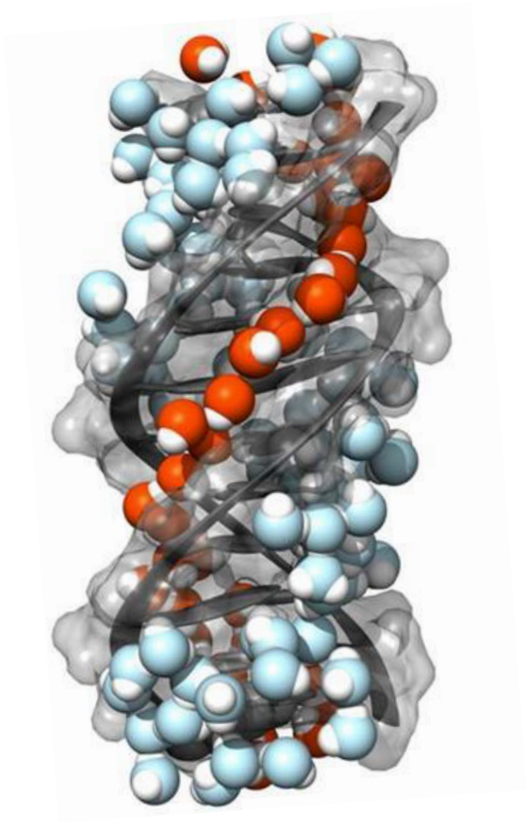
What drives the binding of a protein to DNA ?

Non-covalent interactions!

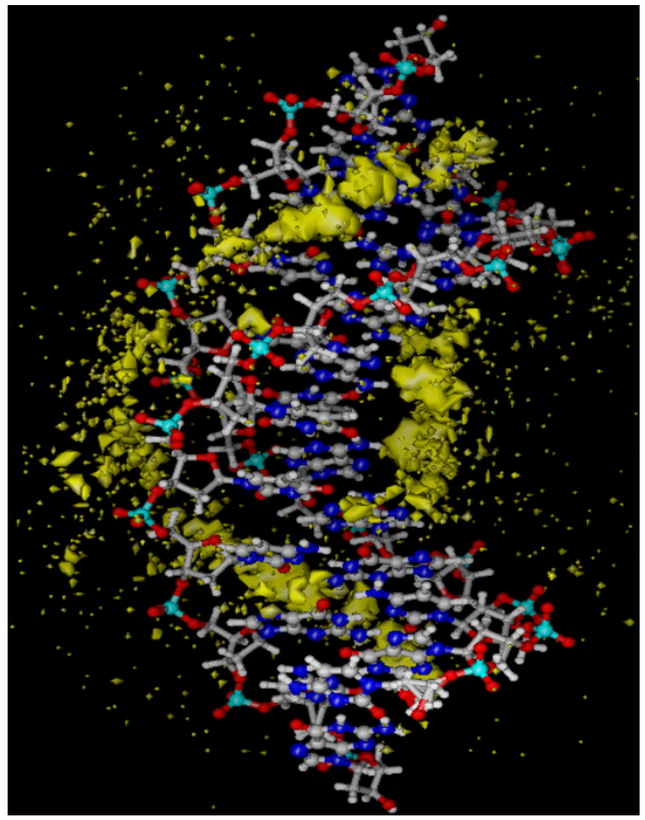
But: Non-covalent interactions can occur between macromolecule and water/ions just as well.

Water molecules (red) at DNA



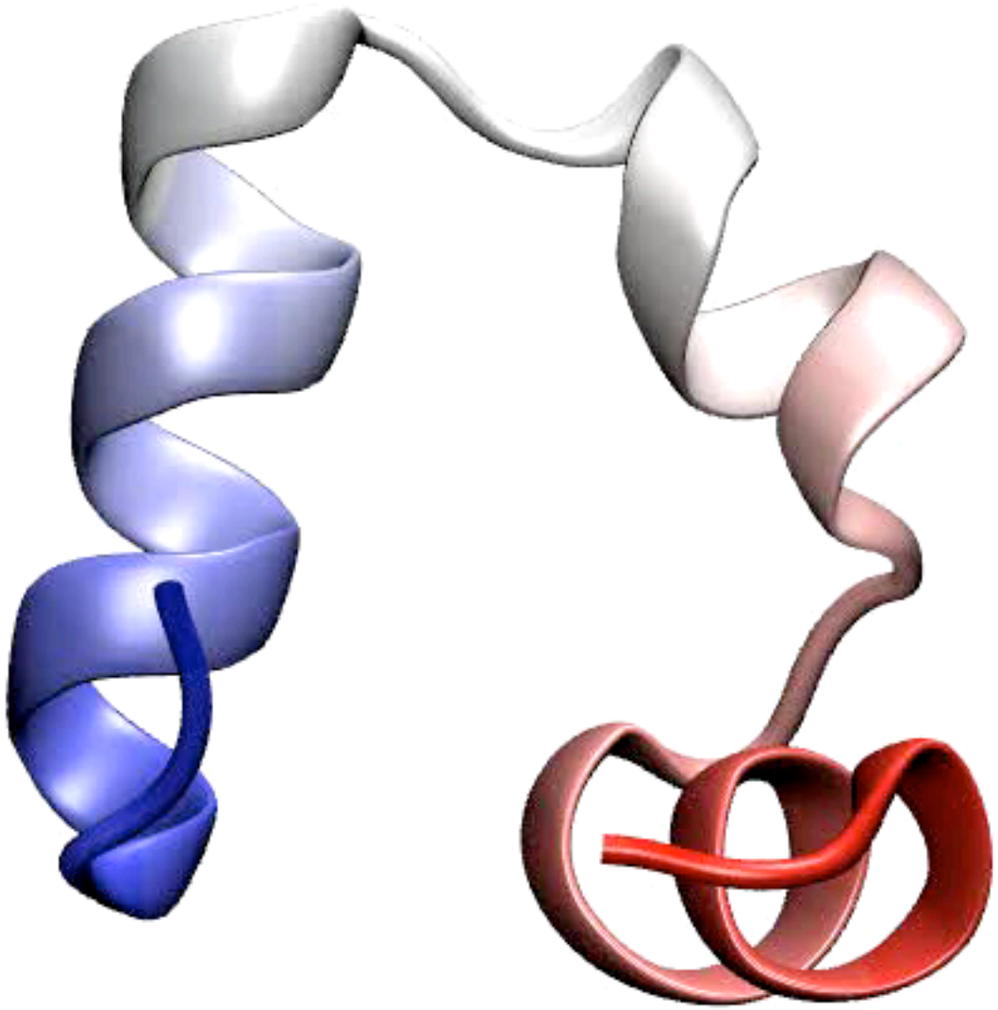
McDermott ACS Centr. Sci., 2017
DOI: 10.1021/acscentsci.7b00100

Regions of high Na^+ (yellow) density



Ponomarev, PNAS 2004,
DOI: 10.1073/pnas.0406435101

What drives the folding of a protein?

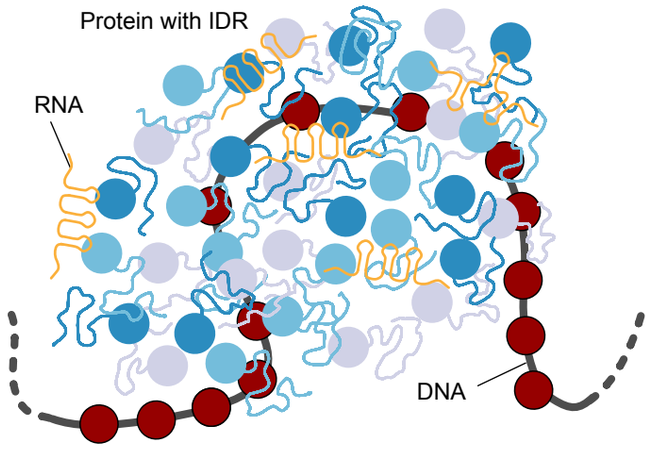


Protein folding as seen in molecular dynamics simulations of the villin protein headpiece (~10 μs time scale)

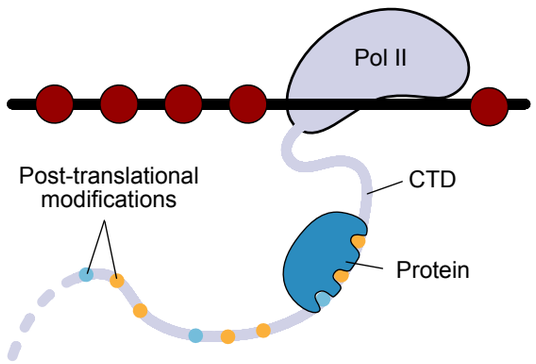
Peter L. Freddolino and Klaus Schulten. *Biophysical Journal*, 97:2338-2347, 2009

Intrinsically disordered protein regions (IDRs) can mediate macromolecular interactions

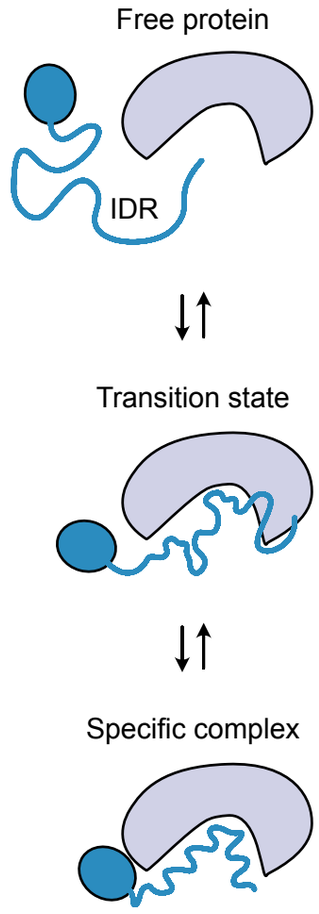
LLPS



Specific interactions with IDR

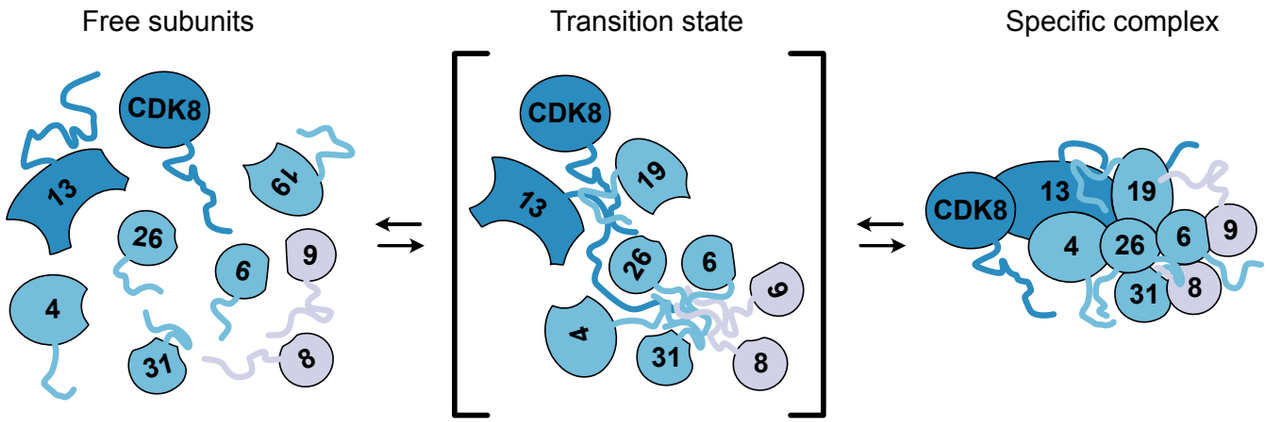


Induced folding interactions



Rippe & Papantonis 2022
Curr Opin Cell Biol

Increasing the kinetic rate for forming a multi-subunit complex



- 25–30% of eukaryotic proteins are predicted to be mostly disordered
- Half of all eukaryotic proteins have long regions (<50 aa) of disorder

Uversky 2021
Ann Rev Biophys

[CONTRIBUTION FROM THE DEPARTMENT OF BIOCHEMISTRY, DUKE UNIVERSITY, MEDICAL CENTER, DURHAM, NORTH CAROLINA]

Contribution of Hydrophobic Interactions to the Stability of the Globular Conformation of Proteins

BY CHARLES TANFORD

RECEIVED APRIL 9, 1962

Tanford, C. *J Am Chem Soc* **84,** 4240–4247 (1962).

Calculating entropy: how probable or disordered is the final state?

Entropy provides that measure (Boltzmann)…

$$S \equiv k_B \ln W$$

- S: Molecular Entropy
- k_B: Boltzmann Constant
- W: Number of microscopic ways in which a particular outcome (macroscopic state) can be attained

For Avogadro number's of molecules…

$$S = \underbrace{(N_{Avogadro} k_B)}_{R \text{ (gas constant)}} \ln W$$

Therefore: the most probable outcome maximizes entropy of <u>isolated systems</u>

Criterion for Spontaneity:

$\Delta S > 0$ (spontaneous)

$\Delta S < 0$ (non-spontaneous)

Unfavorable conformation entropy for protein folding

for the folded state: ~ 1 conformation

$$S_{\text{folded}} = R\ln(1) = 0$$

for the unfolded state: x is the number of flexible points per residue and z is the number or possible orientations of equal energy at each point.

$$S_{\text{unfolded}} = R\ln\left(z^{x}\right)$$

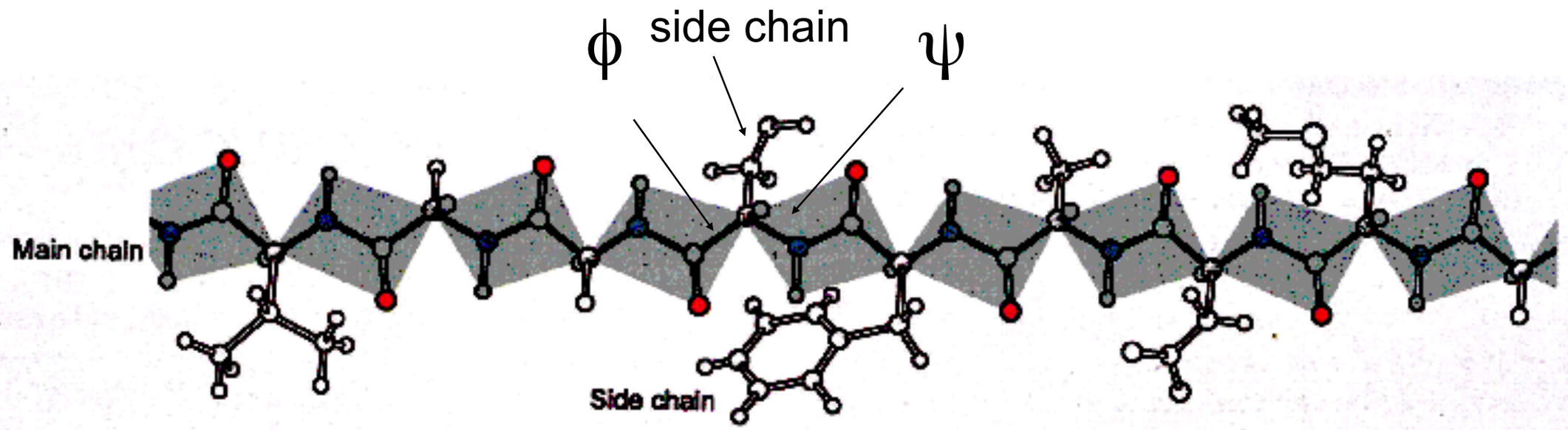
Estimating the unfavorable conformational entropy for protein folding

$$S_{\text{conf}} = R\ln\left(z^{x}\right) \qquad \Delta S_{\text{fold}} = R\ln\left(\frac{W_{\text{unfolded}}}{W_{\text{folded}}}\right)$$

