Effective Transcription with Low-Affinity Binding Sites

Bill Ke, Eric Wu

MIT PRIMES

Mentored by David Darrow (MIT) and Joseph Landsittel (Northwestern)

Fall-Term PRIMES Conference
October 24-25, 2026

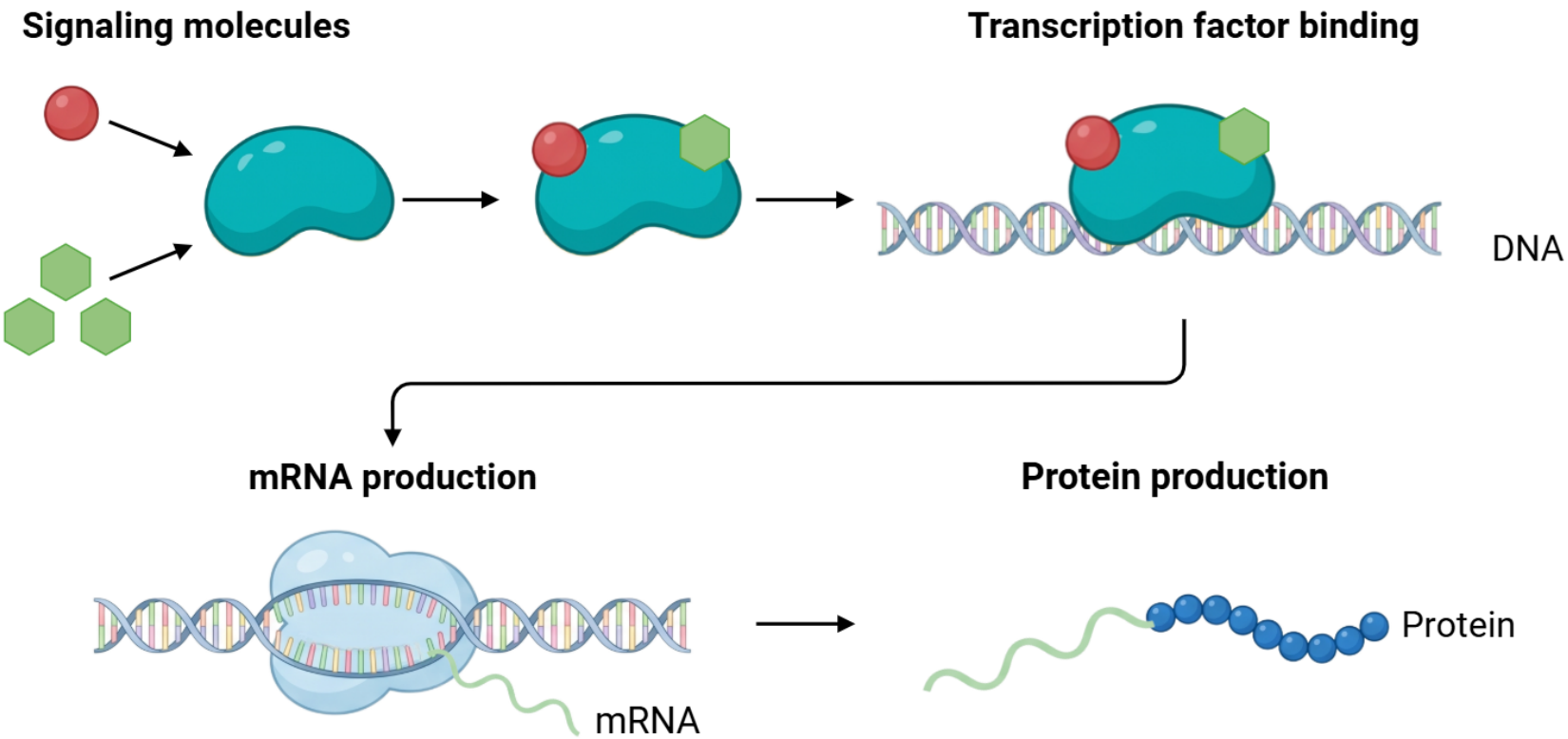
Arctic Foxes





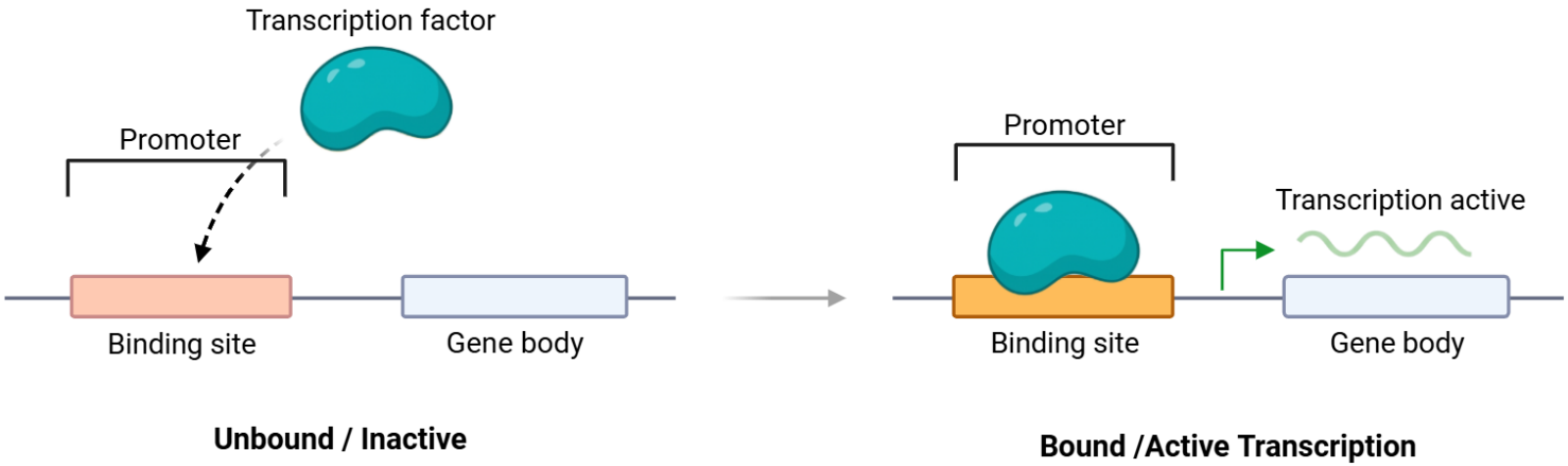
Polar Guidebook (2023), "Why Do Arctic Foxes Change Color?"

