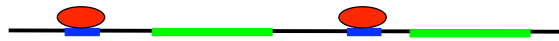


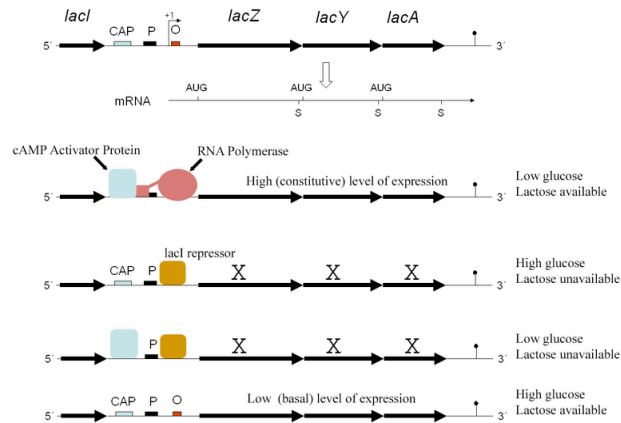
Protein-DNA interactions

Introduction: DNA, genes and regulation



- DNA contains **genes**
- genes need to be turned on and off
- genes are regulated by proteins that bind to specific **regulatory sequences** on DNA.
- proteins that regulate genes are called **transcription factors** (TFs)

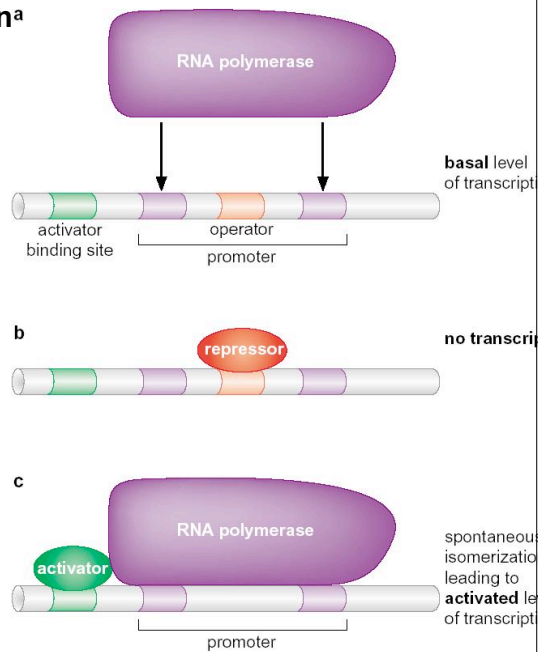
The *lac* Operon and its Control Elements



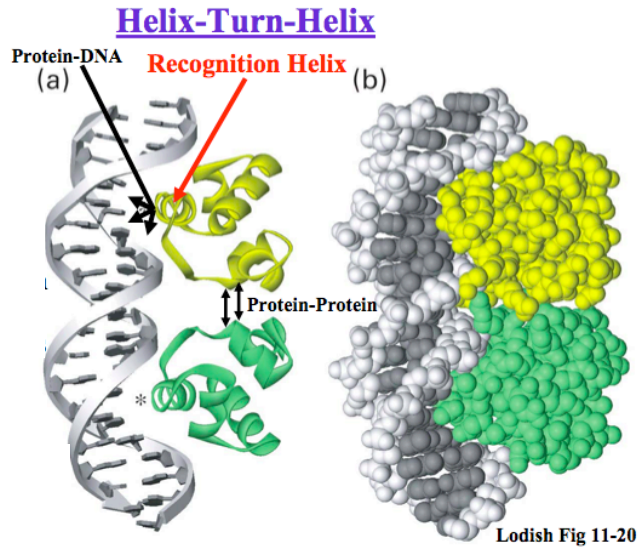
http://en.wikipedia.org/wiki/Lac_operon

Introduction: DNA, genes and regulation^a

- TF can either repress or activate gene expression



Introduction:
DNA, genes and regulation



Introduction:
DNA, genes and regulation

- TFs recognize specific DNA motifs
- protein-DNA complexes are very stable (life time $\sim 10^3$ sec)
- stabilized by interactions between DNA base-pairs and protein amino acids.

