Fractioned site occupancy dependence on the number of binding sites and interaction between transcription factors

Tek Hyung Lee
MIT Department of Chemical Engineering
(Dated: May 17, 2007)

I designed a simple model of interaction between transcription factors which were required for recruitment of RNA polymerase. Applying the idea of Icing Model, total state energy was calculated with interaction energy term. Fractioned occupancy was raised remarkably as the interaction energy and the number of transcription factor binding sites increase.

I. INTRODUCTION

In biological system, gene expression is controlled by several key steps such as transcription factor binding to operator site, RNA polymerase (RNAP) recruitment to the promoter region, transcription and translation [1]. Among them, RNAP recruitment is enhanced or repressed by certain transcription factors (TFs). In order to analyze these activation and repression processes quantitatively, many thermodynamic models based on statistical mechanics were studied. From previous researches, each state of TF binding was interpreted with a probability including Boltzmann weight and the probability of RNAP being bound to the promoter of interest with the help of TF was calculated [2]. RNAP binding models for one transcription factor system such as *lac* promoter of *E.coli*, where one transcription factor (cAMP receptor protein) affects to RNAP recruitment, and dual activator system model such as phage lambda promoter, were also well studied [3]. However, quantitative study for more transcription factors and their interaction is not sufficient and valid model is also required to understand complex gene expression process such as eukaryote system.

In this paper, I designed a simple model which explains fractioned site occupancy dependence on TF interaction and the number of TF binding sites.

II. THEORY AND MODELING

Fractioned site occupancy of TF binding site on DNA affects to the probability that RNAP binds to promoter. As the occupancy increases, the probability for interaction between RNAP and TFs is enhanced and transcription initiation by RNAP also increases. Fractioned site occupancy is influenced by several factors such as TF concentration within cell, the number of TF binding sites and the interaction between TFs. In this analysis, I made some assumptions to design effective model equation. Firstly, I assumed that TFs only bind specific binding sites (i.e. there are no nonspecific binding sites on DNA.) and assumed that each TF can interact with only TF located on the right next site. Secondly, I assumed that all TFs are identical thus the binding energy which TF bind to DNA binding site is also same. With these assumptions, TF interaction model could be simplified easily. Fractioned site occupancy F of the binding site is defined as:

$$F = \frac{TFS}{TFS + S} \tag{1}$$

TFS denotes the number of TF bounded sites where transcription factor binds and S is the

number of free sites. If we think that there are N binding sites, 2^N possible states exist. Since binding state of each site is expressed 0 or 1, where 1 corresponds to binding and 0 corresponds to non-binding, we can generate state vector including total sites. Moreover, we can construct possible state matrix X :

$$X = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 & 0 \\ 1 & 1 & \cdots & 1 & 1 \end{bmatrix} \quad (2)$$

Each row corresponds to possible each state and each column denotes the occupancy of binding site. For example, a row $\begin{bmatrix} 0 & 1 & 0 & 0 & 1 \end{bmatrix}$ means the state where there are 5 binding sites and second and fifth site are occupied by TF. Therefore, matrix X includes all possible binding states. Next, the state probability can be calculated for state where i sites are occupied with Boltzmann weight factor:

$$p(X_i) = \frac{e^{-E(X_i)/kT}}{\sum_{X_i} e^{-E(X_i)/kT}} \quad (3)$$

$E(X_i)$ corresponds to the state energy where i sites of N sites are occupied by TFs and denominator corresponds to the partition function. This energy contains TF binding energy to DNA binding site, energy by TFs translational entropy effect and interaction energy between TFs. It can be expressed as Icing-like model [4]. On each site in certain state, TF bounded site contribute to total state energy by $-\Delta G$ and free site contributes by $kT \ln TF$ (i.e. entropic effect). Interaction energy E_{INT} between TFs could be expressed only if both one site and next site are bounded by TFs (pairing) by following expression:

$$E_{INT} = x_i x_{i+1}(-\varepsilon) \quad (4)$$

x_i denotes state of i th site. For example, if both one and the next site are occupied $x_i = x_{i+1} = 1$ and interaction energy ε contributes to total state energy. Finally, state energy $E(X_i)$ is expressed:

$$E(X_i) = -i\Delta G + (N-i)kT \ln TF + \sum_{j=1}^{N-1} x_j x_{j+1}(-\varepsilon) \quad (5)$$

Fractioned site occupancy F can be expressed with the state probability:

$$F = \frac{1}{N}\sum_{X_i} p(X_i)\cdot(\sum_{j=1}^{N} x_j) = \frac{\sum_{X_i} i e^{-E(X_i)/kT}}{N\sum_{X_i} e^{-E(X_i)/kT}} \quad (6)$$

III. MODEL EVALUATION

With this model, fractioned site occupancy was calculated with the various N , epsilon ε and TF number. I calculated the energy variation of this model by MATLAB 7.1 and I obtained typical TF binding and its interaction energy values ($\Delta G \simeq -20 kcal/mole$ and $\varepsilon \simeq 3 kcal/mole$) from the previous research [5]. Figure 1 shows the effect of different interaction energies between TFs. As the interaction energy ε increases, fractioned site occupancy approaches to 1 more rapidly. Since interaction energy increase means the enhancement of cooperativity, it could be confirmed that fractioned site occupancy depends highly on cooperativity between TFs.

