Parameterization of a Nonlinear Genotype to Phenotype Map Using Molecular Networks

J. Peccoud and K. Vander Velden

Pacific Symposium on Biocomputing 10:284-295(2005)

PARAMETERIZATION OF A NONLINEAR GENOTYPE TO PHENOTYPE MAP USING MOLECULAR NETWORKS

JEAN PECCOUD, KENT VANDER VELDEN

Pioneer Hi-Bred Int'l, Inc.7250 NW 62nd Ave, Johnston, IA 50131-0552,

USA

E-mail: jean.peccoud@pioneer.com , kent.vandervelden@pioneer.com

Mathematical models of networks of molecular interactions controlling the expression of traits could theoretically be used as genotype to phenotype (GP) maps. Such maps are nonlinear functions of the environment and the genotype. It is possible to use nonlinear least square minimization methods to fit a model to a set of phenotypic data but the convergence of these methods is not automatic and may lead to a multiplicity of solutions. Both factors raise a number of questions with respect to using molecular networks as nonlinear maps. A method to fit a molecular network representing a bistable switch to various types of phenotypic data is introduced. This method relies on the identification of the model stable steady states and the estimation of the proportion of cells in each of them. By using environmental perturbations, it is possible to collect time-series of phenotypic data resulting in a smooth objective function leading to a good estimate of the parameters used to generate the simulated phenotypes.

1 Introduction

Pharmacogenomics' ambition is to relate a phenotype, the effect of a drug, to the genotype of patients exposed to environmental conditions partly defined by the drugs they receive [1]. For a geneticist this project requires building a genotype to phenotype map (GP map) of drug effects. Mathematically, a GP map is a function f such as *phenotype* = f(*genotype,environment*). It maps into a phenotypic space, the product of a genetic space generated by the genetic diversity in a population by the space of environmental conditions to which individuals of this population can be exposed [2]. The simplest GP map is the one upon which relies Mendelian genetics. The function is Boolean indicating the presence or absence of a character. The environment is ignored and genes are considered independent of each other. Since most traits are quantitative and not binary, the genetics of quantitative traits relies on a more refined family of GP maps representing the phenotype as linear statistical models. In general multiple loci are assumed to contribute additively to the phenotype. In some cases terms representing digenic interactions are introduced. The effect of the environment on the phenotype is generally decomposed into an additive term and a genotype by environment term [3].

Just like complex interactions between multiple genetic loci generate a diversity of phenotypes for pathologies that were considered monogenic [4], responses to drugs are generally considered multigenic traits [1, 5]. Many of the genetic determinants controlling the response to drugs have been identified by a candidate-gene approach relying on the understanding of the molecular mechanisms

of the drug action and metabolism. Integrating into a mathematical model the network of molecular interactions affecting the response to a drug is therefore an attractive avenue to build the GP map.

Using different approaches, a number of authors have recently demonstrated that it is possible to build mathematical models to predict the phenotype controlled by small artificial gene networks [6-10], larger natural networks [11, 12], or even genome-wide metabolic pathways [13, 14]. In order to use a mathematical model as a GP map it is necessary to bridge the molecular and population-levels views of the genotype-phenotype relationship. When using mass-action models of molecular interactions, it has proved possible to analyze the genetic properties of a molecular network by associating genetic polymorphism with discrete kinetic values of the parameter of each interaction [15]. The possibility of determining the kinetic parameters of each interaction is key to using molecular networks as GP maps.

One way to estimate the GP map parameters is to find a set of parameters minimizing the difference between the phenotype predicted by the model and the observed phenotype. Since the phenotype is a nonlinear function of the parameters, this problem can be addressed by using a nonlinear least-square approach [16, 17]. Nonlinear minimization methods are iterative algorithms that require a set of starting parameter values to converge to a local solution. Different starting values can result in different solutions with different quality of fit. This limitation has the potential to prevent a unique determination of the map parameters. The topology of the molecular network model and the experimental design both contribute to shape the objective function being minimized. The number and geometry of its local minima determines the possibility to find and identify solutions corresponding to the actual parameters' values that generated the set of observed phenotypes. Since for many real molecular networks, it is not possible to explore the entire parameter space, it is possible that no starting parameter values will converge toward the actual parameter set. It is also possible that many starting values will result in many solutions with similar fits making it impossible to distinguish the solutions closest to the actual parameter set. Few authors used nonlinear least-square minimization to estimate GP map parameters [12, 18] and it is likely that a number of people attempted this without success and never published these negative results.

