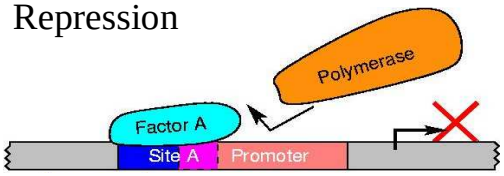
Computing **by** biology

- Survival of cells requires constant monitoring and processing of environmental and cellular information
  - availability of resources: sugars, amino acids, minerals, …
  - external state: temperature, pH, osmotic presure, …
  - internal state: mitosis, myosis, development, …
  - inter-cellular signals: phemerones, hormones, …
- Limited number of cellular components (genes/proteins) requires combinatorial strategies for decision making
  - protein networks: proteins phorphorylating each other
  - gene networks: genes turning each other on/off
- Changing habitats: hard-wired computational processes need to be readily evolvable
  (e.g., programmable rather than dedicated computer)

➔ **Discuss in the context of gene regulation**

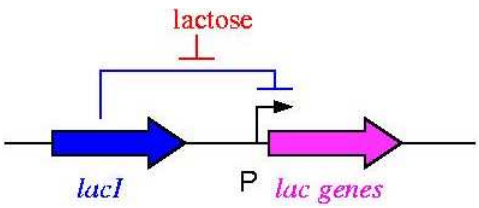
Transcription regulation in bacteria

Repression



Coupling to environment: TF-DNA binding modulated by inducers or phosphorylation

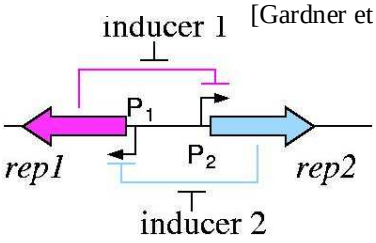
e.g., regulation of the *lac*-operon



Simple genetic circuits

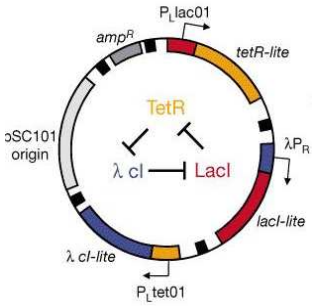
- toggle switch:
  - inducer 1 ON: rep2 ON & rep1 OFF
  - inducer 2 ON: rep1 ON & rep2 OFF

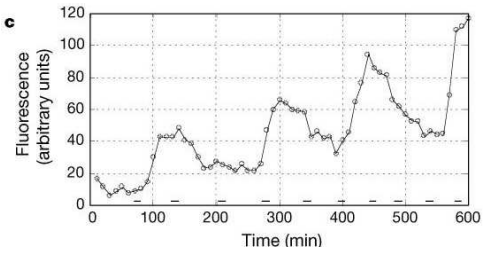
[Gardner et al, '00]



- repressilator

[Elowitz & Leibler, '00]



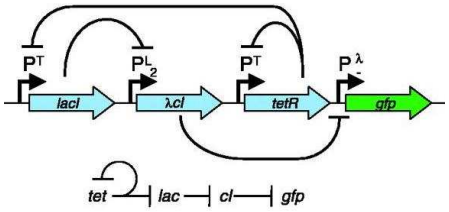


Simple logic devices

[Guet et al, '02]

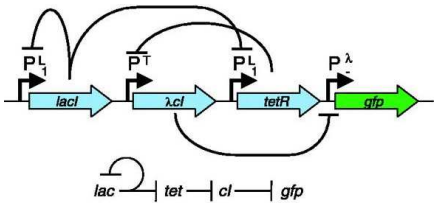
– NAND gate

IPTG/aTc	gfp
lo/lo	ON
lo/hi	ON
hi/lo	ON
hi/hi	OFF



– NOR gate

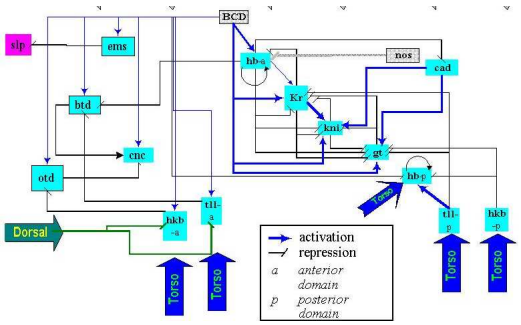
IPTG/aTc	gfp
lo/lo	ON
lo/hi	OFF
hi/lo	OFF
hi/hi	OFF

