A Scalable Cellular Logic Technology Using Zinc-Finger Proteins

Christopher Batten, Ronny Krashinsky, Thomas Knight, Jr.

Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
June 20, 2004

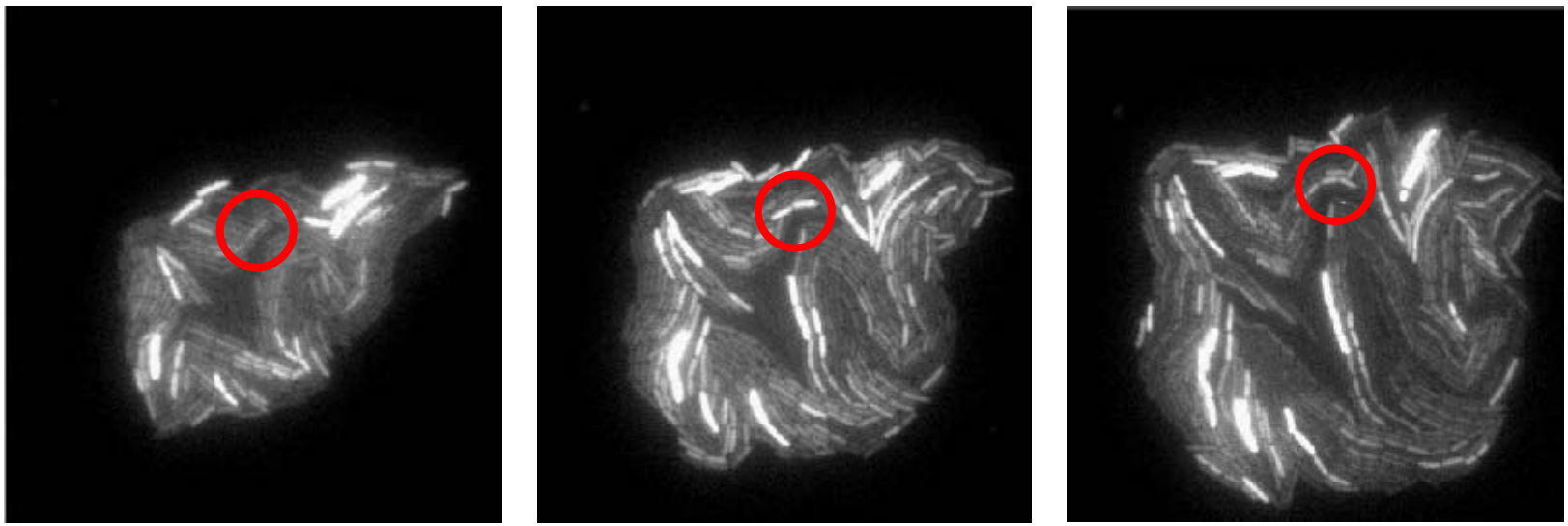
Synthetic Biology

- Synthetic biology hopes to bring engineering practices common in other engineering disciplines to the field of molecular genetics and thus create a novel nanoscale computational substrate
- Advantages
  - Tightly integrated biological inputs and outputs
  - Easily grow thousands of computational engines
  - Natural use of directed evolution
- Disadvantages
  - Speed is on the order of millihertz (tens of seconds)
  - Modest computational capability of each engine

Synthetic biology is not an attempt to replace silicon computing!

Synthetic Biology Applications

- Autonomous biochemical sensors
- Biomaterial manufacturing
- Programmed therapeutics
- Smart agriculture
- Engineered experimental systems for biologists



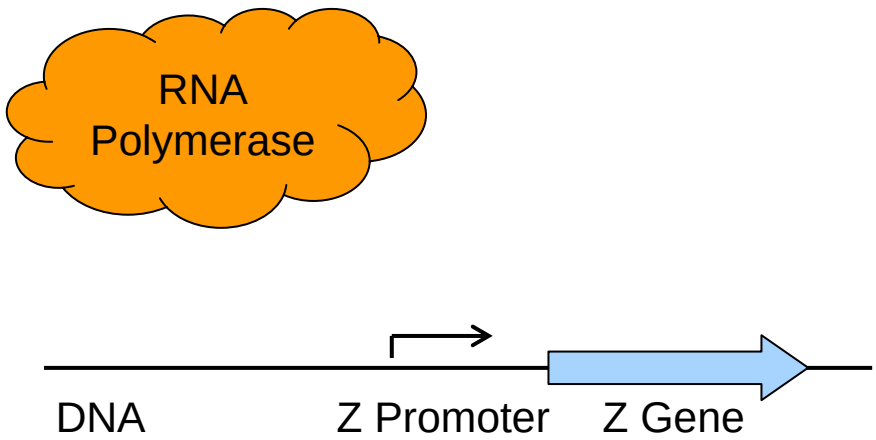
M. Elowitz and S. Leibler
A synthetic oscillatory network of transcriptional regulators
Nature, January 2000

Outline

- Background
  - Protein expression basics
  - Transcription-based cellular logic
  - Zinc-Finger Proteins (ZFPs)
- Proposed ZFP Logic Technology
- Evaluation
  - Analytical model
  - Simulation results
- Future Work and Conclusions

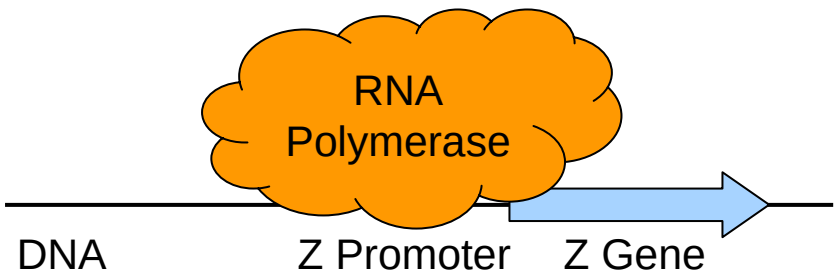
Protein Expression Basics

- RNA polymerase binds to promoter
- RNAP transcribes gene into messenger RNA
- Ribosome translates messenger RNA into protein



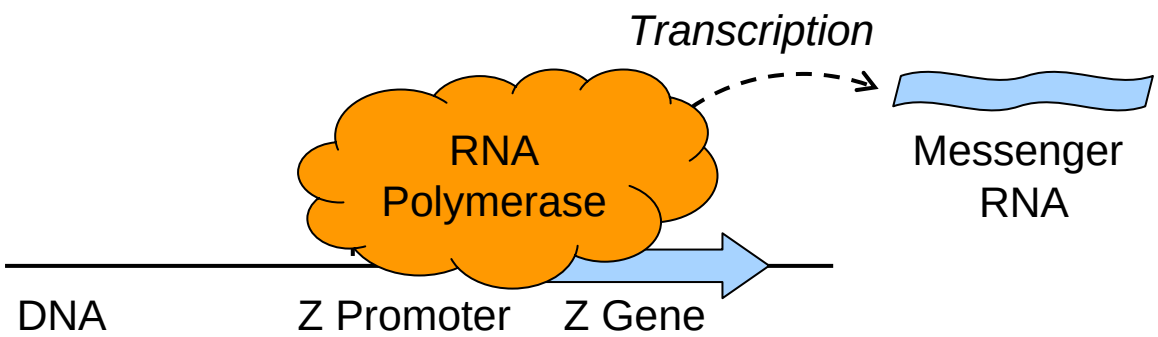
Protein Expression Basics

- **RNA polymerase binds to promoter**
- RNAP transcribes gene into messenger RNA
- Ribosome translates messenger RNA into protein



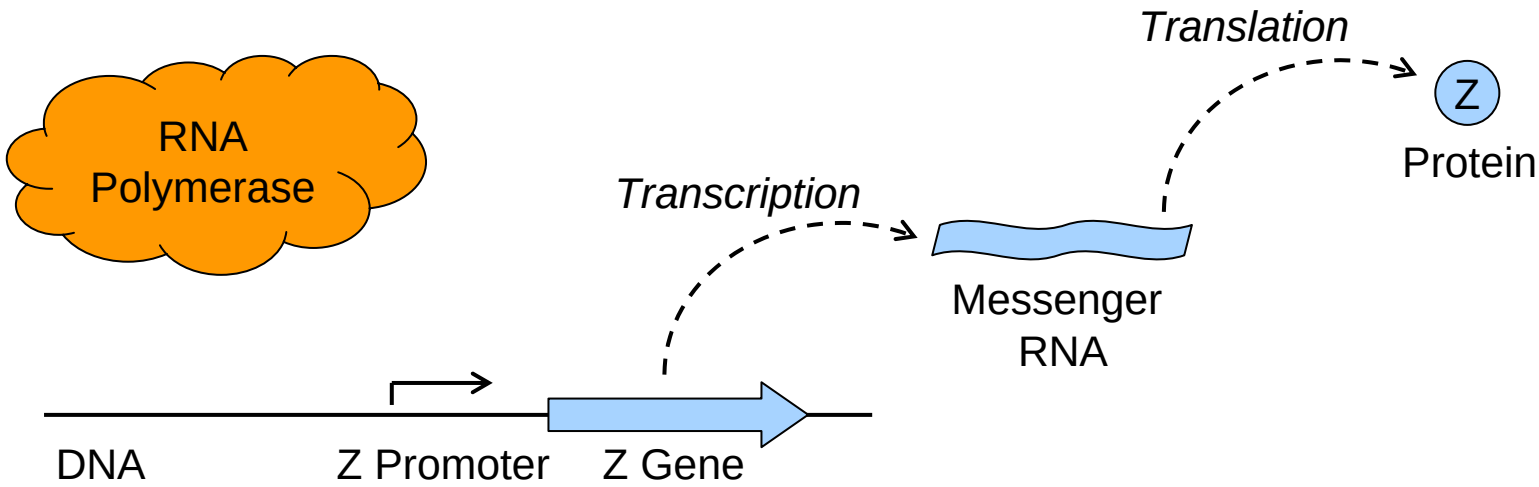
Protein Expression Basics

- RNA polymerase (RNAP) binds to promoter
- **RNAP transcribes gene into messenger RNA**
- Ribosome translates messenger RNA into protein



