J. Biochem. **142,** 135–144 (2007)
doi:10.1093/jb/mvm151

An Analytical Rate Expression for the Kinetics of Gene Transcription Mediated by Dimeric Transcription Factors

Hsih-Te Yang[1,3], Chao-Ping Hsu[2] and Ming-Jing Hwang[3,*]

[1]Institute of Biochemistry and Molecular Biology, National Yang-Ming University; [2]Institute of Chemistry; and [3]Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan

Received April 17, 2007; accepted June 1, 2007; published online July 25, 2007

To model gene transcription kinetics, empirical fitting with the Hill function or S-system is often used. In this study, we derived an analytical expression for gene transcription rates in a manner similar to that developed for enzyme kinetics to describe the kinetics of gene transcription mediated by dimeric transcription factors (TFs) such as Gcn4p, a *Saccharomyces cerevisiae* master gene regulator. We showed that the analytical rate expression and its parameters estimated from several sets of experimental data could accurately reproduce the experimentally measured promoter-binding activity of Gcn4p. Furthermore, the analytical rate expression allowed us to derive analytically, rather than fit empirically, the parameters of the Hill function and S-system for use in modelling transcription kinetics. We found that a plot of gene transcription rate against Gcn4p concentration gave a sigmoidal dose-response curve with a positive co-operativity Hill coefficient (~1.25), in accordance with previous experimental findings on the promoter binding of dimeric TFs. The characteristics of the dose-response curve around the estimated cellular Gcn4p concentration suggest that transcription regulation is efficiently controlled under physiological conditions. This work is a useful initial step towards analytically modelling and simulating complicated gene transcription networks.

Key words: enzyme kinetics, gene transcription modelling, Hill function, S-system, transcription kinetics.

Abbreviations: MM, Michaelis–Menten; PIC, pre-initiation complex; TF, transcription factor.

As data for probing genetic regulatory networks is rapidly accumulating from genome-wide high-throughput experiments, mathematical modelling of gene transcription kinetics is assuming a prominent role. Hill's function (*1*, *2*) and the power-law-rate equations of S-system analysis (*3*), with parameters obtained from fitting experimentally observed dose-response data, are often employed for such studies. Hill function has a long history of been used to model the effect of co-operative binding in enzyme kinetics since the sigmoidal oxygen binding curve of haemoglobin was described by Archibald Hill in 1910 (*1*). In recent years, it has also been adopted to model the promoter binding of transcription activators or repressors during gene transcription processes (*2*). The Biochemical Systems Theory, also known as the S-system, is a mathematical modelling framework with which one can model the dynamics of a biochemical system by using power-law expansions to describe rates of biochemical processes (*3*).

An alternative treatment is the use of Michaelis–Menten (MM) approach (*4*), which involves solving a set of rate equations representing successive individual steps of the biochemical reaction of interest. MM-like expressions have been very successful in modelling enzyme kinetics, and, because an analogy can usually be drawn between enzyme and transcription reactions, MM-like models have also been adopted for modelling transcription kinetics. For example, the physical–chemical model for bacteriophage Lambda gene regulation developed by Shea and Ackers (*5*) describes, essentially, MM kinetics (*6*). More recently, Ronen *et al*. (*7*) showed that it is possible to use an MM-like model to accurately simulate the kinetics of the DNA damage response regulated by LexA in *Escherichia coli*. Furthermore, not only are MM-like models, which describe deterministic kinetics, useful to fit and simulate dynamical profiles of biochemical reactions, they can, under certain conditions, provide a framework to simplify and facilitate simulation of stochastic kinetics (*6*, *8*).

These modelling successes and the attractiveness of using the same framework for comparative studies between enzyme and transcription kinetics provide us with an impetus to further explore the use of MM-like expressions for modelling transcription reactions. In this contribution, we present an MM-like model for gene transcription mediated by dimeric TFs. Dimeric TFs are widely observed in both prokaryotes (*9*) and eukaryotes (*10*) and therefore represent an important class for modelling gene transcription. We showed that, using a set of fundamental rate expressions analogous to the MM rate expressions for enzyme kinetics, together with approximations made from considering mass balance, pre-equilibrium between the monomeric and dimeric forms of the TF, and a quasi-steady state, an analytical

*To whom correspondence should be addressed. Tel: +886-2-2789-9033, Fax: +886-2-2788-7641,
E-mail: mjhwang@ibms.sinica.edu.tw

solution to the rate expressions could be derived. Using the rate expression, we further showed that analytical expressions for the parameters of the Hill function and S-system could be derived in terms of fundamental rate constants. For evaluation, we applied these rate expressions to model the transcription mediated by Gcn4p, a master regulator of gene expression in yeast (*11*), taking advantage of the availability of the well-characterized kinetics parameters and data for its promoter binding pathways (*12*). We showed that the Gcn4p-mediated gene transcription rates computed using the analytical expressions were in good agreement both with experimentally observed data and with results from numerical simulations. These analytical expressions should be useful for modelling and simulating many complicated gene transcriptions and provide a means to estimate and interpret Hill and S-system parameters for modelling gene transcription kinetics.

MATERIALS AND METHODS

Model Settings—The initiation of transcription of a eukaryotic gene requires chromatin remodelling and the ordered assembly of the pre-initiation complex (PIC), a structure consisting of TFs, DNA template, and RNA polymerase (*13*, *14*). In this work, we modelled the process of TF binding and transcription initialization as shown in Fig. 1. In this process, chromatin remodelling, which involves a fast switching to an active state and the subsequent maintenance of this state (*15*), was not considered. In other words, we assumed that, during expression, the promoter sequence of a given gene is always exposed and available for binding by TFs and polymerases. In this binding, TFs may slide along DNA by one-dimensional diffusion or translocate/hop between DNA segments by three-dimensional diffusion before they find their target promoter (*16*). However, to simplify our task and also to accommodate the use of the kinetic parameters available for the Gcnp4 system studied (Table 1), which are estimated from *in vitro* experiments and do not include non-specific bindings (*12*), we furthermore ignore the intermediate steps of TFs searching for specific targets. Using these assumptions and simplifications, we modelled the kinetics of gene transcription on the binding of TFs to promoter, or DNA occupancy, which is a major controlling factor of gene transcription (*17*). Fig. 1 is essentially an extension of the TF monomer and dimer pathway model derived from experimental results of Cranz *et al*. (*12*), to which we added the step of PIC formation (DT_2P_O in Fig. 1). This last step has been modelled in a study in which it was split into several individual processes and modelled separately (*18*). As our aim was to derive an analytical rate expression for gene transcription, we necessarily simplified PIC assembly as a single step with a rate constant k_5, although, in principle, this step can be extended to deal with the individual processes modelled by Wang *et al*. (*18*). In addition to initiation, the process of transcribing a gene requires other reactions, such as elongation, pyrophosphorolysis, arrest, pause, editing, and termination (*13*). However, these are reactions for mRNA synthesis, which may be considered separately from transcription rate modelling (*19*, *20*). In other words, under the condition that mRNA synthesis is normally controlled, we can assume that a copy of mRNA is produced once a PIC is formed. Thus, in our model, the transcription rate is the PIC formation rate, which, in turn, is proportional to the concentration of the TF-promoter complex (DT_2) formed prior to polymerase loading.

