m$

Table 12: List of degradations after dissociations of proteins used in Model 5.

EGFR $\rightarrow \emptyset$

Grb2 $\rightarrow \emptyset$

SOS $\rightarrow \emptyset$

Shc $\rightarrow \emptyset$

Ras.GDP $\rightarrow \emptyset$

GAP $\rightarrow \emptyset$

Raf $\rightarrow \emptyset$

Raf.I $\rightarrow \emptyset$

Raf.A$_m$ $\rightarrow \emptyset$

PP2A $\rightarrow \emptyset$

PAK $\rightarrow \emptyset$

PP5 $\rightarrow \emptyset$

MEK $\rightarrow \emptyset$

MEK$_F$ $\rightarrow \emptyset$

MEK$_S$ $\rightarrow \emptyset$

MEK.p2 $\rightarrow \emptyset$

RKIP $\rightarrow \emptyset$

Raf.I-RKIP $\rightarrow \emptyset$

PKC $\rightarrow \emptyset$

ERK $\rightarrow \emptyset$

ERK.p1 $\rightarrow \emptyset$

ERK.p2 $\rightarrow \emptyset$

TF $\rightarrow \emptyset$

c-Fos.RNA $\rightarrow \emptyset$

c-Fos $\rightarrow \emptyset$

MKP.RNA $\rightarrow \emptyset$

MKP $\rightarrow \emptyset$

c-Fos.p $\rightarrow \emptyset$

RSK $\rightarrow \emptyset$

Table 13: List of proteins used in inference of the reaction rates of Model 5.

Independent	Ras.GDP, Ras.GTP, Raf, Raf.I, Raf.A$_m$, Raf.A-Ras.GTP$_m$, Raf.I-RKIP, Raf.I-RKIP$_m$, Raf.I-RKIP-Ras.GTP$_m$, MEK, MEK$_F$, MEK$_S$, MEK.p2, MEK-RKIP, MEK$_F$-RKIP, MEK$_S$-RKIP, ERK, ERK.p1, ERK.p2, ERK.p2-TF.p2, ERK.p2-RSK.A, Grb2, Shc, Shc$_m$, SOS, Grb2-SOS, Grb2-SOS$_m$, c-Fos, c-Fos.RNA, c-Fos.p, MKP, MKP.RNA, EGFR, RKIP
Dependent	Raf.I$_m$, Raf.I-Ras.GTP$_m$, MEK.p2-RKIP, ERK.p2-RSK.A-TF.p2, Grb2$_m$, Shc-Grb2-SOS$_m$, Shc-Grb2$_m$, MKP.DNA, TF, GAP, PP2A, PAK, PP5, PKC, RSK, c-Fos.DNA, RKIP.p

Table 14: Posterior means (μ), standard deviations (σ), and acceptance ratios (p) for the estimated rate constants c_i for Model 5 from the western blotting data when $\Delta t = 5$ and $\Delta t = 10$ are time units, respectively. The estimates are based on 15,000 MCMC runs after 85,000 burn-in runs.

