Published in final edited form as:

Methods Enzymol. 2011 ; 487: 465–483. doi:10.1016/B978-0-12-381270-4.00016-0.

Inferring mechanisms from dose-response curves

Carson C. Chow[†], Karen M. Ong, Edward J. Dougherty[¶], and S. Stoney Simons Jr.[¶]

[†]Laboratory of Biological Modeling, NIDDK, National Institutes of Health, Bethesda, MD

[¶]Steroid Hormones Section, NIDDK/CEB, National Institutes of Health, Bethesda, MD

Abstract

The steady state dose-response curve of ligand-mediated gene induction usually appears to precisely follow a first-order Hill equation (Hill coefficient equal to 1). Additionally, various cofactors/reagents can affect both the potency and the maximum activity of gene induction in a gene-specific manner. Recently, we have developed a general theory for which an unspecified sequence of steps or reactions yields a first-order Hill dose-response curve (FHDC) for plots of the final product vs. initial agonist concentration. The theory requires only that individual reactions “dissociate” from the downstream reactions leading to the final product, which implies that intermediate complexes are weakly bound or exist only transiently. We show how the theory can be utilized to make predictions of previously unidentified mechanisms and the site of action of cofactors/reagents. The theory is general and can be applied to any biochemical reaction that has a FHDC.

Introduction

Ligand-regulated gene transcription is ubiquitous in biological systems. The dose-response curve of the amount of final gene product expressed versus the amount of ligand present is of crucial importance for development, differentiation and homeostasis. In many cases, the dose-response curve in gene induction obeys a sigmoidal curve but not all sigmoidal curves have the same shape (Goldbeter and Koshland, 1981). For example, a dose-response curve obeying a first-order Hill equation or function (Hill coefficient equal to 1), goes from 10% to 90% of maximum activity over an 81-fold change in ligand concentration whereas the change is only 9-fold in a second-order Hill function, which thus has a different shape (Fig. 1). (A first-order Hill function is sometimes called a Michaelis-Menten function.) The shape and position of a first-order Hill dose-response curve (FHDC) is specified by the potency (i.e., concentration required for 50% of maximal response, or EC_{50}) and maximum activity (A_{max}). These two parameters completely describe the expression of the regulated gene in response to ligand concentration.

The addition of various cofactors can shift the EC_{50} and A_{max} yet preserve the shape of the dose-response curve. These properties put strong constraints on the mechanisms of gene-induction and raise two questions: how can a FHDC arise from a multi-step reaction sequence, and how do cofactors modify potency? To address these questions, we recently developed a general theoretical framework for dose-response curves of biochemical reactions and showed that it is possible for an arbitrarily long sequence of complex-forming reactions to yield FHDCs, provided that a stringent but biologically achievable set of conditions are satisfied. The theory in turn provides a means to make previously unobtainable predictions about the mechanisms and site of action of cofactors that influence the dose-response curve. The FHDC also allows standard methods of enzyme kinetics to be modified for the analysis of FHDCs of arbitrarily long biochemical reaction sequences at steady state.

Although mathematical models have been extensively developed for enzymes, receptor binding, trafficking, and signaling, missing information about downstream steps (such as the phosphorylated proteins and final cellular response) have previously limited mathematical development in this area (Lauffenburger and Linderman, 1993). In contrast, our theory is applicable even when only partial information is available because the constraints of a first-order Hill function and the mechanism of factors permit modeling regardless of whether their position or order in a given cascade of steps is known. The theory also avoids the ‘explosion of parameters’ that usually confounds the search for mathematical models by “telescoping” the unknown intermediate steps to produce a simplified analytical equation with a small set of measurable parameters.

General Theory

The classical explanation for a Hill coefficient of one in steroid-induced gene expression has been that steroid binding to receptor is the rate-limiting step (Baxter and Tomkins, 1971). Consider the reaction $R + S \leftrightarrow RS \rightarrow P$, where R is the steroid receptor, S is the steroid, and P is the final protein product. If the reactions obey mass action kinetics, and the steroid-receptor binding reaction is fast compared to the formation of the product or to the time of product measurement, we can assume that it reaches equilibrium or steady state so that $[RS] = q[R][S]$, where square brackets indicate concentration and q is the affinity or association constant. By mass conservation, $[R] + [RS] = R^T$, where R^T is the total receptor concentration. Combining the steady state and mass conservation equations results in $[R] + q[R][S] = R^T$. Solving for $[R]$ and substituting back into the steady state equation gives

$$[RS] = \frac{qR^T[S]}{1 + q[S]} \quad (1)$$

and the measured protein product is assumed to be proportional to $[RS]$. In this form, the maximal activity ($A_{\max}$) is R^T and the effective concentration for 50% of maximum activity (EC_{50}) is equal to the inverse of the association constant (i.e., dissociation constant) $1/q$. Hence, both EC_{50} and $A_{\max}$ cannot be influenced by downstream cofactors, which is contradicted by experiments (Kim *et al.*, 2006). However, if downstream reactions are included, a FHDC for the final product with respect to the steroid concentration is not guaranteed. For example, steroid-receptor complex binding to DNA initiates a series of transcriptional and translational processes. Even two steps of such a sequence, e.g., $R + S \leftrightarrow RS$, $RS + D \leftrightarrow RSD$, where the second reaction represents binding to DNA, would not generally lead to a FHDC (Ong *et al.*, 2010).