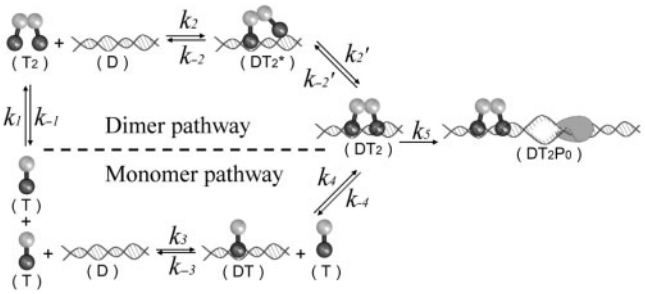

Fig. 1. **Reaction pathways of gene transcription controlled by a dimeric TF.** The TF in its monomeric form is depicted by a dumbbell-shaped symbol, the dark end representing its N-terminal DNA-binding domain and the bright end its C-terminal dimerization domain. The two precursors of the transcription process are promoter DNA (D) and monomeric TF (T); the latter forms a dimer (T_2) under the monomer-dimer pre-equilibrium assumption (see text). Three intermediates are shown: DT is the complex formed from binding of monomer (T) to promoter DNA (D) in the monomer pathway, DT_2^* is that formed by binding of dimer (T_2) to promoter (D) in the dimer pathway, and DT_2 is the mature complex. DT_2 recruits the RNA polymerase II complex, which unwinds the closed double-stranded DNA to form the ultimate transcription PIC (DT_2P_O).

Table 1. **Rate constants used in this work.**

Constant	Value	Units	Ref.
k_1[a]	$(1.6 \pm 0.5) \times 10^7$	$M^{-1}s^{-1}$	(*39*)
k_{-1}[a]	0.1 ± 0.031	S^{-1}	(*39*)
k_2	$(3.0 \pm 1.3) \times 10^8$	$M^{-1}s^{-1}$	(*12*)
k_{-2}	30 ± 12.5	S^{-1}	(*12*)
$k_{2'}$	10 ± 8	S^{-1}	(*12*)
$k_{-2'}$	0.1 ± 0.08	S^{-1}	(*12*)
k_3	$(5.0 \pm 1.3) \times 10^8$	$M^{-1}s^{-1}$	(*12*)
k_{-3}	50 ± 13.6	S^{-1}	(*12*)
k_4	$(5.0 \pm 1.3) \times 10^8$	$M^{-1}s^{-1}$	(*12*)
k_{-4}	0.03 ± 0.008	S^{-1}	(*12*)
k_5[b]	0.121 ± 0.000155	S^{-1}	

[a]The variance is estimated based on that of ΔC_p, the difference in heat capacity between the unfolded monomeric state and the dimeric state. The mean ΔC_p from experiments was $1.98 \pm 0.60\,kJ\,mol^{-1}K^{-1}$ (*39*). We used this ratio (3.3) of the mean and variance to estimate the variance for k_1 and k_{-1}.
[b]Estimated in this work (see Supplement).

According to the scheme depicted in Fig. 1, we can model the kinetics of gene transcription by the following steps of elementary chemical reactions:

$$\frac{d[T]}{dt} = 2k_{-1}[T_2] + k_{-3}[DT] + k_{-4}[DT_2] - 2k_1[T]^2 - k_3[T][D] - k_4[T][DT] \quad (1)$$

$$\frac{d[D]}{dt} = k_{-2}[DT_2^*] + k_{-3}[DT] - k_2[T_2][D] - k_3[T][D] \quad (2)$$

$$\frac{d[T_2]}{dt} = k_1[T]^2 + k_{-2}[DT_2^*] - k_{-1}[T_2] - k_2[D][T_2] \quad (3)$$

$$\frac{d[DT_2^*]}{dt} = k_2[T_2][D] + k_{-2'}[DT_2] - k_{-2}[DT_2^*] - k_{2'}[DT_2^*] \quad (4)$$

$$\frac{d[DT]}{dt} = k_3[D][T] + k_{-4}[DT_2] - k_{-3}[DT] - k_4[T][DT] \quad (5)$$

$$\frac{d[DT_2]}{dt} = k_{2'}[DT_2^*] + k_4[T][DT] - k_{-2'}[DT_2] - k_{-4}[DT_2] - k_5[DT_2] \quad (6)$$

$$\frac{d[DT_2Po]}{dt} = k_5[DT_2] \quad (7)$$

In these rate equations, D denotes promoter DNA, T monomeric TF, T_2 dimeric TF, and P_O polymerase in the open state. The above is a set of ordinary differential equations (ODEs) for time-dependent concentrations of the species involved in the transcription process depicted in Fig. 1. Most of the rate constants of these elementary reactions have been estimated based on experiments that measure the promoter binding of Gcn4p (*12*), but, to estimate the rate of RNA polymerase loading, which has not yet been determined in yeast, we adopted some assumptions and used data from *E. coli* experiments (see Supplement). The experimentally estimated kinetic constants used in this study are given in Table 1.

Numerical Simulation and Approximations at Quasi-Steady State—To characterize the proposed model, we performed a numerical simulation of Equations 1–7, using the estimated rate constants (Table 1) and the same initial concentrations of TF and promoter (see legend to Fig. 2) as used by Cranz *et al*. (*12*). The numerical simulation, solved by an ODE solver (see Supplement), yielded a time series of the concentrations and formation rates of the species involved. Based on the simulation results, shown in Fig. 2a and b, we were able to formulate several approximations and solve for an analytical solution. These approximations and assumptions are described below.

Mass balance—In a closed system, the total amounts of TF and promoter remain constant. Moreover, the promoter concentration is typically much lower than that of TF (since a copy of DNA can express many copies of proteins, i.e. $[D]_0 << [T]_0$; the subscript 0 denotes the initial condition at time 0). Consequently, for practical purposes, we can ignore the comparatively small amount of promoter-containing species that are associated with TF (i.e. $[DT]$, $[DT_2^*]$, $[DT_2]$, and $[DT_2Po]$); therefore,

$$[T] + 2[T_2] \cong [T]_0, \quad (8)$$

As shown in Fig. 2b, throughout the simulation, $[T] + 2[T_2]$ was fairly constant, although there was a small difference from the initial amount used (~90 *vs* 100 nM). This difference, as expected for a dimeric TF, is about twice that of the amount of DNA.

