International Institute for
Applied Systems Analysis
Schlossplatz 1
A-2361 Laxenburg, Austria

Tel: +43 2236 807 342
Fax: +43 2236 71313
E-mail: publications@iiasa.ac.at
Web: www.iiasa.ac.at

Interim Report **IR-05-081**

Optimal Phosphorylation Step Number of Intracellular Signal Transduction Pathway

Jun Nakabayashi (nakabaya@bio-math10.biology.kyushu-u.ac.jp)
Akira Sasaki (asasascb@mbox.nc.kyushu-u.ac.jp)

Approved by

Ulf Dieckmann
Program Leader, ADN

December 2005

Contents

Optimal phosphorylation step number of intracellular signal transduction pathway

Jun Nakabayashi and Akira Sasaki

Department of Biology, Faculty of Sciences,
Kyushu University, Fukuoka 812-8581, Japan
Phone: +81-92-642-2641
Facsimile: +81-92-642-2645
Email: nakabaya@bio-math10.biology.kyushu-u.ac.jp

Abstract

Eukaryotic cells use signal transduction network to respond in specific ways to external signals from their environment. Several signal transduction pathway are composed of multi-step chemical reactions. We here theoretically study what determines the number of kinase phosphorylation steps composing of the intracellular signal transduction cascade. We examine a simple mathematical model for the association and phosphorylation process of kinases in the signal transduction cascade. We focus on the speed of signal transduction as the criterion for determining the optimal response. The present model first reveals that the initial expression level of kinase in each step of the cascade must be the same in the optimal response under the constraint of the constant total kinase concentration. The second conclusion is that the optimal step number of kinase cascade is primarily determined by the ratio of the target concentration of the final phosphorylated kinase in the cascade to that of input signal molecule, C/S. A multi-step phosphorylation can be optimal when the amplification of the final product concentration C relative to the input signal S is sufficiently large. This suggests that multi-step phosphorylation would have evolved to accelerate the speed of transduction of weak signals.

Keywords: kinase cascade, signal transduction pathway, phosphorylation, enzyme kinetics, evolution of gene network

1 Introduction

Eukaryotic cells use signal transduction networks to respond in specific ways to external signals from their environment (Table 1). Several protein-kinase-based signal transduction pathways are composed of successive kinase-phosphorylation steps. For example, MAPK pathway is composed of four steps of phosphorylation reactions. The kinase MAPKKKK, when the associated receptor perceives an environmental change, is phsphorylated and activated. The activated MAPKKKK then phosphorylates the second kinase MAPKKK, which then becomes active and phosphorylates the third kinase MAPKK. The phosphorylation and activation of the last kinase MAPK follows the activation of MAPKK, and regulates the activity of the specific transcription factors and the expression of the target genes. This regulation of the target genes causes the specific and appropriate responses in the cell for the change of environment which is initially received by receptors. In yeast MAPK cascade, Ste20p MAPKKKK that associates the seven transmembrane receptor/G-protein complex is first phosphorylated and activated in response to the stimulation of pheromone (Wu *et al.*, 1995; Leeuw *et al.*, 1998). Ste20p then phosphorylates and activates Ste11p MAPKKK (Drogen *et al.*, 2000). Ste11p activates Ste7p MAPKK, and finally Ste7p activates Fus3p MAPK (Matsuyama *et al.*, 2000). Fus3p regulates the cellular responses for mating (Gustin *et al.*, 1998; Madhani & Fink, 1998). The above summary for the cascade of signal transduction pathways might be too simple to describe the real MAPK pathway in the cells – MAPKs are dephosphorylated and hence inactivated by the specific phosphatases (Theodosiou & Ashworth, 2002); scaffold proteins that bind multiple kinases in the pathway plays an important role.

There are several hypotheses for the adaptive significance of the multi-step kinase cascade in signal transduction pathways: the reaction cascade may amplify weak signals; the successive amplification of intermediates may accelerate the speed of signal transduction; the cooperativity by multi-step reaction may tolerate against the noise in signal reception; the participation of several different kinases may allow alternative entrance points for different signals in differential regulation

of the transduction pathway, and so on. *Xenopus* MAPK cascade that regulates progesterone-induced oocyte maturation behaves like a highly cooperative enzyme – the stimulus/response curve of MAPK signal is steeper than that of Michaelis-Menten process. Either multi-step cascade or positive feedback is necessary to account for the MAPKs' switching response (Huang & Ferrell, 1996; Ferrell & Machleder, 1998). T-cells receive the signal from antigen-presenting cells (APC). T-cell receptor (TCR) on T-cells binds to a specific peptide-MHC (major histocompatibility complex) on the APC. Multi-step modification (protein-protein interaction and phosphorylation) of TCR-MHC complex occurs in the signal transduction pathway to invoke immune response. Mckeithan (1995) showed that nonlinearity due to the intermediate steps of TCR modification were necessary for highly selective (noise-resistant) signaling in TCR signal transduction.

MAPKs are activated by stresses including ischemia reperfusion, neuronal injury, osmotic shock, and exposure to UV irradiation (Karin, 1998). Cells must respond against these cytotoxic environmental changes. Klipp et al analyzed the unbranched model metabolic pathway consisting of n consecutive enzyme-catalyzed mononuclear enzymatic reactions and a series of $n-1$ intermediates, X_i (Klipp *et al.*, 2002). They assumed the final product P represented a biochemical compound whose availability is rate-limiting for the reproduction of an individual. For this linear model pathway, it was shown that wave-like enzyme profiles were optimal for a rapid substrate turnover under the constraint of limited total enzyme amount. These enzyme profiles were in close correlation with observed gene expression data. We therefor focus on the initial speed of the signal transduction as a criteria for the optimality of the signal transduction pathways. This is different from the focus of the Heinrich *et al.* (2002)'s model, in which they used a multi-step phosphorylation cascade model of signal transduction to examine the amplitude of the signal output and the duration of signaling. The problem we pursue in this paper is whether or not the multiple steps of phosphorylation in signal transduction pathway can be optimal. We first introduce a mathematical model of kinase cascade, the signal transduction system with stepwise phosphorylation like MAPK. We then study the optimization of the gene expression level of kinase that maximizes the speed of the signal transduction, and the optimal number of steps composing the cascade of phosphorylation.

2 A model for kinase cascade

We here consider a signal transduction pathway that is made up of the cascade of kinase phosphorylation/activation steps (like MAPK pathway). Each step of signal transduction cascade is composed of the reactions of association of kinases, phosphorylation of kinase in the complex, and the dephosphorylation of kinase. We adopt the following simplifying assumptions: (1) All components are homogeneously distributed in a cell – i.e., we ignore the localized distribution of components in cytological structure like cytoplasm and nucleus. (2) Each kinase has only one phosphorylation site. (3) No protein is newly synthesized through the process of the signal transduction. (4) No protein is degraded through the process of the signal transduction.

