Effective Transcription with Low-Affinity Binding Sites

Shengdao Ke, Zifang Wu

MIT PRIMES

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

Fall-Term PRIMES Conference
October 23-24, 2026

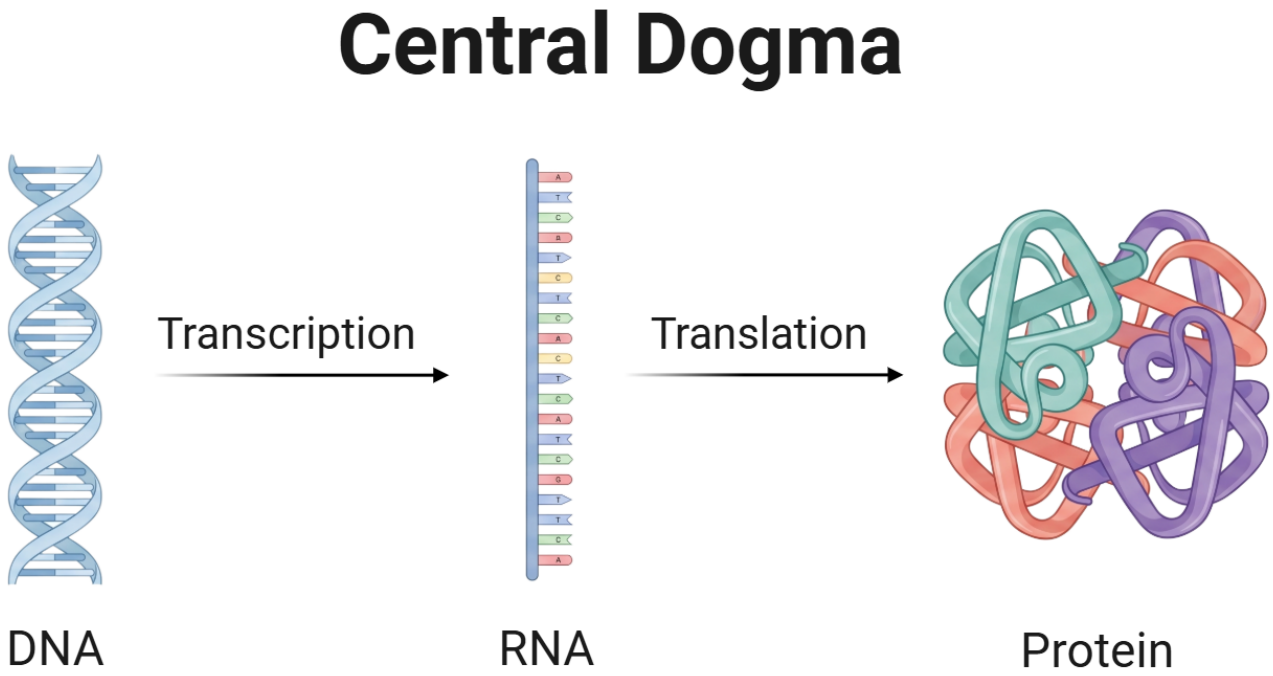
Overview

All cells respond to their environment through regulation. The proteins a cell produces depend on the signals they receive.