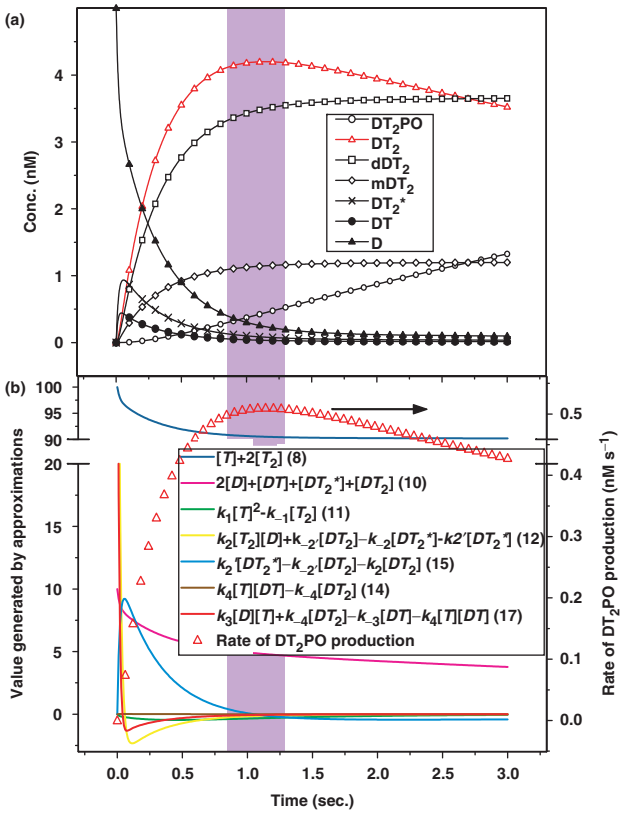

Fig. 2. **Kinetic profiles produced by numerical simulations.** (**a**) Time series of concentration data produced by numerically integrating the proposed set of rate expressions (Equations 1–7) using the kinetic constants listed in Table 1 and the initial concentrations: $[D]_0 = 5$ nM, $[T]_0 = 100$ nM, and zero for all other species (*12*). Corresponding species are: DT_2P_O—open circles, DT_2—red open triangles, dDT_2—squares, mDT_2—diamonds, DT_2^*—crosses, DT—close circles and D—closed triangles. The profiles for mDT_2 (contribution to DT_2 formation of the monomer pathway) and dDT_2 (contribution to DT_2 formation of the dimer pathway) were obtained by treating their respective pathway separately and ignoring the subsequent formation of PIC (DT_2P_O). The two profiles served to check consistency with the previous results (*12*). (**b**) Dynamic profiles produced to examine the validity of the assumptions incorporated for deriving the analytical rate expressions. The rate of DT_2P_O production is shown on the right ordinate. The shaded region near 1 s indicates the time interval when the system reaches a quasi-steady state. Shown are values of the expressions in the left hand side of the equations: dark blue for Equation 8 ($[T] + 2[T_2]$); pink for Equation 10 ($2[D] + [DT] + [DT_2^*] + [DT_2]$); green for Equation 11 ($k_1[T]^2 - k_{-1}[T_2]$); yellow for Equation 12 ($k_2[T_2][D] + k_{-2'}[DT_2] - k_{-2}[DT_2^*] - k_{2'}[DT_2^*]$); brown for Equation 14 ($k_4[DT][T] - k_{-4}[DT_2]$); blue for Equation 15 ($k_{2'}[DT_2^*] - k_{-2'}[DT_2] - k_5[DT_2]$); red for Equation 17 ($k_3[D][T] + k_{-4}[DT_2] - k_{-3}[DT] - k_4[T][DT]$).

Note that this difference will decrease and the approximation of Equation 8 become more acceptable as the $[T]_0/[D]_0$ ratio increases. The ratio of 20, used in the *in vitro* work of Cranz *et al*. (*12*) and followed here, is much smaller than the estimated *in vivo* ratio of ~198 for the Gal4p system (*21*). Note also that the condition of $[D]_0 << [T]_0$ is similar to that of $[E] << [S]$ (E: Enzyme,

S: Substrate) normally observed in enzyme kinetics. It should be noted, however, that this approximation may not be valid for all transcription processes, such as those involving a low-abundance transcription factor binding to a multitude of genes.

The sum of DNA and its containing species should also be equal to the initial amount of promoter DNA ($[D]_0$):

$$[D] + [DT] + [DT_2^*] + [DT_2] + [DT_2P_O] \cong [D]_0 \quad (9)$$

The results of the numerical simulation led us to assume that, near the steady state, $[DT_2P_O] \cong [D]$ (Fig. 2a, time near 1s). This assumption also makes it possible to reduce the number of variables for deriving an analytical solution. Note that this assumption breaks down at very-low and very-high TF concentrations, since, at a very low $[T]_0$, transcription is not activated, the free DNA concentration ($[D]$) will be high, and the active complex concentration $[DT_2P_O]$ will be low, whereas, at a very high $[T]_0$, most of the promoter will be bound and converted to PIC, giving rise to a low $[D]$ and high $[DT_2P_O]$. However, since we are interested in modelling the kinetic behaviour of transcription activation, an intermediate $[T]_0$ similar to that found in cells should be the most important consideration. With this assumption, Equation 9 becomes:

$$2[D] + [DT] + [DT_2^*] + [DT_2] \cong [D]_0 \quad (10)$$

As shown in Figure 2b; after a brief induction, the sum of the terms on the left side of this equation, though slowly decreasing, was maintained around $[D]_0$, the initial promoter concentration (5 nM) used in the simulation.

Pre-equilibrium between the monomeric and dimeric forms of transcription factors—Following the work of Cranz *et al*. (*12*), we assumed there exists a pre-equilibrium between the monomeric and dimeric forms of Gcn4p before its interaction with DNA. It follows that (see Fig. 1):

$$k_1[T]^2 - k_{-1}[T_2] \cong 0 \quad (11)$$

As shown in Fig. 2b, this approximation held well in the numerical simulation.

Quasi-steady state

At about 1 s, the amount of DT_2 reached its maximum (Fig. 2a), as did the rate of DT_2P_O formation (Fig. 2b). At the same time, the concentrations of the intermediates, DT and DT_2^*, approached a very-low constant value (Fig. 2a). These results are characteristics of a steady state, and are similar to the situation in enzyme kinetics, in which both the steady-state and equilibrium state of the enzyme-substrate complex provide important approximations in developing enzyme kinetics models (*22*).

At a quasi-steady state, the net flux of $[DT_2^*]$ and $[DT_2]$ should approach zero:

$$\frac{d[DT_2^*]}{dt} = k_2[T_2][D] + k_{-2'}[DT_2] - k_{-2}[DT_2^*] - k_{2'}[DT_2^*] \cong 0 \quad (12)$$

$$\frac{d[DT_2]}{dt} = k_{2'}[DT_2^*] + k_4[DT][T] - k_{-2'}[DT_2] - k_{-4}[DT_2] - k_5[DT_2] \cong 0 \quad (13)$$

Since the overall rate of the monomer pathway is limited by a flux generation step for the binding of monomeric TF to DNA (because $k_{-3}/k_3 >> k_{-4}/k_4$, Table 1), we could assume that the reactions between $[DT_2]$ and $[DT]$ (involving k_4 and k_{-4}) reach "rapid equilibrium" (*22*) and therefore Equation 13 could be replaced by Equations 14 and 15,

$$k_4[DT][T] - k_{-4}[DT_2] \cong 0 \quad (14)$$

$$k_{2'}[DT_2^*] - k_{-2'}[DT_2] - k_5[DT_2] \cong 0 \quad (15)$$

Figure 2b shows that Equations 12, 14 and 15 are reasonable approximations once the transcription process approaches the quasi-steady state.