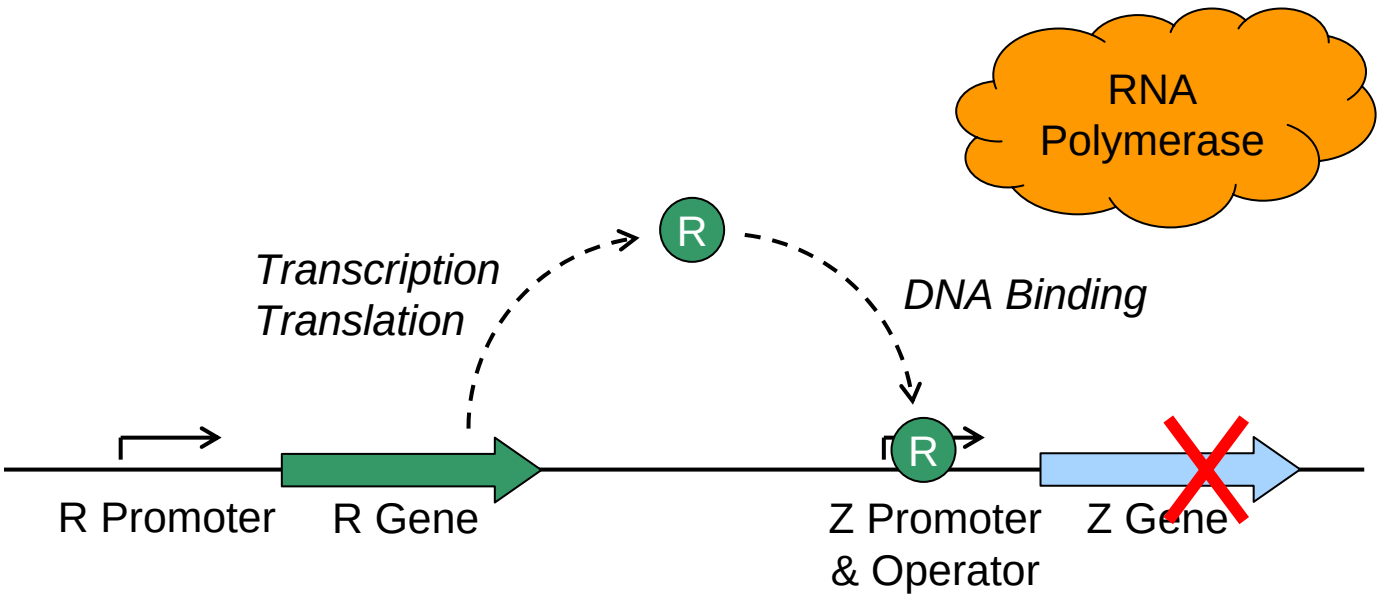
Protein Expression Basics

- RNA polymerase binds to promoter
- RNAP transcribes gene into messenger RNA
- **Ribosome translates messenger RNA into protein**



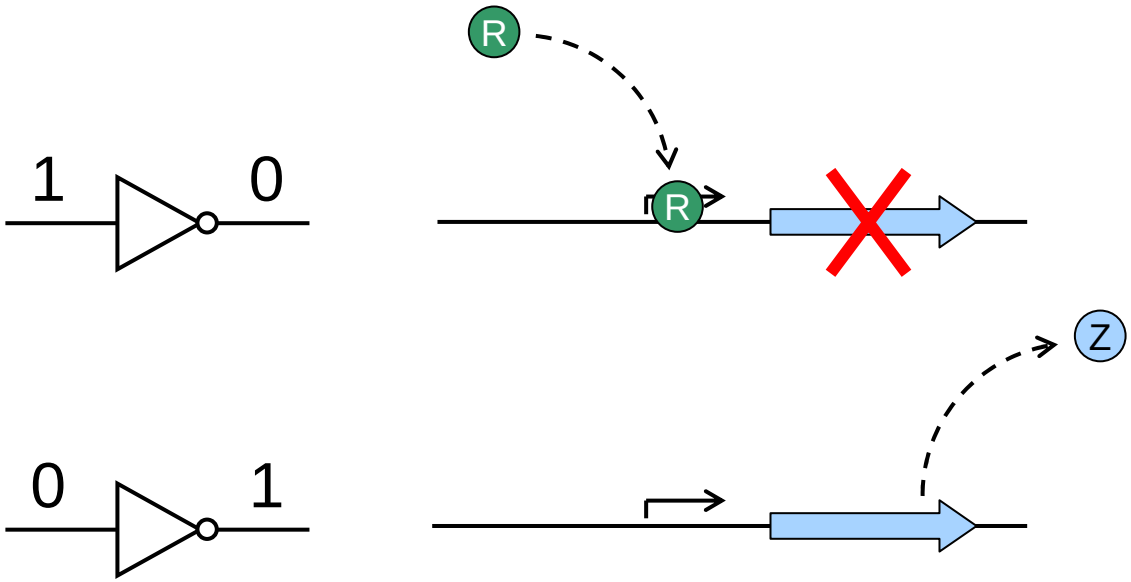
Regulation Through Repression

- Repressor proteins can bind to the promoter and block the RNA polymerase from performing transcription
- The DNA site near the promoter recognized by the repressor is called an **operator**
- The target gene can code for another repression protein enabling regulatory cascades



Transcription-Based Inverter

- Protein concentrations are analogous to electrical wires
- Proteins are not physically isolated, so unique wires require unique proteins



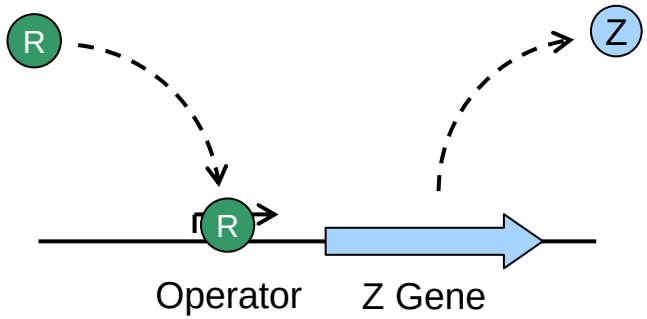
Simple Inverter Model

Chemical Equations

Repressor Binding $R + O \leftrightarrow RO$ $K_{R+R} = (O)(R)/(RO)$

Protein Synthesis $O \rightarrow O + Z$ k_x

Protein Decay $Z \rightarrow$ k_{deg}



Total Concentration Equations

Total Operator $(O_T) = (O) + (RO)$

Total Repressor $(R_T) = (R) + (RO) \approx (R)$ *if* $(R_T) >> (O)$

Transfer Function Derivation

$$\frac{(O)}{(O_T)} = \frac{(O)}{(O) + (RO)} = \frac{1}{1 + (RO)/(O)} = \frac{1}{1 + (R)/K_{R+R}}$$

$$\frac{d(Z)}{dt} = k_x \bullet (O) - k_{deg} \bullet (Z) = 0 \quad \textit{at equilibrium}$$

$$(Z) = \frac{k_x}{k_{deg}}(O) = \frac{k_x}{k_{deg}} \bullet \frac{(O_T)}{1 + (R)/K_{R+R}}$$

Simple Inverter Model

Chemical Equations

Repressor Binding $R + O \leftrightarrow RO$ $K_{R+R} = (O)(R)/(RO)$

Protein Synthesis $O \rightarrow O + Z$ k_x

Protein Decay $Z \rightarrow$ k_{deg}

Total Concentration Equations

Total Operator $(O_T) = (O) + (RO)$

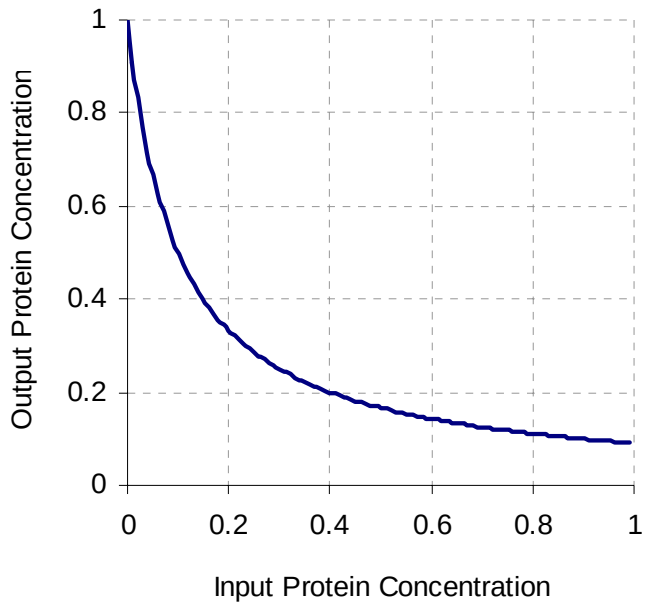
Total Repressor $(R_T) = (R) + (RO) \approx (R)$ *if* $(R_T) >> (O)$

Transfer Function Derivation

$$\frac{(O)}{(O_T)} = \frac{(O)}{(O) + (RO)} = \frac{1}{1 + (RO)/(O)} = \frac{1}{1 + (R)/K_{R+R}}$$

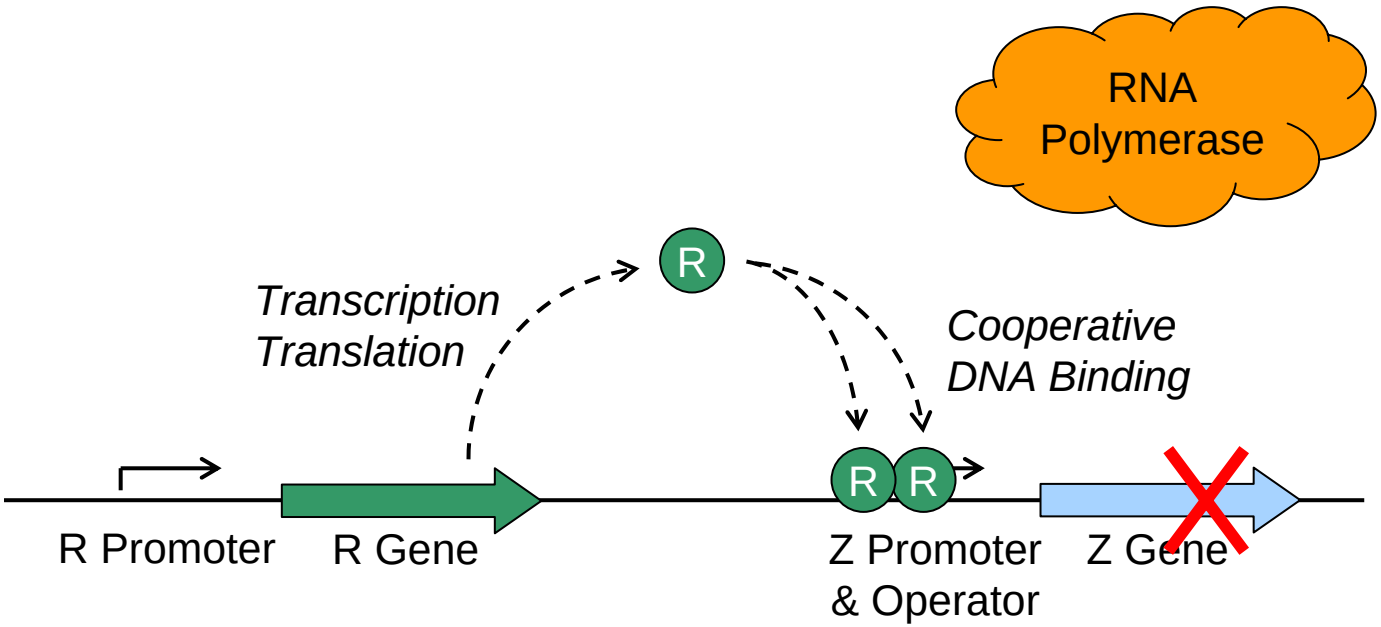
$$\frac{d(Z)}{dt} = k_x \bullet (O) - k_{deg} \bullet (Z) = 0 \quad \textit{at equilibrium}$$