This paper introduces an algorithm to estimate the parameters of a molecular network from time-series of molecular phenotypes collected after an environmental perturbation. The objective function used takes into consideration the possibility that phenotypic data collected at the cell population level result from a random distribution of the cells among multiple stable-steady states. The presence of a positive feed-back loop [19] creates the possibility of multistationarity. Multiple steady states have been observed in artificial gene networks [8-10, 20, 21] but also in natural regulatory networks [11], for which this possibility had not been considered even recently [12].

The algorithm considered in this article is automatic and can be applied to virtually any mass action model of molecular networks without requiring any manual mathematical derivation.

2 Methods

2.1 Model

The model used in this article is a mass action equivalent of a model of a bi-stable switch [9, 20, 21]. In the list of reactions below, G_i and GX_i refer to the active and inactive forms of the ith gene coding for the protein P_i , respectively, while L_i represents the ith ligand and PX_i the ith protein complexed with its ligand.

$$\begin{array}{c} R_{1}:G_{1} & \xrightarrow{k_{1}} \rightarrow G_{1} + P_{1} \\ R_{2}:G_{2} & \xrightarrow{k_{2}} \rightarrow G_{2} + P_{2} \end{array} \end{array} Gene expression \\ R_{3}:P_{1} & \xrightarrow{k_{3}} \rightarrow \varnothing \\ R_{3}:P_{1} & \xrightarrow{k_{3}} \rightarrow \varnothing \\ R_{4}:P_{2} & \xrightarrow{k_{4}} \rightarrow \varnothing \end{array} \end{array} Protein degradation \\ R_{4}:P_{2} & \xrightarrow{k_{4}} \rightarrow \varnothing \end{array} Protein degradation \\ R_{5},R_{6}:G_{2} + 2P_{1} \xrightarrow{k_{5}} GX_{2} \\ R_{5},R_{6}:G_{1} + 2P_{2} \xrightarrow{k_{7}} GX_{1} \\ R_{9},R_{10}:P_{1} + L_{1} \xrightarrow{k_{9}} PX_{1} \\ R_{11},R_{12}:P_{2} + L_{2} \xrightarrow{k_{11}} PX_{2} \end{array}$$
(1)

The time-evolution of the model is represented by mass-action differential equations. The set of coupled differential equations can automatically be derived from the chemical equations Eq. (1) [15].

Mass conservation relationships can be used to eliminate some variables from the model. Assuming that there is only one copy of each of the two genes in the system, the first mass-conservation relation makes it possible to eliminate the repressed forms of the genes. We also assume that the interaction between the small molecules representing the environment and the repressors are much faster than the other reactions. Using a quasi-steady state approximation, we eliminate R_9 to R_{12} from the model. This results in the list of reaction rates below where **X** is the vector representing the state of the system and r_i the rate of the reaction R_i :

with
$$\mathbf{X} = \begin{pmatrix} G_1 \\ P_1 \\ P_2 \\ G_2 \end{pmatrix}, \begin{cases} r_1(\mathbf{X}) = k_1 G_1 & r_5(\mathbf{X}) = k_5 G_2 \left[\frac{1}{1+L_1} P_1 \right]^2 \\ r_2(\mathbf{X}) = k_2 G_2 & r_6(\mathbf{X}) = k_6 (1-G_2) \\ r_3(\mathbf{X}) = k_3 P_1 & r_7(\mathbf{X}) = k_7 G_1 \left[\frac{1}{1+L_2} P_2 \right]^2 \\ r_4(\mathbf{X}) = k_4 P_2 & r_8(\mathbf{X}) = k_8 (1-G_1) \end{cases}$$
 (2)

The differential equation representing the time evolution of the system is derived from this list of reaction rates.