An Analytical Rate Expression—The six concentration variables of ($[T]$, [D], $[T_2]$, $[DT_2^*]$, $[DT]$, and $[DT_2]$) were obtained by solving Equations 8, 10–12, 14 and 15. Using the 'Solve' routine of Mathematica 5.2 (Wolfram Research Inc., Champaign, IL, USA), which can handle equations involving symbolic functions, we obtained two solutions, but one was not physically feasible (non-zero transcription rate at zero $[T]_0$; data not shown). The feasible solution, i.e. the resulting analytical rate expression for gene transcription rate, was:

$$V_{[T]_0} = k_5[DT_2] = \frac{2k_1k_2k_5[D]_0[T]_0^2}{a[T]_0^2 + b[T]_0 + c} \quad (16)$$

where

$a = 2k_1k_2(K_D + 1)$

$b = 8k_1K_N + k_2K_4K_p$,

$c = 2K_NK_p$,

and $K_D = (k_5 + k_{-2'})/k_{2'}$, $K_N = k_{-2}K_D + k_5$, $K_4 = k_{-4}/k_4$. These are three constants for the dissociation of DT_2, with K_D and K_N being mainly associated with the dimer pathway and K_4 with the monomer pathway.

$$K_p = k_{-1} + \sqrt{k_{-1}^2 + 8k_1k_{-1}[T]_0}$$

is a term resulting from the pre-equilibration between $[T]$ and $[T_2]$ (see Supplement). We note that, if the steady-state condition for [DT],

$$\frac{d[DT]}{dt} = k_3[D][T] + k_{-4}[DT_2] - k_{-3}[DT] - k_4[T][DT] \cong 0 \quad (17)$$

and Equation 13 were used in place of Equations 14 and 15, one obtained a much more complex solution that did not allow further analytical manipulation. Moreover, the numerical result of the more complex solution, using the same sets of rate constants for Gcn4p, was essentially the same as that computed using Equation 16 (see Discussion section), further supporting that the assumption of 'rapid equilibrium' made regarding k_4 and k_{-4} in the monomer pathway was reasonable.

Analytical Expression for the Hill Function Parameters—With the analytical rate expression (Equation 16), an analytical expression for the Hill coefficient (H), which measures the degree of co-operativity, and the median effective concentration of regulator (κ), which produces half promoter occupancy, can be derived. The Hill coefficient, H, is defined as (*23*, *24*):

$$H = \frac{\mathrm{d}\ln[Z]}{\mathrm{d}\ln[T]_0} = \frac{\mathrm{d}\ln[Z]}{\mathrm{d}[T]_0} \times [T]_0, \tag{18}$$

where

$$Z = \frac{Y}{(1-Y)} \quad \text{and} \quad Y = \frac{V_{[T]_0}}{V_{[T]_0 \to \infty}}, \tag{19}$$

and Y, the fractional gene transcription rate, represents the fraction of promoter occupied by PIC, while the maximum transcription rate, according to Equation 16, is

$$V_{[T]_0 \to \infty} = \lim_{[T]_0 \to \infty} V_{[T]_0} = \frac{k_5 k_{2'}[D]_0}{k_5 + k_{-2'} + k_{2'}} = \frac{k_5[D]_0}{K_D + 1}, \tag{20}$$

Plugging Equations 20, 19 and 16 into Equation 18, we can obtain an analytical expression for the Hill coefficient as a function of $[T]_0$ and rate constants:

$$H = \frac{\begin{array}{c} k_{-1}(k_1 k_2 K_4 [T]_0 (4k_1[T]_0 + K_P) \\ +4K_N(2k_1 K_P[T]_0 + k_{-1}(4k_1[T]_0 + K_P))) \end{array}}{(K_P - k_{-1})(k_1 k_2 K_4 K_P + 2k_{-1}K_N(4k_1[T]_0 + K_P))} \tag{21}$$

The median effective concentration of regulator (κ) can also be obtained by solving Equation 19 for $[T]_0$ by setting $Z = 1$ (*23*):

$$\kappa = \frac{\begin{array}{c} k_{-1}(1+K_D)(\sqrt{k_1}(k_2 k_4 + 8K_N) + U) \\ +2(k_1^{3/2} k_2 K_4^2 + k_1 K_4 U) \end{array}}{2k_{-1}k_2\sqrt{k_1}(1+K_D)^2}, \tag{22}$$

$$\text{where } U = \sqrt{k_2(k_1 k_2 K_4^2 + 8k_{-1}K_N(1+K_D))}$$

Analytical Expression for the S-system Parameter—The power-low rate equations of the S-system (see Supplement) allow non-integer or even negative (for inhibition) kinetic orders to be used in the modelling of a biochemical process such as a metabolic reaction or gene transcription (*3*). The kinetic order (g) of the S-system is defined as (*25*):

$$g = \left.\frac{\mathrm{d}\ln V_{[T]_0}}{\mathrm{d}\ln[T]_0}\right|_C = \left.\frac{\mathrm{d}V_{[T]_0}}{\mathrm{d}[T]_0}\right|_C \times \frac{[T]_0}{V_{[T]_0}} \tag{23}$$

where C is the TF concentration at which the first-order approximation is made. With this definition and the derived transcription rate expression (Equation 16), the kinetic order of the S-system can also be expressed as a function of $[T]_0$ and rate constants:

$$g = \frac{\begin{array}{c} k_{-1}(k_1 k_2 K_4 [T]_0 (4k_1[T]_0 + K_P) \\ +4K_N(2k_1 K_P[T]_0 + k_{-1}(4k_1[T]_0 + K_P))) \end{array}}{\begin{array}{c} (K_P - k_{-1})(k_1 k_2 K_4 K_P [T]_0 \\ +2k_{-1}(k_1[T]_0(k_2[T]_0(1+K_D) + 4K_N) + K_N K_P)) \end{array}} \tag{24}$$

RESULTS

Validation Against Experimental Data—We tested the derived analytical expressions by comparing them against quantitative experimental data. A series of time-course data measuring the Gcn4p-mediated response in yeast, including the Gcn4p concentration obtained from western blotting and the corresponding *TRP4* promoter binding activity of Gcn4p determined by gel retardation assay, has been reported (*26*). When the reported arbitrary concentration units were multiplied by a factor of 3, we not only produced the best comparison with the measured transcriptional activities, but the Gcn4p concentrations obtained were consistent with the estimated range for the cellular concentration (see Supplement). For the Gcn4p-mediated response, the time course of the normalized amount of Gcn4p-bound *TRP4* promoter measured by Albrecht *et al*. (*26*) was considered as a normalized transcription rate according to our model (Equation 7), which could then be directly compared with Y, the fractional gene transcription rate, as defined by Equation 19, or

$$Y = \frac{V_{[T]_0}}{V_{[T]_0 \to \infty}} = \frac{2k_1 k_2 (K_D + 1)[T]_0^2}{a[T]_0^2 + b[T]_0 + c}, \tag{25}$$

where $[T]_0$ in this comparison is the Gcn4p concentration in arbitrary units multiplied by a factor of 3. The comparison is shown in Fig. 3.

