Oct 1st, 12:00 AM

Modeling Transcriptional Output Controlled by a Potential Enhancer for the Twin of Eyeless Gene in D. melanogaster

Cristina Santo
Clark University, csanto@clarku.edu

Interactions Between Transcription Factors of *D. melanogaster From Zone 1 of DNA*

Cristina Santo

Introduction

DNA transcription is a vital process that regulates gene expression in all living organisms, these genes can be turned on and off to control life processes [3]. Cis-regulatory modules (CRMs), also known as enhancers, which are defined as a non-coding region of DNA, and these enhancers usually have clusters of binding sites for transcription factors (TFs) [2]. When a TF binds to an enhancer, it can either activate or repress gene transcription of that specific gene, or display cooperativity (two activators) or quenching (one activator and one repressor) when TFs are adjacently bound [4].

But in the actual realm of DNA transcription, the question remains if a particular configuration with bound TFs will activate or repress transcription [4]. During the summer, thermodynamic models were created to measure gene expression of *toy*, the gene responsible for correct eye development in *D. melanogaster* in Zone 1 of DNA.

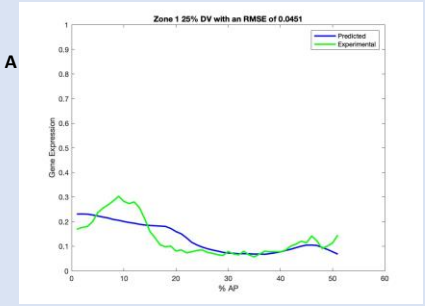
Figure 1. Transcription factors binding to an enhancer (CRM) [1]

Methods

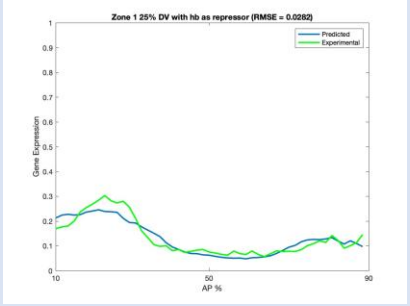
1. The DNA sequence for Zone 1 was obtained from Nourie's thesis by inserting the coordinates from the thesis into the Genome Browser [5]. From Conrad's graphs, the goal was to achieve a high peak toward the anterior axis and a lower peak toward the posterior axis with repression in between.
2. A bioinformatic algorithm was used to find the predicted binding sites for 5 specific transcription factors: bicoid (bcd), hunchback (hb), caudal (cad), knirps (kni), and kruppel (kr), with consideration to the forward and reverse strands.
3. One model was made where *hunchback* was just an activator and the other where *hunchback* was just a repressor. All the possible states for each adjacently bound TF were noted and a function could be made using MatLab. The equation used for the function was all successful states divided by all possible states [A]. Some of these states included quenching and cooperativity.
4. Another function was created with a threshold of i<15, meaning all data points before 15 considered hb an activator, but when the data points exceeded 15, hb was considered a repressor.
5. With each of the graphs that resulted, parameters such as binding affinity, cooperativity, and quenching were modified to best match the graph from Regan's thesis.
6. For each of the graphs, the predicted data was compared with experimental for 25% DV (protein concentrations) and 50% DV (mRNA concentrations) by using fminsearchbnd which comprised of a root mean square error equation which determined the error between the predicted and experimental data and provided the closest parameter values using the predicted values.

Results

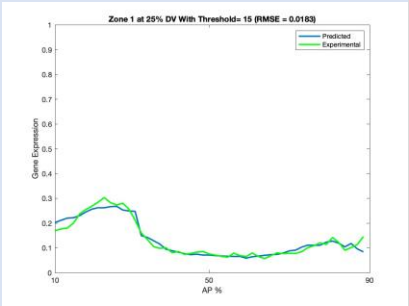
A



B



C



D

25% *hb* repressor	
Variables	Parameter Values
K1 = R1= kni	27.8369
K2 = A1= hb	0
K3 = A2 = bcd	12.4434
K4 = A3 = cad	0
K5 = R2 = kr	7.7185
C3= bcd * cad	0.0739
Q2 = kni * bcd	0.0614
Q3 = kni * cad	0.8358
Q5 = bcd * kr	0.7297
Q6 = cad * kr	0.1474
Q7 = hb * cad	0.1438
Q8 = hb * cad	0.0301
RMSE	0.0282

