

Topic 1: Protein-DNA Interaction

- Goals:
 - find DNA binding target seq for each transcription factor (TF)
 - find the affinity of a TF to its DNA target as a function of its cellular concentration *in vivo*
 - find how the TF-DNA affinity may be modulated by the target sequence
- Problems:
 - thousands of TFs each with distinct target sequence; only a few characterized in detail experimentally
 - *ab initio* molecular calculation difficult even when TF-DNA co-crystal structure available
 - need to deal with the entire genomic DNA seqs *in vivo*

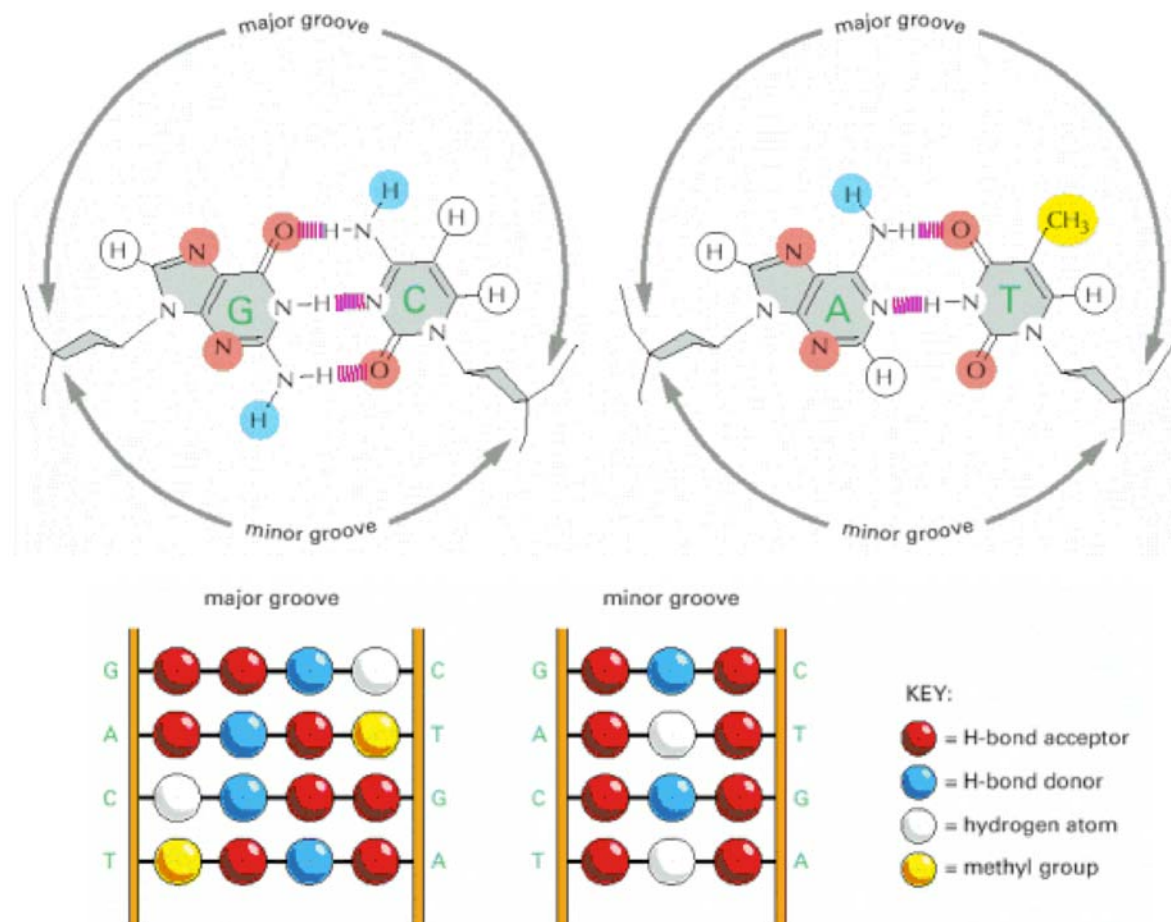
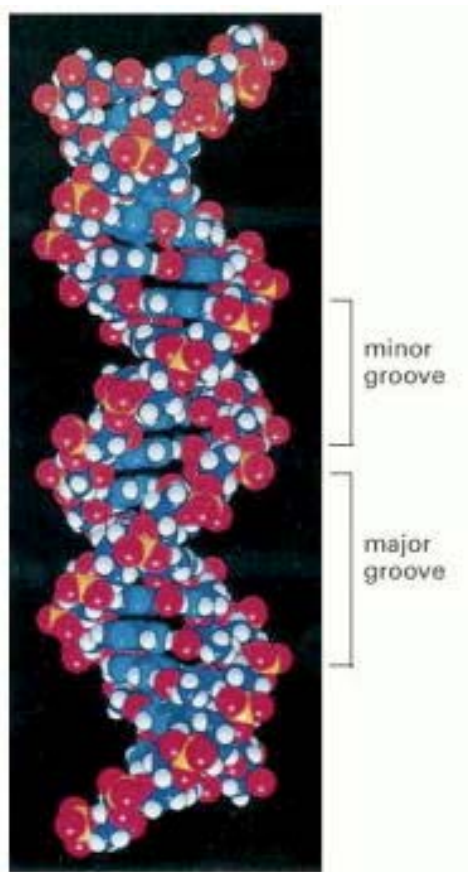
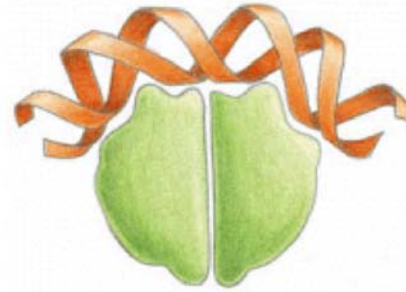
Statistical physics:

- ➔ ways to think quantitatively about TF-DNA interaction in the absence of detailed microscopic information
- ➔ link from molecule to function

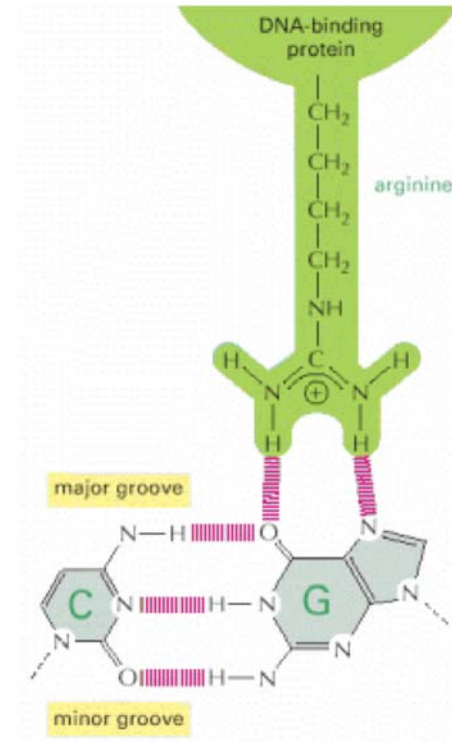
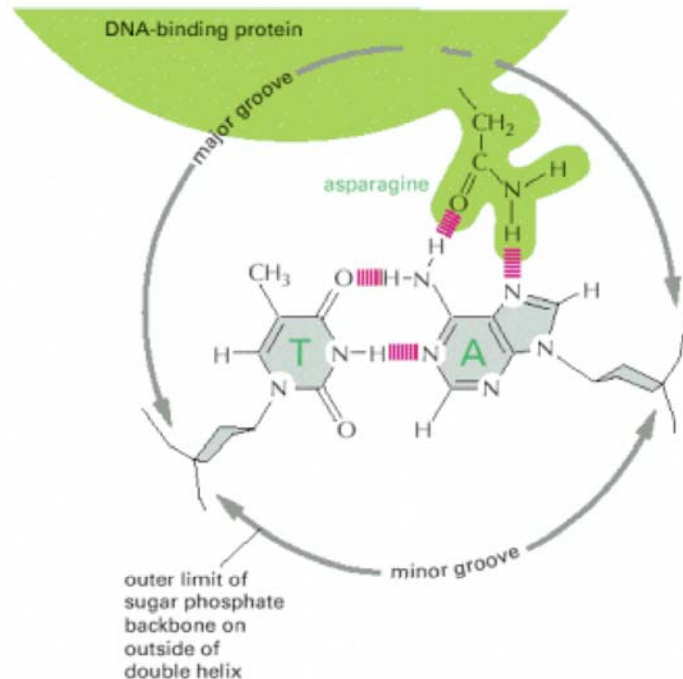
A. Empirical facts