Under these assumptions, we construct the model with multi-step kinase-kinase interactions and phosphorylation/dephophorylation illustrated in Fig. 1. Signal molecule S is given from outside the cell. S interacts with the intracellular kinase E_1 to make a complex SE_1 with the association constant α_1, and the complex dissociates with the dissociation constant γ_1. E_1 in the complex SE_1 is then phosphorylated with the reaction constant β_1 to become an activated kinase (E_1p), which is subject to dephosphorylation with the rate δ. The active kinase E_1p interacts with another kinase E_2 in the downstream of the cascade with the association constant α_2 to form a complex E_1pE_2. The complex dissociates with the rate γ_2, and the kinase E_2 in the complex is phosphorylated and activated with the association constant β_2, and so forth. Generalization to arbitrary number of steps is simple: the active kinase of steps $i-1$ ($E_{i-1}p$) interacts with the downstream kinase (E_i) with the association constant α_i to form a complex ($E_{i-1}pE_i$), and the complex dissociates with the rate γ_i. The kinase E_i in the complex is phosphorylated and activated (E_ip) with the reaction constant β_i, and the activated kinase is dephosphorylated with the constant rate δ. The active kinase then interacts with the free kinase (E_{i+1}) in the further downstream step in the cascade. The reaction proceeds step by step until the last kinase is activated. The extent to which the signal transduction proceeds is represented by the activity of the last kinase. The reactions described

above can be rewritten as

$$
\begin{aligned}
&S + E_1 \underset{\gamma_1}{\overset{\alpha_1}{\rightleftharpoons}} SE_1 \\
&SE_1 + p \xrightarrow{\beta_1} S + E_1p \\
&E_1p \xrightarrow{\delta} E_1 + p
\end{aligned}
$$

for the reactions in the first step of the cascade, which describes the process until the first-step kinase (associated with the receptor) is phosphorylated by signal, and

$$
\begin{aligned}
&E_{i-1}p + E_i \underset{\gamma_i}{\overset{\alpha_i}{\rightleftharpoons}} E_{i-1}pE_i \\
&E_{i-1}pE_i + p \xrightarrow{\beta_i} E_{i-1}p + E_ip \\
&E_ip \xrightarrow{\delta} E_i + p
\end{aligned}
$$

for the reactions in the i-th step of the cascade ($i = 2, 3, \cdots, n$, where n is the number of steps making up of the cascade).

Our model is described by the time change of the concentrations x_i, y_i, z_i of free kinases E_i, the complex $E_{i-1}pE_i$, and the phosphorylated kinase E_ip in each step $i (i = 1, 2, \cdots, n)$. The signal molecules are given from outside the cell at time $t = 0$, which initiates the signal transduction. We assume that the concentration S of the signal is kept constant after $t = 0$. These yield the differential equations:

$$
\begin{aligned}
\frac{dx_i}{dt} &= -\alpha_i z_{i-1} x_i + \delta z_i + \gamma_i y_i, \\
\frac{dy_i}{dt} &= \alpha_i z_{i-1} x_i - \beta_i y_i p - \gamma_i y_i, \\
\frac{dz_i}{dt} &= \beta_i y_i p - \delta z_i - \alpha_{i+1} z_i x_{i+1} + (\beta_{i+1} p + \gamma_{i+1}) y_{i+1},
\end{aligned}
\tag{1}
$$

($i = 1, 2, \cdots, n$) with $z_0 = S$. The time course of the concentrations of each phosphorylated kinases are shown in Fig. 2. If, as expected, the phosphate concentration p is much larger than that of the kinases and complexes, the speed of destruction $\beta_i y_i p$ of the complex for a moderate y_i is much larger than its production speed $\alpha_i z_{i-1} x_i$. The concentration y_i of the complex must then be very small, and be its time derivative dy_i/dt as well. Thus $dy_i/dt = \alpha_i z_{i-1} x_i - (\gamma_i + \beta_i p) y_i \simeq 0$. This

approximation can be justified by noting that the phosphates in the cell are supplied from ATP, and that its concentration is sufficiently large and the substrate bound to active kinase is rapidly phosphorylated. We can then approximate (1) as

$$\begin{aligned} \frac{dx_i}{dt} &= -\alpha_i z_{i-1} x_i + \delta z_i + \gamma_i \frac{\alpha_i z_{i-1} x_i}{\gamma_i + \beta_i p} z_{i-1} x_i \\ &= -\tilde{\alpha}_i z_{i-1} x_i + \delta z_i, \\ \frac{dz_i}{dt} &= \tilde{\alpha}_i z_{i-1} x_i - \delta z_i, \end{aligned} \tag{2}$$

where $\tilde{\alpha}_i \equiv \alpha_i \beta_i p/(\gamma_i + \beta_i p)$. Simplified version (2) of our model corresponds to the linear signaling cascades model of Heinrich *et al.* (2002). In their model, the receptor activation is assumed to decay exponentially: $S(t) = \exp(-\lambda t)$, but in our model, we analyze the limiting case of slow signal decomposition by which the receptor is kept active, namely $\lambda = 0$.

As we have assumed that no protein is decomposed or newly synthesized in the early response of pathway to signal, and that the complex concentration is sufficiently small ($y_i \approx 0$), the total concentration of enzymes are preserved as $x_i + z_i = e_i$, where e_i is the initial concentration of free kinase E_i. Using this conservation relationship, the dynamics (2) is further simplified as

$$\frac{dz_i}{dt} = \tilde{\alpha}_i z_{i-1}(e_i - z_i) - \delta z_i, \tag{3}$$

$(i = 1, 2, \cdots, n)$ with $z_0 = S =$ constant. The system (3) gives the basis for the analysis below, but numerical integration of (1) is also fulfilled to examine the accuracy of the approximation.

Our interest is to know the extent to which signal is transduced in the cell after it is exposed to an environment change. We denote the concentration $z_i(t)$ of the phosphorylated kinase relative to that e_i of the initial free kinase by $\xi_i(t) = z_i(t)/e_i$, which we call the activity level of kinase i. The basic model (3) in terms of ξ_i is expressed as

$$\frac{d\xi_i}{dt} = \tilde{\alpha}_i e_{i-1} \xi_{i-1}(1 - \xi_i) - \delta \xi_i, \tag{4}$$

$(i = 1, 2, \cdots, n)$ with $\xi_0 = 1$. Shortly after the signal transduction is started, ξ_i is still sufficiently

smaller than one. We can then linearize (4) as

$$\frac{d\xi_i}{dt} = \tilde{\alpha}_i e_{i-1}\xi_{i-1} - \delta\xi_i. \tag{5}$$

By introducing $\zeta_i(t) = \xi_i(t)\exp(\delta t)$, (5) can be solved recursively from

$$\zeta_i = \tilde{\alpha}_i e_{i-1}\int_0^t \zeta_{i-1}(\tau)d\tau, \qquad (i = 1, 2, \cdots, n),$$

and $\zeta_0 = \exp(\delta t)$. We then have the time dependent solution of the linearized dynamics (5) for the activity level ξ_i of kinase i, and then we obtain the concentration $z_i(t) = e_i\xi_i(t)$ of the phosphorylated kinase i:

$$z_i(t) = \left(\prod_{j=1}^{i}\tilde{\alpha}_j\prod_{j=0}^{i}e_j\right)\frac{1}{\delta^i}\left[\exp(\delta t) - \sum_{j=0}^{i-1}\frac{(\delta t)^j}{j!}\right]\exp(-\delta t), \quad (1, 2, \cdots, n), \tag{6}$$

after the signal level is raised at $t = 0$:

$$z_0 = \begin{cases} 0, & (t < 0), \\ S, & (t \geq 0). \end{cases}$$

Much simpler expressions of z_i's are obtained if we further assume that the time t of our interest is sufficiently shorter than the relaxation time $1/\delta$ of the dephosphorylation ($t \ll 1/\delta$):

$$z_1(t) = \tilde{\alpha}_1 S e_1 t, \qquad z_2(t) = \frac{\tilde{\alpha}_1\tilde{\alpha}_2 S e_1 e_2}{2}t^2, \qquad \cdots,$$

and

$$z_n(t) = \left(\prod_{i=1}^{n}\tilde{\alpha}_i\prod_{i=0}^{n}e_i\right)\frac{t^n}{n!}. \tag{7}$$