$$(Z) = \frac{k_x}{k_{deg}}(O) = \frac{k_x}{k_{deg}} \bullet \frac{(O_T)}{1 + (R)/K_{R+R}}$$



Cooperativity

- Cooperative DNA binding is where the binding of one protein increases the likelihood of a second protein binding
- Cooperativity adds more non-linearity to the system
  - Increases switching sensitivity
  - Improves robustness to noise



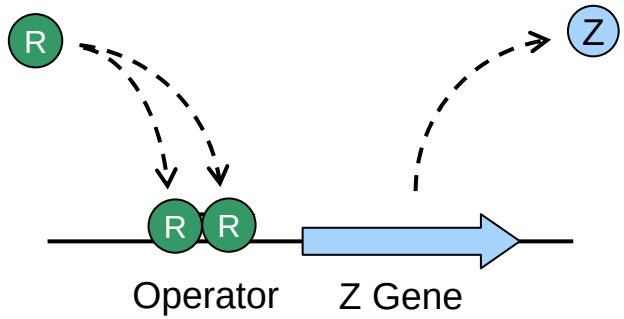
Cooperative Inverter Model

Chemical Equations

Coop Binding $R + R + O \leftrightarrow R_2O$ $K_{R2O} = (O)(R)^2/(R_2O)$

Protein Synthesis $O \rightarrow O + Z$ k_x

Protein Decay $Z \rightarrow$ k_{deg}



Total Concentration Equations

Total Operator $(O_T) = (O) + (R_2O)$

Total Repressor $(R_T) = (R) + 2\bullet(R_2O) \approx (R)$ *if* $(R_T) >> (O)$

Transfer Function Derivation

$$\frac{(O)}{(O_T)} = \frac{(O)}{(O) + (RO)} = \frac{1}{1 + (RO)/(O)} = \frac{1}{1 + (R)^2/K_{R20}}$$

$$\frac{d(Z)}{dt} = k_x \bullet (O) - k_{deg} \bullet (Z) = 0 \quad \textit{at equilibrium}$$

$$(Z) = \frac{k_x}{k_{deg}}(O) = \frac{k_x}{k_{deg}} \bullet \frac{(O_T)}{1 + (R)^2/K_{R+R}}$$

Cooperative Non-Linearity

Cooperative Inverter Model

Chemical Equations

Coop Binding $R + R + O \leftrightarrow R_2O$ $K_{R2O} = (O)(R)^2/(R_2O)$

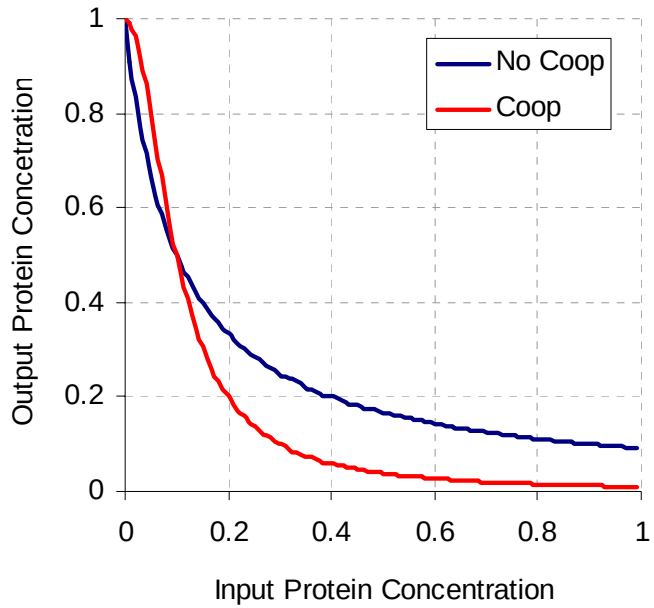
Protein Synthesis $O \rightarrow O + Z$ k_x

Protein Decay $Z \rightarrow$ k_{deg}

Total Concentration Equations

Total Operator $(O_T) = (O) + (R_2O)$

Total Repressor $(R_T) = (R) + 2\bullet(R_2O) \approx (R)$ *if* $(R_T) >> (O)$



Transfer Function Derivation

$$\frac{(O)}{(O_T)} = \frac{(O)}{(O) + (RO)} = \frac{1}{1 + (RO)/(O)} = \frac{1}{1 + (R)^2/K_{R20}}$$

$$\frac{d(Z)}{dt} = k_x \bullet (O) - k_{deg} \bullet (Z) = 0 \quad \textit{at equilibrium}$$

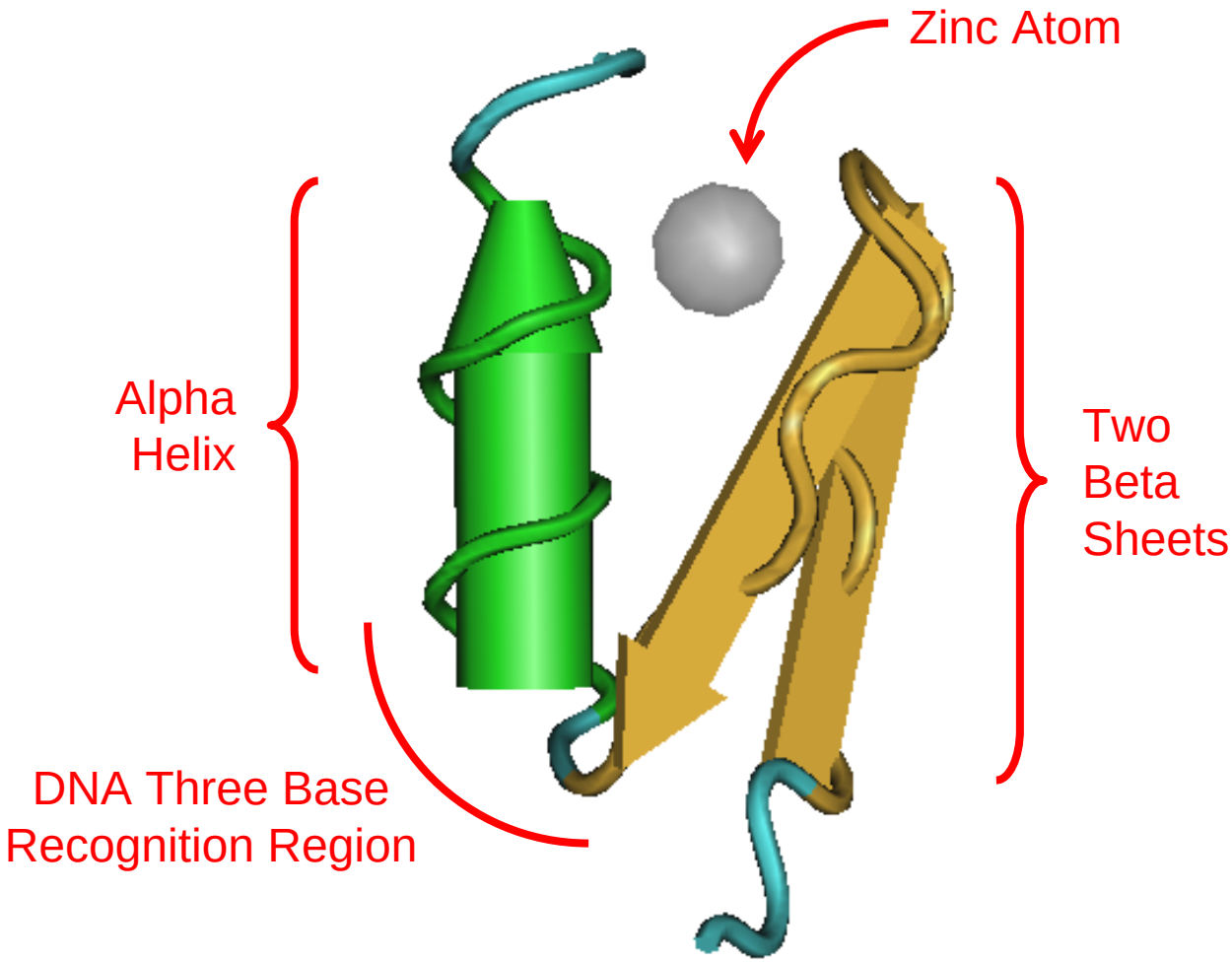
$$(Z) = \frac{k_x}{k_{deg}}(O) = \frac{k_x}{k_{deg}} \bullet \frac{(O_T)}{1 + (R)^2/K_{R+R}}$$