Cellular Regulation



Created with BioRender

Transcription Factor Binding



Created with BioRender

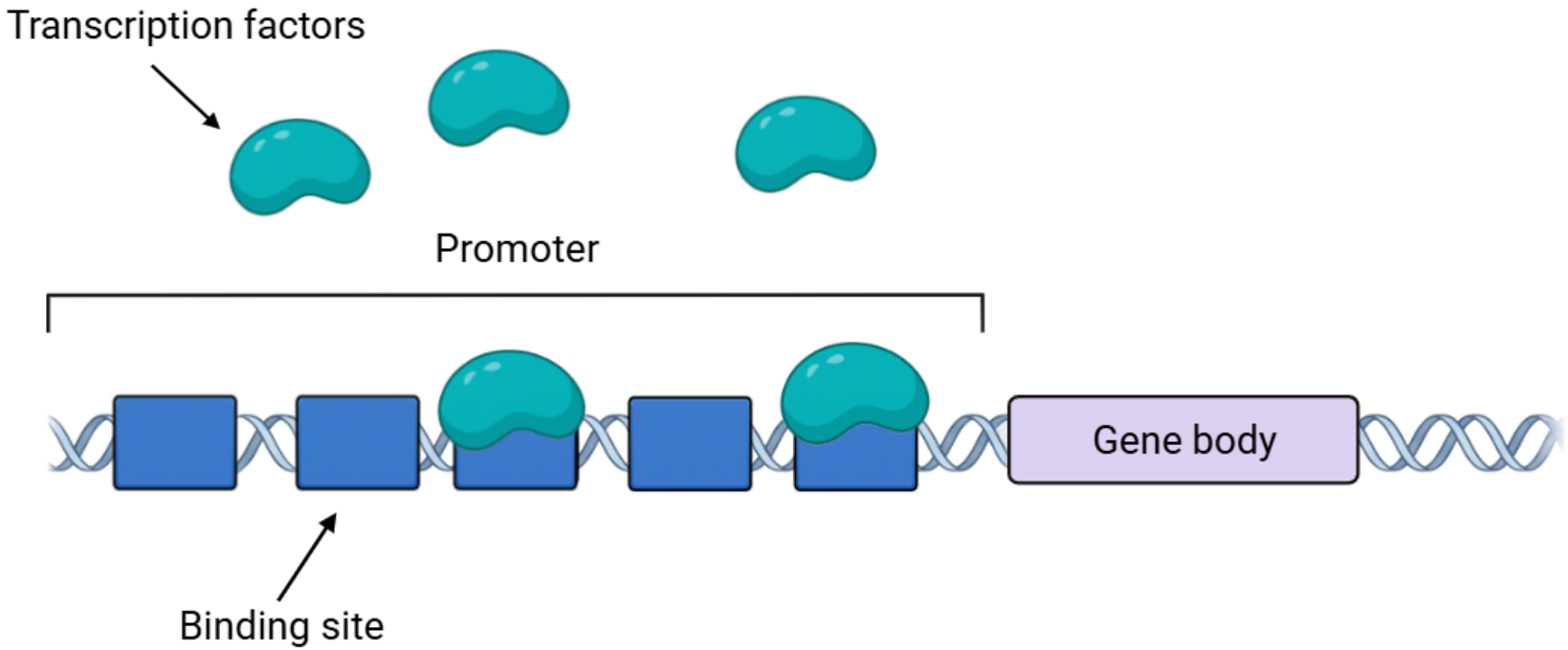
Prokaryotes vs Eukaryotes

Characteristic	Bacteria	Multicellular Eukaryotes
Genome length	10^6–10^7 bp	10^8–10^{10} bp
Regulatory architecture	Relatively compact	Clusters of sites are often dispersed
Number of each TF	1–10	10^3–10^5

Source: Wunderlich and Mirny (2009).

Prokaryotes are better understood than eukaryotes

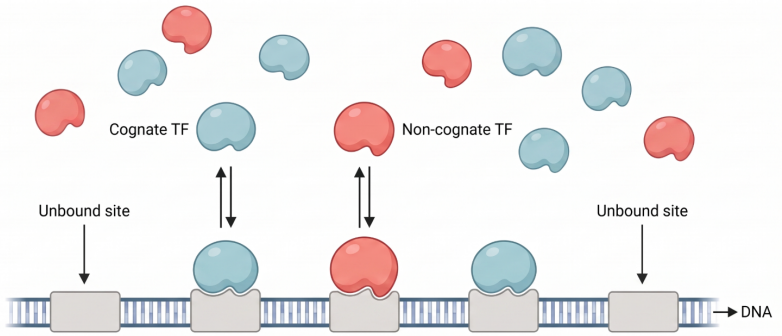
Low-affinity binding sites



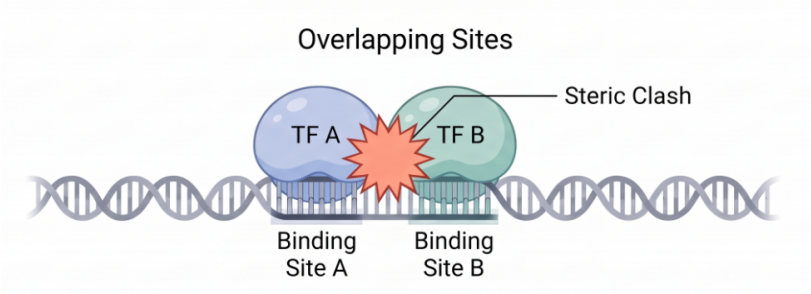
Created with BioRender

Recent evidence shows the relevance of low-affinity binding sites

Issues with low-affinity binding sites



Specificity



Spacing

Created with BioRender

Given this, why do they still evolve?