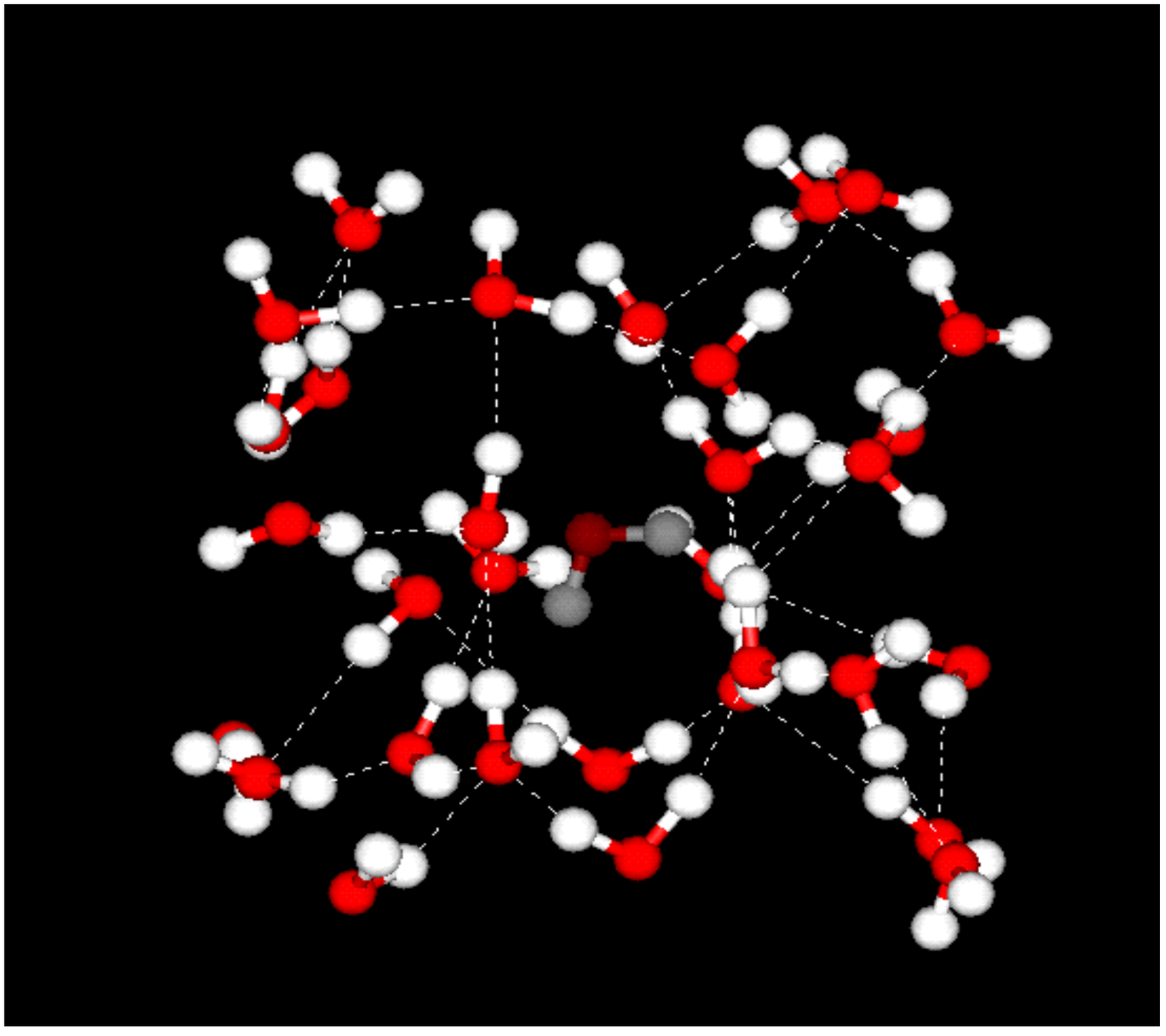
Tanford 1962: For three flexible positions (ϕ , ψ, side chain) with two possible orientations each we have 2^3 conformations per residue:

ΔS_{fold} = 8.31 J·mol^{-1}·K^{-1}·ln (⅛) = 17.3 J·mol^{-1}·K^{-1} or “entropy units” (e.u.).

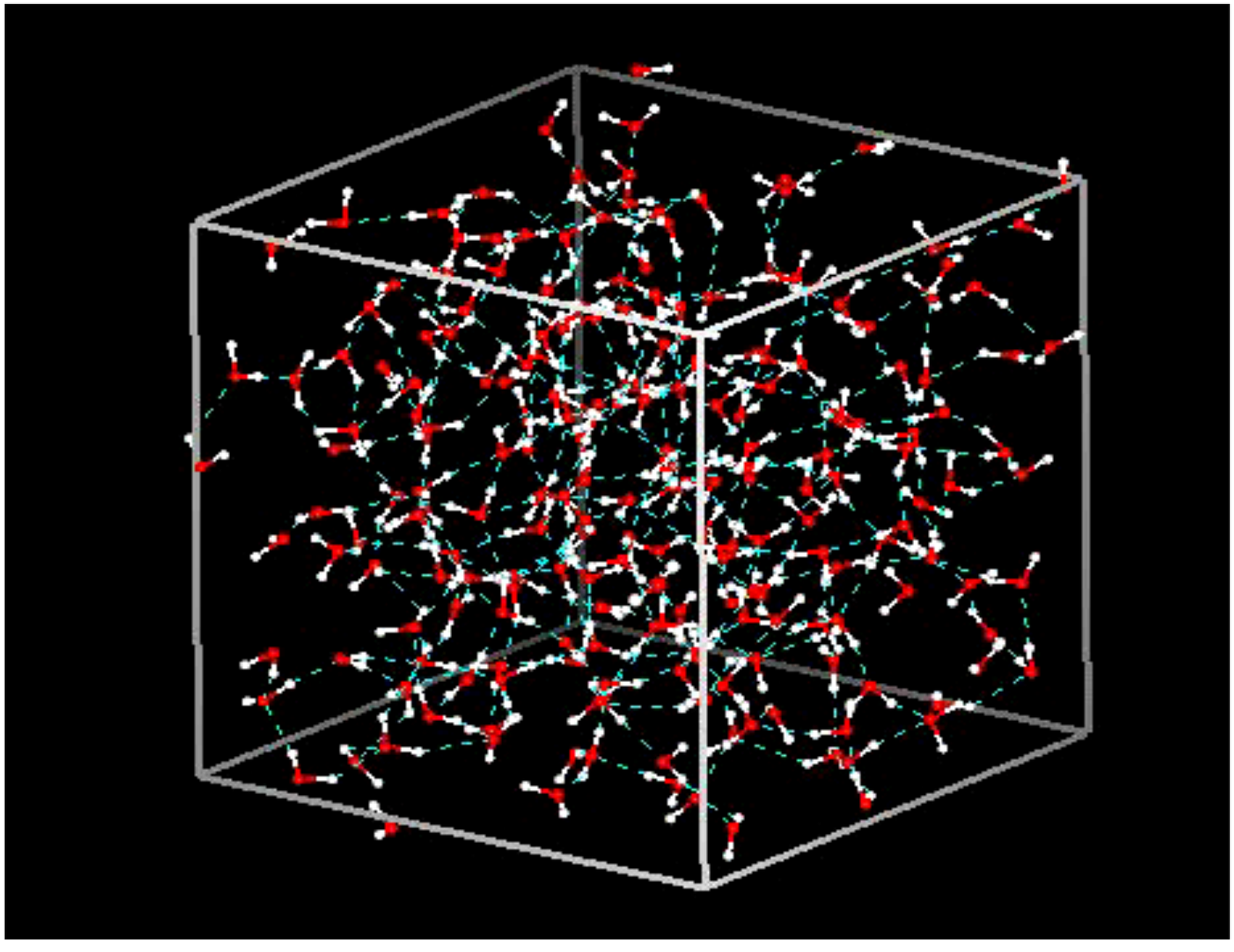
at 25 C: -TΔS =1.2 kcal/mol or 5.2 kJ/mol



Hydrogen bonding of liquid water

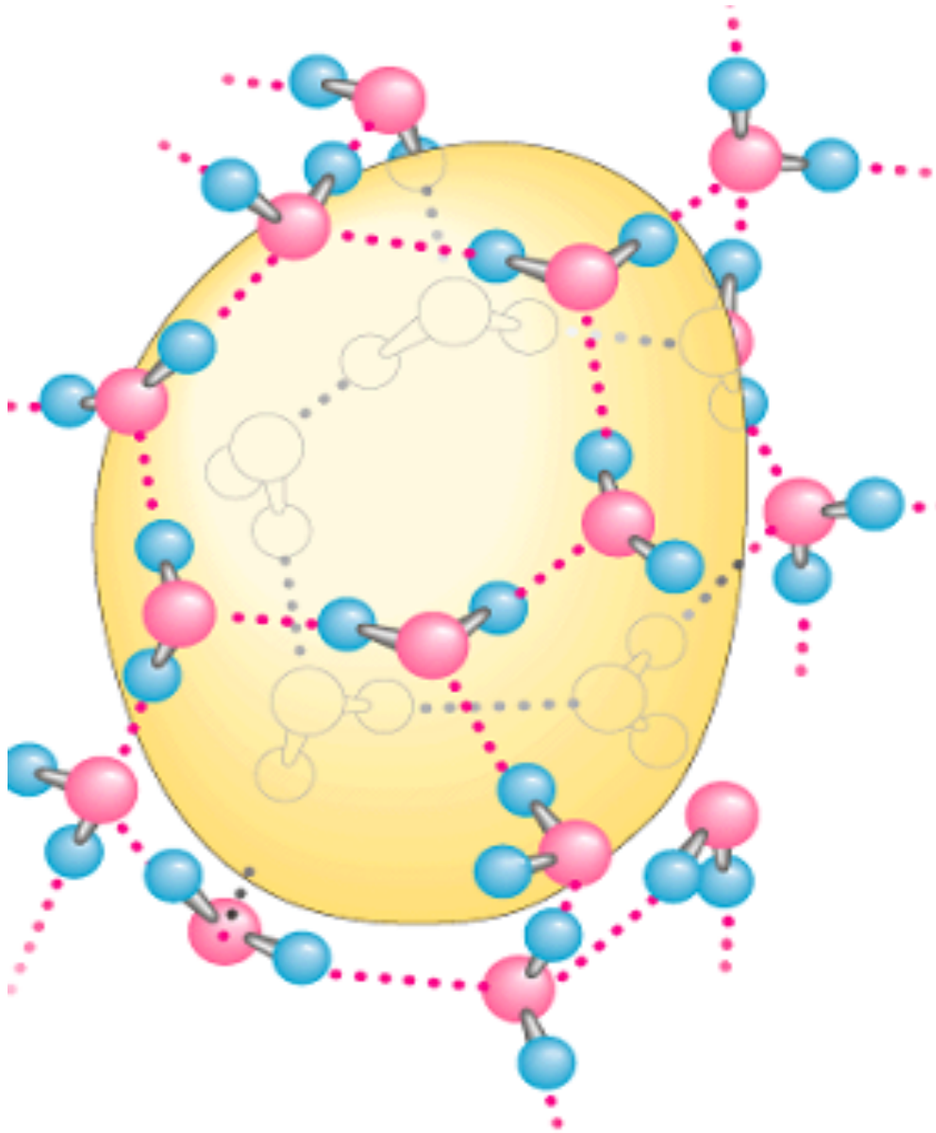


Hydrogen bonding of liquid water



The hydrophobic effect drives protein-protein and protein-DNA/RNA interactions in water

- Minimization of non-polar/water surface area leads to stability
- Complex mixture of physical properties
- Entropic contribution most significant
- Water must form a “cage” structure around non-polar surfaces

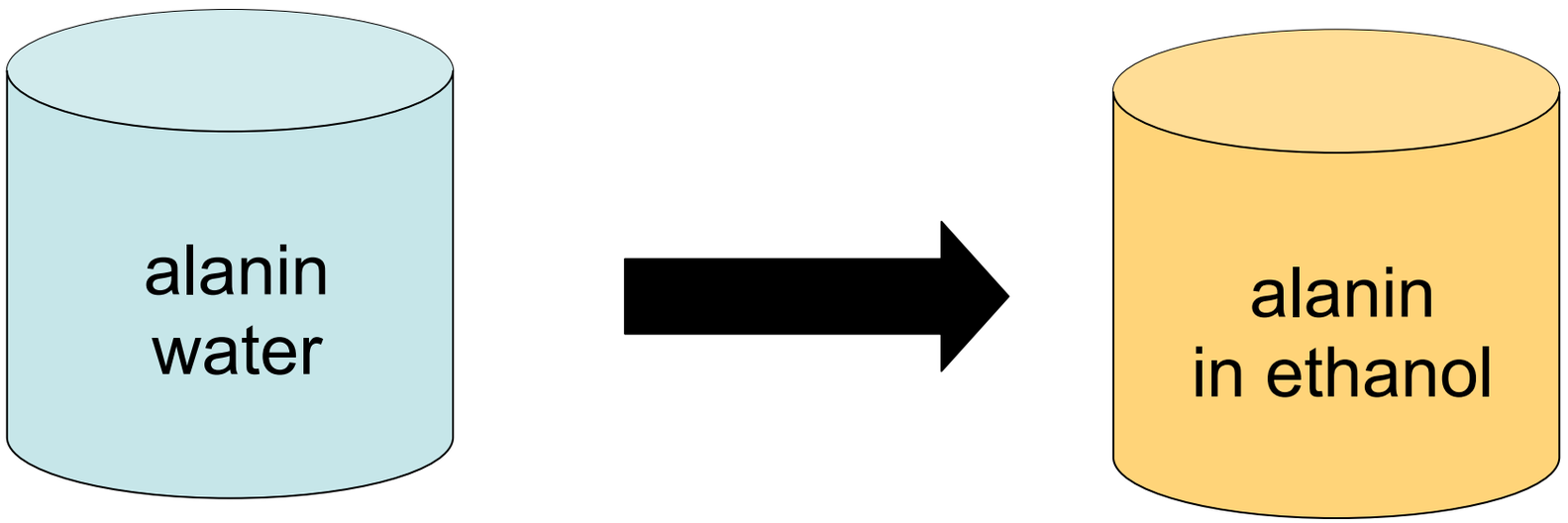


Measure solubility of amino acid in ethanol (= inside the folded protein) and in water (= unfolded state or at the protein surface)