2.2 Numerical Identification of the Steady States

The most generic way of finding steady states is to find the solutions of Eq. (3) below. The notation below indicates that the reaction rates depend on the parameterization of the model, $\mathbf{K} = (k_1, ..., k_8)$, and the environment, $\mathbf{E} = (L_1, L_2)$:

$$\frac{\mathrm{d}\mathbf{X}}{\mathrm{d}t} = \mathbf{F}(\mathbf{X}, \mathbf{E}, \mathbf{K}) = \begin{pmatrix} r_8(\mathbf{X}) - r_7(\mathbf{X}) \\ r_1(\mathbf{X}) - r_3(\mathbf{X}) + 2(r_6(\mathbf{X}) - r_5(\mathbf{X})) \\ r_2(\mathbf{X}) - r_4(\mathbf{X}) + 2(r_8(\mathbf{X}) - r_7(\mathbf{X})) \\ r_6(\mathbf{X}) - r_5(\mathbf{X}) \end{pmatrix} = 0$$
(3)

Roots can be determined by minimizing $\|\mathbf{F}(\mathbf{X})\|$ starting from any point in the model state space. Since Eq. (3) is nonlinear, it is not possible to analytically find its solutions. In order to alleviate this limitation, a grid of starting points is created in a region of the state space expected to include all the biologically relevant steady states of the model.

Variables corresponding to conserved molecules are bounded by the initial conditions. Assuming that each gene in the model has a single copy, then $0 \le G_i \le 1$ with i = 1, 2. The asymptotic values of the non-conserved molecules, i.e. proteins in this case, is somewhere between 0 and $k_{production}/k_{degradation}$, the asymptotic value corresponding to the maximum expression of the gene. Therefore, in the case of the model considered here, all the steady states are expected to be within $V = [0,1] \times [0, k_1/k_3] \times [0, k_2/k_4] \times [0, 1]$.

It is therefore possible to regularly sample V with a user-specified resolution. By starting the minimization algorithm from each point in this grid, a numerical solution to Eq. (3) will generally be found for each starting point. Numerical errors and differences of convergence toward the same limits will result in minor numerical differences between solutions reached from different starting conditions. If the distance between a solution and another previously found solution is less than some specified value, it is assumed that they are identical. After the scan of V is complete, the stability of the steady states is analyzed by computing, at the steady state, the eigenvalues of the Jacobian matrix associated to Eq. (3). If the real parts of all eigenvalues are negative, then the steady state is stable.

2.3 Fitting to Asymptotic Phenotypes

In the context of this article, "asymptotic phenotypes" refers to phenotypic data collected in the stationary regime [12, 22] in different environments E_j with j = 1, ..., v. Since in general, all variables of the model cannot be observed, the number of data points collected in each environment μ is less than M the total number of state-variables of the model. It is convenient to represent asymptotic phenotypes as a $\mu X v$ matrix **P**. Now that the experimental data set is structured, it is necessary to generate a predicted phenotype Q(K) corresponding to a given set of parameters **K**. Assuming that it is possible to compute Q(K), then the least-square distance that needs to be minimized to fit the model to the phenotypes, d(K, P), is:

$$d(\mathbf{K}, \mathbf{P}) = \sum_{i=1}^{\mu} \sum_{j=1}^{\nu} \left[Q_i(K, E_j) - P_i(E_j) \right]^2$$
(4)

Computing the predicted phenotype for a specified environment and set of parameters is immediate if they result in a single stable steady state **S**. In this case: $\mathbf{S}_i(\mathbf{K}, \mathbf{E}_i) = \mathbf{Q}_i(\mathbf{K}, \mathbf{E}_i)$ $i = 1, ..., \mu$ $j = 1, ..., \nu$.