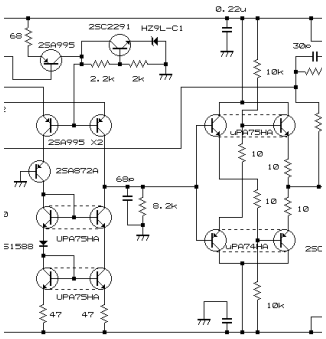


[Kauffman, Thomas & D'ari]

➔ complex computations from many genes linked by intricate connections (gene regulatory network)

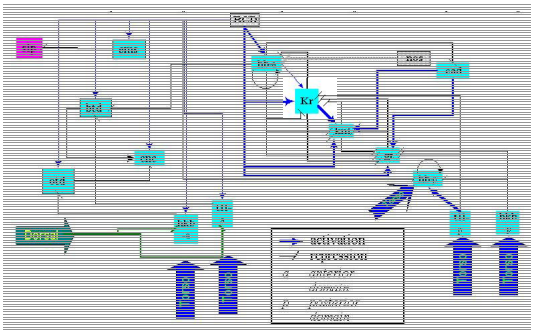
Gene Regulatory Network

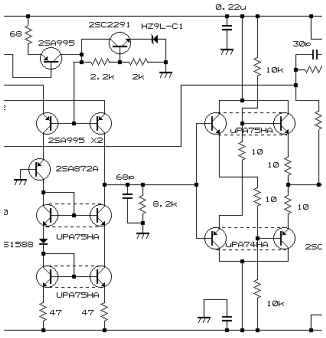




node	gene	transistor
# nodes	1 ~ 1000	10 ~ 10^7
speed of node	~ 10 min	10^{-9} sec
# inputs/node	1 ~ 10	1 ~ 2
network complexity	combinatorial signal integration from complex transcription control	iterated cascades from complex network wiring

Gene Regulatory Network





combinatorial integration and control

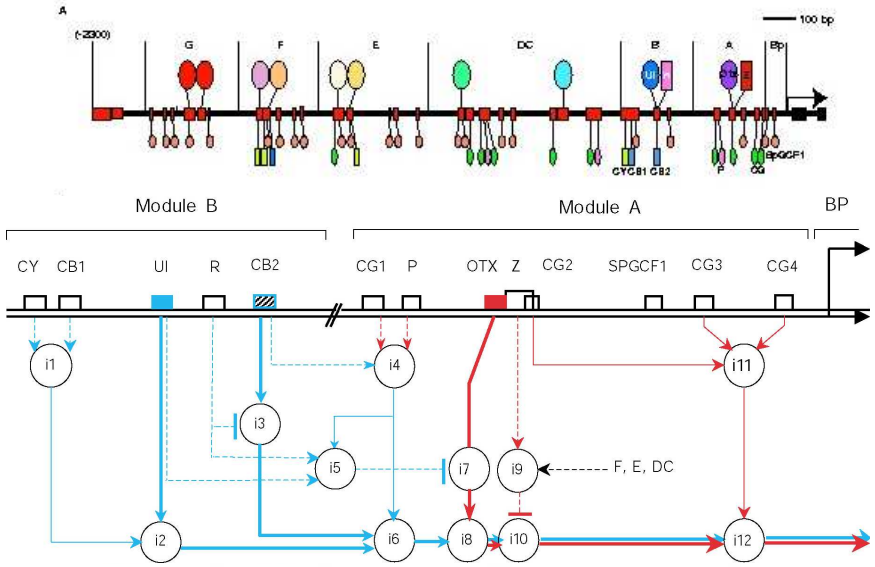
- same gene can exert different effects on different genes
- different combination of gene products
  work together to regulate the same gene

potential and limitations?

➔ focus on transcription control at a single node

Complex transcription control at a node

[Davidson et al]



Molecular basis of complex transcription control

- complex protein-protein interactions
  (**combinatorial** control ???)
- alternative: **regulated recruitment** [Ptashne & Gann '97]
  - glue-like interaction between TFs/RNAP
  - arrange DNA binding sites/strengths to accomplish desired control functions
  (what are possible? how?)

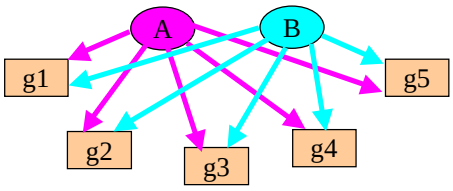
This study:

- take regulated recruitment as starting point
- include thermodynamics of TF/DNA/RNAP interaction
- implement control functions of increasing complexity

Results:

→ a wide class of complex control functions implementable
→ recipe for selecting the strengths and positions of regulatory seq's
→ transcription control system = programmable molecular computer