Cooperative Non-Linearity

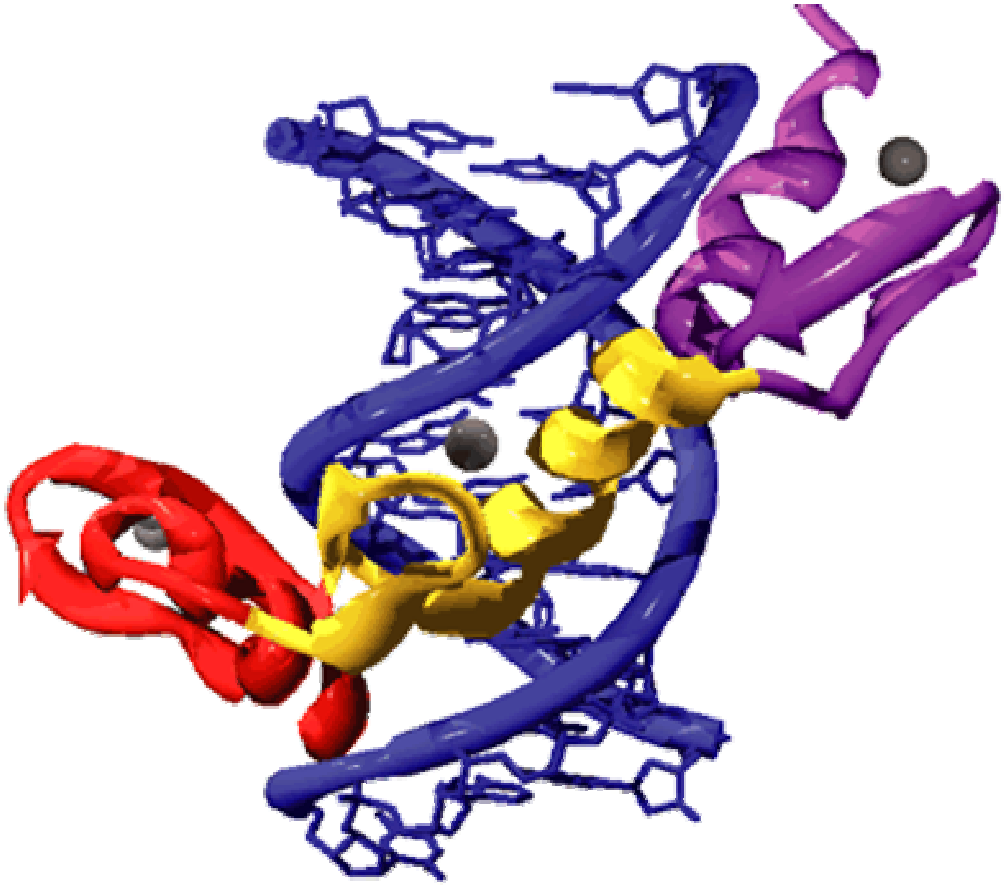
Cellular Logic Summary

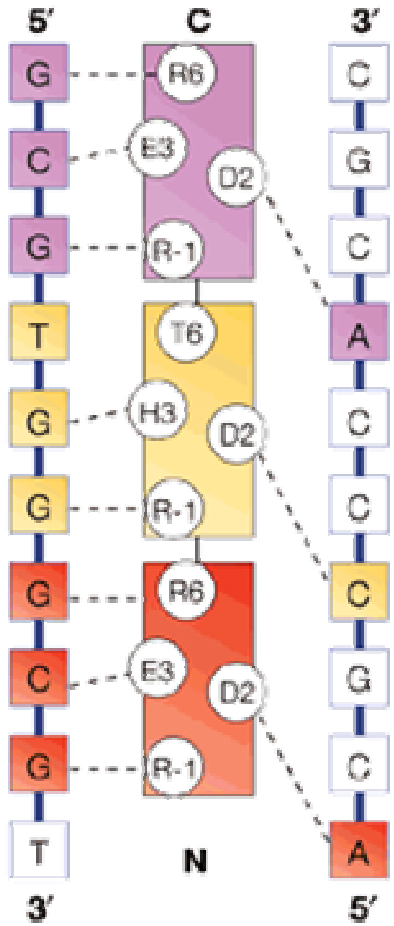
- Current systems are limited to less than a dozen gates
  - Three inverter ring oscillator [Elowitz00]
  - RS latch [Gardner00]
  - Inter-cell communication [Weiss01]
- A natural repressor-based logic technology presents serious scalability issues
  - Scavenging natural repressor proteins is time consuming
  - Matching natural repressor proteins to work together is difficult
- Sophisticated synthetic biological systems require a scalable cellular logic technology with good cooperativity
  - Zinc-finger proteins can be engineered to create many unique proteins relatively easily
  - Zinc-finger proteins can be fused with dimerization domains to increase cooperativity
  - A cellular logic technology of only zinc-finger proteins should hopefully be easier to characterize

Single Zinc-Finger Structure



Poly-Finger ZFPs





A.C. Jamieson, J.C. Miller, and C.O. Pabo.
Drug discovery with engineered zinc-finger proteins.
Nature Reviews Drug Discovery, May 2003

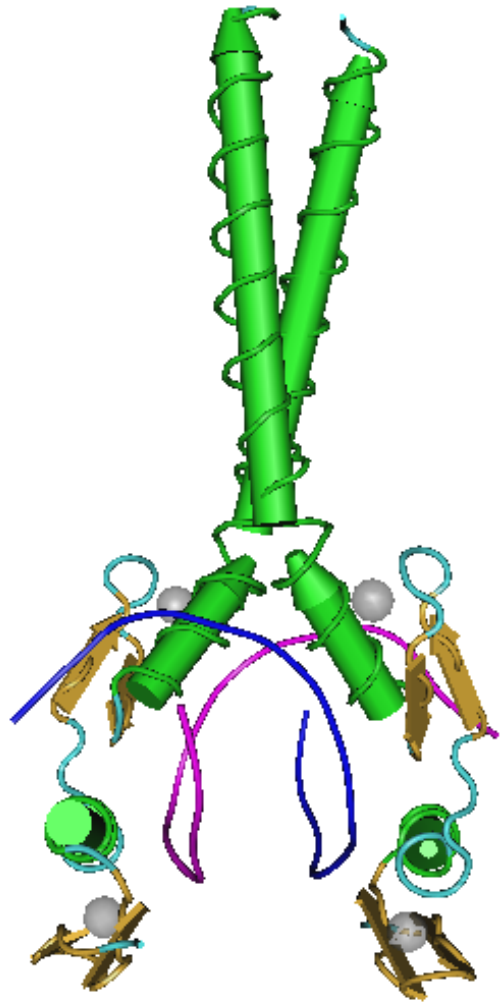
Engineering ZFPs

- Early hopes for a code to simply map amino-acid residues to DNA bases have not materialized [Choo94]
- Some success has been had engineering ZFP fingers to recognize GNNG sequences [Dreier00, Segal99]
- These GNNG fingers can then be easily composed into poly-finger ZFPs
- Recent work has broadened these techniques to include ANNA fingers [Dreier01]

We are nearing the point where an appropriate poly-finger ZFP can be easily composed from a library of fingers to recognize almost any DNA sequence

Engineering ZFP Dimers

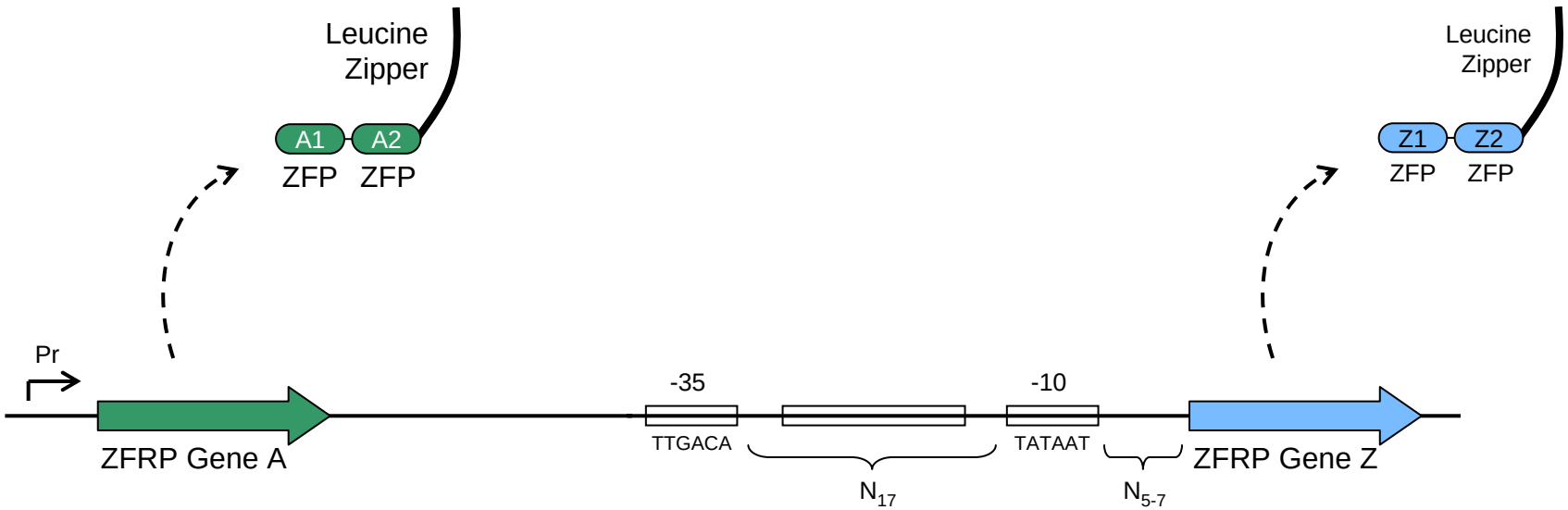
- Dimerization is the natural phenomenon where two proteins bind together
- Dimerization is a form of cooperative DNA binding and increases cooperativity
- Two-finger ZFPs have been fused to GCN4 leucine zipper dimerization domains to create cooperative ZFP DNA binding proteins [Wolfe00]