We present

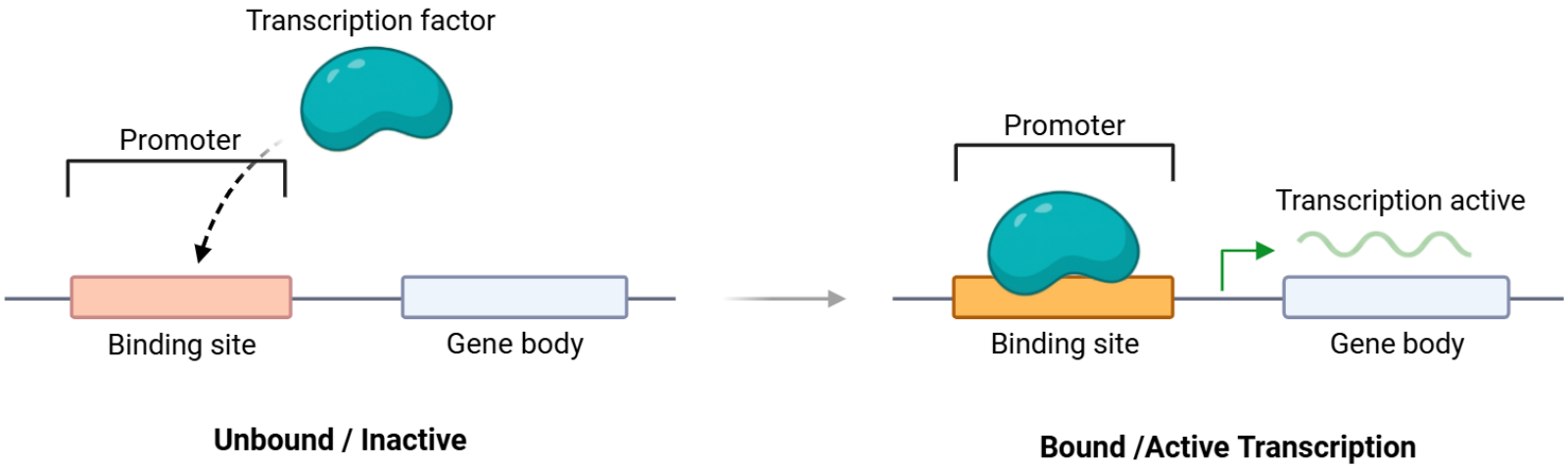
- the biological process behind regulation
- a theoretical framework of information transfer in cells
- results