Examples of combinatorial control



	AND	OR	NAND	XOR	EQ
A/B	g1	g2	g3	g4	g5
lo/lo	OFF	OFF	ON	OFF	ON
lo/hi	OFF	ON	ON	ON	OFF
hi/lo	OFF	ON	ON	ON	OFF
hi/hi	ON	ON	OFF	OFF	ON

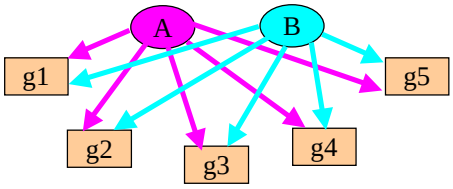
simple control fn

"complex" control fn

cellular conc. of activated TFs
low: [TF] ~ 1 nM
high: [TF] ~ 1000 nM
via phosphorylation, inducer-binding, …

What logic functions are implementable given the toolkit?

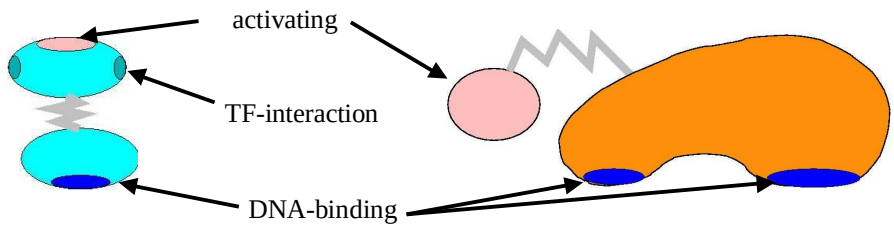
Examples of combinatorial control



	AND	OR	NAND	XOR	EQ
A/B	g1	g2	g3	g4	g5
lo/lo	OFF	OFF	ON	OFF	ON
lo/hi	OFF	ON	ON	ON	OFF
hi/lo	OFF	ON	ON	ON	OFF
hi/hi	ON	ON	OFF	OFF	ON

➔ bacterial transcription machinery sufficient Lego set?

Transcription Factor RNA polymerase complex

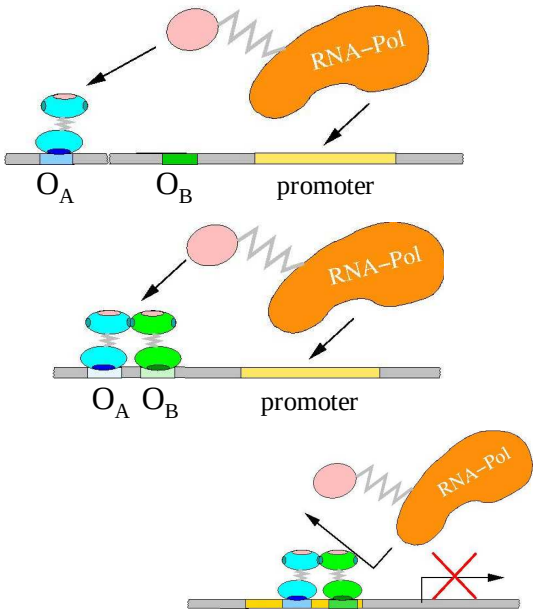


Qualitative constructs of simple control functions

A/B	**OR**
lo/lo	OFF
lo/hi	ON
hi/lo	ON
hi/hi	ON

A/B	**AND**
lo/lo	OFF
lo/hi	OFF
hi/lo	OFF
hi/hi	ON

A/B	**NAND**
lo/lo	ON
lo/hi	ON
hi/lo	ON
hi/hi	OFF



More complex control functions, e.g., **XOR** ?

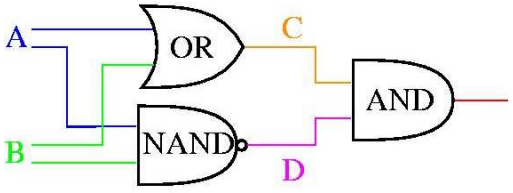
A/B	**XOR**	**OR**	**NAND**
lo/lo	OFF	OFF	ON
lo/hi	ON	ON	ON
hi/lo	ON	ON	ON
hi/hi	OFF	ON	OFF

cannot be implemented by overlapping A and B sites

[cf: linear perceptron (Minsky '69)]

XOR(A,B) = (A **OR** B) **AND** $\overline{(A \textbf{ AND } B)}$

Gene cascade



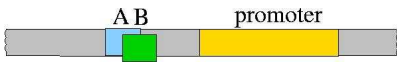
problems:

- need a gene for each intermediate result
- multiple rounds of gene expression: noise + delay
- synchronization difficult

More complex control functions, e.g., **XOR** ?