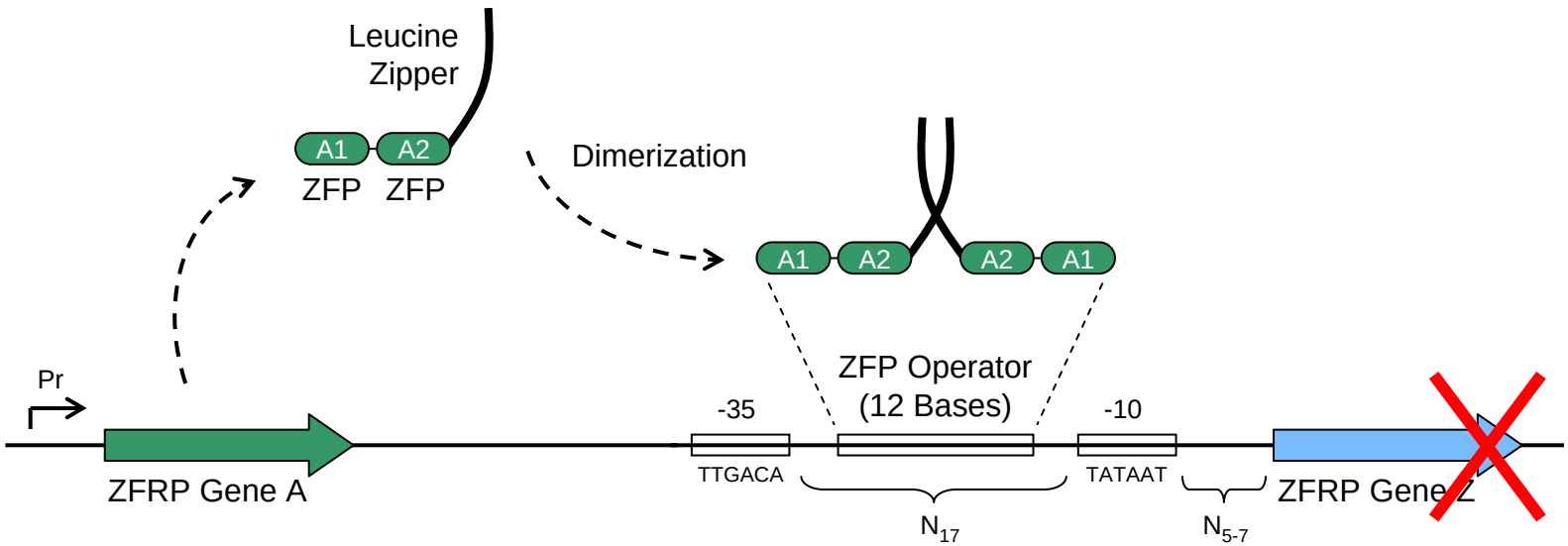
Proposed ZFP Logic Technology

- Use two-finger ZFPs fused to a GCN4 leucine zipper as basic repressor monomer
- Each gate/wire has a unique engineered ZFP
- Why two-finger monomers?
  - Recognizes 6 base pairs permitting an encoding space suitable for hundreds of gates
  - Specificity suitable for *E. coli* genome
  - Affinity suitable for biologic circuit dynamics
- Since all gates have identical leucine zipper dimerization domains, monomers from different gates could dimerize causing **inter-gate interference**

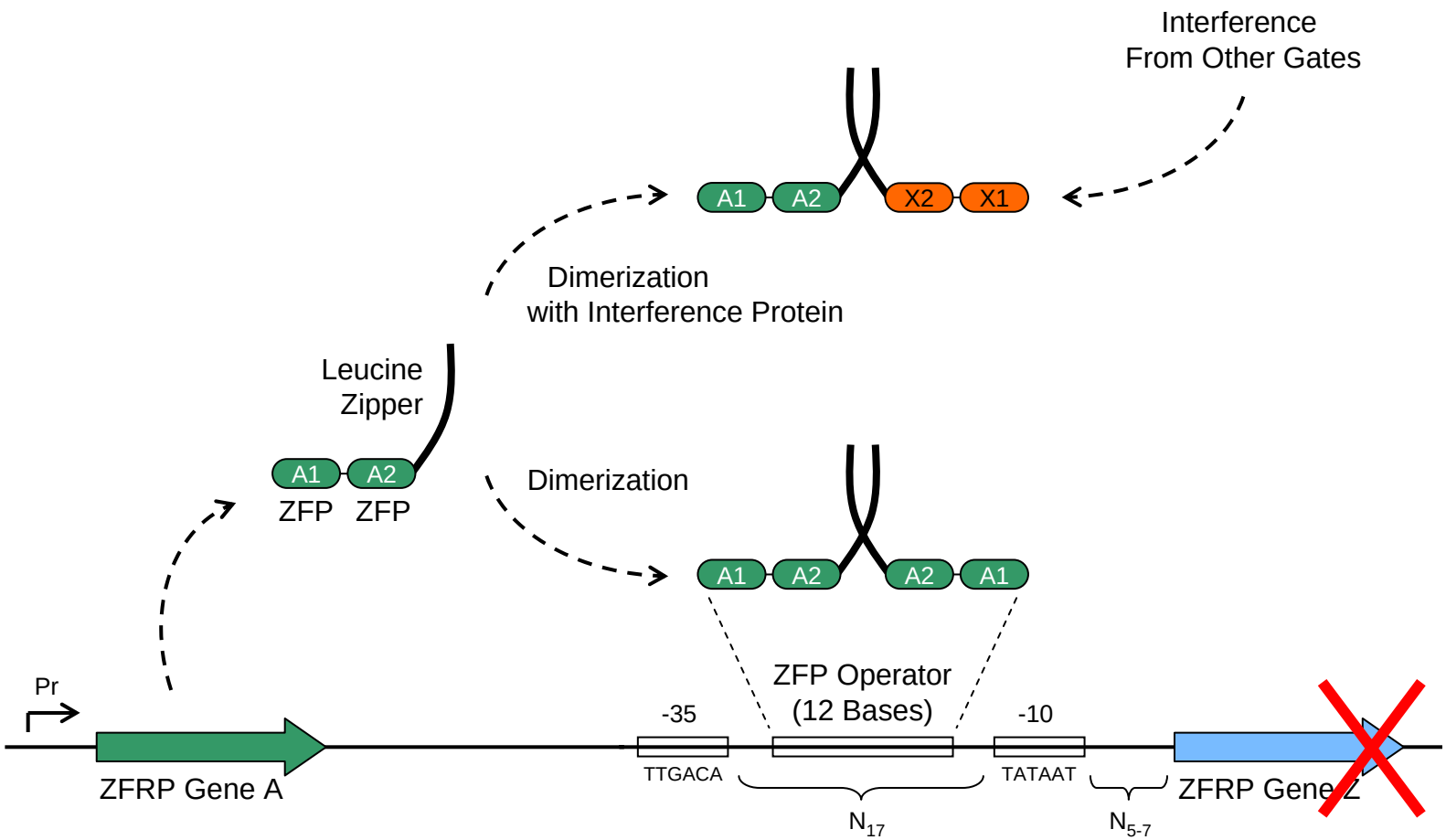
Proposed ZFP Logic Technology



Proposed ZFP Logic Technology



Proposed ZFP Logic Technology



Analytical Model

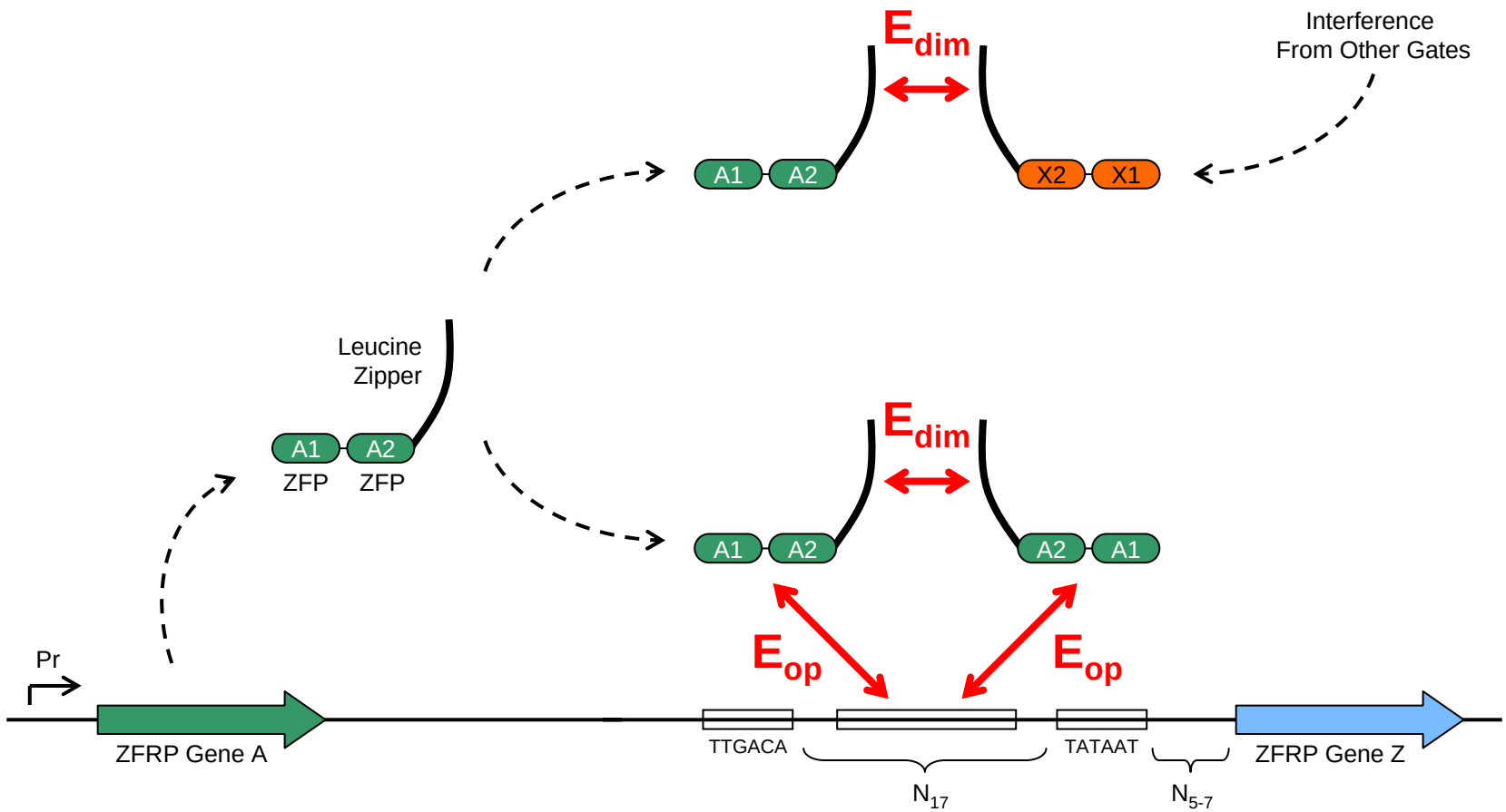
Dimerization	$R + R$	$\leftrightarrow$	R_2	K_{R+R}	=	$(R)^2/(R_2)$	=	$e^{Edim/RT}$
Dimer Binding	$O + R_2$	$\leftrightarrow$	R_2O	K_{R2+O}	=	$(O)(R_2)/(R_2O)$	=	$e^{2Eop/RT}$
Monomer Binding	$O + R$	$\leftrightarrow$	OR	K_{R+R}	=	$(O)(R)/(OR)$	=	$e^{Eop/RT}$
Monomer Binding	$R + O$	$\leftrightarrow$	RO	K_{R+R}	=	$(O)(R)/(RO)$	=	$e^{Eop/RT}$
Cooperative Binding	$OR + R$	$\leftrightarrow$	R_2O	K_{OR+R}	=	$(OR)(R)/(R_20)$	=	$e^{(Eop+Edim)/RT}$
Cooperative Binding	$RO + R$	$\leftrightarrow$	R_2O	K_{RO+R}	=	$(RO)(R)/(R_20)$	=	$e^{(Eop+Edim)/RT}$
Protein Synthesis	O	$\rightarrow$	$O + Z$	k_x				
Protein Decay	Z	$\rightarrow$		k_{deg}				
Dimerization	$X + X$	$\leftrightarrow$	X_2	K_{X+X}	=	$(X)^2/(X_2)$	=	$e^{Edim/RT}$
Inter-Gate Interference	$X + R$	$\leftrightarrow$	XR	K_{X+R}	=	$(X)(R)/(XR)$	=	$e^{Edim/RT}$