The Central Dogma



Created with BioRender

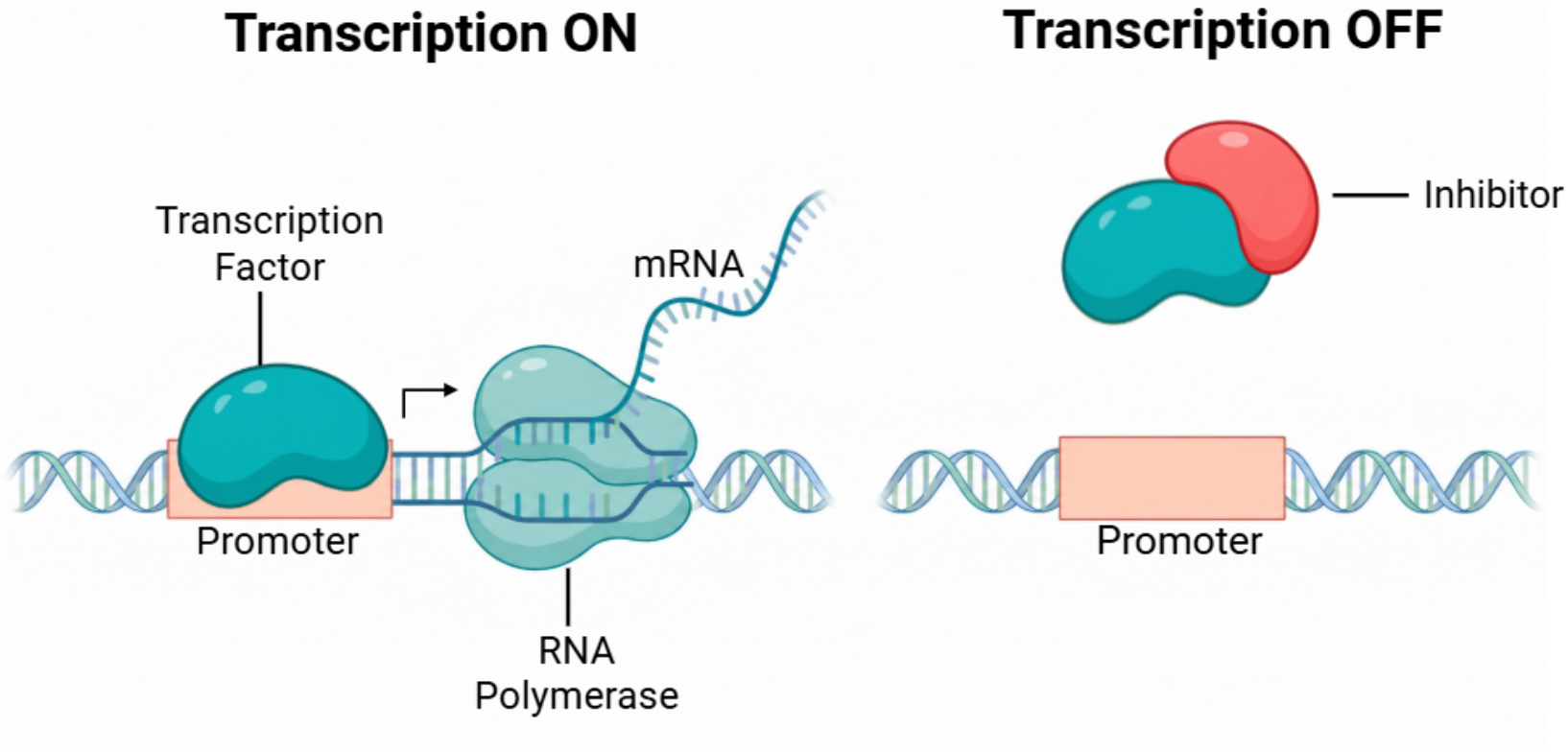
Transcription Factor Binding



Created with BioRender

- Transcription factors attach to specific binding sites in the promoter

Signaling molecules



Created with BioRender

- Signaling molecules influence how transcription factors bind

Full Picture of Transcriptional Regulation

Internal or External Signals$\longrightarrow$ Transcription Factors $\longrightarrow$ Promoter Binding$\longrightarrow$ mRNA Transcription$\longrightarrow$ Protein Production