1. Transcription Factors

- size: ~5nm (10-20 bp)
- molecular basis of sequence recognition



- contact between TF and DNA

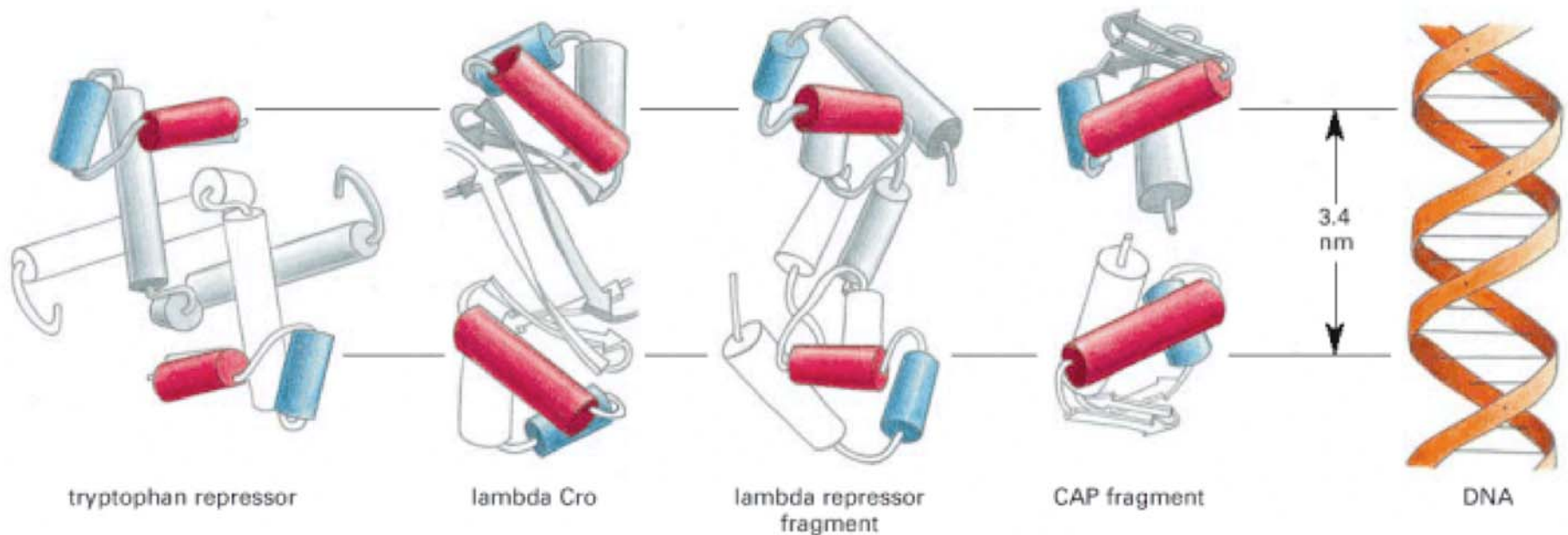


- ➔ structure of a TF must place the appropriate amino acids next to the base pairs they contact

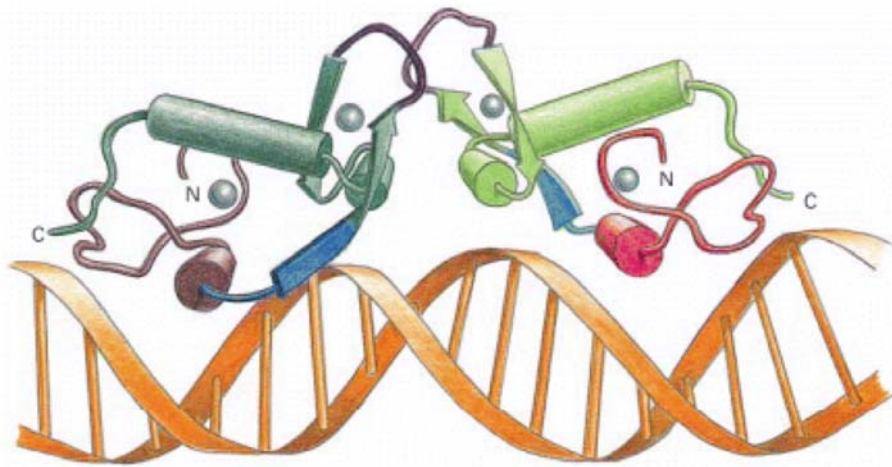
- various molecular strategies
 - Helix-Turn-Helix