In conditions where the model has two stable steady states S and T, then the observed phenotype P is likely to result from a distribution of cells in the two steady states. So, instead of having a direct correspondence between the predicted phenotype and the observed phenotype, the predicted phenotype is a weighted average of the two stable steady states. What is not known, though, is the proportion of cells in each of the steady states. This proportion needs to be estimated by solving a linear constrained least-square problem:

$$\mathbf{Q}(\mathbf{K}, \mathbf{E}) = \min_{\alpha \in [0,1]} \left(\alpha \mathbf{S}(\mathbf{K}, \mathbf{E}) + (1 - \alpha) \mathbf{T}(\mathbf{K}, \mathbf{E}) - \mathbf{P}(\mathbf{E}) \right)$$
(5).

This approach can be generalized to more than two stable steady states.

2.4 Fitting to a Time Series of Phenotypes

Observing the model state variables at different points in time is a natural way of collecting data characterizing the model dynamics [9, 20]. Many experimental designs can lead to this type of data. Only a single simple experiment is considered in this paper but it demonstrates that system multi-stationarity needs to be considered to properly analyze the data.

A cell population is placed in a first environment E_1 until it reaches a stationary regime indicated by the stabilization of the phenotype. An instantaneous perturbation is applied to the environment creating a new environmental condition E_2 . Phenotypic data are recorded at different time points while the population stabilizes toward a new stationary regime. For instance, cells can be grown in absence of ligands. One of the ligands is added to the growth medium creating a new environment. Samples of cell culture are taken and phenotyped at different points in time after the ligand has been added. This design can be generalized to multiple environmental perturbations. E_{1j} and E_{2j} refer to first and second environments of the jth perturbation. The first phenotype of each time series is collected in the stationary regime before the perturbation is applied. All other phenotypes are collected in the second environment and are indexed by the instant of observation. Similarly, it is necessary to compute a series of predicted phenotypes corresponding to the series of experimental data. The distance between the predicted and the observed phenotypes is computed by summing the distance over all time-points:

$$d\left(\mathbf{K},\mathbf{P}\right) = \sum_{k=1}^{\tau} \sum_{i=1}^{\mu} \sum_{j=1}^{\nu} \left[\mathbf{Q}_{i}\left(\mathbf{K},\mathbf{E}_{2,j},t_{k}\right) - \mathbf{P}_{i}\left(\mathbf{E}_{2,j},t_{k}\right) \right]^{2}$$
(6)

Let $G(\mathbf{X}_0, \mathbf{K}, \mathbf{E}, t)$ be the solution of Eq. (3) starting from \mathbf{X}_0 . Computing the predicted phenotype for a specified environmental perturbation and set of parameters is immediate if the initial environment and parameter set result in a single stable steady state $\mathbf{S}(\mathbf{K}, \mathbf{E}_1)$. In this case the predicted phenotypes are extracted from the solution of Eq. (3) starting at $\mathbf{S}(\mathbf{K}, \mathbf{E}_1)$: $G(\mathbf{S}(\mathbf{K}, \mathbf{E}_1), \mathbf{K}, \mathbf{E}_2, t_k) = \mathbf{Q}_i(\mathbf{K}, \mathbf{E}_{2,i}, t_k)$. If the parameter set leads to two steady states in the initial environment $\mathbf{S}(\mathbf{K}, \mathbf{E}_1)$ and $\mathbf{T}(\mathbf{K}, \mathbf{E}_1)$, then it is possible to estimate the proportion $\hat{\alpha}$ just as in Eq. (5). The predicted phenotype would then be a weighted average of trajectories starting from the two initial conditions \mathbf{S} and \mathbf{T} .

2.5 Application

The number of variables observed in the phenotype and the number of environments where the phenotypes are observed are likely to have a significant impact on the possibility to match the model with phenotypic data. So, phenotypic data were simulated in different numbers of environments and by recording different numbers of observed variables.

Twelve series of phenotypic data were generated using the same set of parameters. The first 6 phenotypes were asymptotic phenotypes. The second group of 6 phenotypes were time series.

In both cases (asymptotic and time series), three of the phenotypes consisted in the observation of one protein, P_1 . In the remaining three phenotypes the values of both proteins were recorded in the phenotype.