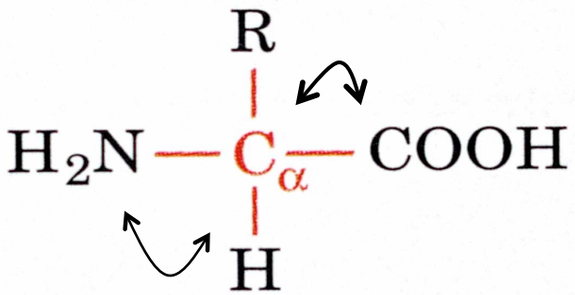
$$\text{partition coefficient } K_{\mathrm{D}} = \frac{\text{solubility } (Alanin_{\mathrm{EtOH}})}{\text{solubility } (Alanin_{\mathrm{H_2O}})}$$

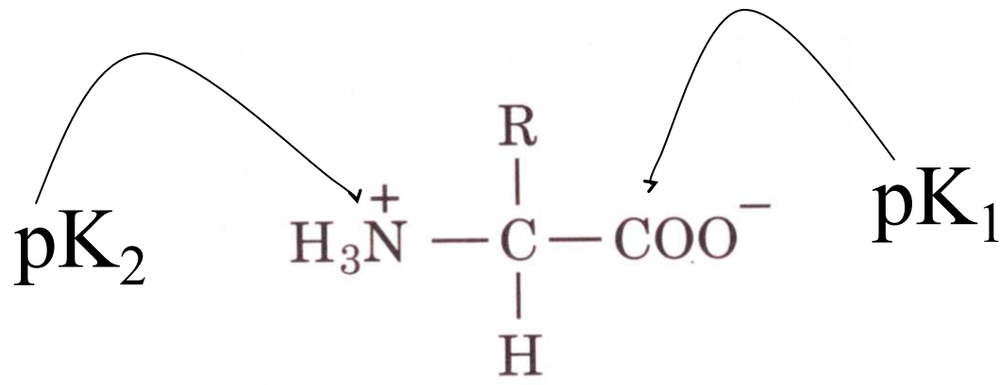
Calculate the free energy from transferring an amino acid from water to ethanol

$$\Delta G_{\mathrm{tr}} = -RT\ln\left(\frac{N_{\mathrm{EtOH}}}{N_{\mathrm{H_2O}}}\right) = -RT\ln\left(K_{\mathrm{D}}\right)$$



Free amino acids carry a positive and a negative charge that is not present in the peptide chain





α amino acids because of the α carboxylic and α amino groups pK_1 and pK_2 respectively pK_R is for R group pK's

$pK_1 \approx 2.2$ while $pK_2 \approx 9.4$

In the physiological pH range, both carboxylic and amino groups are completely ionized

TABLE I[a]

FREE ENERGY CHANGE IN CALORIES PER MOLE FOR TRANSFER FROM ETHANOL TO WATER AT 25°

Tanford 1962

	ΔF_t, whole molecule	Δf_t, side chain contribution	
Non-polar side chains			
Glycine	−4630	0	
Alanine	−3900	+ 730	
Valine	−2940	+1690	
Leucine	−2210	+2420	
Isoleucine	−1690[b]	+2970[b]	
Phenylalanine	−1980	+2650	
Proline	−2060[c]	+2600[c]	
Other side chains			
Methionine	−3330	+1300	
Tyrosine	− 930[d]	+2870[d]	
Threonine	−4190	+ 440	
Serine	−4590	+ 40	
Asparagine	−4640	− 10	
Glutamine	−4730	− 100	
Aspartic acid[e]	−4090	+ 540	uncharged
Glutamic acid[e]	−4080	+ 550	uncharged

Burying a charged amino acid in the interior (Born expression)

$$W_B = \frac{q^2}{4\pi\varepsilon_0 r}\left(\frac{1}{\varepsilon_1} - \frac{1}{\varepsilon_2}\right)$$

W_B is the free energy of transfer in moving a charged body from a region with a relative dielectric constant ε_2 to a medium with a with a relative dielectric constant ε_1. The parameter r is the radius of the charge.

q (charge of an electron) = 1.60 x 10^{-19} C
dielectric constant in vacuum ε_0 = 8.85 x 10^{-12} C^2 J^{-1} m^{-1}
r is ionic radius, with is typically 1-2 Å

For ε_1 = 2-8 and ε_2 = 80 (H_2O) => ΔG_{tr} = +30 to 50 kcal/mol

Sharp, K.A. and Honig, B. (1990) Electrostatic interactions in macromolecules: theory and applications. Annu Rev Biophys Biophys Chem, 19, 301-332.

Tanford 1962

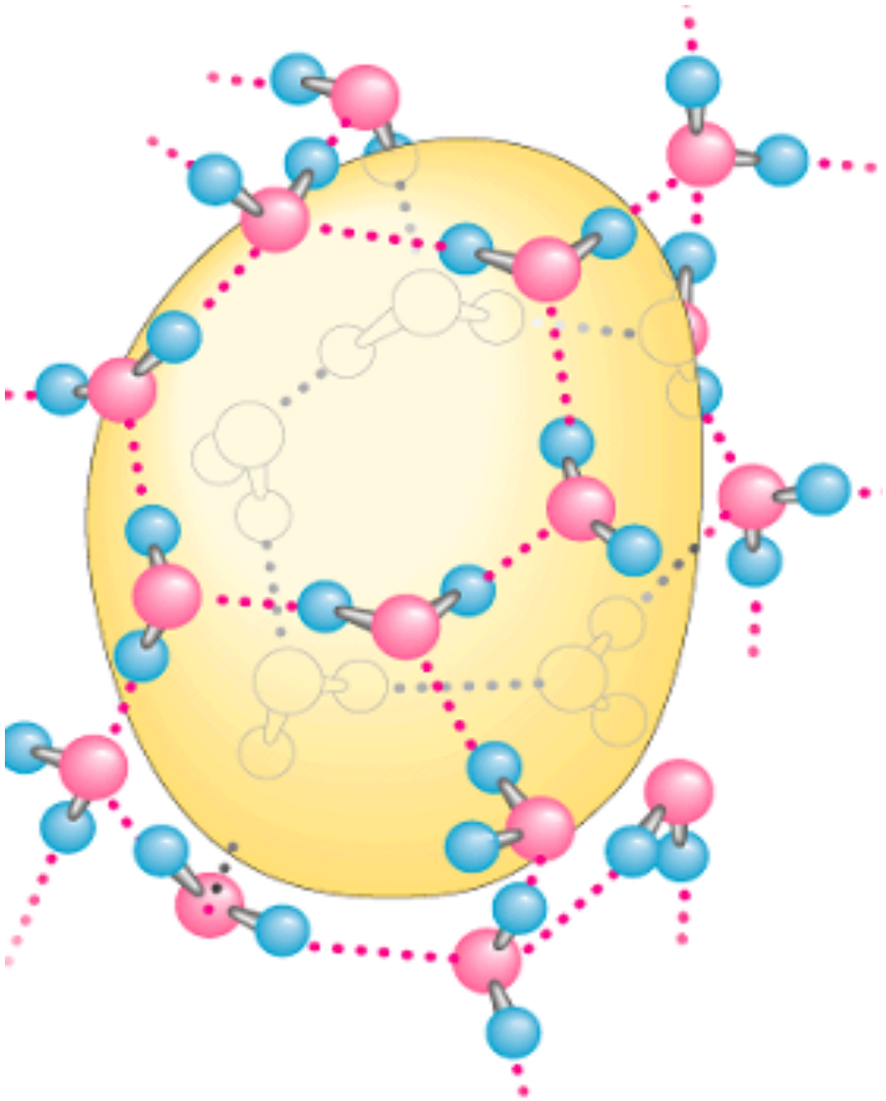
TABLE III

CONTRIBUTION OF THE MOST IMPORTANT HYDROPHOBIC INTERACTIONS TO THE FREE ENERGY OF UNFOLDING AT 25°

Side chain	Δf_u per side chain, cal./mole	Number present in myoglobin[a]	Number present in β-lactoglobulin[b]	Number present in ribonuclease[c]
Tryptophan	3000	2	2	0
Isoleucine	2970	9	10	3
Tyrosine	2870	3	4	6
Phenylalanine	2650	6	4	3
Proline	2600	4	8	4
Leucine	2420	18	22	2
Valine	1690	8	10	9
Lysine	1500	19	15	10
Methionine	1300	2	4	4
Alanine	730	17	14	11
Arginine	730	4	3	4
Threonine	440	5	8	10
Total number of residues		153	162	124
$-T\Delta S_{conf}$, kcal./mole		−184	−194	−149
$\Sigma\Delta f_u$, kcal./mole		+173	+192	+100

conformation entropy
hydrophobic effect

The hydrophobic effect drives protein-protein and protein-DNA interactions in water

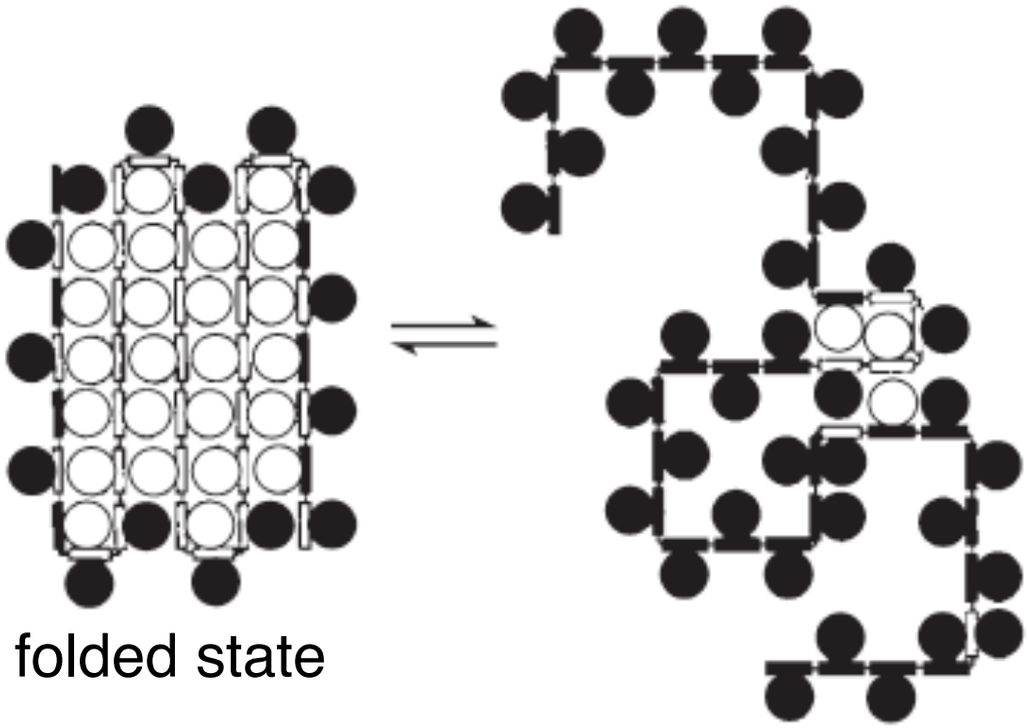


- Minimization of non-polar/water surface area
- Entropic contribution most significant
- Water“cage” around non-polar surfaces

Side-chain of amino acid residue	A^{np} (Å^2)	A^{pol} (Å^2)
Asp	48	58
Gln	53	91
Glu	61	77
Lys	119	48
Ala	67	—
Val	117	—
Leu	137	—
Ile	140	—

J. Mol. Biol. (1990) **213**, 375–384

Protein folding minimizes water accessible unpolar surfaces



folded state

random coil

Entropy change from
the number of states

$$\Delta S_{\text{fold}} = R \ln\left(\frac{W_{\text{unfolded}}}{W_{\text{folded}}}\right)$$

Folded: ~1 conformation
Unfolded: ~2^3 states (per aa)

ΔS_{fold} (2^3 states) = 8.31 J·mol^{-1}·K^{-1}·ln (⅛)
= 17.3 J·mol^{-1}·K^{-1} or “entropy units” (e.u.).

-T(25 ºC) x ΔS =1.2 kcal/mol or 5.2 kJ/mol

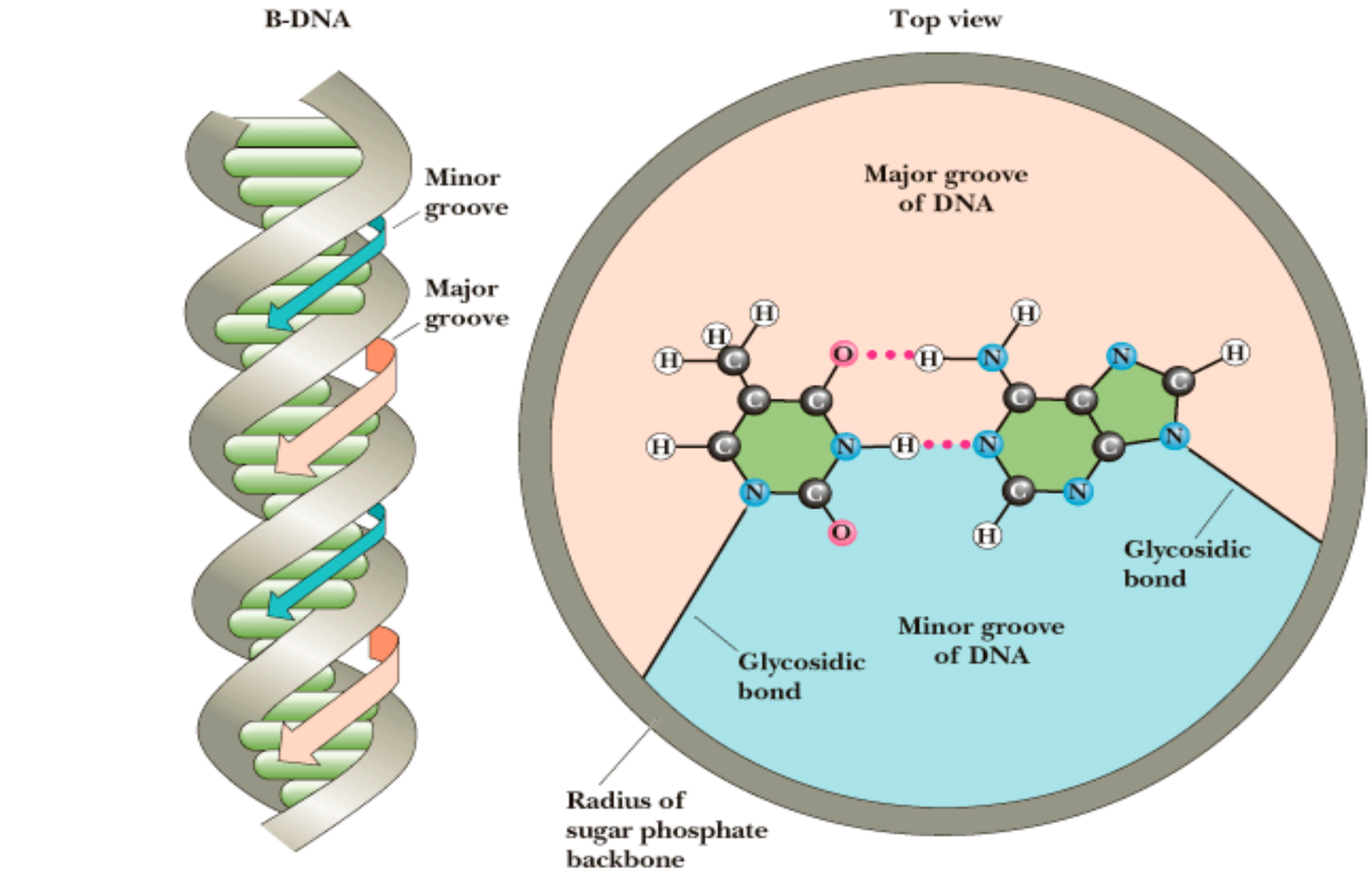
“Burying” an unpolar side chain: Leu = -2.4 kcal/mol, Val = -1.7 kcal/mol