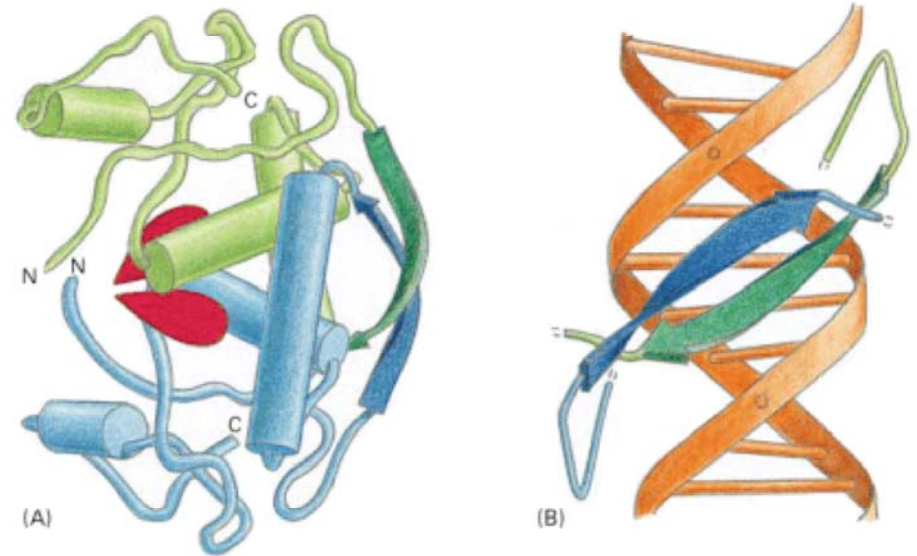
well-known examples in bacteria (note: homodimers)



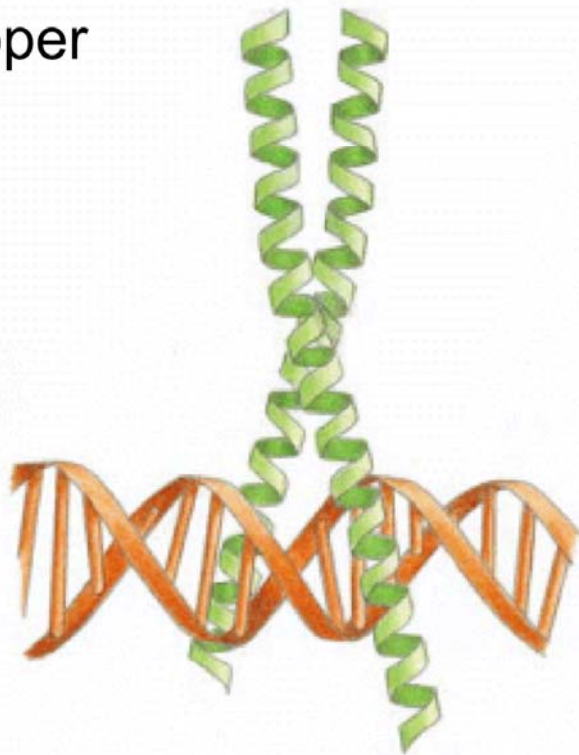
– zinc-finger domain



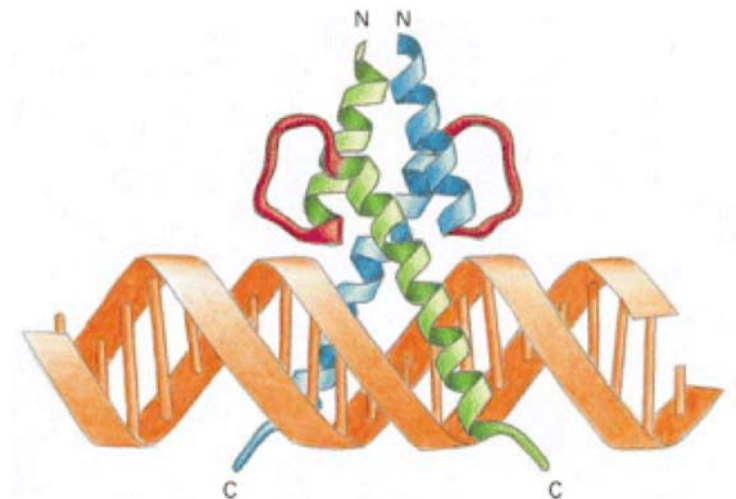
– beta-sheets



– leucine zipper



– helix-loop-helix



2. DNA binding sequences

- typically 10-20 bp in bacteria

protein	target sequence
lac repressor	5' AATTGTGAGCGGATAACAATT 3' TTAACACTCGCCTATTGTTAA
CRP	TGTGAGTTAGCTCACT ACACTCAATCGAGTGA
λ repressor	TATCACCGCCAGAGGTA ATAGTGGCGGTCTCCAT