Reaction	$\Delta t = 5$ unit			$\Delta t = 10$ unit		
	μ	σ	p	μ	σ	p
c_1	1.863×10^{-2}	2.259×10^{-3}	0.054	1.081×10^{-2}	1.429×10^{-3}	0.056
c_2	1.049×10^{-2}	1.398×10^{-3}	0.054	9.110×10^{-3}	1.413×10^{-3}	0.056
c_3	2.290×10^{-2}	2.565×10^{-3}	0.054	1.684×10^{-2}	2.156×10^{-3}	0.056
c_4	1.088×10^{-2}	1.552×10^{-3}	0.103	1.265×10^{-2}	2.397×10^{-3}	0.159
c_5	1.195×10^{-2}	9.738×10^{-3}	0.103	2.233×10^{-2}	1.618×10^{-2}	0.159
c_6	4.142×10^{-3}	3.592×10^{-3}	0.103	9.326×10^{-3}	7.210×10^{-3}	0.159
c_7	4.841×10^{-3}	4.184×10^{-3}	0.122	4.965×10^{-3}	4.746×10^{-3}	0.058
c_8	2.631×10^{-3}	2.500×10^{-3}	0.122	4.854×10^{-3}	4.021×10^{-3}	0.058
c_9	1.196×10^{-1}	4.568×10^{-3}	0.121	2.617×10^{-2}	3.437×10^{-3}	0.058
c_{10}	3.140×10^{-4}	3.268×10^{-4}	0.085	7.943×10^{-4}	6.569×10^{-4}	0.047
c_{11}	1.028×10^{-1}	1.711×10^{-2}	0.085	2.136×10^{-2}	8.948×10^{-3}	0.047
c_{12}	2.822×10^{-2}	1.665×10^{-2}	0.085	1.347×10^{-2}	9.215×10^{-3}	0.047
c_{13}	4.479×10^{-2}	6.623×10^{-3}	0.067	3.345×10^{-2}	7.950×10^{-3}	0.083
c_{14}	3.440×10^{-4}	3.996×10^{-5}	0.067	2.313×10^{-4}	5.119×10^{-5}	0.083
c_{15}	8.938×10^{-2}	1.747×10^{-3}	0.067	4.783×10^{-2}	1.240×10^{-3}	0.083
c_{16}	10.798	1.331×10^{-1}	0.099	6.339	1.025×10^{-1}	0.082
c_{17}	1.423	4.093×10^{-2}	0.099	7.644×10^{-1}	4.013×10^{-2}	0.082
c_{18}	7.969×10^{-2}	1.293×10^{-3}	0.100	3.156×10^{-2}	8.331×10^{-4}	0.082
c_{19}	2.100×10^{-1}	3.106×10^{-2}	0.174	1.387×10^{-1}	1.687×10^{-2}	0.151
c_{20}	1.077	2.345×10^{-2}	0.172	7.927×10^{-1}	2.210×10^{-2}	0.151
c_{21}	1.347×10^{-2}	4.106×10^{-4}	0.173	3.572×10^{-3}	1.593×10^{-4}	0.151
c_{22}	1.501×10^{-2}	4.068×10^{-3}	0.044	8.299×10^{-3}	2.298×10^{-3}	0.021
c_{23}	6.456×10^{-5}	1.406×10^{-5}	0.044	1.800×10^{-5}	7.040×10^{-6}	0.021
c_{24}	2.835×10^{-1}	7.494×10^{-3}	0.044	6.827×10^{-2}	2.363×10^{-3}	0.021
c_{25}	1.413×10^{-4}	2.416×10^{-5}	0.066	5.732×10^{-5}	2.288×10^{-5}	0.053
c_{26}	2.421×10^{-4}	4.300×10^{-5}	0.066	1.039×10^{-4}	3.061×10^{-5}	0.053
c_{27}	8.674×10^{-2}	8.586×10^{-3}	0.066	2.961×10^{-1}	9.671×10^{-3}	0.053
c_{28}	2.198×10^{-3}	4.130×10^{-4}	0.056	6.217×10^{-4}	8.434×10^{-5}	0.001
c_{29}	1.642×10^{-4}	4.870×10^{-5}	0.056	1.869×10^{-5}	9.271×10^{-7}	0.001
c_{30}	8.056×10^{-3}	2.117×10^{-4}	0.056	1.682×10^{-3}	4.059×10^{-5}	0.001
c_{31}	1.677×10^{-1}	3.631×10^{-2}	0.132	2.503×10^{-1}	3.450×10^{-2}	0.088
c_{32}	1.327×10^{-2}	4.927×10^{-4}	0.131	5.607×10^{-3}	7.904×10^{-4}	0.089
c_{33}	4.389×10^{-2}	6.062×10^{-3}	0.132	2.056×10^{-2}	5.418×10^{-3}	0.089
c_{34}	2.663×10^{-2}	5.283×10^{-3}	0.078	1.750×10^{-2}	4.730×10^{-3}	0.053
c_{35}	2.251×10^{-1}	3.812×10^{-2}	0.078	1.992	7.295×10^{-2}	0.052
c_{36}	2.270×10^{-2}	4.079×10^{-3}	0.078	1.182×10^{-2}	2.967×10^{-3}	0.053
c_{37}	2.059×10^{-2}	3.057×10^{-3}	0.092	1.122×10^{-2}	2.647×10^{-3}	0.051
c_{38}	12.505	2.846×10^{-1}	0.091	8.464	2.842×10^{-1}	0.051
c_{39}	12.649	2.135×10^{-1}	0.091	7.507	4.391×10^{-1}	0.051
c_{40}	6.416×10^{-2}	8.716×10^{-4}	0.048	6.051×10^{-3}	4.100×10^{-4}	0.031
c_{41}	1.381×10^{-3}	3.770×10^{-5}	0.048	3.411×10^{-4}	1.213×10^{-5}	0.031
c_{42}	4.706×10^{-1}	3.624×10^{-3}	0.048	9.657×10^{-1}	1.126×10^{-3}	0.031
c_{43}	1.138×10^{-1}	4.391×10^{-3}	0.073	2.301×10^{-2}	1.193×10^{-3}	0.034
c_{44}	1.805×10^{-2}	6.337×10^{-4}	0.073	5.759×10^{-3}	2.667×10^{-4}	0.033