K : Equilibrium dissociation constant

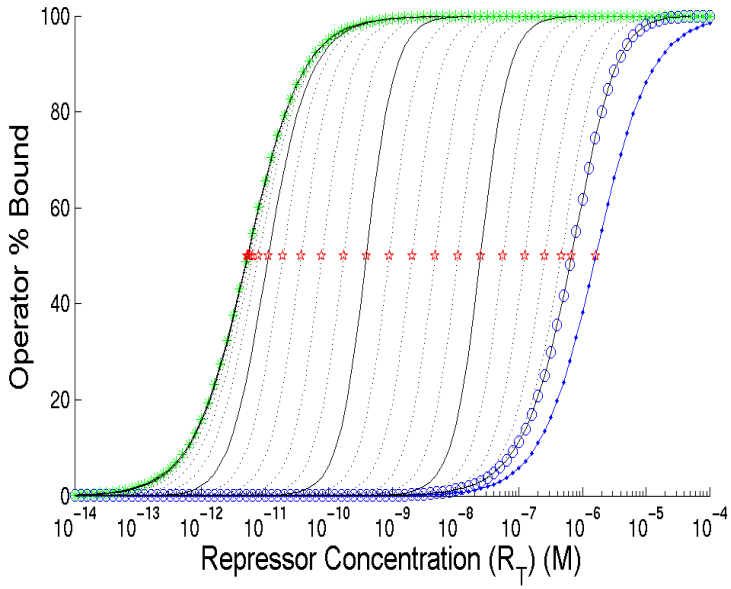
k : Dynamic rate constant

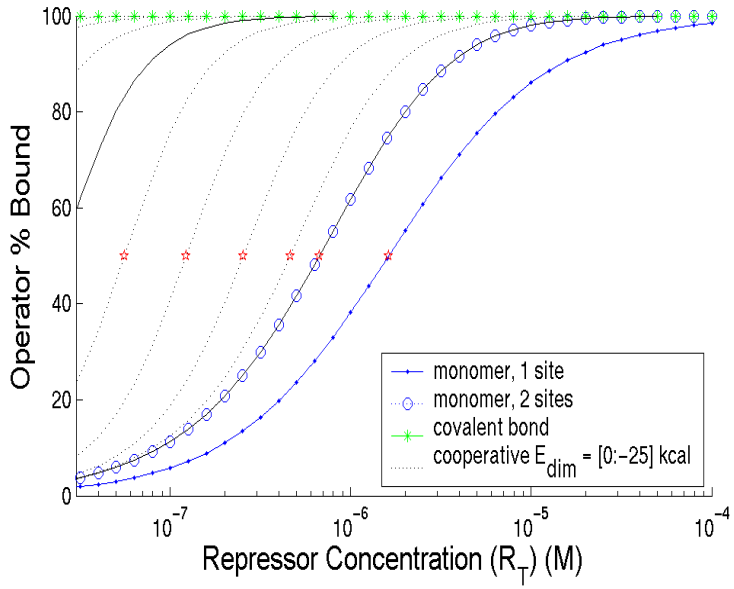
E : Binding energy or change in potential energy caused by the reaction
More negative E means the reaction is more likely to occur