Strickland and Loeb provided a possible solution by examining a two-step process for a system dependent on a secondary mediator such as cyclic AMP (Strickland and Loeb, 1981; Loeb and Strickland, 1987). They showed that if a step following steroid-receptor binding forms an intermediate product Y , where Y itself is also a first order Hill function of $[RS]$, then the result is that $[Y]$ is a first order Hill function of $[S]$ and both $A_{\max}$ and EC_{50} are modifiable by the intermediate process. Hence, if the steady state concentrations of a sequence of reactions are such that the concentration of the product of each reaction is a FHDC with respect to the prior product, then the concentration of the final product will have a FHDC with respect to any prior product (Ong *et al.*, 2010).

We utilized this observation and developed a general theoretical framework of ligand-induced gene expression that could quantitatively model the dose-response curve of the concentration of the expressed protein product as a function of the concentration of applied ligand in the presence of various cofactors or reagents. We modeled gene induction as a

series of steps or reactions that obey mass-action. These reactions include all the processes thought to be involved in gene expression, starting at steroid binding to receptor, ending in the translation of mRNA into protein product, and including, for example, processes such as transcription factor binding to DNA, phosphorylation, methylation, ubiquitination, the uncoiling or unwinding of DNA, the release of paused polymerase, and mRNA processing. We presumed that enough time has elapsed between the addition of the ligand (and any other cofactors/reagents) and the measurement of the final product so that the reactions reach a state of steady state or near steady state. The problem then reduces to solving the system of equations given by steady state and mass conservation equations for all the reactions involved.

Our theory is applied to a sequence of n binary reactions of the form $Y_{i-1} + X_i \leftrightarrow Y_i$, where $i = 1, 2, \dots, n$, is an index for a reaction. We identify Y_0 as the steroid, X_1 as the receptor and Y_1 as the receptor-steroid complex. We call the subsequent X variables activating factors or *activators* and the Y variables *products*. The activators correspond to added or intrinsic agents, including coactivators, comodulators, or small molecules that facilitate an individual reaction locally. We generalize to factors that inhibit local reactions later. The products correspond to states or complexes in the transcription-translation process that are formed or induced by the cofactors and prior products. The reactions need not be reversible; they only should reach steady state. For example, reactions could also have the form $Y_{i-1} + X_i \rightarrow Y_i$, $Y_i \rightarrow *$, where the second step indicates decay or inactivation without addition of a cofactor.

Under steady state conditions governed by mass-action principles, the concentrations obey

$$[Y_i] = q_i [X_i][Y_{i-1}] \text{ and mass conservation implies that } [X_i] + \sum_{k=i}^n [Y_k] = X_i^T \text{ for } i = 1, 2, \dots, n,$$

where the n association constants q_i and the n total concentrations X_i^T are free parameters. The dose-response curve is given by solving the concentration and mass conservation equations simultaneously to obtain $[Y_n]$ as a function of $[Y_0]$. In general, obtaining the functional relationship will require the solution of a 2^{n-1} -degree polynomial that will not yield a FHDC. In order to retain a Hill coefficient of one, each reaction must “disentangle” from the other reactions so that the dose-response curve between any two consecutive products has a Hill coefficient of one and thus any downstream product is always a first order Hill function of any upstream product.

A mechanism that could lead to this condition is that cofactors only act transiently but produce a lasting response. Transient binding leads to an effective reduction in the average concentration of products. Reactions could then have the form $Y_{i-1} + X_i \rightarrow Y_{i-1} + Y_i$. For example, the steroid-receptor complex could bind transiently to DNA but affect the DNA state (e.g., methylation, ubiquitination, uncoiling, untwisting, etc.), facilitate the binding of another cofactor, or alter the mRNA state during translation (Kim *et al.*, 2006). This could also be a multi-step process as in enzyme-substrate reactions (Fromm, 1975; Segel, 1993). The fact that a transient binding event can lead to a delayed action implies that the effect of a factor on the dose-response curve may occur downstream from where the factor actually binds. Considerable experimental evidence has been recently advanced in support of transient binding (dubbed “hit-and-run”) of GR to endogenous genes (Nagaich *et al.*, 2004; Stavreva *et al.*, 2004). Finally, concentration-limiting and transient-binding or hit-and-run mechanisms could both be present at various stages of the reaction sequence.

The result is that any given product is a first order Hill function of any upstream product with the form

$$[Y_i] = \frac{v_i[Y_{i-1}]}{1 + w_i[Y_{i-1}]} \quad (2)$$

where v_i and w_i are defined in Table I (Ong *et al.*, 2010). There is, however, one step that we call the concentration limiting step (CLS), where products after this step obey the form $[Y_i] = v_i[Y_{i-1}]$. In other words, the CLS is the step in which the concentration of the product is much smaller than that for any other reaction. Therefore, the CLS is the equilibrium analogue of the kinetic rate-limiting step, which is that step for which the forward reaction rate is the slowest. For all steps after the CLS, the free concentration of any activator is equal to its total concentration. The location of the CLS depends on the parameters and the experimental situation. It also need not exist for every reaction sequence.