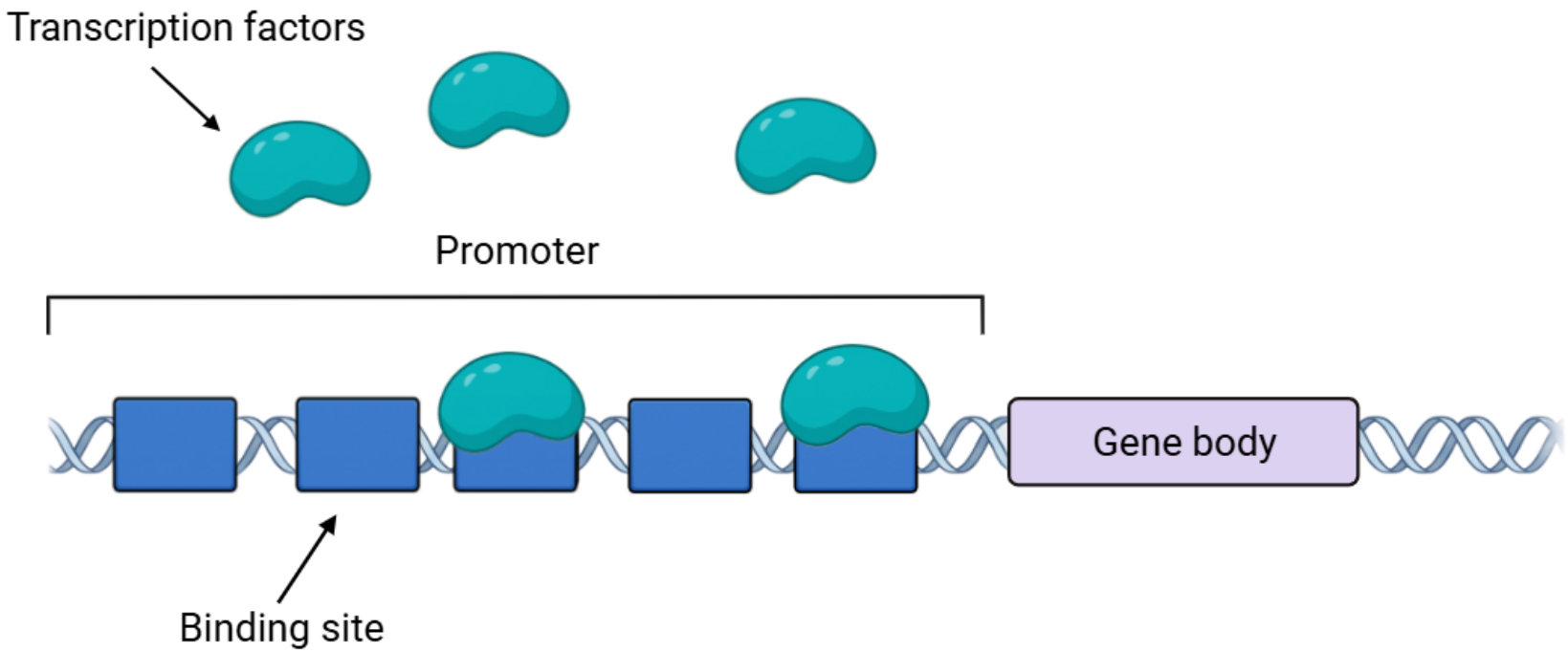
Prokaryotes vs Eukaryotes

Characteristic	**Bacteria**	**Multicellular Eukaryotes**
Genome length	$1–9 \times 10^6$ bp	$10^8–10^{10}$ bp
Regulatory architecture	Relatively compact	Often dispersed, with distant enhancers
Lower bound TF copy number	1–10	$10^3–10^5$

Source: Wunderlich and Mirny (2009).

- Prokaryotic gene regulation in E. coli is extensively studied
- Advantages of eukaryotic gene regulatory mechanics is not well-understood