Because our aim is to examine how fast the kinase cascade can respond to a signal, such an approximation can provide enough information on the determination of the optimal step number and the optimal kinase expression levels. Fig. 3 illustrates how accurate is the approximation (7) when compared with the exact solutions of the nonlinear dynamics (3) and of the full dynamics (1). We can see that the approximation (7) works quite well as long as z_n remains small. We thereafter examine the minimization of the time τ_n at which the kinase activity level z_n of the last kinase in the cascade reaches a certain (small) prefixed level C:

$$\tau_n = \left(\frac{n!C}{\tilde{\alpha}_1\tilde{\alpha}_2\cdots\tilde{\alpha}_n S e_1 e_2\cdots e_n}\right)^{1/n} \rightarrow \text{minimum}. \tag{8}$$

In Appendix, we describe how the parameter values we used in Fig. 3 are estimated from the data reported by Schoeberl *et al.* (2002), and discuss the parameter regions in which our approximation (7) gives accurate approximation to the full kinetic model (1) with nonlinearity and dephosphorylation.

3 Optimal design of kinase cascade

3.1 Optimal initial expression levels of kinases

We first examine the optimal initial expression levels $e_1, e_2, \cdots e_n$ of kinase $1, 2, \cdots, n$ that minimize the time τ_n at which the activity level of the last kinase reaches a certain prefixed value. We consider the minimization of τ_n under such a constraint that the total expression level of kinases is kept constant: $\sum_{i=1}^{n} e_i = E = \text{const}$. The response time τ_n is then minimized under the above constraint by expressing all kinases in equal concentration:

$$e_i^* = E/n, \qquad (i = 1, 2, \cdots, n). \tag{9}$$

It is interesting to note that the optimal initial expression levels of kinases are independent of their association constants (α_i's). In other words, the kinases in a cascade must be equally abundant, irrespective of their efficiencies in being phosphorylated, to maximize the speed of signal transduction. The data on the expression levels of MEK1 and ERK1 (Schoeberl *et al.*, 2002) seems to support this claim, as we discuss later.

3.2 Optimal number of steps that make up a kinase cascade

In this section, we consider the optimal number of phosphorylation steps that make up a signal transduction pathway. We have already shown that for a fixed number n of phosphorylation steps the speed $1/\tau_n$ of signal transduction is maximized under a constant total expression of kinases when each kinase has the same initial expression level: $e_i^* = E/n$, $(i = 1, 2, \cdots, n)$. If we substitute

this into the expression (8) of the response time τ_n, we have

$$\tau_n = \left(\frac{n!C}{\tilde{\alpha}_1 \tilde{\alpha}_2 \cdots \tilde{\alpha}_n S} \right)^{1/n} \frac{n}{E}. \tag{10}$$

Now we examine the optimal number n^* of phosphorylation steps that minimizes the time τ_n until the concentration of the final kinase reaches a certain level C (Fig. 4).

To this end we study how the ratio τ_n/τ_{n+1} depends on n. The optimal step number n^* is obtained when the ratio

$$\frac{\tau_n}{\tau_{n+1}} = \left(\frac{n!C}{S} \frac{(\tilde{\alpha}_{n+1})^n}{\prod_{i=1}^n \tilde{\alpha}_i} \right)^{1/n(n+1)} n(n+1)^{-(n+2)/(n+1)} \tag{11}$$

crosses unity, which illustrates that there is an intermediate optimal step number that minimizes τ_n when the amplification of final product concentration relative to the input signal, C/S, is sufficiently large. For simplicity we assume that the association constants are the same for all kinases: $\tilde{\alpha}_i = \tilde{\alpha}$. Applying the Stirling's approximation $n! \approx n^n e^{-n}$ in (10) for large n, we see that the optimal phosphorylation steps n^* approaches to

$$n^* \approx \frac{1}{2} \log \left(\frac{C}{S} \right), \tag{12}$$

when n^* is increased by increasing the amplification ratio C/S of final product to input signal. As consistent with this formula, the optimal step number n^* increases approximately linearly with the logarithm of the amplification ratio of the final product concentration C relative to the input signal S (Fig. 6).

By having greater number n of steps in the cascade, the greater is the nonlinearity in the initial response – the final product concentration initially increases with the n-th power of time as shown in (7). A greater nonlinearity (a larger number of steps in the cascade) implies a greater acceleration in the response, which merits speeding up the response when an initially small signal concentration must be amplified to a significant level in the final product (i.e., if C/S is sufficiently large). Conversely, if the amplification factor is small, on the other hand, the cascade with a smaller number of steps has a faster response.

Linear approximation accurately predicts the response time τ_n, as long as the target concentration C of final product is small. Our analysis then reveals that a multiple step cascade becomes the optimal if the concentration of input signal S is sufficiently small. However, linear analysis may fail if the target concentration C becomes large. In the next section we numerically examine the optimal step number of cascade using the full dynamics (1) with all nonlinear terms and dephosphorylation process intact.

4 Accuracy of linear approximation

We here confirm the accuracy of the approximation used in the previous sections. By numerically integrating the system (1) which retains all nonlinear terms and dephosphorylation process, the time τ_n until half of the kinases in the last step of the cascade are activated is plotted against the number of steps n of cascade (Fig. 5). We found that the optimal step numbers obtained in the previous sections (Fig. 4) slightly underestimate the exact optimums (Fig.5), but otherwise the approximation is quite efficient in predicting the dependency of τ_n on the step number n of cascade, and the dependency of the optimal step number on the amplification ratio C/S of final product to input signal.