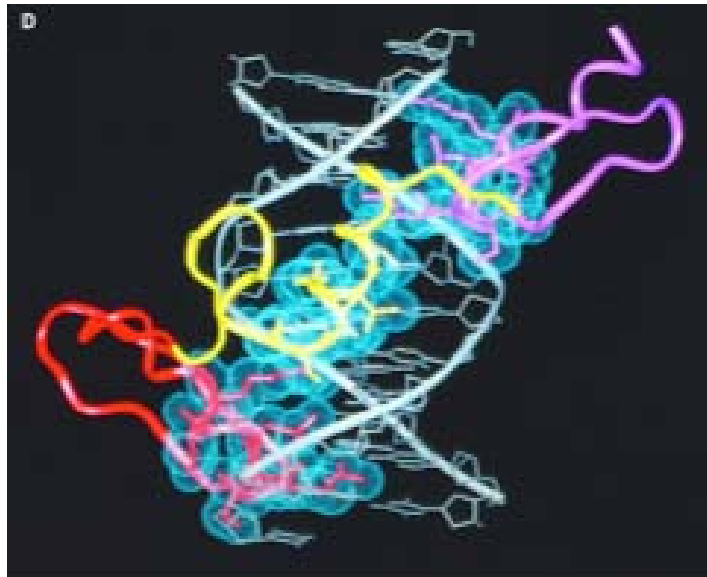
- lots of sequence variants
- consensus sequence** often palindromic
- common to have 2~3 mismatches from the core consensus sequence
-- “fuzzy” binding motif

ATTCTGTAAACAGAGATCACACAAA
CCTTTGTGATCGCTTTCACGGAGC
AAAACGTGATCAACCCTCAATTT
AACTTGTGGATAAAAATCACGGTCT
GTTTTGTTACCTGCCCTCTAACTTT
TTAATTTGAAAATTGGAATATCCA
AATTTGCGATGCGTCGCGCATTTT
TTAATGAGATTTCAGATCACATATA
AATGTGTGCGGCAATTCACATTTA
GAAACGTGATTTTCATGCGTCATTT
AAATGACGCATGAAAATCACGTTTC
TTGCTGTGACTCGATTACGGAAGT
TTTTTGTGGCCTGCTTCAAACTTT
GAATTGTGACACAGTGCAAAATTC
ATAATGTTATACATATCACTCTAA
CGATTGTGATTTCGATTACATTTA
GTTTTGTGATGGCTATTAGAAATT
GAACTGTGAAACGAAACATATTTT
AATGTGTGTAAACGTGAACGCAAT
TTTGTGTGATCTCTGTTACAGAAT
GTAATGTGGAGATGCGCACATAAA
TTTTTGTGAAGCAACATCACGAAAT
TTAATGTGAGTTAGCTCACTCATT
ATTATTTGCACGGCGTCACACTTT
ATTATTTGAACCAGATCGCATTAC
TAATTGTGATGTGTATCGAAGTGT
....TGTGA.....TCACA....

3. TF-DNA interaction

- passive (no energy consumption)
- strong electrostatic attraction indept of binding seq
e.g., $[TF - DNA] > 10 \times [TF]_{free}$ for LacI in 0.1M salt
→ non-specific binding: $G_{ns} - G_{cyto} \simeq -15RT$
($RT \approx 0.62$ kcal/mole at 37C)
- additional energy gained from hydrogen bonds to **preferred** sequences
strongest binder: $G^* - G_{ns} \simeq -15RT$
- graded increase in binding energy for sequences with partial match to the preferred sequence

- relative binding affinity for Mnt



Binding energies for Mnt (in $k_B T \equiv 1$):

pos.	10	11	12	13	14	15	16	17
<i>A</i>	1.8	2.4	1.6	1.0	0	2.1	0.8	1.1
<i>C</i>	2.4	1.9	4.2	2.1	0.3	0	0	0
<i>G</i>	0	1.6	0	0	1.2	3.2	1.0	1.2
<i>T</i>	3.0	0	2.2	2.2	0.6	2.2	0.7	0.3

(D.S. Fields, Y. He, A.Y. Al-Uzri & G. Stormo, 1997)

(from competitive binding expts)

- weak energetic preference -- **weak specificity**
- similar results for other TFs studied (e.g., LacI, λ -CI, λ -Cro)
- double mutation: binding energy **approx additive**

- Issues to be addressed here:
 - range of TF-DNA affinity *in vivo*
 - dependence of this affinity on variation in target sequence
 - why weak specificity of TF-DNA interaction?
[“design rule” for TF]
 - why fuzzy motifs
[choice of DNA targets]
- Issues not addressed:
 - what is the target sequence of a given TF
[can be probed experimentally]
 - fluctuations in TF-DNA binding