Each of the individual first order Hill functions (2) are substituted into each other sequentially to obtain the formula for the concentration of the final protein product in terms of the concentration of any previous product $[Y_{b-1}]$:

$$[P] = \Gamma[Y_{cls}] = \frac{\Gamma V_b^{cls}[Y_{b-1}]}{1 + W_b^{cls}[Y_{b-1}]}, \quad (3)$$

where $\Gamma = \sum_{k=cls}^n a_{k-cls} V_{cls+1}^k$, $V_b^m = \prod_{i=b}^m v_i$, $W_b^m = \sum_{i=b}^m w_i \prod_{j=b}^{i-1} v_j$, with the convention

$\prod_{i=a}^n x_i = x_a x_{a+1} \cdots x_n$, and $\prod_{i=a}^n x_i = 1$ if $n < a$. Note that $W_b^m = W_b^{cls}$, for $m \geq cls$. As is evident from comparison with (1), $A_{\max} = V_b^m / W_b^m$ and $EC_{50} = 1 / W_b^m$. For steps $m \geq cls$, the denominators of all products $[Y_m]$ are the same. Thus, the sum of any number of products $[Y_{cls}]$ to $[Y_n]$ will preserve first order Hill form. This implies that all reaction sequences with the same W_b^m can be considered to be *parallel pathways*. From a Boolean logic perspective, the cofactors prior to the CLS are equivalent to an “AND” operation in the sense that all cofactors are necessary for gene-induction and cofactors summed in parallel pathways after the CLS are equivalent to an “OR” operation, in the sense that any of the cofactors is sufficient to produce the final product.

The parameters V_b^m and W_b^m are functions of parameters of all the reactions leading up to the final product. Hence, a cofactor influencing any reaction in the entire sequence can affect both the $A_{\max}$ and the EC_{50} . In total there are $4n - cls + 1$ parameters, namely the a , ε , q , and X^T parameters, and the choice of these parameters specifies a *model* of the data. In general, all of these parameters and even the number of steps will not be known. However, the theory is still useful because the number of parameters necessary to specify the model can be reduced by exploiting the property of the first order Hill function that allows any sequence of steps to collapse or “telescope” into a single first order function with an effective $A_{\max}$ and EC_{50} (Ong *et al.*, 2010). Each effective reaction can itself take part in another reaction sequence and in this way arbitrarily complex reaction sequences that maintain a FHDC can be constructed. This telescoping property will be exploited below when comparing the model predictions to the data. In addition, the reaction sequence has a “modular” structure in that reactions, or even a sequence of reactions, can be inserted or deleted without affecting the dose-response curve shape. Thus, different genes could possibly mix and match different pieces of the reaction sequence and each have a first order response but with a different $A_{\max}$ and EC_{50} .

Inhibitors

One property of the FHDC-preserving sequence of reactions is that any type of reaction whose output product has a FHDC with respect to its input product can be inserted into any position of the sequence. Thus, the results of enzyme kinetics theory can be adapted and applied (Fromm, 1975; Segel, 1993). For example, an inhibitor acting at step i leading to product $[Y_i]$ is represented by the following reaction scheme.

where $0 \leq \alpha \leq 1$, $0 \leq \gamma \leq 1$, and $0 \leq \beta$. The case of $\alpha = 0$ (i.e., reactions $Y_i + I_i \leftrightarrow Y'_i$ and $Y_{i-1} + Z_i \leftrightarrow Y'_i$ are missing) is called *competitive* inhibition. The case of $\gamma = 0$ (i.e., reactions $X_i + I_i \leftrightarrow X'_i$ and $Y_{i-1} + X'_i \leftrightarrow Y'_i$ are missing) is called *uncompetitive* inhibition and the case of $\alpha = \gamma$ is called *noncompetitive* inhibition. We refer to the case where α and γ are both nonzero as *mixed* inhibition. Sub-classes of each type of inhibition occur depending upon the value of β . Those situations in which $\beta = 0$ involve linear inhibition (e.g., linear competitive); and, when $\beta > 0$, it is called *partial* inhibition (e.g., partial competitive). Solving this system, and assuming that the two products $[Y_i^*]$ and $[Y'_i]$ are then summed via $[Y_i] = [Y_i^*] + \beta[Y'_i]$, gives a final expression with the form (2) but with different parameters as are listed in Table I. One should note that although I is called an inhibitor, its actions need not be inhibitory or repressive on the final product. For example if $\beta > 0$, I can be activating if it diverts the output from a lower yield product to a higher yield one. In fact, a partial uncompetitive inhibitor is a special case of an activator acting after the CLS. This is because the reaction $Y_i + I_i \leftrightarrow Y'_i$ is an activation equation and $[Y'_i]$ is proportional to $[Y_i^*]$. This allows the two products to be directly summed as activators after the CLS.