Overview

In the rest of our presentation, we cover

- information-theoretic rationale for weak binding site clusters,
- models of gene expression,
- numerical results.

Information Transfer

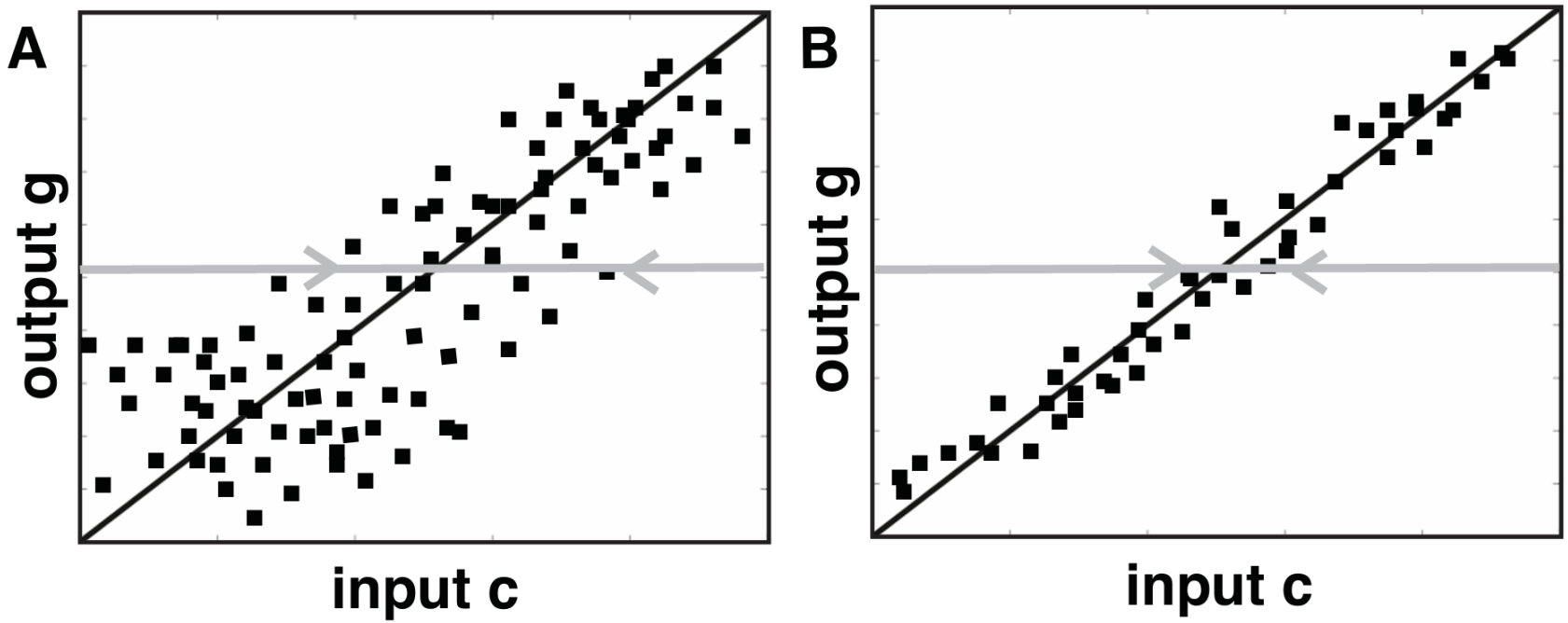


Figure from Tkacik and Walczak (2018)

We can quantify how much information the environment transfers to the gene through information theory.

Mutual Information and Entropy

Definition. *Shannon entropy* is a measure of the uncertainty in a continuous probability distribution $P(c)$, defined as

$$S = -\int dc\, P(c) \log_2 P(c).$$

Definition. *Mutual information* is defined as a difference in entropies

$$I(c; g) = \int dc\, P(c)(S[P(g)] - S[P(g|c)]).$$

Example (Guessing the number game)

Both entropy and mutual information are measured in *bits*. A bit is equivalent to the information gained from a yes/no question.