Fitting the Hill function or S-system (Equations 28 and 29 in Supplement) to experimentally observed data for transcription activities is a common practice. Fig. 3 also presents such a fit for both the Hill function and S-system (see Supplement for fitting procedures). As shown in Fig. 3, using the kinetic constants estimated in Table 1 and the scaling factor of 3, Equation 25 reproduced the experimental data well, as did the direct fitting with the Hill function and S-system. Furthermore, as shown in Table 2, the values for the parameters of the Hill function and S-system computed from the analytical expressions (Equations 21, 22 and 24) using the kinetic constants of Table 1 and the estimated cellular concentration of Gcn4p agreed well with those determined by directly fitting the two empirical functions to experimental data. The derived value for the Hill coefficient (~1.25) suggested a positive co-operativity in Gcn4p binding, a result in accordance with previous experimental studies (*12*, *27*). These results indicated that our analytical expressions were capable of accurately characterizing the kinetics of Gcn4p binding to promoter in yeast.

Sensitivity of the Analytical Expressions to the Estimated Kinetic Constants—To gauge the sensitivity of the analytical expressions to the estimated kinetic constants (Table 1), we imposed a Gaussian distribution on each of the estimated rate constants according to their measurement uncertainties, and computed, from the analytical expressions, the parameters of the Hill function and S-system for 10 000 randomly selected sets of these kinetic constants (see Table 1 and legend to Fig. 4). As shown in Fig. 4, the statistical distribution

of these parameters (H, g and κ) encompassed, within a 95% confidence interval, the values that best fit the experimentally observed transcription activities (i.e. Fig. 3), showing that our model and the derived analytical expressions were fairly robust. The parameter κ varied in a much larger range, and this is because a rate constant, k_2, in the denominator of Equation 22 can come close to zero during the Gaussian random sampling, whereas the large variation of g reflects the fact that S-system has one less parameter than Hill function to describe the kinetics.

Dose-Response Curves—To further evaluate the analytical expressions, we repeated the numerical simulation of Fig. 2a at various initial TF concentrations. Recording the rate of DT_2P_O formation, or, equivalently, $k_5[DT_2]$ (equation (7)), at the quasi-steady state for each $[T]_0$ simulation run, a dose (TF concentration) – response (rate of transcription) curve was produced from these numerical simulations. Fig. 5 shows that, without fitting any parameters, the analytical rate expression (Equation 16) reproduced the numerical simulations well, with deviations only at very high TF concentrations. The curves produced by the analytical expressions for the Hill function (using Equations 21 and 22) and the S-system (using Equation 24) were also as good as those fitted directly against the numerical simulations (see Supplement for fitting procedures). Likewise, as summarized in Table 2, the values of the Hill function and S-system parameters computed from the analytical expressions agreed well with those obtained from fitting to experimental data (Fig. 3) or to numerical simulations (Fig. 5).

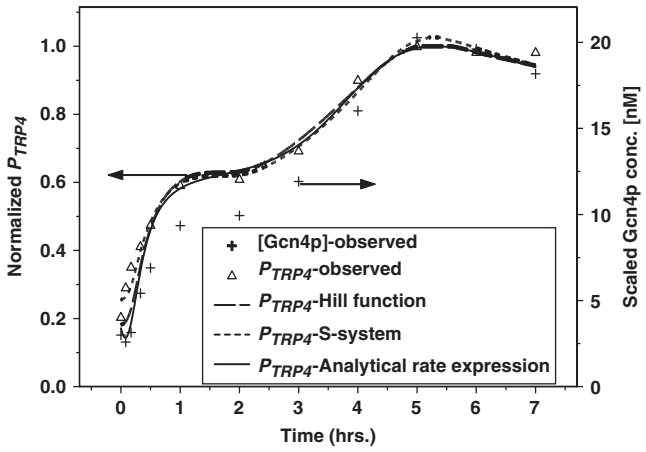

Fig. 3. **Gcn4p-mediated transcription activities: model *vs* experiments.** The observed Gcn4p concentrations (crosses; right ordinate) are those given in arbitrary units by Albrecht *et al*. (*26*) multiplied by a factor of three (see text and Supplement). The transcription activities are the normalized promoter (*TRP4*) binding activities of Gcn4p measured by Albrecht *et al*. (triangles). The transcription activity curves for the Hill function (dashed curve) and S-system (dotted curve) were derived by fitting to the experimental data, while that for the analytical rate expression (solid curve) was computed by Equation 25 using various observed Gcn4p concentrations as input.

It has been suggested from metabolic control analysis (MCA) (*28*) that the elasticity (ε), or kinetic order (g) (*25*), of an irreversible enzyme is related to the Hill coefficient (H) by the fractional saturation (Y) with substrate,

$$\varepsilon(\text{or } g) = (1 - Y) \times H \tag{26}$$

Analogously, the kinetic order of a gene transcription reaction is the elasticity of a gene promoter (enzyme) to a TF (substrate). As shown in Fig. 6, the analytical expressions derived for gene transcription kinetics closely followed this MCA relationship, indicating that, as in metabolic enzyme kinetics, g is similar to H at low transcription rates, but becomes progressively smaller as the transcription rate increases. However, unlike metabolic enzyme kinetics, our analytical expressions for gene transcription predicted high H and g values ($\sim$2.0 *vs* $\sim$1.0 for metabolic enzyme kinetics (*28*)) at very low TF concentrations (10^{-3}–10^{-2} nM). This may reflect a fundamental difference between the two systems: in gene transcription mediated by a dimeric TF, the production of mRNA requires the binding of two TF molecules, while in enzyme kinetics, at a low substrate concentration, the binding with one substrate can still lead to a product molecule.

DISCUSSION

Above, we showed that a simple analytical rate expression can be derived to accurately model the dynamics of gene transcription mediated by a dimeric TF, such as Gcn4p, a master gene regulator of yeast. Although in the present work we have focused only on the promoter binding of Gcn4p, a homodimer TF, it should not be difficult to extend our work to model a heterodimeric gene transcription system such as Fos/Jun (*10*) and its more complex combinatorial binding (*29*). In this extension, the rate expression would involve the concentration of two different TFs, [Fos] $\times$ [Jun], and at least two,

Table 2. **Hill and S-system parameters derived by fitting to experimentally observed data or to numerical simulations, or from computation with the analytical expressions.**

Model	Parameters	Fit to experiments (Fig. 3)	Fit to simulations (Fig. 5)	Analytical expressions	
				$[T]_0$ at 9.5 nM[a]	Distribution of parameters (Fig. 4) Median (range of the middle 95%)
S-system	Kinetic order (g)	0.69	0.87	0.84	0.85 (0.41–1.27)
Hill	Hill coefficient (H)	1.24	1.26	1.27	1.27 (1.16–1.43)
	Kappa (κ), nM	15.17	16.78	16.05[b]	16.85 (5.46–165.05)

[a]Estimated cellular concentration of Gcn4p (See Supplement).
[b]Calculated from Equation 22, which is independent of $[T]_0$.