Application of model to data

The telescoping property of reaction sequences preserving the FHDC renders application of the theory to experimental data tractable because extraneous and unknown cofactors can be collapsed into a small number of unknown parameters that can be determined experimentally. The result is that there are at most a small number of possible models that can explain a given data set. The models can also be expanded to account for more data. For example, suppose that we are interested in the effect of an activator X_i or inhibitor I_i acting at step i on the induction of the final product. There are only three different model forms that equation (3) can take, namely if the step appears before, at, or after the CLS. The general form of the final product with the activator (X) and inhibitor (I) mathematically expressed is

$$[P] = \frac{(D_0(1+q'_i\gamma[I_i]) + D_1X_i^T(1+q'_i\alpha\beta[I_i]))[Y_0]}{1+q'_i\gamma[I_i] + (D_2(1+q'_i\gamma[I_i]) + D_3(D_4+q'_i\alpha[I_i]) + D_5X_i^T(1+q'_i\alpha\beta[I_i]))[Y_0]} \quad (4)$$

where the D parameters are functions of the parameters of the “hidden” reactions (Ong *et al.*, 2010). Although they can be derived exactly (Ong *et al.*, 2010) in terms of the parameters of the reactions of the entire sequence, they can be treated as effective parameters that can be estimated from the data. Equation (4) shows how the addition of an activator or inhibitor will directly influence the $A_{\max}$ and EC_{50} of the dose-response curve.

We can then infer mechanisms using (4) by observing how a given cofactor influences $A_{\max}$ and EC_{50} . With the addition of a cofactor, $A_{\max}$ and EC_{50} can both either increase, decrease, or not change, leading to 8 different possibilities, not including the trivial case of both not changing. The possible mechanisms leading to these outcomes is listed in Table II (Ong *et al.*, 2010).

Inferring cofactor mechanisms by direct fitting to data

A given cofactor could be an activator or one of several types of inhibitor and act before, at or after the CLS. The theory can generate an analytical model equation for the dose-response curve for protein product with respect to steroid concentration for each of these possible mechanisms. This is also not limited to the action of single cofactors. Analytical model equations can be generated for any number of cofactors. The mechanism of a cofactor can then be inferred by determining which model best fits the data. The most direct approach, which we describe in this section, is to fit the many possible models directly to the data. In the next section, we show how graphical methods adapted from enzyme kinetics theory can be utilized to predict the mechanism.

Here, we use the example of the combined action of comodulator Ubc9 and glucocorticoid receptor (GR) on the dose-response curve of dexamethasone. As previously reported (Kaul *et al.*, 2002; Cho *et al.*, 2005; Kim *et al.*, 2006; Szapary *et al.*, 2008) and displayed in Fig. 2, the effects of increased Ubc9 depend critically on GR concentration and are independent of its sumoylation activity (Kaul *et al.*, 2002; Cho *et al.*, 2005). With low levels of GR, Ubc9 robustly increases $A_{\max}$ while perturbing EC_{50} marginally. At higher GR concentrations, there is less proportional increase in $A_{\max}$ and a much greater decrease in EC_{50} . We can determine possible mechanisms by examining the models that include the combined effect of steroid, receptor, and Ubc9. The constraints for the model are that steroid and receptor are in the first reaction at high GR concentrations and that exogenous Ubc9 increases the $A_{\max}$ and decreases the EC_{50} (see Fig. 3). Matching this behavior to one of the 8 possibilities in Table II shows that this combination of responses must be due to an activator acting before or after the CLS. We, therefore, conclude that Ubc9 is an activator that acts either before or after the CLS. (Note that an activator after the CLS is mathematically equivalent to a partial uncompetitive inhibitor).

The general model for two activators acting before or after the CLS can be written in the form (Ong *et al.*, 2010)

$$A = \frac{\alpha K_1 K_2 D^T R^T (K_3 + K_4 U^T) S^T}{1 + K_1 (1 + K_2 R^T (1 + K_3 + K_4 U^T)) S^T}, \quad (5)$$

where A represents the product (Luciferase activity), U^T represents the total Ubc9 concentration, R^T represents the total steroid receptor concentration, and $[S]$ is the free steroid concentration, which we assume is approximated by the total steroid concentration. We used a Bayesian Markov Chain Monte Carlo method to obtain the free parameter values although any fitting algorithm could be used (details are given in the Appendix). The posterior statistics of the parameters are in Table III. The fit of the predicted data (Fig. 3) is

surprisingly good considering that we assumed a linear relationship between plasmids added to the culture media and their expressed protein concentrations. Parameters K_3 and K_4U^T reflect the amount of product without and with applied Ubc9, respectively. The model fit (see Methods M3 and Table III) finds that K_3/K_4U^T is ~ 1 . This ratio has different implications depending on where Ubc9 acts. If Ubc9 acts before the CLS then this implies that the ratio of endogenous to exogenous Ubc9 in the cell is near 1. However, qRT-PCR measurements give a ratio of ~ 100 (unpublished results), which rules out this possibility. Thus we conclude that Ubc9 is an activator that acts after the CLS, just as was previously proposed (Kaul *et al.*, 2002).

Inferring mechanisms using graphical analysis

Graphical methods have been extensively used in classical enzyme kinetics to extract parameters and infer mechanisms (Segel, 1993). Here we show how plots of $A_{\max}/EC_{50}$, as a function of the concentration of a cofactor, can be used to predict the mechanism of a cofactor.