Table 14: Continued.

Reaction	$\Delta t = 5$ unit			$\Delta t = 10$ unit		
	μ	σ	p	μ	σ	p
c_{45}	6.633×10^{-4}	2.036×10^{-5}	0.074	1.478×10^{-4}	6.480×10^{-6}	0.033
c_{46}	2.627×10^{-3}	3.037×10^{-4}	0.092	2.230×10^{-3}	3.253×10^{-4}	0.073
c_{47}	4.044×10^{-4}	5.053×10^{-5}	0.092	1.440×10^{-4}	3.287×10^{-5}	0.073
c_{48}	7.617×10^{-1}	1.186×10^{-2}	0.092	5.006×10^{-1}	1.152×10^{-2}	0.073
c_{49}	2.143×10^{-1}	2.466×10^{-2}	0.109	7.972×10^{-2}	1.902×10^{-2}	0.119
c_{50}	1.464×10^{-3}	2.639×10^{-4}	0.109	1.393×10^{-3}	3.305×10^{-4}	0.119
c_{51}	2.311×10^{-2}	4.525×10^{-3}	0.109	7.878×10^{-3}	3.686×10^{-3}	0.119
c_{52}	2.390×10^{-2}	2.650×10^{-3}	0.080	1.318×10^{-2}	2.303×10^{-3}	0.066
c_{53}	1.566×10^{-4}	2.367×10^{-5}	0.080	5.894×10^{-5}	2.746×10^{-5}	0.066
c_{54}	5.548×10^{-3}	3.231×10^{-3}	0.080	2.575×10^{-3}	1.542×10^{-3}	0.066
c_{55}	4.864×10^{-2}	6.391×10^{-3}	0.181	3.724×10^{-2}	8.192×10^{-3}	0.154
c_{56}	4.417×10^{-2}	1.729×10^{-3}	0.180	2.874×10^{-2}	1.886×10^{-3}	0.154
c_{57}	3.768×10^{-1}	6.189×10^{-3}	0.180	2.040×10^{-1}	7.410×10^{-3}	0.153
c_{58}	5.887	8.624×10^{-2}	0.063	9.909×10^{-1}	9.228×10^{-2}	0.009
c_{59}	2.541×10^{-2}	1.612×10^{-2}	0.063	4.889×10^{-2}	1.827×10^{-2}	0.009
c_{60}	3.079×10^{-4}	2.914×10^{-4}	0.063	2.358×10^{-4}	1.688×10^{-4}	0.009
c_{61}	2.863×10^{-2}	1.563×10^{-2}	0.156	2.475×10^{-2}	9.511×10^{-3}	0.045
c_{62}	7.608×10^{-1}	3.301×10^{-2}	0.156	1.042×10^{-1}	1.195×10^{-2}	0.045
c_{63}	3.469×10^{-2}	6.047×10^{-3}	0.156	3.585×10^{-2}	7.826×10^{-3}	0.045
c_{64}	4.354	1.214×10^{-1}	0.076	6.045×10^{-1}	4.787×10^{-2}	0.022
c_{65}	35.422	5.094×10^{-1}	0.076	20.302	8.049×10^{-1}	0.022
c_{66}	1.784×10^{-2}	2.261×10^{-3}	0.076	7.750×10^{-3}	1.443×10^{-3}	0.022