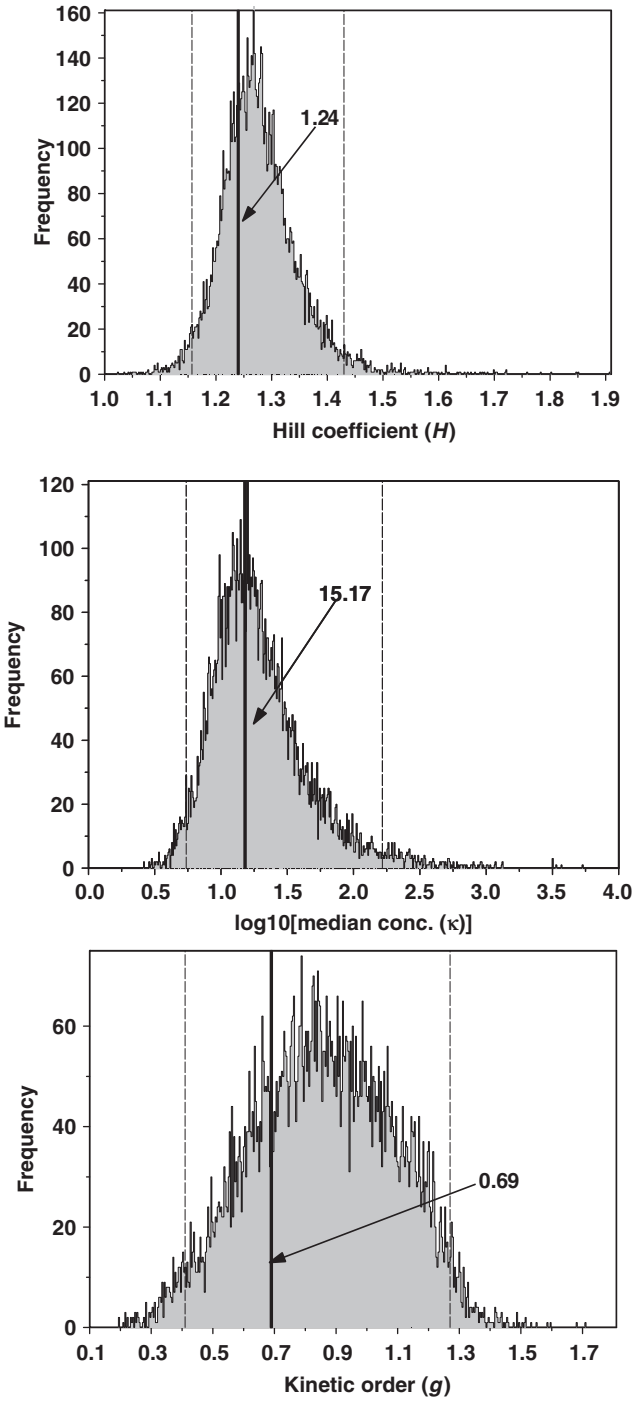

Fig. 4. **Distribution of the Hill function and S-system parameters.** These statistical distributions were obtained by computing the analytical expressions (Equations 21, 22 and 24) at $[T]_0 = 9.5$ nM and using each of 10 000 randomly generated sets of rate constants produced by imposing a Gaussian distribution on the mean and one standard deviation of each rate constant listed in Table 1. The regions bracketed by the dashed lines represent the 95% confidence interval of the distribution. The vertical lines are the values obtained by direct fitting to the experimental data as described in Fig. 3.

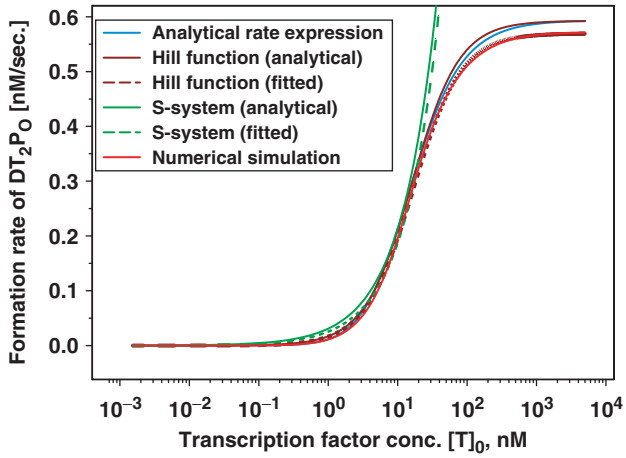

Fig. 5. **Dose-response curves: analytical expressions *vs* numerical simulations.** The rate of PIC formation was plotted as a function of TF concentration for numerical simulations (red solid line, see text), the analytical rate expression (Equation 16, blue solid line), and both direct fitting and analytical computations of the Hill function (both brown; fitted results in a dashed line and analytical results in a solid line) and of S-system (green curves). For the direct fitting of S-system, only the range of experimentally measured Gcn4p concentrations, 2.73–20.50 nM (see Fig. 3), was fitted, since S-system, with just two parameters, operates as a local approximation [Equation 29 in Supplement and ref. (*25*)].

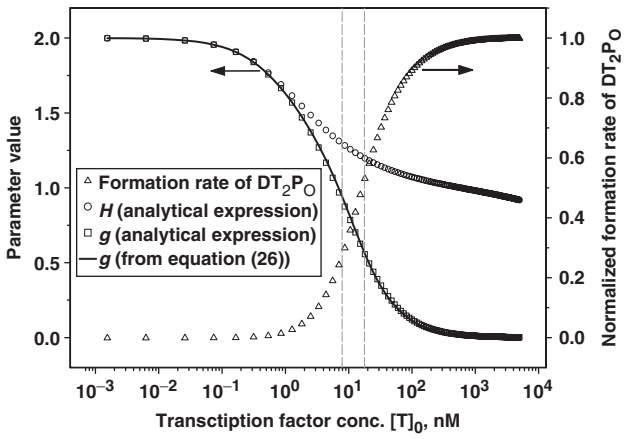

Fig. 6. **Relationship between the Hill coefficient and the kinetic order of the S-system.** The analytical expressions, Equations 21 and 24, were used to compute H (circles) and g (squares), respectively, using the rate constants in Table 1 and a TF concentration ranging from 0.001 to 5000 nM. The solid line is the relationship derived from a MCA($g = (1 - Y) \times H$) (*28*). Between the two dashed vertical lines is a region for the estimated range of cellular Gcn4p concentrations of 7.87–17.73 nM (see Supplement). The numerically simulated dose-response curve from Fig. 5 (formation rate on right ordinate; triangles) is shown for reference.

not one, monomer pathways (Jun binding first and Fos binding first, respectively).