From equation (4) we find that

$$\frac{A_{\max}}{EC_{50}} = \frac{D_0(1+q'\gamma[I]) + D_1X^T(1+q'\alpha\beta[I])}{1+q'_i\gamma[I]} \quad (6)$$

where all the free parameters are nonnegative and $D_0 = 0$ if the factors act at or before the CLS. The cofactor is either an activator or an inhibitor with concentration X^T or $[I]$, respectively. If the cofactor is an inhibitor then it can be competitive, uncompetitive, noncompetitive, or mixed, and in each case it can further be partial or linear. The plot of $A_{\max}/EC_{50}$ vs. cofactor can be either linear with a positive slope or nonlinear. Figure 4 shows a decision tree that can be followed to determine the mechanism and site of the cofactor based on the shape of the plot. In some cases, other plots are needed to refine the decision, such as $EC_{50}/A_{\max}$ vs. cofactor and $1/((A_{\max}/EC_{50}) - (1/G))$ vs. cofactor, where G is a constant that the nonlinear curve in $EC_{50}/A_{\max}$ vs. cofactor approaches.

An example of the utility of these plots, which offer an alternative approach to the above direct fitting to data, is shown in Fig. 5 for Ubc9. The interpretation of the linear plot for $A_{\max}/EC_{50}$ vs. different concentrations of Ubc9 at a constant amount of cotransfected GR depends upon the position of the true y-intercept. As seen from the Western blot of cell cytosols before and after transfection with Ubc9 plasmid (see insert), 150 ng of transfected Ubc9 plasmid gives at least 10-fold more Ubc9 protein than the endogenous level of Ubc9. This means that the true zero for this plot (i.e., where there is no Ubc9 in the cells) is at the dashed vertical line or even closer to the labeled zero position. Therefore, the plot for $A_{\max}/EC_{50}$ intersects the true $y = 0$ line (dashed) at $y > 0$, in which case Ubc9 is identified as an activator acting after the CLS. This is the same conclusion that was reached using the above method.

Another example concerns the modification of GR induction of the synthetic reporter GREtkLUC by varying concentrations of the coactivator TIF2 in U2OS cells (Fig. 6). The $A_{\max}$ and EC_{50} were determined by exact fits of the amount of induced luciferase activity above basal level to a FHDC for three steroid concentrations (Fig. 5A; data is of ED11-53: this information is for reference only and not for publication). The linear plot for $A_{\max}/EC_{50}$ vs. different concentrations of TIF2 intersects the x-axis at -20 ng of TIF2 plasmid (Fig. 5B). Importantly, densitometric analysis of Western blots for TIF2 from identically treated cells (see insert; ED12-35) indicates that the total TIF2 protein after transfection with 20 ng

of TIF2 plasmid is 2.56 times the endogenous protein level, in which case the endogenous TIF2 is equivalent to 13 ng of TIF2 plasmid. This close agreement between the amount of endogenous TIF2 as determined by Western blot and the x-axis intercept in Fig. 6B leads us to conclude that the plot of $A_{\max}/EC_{50}$, adjusted for total TIF2 concentration that includes the endogenous contribution, would have a y-intercept value equal to zero, in which case (as seen in Fig. 4) TIF2 is acting as an activator before the CLS. This conclusion confirms the widely accepted view of how TIF2 acts. In addition, these assays for Ubc9 and TIF2 say when each factor is acting. Such information about when a factor functions is a unique outcome of our model. Current methods (e.g., ChIP assays) reveal when (and where on the genome) a factor binds but are silent regarding when the factor exerts its biological activity.

Discussion

We have described methods for inferring biological mechanisms for the actions of cofactors in gene-induction reactions based on a theoretical framework for the dose-response curve of gene induction by steroid receptors. Our theory does not require a detailed description of each participating step and/or factor. Nevertheless, it correctly predicts the shape of the dose-response curve. One important result is that all steps before and after an arbitrarily chosen step will telescope down, thereby permitting exact solutions to the equations that can be fit directly to the experimental data.

Our theory was constructed from first principles, starting with the experimental observation that the dose-response curve for steroid action has a Hill coefficient of one. It is strong support for the validity of our theory that, with such minimal initial experimental constraints, the model successfully yields FHDC curves and accounts for all of the different combinations of changing $A_{\max}$ and EC_{50} that we have observed experimentally (Ong *et al.*, 2010). The theory also explains how steps well downstream of steroid binding can alter the $A_{\max}$ and EC_{50} (Kim *et al.*, 2006; Tao *et al.*, 2008) and provides theoretical corroboration for our previous hypothesis of Ubc9 acting downstream of both GR and a rate-limiting step (Kaul *et al.*, 2002), which is further support for the validity of our theory. Furthermore, this theory permits the adaptation of the graphical methods of enzyme kinetics for studying steroid hormone action, as witnessed by the success with analyzing the different mechanisms by which Ubc9 and TIF2 each increase the $A_{\max}$, and decrease the EC_{50} , of GR-mediated transactivation. With these new methods, one can now determine mechanistic details of cofactors, such as when cofactor activity is expressed, that are otherwise inaccessible.

A stringent test for the application of our theory is the invariance of an 81-fold change between 0.1 and 0.9 of maximal induction under all conditions (Goldbeter and Koshland, 1981). This does not hold for a general complex-forming sequence of reactions. Thus, any reaction scheme that is not described by our theory would be unlikely to produce the 81-fold change. Conversely, the observation of a dose-response curve that requires appreciably more or less than an 81-fold change in steroid to change from 0.1- to 0.9-fold of maximal induction (or repression) indicates the presence of at least one step that does not follow first order Hill plot kinetics. Such observations appear to be more common among endogenous genes (He and Simons; Jr., 2007) and provide clues for the unique steroid-regulated behavior of these genes.