A/B	**XOR**	**OR**	**NAND**
lo/lo	OFF	OFF	ON
lo/hi	ON	ON	ON
hi/lo	ON	ON	ON
hi/hi	OFF	ON	OFF

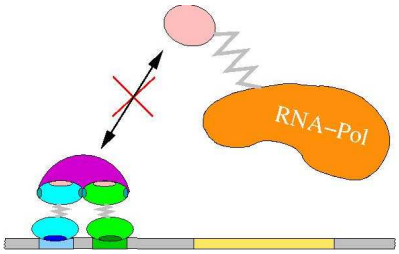
cannot be implemented by overlapping A and B sites



[cf: linear perceptron (Minsky '69)]

XOR(A,B) = (A **OR** B) **AND** $\overline{(A \textbf{ AND } B)}$

Allosteric or cofactor-mediated:

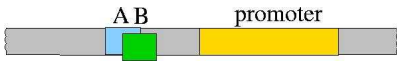


problem:
- lose combinatorial control
  e.g., can't implement **AND** elsewhere

More complex control functions, e.g., **XOR** ?

A/B	**XOR**	**OR**	**NAND**
lo/lo	OFF	OFF	ON
lo/hi	ON	ON	ON
hi/lo	ON	ON	ON
hi/hi	OFF	ON	OFF

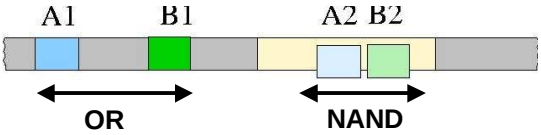
cannot be implemented by overlapping A and B sites



[cf: linear perceptron (Minsky '69)]

XOR(A,B) = (A **OR** B) **AND** $\overline{(A \textbf{ AND } B)}$

Regulated recruitment:

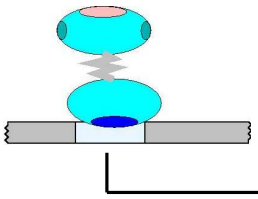


- integrates **OR** and **NAND** into a single regulatory region
- modular and evolvable
- but requires fine balance between activating/repressive effects

  → need quantitative characterization

Quantitative modeling:

- specific protein-DNA binding:



binding probability: $p = \frac{n}{n + \tilde{n}}$

n: TF conc. [1~1,000/cell]

$\tilde{n}$: half-max binding conc.

can be tuned [1~10,000/cell]

$\tilde{n}$ dependent on { binding sequence S ; TF-DNA binding energy $E(S)$; rest of the genome }

approx form of interaction: $E(S) = \varepsilon \times |S\text{-}S^*|$

mismatches from best binder S^*

discrimination energy ~ 1 kcal/mole

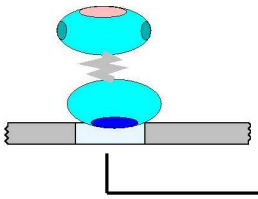
➔ $\tilde{n}$ = 1 ~ 10,000/cell for $|S\text{-}S^*| = 0 \sim 5$ mismatches

[Gerland, Moroz, TH, PNAS '02]

Quantitative modeling:

programmable molecular interaction

- specific protein-DNA binding:



binding probability: $p = \frac{n}{n + \tilde{n}}$

n: TF conc. [1~1,000/cell]

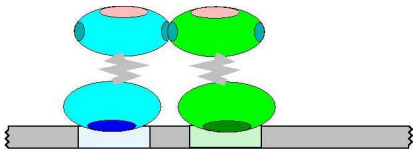
$\tilde{n}$: half-max binding conc.

can be tuned [1~10,000/cell]

- protein-protein interaction: tunable via site placements

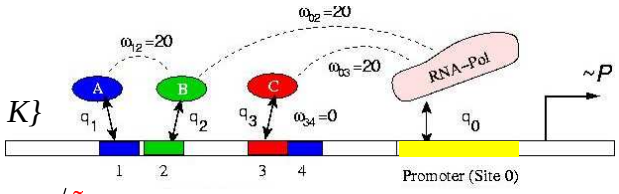
quantify by coop. factor ω

no interaction	$\omega = 1$
repulsive	$\omega = 0$
cooperative	$\omega = 20$



[longer-ranged for coop. TF-RNAP interaction]

Integration:



- binding sites: $j=\{1, \ldots, K\}$
- target TF's: $\alpha(j)$
- binding site affinity: $n_{\alpha(j)}/\tilde{n}_j \equiv q_j$
- promoter affinity: $n_P/\tilde{n}_P \equiv q_0$
- TF-TF and TF-RNAP interaction: $\omega_{i,j} = \{ 0, 1, 20 \}$
- "expression level" ~ equilibrium promoter occupation prob P