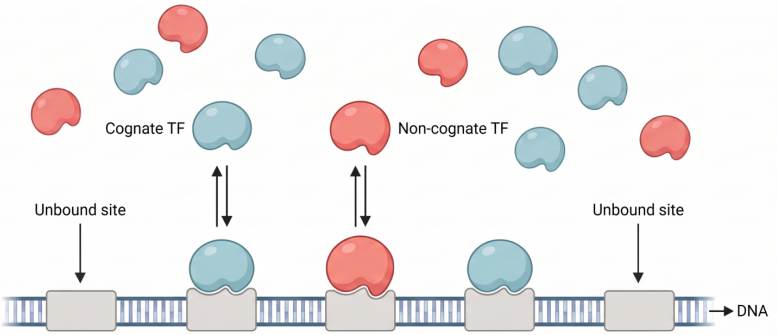
Low-affinity binding sites



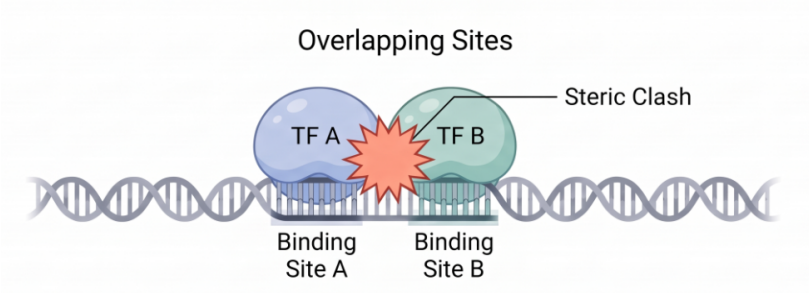
Created with BioRender

- Recent evidence shows the relevance of low-affinity binding sites

Issues with low-affinity binding sites



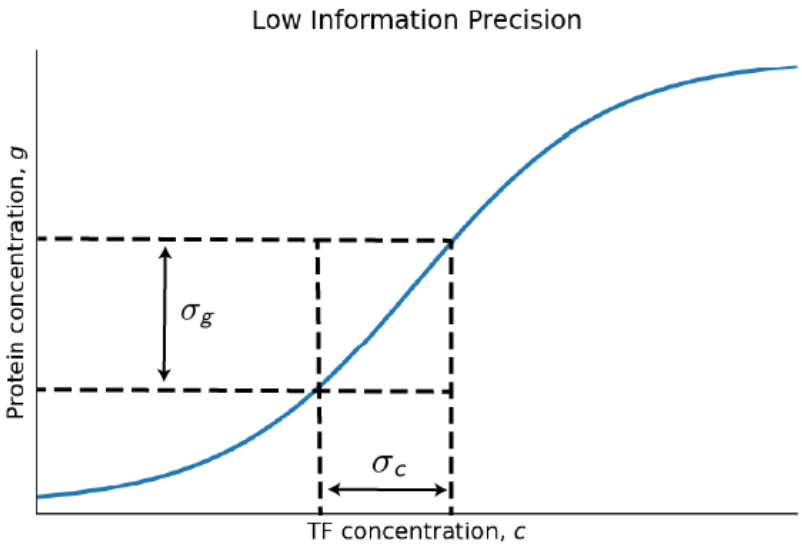
Specificity

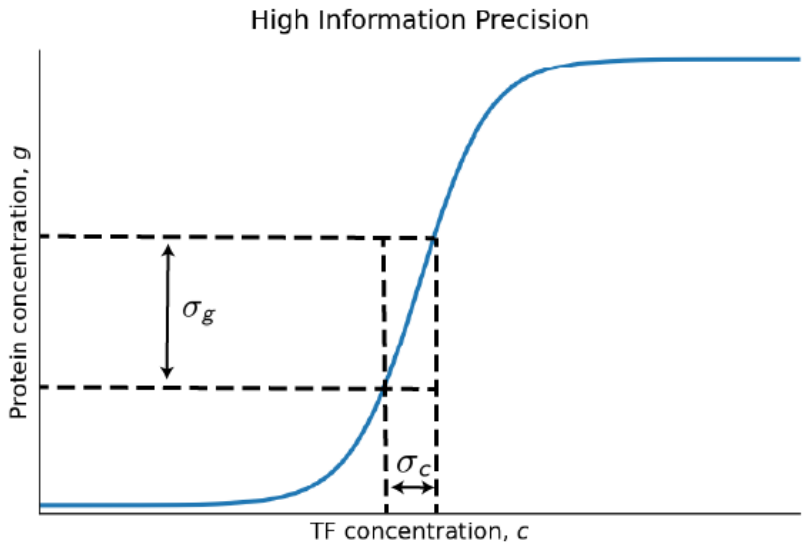


Spacing

- This raises the question: why do they evolve through natural selection?

Information Transfer





- 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*, measured in *bits*, is defined as a difference in entropies

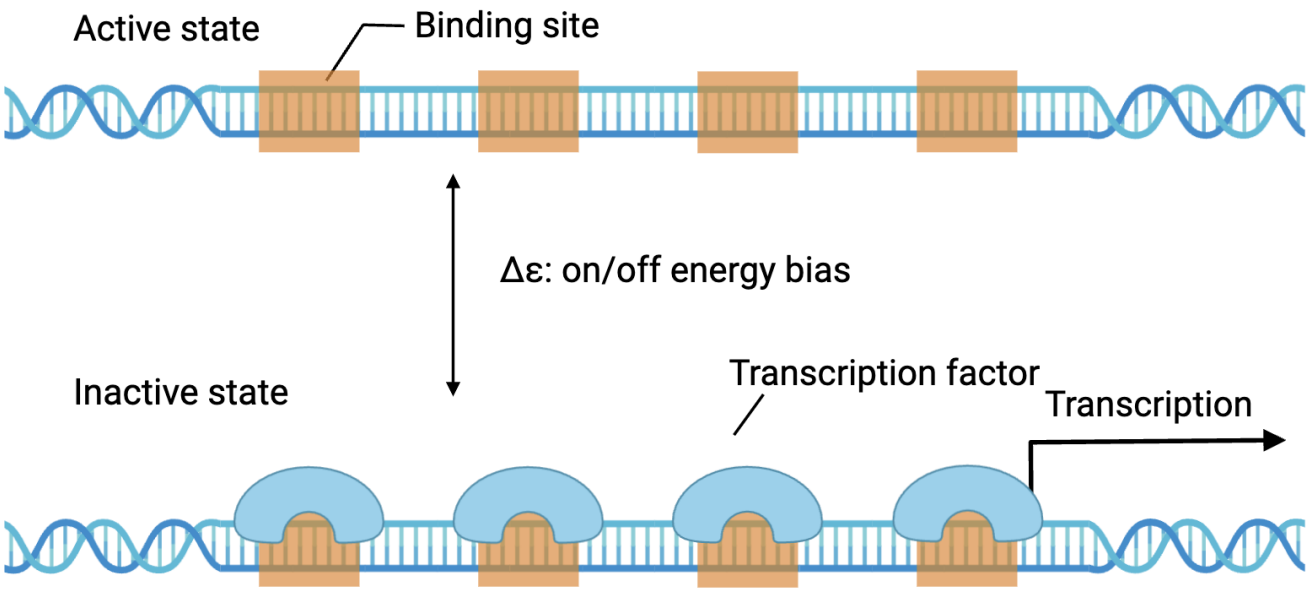
$$I(c; g) = \int dc\, P(c)(S[P(g)] - S[P(g|c)]).$$