Finally, our theory provides a cautionary tale for pure bottom-up modeling and blind high throughput data gathering. Our theory shows that essentially all the reactions in the network determine the $A_{\max}$ and EC_{50} of the dose-response curve. Thus measuring any incomplete subset of the reaction constants is not able to predict the properties of the dose-response curve. However, the theory also shows that if the average concentration of downstream

products is small then the dose-response curve is controlled by only a small number of effective parameters that can be measured directly. The implication is that detailed information about biological mechanisms can be inferred by combining a theoretical framework with a relatively small number of directed experiments.

Acknowledgments

This research was supported by the Intramural Research Program of the NIH, NIDDK.

References

- BAXTER JD, TOMKINS GM. Specific cytoplasmic glucocorticoid hormone receptors in hepatoma tissue culture cells. *Proc Natl Acad Sci USA*. 1971; 68:932–937. [PubMed: 4396919]
- CHO S, KAGAN BL, BLACKFORD JA JR, SZAPARY D, SIMONS SS JR. Glucocorticoid receptor ligand binding domain is sufficient for the modulation of glucocorticoid induction properties by homologous receptors, coactivator transcription intermediary factor 2, and Ubc9. *Mol Endo*. 2005; 19:290–311.
- FROMM HJ. *Initial Rate Enzyme Kinetics*. 1975
- GOLDBETER A, KOSHLAND DEJ. An amplified sensitivity arising from covalent modification in biological systems. *Proc Natl Acad Sci U S A*. 1981; 78:6840–6844. [PubMed: 6947258]
- GREGORY PC. *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica Support*. 2005:312–351.
- HE Y, SIMONS SS JR. STAMP: a novel predicted factor assisting TIF2 actions in glucocorticoid receptor-mediated induction and repression. *Mol Cell Biol*. 2007; 27:1467–1485. [PubMed: 17116691]
- KAUL S, BLACKFORD JA JR, CHEN J, OGRYZKO VV, SIMONS SS JR. Properties of the glucocorticoid modulatory element binding proteins GMEB-1 and -2: potential new modifiers of glucocorticoid receptor transactivation and members of the family of KDWK proteins. *Mol Endocrinol*. 2000; 14:1010–1027. [PubMed: 10894151]
- KAUL S, BLACKFORD JA JR, CHO S, SIMONS SS JR. Ubc9 is a novel modulator of the induction properties of glucocorticoid receptors. *J Biol Chem*. 2002; 277:12541–12549. [PubMed: 11812797]
- KIM Y, SUN Y, CHOW C, POMMIER YG, SIMONS SS JR. Effects of acetylation, polymerase phosphorylation, and DNA unwinding in glucocorticoid receptor transactivation. *J Steroid Biochem Molec Biol*. 2006; 100:3–17. [PubMed: 16723222]
- LAUFFENBURGER DA, LINDERMAN JJ. *Receptors: models for binding, trafficking, and signalling*. 1993
- LOEB JN, STRICKLAND S. Hormone binding and coupled response relationships in systems dependent on the generation of secondary mediators. *Mol Endocrinol*. 1987; 1:75–82. [PubMed: 2842661]
- NAGAICH AK, WALKER DA, WOLFORD R, HAGER GL. Rapid periodic binding and displacement of the glucocorticoid receptor during chromatin remodeling. *Mol Cell*. 2004; 14:163–174. [PubMed: 15099516]
- ONG KM, BLACKFORD JA JR, KAGAN BL, SIMONS SS JR, CHOW CC. A new theoretical framework for gene induction and experimental comparisons. *Proc Natl Acad Sci U S A*. 2010; 107:7107–7112. [PubMed: 20351279]
- SEGEL IH. *Enzyme Kinetics: behavior and analysis of rapid equilibrium and steady-state enzyme systems*. 1993
- SONG L-N, HUSE B, RUSCONI S, SIMONS SS JR. Transactivation specificity of glucocorticoid vs. progesterone receptors: role of functionally different interactions of transcription factors with amino- and carboxyl-terminal receptor domains. *J Biol Chem*. 2001; 276:24806–24816. [PubMed: 11333273]
- STAVREVA DA, MULLER WG, HAGER GL, SMITH CL, MCNALLY JG. Rapid glucocorticoid receptor exchange at a promoter is coupled to transcription and regulated by chaperones and proteasomes. *Mol Cell Biol*. 2004; 24:2682–2697. [PubMed: 15024059]