Let $\sigma_j=\{0,1\}$ represent the occupation state of each site j

→ weight for TF binding: $W[\sigma_1,\ldots,\sigma_K] = \prod_{j=1}^{K} q_j^{\sigma_j} \prod_{i<j} \omega_{i,j}^{\sigma_i\sigma_j}$

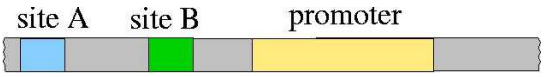
→ weight for RNAP binding: $Q[\sigma_1,\ldots,\sigma_K] = q_0 \prod_{j=1}^{K} \left[1-\sigma_j\,\delta(\omega_{0,j},0)\right] \left[1+\omega \sum_{j=1}^{K} \sigma_j\,\delta(\omega_{0,j},\omega)\right]$

➡ $P\{n_\alpha,\tilde{n}_j,\omega_{i,j}\} = \dfrac{\langle Q\rangle_W}{1+\langle Q\rangle_W}$

Task: find $\{\tilde{n}_j,\omega_{i,j}\}$ to implement the desired $P\{n\}$

Response characteristics

OR gate:



-- weak promoter

$n_P/\tilde{n}_P = 0.05$

-- strong A and B sites

$\tilde{n}_A = \tilde{n}_B = 100$

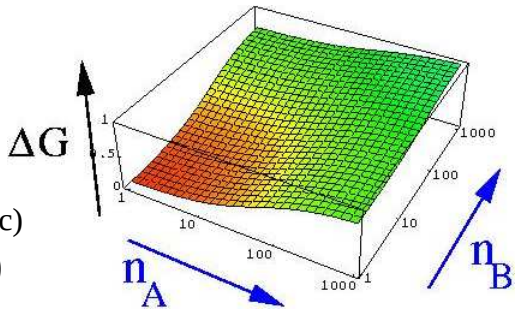
-- TF-RNAP interaction
coop (but not synergistic)

$\omega_{A-P} = \omega_{B-P} = 20$

-- no TF-TF interaction

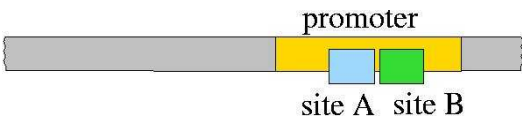
$\omega_{A-B} = 1$

→ Polymerase binding prob: $P(n_A,n_B)$



fold change: $\Delta G = 8 \sim 9$

NAND gate:



-- strong promoter

$n_P / \tilde{n}_P = 100$

-- strong A and B sites

$\tilde{n}_A = \tilde{n}_B = 150$

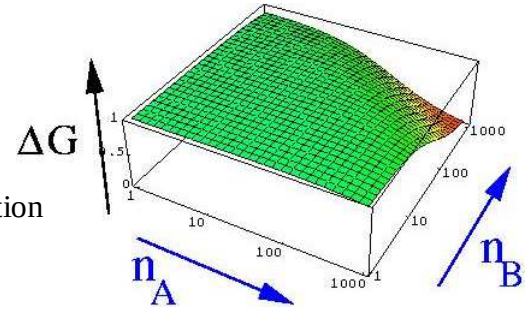
-- repress. TF-RNAP interaction

$\omega_{A-P} = \omega_{B-P} = 0$

-- coop. TF-TF interaction

$\omega_{A-B} = 20$

→ Polymerase binding prob: $P(n_A, n_B)$

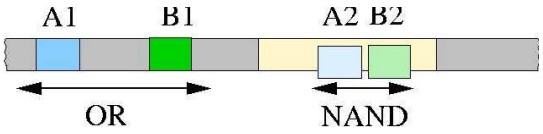


fold change: ΔG = 12

→ used as two hybrid scheme to detect interaction

[Schnarr et al, '98]