Example (Guessing the number game)

Each bit is equivalent to the information gained from a yes/no question.

A noise-considering model of transcription

- Given some TF concentration c, how can we model outputted normalized gene expression g?
- A promoter has n sites with active-state dissociation constant $K_{d,\text{on}}$, with larger values meaning weaker binding. The inactive-state constant, $K_{d,\text{off}}$, is significantly larger.
- The energy bias between the two states is $\Delta\epsilon$.



Created using BioRender

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.

The mean response is calculated through:

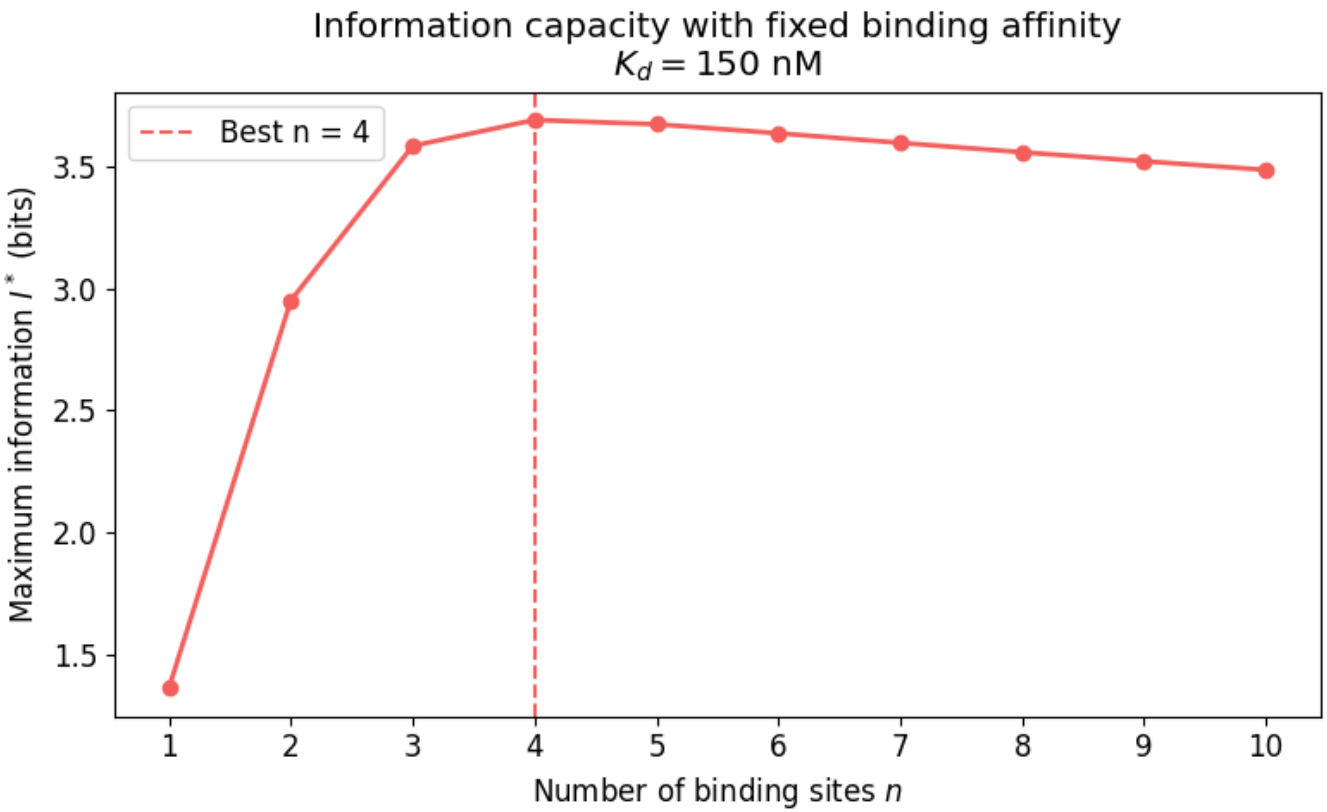
$$\bar{g}(c) = \frac{1}{1 + e^{\Delta\varepsilon}(1 + c/K_d)^{-n}}. \tag{1}$$