- STRICKLAND S, LOEB JN. Obligatory separation of hormone binding and biological response curves in systems dependent upon secondary mediators of hormone action. *Proc Natl Acad Sci USA*. 1981; 78:1366–1370. [PubMed: 6262790]
- SZAPARY D, HUANG Y, SIMONS SS JR. Opposing effects of corepressor and coactivators in determining the dose-response curve of agonists, and residual agonist activity of antagonists, for glucocorticoid receptor regulated gene expression. *Mol Endocrinol*. 1999; 13:2108–2121. [PubMed: 10598585]
- SZAPARY D, SONG LN, HE Y, SIMONS SS JR. Differential modulation of glucocorticoid and progesterone receptor transactivation. *Mol Cell Endocrinol*. 2008; 283:114–126. [PubMed: 18215457]
- SZAPARY D, XU M, SIMONS JR SS. Induction properties of a transiently transfected glucocorticoid-responsive gene vary with glucocorticoid receptor concentration. *J Biol Chem*. 1996; 271:30576–30582. [PubMed: 8940029]
- TAO YG, XU Y, XU HE, SIMONS SS JR. Mutations of glucocorticoid receptor differentially affect AF2 domain activity in a steroid-selective manner to alter the potency and efficacy of gene induction and repression. *Biochemistry*. 2008; 47:7648–7662. [PubMed: 18578507]

Appendix

Model fit to data and parameter estimates for Ubc9 and glucocorticoid receptor

We fit the model for two activators (glucocorticoid receptor (R^T) and Ubc9 (U^T) and steroid (S^T) (dexamethasone) (given by equation (4)) to experimental data. The data set consisted of Luciferase activity as a function of S^T , R^T and U^T . In all there were 60 data points taken each in triplicate (Ong *et al.*, 2010). The concentrations of S^T , R^T and U^T in equation (5) are inputs expressed in units of nM. We then fit the model to the data to obtain the K parameters and α . D^T was an extraneous free parameter that was fixed to an arbitrary number. The proportionality constant α represents, the luminosity per mole of output protein luciferase and the proportionality constant of luciferase to the amount of final complex.

The experiments utilize plasmids for GR and Ubc9. In order to fit the model to the data, we assumed that the amount of protein expressed is proportional to the amount of plasmids added. The actual proportionality constant is not important for the fits but for convenience we estimated the concentrations of protein expressed based on the amount of added plasmids. Assuming 660 g/mol per base pair, a well volume of .00033 L, and 6800 b.p. per pSG5/GR plasmid and 5163 b.p./Ubc9 plasmid, we calculated the concentration of plasmid in nM from the mass in ng of plasmid added. We assumed that the concentration of the actual number of proteins translated is directly proportional to the concentration of the plasmid. From these estimates, we equated 0.1, 2, 10 and 25 ng of GR plasmid with 6.75E-5, 0.00135, 0.00675 and 0.0169 nM of GR, and 135 and 175 ng of Ubc9 plasmid with 0.120053 and 0.155625 nM. Again, these values are not important for the model fits. They only change the scale of the parameters. The only assumption that is important is that the amount of plasmid added is proportional to the amount of protein expressed in the cells.

The data was fit using a Bayesian Markov-Chain Monte Carlo method, specifically a variant of the Metropolis-Hastings Algorithm with parallel tempering (Gregory, 2005). Initial priors for the parameters (K_1 , K_2 , K_3 , K_4 , α , D^T) were (0.04285, 1335, 0.346736, 3.549, 101084, 0.080007) and bounded to the range ($[0, 100]$, $[0, \infty]$, $[0, \infty]$, $[0, \infty]$, $[0, \infty]$, $[0, 0.080007]$) with guess ranges of (1, 10, 30, 30, 100000, 0.1). The upper bound of D^T was determined from the concentration of luciferase plasmid and was forced to be below that value. These guess ranges were determined empirically by trial and error to give a reasonable acceptance rate of the algorithm. The parallel tempering was run at different inverse “temperatures” (β)

of 0.00001, 0.001, 0.1, 0.4, 0.7, and 1. The Monte Carlo algorithm was run for 100,000 iterations at each value of β and the first half of the results were discarded, resulting in 77.2%, 75.4%, 18.2%, 16.2%, 15.9%, and 15.6% acceptance rates, respectively. χ^2 values at each β were calculated using the last half of the trial fits (i.e., the last 50,000 values) to allow the transients to decay. The log-likelihood versus β was integrated to obtain the true chi-squared of -2066.27 for this model.

Fig. 1. Shapes of different Hill plots. Computer-generated dose-response curves are shown with Hill coefficients “n” of 0.5, 1, and 2. The dashed lines show 10% and 90% of full activity, which requires a change in ligand concentration of 6561 for $n = 0.5$, 81 for $n = 1$, and 9 for $n = 2$.

Fig. 2.

Varying GR concentration alters the effects of Ubc9 on the A_{max} and EC_{50} . CV-1 cells were co-transfected with GREtkLUC plus the indicated amounts of GR ± Ubc9 plasmids and the averages of triplicate luciferase activities were plotted. The best-fit FHDR curve was determined by Kaleidagraph ($R^2 > 0.96$). The vertical line indicates the EC_{50} for each dose-response curve.

Fig. 3.

Dose-response curves with different amounts of GR and Ubc9. The induction by Dex of transiently transfected with GREtkLUC reporter and different amounts of GR (for A, B, C, and D respectively: 0.1, 2, 10, and 25 ng GR plasmid) $\pm$ Ubc9 plasmid in CV-1 cells was determined as in Fig. 2. Solid Circle = 0 ng Ubc9, Open Square = 135 ng Ubc9, Open Diamond = 175 ng Ubc9. Curves were determined by best fit to mathematical models by a Bayesian Markov Chain Monte Carlo method.