“Burying” a negative/positive charge: ΔG_{tr} = +30 to 50 kcal/mol

Making interactions between proteins and double-stranded DNA

B-DNA – major & minor grooves

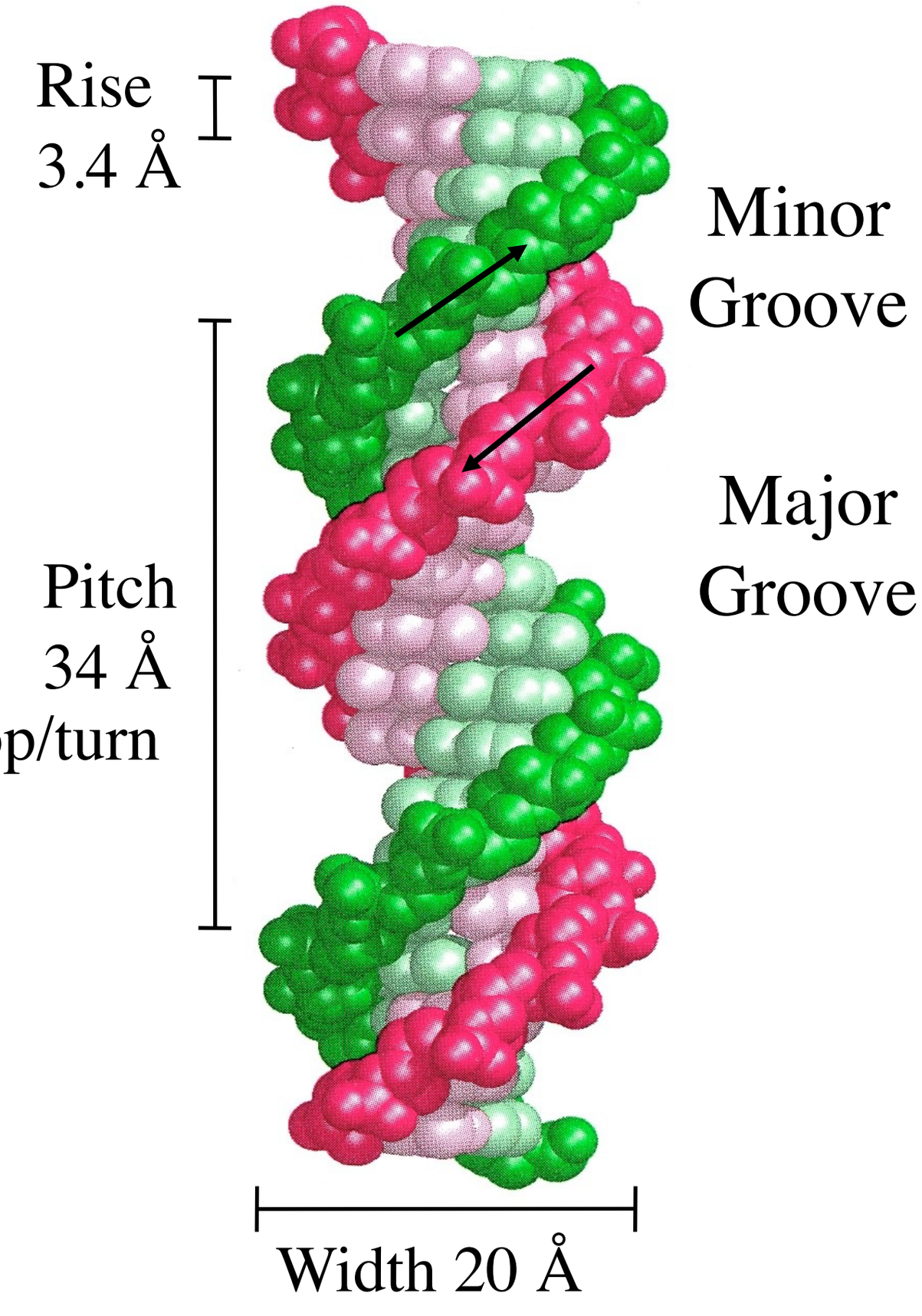
Garrett & Grisham: Biochemistry, 2/e
Figure 12.11

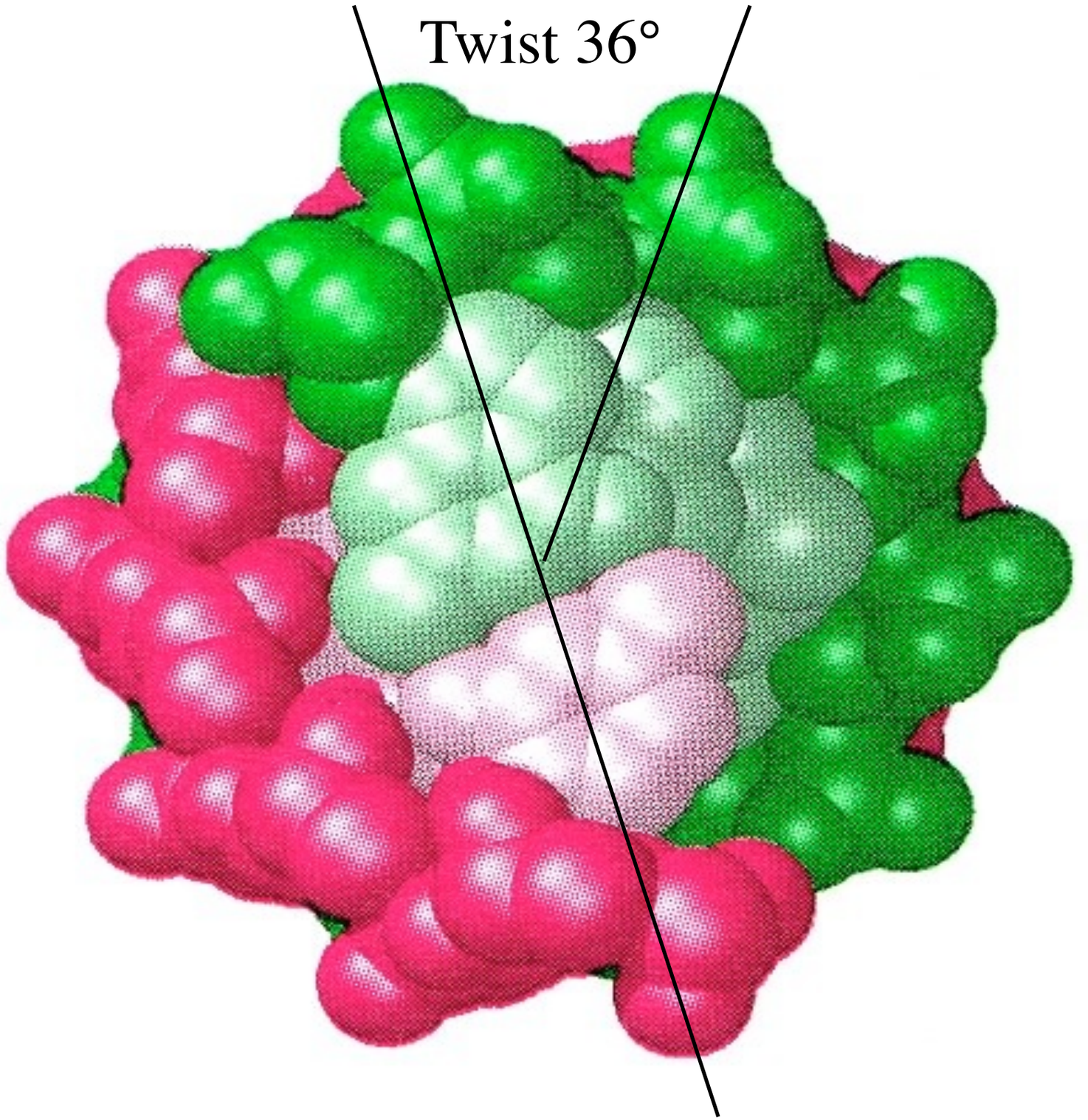


Saunders College Publishing

G & G 12.11

B-DNA: A right Handed double helix

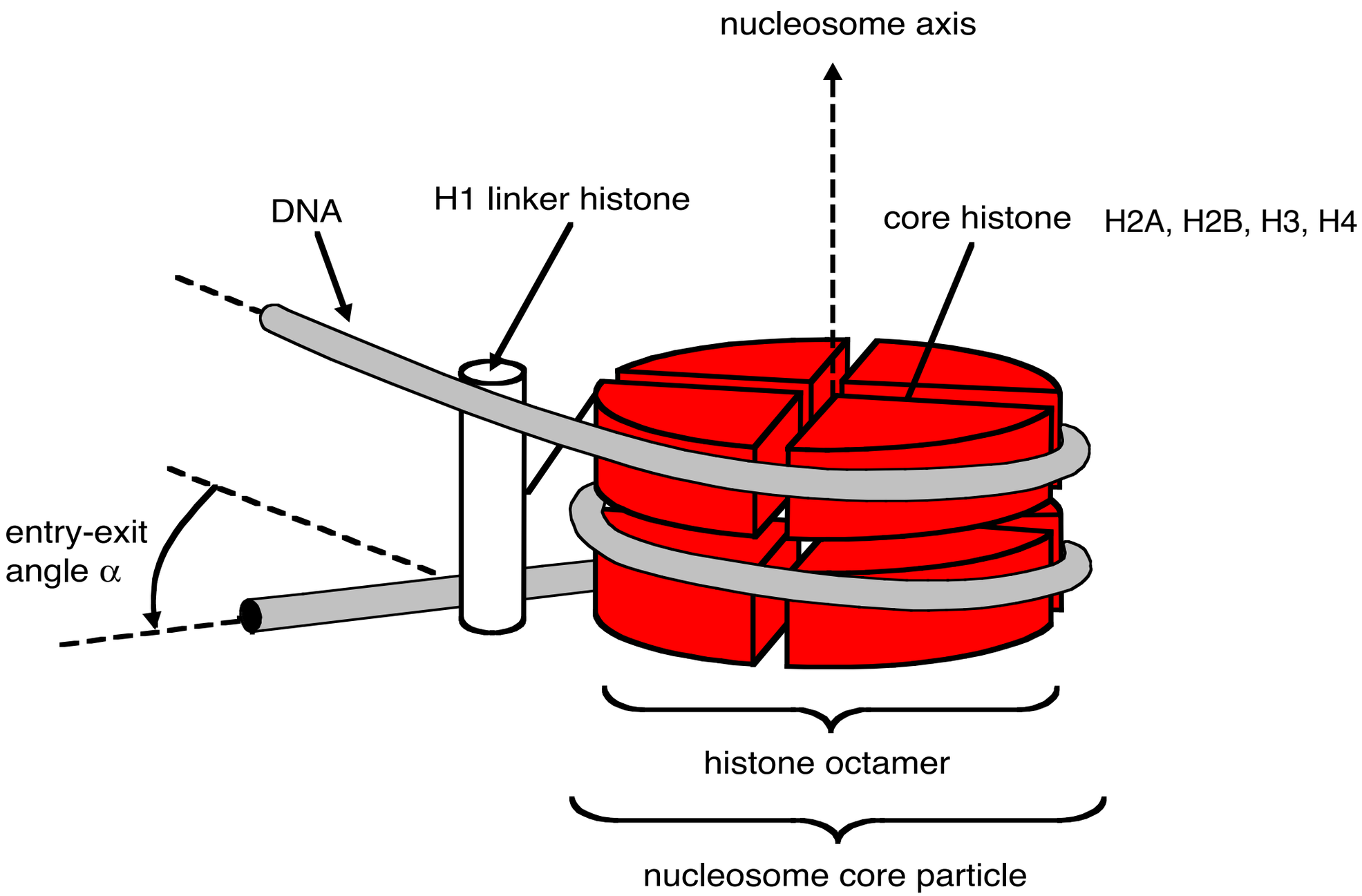



Twist 36°

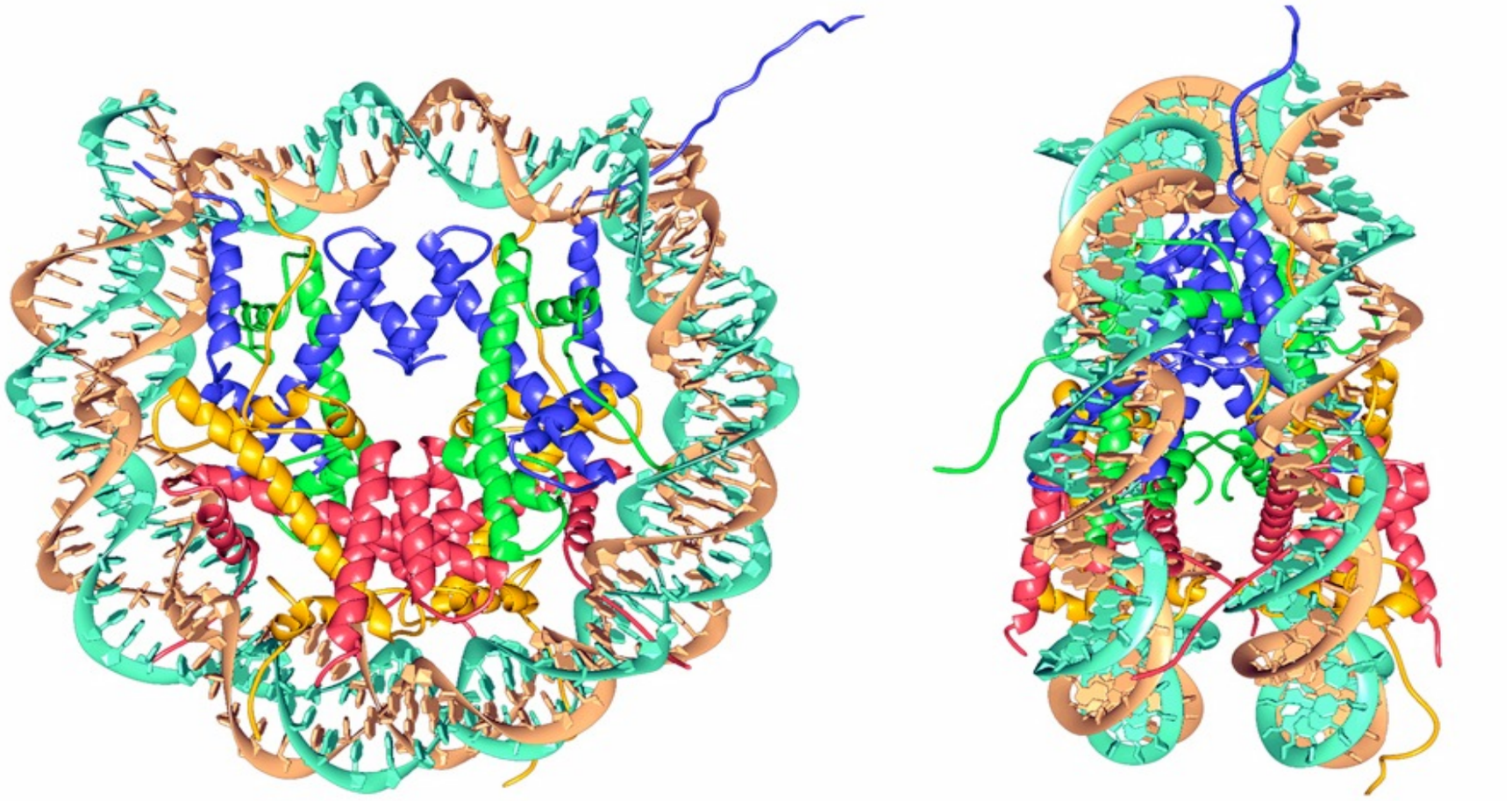
Protein-DNA interaction

- Sequence independent
  – may interact with the negatively charged sugar-phosphate backbone
- Sequence dependent
  – need to recognize the bases *in the double-helical structure* (don't have access to the atoms involved in base pair H-bonds)