E

25% *hb* thresh	i<thresh
Variables	Parameter Values
K1 = R1 = kni	63.0235
K2 = A1 = hb	81.4302
K3 = A2 = bcd	7.6609
K4 = A3 = cad	0.0709
K5 = R2 = kr	90.9455
K6 = R3 = hb	30.4997
C1 = hb * bcd	1.2086
C2 = hb * cad	3.1396
C3 = bcd * cad	0.0056
Q1 = kni * hb	0.1469
Q2 = kni * bcd	0.0026
Q3 = kni * cad	0.9604
Q4 = hb * kr	0.9961
Q5 = bcd * kr	0.1847
Q6 = cad * kr	0.1634
Q7 = hb * cad	0.0095
Q8 = hb * cad	0.4989
RMSE	0.0183

Figure 3. (A) Graph with *hb* as an activator throughout. **(B)** Graph with *hb* as a repressor throughout. **(C)** Graph with a threshold of 15 where i<thresh has *hb* as an activator and i>thresh has *hb* as a repressor. **(D)** Table with parameter values and successful states at 25% DV where *hb* is a repressor. **(E)** Table with parameter values and successful states at 25% DV when *hb* has a threshold of 15.

Discussion

1. The RMSE value when setting a threshold at 25% DV gave the lowest value, suggesting that *hunchback* could act as both an activator and a repressor [C]. If *hunchback* can act as both an activator and a repressor, additional components may be involved, such as the location of the threshold on the anterior-posterior axis. The case may be that at particular points of the embryo, *hb* functions best as an activator or a repressor or utilizes aid from other transcription factors. For example, another transcription factor may need to be accounted for with the mRNA concentrations, leading to better values.
2. The RMSE indicates when *hb* interacts with *bcd* or *cad*, the interaction of two activators and the quenching in this sense allows *toy* in Zone 1 to have a high expression when *hb* is an activator, and lower expression when *hb* is a repressor [E].
3. Furthermore, for value Q8, where *hb* and *cad* interact, the closer the Q value is to 0, the stronger the repressor is [D]. This indicates that when *hunchback* and *caudal* interact, *hb* represses *cad* so much that there is no expression [D]. When *hb* is a repressor, even though the quenching between *knirps* and *bicoid* indicates more activation, the quenching between *hb* and *cad* seems to cancel out all possible gene expressions of *cad* [D].
4. Since only five transcription factors were used with the changing variable of *hb*, there that other corepressors and coactivators are aiding in the behaviors of *hb*. With coactivators, the transcription rate would increase by binding to an activator and vice versa for corepressors. In addition, there is the chance that there are more transcription factors present that have not been included, and there are most likely more binding sites. Depending on the order these binding sites may be in, this could significantly alter the figures made and any conclusions that have been previously drawn. Also, there may be more than one binding site in varying locations for the five TFs that were already involved that the bioinformatic algorithm did not recognize.

Future Direction

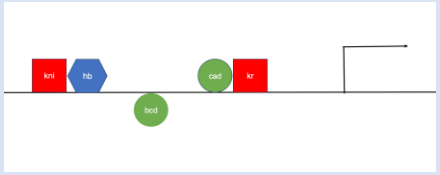
1. Other transcription factors that were not used will be considered.
2. Determining whether there are multiple binding sites for a single transcription factor.
3. Only using protein concentrations instead of mRNA concentrations.

Figure 2. (A) Equation used to express gene expression. **(B)** Model with transcription factors in order of position

A

$$[mRNA] \propto \frac{K_A[A]+QK_AK_R[A][R]}{1+K_A[A]+K_R[R]+K_AK_R[A][R]}$$

B



I'd like to thank Professor Dresch, Lindsay Carlson, and Eleanor Verhagen for their help and guidance throughout the course of this summers research. I'd also like to thank the Biology department for giving me the privilege of participating in research.

1. Adcock IM, Caramori G (2009) Asthma and COPD (Second Edition)
2. Conrad R (2020) Development of Image Processing Pipeline for Analysis of *twin of eyeless* Enhancers in *D. melanogaster* Embryos
3. Dresch JM and Drewell RA (2012) Decoding the *cis*-regulatory grammar behind enhancer architecture. *Genomics III: Methods, Techniques and Applications* (iConcepts Press, book chapter) ISBN: 978-1-922227-096.
4. Drewell RA, Nevarez MJ, Kurata JS, Winkler LN, Li L, Dresch JM (2014) Deciphering the combinatorial architecture of a *Drosophila* homeotic gene enhancer. *Mechanisms of Development.* 131:68-77.
5. Nourie LL (2019) Analyzing the Regulation of *twin of eyeless* Expression During Early *Drosophila melanogaster* Development