The asymptotic phenotypes were simulated in three different numbers of environments (3, 5, and 9 environments). Environments are represented by the concentrations of the two ligands, (L_1, L_2) . The first three environments were: (0,0), (10¹, 0), and (0, 10¹). In the 5 environments experiments, (1, 0) and (0, 1) were added to the first 3 environments. In the 9 environments experiments (10⁻¹, 0), (10⁻², 0), (0, 10⁻¹), and (0, 10⁻²) were added to the five previous environments.

The times series phenotypes are transitions between two environments. In the first experiment, the transition from (10, 0) to (0, 10) was simulated. In the 2-transition experiment, the transition from (5, 0) to (0, 5) was added. In the 3-transition experiment, the transition from (1, 0) to (0, 1) was added to the two previous transitions.

The same set of 25 initial parameter values was used to fit the model to the asymptotic and time-series phenotypes resulting in a series of 300 optimizations.

3 Results

3.1 Numerical Identification of Steady States

The method to find the steady states of the model works well on this model. By using only the 8 "corners" of V, it seems that all the steady states of the system were found. Increasing the resolution of the grid did not result in a larger list of steady states. Depending on the environment and parameter values, two types of regimes were found: a single stable steady state or two stable steady states and one unstable steady state.

In the least-square minimization procedures, the specificity of this network made it possible to use only two initial conditions $(0.5, k_1/k_3, 0, 0.5)$ and $(0.5, 0, k_2/k_4, 0.5)$ to find the stable steady states of the system. This simplification speeds up the optimization process that often requires hundreds or even thousands of steady state determinations. These two initial conditions do not allow the identification of the unstable steady states of the system and this approach may not be applicable to other models.

A bifurcation diagram was generated by computing the steady states (stable and unstable) of the model over a range of L_1 concentrations in order to verify the steady state identification procedure while the concentration of the second ligand was kept at 0. The system is bi-stable for low concentrations of L_1 and beyond a critical concentration, the system becomes mono-stable. This result is consistent with the bifurcation diagram of a similar model [20] and also with our own bifurcation analysis run in XPP/AUT [23]. The positions of the stable steady states are not very much affected by the concentration of L_1 , except in the vicinity of the critical concentration. This indicates the robustness of the phenotype to environmental perturbation.

3.2 Fitting to Asymptotic Phenotypes

An exploration of the neighborhood of the original set of parameters used to generate the phenotypes indicated that initial conditions very close to the original parameter set could not lead to a good fit (data not shown). This indicated that the objective function was rough and may be difficult to minimize. It turned out that convergence was much easier to achieve than initially anticipated. When the phenotype included the two protein concentrations a good fit was achieved for 1/3 of the initial conditions.

This can be explained by observing that an infinite number of parameterizations have the same steady states. Solutions of Eq. (3) verify:

$$r_{8}(\mathbf{X}) - r_{7}(\mathbf{X}) = k_{8}(1 - G_{1}) - k_{7}G_{1}\left[\frac{1}{1 + L_{2}}P_{2}\right]^{2} = 0$$

$$r_{1}(\mathbf{X}) - r_{3}(\mathbf{X}) = k_{1}G_{1} - k_{3}P_{1} = 0$$

$$r_{2}(\mathbf{X}) - r_{4}(\mathbf{X}) = k_{2}G_{2} - k_{4}P = 0$$

$$r_{6}(\mathbf{X}) - r_{5}(\mathbf{X}) = k_{6}(1 - G_{2}) - k_{5}G_{2}\left[\frac{1}{1 + L_{1}}P_{1}\right]^{2} = 0$$
(7)

In other words, the minimization problem defined by asymptotic phenotypes is unidentifiable. It is not possible to estimate the 8 kinetic parameters but only the 4 equilibrium constants.

3.3 Fitting to Time Series of Phenotypes

The convergence criteria used in this case was a root mean square of residuals less than 10^{-1} . Using this criterion, 14 convergences were observed (9% of the 150 optimizations using time series phenotypes) that can be broken down into 13% of convergence when only one protein is observed and 5% when both proteins are recorded. These rates of convergence need to be confirmed by analyzing a larger number of initial conditions using a faster implementation of this algorithm. However, they are surprisingly high and indicative of a relatively smooth performance function.