Histone octamer - Nucleosome



Nucleosome crystal structure

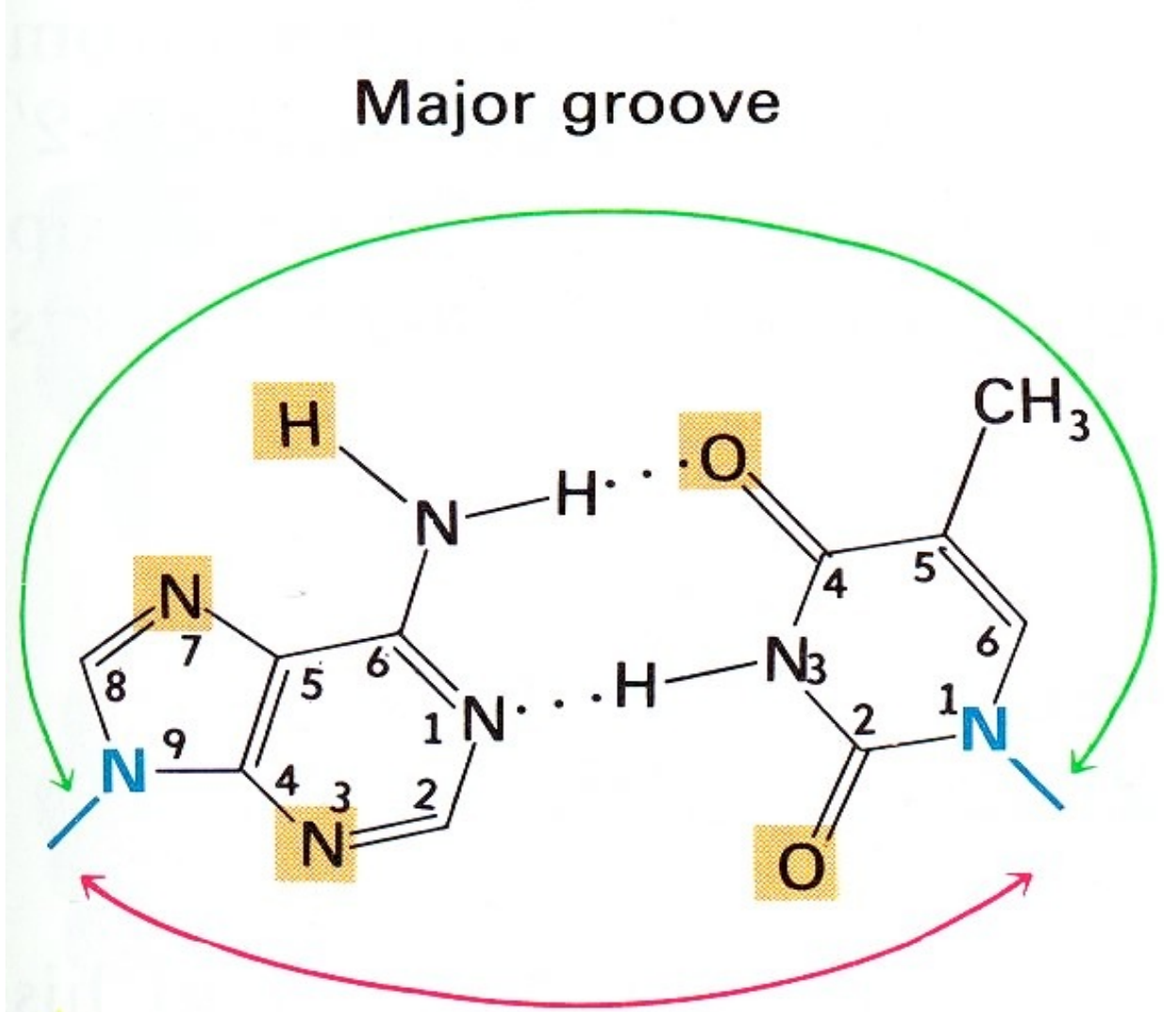


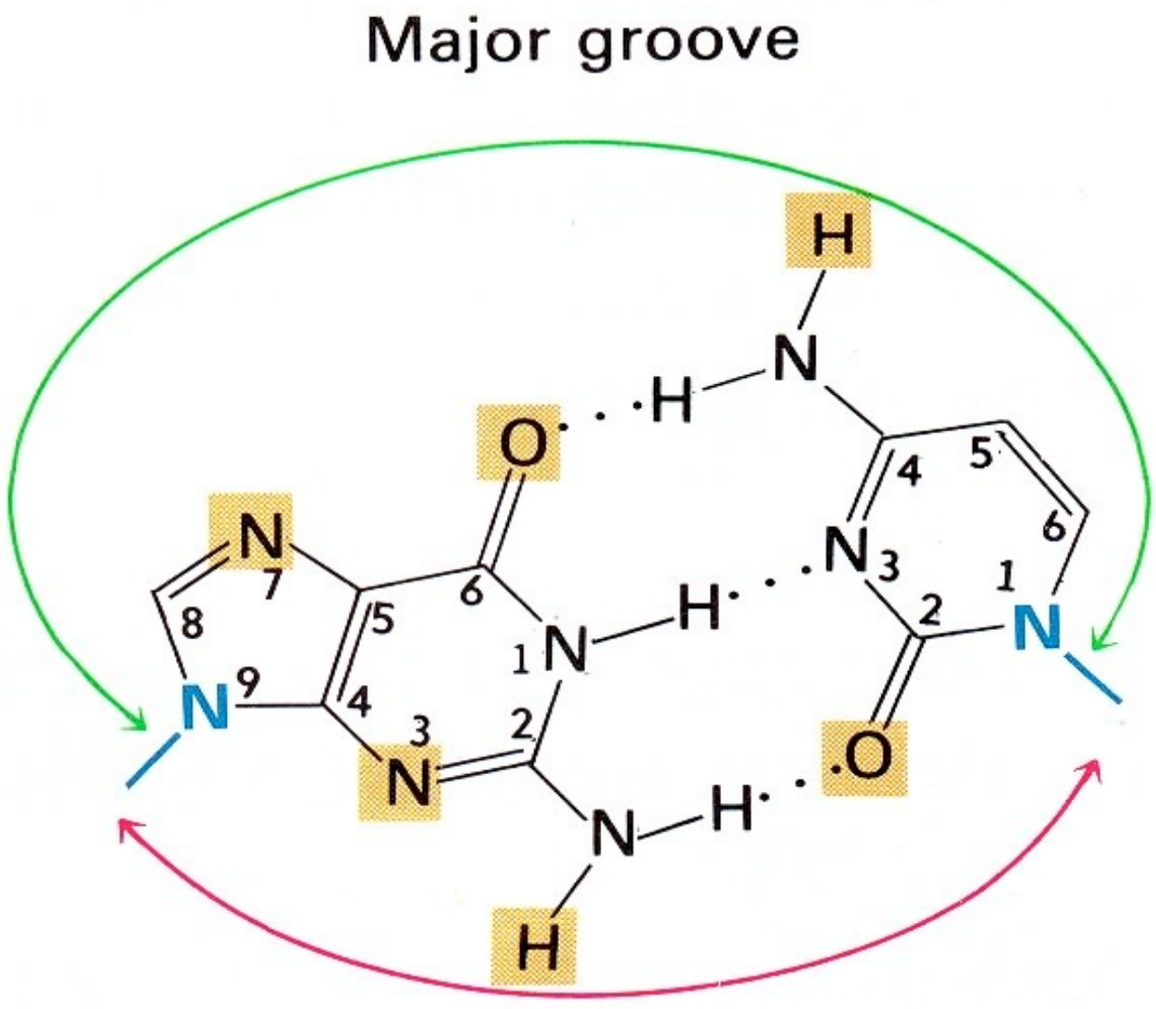
Histones in the nucleosome

- Histone proteins are Lys & Arg rich and highly positiviely charge (“basic”)
- 2 copies of H2A-H2B dimer and (H3-4)2 tetramer form octamers
- Basic histones interact with the negatively charged DNA phosphates
- The DNA is wrapped around protein core in ~1.7 turns

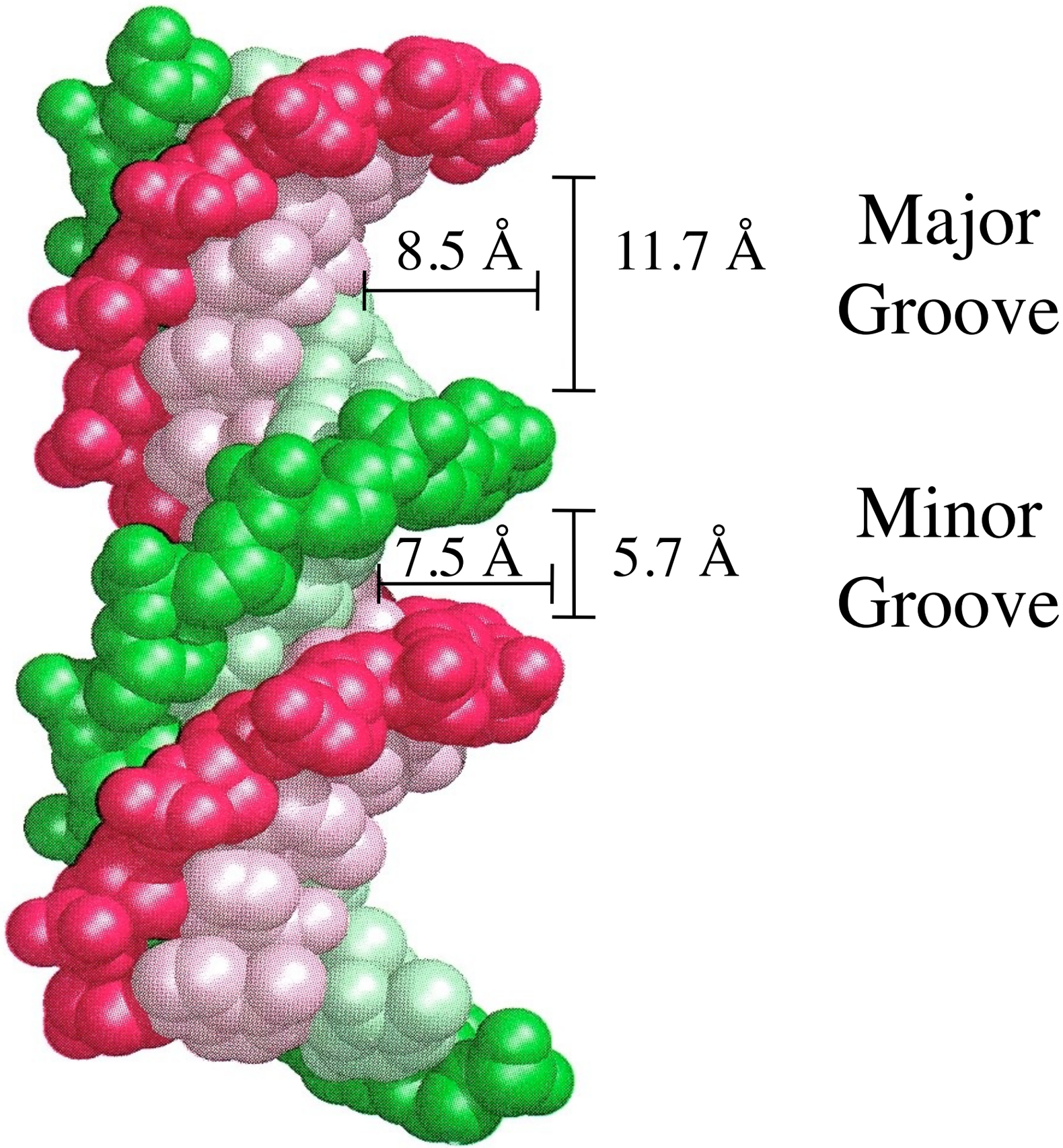
Specific binding of proteins to DNA

Which groove?





- H-bond donors and aceptors different in major groove for 2 types of base pair but very similar in the minor groove
- Minor groove too small to accommodate large protein probe


8.5 Å
11.7 Å
Major
Groove
7.5 Å
5.7 Å
Minor
Groove

Which groove?