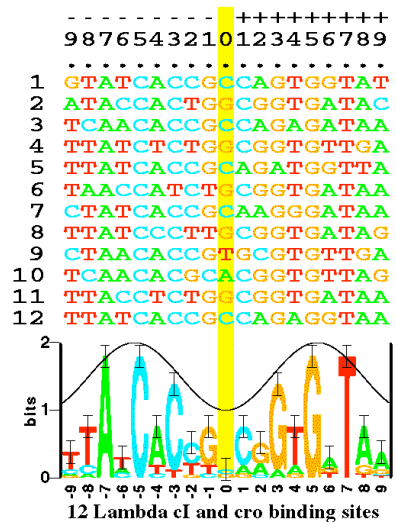
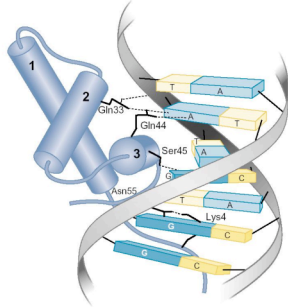


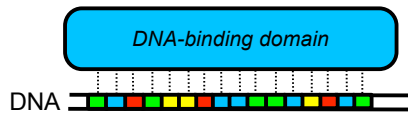
Fig. 1. Some aligned sequences and their sequence logo. At the top of the figure are listed the 12 DNA sequences from the λ - and λ -cro control regions in bacteriophage lambda. These are bound to

Protein-DNA interaction energy



- Specific (sequence-dependent) and non-specific interactions
- Specificity is achieved by formation of local interactions with DNA

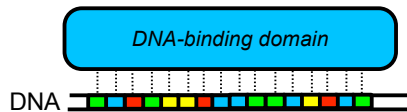
Model



Assumption:
Independent contributions of individual base-pairs.

Protein-DNA interaction energy

Model



$$E = \sum_{i=1}^l e(i, b_i)$$

Example: *Mnt* protein weight matrix
(Stormo & Fields, TIBS 23, p. 109, 1998)

	11	12	13	14	15	16	17	18	19
A	0.26	0.37	0.91	0.31	0.058	-0.29	0.55	0.14	0.39
C	-0.16	0.87	0.58	1.90	0.74	-0.11	-0.72	-0.33	-0.33
G	-0.16	-0.68	0.36	-0.68	-0.56	0.45	1.30	0.30	0.45
T	0.26	1.20	-0.63	0.68	0.85	0.13	0.74	0.11	-0.16

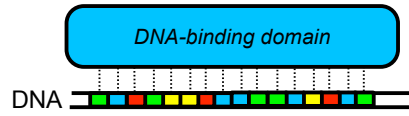
(a)

	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
a	g	g	t	c	c	a	c	g	g	t	g	g	a	c	c	t	

(b)

Protein-DNA interaction energy

Model



$$E = \sum_{i=1}^l e(i, b_i)$$

CGATGGACTT
 CCATGGCCCT
 GGATGGGCTT
 GGATGGTCCT
 GGATGCACCT

$$e(i, x) = -\beta \log(f(i, x) / g(x))$$

	11	12	13	14	15	16	17	18	19
A	0.26	0.37	0.91	0.31	0.058	-0.29	0.55	0.14	0.39
C	-0.16	0.87	0.58	1.90	0.74	-0.11	-0.72	-0.33	-0.33
G	-0.16	-0.68	0.36	-0.68	-0.56	0.45	1.30	0.30	0.45
T	0.26	1.20	-0.63	0.68	0.85	0.13	0.74	0.11	-0.16

	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
a	g	g	t	c	c	a	c		g	g	t	g	g	a	c	c	t

GIVEN THIS ENERGETICS, CAN A PROTEIN FIND ITS SITE?