Dimerization and Operator Energy



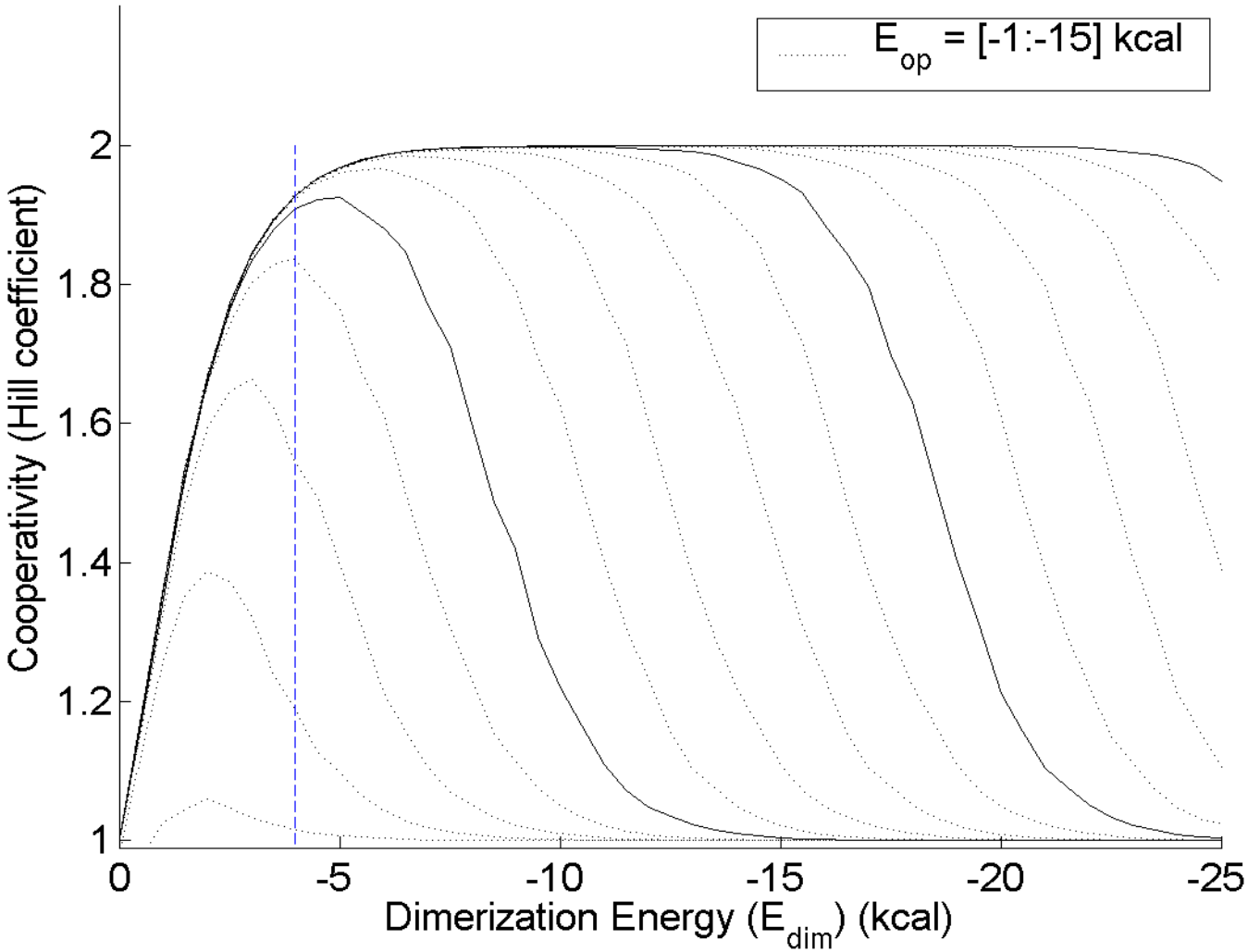
Percent Operator Bound



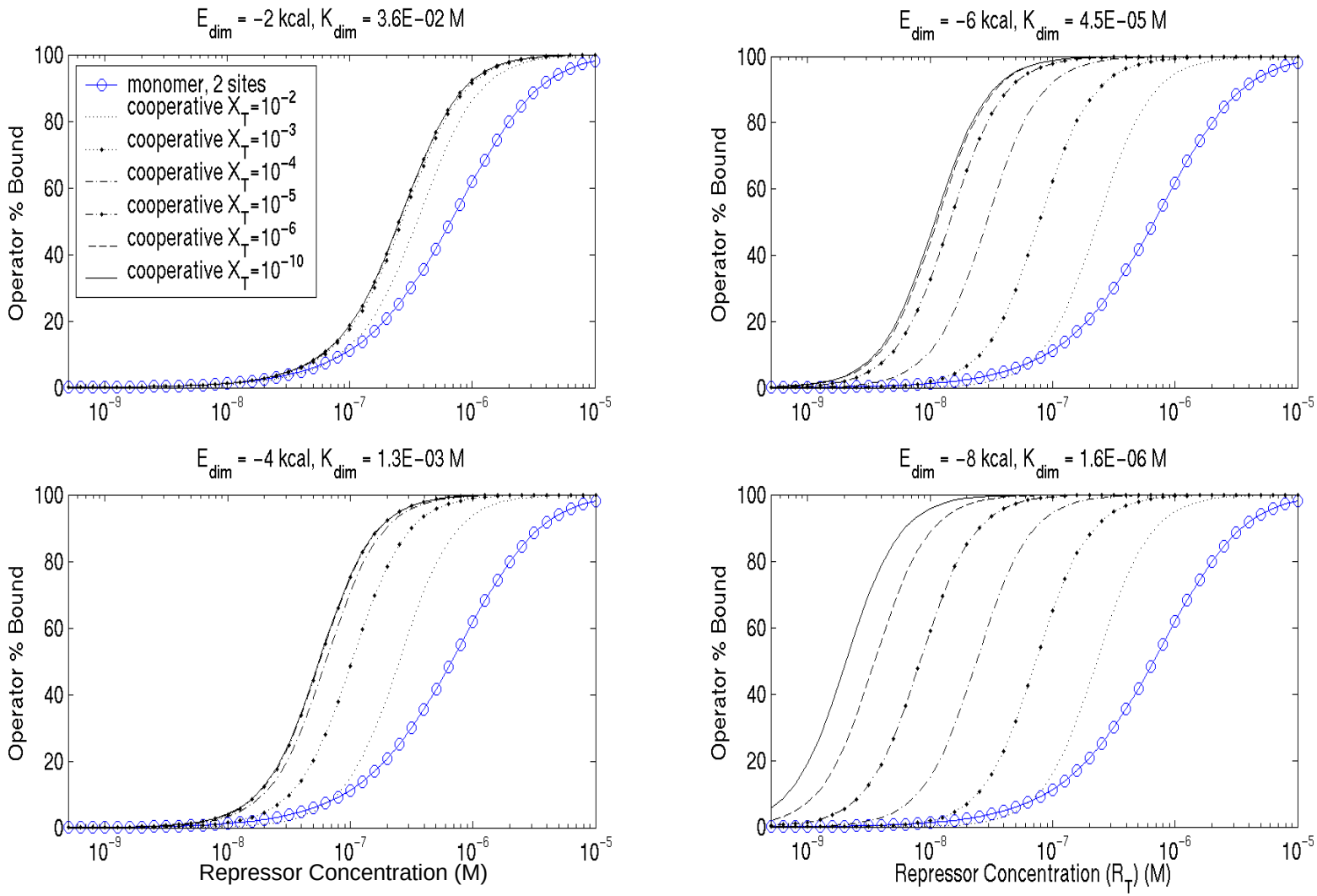


- For **very low dimerization energies**, system approaches uncooperative repressor monomer system
- For **very high dimerization energies**, system approaches uncooperative covalently bonded repressor system
- For **moderate dimerization energies**, the system is cooperative ie. the slope of the curve is steeper than for the uncooperative systems

Cooperativity



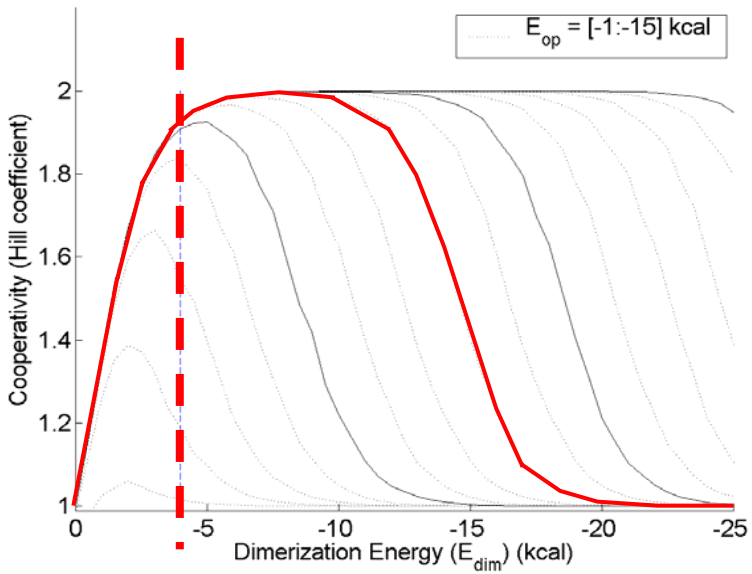
Inter-Gate Interference

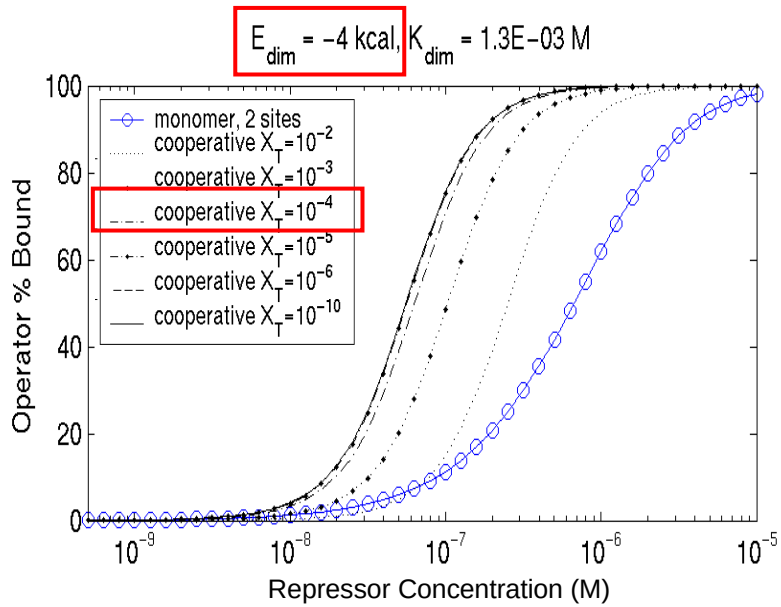


Desired Dimerization Energy

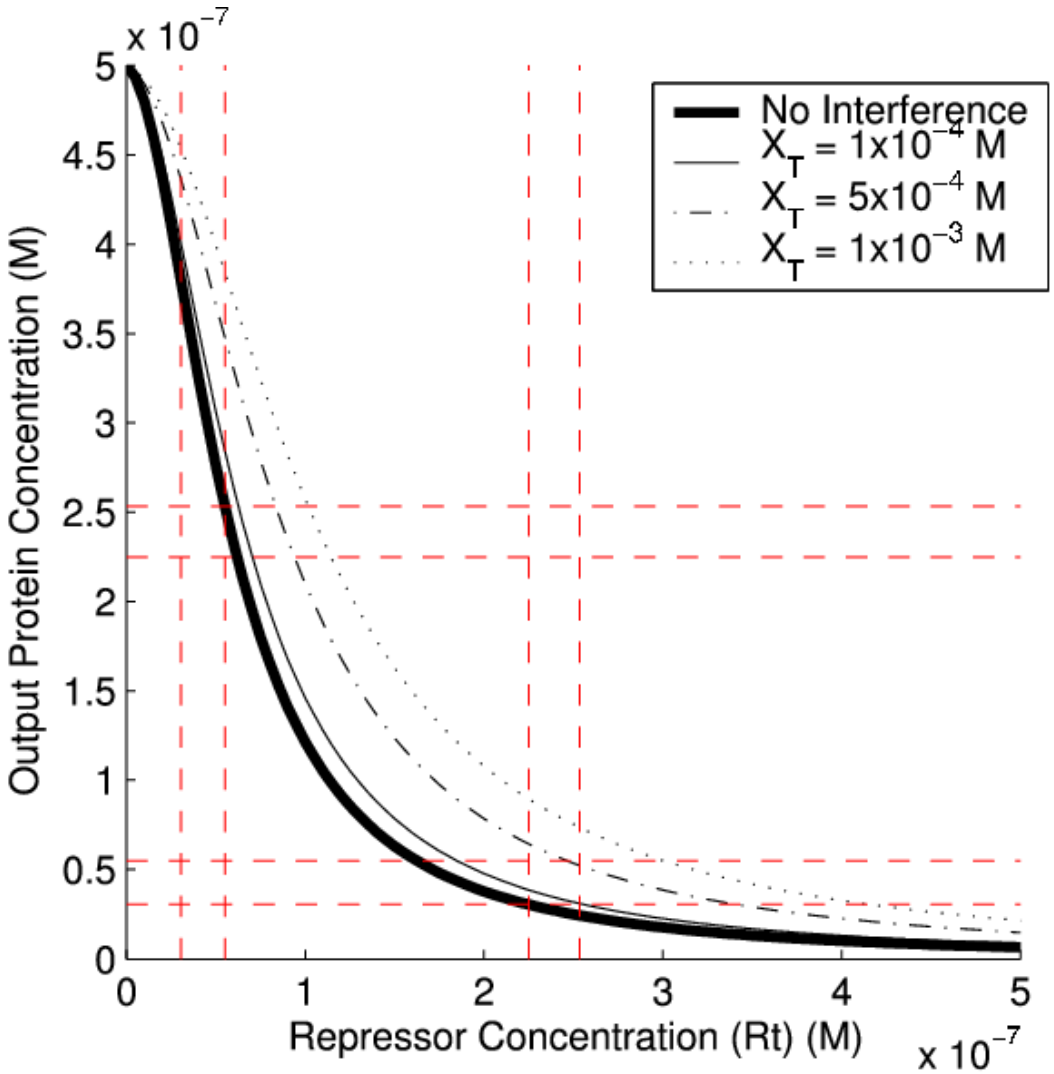
- Tradeoffs in setting the dimerization energy
  - Stronger dimerization energy increases cooperativity
  - Stronger dimerization energy increases inter-gate interference

We desire the weakest dimerization energy which still achieves the maximum cooperativity