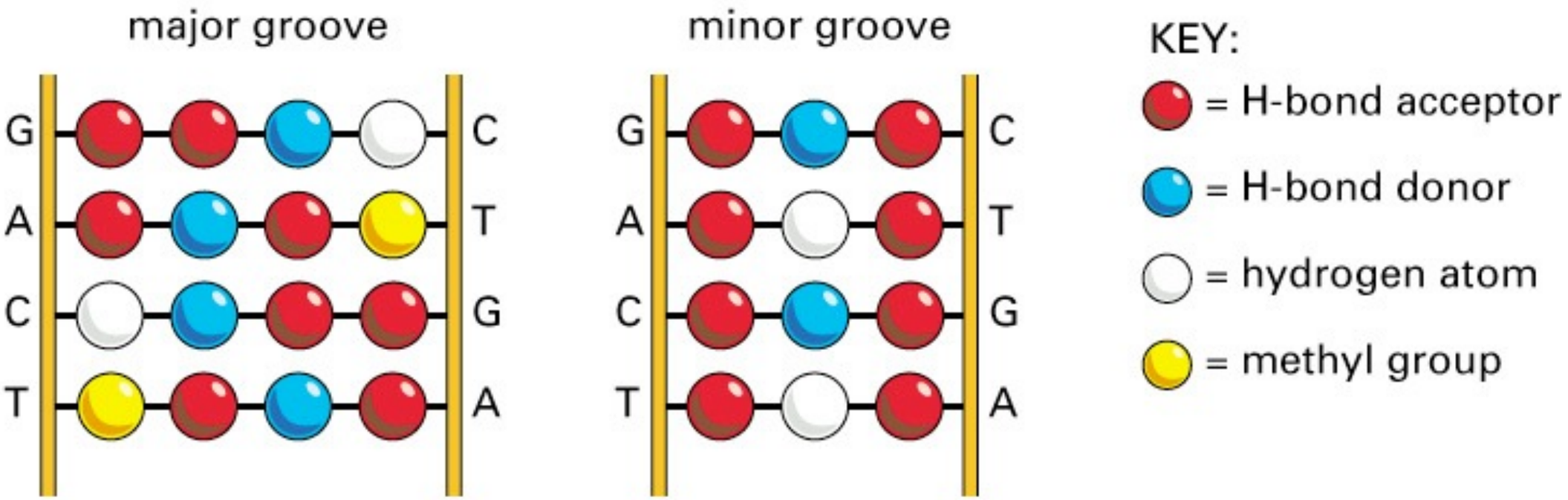
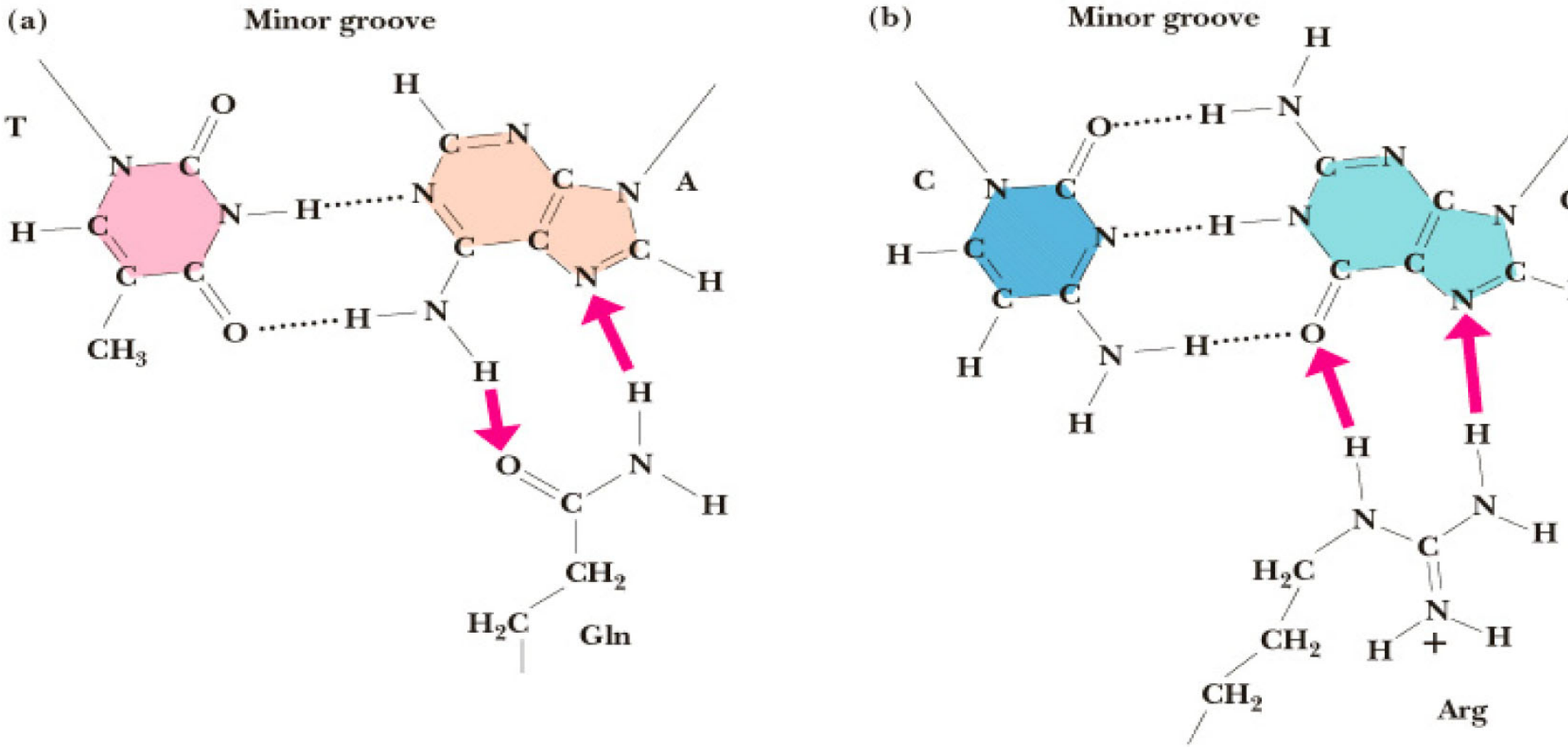


Figure 7–8. Molecular Biology of the Cell, 4th Edition.

- H-bond donors and aceptors different in major groove for 2 types of base pair but very similar in the minor groove
- Minor groove too small to accommodate large protein probe

Which groove?



- H-bond donors and aceptors different in major groove for 2 types of base pair but very similar in the minor groove
- Minor groove too small to accommodate large protein probe

Hydrogen bonding between aspargaine and adenine

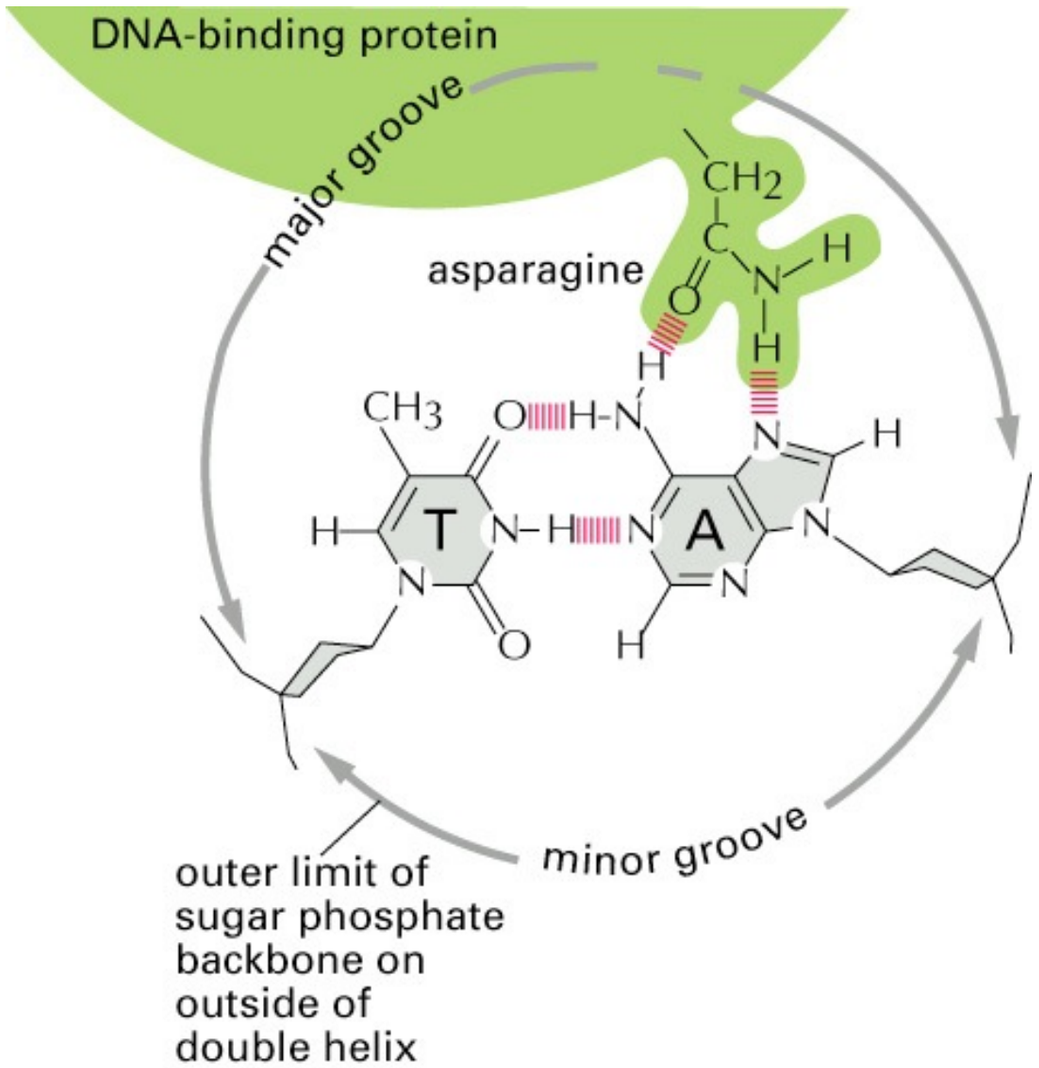


Figure 7–12. Molecular Biology of the Cell, 4th Edition.

α Helices and DNA - a perfect fit

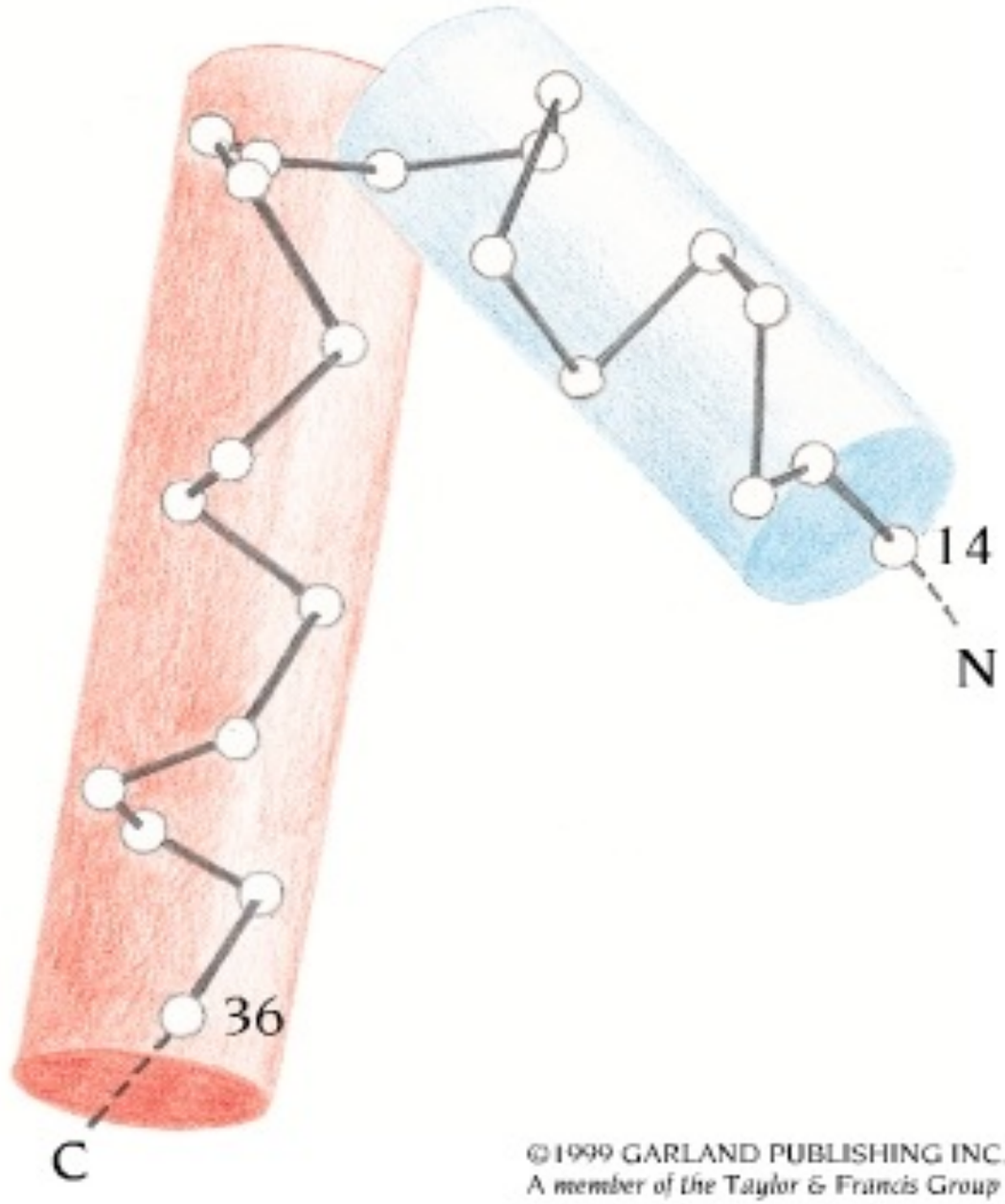
- DNA-binding proteins often have an a-helical segments that fit directly into the major groove of B-form DNA

- Diameter of helix is 1.2 nm (12 Angstroms)
  - Major groove of DNA is about 1.2 nm wide and 0.6 to 0.8 nm deep

- Proteins can recognize specific sites (sequences) in DNA

The helix-turn-helix motif

- Generally bind as dimers to dyad-symmetric sites on DNA
- All contain two alpha helices separated by a loop with a beta turn
- The C-terminal helix fits in major groove of DNA
- N-terminal helix stabilised by hydrophobic interactions with C-terminal helix