XOR from multiple binding sites



-- weak promoter

$n_P / \tilde{n}_P = 0.1$

-- strong sites

$\tilde{n}_{A1} = \tilde{n}_{B1} = 200$

-- weak sites

$\tilde{n}_{A2} = \tilde{n}_{B2} = 900$

-- TF-RNAP interaction

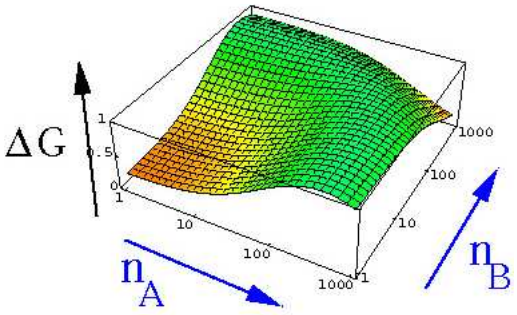
$\omega_{A1-P} = \omega_{B1-P} = 20$

$\omega_{A2-P} = \omega_{B2-P} = 0$

-- TF-TF interaction

$\omega_{A1-B1} = 1$

$\omega_{A2-B2} = 20$



fold change: ΔG = 4 ~ 5

XOR from multiple promoters

-- weak promoter

$n_P / \tilde{n}_{P1} = n_P / \tilde{n}_{P2} = 0.05$

-- weak sites

$\tilde{n}_{A1} = \tilde{n}_{B2} = 500$

-- strong sites

$\tilde{n}_{A2} = \tilde{n}_{B1} = 100$

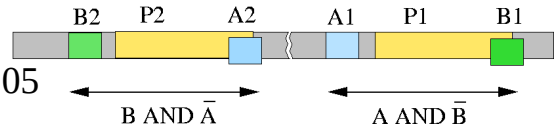
-- TF-RNAP interaction

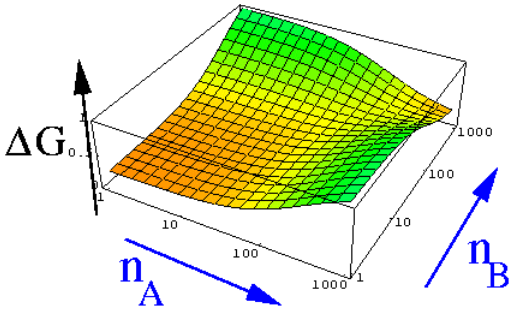
$\omega_{A1-P1} = \omega_{B2-P2} = 20$

$\omega_{A2-P2} = \omega_{B1-P1} = 0$

-- TF-TF interaction

$\omega_{A-B} = 1$





fold change: $\Delta G = 4 \sim 5$

EQ gate?

A/B	**EQ**
lo/lo	ON
lo/hi	OFF
hi/lo	OFF
hi/hi	ON

- strong promoter
- need **multiple** ways of repression

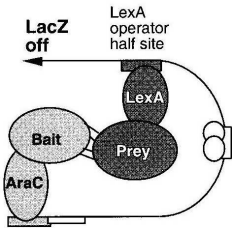
A (low) B (high) or A (high) B (low)

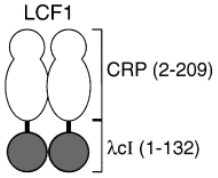
→ competitive promoter binding awkward due to limited promoter size

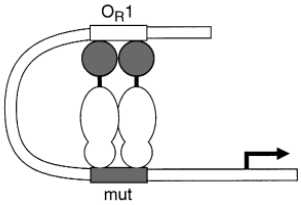
→ possible solution: interaction at a distance

e.g., DNA-looping via dimers (AraC, GalR, MelR,…)

heterodimers: [A. Hochschild et al]

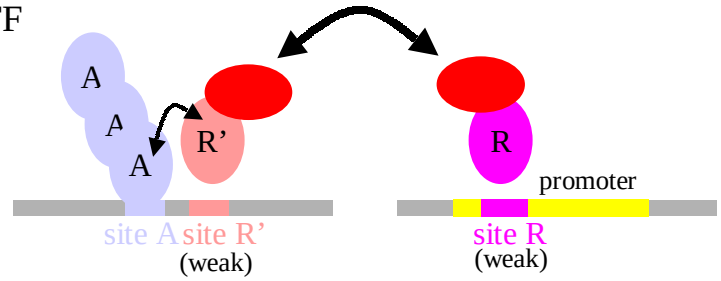




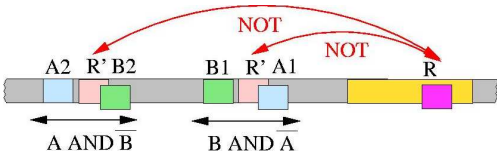


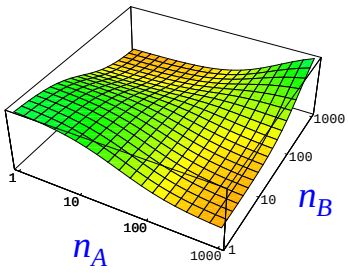
e.g., distal repression by heterodimer pair R and R'

control by TF



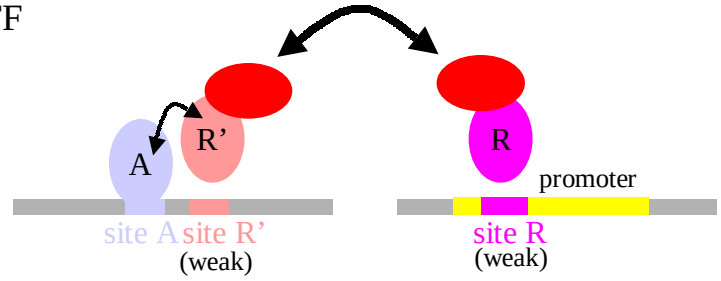
Implementation of **EQ** gate: Effective cascade w/o new genes!