(19) $\text{Raf.I}_m \rightarrow \text{Raf}$

(20) $\text{Raf.I-Ras.GTP}_m \rightarrow \text{Raf.I}_m + \text{Ras.GTP}$

(21) $\text{Raf.A}_m + \text{MEK} \rightarrow \text{Raf.A}_m + \text{MEK.p2}$

(22) $\text{PAK} + \text{MEK} \rightarrow \text{PAK} + \text{MEK}_F$

(23) $\text{MEK}_F + \text{Raf.A}_m \rightarrow \text{MEK.p2} + \text{Raf.A}_m$

(24) $\text{Raf.I} + \text{RKIP} \rightarrow \text{Raf.I-RKIP}$

(25) $\text{Raf.I-RKIP} + \text{Ras.GTP} \rightarrow \text{Raf.I-RKIP}_m + \text{Ras.GTP}$

(26) $\text{Raf.I-RKIP}_m + \text{Ras.GTP} \rightarrow \text{Raf.I-RKIP-Ras.GTP}_m$

(27) $\text{MEK} + \text{RKIP} \rightarrow \text{MEK-RKIP}$

(28) $\text{MEK}_F + \text{RKIP} \rightarrow \text{MEK}_F\text{-RKIP}$

(29) $\text{MEK}_S + \text{RKIP} \rightarrow \text{MEK}_S\text{-RKIP}$

(30) $\text{MEK.p2} + \text{RKIP} \rightarrow \text{MEK.p2-RKIP}$

(31) $\text{PKC} + \text{Raf.I-RKIP} \rightarrow \text{PKC} + \text{Raf.I} + \text{RKIP.p}$

(32) $\text{ERK.p2} + \text{Raf.I-RKIP} \rightarrow \text{ERK.p2} + \text{Raf.I} + \text{RKIP.p}$

(33) $\text{Raf.I-RKIP-Ras.GTP}_m \rightarrow \text{Raf.I-RKIP}_m + \text{Ras.GTP}$