The helix-turn-helix motif: homeodomain transcription factor

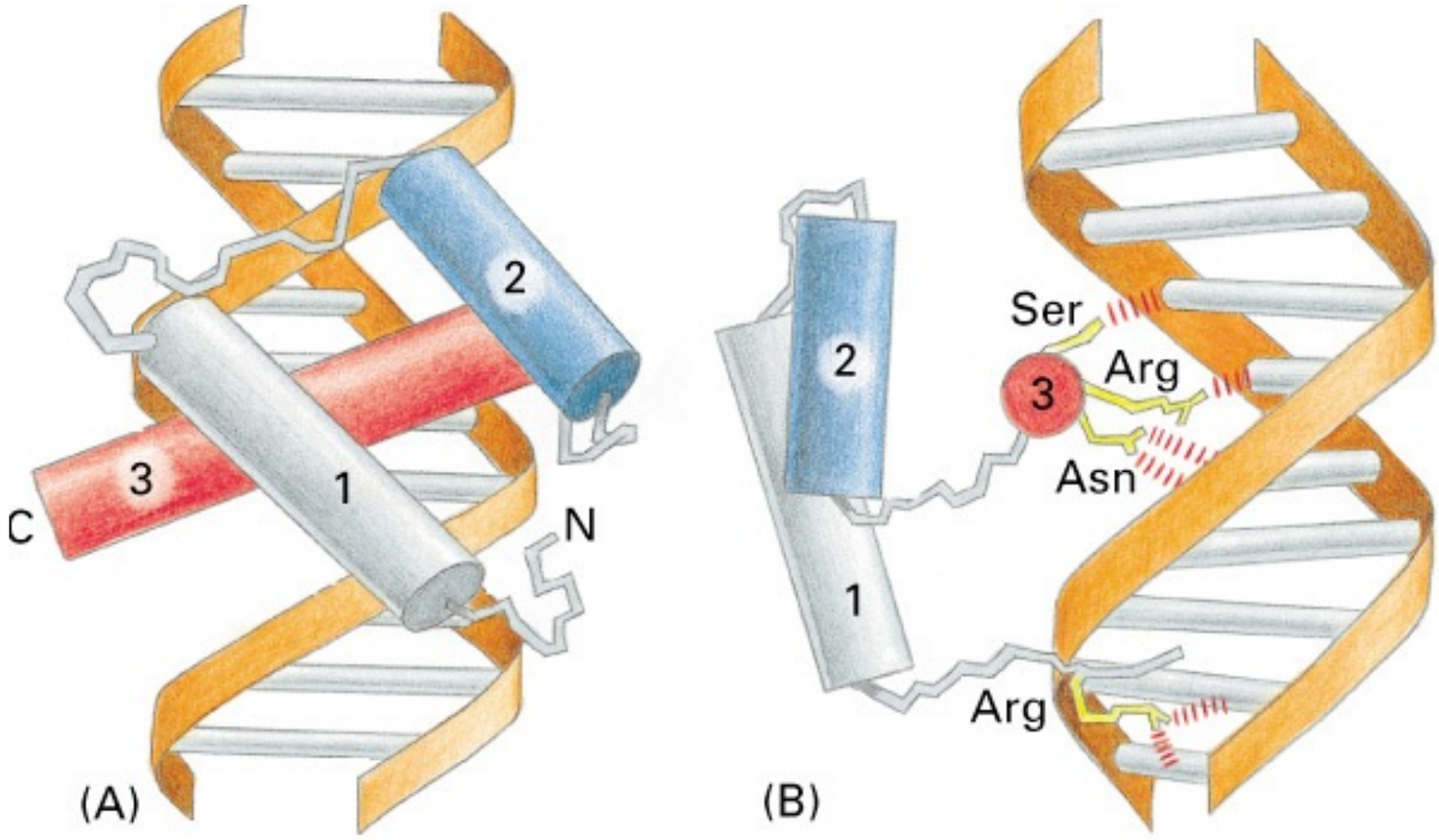


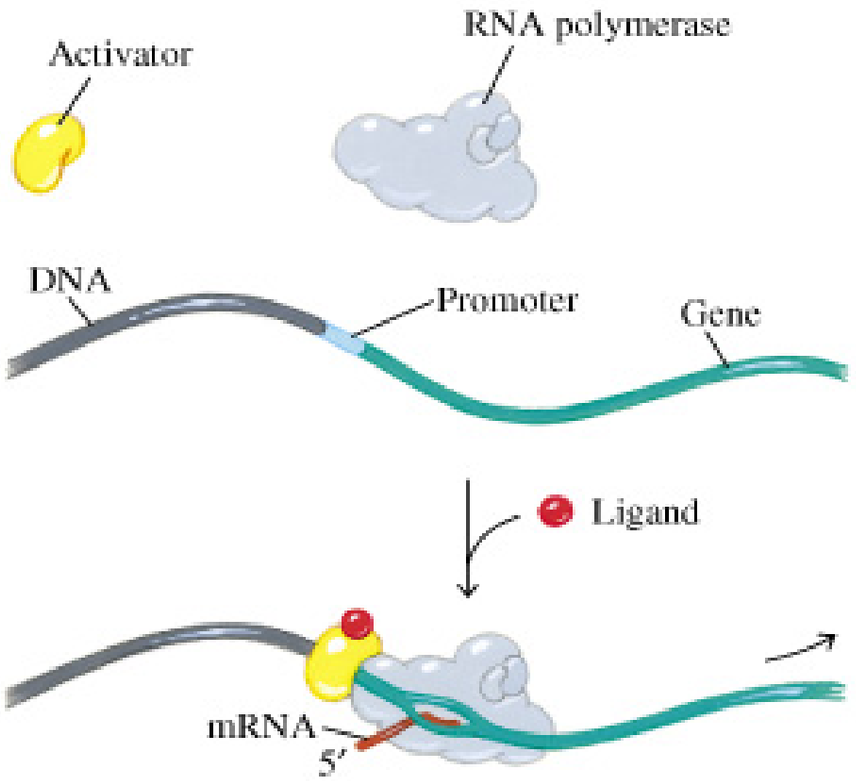
Figure 7–16. Molecular Biology of the Cell, 4th Edition.

Protein dimer binding to a symmetric DNA site

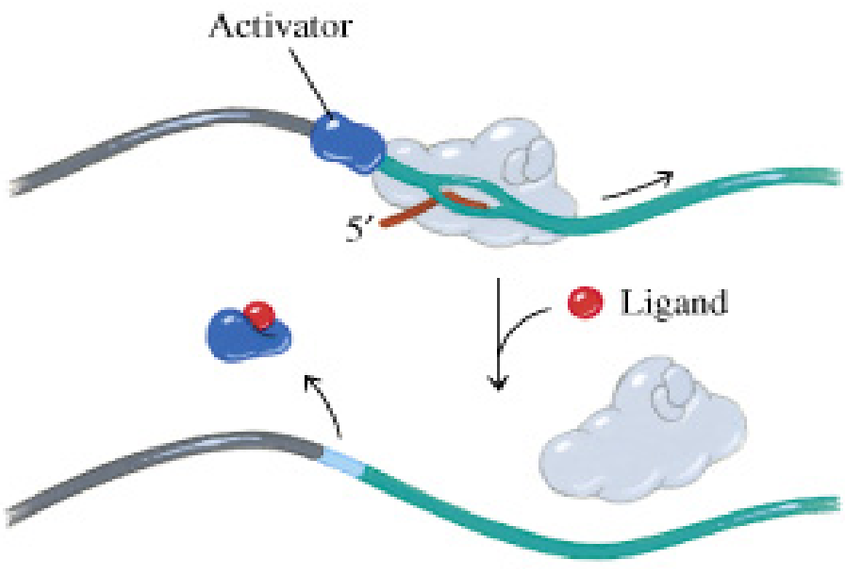
Figure 7–15. Molecular Biology of the Cell, 4th Edition.

Strategies for transcription regulation in bacteria

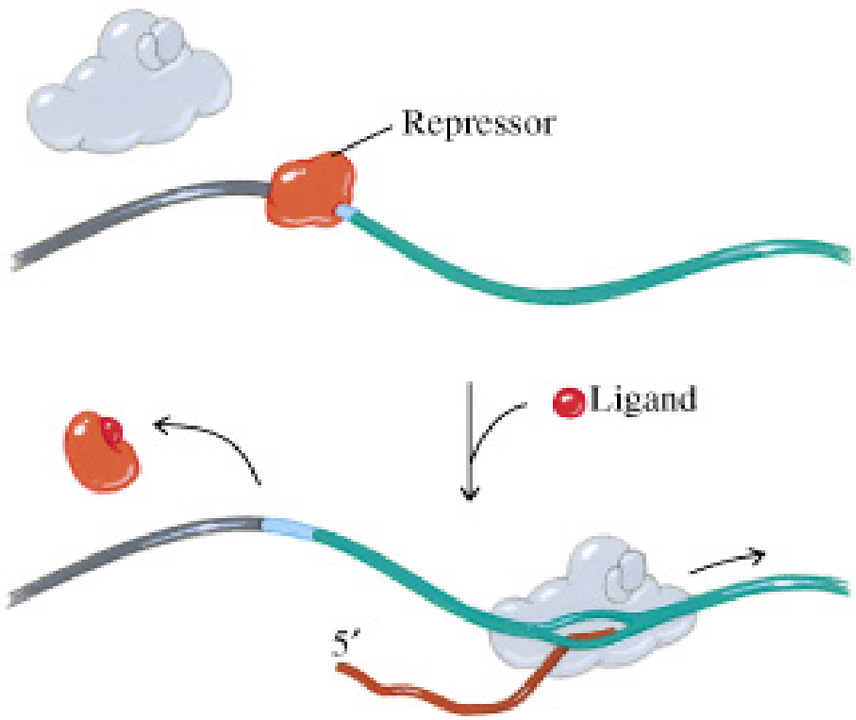
(a) An activator with bound ligand stimulates transcription.



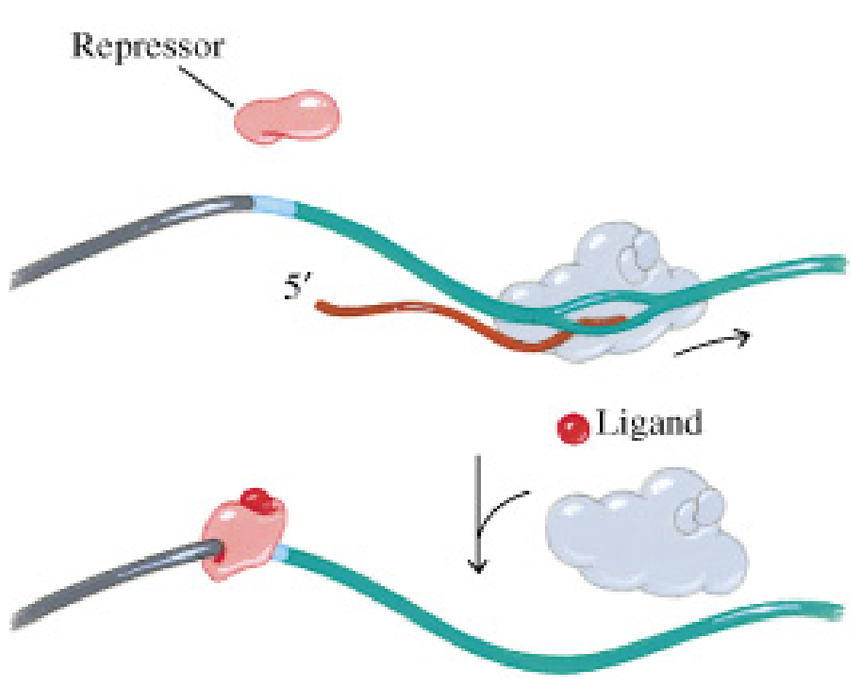
(b) An activator stimulates transcription. In the presence of ligand, the activator is inhibited.



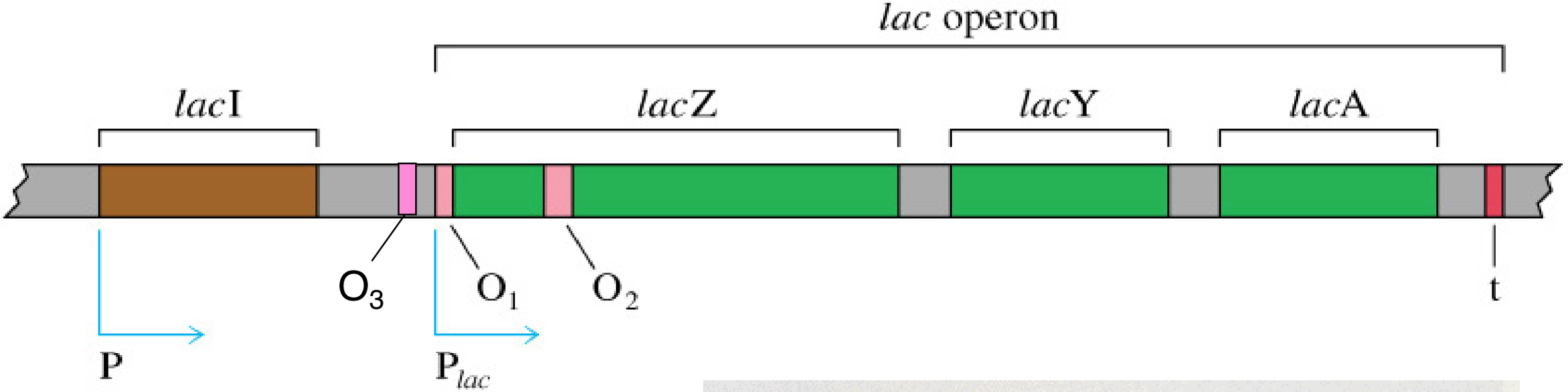
(c) A repressor prevents transcription. Binding of ligand (inducer) to the repressor inactivates the repressor and allows transcription.



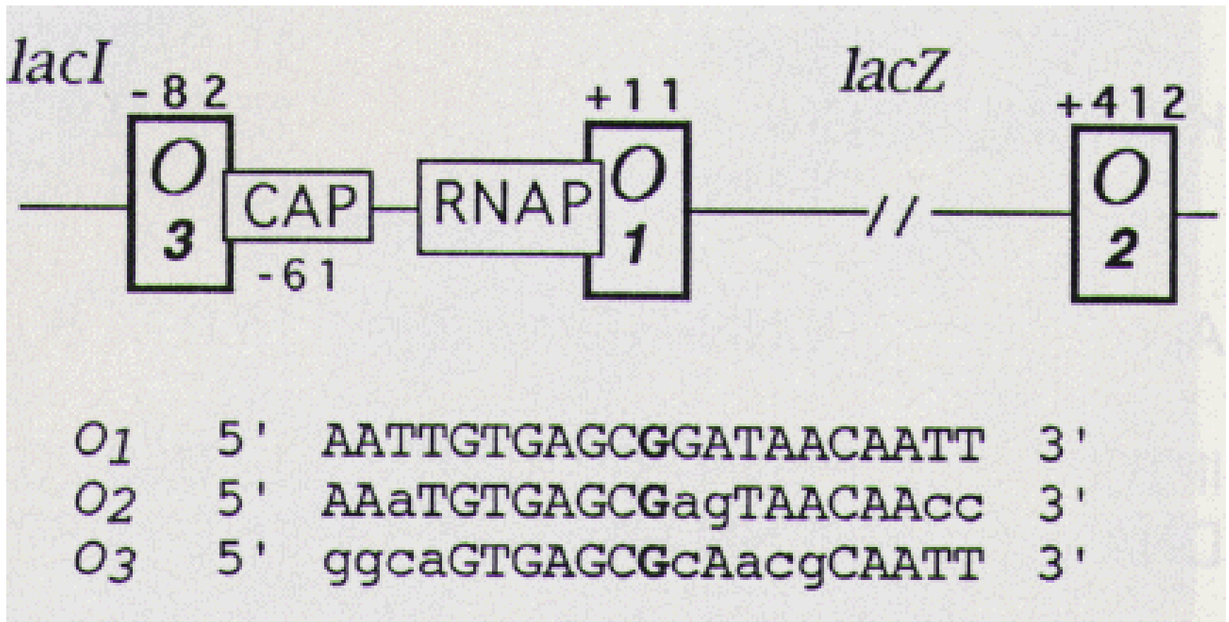
(d) In the absence of ligand, the repressor does not bind to DNA. Repression occurs only when ligand (corepressor) is present.