Fig. 4. Decision tree for determining factor mechanism from plots of $A_{\max}$ and EC_{50} . See text for details.

Fig. 5.

Graphical analysis of Ubc9 activity with GR. The $A_{\max}$ and EC_{50} were determined from dose-response curves for GR (10 ng) induction of cotransfected reporter (GREtkLUC) in the presence of the indicated amounts of Ubc9 plasmid as in Fig. 2. The ratio of $A_{\max}/EC_{50}$ was plotted vs. Ubc9 to give a computer-generated straight line ($R^2 = 0.99$) using Kaleidagraph. The insert shows the Western blot detection by anti-Ubc9 antibody of Ubc present in parallel treated cells that had been transfected with 150 ng of Flag/Ubc9 plasmid or an equimolar amount of empty Flag vector. From these Western blot data, it is calculated that the endogenous Ubc9 is $\leq$ the equivalent of 15 ng of Flag/Ubc9 plasmid. The dashed vertical line therefore represents the theoretical point where there is no Ubc9 in the cells.

Fig. 6.

Graphical analysis of TIF2 coactivator activity with GR. (A) Dose-response curves for GR transactivation of GREtkLUC. The program Kaleidagraph was used to fit FHDC curves to the raw Luciferase data for GR induction of transiently transfected GREtkLUC reporter and the indicated amounts of cotransfected TIF2 plasmid in U2OS cells with the indicated amounts of co-transfected TIF2 plasmid ($R^2 > 0.998$). The A_{max} and EC_{50} values generated by the curve-fitting program were then used to plot $A_{max}/(100 \times EC_{50})$ in (B). The straight line represents the computer-generated best fit ($R^2 = 0.94$) using Kaleidagraph. The insert in B shows the Western blots with anti-TIF2 antibody of lysates from cells that had been transfected under the parallel conditions with 0 or 20 ng of TIF2 plasmid.

Table I

Values for v_i and w_i for activator and inhibitor at position i before, at, or after CLS

Position	Activator	Activator with Inhibitor
Before CLS $i < cls$	$v_i = q_i X_i^T$ $w_i = q_i \varepsilon_i$	$v_i = \frac{q_i X_i^T (1 + \alpha_i \beta_i q_i' [I_i])}{1 + \gamma_i q_i' [I_i]}$ $w_i = \frac{q_i (\varepsilon_i + \alpha_i q_i' [I_i])}{1 + \gamma_i q_i' [I_i]}$
At CLS $i = cls$	v_i same as before CLS $w_i = q_i \sum_{k=i}^n \varepsilon_k \prod_{j=i+1}^k v_j$	v_i same as before CLS $w_i = \frac{q_i \left(\sum_{k=i}^n \varepsilon_k \prod_{j=i+1}^k v_j + \alpha_i q_i' [I_i] \right)}{1 + \gamma_i q_i' [I_i]}$
After CLS $i > cls$	v_i same as before CLS $w_i = 0$	$v_i = \frac{q_i X_i^T}{1 + \gamma_i q_i' [I_i]}$ $w_i = 0$

Table IIEffect of cofactors on $A_{\max}$ and EC_{50}

$A_{\max}$	EC_{50}	Cofactors and context	Mechanism and position
Decrease	Decrease	CBP (with GR) (Szapary <i>et al.</i> , 1999), NCoR (with PR) (Song <i>et al.</i> , 2001)	L and U or A after CLS
Decrease	Increase	GMEB2, NCoR (with GR) (Kaul <i>et al.</i> , 2000), CPT, H8, DRB (with high GR and Ubc9) (Kim <i>et al.</i> , 2006)	C or L before or at CLS
Decrease	No Change	DRB, H8 (with high GR), VPA (with high GR and Ubc9) (Kim <i>et al.</i> , 2006)	P and U before CLS or A after CLS
Increase	Decrease	TIF2 (with GR) (Szapary <i>et al.</i> , 1996; Szapary <i>et al.</i> , 1999), Ubc9 (with high GR) (Kaul <i>et al.</i> , 2002; Kim <i>et al.</i> , 2006)	A before or after CLS
Increase	Increase	Not observed but predicted	C after CLS
Increase	No Change	TSA, VPA, Ubc9 (with low GR) (Kaul <i>et al.</i> , 2002; Kim <i>et al.</i> , 2006)	A at or after CLS
No Change	Decrease	SRC-1 (with GR) (Szapary <i>et al.</i> , 1999)	A after CLS
No Change	Increase	TIF2 siRNA (with GR) (Luo and Simons, Human Immunology, in press)	L or C anywhere

Note: Abbreviations: A, activator; I inhibitor; L, linear inhibitor; P, partial inhibitor; C, competitive inhibitor; U, uncompetitive inhibitor.

Table III

Posterior statistics of parameters for Ubc9 fits

Parameter	Mean	Median	Standard Deviation	Interquartile Range	Max Likelihood Value
K_1	0.041	0.040	0.0012	0.00076	0.039
K_2	1435.08	1430.32	60.47	75.62	1484.67
K_3	0.37	0.376871	0.00730698	0.0071	0.38
K_4	4.46	4.45226	0.389866	0.5133	4.45
α	150807	160449	35508.1	64995.3	183930
Dt	0.054	0.047	0.013	0.025	0.041