Several features of the derived analytical rate expression (Equation 16) are worthy of comment. Firstly, Equation 16 is similar to the rate laws derived for enzyme reactions, which are often expressed as a ratio of polynomials of substrate concentration (*30*). In our equation, both the numerator and denominator are second order polynomials of TF concentration, reflecting a complex kinetics probably arising from the nature of a dimeric TF. This is analogous to enzyme reactions, in which the power of the polynomials equals the number of interacting substrate molecules (*22*). Secondly, Equation 16 satisfies the expected lower limit, i.e. $\lim_{[T]_0 \to 0} V_{[T]_0} = 0$, since the rate of transcription should be zero when TF concentration is zero. Thirdly, at an extremely high TF concentration, in an ideal situation, all promoters are bound prior to the formation of PIC, giving rise to $[DT_2] \cong [D]_0$ and a maximal transcription rate $V_{\max} \cong k_5[D]_0$ (from Equation 7). Given that $k_{2'} >> k_5, k_{-2'}$ (Table 1), hence $k_D << 1$, Equation 20 is in accordance with this upper limit condition. Note that, in this hypothetical situation, the concentration of DT_2P_O is assumed to be zero; however, in real situations, its concentration is finite and this will result in a smaller maximal transcription rate than that predicted by the analytical expression. In other words, at high TF concentrations, which may render a low $[D]$ and non-negligible $[DT_2P_O]$ and consequently invalidate the assumption of $[DT_2P_O] \cong [D]$ (see Materials and Methods section), our analytical rate expression will overpredict the transcription rate, as observed in Fig. 5.

A key foundation of our model is the employment of a quasi-steady-state condition, as opposed to the equilibrium state used previously (*31*). Both equilibrium state and steady state are important approximations of enzyme kinetics. As Segel (*22*) pointed out, an equilibrium model is actually a special case of the steady-state treatment for conditions which in our case would be $k_5 << k_{-2'}$. Our values for k_5 and $k_{-2'}$ were quite similar (0.121 *vs* 0.1; Table 1), suggesting that the equilibrium condition may not be generally employable for modelling transcription kinetics. In enzyme kinetics, equilibrium is usually established rapidly (*28*) and therefore the equilibrium model can easily be employed. In gene transcription, however, equilibrium is typically reached over a long time scale [several minutes in the work of Cranz *et al*. (*12*)], whereas our simulation showed that a quasi-steady state could occur in only a few seconds (Fig. 2). Thus, in working with typical experimental results that dissect transcription dynamics on a timescale of hours, equilibrium and steady-state approaches are both acceptable. Indeed, an equilibrium model that ignores the influence of k_5 yielded a dose-response curve overall similar to that of the quasi-steady-state model (see Fig. 7). However, this is not a general result, since the equilibrium model ignores an essential part of the reaction flux (k_5), which did elicit noticeable difference in the details of the resulting dose-response curves, particularly in the region of physiologically relevant concentrations

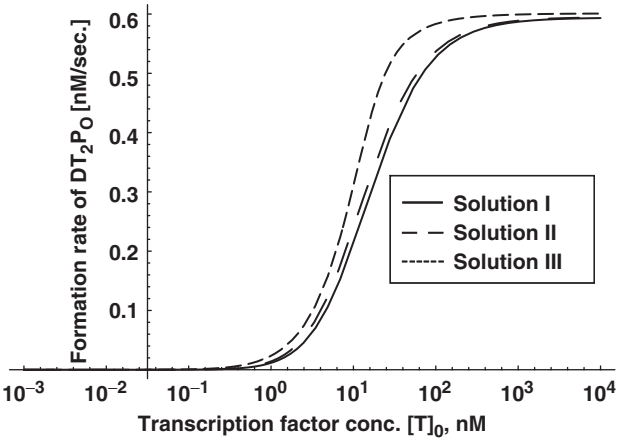

Fig. 7. **Dose-response curves using different approximations**. Dose-response curves from three different solutions to equations 1–7 are plotted. The solution derived in the main text (Equation 16) is solution I. In solution II, Equations 13 and 17 were used in place of Equations 14 and 15 to solve for the quasi-steady-state approximation. This resulted in a much more complicated formula (not shown) than Equation 16. Solution III is an equilibrium model where the reaction flux described in Equation 7 was ignored, and all the other reactions were at their equilibrium.

(Fig. 7). It follows that if the experimental measurements were made in a very short timescale for systems such as the gene activity regulated by Gcn4p studied here, the equilibrium approach may not be as applicable as the quasi-steady-state approximation.

The Gcn4p concentration was estimated to be ~9.5 nM for yeast after 2 h of amino acid starvation (see Supplement). This estimated value is lower than, but reasonably close to, the median effective concentration ($\kappa = 15.17$ nM) derived from directly fitting the Hill function to experimentally observed Gcn4p-mediated transcription activities (Fig. 3). This is similar to the observed substrate concentrations in enzyme kinetics, where *in vivo* levels of substrate are usually similar to, or lower than, the half-effective concentration (*32*, *33*). These *in vivo* levels of the TF concentration may underlie a physiological mechanism for transcriptional regulation, since, at such concentrations, a small change in TF levels could lead to a large change in transcription rate (Fig. 6). Thus, transcription appears to be efficiently controlled physiologically, and TF concentration probably plays a key role in this control.

Obviously, our model gives a highly simplified description of a transcription process that often involves co-operative binding of multiple transcription activators and repressors. Moreover, individual transcriptions could sum up to generate an all-or-none transcription switch in a population of cells (*2*). Stochastic simulations have illustrated that fluctuations in TF binding could play a role in such a binary response in the expression of an inducible gene (*34*), and pulses in mRNA production contribute to the noise of gene expression (*35–37*). Nevertheless, although our deterministic model simply describes the averaged behaviour of gene transcription in a cell, models of random perturbation can be introduced

into the deterministic analytical expression to allow a simplified approach for simulating stochastic chemical kinetics (*6*). The analytical expressions could also provide a rate law-based foundation for employing the Hill function and S-system to model gene transcription kinetics and allow estimation of their parameters in the absence of experimental dose-response data. It should also be possible to incorporate other modelling works (*e.g. 16*, *18*, *38*) to derive a more elaborate expression that can be applied to describe *in vivo* transcription processes and their kinetics better and in more details.

Supplementary data are available at JB Online.

We thank Dr. Feng-Sheng Wang, Department of Chemical Engineering, National Chung-Cheng University, for providing the HDE program. We are grateful to Dr. Feng-Sheng Wang, Dr. Chiou-Hwa Yu of the National Health Research Institute, and members of the Hwang lab for many stimulating and helpful discussions. This work was supported by financial aid from the Academia Sinica and NSC (to CPH and MJH).

