The Importance of Hierarchical Time Scales on the Nonlinear Dynamics of Transcription Factor Complex Formation

Shankar Mukherji
Harvard University/Massachusetts Institute of Technology,
Division of Health Science and Technology Cambridge, MA 02139
mukherji@mit.edu
(Dated: May 12, 2005)

We present a simple model of the formation of transcription factor binding comlexes. The model consists of a set of six coupled nonlinear differential equations obtained from knowledge of the fundamental processes that make up the complexing event. The equations are subjected to a fixed point analysis, and we find that the fixed points exhibit many bifurcations. The equations are then numerically integrated to obtain time trajectories for the relevant species. The numerical integration shows that if there exists a hierarchy of time scales separating the basic complex formation processes, transcriptional activity is stable even with low copy numbers of the transcription factor binding proteins.

INTRODUCTION

The regulation of gene expression is among modern biology's foremost problems. As years of data on various regulatory schemes have built up, there has been a shift in the biological community from focussing purely on the molecular actors that regulate the transcription and translation of genes to a broader approach that aims to classify the schemes in terms of modules, which can then be hierarchically assembled to reconstruct the entirely regulatory network of a living system [1].

A particularly fruitful line of attack in studying such modular, biological networks has been achieved from a dynamical systems theory point of view. Within this framework, the relevant species are recast as dynamical variables governed by a set of coupled, nonlinear ordinary differential equations. Such nonlinear dynamical systems have been shown to model such biologically relevant situations as the asychrony-synchrony dynamical phase transition in firefly lighting [2] as well as mutually repressive genes that yield oscillatory behaviour [3].

In the model presented below, the differential equations that will be analysed according to the techniques of nonlinear dynamics are derived from simple chemical kinetic rate equations. Following the specification of the model, the overarching theme of the analysis is the generation of distinct, hierarchical time scales from the strongly coupled reactions.

TRANSCRIPTION FACTOR COMPLEXING

A widely encountered method of regulating gene expression levels involves the formation of complicated transcription factor complexes at the promoter regions of genes. Under such a scheme, gene activation or repression does not occur until all the appropriate transcription factors are recruited to the site of the promoter region of the gene. Concrete examples of such transcription factor complexing events in real biological systems include the interaction of the haploid α-cell MCM1 protein with the homeodomain-containing protein MATα2 that act in repressing the genes specifying the **a** mating type, and the interaction of SOX proteins with POU elements in controlling such diverse processes as the maintainance of the pluripotency of stem cells, the development of the immune system and cell type during the development of the immune system [4].

In our study below, we will consider processes in which the protein-protein interaction that operates on the transcription factor binding DNA boosts the residence time of the transcription factor on the DNA compared to the residence time of the uncomplexed residence time. Such phenomena are observed in biological cells, leading to the hypothesis that perhaps the boosted residence time explains why transcription factor binding proteins exist in such low copy number in the endogenous cell [5]. The claim is that the low copy number of the transcription factor binding protein serves as a filter for gene expression levels by imposing a theoretical maximum rate of transcription; presumably then the mean copy number of the transcription factor binding protein would be tuned and modulated over time by evolution. Our goal below is to examine the validity of the claim that transcription factor binding protein-induced boosts of residence time of transcriptional complexes on the DNA can exhibit reasonable biochemical behaviour even with low copy number of the transcription factor binding protein itself. We will attack the problem by examining the fixed point structure of the simple model and then perform numerical integrations of the differential equations to validate the fixed point picture.