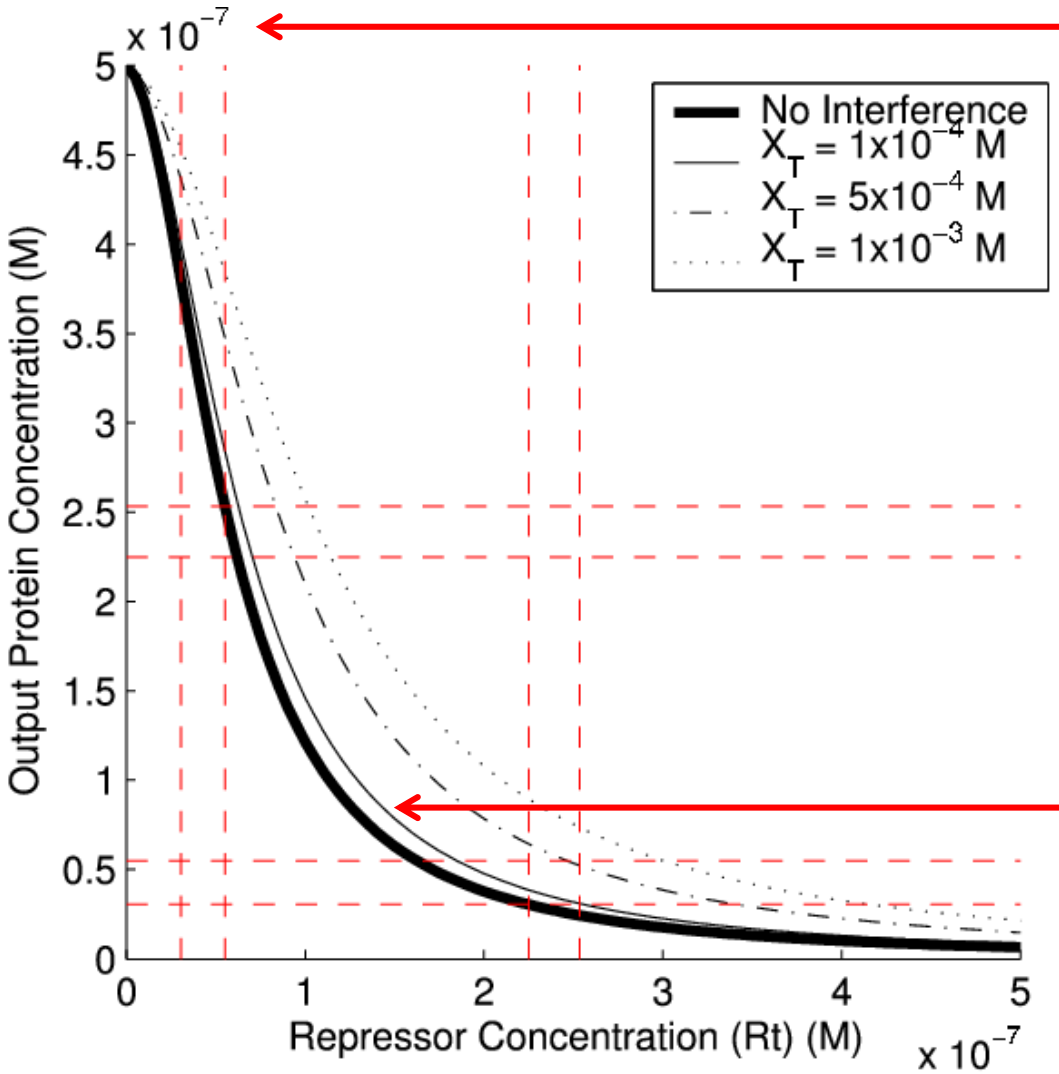




Transfer Curve and Interference



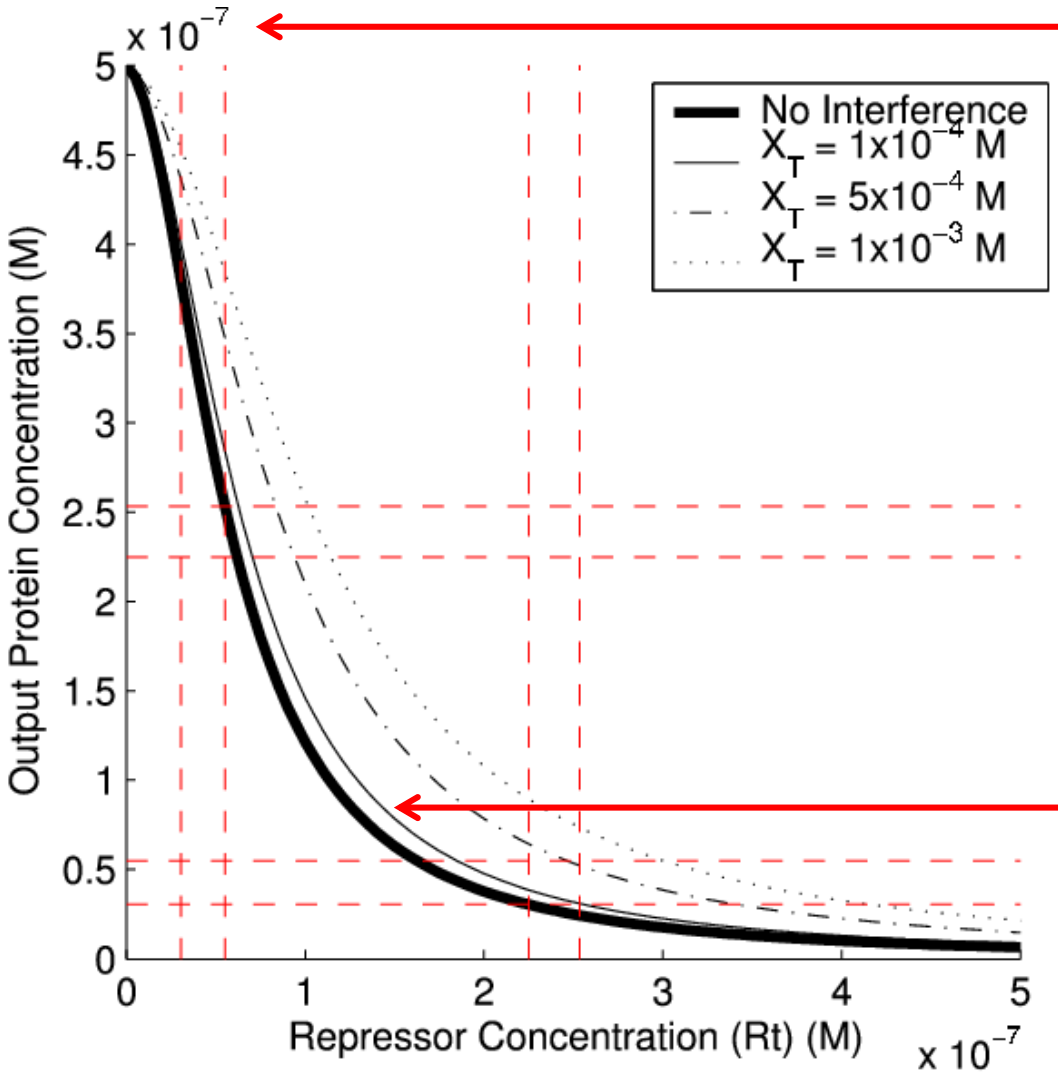
Transfer Curve and Interference



Max output protein concentration per gate is 5×10^{-7} M

Inter-gate interference must be below 10^{-4} M

Transfer Curve and Interference

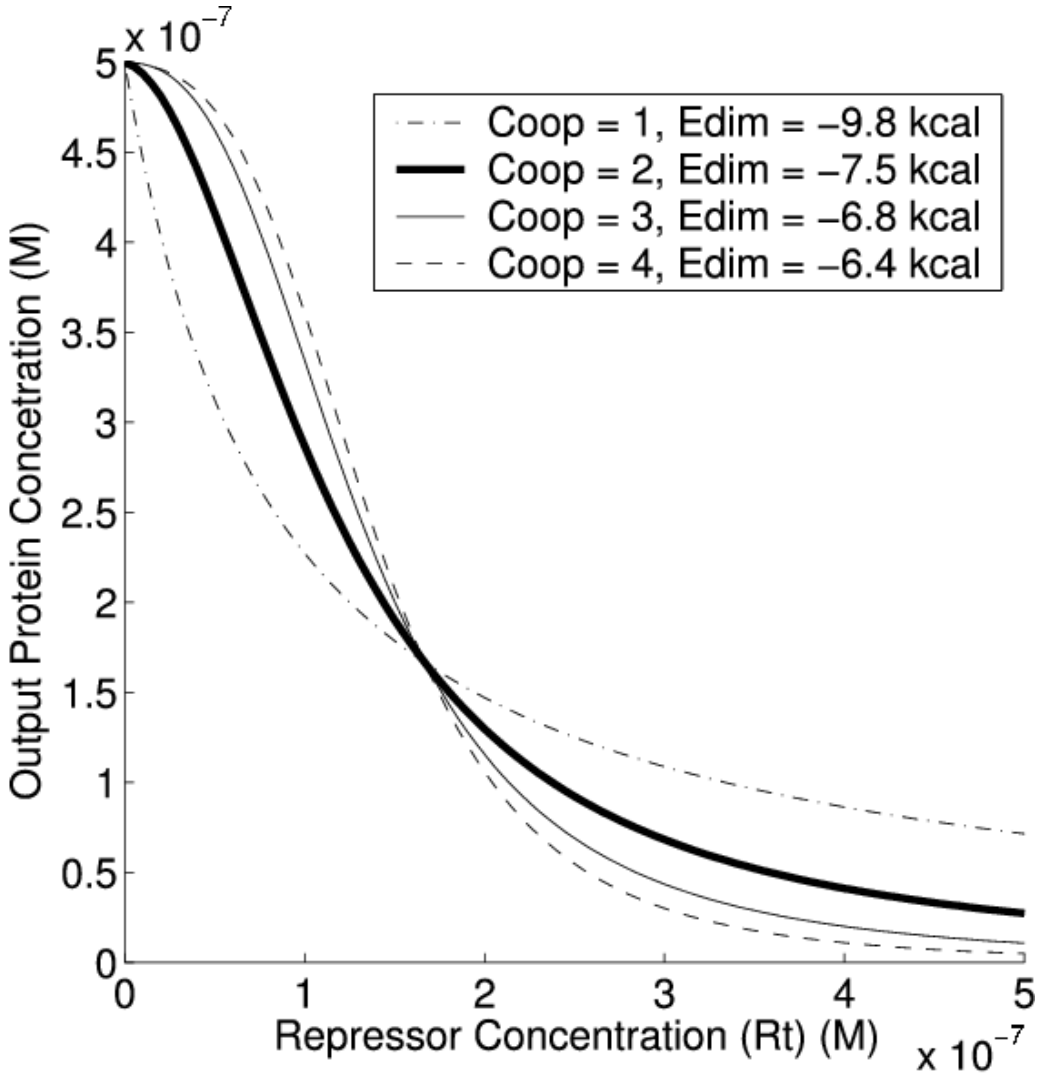


Max output protein concentration per gate is 5×10^{-7} M

To first order, could have $10^{-4} / 5 \times 10^{-7} \approx 200$ gates

Inter-gate interference must be below 10^{-4} M

Transfer Curve and Cooperativity



Future Work

- Model and Design Improvements
  - Model system transient response
  - Model stochastic effects
  - Design a system with increased cooperativity
- Implementation
  - Simple test circuits to investigate use of two finger ZFP dimer as a cooperative repressor in *E. coli*
  - Engineered zinc-finger system with heterodimers to implement more complex logic gates

Conclusions

- Current natural repressor-based biological circuits are limited to less than a dozen gates
- A cellular logic technology based on zinc-finger proteins should enable hundreds of gates
- Careful engineering of the dimerization energy can help mitigate inter-gate interference without sacrificing cooperativity