REFERENCES

1. Hill, A.V. (1910) The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J. Physiol.* **40**, 4–7
2. Rossi, F.M.V., Kringstein, A.M., Spicher, A., Guicherit, O.M., and Blau, H.M. (2000) Transcriptional control: rheostat converted to on/off switch. *Mol. Cell* **6**, 723–728
3. Savageau, M.A. (1976) *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology* Addison-Wesley Publishing Company, Reading, MA
4. Michaelis, L. and Menten, M.L. (1913) Kinetics of invertase action. *Biochem. Z* **49**, 369
5. Shea, M.A. and Ackers, G.K. (1985) The OR control system of bacteriophage lambda: A physical-chemical model for gene regulation. *J. Mol. Biol.* **181**, 211–230
6. Rao, C.V. and Arkin, A.P. (2003) Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. *J. Chem. Phys.* **118**, 4999–5010
7. Ronen, M., Rosenberg, R., Shraiman, B.I., and Alon, U. (2002) Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. *Proc. Natl Acad. Sci. USA* **99**, 10555–10560
8. Goutsias, J. (2005) Quasi-equilibrium approximation of fast reaction kinetics in stochastic biochemical systems. *J. Chem. Phys.* **122**, 184102
9. Wagner, R. (2000) *Transcription Regulation in Prokaryotes.* Oxford Univerity Press, Oxford
10. Kohler, J.J. and Schepartz, A. (2001) Kinetic studies of Fos.Jun.DNA complex formation: DNA binding prior to dimerization. *Biochemistry* **40**, 130–142
11. Hinnebusch, A.G. and Natarajan, K. (2002) Gcn4p, a master regulator of gene expression, is controlled at multiple levels by diverse signals of starvation and stress. *Eukaryotes Cell* **1**, 32
12. Cranz, S., Berger, C., Baici, A., Jelesarov, I., and Bosshard, H.R. (2004) Monomeric and dimeric bZIP transcription factor GCN4 bind at the same rate to their target DNA site. *Biochemistry* **43**, 718–727
13. Greive, S.J. and von Hippel, P.H. (2005) Thinking quantitatively about transcriptional regulation. *Nat. Rev. Mol. Cell. Biol.* **6**, 221–232
14. Shanblatt, S.H. and Revzin, A. (1984) Kinetics of RNA polymerase-promoter complex formation: effects of nonspecific DNA-protein interactions. *Nucleic Acids Res.* **12**, 5287–5306
15. Kingston, R.E. and Narlikar, G.J. (1999) ATP-dependent remodeling and acetylation as regulators of chromatin fluidity. *Genes Dev.* **13**, 2339–2352
16. Berg, O.G., Winter, R.B., and von Hippel, P.H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* **20**, 6929–6948
17. Xu, H.E., Kodadek, T., and Johnston, S.A. (1995) A single GAL4 dimer can maximally activate transcription under physiological conditions. *Proc. Natl Acad. Sci. USA* **92**, 7677–7680
18. Wang, J., Ellwood, K., Lehman, A., Carey, M.F., and She, Z.S. (1999) A mathematical model for synergistic eukaryotic gene activation. *J. Mol. Biol.* **286**, 315–325
19. Monk, N.A. (2003) Oscillatory expression of Hes1, p53, and NF-κB driven by transcriptional time delays. *Curr. Biol.* **13**, 1409–1413
20. Arnold, S., Siemann, M., Scharnweber, K., Werner, M., Baumann, S., and Reuss, M. (2001) Kinetic modeling and simulation of in vitro transcription by phage T7 RNA polymerase. *Biotechnol. Bioeng.* **72**, 548–561
21. Verma, M., Bhat, P.J., and Venkatesh, K.V. (2004) Expression of GAL genes in a mutant strain of Saccharomyces cerevisiae lacking GAL80: quantitative model and experimental verification. *Biotechnol. Appl. Biochem.* **39**, 89–97
22. Segel, I.H. (1993) *Enzyme Kinetics: Behavior and analysis of rapid equilibrium and steady state enzyme systems.* Wiley Classics Library Edition Published, New York
23. Wyman, J. (1963) Allosteric effects in hemoglobin. *Cold Spring Harbor Symp. Quant. Biol.* **28**, 483–489
24. Goldbeter, A. (1976) Kinetic cooperativity in the concerted model for allosteric enzymes. *Biophys. Chem.* **4**, 159–169
25. Voit, E.O. (2000) *Computational Analysis of Biochemical Systems* Cambridge University Press, Cambridge, UK
26. Albrecht, G., Mosch, H.U., Hoffmann, B., Reusser, U., and Braus, G.H. (1998) Monitoring the Gcn4 protein-mediated response in the yeast Saccharomyces cerevisiae. *J. Biol. Chem.* **273**, 12696–12702
27. Kim, B. and Little, J.W. (1992) Dimerization of a specific DNA-binding protein on the DNA. *Science* **255**, 203–206
28. Fell, D. (1997) *Understanding the Control of Metabolism.* Portland Press, London and Miami
29. Remenyi, A., Scholer, H.R., and Wilmanns, M. (2004) . Combinatorial control of gene expression. *Nat. Struct. Mol. Biol.* **11**, 812–815
30. Savageau, M.A. (1969) Biochemical systems analysis. I. Some mathematical properties of the rate law for the component enzymatic reactions. *J. Theor. Biol.* **25**, 365–369
31. Bolouri, H. and Davidson, E.H. (2003) Transcriptional regulatory cascades in development: initial rates, not steady state, determine network kinetics. *Proc. Natl Acad. Sci. USA* **100**, 9371–9376
32. Hochachka, P.W. and Somero, G.N. (2002) *Biochemical Adaptation: Mechanism and Process in Physiological Evolution.* Oxford University Press, New York
33. Cornish-Bowden, A. (1976) The effect of natural selection on enzymic catalysis. *J. Mol. Biol.* **101**, 1–9
34. Pirone, J.R. and Elston, T.C. (2004) Fluctuations in transcription factor binding can explain the graded and binary responses observed in inducible gene expression. *J. Theor. Biol.* **226**, 111–121

35. Golding, I. and Cox, E.C. (2004) RNA dynamics in live Escherichia coli cells. *Proc. Natl Acad. Sci. USA* **101**, 11310–11315
36. Golding, I., Paulsson, J., Zawilski, S.M., and Cox, E.C. (2005) Real-time kinetics of gene activity in individual bacteria. *Cell* **23**, 1025–1036
37. Chubb, J.R., Trcek, T., Shenoy, S.M., and Singer, R.H. (2006) Transcriptional pulsing of a developmental gene. *Curr. Biol.* **16**, 1018–1025
38. Ferreiro, D.U. and de Prat-Gay, G. (2003) A protein-DNA binding mechanism proceeds through multi-state or two-state parallel pathways. *J. Mol. Biol.* **331**, 89–99
39. Durr, E., Jelesarov, I., and Bosshard, H.R. (1999) Extremely fast folding of a very stable leucine zipper with a strengthened hydrophobic core and lacking electrostatic interactions between helices. *Biochemistry* **38**, 870–880