An interesting property of the system (1) which is overlooked in the linear analysis is that the conversion efficiency in the signal transduction is improved by increasing the step number of cascade in the presence of dephosphorylation process (Fig 2). Indeed, we can see that the fraction of final step kinase which is to be eventually activated, $\hat{z}_n/e_n = z_n(\infty)/e_n$, monotonically increases with the step number n of the cascade, if association and dessociation rates remain the same in all steps. This is shown by setting the right hand sides of (1) to zero (i.e., $dx_i/dt = dy_i/dt = dz_i/dt = 0$ for $i = 1, 2, \cdots, n$) and noting $x_i + y_i + z_i = e_i$, which yields

$$\hat{z}_i = \frac{e_i \beta_i p \alpha_i \hat{z}_{i-1}}{\alpha_i \hat{z}_{i-1}(\beta_i p + \delta) + (\beta_i p + \gamma_i)\delta}.$$

As in the previous sections we assume for simplicity that the association and dissociation constants

(α_i's and β_is) and the expression levels of kinases (e_is) are the same for all steps: $\alpha_i = \alpha, \beta_i p = B, \gamma_i = \gamma$ for $i =, 2, \cdots n$. By defining

$$F(z) = \frac{eB\alpha z}{(B+\delta)\alpha z + (B+\gamma)\delta}$$

and noting that $\hat{z}_0 = S$, $\hat{z}_i$ is solved recursively as

$$\hat{z}_1 = F(S), \quad \hat{z}_2 = F(F(S)), \quad \cdots, \quad \hat{z}_n = F^n(S), \quad \cdots . \tag{13}$$

If we assume that the dephosphorylation rate δ is not too large, $\delta < eB\alpha/(B+\gamma)$, then $F(0) > 1$, and hence, starting from a sufficiently small signal concentration S, $\hat{z}_n$ monotonically increases towards $\hat{z}_\infty = [eB\alpha - (B+\gamma)\delta]/\alpha(B+\delta)$ as n increases. Thus we conclude that another advantage of having more steps in signal transduction pathway is to amplify the saturation level of the activated kinase.

5 Discussion

We here introduced a simple mathematical model for signal transduction pathway that can be made up of a cascade of phosphorylation reactions of kinases, and examined its optimum architecture in terms of the response time to signal. Our particular interest is under what conditions having multiple step cascades becomes advantageous over a single step transduction. The most important finding is that the multiple steps can accelerate the speed of response by amplifying the concentrations of kinases that intervene the signal and the final product. By having greater number n of steps in the cascade, the greater is the nonlinearity in the initial response – the final product concentration initially increases with the n-th power of time. The nonlinearity, not only guarantees switch-like response to a signal as noted previously, but also speeds up the response when the initial signal concentration is low relative to the target concentration of final product (i.e., if C/S is large). This yields an immediate prediction that a kinase cascade should have a greater number of steps if the signal concentration is low relative to the target concentration of the final product that is necessary for switching the response on. Though no experimental result is available yet, it is quite

interesting if the amplification ratio C/S of final product to input signal is greater in a multi-step phosphorylation (e.g. MAPK) pathway, than in a direct phosphorylation (e.g. SMAD) pathway.

The second conclusion of the model that, in a multiple step pathway, the optimum expression levels of each kinase must be the same irrespective of any difference in their association constants. The optimality here is the same as we discussed the optimal steps (but with an additional constraint for the cost of total kinase expression as discussed in the text) – maximizing the speed of response, or minimizing the time for the final product to reach a certain concentration in response to a given dose of signal. The data on the expression levels of MEK1 and ERK1 (Schoeberl *et al.*, 2002) seems to support this claim, though more studies are needed to test the prediction. Indeed, the expression levels of MEK1 and ERK1 (analogues of MAPKK and MAPK) are found to be almost the same in HeLa cell: 2.2×10^7 and 2.1×10^7 molecules per cell respectively (Schoeberl *et al.*, 2002), agreeing with out result. However, it was also shown that the concentrations of MAPKK and MAPK are much lower than that of MAPKKK (Huang & Ferrell, 1996). The latter fact might suggest that the expression levels of kinases may be determined not only by the initial speed but also by the amplitude and/or the duration of response, which we excluded from our objective function. In this report, we focus on the relationship between the number of steps in a kinase cascade and the speed at which the final kinase is activated after the cell is exposed to a signal.

Recently, the extent of phosphorylation becomes able to be detected by using fluorescence protein. Violin *et al.* reported that the phosphorylation of reporter protein (bound by the fluorescence proteins) causes the change in the fluorescence resonance energy transfer (FRET), allowing the real time imaging of PKC (protein kinase C) activation (Violin *et al.*, 2003). If we generate a specific reporter protein that cause the change in FRET by phosphorylation of each kinase in the MAPK cascade, we can measure the real time activity change of each step in the MAPK cascade. The reaction constants of phosphorylation, for example, may be estimated by the data obtained from such experiments. By changing the initial expression level of each kinase and/or the step number of cascade, we can calculate the speed of the signal transduction, and therefore be able to

test our hypotheses derived in this paper.

There are a number of potential limitations in our model. First, when we discuss the optimal architecture of kinase cascade, we focussed only on the speed of the response to a signal. Second, we made a number of simplifying assumptions on the kinetic reactions. The present model has the same degree of rigor as the other kinetic models of kinase phosphorylation (Heinrich *et al.*, 2002), though in some aspects our model takes a greater detail into account and in other aspects the reverse is true. For example, a simplified version of our model assuming quasi-steady state for the complex $E_{i-1}pE_i$, Eq. (2) in the text) corresponds to the Heinrich et al. (2002)'s linear model with slow signal decomposition (see the main text). We numerically confirmed how our simplifying assumptions (linearization of the process, quasi-steady state assumption based on the abundance of phosphates in the cell, and the neglect of dephosphorylation) affect the results, and found that none of them significantly changes the results in the time scale of our interest. Third, our model works, only when the parameter of dephosphorylation fall below a certain value (see appendix). When the concentration of phosphatase is estimated based on Schoeberl's data, the phosphorylation can not proceed. Because the reported concentration of the phosphatase varies depending on the cell types, our model may be applicable to the actual system, only when the concentration of phosphatase is small.

To conclude, we have revealed from a simple mathematical model for the signal transduction pathway that having multiple step kinase cascade makes the pathway to respond to signal faster than the single-step pathway, if the signal concentration S relative to the target concentration C of final product is low. The optimum number n^* of kinase cascade steps increases approximately linearly with the logarithm of the amplification of the final product concentration C relative to the input signal S .

Appendix: Parameter estimation

The parameters we used in this paper are taken from the data of the reaction between MEK1 (MAPKK) and ERK1 (MAPK) reported in Schoeberl *et al.* (2002). The association and dissociation constants of phosphorylated MEK1 and ERK1 in their paper correspond to α and γ respectively in our model, and are estimated as $\alpha = 0.11(\mu M^{-1}s^{-1})$ and $\gamma = 0.033(s^{-1})$. The reaction constant of phosphorylation in Schoeberl *et al.* (2002) corresponds to βp (β times the concentration of phosphate) in our paper, and is estimated as $\beta p = 16.0(s^{-1})$. Because we assume in this paper that all receptors are activated when the signal transduction starts, the expression level of EGF receptor in their paper corresponds to S $(= e_0)$ in our model. The expression levels of MEK1 and ERK1 correspond to the initial expression levels, e_1 and e_2, of the first and the second kinases in the cascade. They are then estimated as $S = 0.08(\mu M)$, $e_1 = e_2 = 1.0\,(\mu M)$. The dephosphorylation process is modeled in our paper as if it occurs spontaneously, but in reality it depends on phosphatase. However, if the complex of phosphorylated kinase and phosphatase dissociates sufficiently quickly, we can regard δ in our model as the product of dissociation constant d of this reaction and the concentration of phosphatase: $\delta = d \times$ [phosphatase]. Schoeberl *et al.* (2002)'s estimations are $d = 1.875\,(\mu M^{-1}s^{-1})$ and [phosphatase3] $= 0.5\,(\mu M)$ for the reaction between ERK1 and phosphatase3 in HeLa cell, which yields $\delta_{\mathrm{HeLa}} = 0.9375\,(s^{-1})$. The ratio of expression levels of the phosphatase to that of the kinase varies from 1 (Schoeberl *et al.*, 2002) to 1/500 (Hatakeyama & Konagaya, 2003), depending on the cell types, though estimates on α, γ, βp are relatively stable. For example, in CHO cell, [phosphatase]/[kinase] $= 1/500$. We numerically examined various [phosphatase]/[kinase] ratios and found that if this ratio is as large as 1/10, the dephosphorylation rate becomes too large to account for a high concentration of phosphorylated kinase (indeed, with the ratio 1/10, the concentrations of phosphorylated kinases are less than 1/1000 of initial concentrations of inactive kinases). We therefore restricted our analysis for the case of [phosphatase]/[kinase] $= 1/100$ or less, with which phosphorylated kinase concentration can become sufficiently large, and our approximation (7) agree quite well with full kinetic model (1). In