MODEL

As mentioned above, the mathematical model used to capture the dynamics of the transcription factor binding

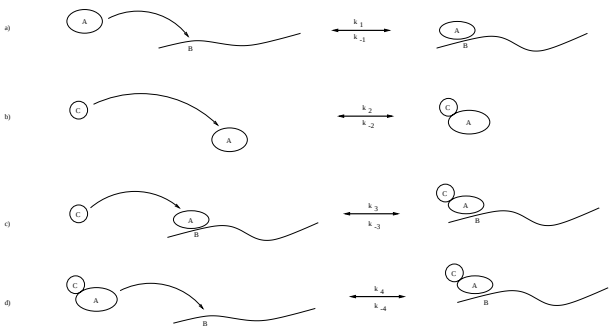

FIG. 1: These four steps form the fundamental processes in the formation of transcription factor complexes. The first step is the binding of the transcription factor, A, to the DNA binding site B. The second step is the binding of the transcription factor binding protein C to the transcription factor itself. The third and fourth processes depict the formation of the binding complex given either protein-protein or protein-DNA precursors. We will be particularly interested in the case in which the first two processes operate mainly in the reverse direction, while the last two processes are stable and operate slowly in the reverse direction.

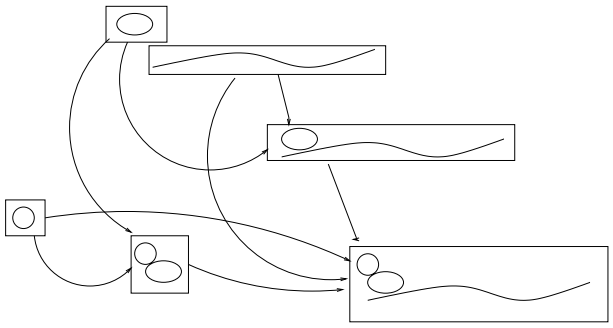

FIG. 2: The dynamics on the network depend crucially on the topology formed by the elements of the reactions. The nodes (boxes) represent reactant species while the edges represent paths that nodes can take to form other nodes. Of crucial importance are symmetry breaking elements in the network, which allow for rich phase diagrams including oscillations and saddle points.

process are obtained from simple considerations of chemical reactions. The model's underlying chemical reactions are depicted in Figure 2 and its topological structure is depicted in Figure 3.

$$\frac{d[A]}{dt} = k_{-1}[AB] + k_{-2}[AC] - (k_1 + k_2)[A] \quad (1)$$

$$\frac{d[B]}{dt} = k_{-1}[AB] + k_{-4}[ABC] - (k_1 + k_4)[B] \quad (2)$$

$$\frac{d[C]}{dt} = k_{-3}[ABC] + k_{-2}[AC] - (k_2 + k_3)[C] \quad (3)$$

$$\frac{d[AB]}{dt} = k_1[A][B] - k_{-1}[AB] \quad (4)$$

$$\frac{d[AC]}{dt} = k_2[A][C] - k_{-2}[AC] \quad (5)$$

$$\frac{d[ABC]}{dt} = k_3[C][AB] + k_4[B][AC] - (k_{-3} + k_{-4})[ABC] \quad (6)$$

There are several important points to note about the model. First it is completely deterministic and its dynamical variables are allowed to vary continuously. In the low copy number regime of the transcription factor binding protein (species [C]), both of these assumptions are in actuality violated - the low copy number endows the system with stochasticity in actuality and requires integer-valued dynamical variables.

Second, we chose to handle the terms reflecting the transcription factor binding sites in a manner similar to the diffusing proteins. This assumption is based on the observation that the transcription factor binding sites have many of the same characteristics as the proteins in the model. There are typically several binding sites on the DNA, and therefore one can define a 'copy-number' in the same manner as one defines a 'copy-number' for the relevant proteins. The significant difference between the protein-DNA and protein-protein interaction is the method by which each interaction is found in the space of possible interactions. As highlighted in [6], the transcription factors typically search out binding sites by mixing 1D diffusion along the DNA and 3D diffusion in the space surrounding the DNA, whereas it is not understood how transcription factor binding proteins specifically search for their targets. In our model we are making gross enough approximations by allowing our dynamical variables to be continuously valued so that such fine details on search methods can be ignored.