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

Several weak sites can improve transcription

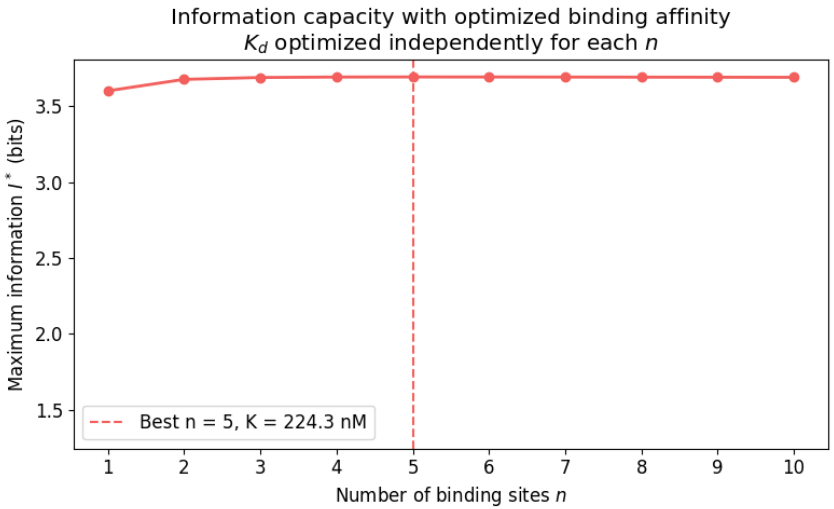
- Define capacity I^* as the maximum mutual information obtained by choosing the input probability density $P(c)$.
- Having multiple low-affinity binding sites can produce outputs comparable to a singular high-affinity binding site.

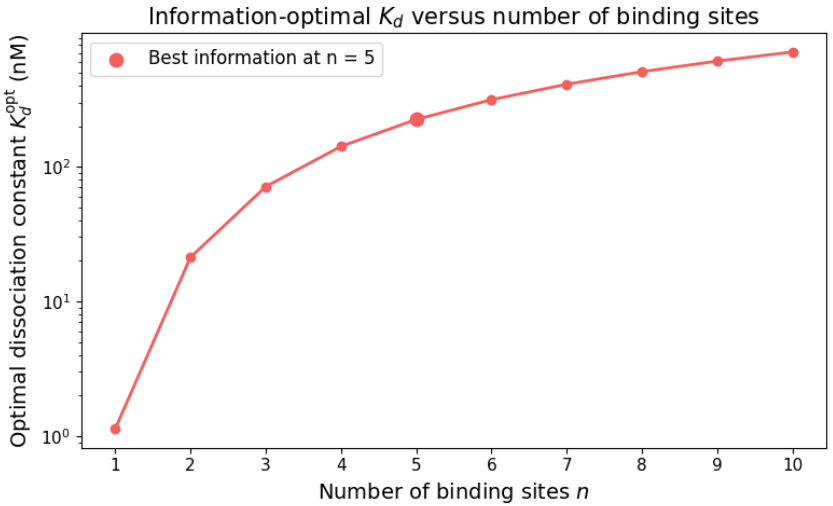
Several weak sites can improve transcription (cont.)



- At $K_d = 150$ nM, we optimize input distribution for each n.
- When $n = 1$, $I^* = 1.36$ bits. When $n = 4$, $I^* = 3.69$ bits.

Several weak sites can improve transcription (cont.)