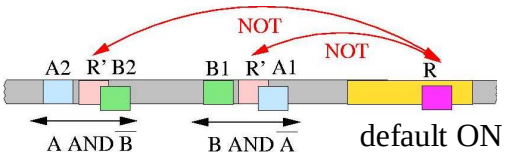


e.g., distal repression by heterodimer pair R and R'

control by TF

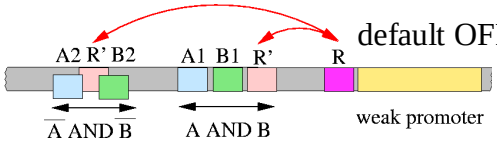


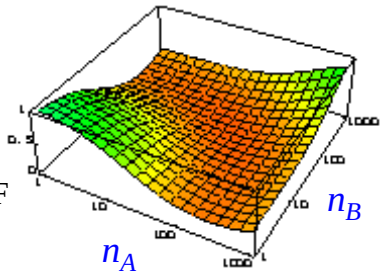
Implementation of **EQ** gate Effective cascade w/o new genes!



default ON

default OFF





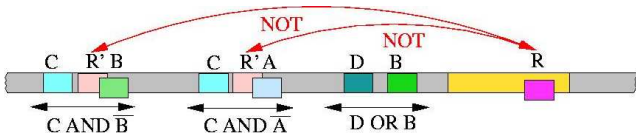
Specific vs. generic interaction

- so far, only discussed integration of signals carried by two TF's (A & B)
  - combinatorial integration from choice and placement of DNA binding sites
- same constructs can be used to control genes with different pairs of TF's (e.g., A & C, B & C) if similar TF-TF interaction exist
  - generic interaction → combinatorial control
- full power of combinatorial integration/control may not be necessary

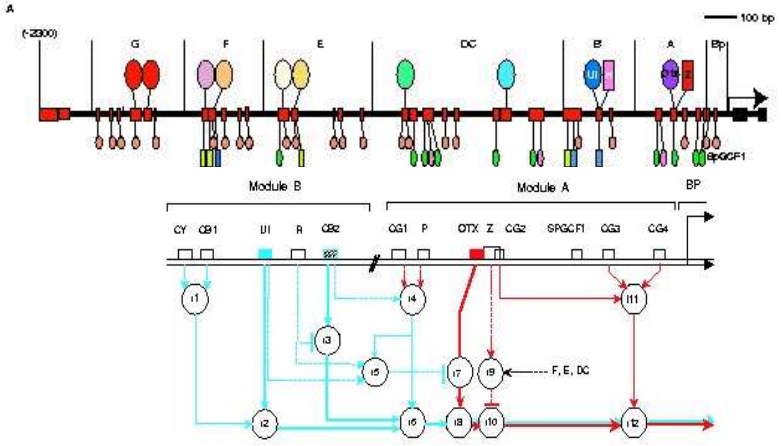
→ expect unnecessary interaction to decay away

→ remaining interaction strengthened if needed ("specific")

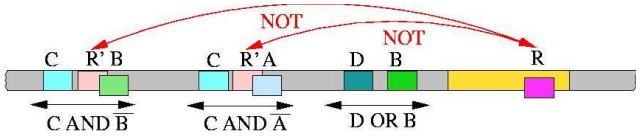
Generalized control architecture w/ multiple TF's



→ naturally modular [cf: Davidson]



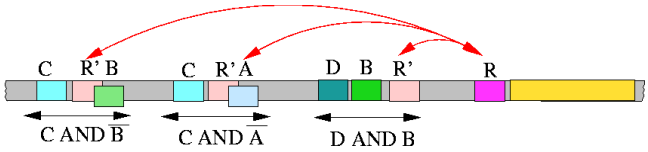
Generalized control architecture w/ multiple TF's



Output = […**OR**…**OR**..] **& NOT**[…**&**…**&**..] **& NOT**[…**&**…**&**..]

phenotype: dominant repression

Conjunctive Normal Form



Output = […**&**…**&**..] **OR** […**&**…**&**..] **OR** […**&**…**&**..]