FIXED POINT ANALYSIS

Borrowing from the techniques of dynamical systems theory, we proceed as follows. First we set the left hand side of the coupled set of equations to 0 to obtain the fixed points of the dynamics. Then we construct the Jacobian matrix from the differential equations, evaluate the Jacobian matrix at each fixed point, and diagonalise the resulting matrix to complete the linear stability analysis in a neighbourhood around the fixed point. An analysis of the eigenvalues shows that all three of the fixed points behave as unstable saddle points - the results of the fixed point analysis are noted in Table 1 in a truncated fashion.

The presence of three unstable saddle points indicates the possibility of saddle-switching phenomenon within the dynamics of the system. The presence of saddle-switching would indicate that the phase portraits capturing the dynamics are not structurally stable and that any small perturbation in any of the parameters could affect the topology of the phase portrait. This idea, however,

TABLE I: From the linear stability analysis, we construct a table that characterises the linearised dynamics in the neighbourhood surround the fixed points. We see that the nontrivial fixed points both exhibit bifurcation behaviours as they both can display qualitatively different local dynamics.

Fixed Point	Stability	
[0,0,0,0,0,0]	weak saddle point	
'+' fixed point	saddle point	stable spiral
'-' fixed point	saddle point	stable spiral

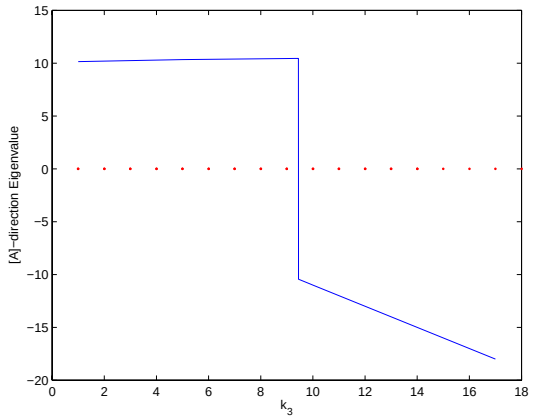

FIG. 3: The bifurcation depicted in this figure is highly representative of the types of bifurcations taking place in the system. As the k_3 parameter is tuned through a critical value of approximately 9.447, the eigenvalue describing the flows along the [A] direction reverses in sign (at the same time the [C] direction also reverses sign, hence the term bifurcation).

remains outside the scope of this study and warrants further investigation.

NUMERICAL ANALYSIS

The numerical portion of the study of this simple model of transcription factor complex formation involves two distinct components. The first is numerical studies of the algebraic concepts underlying the existence of fixed points; namely we will track the fixed point as it gains stability in one eigendirection at the expense of losing stability in another as we scan across different regimes of rate constants. The second component of the numerical analysis involves studying time trajectories obtained from numerical integration of the coupled differential equations and validating our hypothesis relating hierarchical time scales of protein binding and the ability to 'get away with' low copy numbers of relevant proteins.

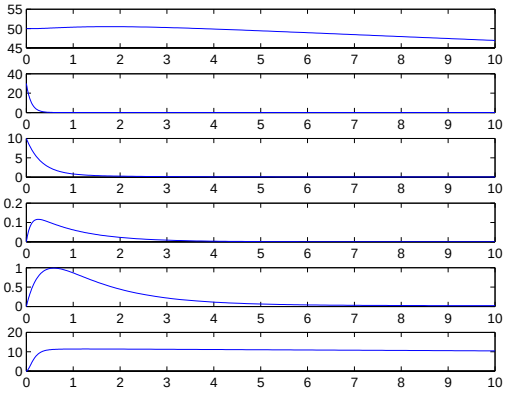

FIG. 4: We present time course trajectories for the 6 reactive species in the network. The top trace is the time course data for A, the second trace is for B, the third trace for C, the fourth trace for A bound to B, the fifth trace for A bound to C, and the last trace depicts the lifetime of the transcription factor complex as a whole. See text for details on the importance of the apparent time lags built into the network.