(34) $\text{Raf.I-RKIP}_m \rightarrow \text{Raf.I-RKIP}$

(35) $\text{MEK-RKIP} \rightarrow \text{MEK} + \text{RKIP}$

(36) MEK_F-RKIP $\rightarrow$ MEK_F + RKIP

(37) MEK_S-RKIP $\rightarrow$ MEK_S + RKIP

(38) MEK.p2-RKIP $\rightarrow$ MEK.p2 + RKIP

(39) RKIP.p $\rightarrow$ RKIP

(40) MEK.p2 + ERK $\rightarrow$ MEK.p2 + ERK.p1

(41) MEK.p2 + ERK.p1 $\rightarrow$ MEK.p2 + ERK.p2

(42) MEK.p2-RKIP + ERK $\rightarrow$ MEK.p2-RKIP + ERK.p1

(43) MEK.p2-RKIP + ERK.p1 $\rightarrow$ MEK.p2-RKIP + ERK.p2

(44) ERK.p2 + MEK $\rightarrow$ ERK.p2 + MEK_S

(45) MEK_S + Raf.A$_m$ $\rightarrow$ MEK.p2 + Raf.A$_m$

(46) ERK.p2 + Shc-Grb2-SOS$_m$ $\rightarrow$ ERK.p2 + Shc-Grb2$_m$ + SOS

(47) ERK.p2 + Grb2-SOS$_m$ $\rightarrow$ ERK.p2 + Grb2$_m$ + SOS

(48) Shc-Grb2$_m$ $\rightarrow$ Shc + Grb2

(49) Grb2$_m$ $\rightarrow$ Grb2

(50) ERK.p2 + TF $\rightarrow$ ERK.p2-TF.p2

(51) ERK.p2-TF.p2 + c-Fos.DNA $\rightarrow$ ERK.p2-TF.p2 + c-Fos.DNA + c-Fos.RNA

(52) c-Fos.RNA $\rightarrow$ c-Fos

(53) ERK.p2 + c-Fos $\rightarrow$ ERK.p2 + c-Fos.p

(54) ERK.p2-TF.p2 + MKP.DNA $\rightarrow$ ERK.p2-TF.p2 + MKP.DNA + MKP.RNA

(55) MKP.DNA $\rightarrow$ MKP

(56) MKP + ERK.p2 $\rightarrow$ MKP + ERK

(57) ERK.p2 + RSK $\rightarrow$ ERK.p2-RSK.A

(58) ERK.p2-RSK.A + TF $\rightarrow$ ERK.p2-RSK.A-TF.p2

(59) ERK.p2-RSK.A-TF.p2 + c-Fos.DNA $\rightarrow$ ERK.p2-RSK.A-TF.p2 + c-Fos.DNA + c-Fos.RNA

(60) ERK.p2-RSK.A + c-Fos $\rightarrow$ ERK.p2-RSK.A + c-Fos.p

(61) ERK.p2-RSK.A-TF.p2 + MKP.DNA $\rightarrow$ ERK.p2-RSK.A-TF.p2 + MKP.DNA + MKP.RNA

(62) MKP + ERK.p2-RSK.A $\rightarrow$ MKP + ERK + RSK

(63) ERK.p2-TF.p2 $\rightarrow$ ERK.p2 + TF

(64) ERK.p2-RSK.A $\rightarrow$ ERK.p2 + RSK

(65) ERK.p2-RSK.A-TF.p2 $\rightarrow$ ERK.p2-RSK.A + TF

(66) EGFR $\rightarrow$ $\emptyset$.

G. List of Protein States Used in the MAPK Pathway

We use the following substrates in the description of the MAPK pathway with different models.

(1) Ras protein states.

  (a) Ras.GDP: the inactive Ras protein near the cell membrane.
  (b) Ras.GTP: the active Ras near the cell membrane.

(2) Raf protein states

  (a) Raf: the inactive and nonphosphorylated Raf protein in the cytosol.
  (b) Raf.A$_m$: the active Raf phosphorylated on the S338 and the S471 binding sites near the cell membrane.
  (c) Raf.A-Ras.GTP$_m$: the complex of the active Raf and Ras.GTP near the cell membrane.
  (d) Raf.I: the inactive Raf phosphorylated on the S259 binding site in the cytosol.
  (e) Raf.I$_m$: the inactive Raf phosphorylated on the S259 binding site and recruited from the cytosol to the cell membrane by the Ras.GTP protein.
  (f) Raf.I-Ras.GTP$_m$: the complex of the inactive Raf and Ras.GTP near the cell membrane.
  (g) Raf.I-RKIP: the complex of the inactive Raf and RKIP, whose binding site is S338, in the cytosol.
  (h) Raf.I-RKIP$_m$: the complex of the inactive Raf and RKIP which is recruited to the membrane by the Ras.GTP protein.
  (i) Raf.I-RKIP-Ras.GTP$_m$: the complex of the inactive Raf, RKIP, and Ras.GTP near the cell membrane.

(3) MEK protein states.

  (a) MEK: the inactive and nonphosphorylated MEK protein in the cytosol.
  (b) MEK$_F$: the inactive MEK in the cytosol which is monophosphorylated by the activator PAK on the S298 binding site.
  (c) MEK$_S$: the inactive MEK in the cytosol which is monophosphorylated by the active ERK on the T292 binding site.
  (d) MEK.p2: the double-phosphorylated MEK (active MEK) on the S218 and S222 binding sites in the cytosol.
  (e) MEK-RKIP: the complex of the MEK and RKIP proteins in the cytosol.
  (f) MEK$_F$-RKIP: the complex of the MEK$_F$ and RKIP proteins in the cytosol.
  (g) MEK$_S$-RKIP: the complex of the MEK$_S$ and RKIP proteins in the cytosol.
  (h) MEK.p2-RKIP: the complex of the active MEK.p2 and RKIP proteins in the cytosol.

(4) ERK protein states.

  (a) ERK: the inactive and nonphosphorylated ERK protein in the cytosol.
  (b) ERK.p1: the inactive, monophosphorylated ERK in the cytosol.