Fig 3, we adopted [phosphatase]/[kinase] = 1/100, which yields $\delta = 0.01875(s^{-1})$. The time scale of our interest, until the activated kinase concentration reaches a target level (say, 20% activation), is about 20 seconds, which is less than $1/\delta = 53$ seconds. If we assume [phosphatase]/[kinase] = 1/500, $1/\delta = 265$ seconds. The reaction time, about half to one minute, of the process using these estimated parameters might look too short. The reaction time is relatively insensitive to δ (which is the parameter showing the greatest variation over cell types), depending mainly on other parameters with relatively stable estimates. Note however that our model focuses on the very last stage, kinase cascade, of signal transduction, disregarding the time required for the receptors to be activated.

References

Attisano, L. & Wrana, J. L. (2002). Signal transducvtion by the TGF-β superfamily. *Science,* **296**, 1646–1648.

Castelli-Gair Hombria, J. & Brown, S. (2002). The fertile field of Drosophila JAK/STAT signaling. *Curr. Biol.* **12**, R569–R575.

Drogen, F., O'Rourke, S. M., Stucke, V. M., Jaquenoud, M., Neiman, A. M. & Peter, M. (2000). Phosphorylation of the MEKK Ste11p by the PAK-like kinase Ste20p is required for MAP kinase signaling *in vivo*. *Curr Biol.* **10**, 630–639.

Ferrell, Jr., J. E. & Machleder, E. M. (1998). The biochemical basis of an all-or-none cell fate switch in *Xenopus* oocytes. *Science,* **280**, 895–898.

Gustin, M. C., Albertyn, J., Alexander, M. & Davenport, K. (1998). MAP kinase pathways in the yeast *Saccharomyces cerevisiae*. *Microbiology and Molelular Biology Review,* **62**, 1264–1300.

Hatakeyama, M., Kimura, S. Naka. T. Kawasaki, T. Yumoto, N. Ichikawa, M. Kim, JH. Saito, K. Saeki, M. Shirouzu, M. Yokoyama, S. & Konagaya, A. (2003). A computational model on the modulation of mitogen-activated protein kinase (mapk) and akt pathways in heregulin-induced erbb signaling. *Biochem. J.* **373**, 451–463.

Heinrich, R., Neel, B. G. & Rapoport, T. (2002). Mathematical models of protein kinase signal transduction. *Mol. Cell.* **9**, 957–970.

Huang, C.-Y. F. & Ferrell, Jr., J. E. (1996). Ultrasensitivity in the mitogen-activated protein kinase-cascade. *Proc. Natl. Acad. Sci. USA,* **93**, 10078–10083.

Karin, M. (1998). Mitogen-activated protein kinase cascades as regulators of stress responses. *Ann. N. Y. Acad. Sci.* **851**, 139–146.

Klipp, E., Heinrich, R. & Holzhutter, H. (2002). Prediction of temporal gene expression. metabolic optimization by re-distribution of enzyme activities. *Eur. J. Biochem,* **269**, 5406–5413.

Leeuw, T., Wu, C., Schrag, J. D., Whiteway, M., Thomas, D. Y. & Leberer, E. (1998). Interaction of a G-protein β-subunit with a conserved sequence in Ste20/PAK family protein kinases. *Nature,* **391**, 191–195.

Madhani, H. D. & Fink, G. R. (1998). The riddle of MAP kinase signaling specificity. *Trends Genet.* **14**, 151–155.

Matsuyama, A., Yabana, N., Watanabe, Y. & Yamamoto, M. (2000). *Schizosaccharomyces pombe* Ste7p is required for both promotion and withholding of the entry to meiosis. *Genetics,* **155**, 539–549.

McKeithan, T. (1995). Kinetic proofreading in t-cell receptor signal transduction. *Proc. Natl. Acad. Sci. USA,* **92**, 5042–5046.

Schoeberl, B., Eichler-Jonsson, C., Gilles, E. D. & Muller, G. (2002). Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat. Biotechnol.* **20**, 370–375.

Theodosiou, A. & Ashworth, A. (2002). MAP kinase phosphatases. *Genome Biol.* **3**, REVIEWS3009.

Violin, J., Zhang, J., Tsien, R. & Newton, A. (2003). A genetically encoded fluorescent receptor reveals oscillatory phosphorylation by protein kinase c. *J. Cell Biology.* **161**, 899–909.

Wu, C., Whiteway, M., Thomas, D. Y. & Leberer, E. (1995). Molecular characterization of Ste20p, a potential mitogen-activated protein or extracellular signal-regulated kinase kinase (MEK) kinase kinase from *Saccharomyces cerevisiae. J. Biol. Chem.* **270**, 15984–15992.

Signal transducer	Ligand	# phosphorylation steps	Reference
MAPK	mating pheromone	4	Elion 2002
CAM-kinase	BDNF	3	Soderling 1999
SMAD	TGF-beta	1	Attisano & Wrana 2002
STAT	cg5988	1	Castelli-Gair Hombria & Brownt 2002

Table 1: Various signal transduction pathway and their phosphorylation step number.