A model of transcription

- The input c, is the concentration of the intended (cognate) transcription factor.
- The output, g, is normalized gene expression on a scale from 0 to 1.
- Even at the same input, molecular fluctuations can produce different outputs.

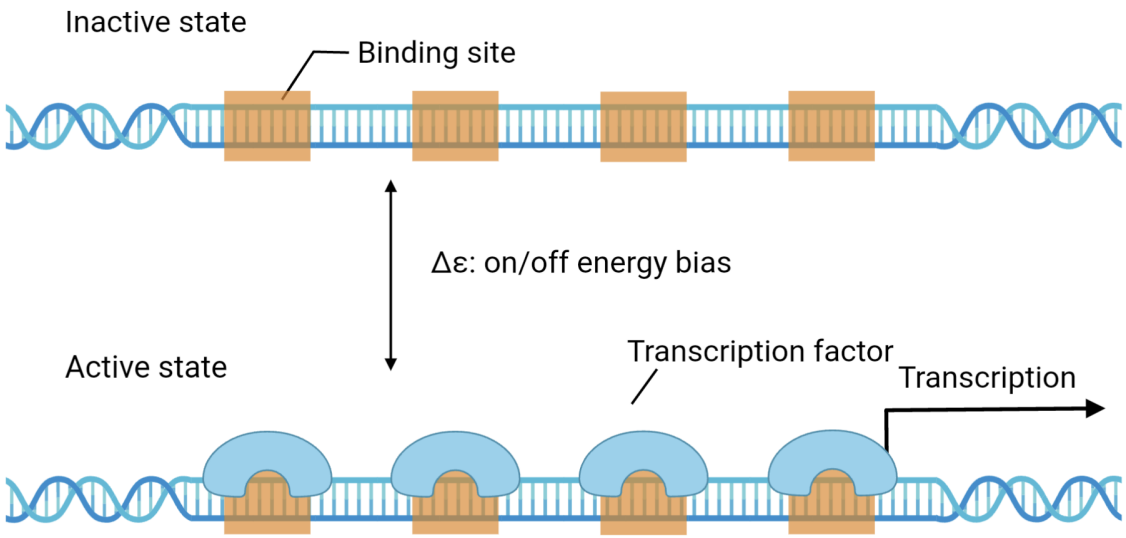
Question

Can a cluster of weak binding sites reliably respond to changes in TF concentration?

A model of transcription (cont.)

The promoter switches between active and inactive states. We have:

- n: number of binding sites,
- K_d: dissociation constant in the active state, larger K_d means weaker binding,
- $\Delta\varepsilon$: inactive-to-active energy bias



The Monod-Wyman-Changeux model

- The Monod-Wyman-Changeux (MWC) model describes the mean expression $\bar{g}$ at a fixed input.
- However, expression fluctuates due to factors such as protein synthesis and transcription factor diffusion.
- Assuming production proportional to promoter activity, normalized mean protein abundance equals the probability of being active.

Biological quantities

- c: signaling TF concentration
- $\bar{g}(c)$: mean normalized protein abundance

Mean protein response

$$\bar{g}(c) = \frac{1}{1 + e^{\Delta\varepsilon}(1 + c/K_d)^{-n}}. \quad (1)$$

Gene responses and noise determines how well we can distinguish inputs.

Information capacity

For a fixed promoter and input range, define its information capacity as:

$$I^* = \max_{P(c),\ 0 \leq c \leq c_{\max}} I(c; g),$$

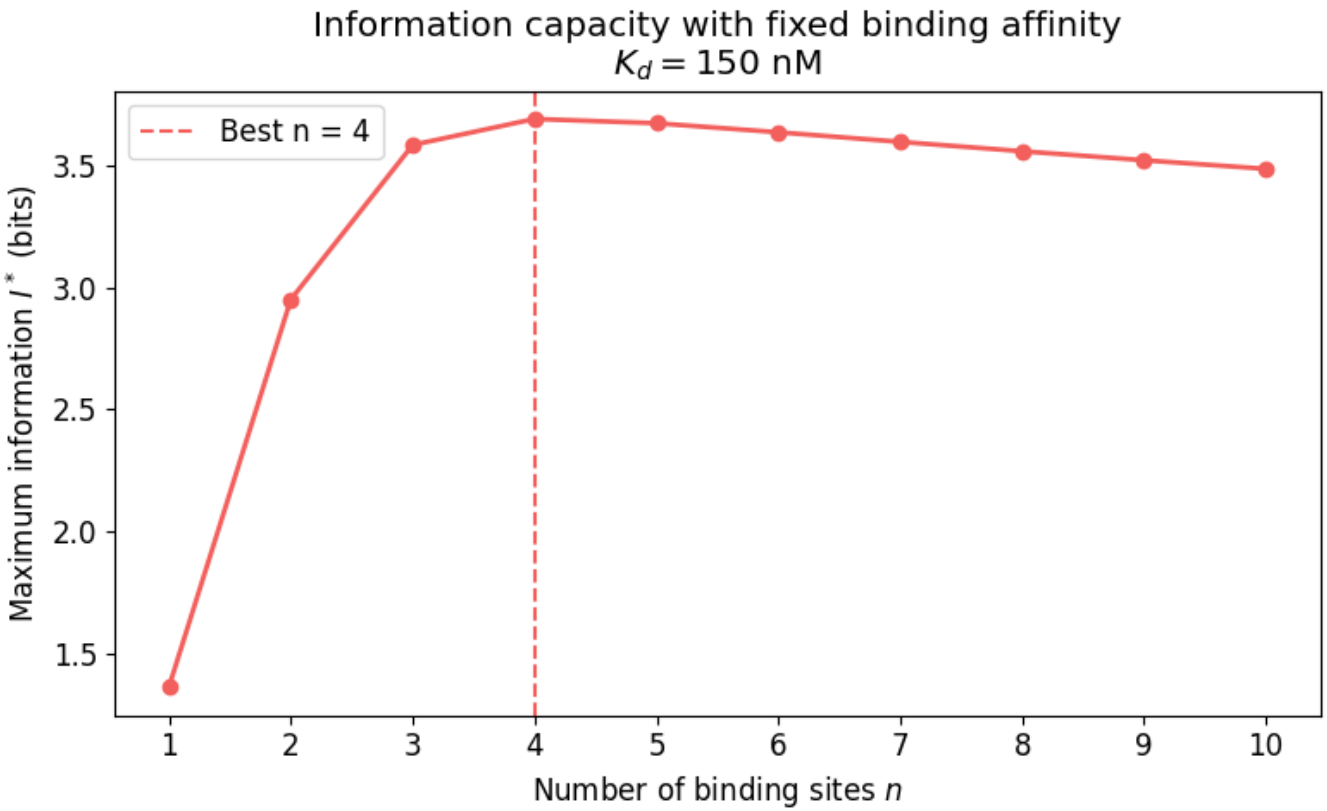
where

- $P(c)$ is the probability density of each input concentration occurring,
- $I(c; g)$ is the mutual information between input and output,
- $c_{\max}$ is the largest possible input concentration.

Evolution and optimality

Due to evolution, selection produces favorable responses to the environment. Capacity compares what different promoter architectures can theoretically transmit under optimal input distributions.