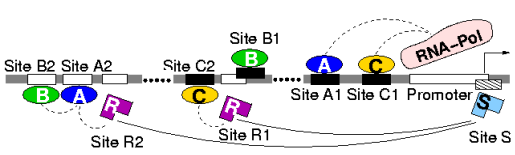
phenotype: enhancer autonomy

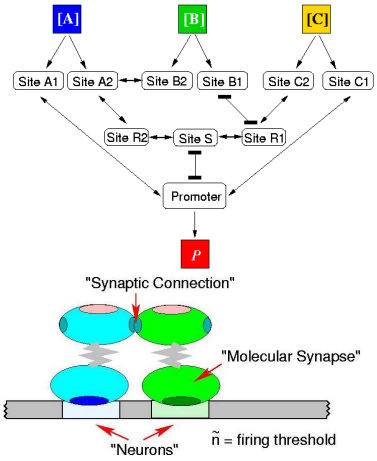
Disjunctive Normal Form

→ all logic functions reducible to minimal CNF or DNF

→ recipe for constructing arbitrary regulatory logic functions (programmable molecular computer!)

Molecular computer ⟷ "Neural network"





TF concentration (n)	input
TF binding sites (j)	neurons
promoter activity (P)	output
binding strength ($\tilde{n}$)	firing threshold
TF-TF interaction ($\omega_{i,j}$)	synapse

- symmetric interaction: "recurrent network"
- heterodimer sites: "hidden units"

➔ **"Molecular Boltzmann machine"**

"learning" = mutation + selection

Ingredients for complex transcription control

- programmable protein-DNA interaction
- weak, glue-like interaction between nearby proteins
- long-distance activation/repression

→ already possible w/ bacteria transcription apparatus
[experimental effort to synthesize logic gates]

→ but only simple controls found/characterized (so far)

not needed? or other limitations?

-- Promiscuity of glue-like interaction? Okay for weak interaction

-- Inter-genic cross talk? Big problem for bacteria
necessary evil due to long-distance interaction

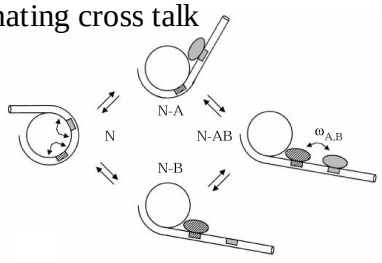
→ **complex transcription control in bacteria only possible for isolated genes (separated far apart or on plasmids)**

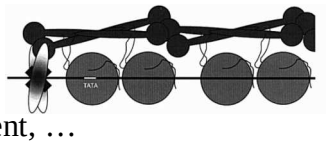
Ingredients for complex transcription control

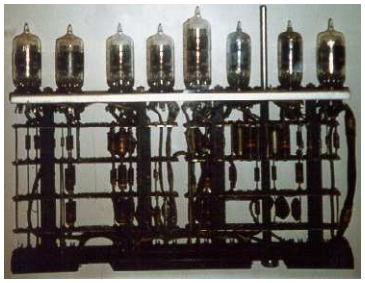
- programmable protein-DNA interaction
- weak, glue-like interaction between nearby proteins
- long-distance activation/repression

Eukaryotes: (merely) a superior implementation platform

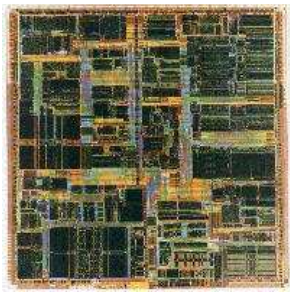
- **insulating elements**: crucial for eliminating cross talk
- generic cooperative interaction mediated by **nucleosomes**
  [Polach & Widom, '96]
  → physical attraction not necessary
- **short-range quenching** [M. Levine]
- distal repression via recruitment of **chromatin modification** agents
- physical sequestation, indep recruitment, …







vacuum tube



pentium chip

Different implementation of the same principle!

Summary and Outlook:

- genetic computing: broad but shallow
- cis-regulatory system: programmable molecular network (molecular Boltzmann machine!)
  - modular organization (CNF or DNF)
  - recipe for coding cis-regulatory logics

  → evolution of complex control functions
- incorporate feedback

  → simple constructs of memory, counter, differentiator, …
- ingredients for complex transcription control:
  - specific protein-DNA interaction
  - weak, glue-like protein-protein interaction
  - long distance activation/repression
  - insulation of inter-genic, inter-module cross talk
- *phenomenological* model of combinatorial gene regulation

  → relate binding site information to gene expression patterns

Collaborators:
-- Nicolas Buchler (UCSD)
-- Ulrich Gerland (UCSD)

Thanks to:
Antoine Danchin (Pasteur)
Eric Davidson (Caltech)
Bill Loomis (UCSD)
Andrew Murray (Harvard)
Mark Ptashne (Sloan-Kettering)
John Reinitz (Stony Brook)
Jon Widom (Northwestern)
…
ITP

Support: NSF, Burroughs-Wellcome, DAAD