'Bifurcations'

Typically studies of bifurcations of fixed points involves the switching of fixed points from one type to another as different parameters are varied. In the case of our transcription factor binding model, however, the bifurcation behavior is mainly confined to the pairwise switching of stable and unstable manifolds making up a saddle point. The special case [$k_1 = k_2 = k_{-3} = k_{-4} = 1$, $k_{-1} = k_{-2} = k_3 = k_4 = 10$] does have two complex eigenvalues with negative real parts. The resultant spiral oscillations that are stable in the plane of oscillation and three of the other four dimensions, but have a positive eigenvalue in the [ABC] dimension and thus forms a helical spiral that has no analogue in the two dimensional flows that forms the basis of our language of fixed point behaviours.

A sample of the bifurcation behaviour is depicted in Figure 3.

Time Trajectories

Since the differential equations presented in the model are impossible to integrate by hand, we turn instead to the MATLAB routine ode45. In particular we feed ode45 parameters in our region of interest (in which the residence time of [A] alone on [B] much shorter than the residence time of [AC] on [B], as encoded in the kinetic rate parameters) as well as initial species numbers in our region of interest (in particular setting [B] to be a low value to determine whether or not transcription factor complexes can achieve long residence times on DNA binding

sites despite low copy numbers of a transcription factor binding protein) and then simultaneously integrate the equations over time using a (4,5) Runge-Kutta formula. An example of a time trace with parameters chosen to fit the criteria mentioned above is included in Figure 4.

In examing this trace we see exactly the hierarchy of time scales that we were expecting. The initially high concentration of [A] (seen in the first row) quickly quenches the 'concentration' of [B] and [C] (second and third row respectively). The [AB] and [AC] species take some time to reach a maximum value and then they themselves begin to decay in an intermediate time scale in favour of the slowly rising [ABC] species, which does not appreciably decay over the time interval plotted once it has reached its maximum value (which is determined by the initial number of B and the fast decay of [AB] to [A]; this also explains the relatively slow decay of [A] over the time scale plotted above).

CONCLUSIONS

The numerical simulations of the time trajectory unambiguously show that a hierarchy of kinetics can lead to stable transcription levels even when the transcription factor binding protein that stabilises the whole ensemble is present in low copy number. The dynamics, however, are determined by the region in parameter space that the system is operating in - and since the system's fixed points exhibit bifurcations, the dynamics can be very sensitive to small perturbations in the parameter (i.e., if the system is operating near a critical parameter value). That said, the bifurcations are dominated by saddle point to saddle point transitions, therefore the different time plots 'switch places' between different species and do not actually display much qualitatively different phenomena. There is at least one neighbourhood in parameter space in which stable spirals develop in phase space - and it happens to coincide with biologically relevant parameter values for the kinetic rate constants built into the model.

For further study, however, the considerations of integer-valued copy number and the stochasticity demanded therein must be taken seriously. A straightforward extension to the project could involve use of the Gillespie algorithm to build a detailed timestep-by-timestep model of the stochastic dynamics in order to get a more realistic picture of what would happen in a real biological cell.

[1] Hartwell, L.H., et al., "From Molecular to Modular Biology,"Nature(402), pp. C47-C52, [1999]

[2] Mirollo, R.E. and Strogatz, S.H., "Synchronization of pulse-coupled biological oscillators," SIAM J. Appl. Math.:(50), pg. 1645, [1990]

[3] Elowitz, M. and Leibler, S., "A Synthetic Oscillatory Network of Transcriptional Regulators," Nature(403): pp. 335-338, [2000]

[4] Remenyi, A. et al. "Combinatorial Control of Gene Expression," Nature Structural and Molecular Biology 11(9): pp. 812-815, [2004]

[5] Slutsky, M. E-mail communication

[6] Spirin, V. and Mirny, L.A., "Protein Complexes and Functional Modules in Molecular Networks," PNAS 100(21): pp. 12123-12128, [2003]

The author would like to thank Michael Slutsky for helping him making the problem concrete.