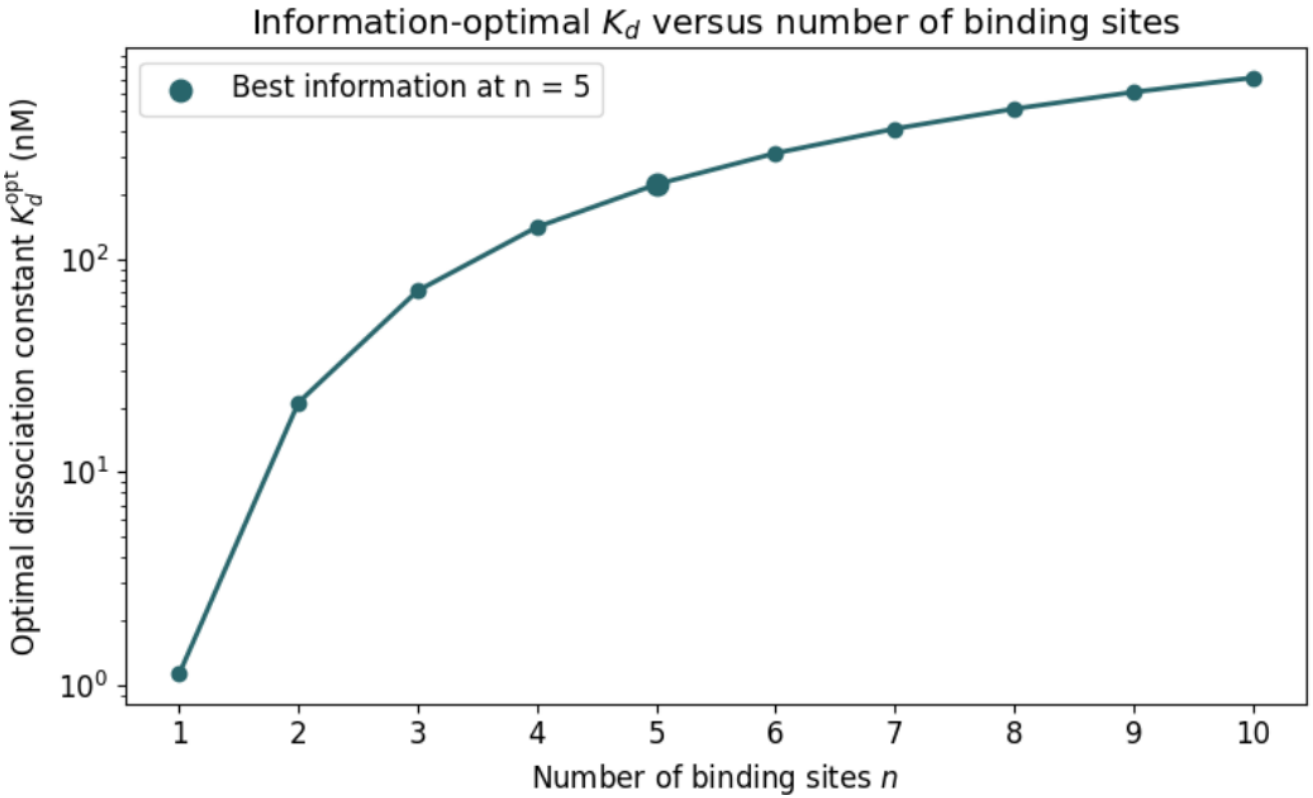
Several weak sites can improve transcription



- At $K_d = 150$ nM, we optimize input distribution for each n.
- When $n = 1$, $I^* \approx 1.4$ bits. When $n = 4$, $I^* \approx 3.7$ bits.

Increasing the number of binding sites doesn't necessarily increase information.

Several weak sites can improve transcription (cont.)

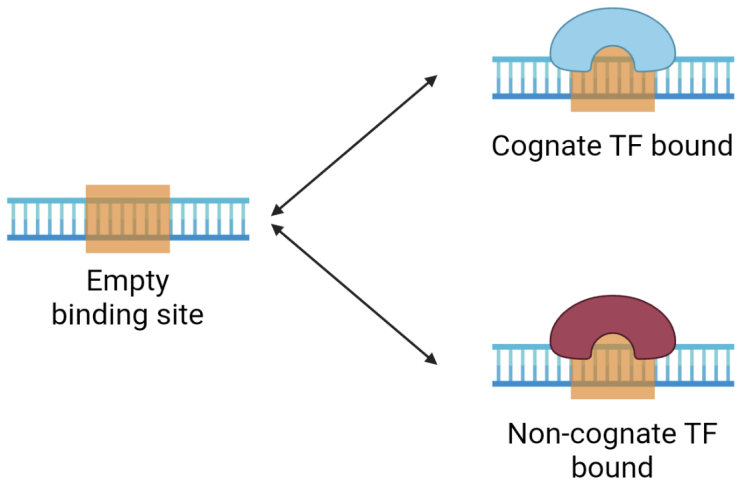


- At each site number, we find the affinity that results in maximal information capacity.
- As n increases, optimal K_d also increases.