(c) ERK.p2: the double-phosphorylated ERK (active ERK) in the cytosol.

(d) ERK.p2-RSK.A: the complex of the active ERK and active RSK, which is activated by the ERK protein, in the nucleus.

(e) ERK.p2-RSK.A-TF.p2: the complex of the active ERK, active RSK, and double-phosphorylated transcription factor in the nucleus.

(f) ERK.p2-TF.p2: the complex of the active ERK and a transcription factor (like Elk or SAP proteins), which is double-phosphorylated by the active ERK, in the nucleus.

(5) Grb2, Shc, and SOS protein states.

(a) Grb2: A protein in the cytosol.

(b) $Grb2_m$: the Grb2 protein near the cell membrane after the dissociation of the SOS protein by the active ERK.

(c) Grb2-SOS: the complex of the Grb2 and SOS proteins in the cytosol.

(d) $Grb2-SOS_m$: the complex of the Grb2 and SOS proteins near the cell membrane, where it is able to activate the Ras protein.

(e) Shc: a protein in the cytosol.

(f) Shc_m: the Shc protein near the cell membrane after the activation of the EGF receptor.

(g) $Shc-Grb2_m$: the complex of the Shc and Grb2 proteins near the cell membrane after the dissociation of the SOS protein by the active ERK.

(h) $Shc-Grb2-SOS_m$: the complex of the Shc, Grb2, and SOS proteins near the cell membrane, where it is able to activate the Ras protein.

(i) SOS: a protein, which is an exchange factor, in the cytosol.

(6) c-Fos and MKP protein states.

(a) c-Fos: a protein in the nucleus.

(b) c-Fos.DNA: the gene sequence of the c-Fos protein.

(c) c-Fos.p: the c-Fos gene phosphorylated by the ERK protein.

(d) c-Fos.RNA: the transcription of the c-Fos gene into the messenger RNA (mRNA).

(e) MKP: a protein in the cytosol.

(f) MKP.DNA: the gene sequence of the MKP protein.

(g) MKP.RNA: the transcription of the MKP gene into mRNA.

(7) Other proteins

(a) EGF: a protein which triggers the activation of the pathway by attaching its receptor (EGFR) in the cell membrane.

(b) EGFR: a receptor that is equated with *activated tyrosine kinase receptors.*

(c) GAP: a protein near the cell membrane.

(d) PAK: a protein near the cell membrane.

(e) PKC: a protein in cytosol.

(f) PP2A: a protein near the cell membrane or in cytosol.

(g) PP5: a protein near the cell membrane.

(h) RKIP: a protein in the cytosol.

(i) RKIP.p: the RKIP monophosphorylated either by the PKC or ERK protein on the binding sites S153 and S99, respectively.

(j) RSK: an inactive protein in the cytosol.

(k) RSK.A: the active RSK, which is activated by the ERK.p2 protein.

(l) TF: a transcription factor (like Elk or SAP proteins), which will be double-phosphorylated by the active ERK, in the nucleus.

Acknowledgments

References

[1] W. Kolch, "Meaningful relationships: the regulation of the Ras/Raf/MEK/ERK pathway by protein interactions," *Biochemical Journal*, vol. 351, no. 2, pp. 289–305, 2000.

[2] W. Kolch, M. Calder, and D. Gilbert, "When kinases meet mathematics: the systems biology of MAPK signalling," *FEBS Letters*, vol. 579, no. 8, pp. 1891–1895, 2005.

[3] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, 2008.

[4] S. Reinker, R. M. Altman, and J. Timmer, "Parameter estimation in stochastic biochemical reactions," *IEE Proceedings*, vol. 153, no. 4, pp. 168–178, 2006.

[5] A. Golightly and D. J. Wilkinson, "Bayesian inference for nonlinear multivariate diffusion models observed with error," *Computational Statistics & Data Analysis*, vol. 52, no. 3, pp. 1674–1693, 2008.

[6] R. J. Boys, D. J. Wilkinson, and T. B. L. Kirkwood, "Bayesian inference for a discretely observed stochastic kinetic model," *Statistics and Computing*, vol. 18, no. 2, pp. 125–135, 2008.