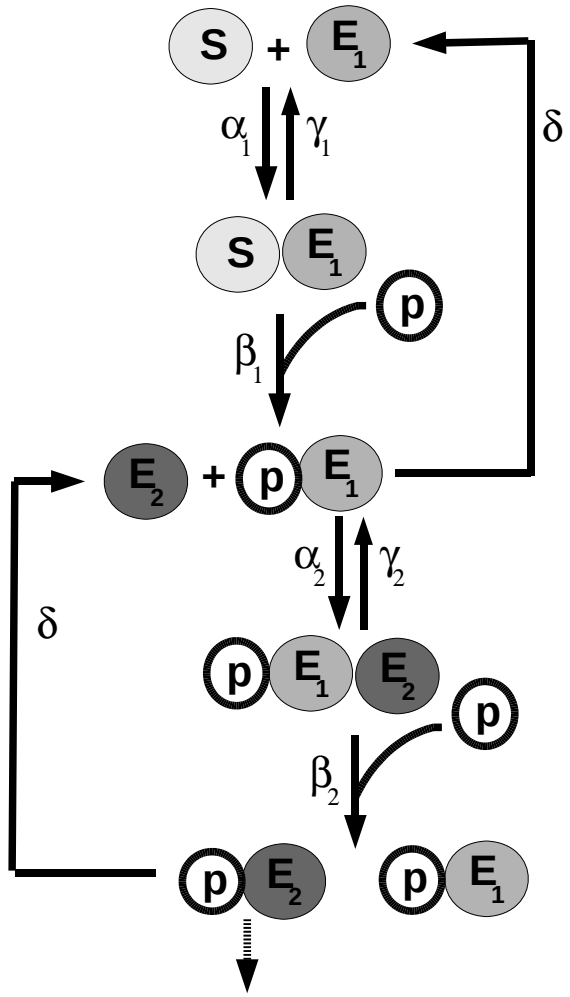


Figure 1: Schematic view of the signal transduction pathway. Signal molecule S interact with the kinase E_1 to make a complex SE_1 with association constant α_1. The complexSE_1 dissociate with dissociation constant γ_1. E_1 in the complex SE_1 is phosphorylated to become an active kinase E_1p. Active E_1p interact with another kinase E_2 in a downstream step of the cascade to form a complex E_1pE_2. E_2 is phosphorylated and activated, and activated kinase interact with still downstream kinase. Activated kinases are dephosphorylated with rate constant δ. The concentration of phosphate is sufficiently large compare with kinases.

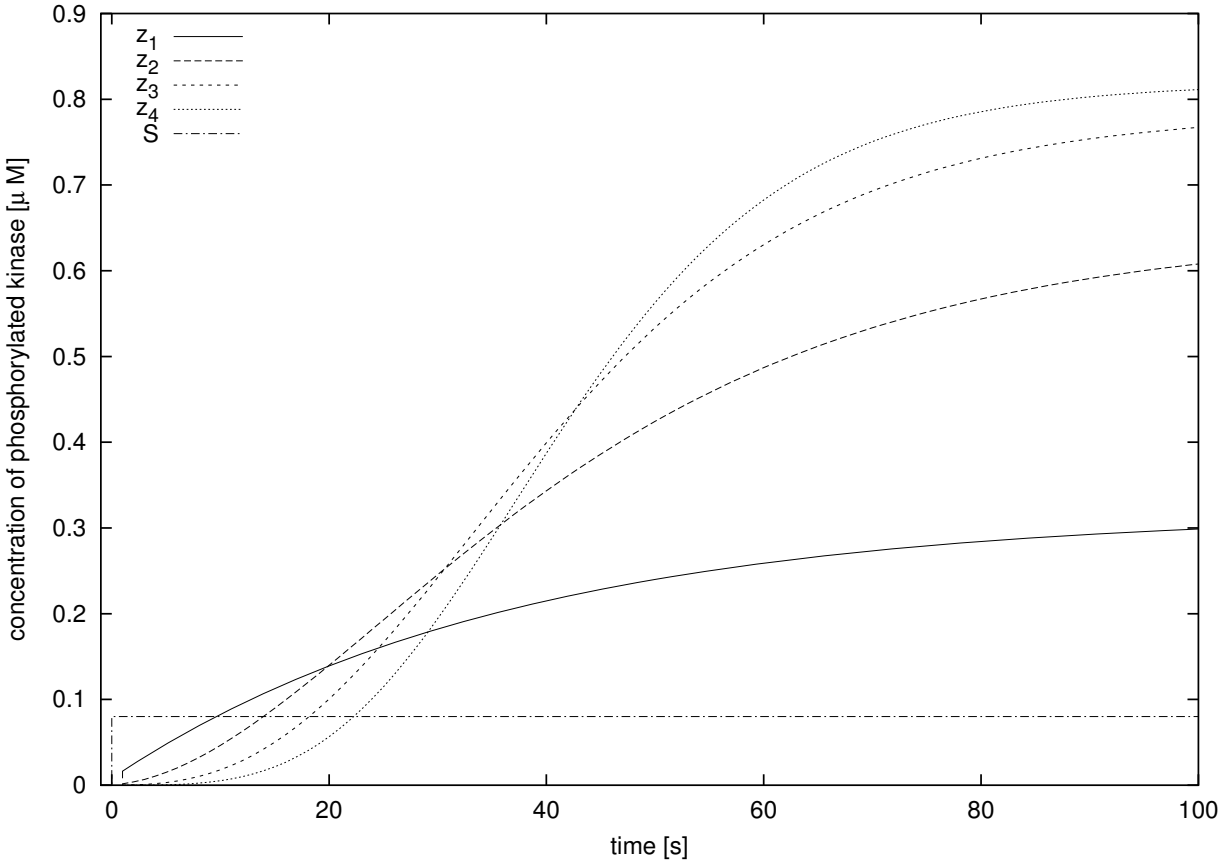
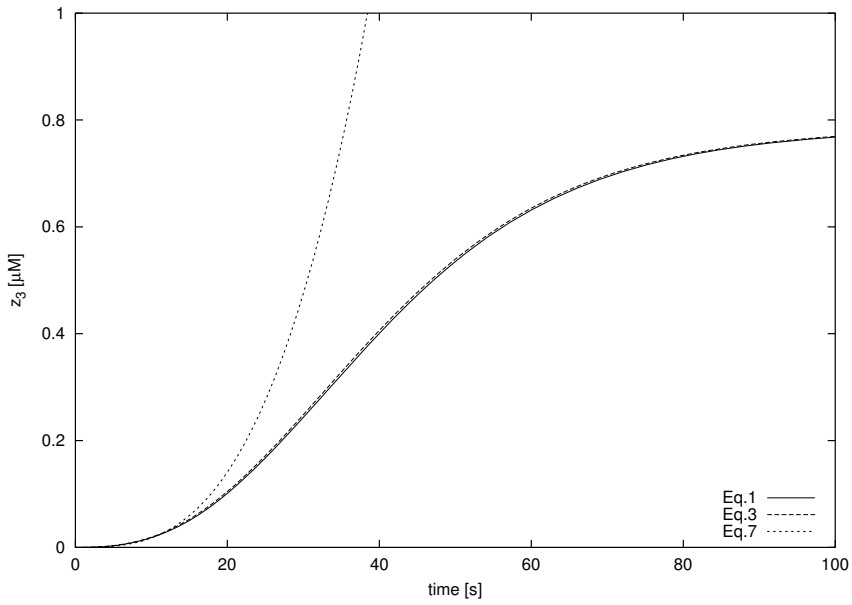


Figure 2: The time course of the concentration of the phosphorylated kinase. The signal transduction start at $t = 0$, when all receptors become active. Parameters: $\alpha_i = 0.11(\mu\mathrm{M}^{-1}\mathrm{s}^{-1})$, $\beta p = 16.0(\mathrm{s}^{-1})$, $\gamma = 0.033(\mathrm{s}^{-1})$, $e_i = 1.0(\mu\mathrm{M})$, $(i = 1, 2, \cdots, n)$, $S = e_0 = 0.08(\mu\mathrm{M})$, and $\delta = 0.01875(\mathrm{s}^{-1})$. The saturation level of kinase in each step is $z_1(\infty) = 0.31$, $z_2(\infty) = 0.65$, $z_3(\infty) = 0.79$, and $z_4(\infty) = 0.82$, respectively, which is calculated from Eq. (13) in the text.