Introducing competition

Definition

A cognate TF is the intended regulator of a promoter. A non-cognate TF is another, unintended TF that can bind to the same sites.

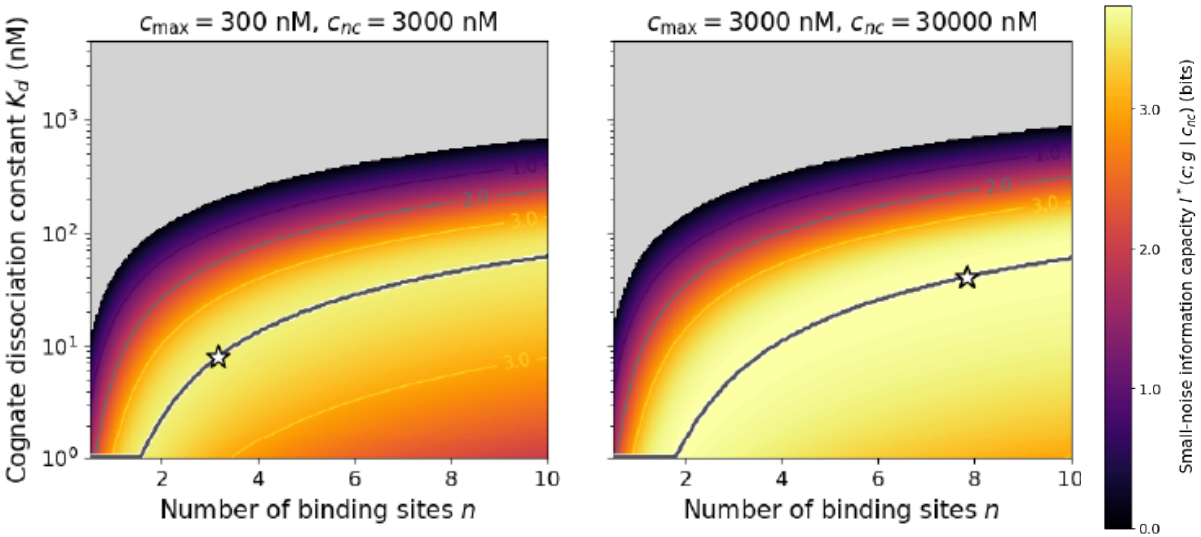


- Only one transcription factor can occupy a binding site at a time.
- Non-cognate transcription factors compete for the same binding sites as cognate transcription factors.

Competitor TFs adds noise, which makes expression fluctuate.

Competition's effect on the preferred architecture

Let $\bar{c}_{\mathrm{nc}}$ equal the mean competitor concentration. We set $\bar{c}_{\mathrm{nc}} = 10 c_{\mathrm{max}}$.



Graphical explanations

Color corresponds to capacity. The gray area is outside the valid range of our approximations. The stars are the information maxima.

In the higher concentration example (right), configurations favor more binding sites and weaker optimal binding.

Conclusion

Through our study, we find that

- many low-affinity binding sites can transmit information efficiently,
- in the presence of competition, at higher concentrations, architectures involving more binding sites are favored.

Any questions?

Acknowledgments

- We are grateful to our mentors David Darrow and Joseph Landsittel for their invaluable guidance and feedback throughout the research process.
- We thank the PRIMES program for organizing this research opportunity and conference.

References



Kieren. (2023, October 29). Why do Arctic foxes change color? [Summer vs winter coat]. Polar Guidebook. https://polarguidebook.com/do-arctic-foxes-change-color/



Tkačik, G., & Walczak, A. M. (2011). Information transmission in genetic regulatory networks: A review. Journal of Physics: Condensed Matter, 23(15), Article 153102. https://doi.org/10.1088/0953-8984/23/15/153102



Wunderlich, Z., & Mirny, L. A. (2009). Different gene regulation strategies revealed by analysis of binding motifs. Trends in Genetics, 25(10), 434–440. https://doi.org/10.1016/j.tig.2009.08.003

End of Presentation

THANK YOU!