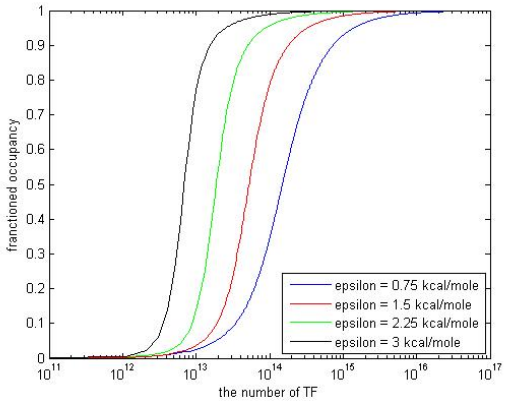

FIG.1. fractioned site occupancy dependence on the interaction energy between TFs in case of 5 binding sites ($N = 5$) where interaction energy ε between TFs is independent with site

Figure 2 shows the dependence of fractioned site occupancy on the number of binding sites. As the number of binding sites increase, occupancy increases. One possible reason for this result is that the increase of binding sites augments the probability of both TFs approach to binding sites and interaction between TFs. Therefore, the probability of complex formation for the RNAP recruitment increases.

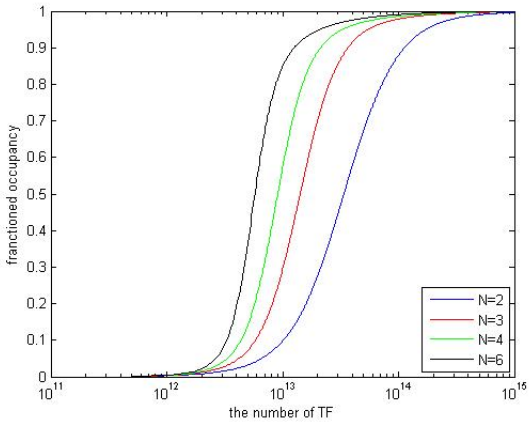

FIG.2. fractioned site occupancy dependence on the number of binding sites ($\varepsilon = 3\ kcal/mole$)

Finally, I changed the each interaction energy between TFs to investigate the effect of various interaction energies. For example, as the distance between TF and RNAP binding site increase, the interaction energy between TFs was changed from $\varepsilon_1 = 3kcal/mole$ to $\varepsilon_3 = 5kcal/mole$ or from $\varepsilon_1 = 3kcal/mole$ to $\varepsilon_3 = 1kcal/mole$. The effect by this variation can be evaluated by Figure 3 and 4. In exist of increasing interaction energy, fraction site occupancy increased more rapidly.

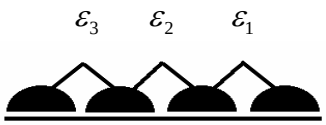

Fig. 3. TF binding with different interaction energy ε where hemispheres denote TF and underline means DNA

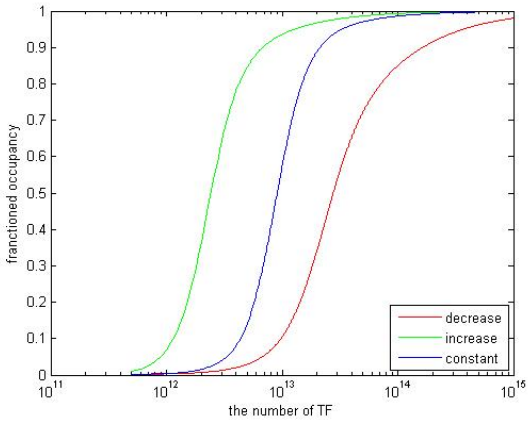

FIG.4. fractioned site occupancy dependence on various interaction energy where the number of sites is 4 ($N = 4$) and where 'increase' or 'decrease' means that interaction energy ε between TFs increases from 3 to $5kcal/mole$ or decreases from 3 to $1kcal/mole$ as the TF binding site goes farther from RNAP binding site

IV. CONCLUSION

Fractioned site occupancy is an important parameter to understand RNAP recruitment. In this project, TF interaction raised site occupancy probability. In reality there are many kinds of TF which contribute to transcription initiation with various interactions. Therefore, assumptions in this paper should be more developed. Nevertheless, this Icing-like model will be useful to study cooperativity between TFs. With valid and reasonable assumption in diverse biological environment, this model can be applied to understand gene expression mechanism.

[1] M. Ptashne and A. Gann, Genes and Signals (Cold Spring Harbor, New York, 2004)
[2] L. Bintu, N. Buchler, H. Garcia, U. Gerland, T. Hwa, J. Kondev and R.phillips, Curr Opin Gener Dev **15**, 116 (2005)
[3] L. Bintu, N. Buchler, H. Garcia, U. Gerland, T. Hwa, J. Kondev, T. Kuhlman and R. Phillips, Curr Opin Gener Dev **15**, 125 (2005)
[4] www.plmsc.psu.edu/~www/matsc597c-1997 /phases/Lecture3/node1.html
[5] K. Koblan and G. Ackers, Biochemistry **31**, 57 (1992)