[7] J. R. Banga, "Optimization in computational systems biology," *BMC Systems Biology*, vol. 2, article 47, pp. 1–7, 2008.

[8] J. M. Bower and H. Bolouri, *Computational Modelling of Genetic and Biochemical Networks*, Massachusetts Institute of Technology, 2nd edition, 2001.

[9] A. Golightly and D. J. Wilkinson, "Bayesian inference for stochastic kinetic models using a diffusion approximation," *Biometrics*, vol. 61, no. 3, pp. 781–788, 2005.

[10] D. J. Wilkinson, *Stochastic Modelling for Systems Biology*, Chapman & Hall/CRC, Boca Raton, Fla, USA, 2006.

[11] N. G. VanKampen, *Stochastic Processes in Physics and Chemistry*, Elsevier, 1981.

[12] H. Risken, *The Fokker-Planck Equation*, vol. 18, Springer, Berlin, Germany, 1984.

[13] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *Journal of Physical Chemistry*, vol. 81, no. 25, pp. 2340–2361, 1977.

[14] D. J. Wilkinson, "Parallel Bayesian computation," in *Handbook of parallel computing and statistics*, vol. 184, pp. 477–508, Chapman & Hall/CRC, Boca Raton, Fla, USA, 2006.

[15] R. Khanin, V. Vinciotti, V. Mersinias, C. P. Smith, and E. Wit, "Statistical reconstruction of transcription factor activity using Michaelis-Menten kinetics," *Biometrics*, vol. 63, no. 3, pp. 816–823, 2007.

[16] G. Sanguinetti, M. Rattray, and N. D. Lawrence, "A probabilistic dynamical model for quantitative inference of the regulatory mechanism of transcription," *Bioinformatics*, vol. 22, no. 14, pp. 1753–1759, 2006.

[17] H. Bolouri, *Computational Modelling of Gene Regulatory Networks—A Primer*, Imperial College Press, 2008.

[18] V. Vyshemirsky, M. Girolami, A. Gormand, and W. Kolch, "A Bayesian analysis of the ERK signalling pathway," Tech. Rep. TR-2006-227, University of Glasgow, 2006.

[19] J. Vera, J. Bachmann, A. C. Pfeifer et al., "A systems biology approach to analyse amplification in the JAK2-STAT5 signalling pathway," *BMC Systems Biology*, vol. 2, no. 38, pp. 1–13, 2008.

[20] K. Yeung, T. Seitz, S. Li et al., "Suppression of Raf-1 kinase activity and MAP kinase signalling by RKIP," *Nature*, vol. 401, pp. 173–177, 1999.

[21] V. Purutçuoğlu and E. Wit, "Bayesian inference for the MAPK/ERK pathway by considering the dependency of the kinetic parameters," *Bayesian Analysis*, vol. 3, no. 4, pp. 851–886, 2008.

[22] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer, Berlin, Germany, 3rd edition, 1993.

[23] B. Eraker, "MCMC analysis of diffusion models with application to finance," *Journal of Business & Economic Statistics*, vol. 19, no. 2, pp. 177–191, 2001.

[24] B. P. Carlin and T. A. Louis, *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman & Hall/CRC, 2nd edition, 2000.

[25] S. Chib, F. Nardari, and N. Shephard, "Analysis of high dimensional multivariate stochastic volatility models," *Journal of Econometrics*, vol. 134, no. 2, pp. 341–371, 2006.

[26] D. Hames and N. Hooper, *Instant Notes: Biochemistry*, BIOS Scientific Publishers, 2000.

[27] R. L. Switzer and L. F. Garrity, *Experimental Biochemistry*, W.H. Freeman and Company, New York, NY, USA, 3rd edition, 2003.

[28] N. Fedoroff and W. Fontana, "Genetic networks: small numbers of big molecules," *Science*, vol. 297, no. 5584, pp. 1129–1131, 2002.

[29] D. A. Hume, "Probability in transcriptional regulation and its implications for leukocyte differentiation and inducible gene expression," *Blood*, vol. 96, no. 7, pp. 2323–2328, 2000.

[30] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data*, John Wiley & Sons, New York, NY, USA, 1990.