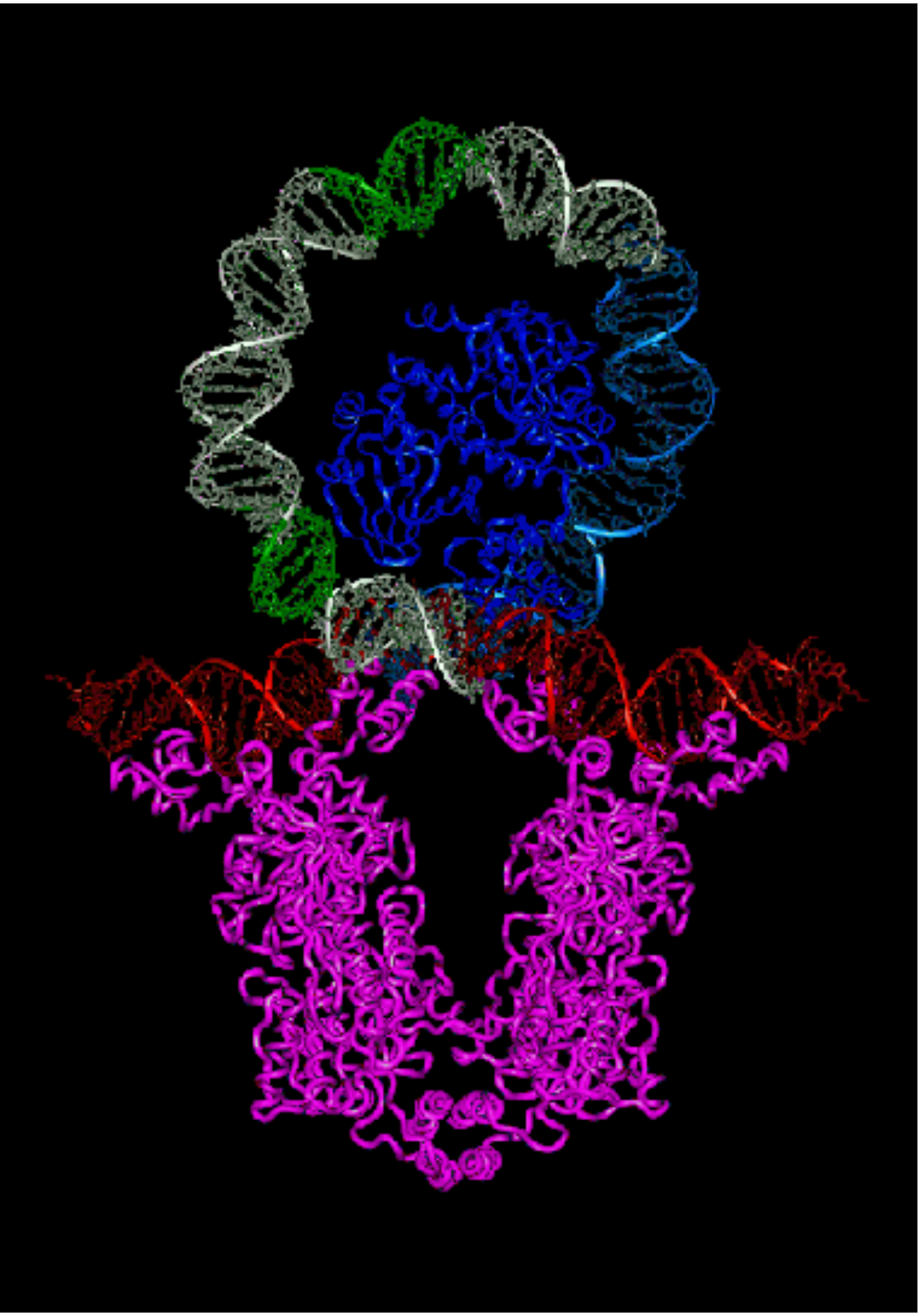


Organization of the genes regulated by Lac repressor, a transcription repressor protein in the bacterium E. coli



Lac repressor binds to the operators O_1, O_2 and O_3





Model for the complex of CAP and LacI at the lac operator

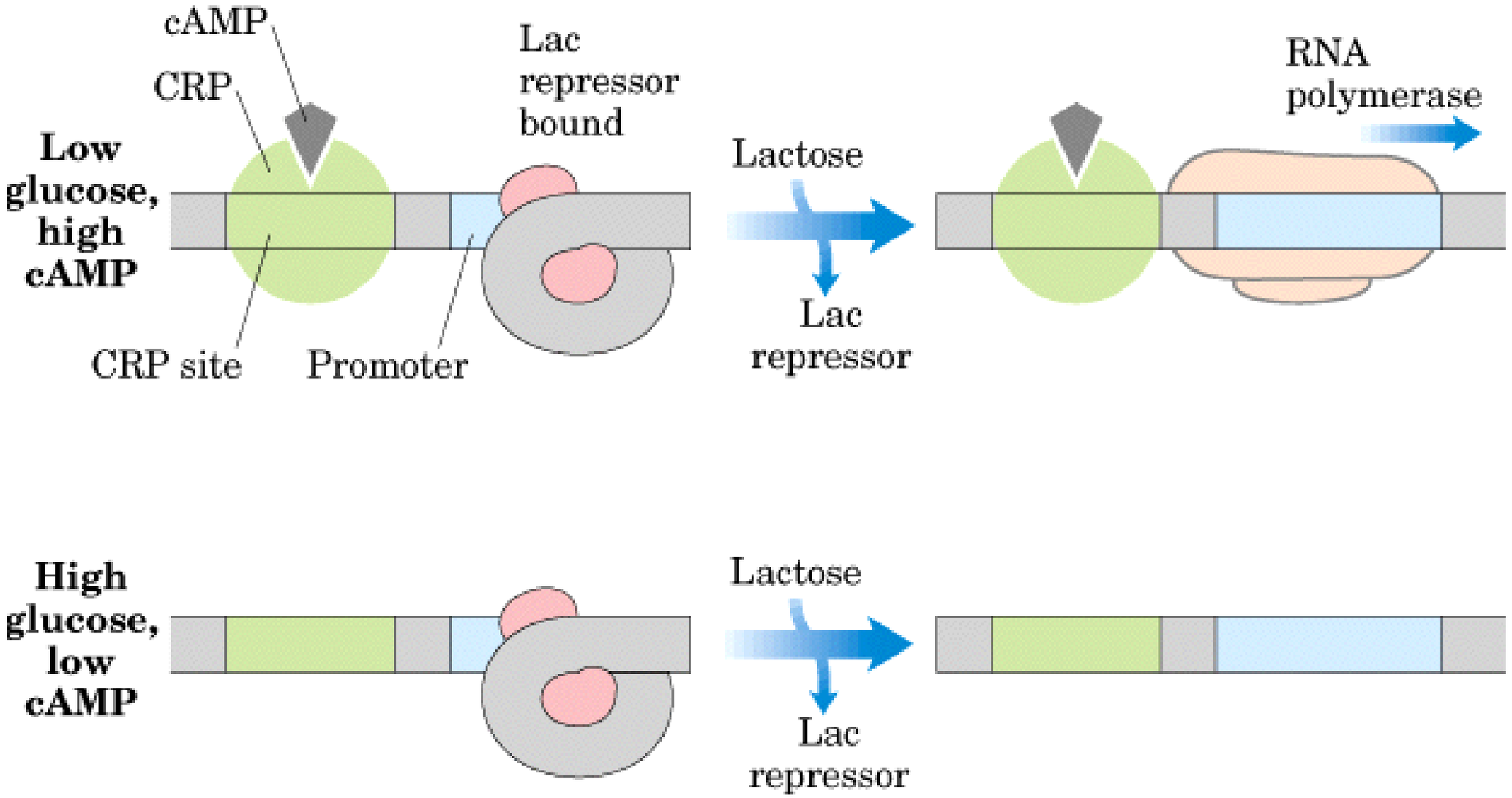
← CAP

low glucose and low lactose
=> both CAP and LacI bound
=> repression

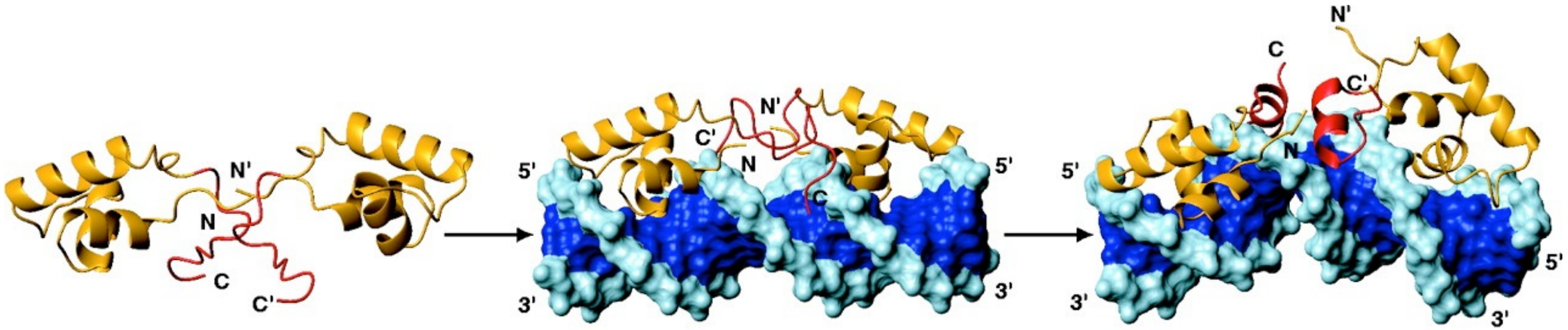
← Lac repressor bound to operator sites O1 and O3

Turning a gene on/off by binding of an activator/repressor

“On” and “off” states of the E. coli lac repressor



The hinge region (50-62 in red) of Lac-DBD is folded only in the specific complex with DNA



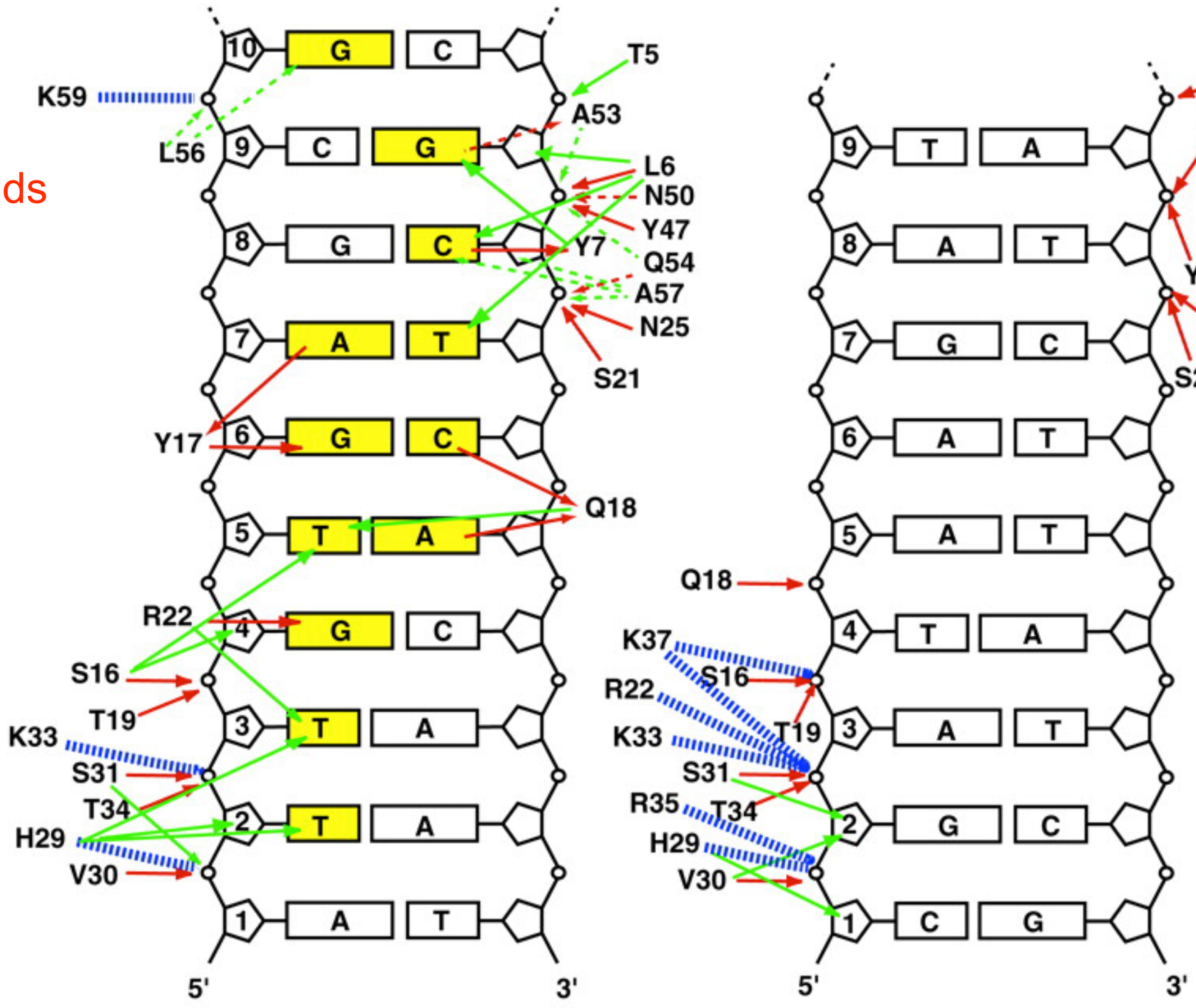
- folding of hinge region with specific contacts in minor groove
- specific interactions major groove
- less electrostatic interactions
- curvature of DNA

Specific (left) and nonspecific (right) protein-DNA contacts of Lac-DBD repressor with DNA

red:
hydrogen bonds

green:
hydrophobic
contacts

dashed blue:
electrostatic
contacts



Specific complex

Non-specific complex

Features of protein DNA interactions

- Unspecific electrostatic interactions of positively charge Arg and Lys with negatively charged sugar-phosphate DNA backbone
- Sequence recognition by non-covalent contacts between protein residues and bases
- Most contacts are in the major groove of DNA
- No simple recognition code between amino acid and base pair
- Frequently dimers that recognize palindromic sequence motifs
- 80% of regulatory proteins can be assigned to one of three classes:
  - helix-turn-helix (HTH)
  - zinc finger (Zn-finger)
  - leucine zipper (bZIP)
  - helix-loop-helix (HLH)
- In addition to DNA-binding domains, these proteins often possess other domains that interact with other proteins (“activation domain”)