All optimization solutions were indexed (not shown) for further analysis. In some cases very similar solutions were found. For instance solution 13 is very close to solution 14 and solution 11 is very close to solution 12. It is worth observing that if solutions 11, 13, and 14 all originated from the same initial condition, solution 12 was found using a different initial condition. Also solutions 11 and 12 are not very far in the parameter space from solutions 13 and 14. Solution 6 is also located in the

same area. Interestingly, these 5 solutions are all very close to the original set of parameters used to generate the phenotype. The solutions were verified by plotting the time course of the two protein concentrations and the profiles are consistent with the objective function used to generate the solutions. Protein concentrations corresponding to solution 11 were plotted over a wide range of initial conditions. Visually they are indistinguishable from the plots generated by the original set of parameters (Figure 1).



Figure 1: In order to visually assess the quality of the fit, the ODE was integrated using two solutions of the time-series optimization experiment and the original set of parameters used to generate the simulated phenotypes. The initial condition for the integration was set to (1, 0, 10, 0) and the environment to (0, 10). Solution 6 (top) was found when only one protein level was used in the phenotype. It is interesting to see that the fit for P_1 is better than the fit for P_2 . The RMS computed using the two protein concentrations at the 11 time points is 0.83. Solution 11 (bottom) gives a very good fit of both of the protein expression profiles leading to a RMS of 0.06. It is necessary to zoom in on specific region of the plot to be able to visually distinguish the trajectories generated by the original parameter set and the trajectories generated by the parameters of Solution 11.

4 Discussion

4.1 Results

Even though this work focuses on a single molecular network model, results presented here are likely to be relevant to other models.

- The specific structure of molecular networks makes it possible to search for steady states in a limited volume of the model state space.
- The possibility of multi-stability should always be considered. In a population of cells observed in a stationary regime, cells can be randomly distributed between multiple steady states. Therefore, the measurement of a gene activity at the cell population level is a weighted average of the molecule concentrations corresponding to the different stable steady states of the model. For a given set of parameter values, different repartitions of the cells in the different steady states leads to different qualities of fit between the model parameterization and the observed phenotypes. In the context of this paper, a linear minimization step was introduced to find the repartition minimizing the distance between the model and the experimental data.
- Asymptotic phenotypic data can only lead to the determination of the equilibrium constants but not the kinetic constants.
- Environmental perturbations can be used to collect time-series of phenotypic data. The relaxation profile observed is a weighted average of trajectories originating from the different stable steady states in the first environment.

4.2 Necessary improvements of the algorithm

In order for this method to be used for routine analysis it will be necessary to address a few issues.

- The steady state finding algorithm needs to be systematically validated. In some cases very stiff parameter sets hampered the convergence of the steady state identification procedure. The reasons for this behavior need to be understood. Since the steady state identification algorithm is the bottleneck of the whole optimization process it is worth trying to improve it.
- Determining the stability of the steady states is also an important step of the algorithm. Numerical errors prevent an accurate determination of the stability in the vicinity of critical points. It is not clear what is the impact of this issue on the outcome of the minimization process. Limit cycles are not considered in this algorithm.
- The local optimization method described in this paper needs to be coupled to a global search strategy to explore the parameter space more systematically.
- In cases where the time of sampling cannot be controlled, it could be necessary to take the actual sampling time into consideration when fitting the model to the data.

• A random term representing the measurement error needs to be added to the phenotypic data. The effect of this term on the convergence of the least-square minimization should be characterized. The addition of an error term would transform the least-square minimization problem into a nonlinear regression problem that could lead to computing confidence intervals for the parameter estimates.

4.3 Research directions

We are working on a generalization of this algorithm to handle phenotypic data collected on a multiplicity of genotypes just like several environmental conditions have been considered in this paper. Along the same line, the current model assumes only one copy of each gene. Introducing a diploid genome with two homologous copies of each gene would require predicting the phenotype of heterozygous individuals, which requires developing a model of dominance at the parameter level. If only homozygous individuals are considered or a total dominance is assumed, the model would remain unchanged.