[31] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, Chapman & Hall/CRC, 2nd edition, 2004.

[32] S. Park, O. Rath, S. Beach et al., "Regulation of RKIP binding to the N-region of the Raf-1 kinase," *FEBS Letters*, vol. 580, no. 27, pp. 6405–6412, 2006.

[33] W. Kolch, "MAPK signalling networks," Oncogene Course Notes, 2005.

[34] W. Kolch, "MAPK signalling networks," RTK Consortium Presentation, 2005.

[35] R. J. Orton, O. E. Sturm, V. Vyshemirsky, M. Calder, D. R. Gilbert, and W. Kolch, "Computational modelling of the receptor-tyrosine-kinase-activated MAPK pathway," *Biochemical Journal*, vol. 392, no. 2, pp. 249–261, 2005.

[36] B. Schoeberl, C. Eichler-Jonsson, E. D. Gilles, and G. Müller, "Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors," *Nature Biotechnology*, vol. 20, no. 4, pp. 370–375, 2002.

[37] B. N. Kholodenko, O. V. Demin, G. Moehren, and J. B. Hoek, "Quantification of short term signaling by the epidermal growth factor receptor," *Journal of Biological Chemistry*, vol. 274, no. 42, pp. 30169–30181, 1999.

[38] F. A. Brightman and D. A. Fell, "Differential feedback regulation of the MAPK cascade underlies the quantitative differences in EGF and NGF signalling in PC12 cells," *FEBS Letters*, vol. 482, no. 3, pp. 169–174, 2000.

[39] O. Elerian, S. Chib, and N. Shephard, "Likelihood inference for discretely observed nonlinear diffusions," *Econometrica*, vol. 69, no. 4, pp. 959–993, 2001.

[40] A. Golightly and D. J. Wilkinson, "Bayesian sequential inference for stochastic kinetic biochemical network models," *Journal of Computational Biology*, vol. 13, no. 3, pp. 838–851, 2006.

[41] V. Purutçuoglu, "Inference of the stochastic MAPKpathway by modified diffusion bridge method," *Central European Journal of Operations Research*, pp. 1–15, 2012.

[42] M. Li, "Fractal time series—a tutorial review," *Mathematical Problems in Engineering*, vol. 2010, Article ID 157264, 26 pages, 2010.

[43] S. Havlin, S. V. Buldyrev, A. L. Goldberger et al., "Fractals in biology and medicine," *Chaos, Solitons and Fractals*, vol. 6, pp. 171–201, 1995.

[44] D. Craciun, A. Isvoran, and N. M. Avram, "Long range correlation of hydrophilicity and flexibility along the calcium binding protein chains," *Physica A*, vol. 388, no. 21, pp. 4609–4618, 2009.

[45] S. C. Lim and S. V. Muniandy, "On some possible generalizations of fractional Brownian motion," *Physics Letters. A*, vol. 266, no. 2-3, pp. 140–145, 2000.

[46] S. C. Kou and X. S. Xie, "Generalized langevin equation with fractional gaussian noise: subdiffusion within a single protein molecule," *Physical Review Letters*, vol. 93, no. 18, pp. 1806031–4, 2004.

[47] M. Li, W. Zhao, and S. Chen, "Mbm-based scalings of traffic propagated in internet," *Mathematical Problems in Engineering*, vol. 2011, Article ID 389803, 21 pages, 2011.

[48] M. Li and W. Zhao, "Quantitatively investigating the locally weak stationarity of modified multifractional gaussian noise," *Physics A*, vol. 391, no. 24, pp. 6268–6278, 2012.

[49] R. F. Peltier and J. Levy-Vehel, "A new method for estimating the parameter of fractional brownian motion," Tech. Rep. RR, 2696:1–4, INRIA, 1994.

[50] J.-P. Bouchaud and A. Georges, "Anomalous diffusion in disordered media: statistical mechanisms, models and physical applications," *Physics Reports*, vol. 195, no. 4-5, pp. 127–293, 1990.

[51] D. Gamerman and H. F. Lopes, *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Chapman & Hall/CRC, Boca Raton, Fla, USA, 2nd edition, 2006.