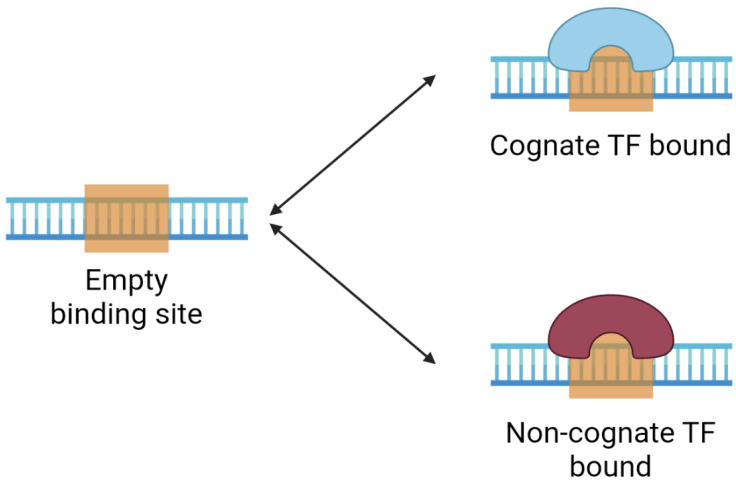




- If we can optimize affinity, one strong site and a cluster of weak sites achieve comparable maximum information.
- Somewhat unintuitively, multiple small weak sites fixes their affinity disadvantage and is comparable with a strong binding site.

Introducing competition

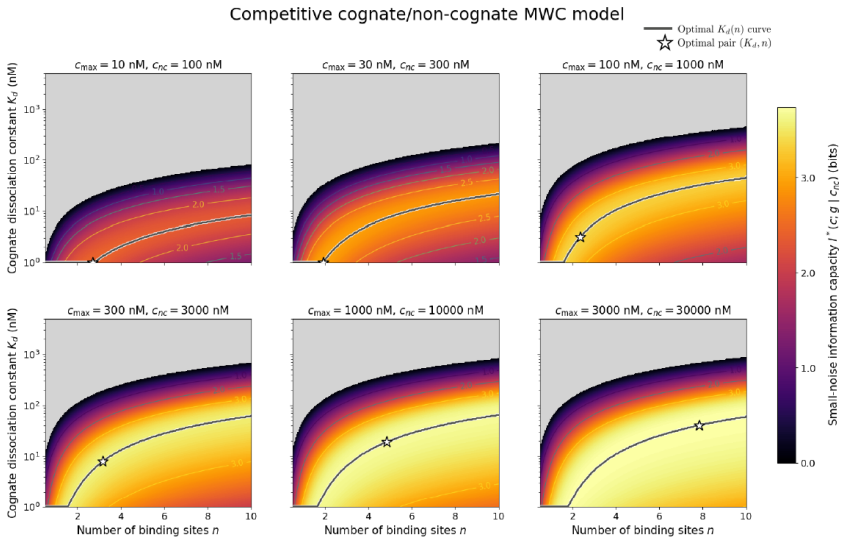
- Only one transcription factor can occupy a binding site at a time.
- Non-cognate transcription factors compete for the same binding sites as cognate transcription factors.



Created using BioRender

Competitor TFs binding and binding adds noise, which makes expression fluctuate.

Competition's effect on performance



Created using BioRender

At higher concentrations, configurations favor more binding sites and weaker optimal binding.

What are the methods we are using?

-statistical mechanics -stochastic model of gene transcription -calculate mutual information from these things and maximizing results. -perhaps put an equation (maybe maximal information transcription or MWC model after explaining everything?)

Result: weaker binding site = need multiple

If you only have one binding site, you have much worse information transmission. Makes up for it for having more. Somewhat unintuitively, it fixes it and is comparable with a strong binding site.

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


S. Black, *On dementors and other immortal creatures*, Proc. Hog. Magic Soc. **75** (1994) 321–401.


M. McGonagall, *Powerful witches: their potential and limitations*, J. Mathemagic **69** (1975) 1–19.


N. Scamander, *Harmless and harmful creatures*, J. Fantast. Creatures **58** (1962) 449–468.


J. Potter and L. Potter, *On special classes of protection spells*. In: Witchcraft Theory and Applications (Eds. A. Dumbledore and S. Snape), pp. 251–279. Lecture Notes in Mathemagic, vol. 1320. Hogsmeade, Scottish Highlands, 1978.

End of Presentation

THANK YOU!