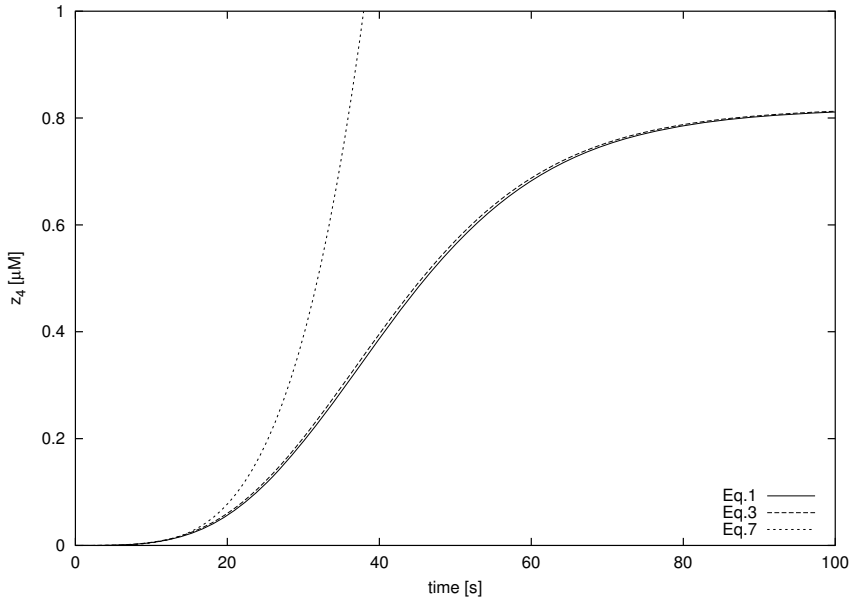


Figure 3: The concentration of the final product (the concentration of phosphorylated kinase at the last stage of cascade z_n) as a function of time. When z_n is sufficiently small, $z_n(t)$ (solid line) calculated from the non-linear system (1) is consistent with those (dashed lines) obtained from the two linearized systems, (3) with dephosphorylation term and (7) ignoring dephosphorylation term. (A) $n = 3$, (B) $n = 4$. The trajectories of the two linearized systems with and without dephosphorylation are too close with each other in this time scale to distinguish the difference by eyes. The same parameters used in Fig2 are used in this figure.

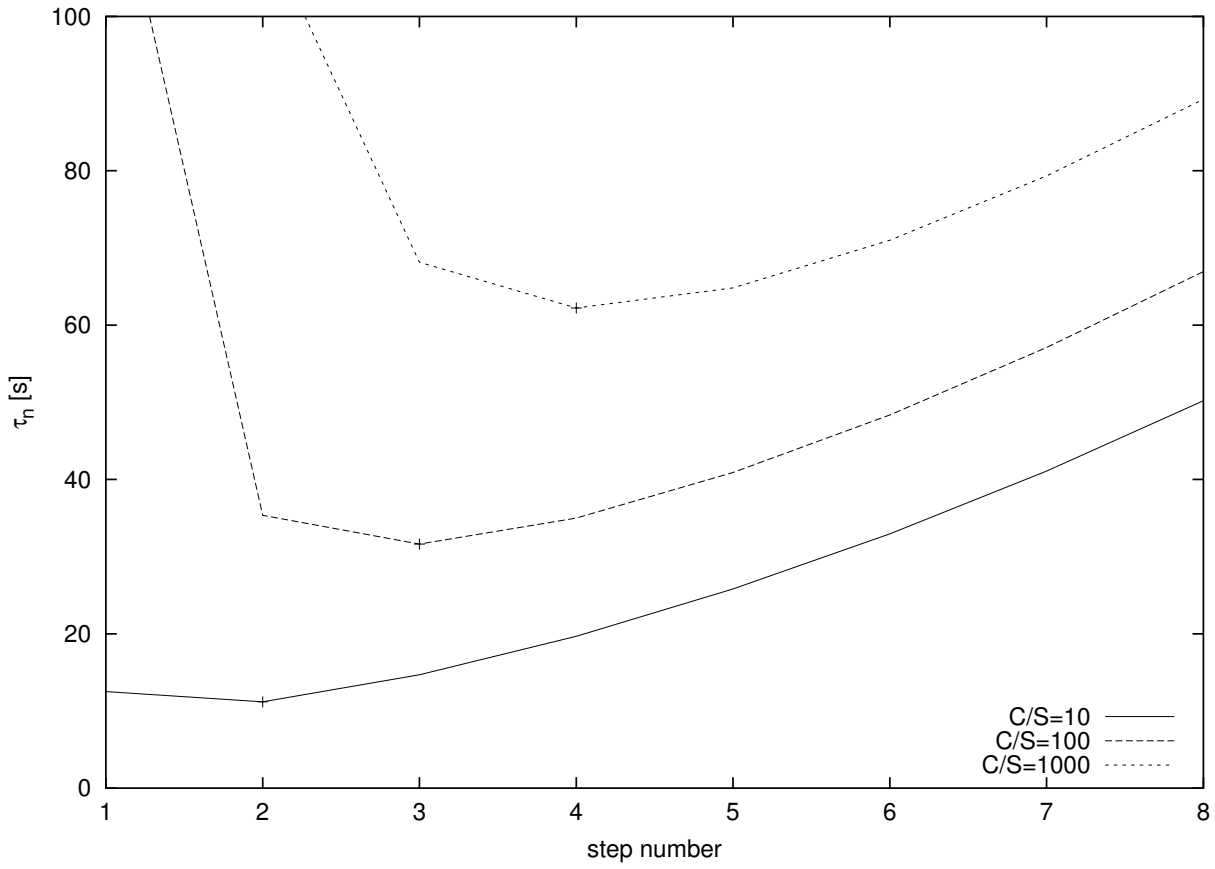


Figure 4: The time τ_n until the concentration of final product reaches at a prefixed value C as a function of the number n of steps that make up a cascade. τ_n is calculated from Eq. (10). The optimal phosphorylation step number is determined by minimizing τ_n. Parameters: $\tilde{\alpha}_i = \tilde{\alpha} = 0.1$ $i = 1, 2, \ldots, n$), $E = 8.0(\mu\text{M})$,and $C/S = 10, 100, 1000$. According to Fig 3, C must be on the order of 0.01μM and time $\leqslant$ sec to apply the linear approximation. This translates to $S = 0.001\mu$M when $C/S = 10$, which is about 315 molecules per cell (assume 5 micron radius).