Geneticists have been building models of the genotype to phenotype relationship for traits of other organisms for more than a century. By deciphering networks of molecular interactions, they hope to be able to build nonlinear GP maps inspired by the mechanisms controlling the expression of complex traits. It is expected that these maps would capture epistatic interactions between the genetic determinants contributing to these traits. Such a map would help a plant breeder define more effective breeding strategies using molecular markers to manipulate alleles of genes contributing to trait variations or using transgene to introduce new sources of genetic variation, a human geneticist better understand how multiple genes can contribute to the development of a pathology, and pharmacogeneticists to customize a medication to the genotype of their patients. Mathematical methods, such as those described here, are needed to analyze molecular data. The next challenge may be to find ways of associating macroscopic phenotypes such as a patient response to a treatment, with the molecular data we collect and analyze.

Acknowledgements

We are grateful to Mark Cooper, Chris Winkler, Dean Podlich, David Bickel, Bard Ermentrout and four anonymous reviewers for valuable comments and suggestions. This work would not have been possible without the support of Lane Arthur and Bob Merrill.

References

1. W. E. Evans, M. V. Relling, Nature 429, 464 (2004).

- M. Cooper, S. C. Chapman, D. W. Podlich, G. L. Hammer, *In Silico Biol.* 2, 151 (2002).
- 3. D. S. Falconer, T. F. C. MacKay, *Quantitative Genetics* (Longman Group Ltd., Harlow (U.K.), ed. 4, 1996).
- 4. D. J. Weatherall, Nat. Rev. Genet. 2, 245 (2001).
- 5. W. E. Evans, H. L. McLeod, N Engl J Med 348, 538 (2003).
- 6. J. Hasty, D. McMillen, J. J. Collins, Nature 420, 224 (2002).
- 7. M. KAErn, W. J. Blake, J. J. Collins, Annu. Rev. Biomed. Eng. 5, 179 (2003).
- 8. M. R. Atkinson, M. A. Savageau, J. T. Myers, A. J. Ninfa, Cell 113, 597 (2003).
- 9. B. P. Kramer et al., Nat. Biotechnol. (2004).
- F. J. Isaacs, J. Hasty, C. R. Cantor, J. J. Collins, *Proc. Natl. Acad. Sci. U. S. A* 100, 7714 (2003).
- E. M. Ozbudak, M. Thattai, H. N. Lim, B. I. Shraiman, A. van Oudenaarden, *Nature* 427, 737 (2004).
- Y. Setty, A. E. Mayo, M. G. Surette, U. Alon, *Proc. Natl. Acad. Sci. U. S. A* 100, 7702 (2003).
- M. W. Covert, E. M. Knight, J. L. Reed, M. J. Herrgard, B. O. Palsson, *Nature* 429, 92 (2004).
- 14. K. R. Patil, M. Akesson, J. Nielsen, Curr. Opin. Biotechnol. 15, 64 (2004).
- 15. J. Peccoud et al., Genetics 166, 1715 (2004).
- 16. D. M. Bates, D. G. Watts, *Nonlinear regression analysis and its applications* (John Wiley & Sons, New-York, 1988).
- 17. J. E. Dennis, R. B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations* (Prentice Hall, Inc., Englewood Cliffs, NJ, 1983).
- M. Ronen, R. Rosenberg, B. I. Shraiman, U. Alon, *Proc. Natl. Acad. Sci. U. S. A* 99, 10555 (2002).
- 19. D. Thieffry, R. Thomas, Pac. Symp. Biocomput. 77 (1998).
- 20. T. S. Gardner, C. R. Cantor, J. J. Collins, Nature 403, 339 (2000).
- 21. R. N. Tchuraev, I. V. Stupak, T. S. Tropynina, E. E. Stupak, *FEBS Lett.* **486**, 200 (2000).
- 22. C. C. Guet, M. B. Elowitz, W. Hsing, S. Leibler, Science 296, 1466 (2002).
- 23. B. G. Ermentrout, Simulating, Analyzing, and Animating Dynamical Systems: A Guide to Xppaut for Researchers and Students (SIAM, ed. 1, 2002).