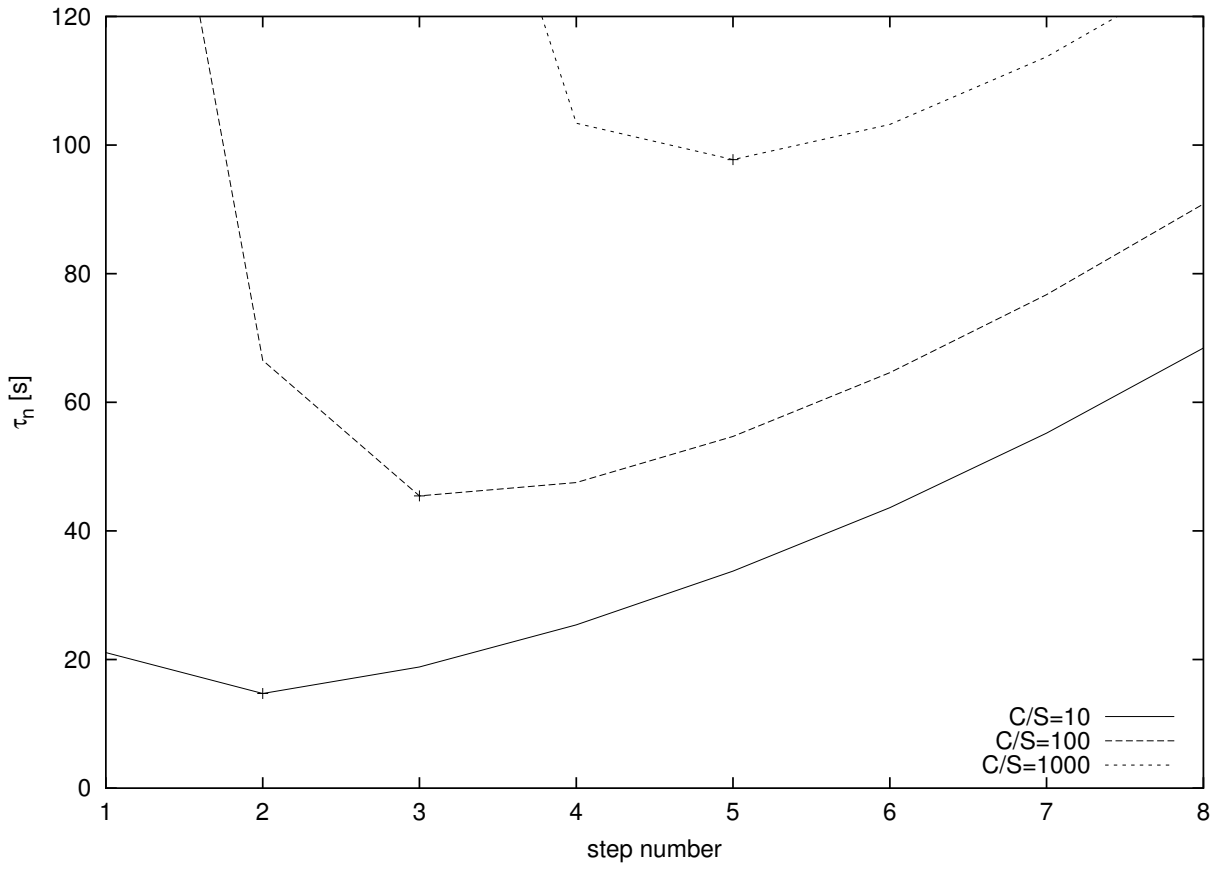


Figure 5: The time τ_n until the concentration of final product reaches a prefixed value C. τ_n is calculated from Eq. (1). The optimal phosphorylation step number is determined by minimizing τ_n. C is set to take a high value (50% of initial concentration of kinase E/n). The optimal step numbers for a given C/S is slightly larger than those predicted from linear approximation (see Fig. 4). Parameters: $\alpha_i = 0.1(\mu\mathrm{M}^{-1}\mathrm{s}^{-1})$, $\beta p = 16.0(\mathrm{s}^{-1})$, $\gamma = 0.033(\mathrm{s}^{-1})$, $\delta = 0.01875(\mathrm{s}^{-1}), E = 8.0(\mu\mathrm{M})$, and $C/S = 10, 100, 1000$.

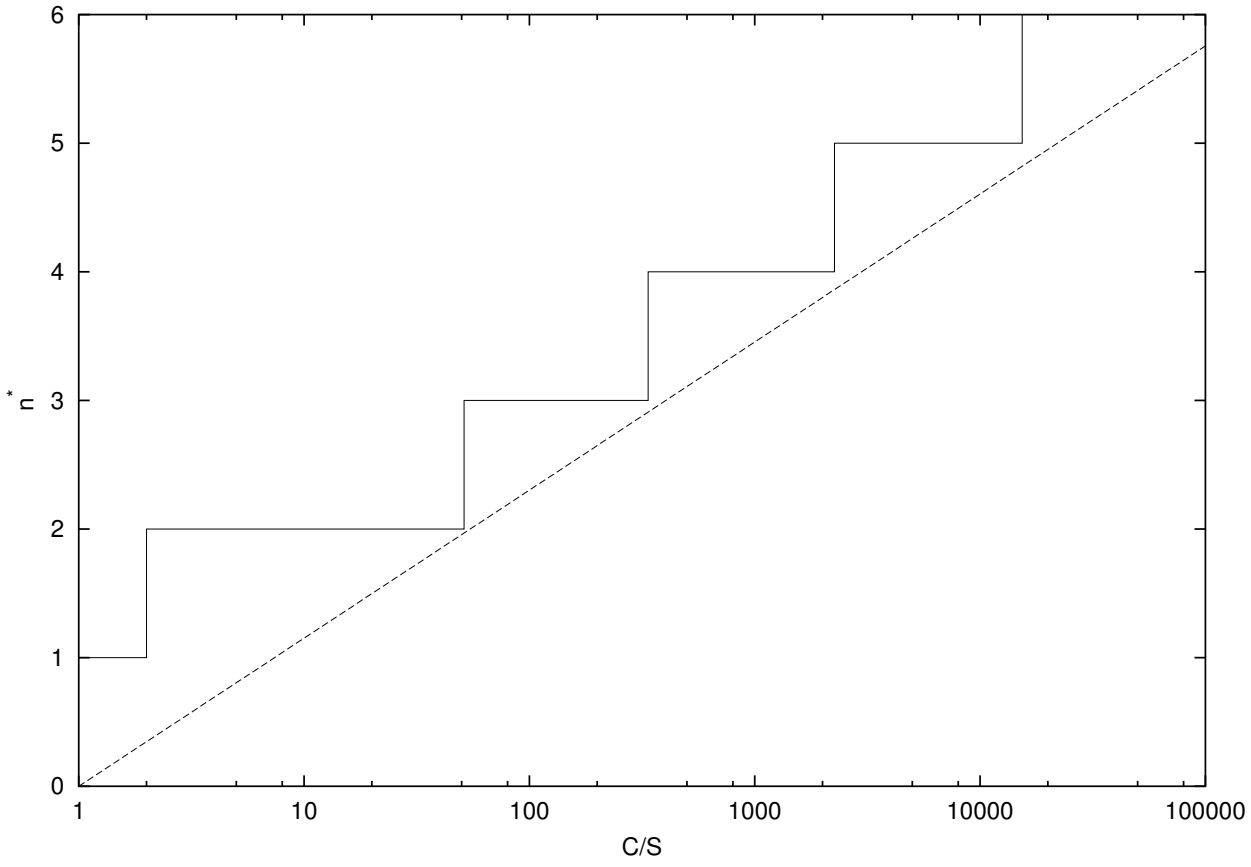


Figure 6: The optimal phosphorylation step number n^* obtained from (11) (solid lines) and its approximate formula (12) for large n (dashed line), plotted against C/S (scaled logarithmically). When the ratio C/S is sufficiently large, the multiple phosphorylation steps become optimal, and the optimal step number increases approximately linearly with $\log(C/S)$.