MIC/BIO/BCH522

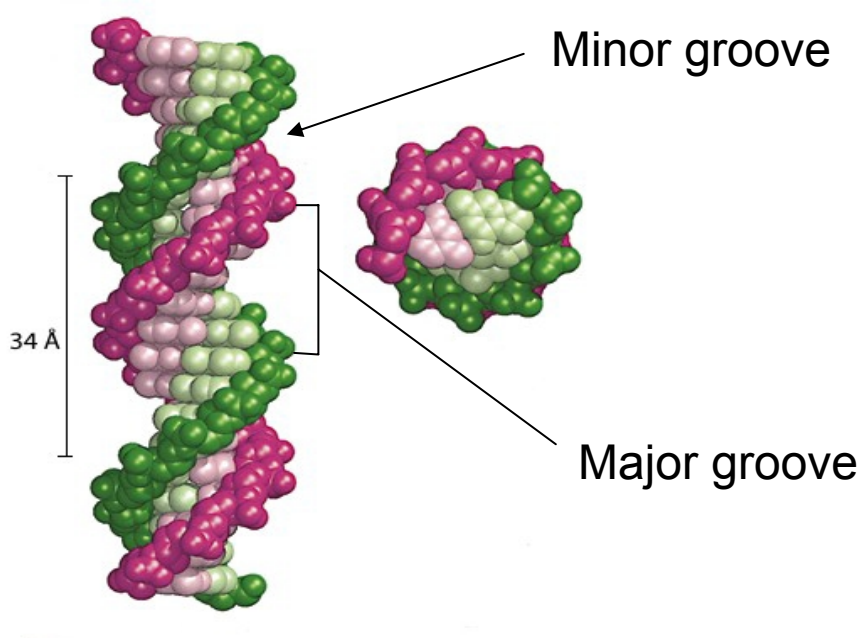
Spring 2006

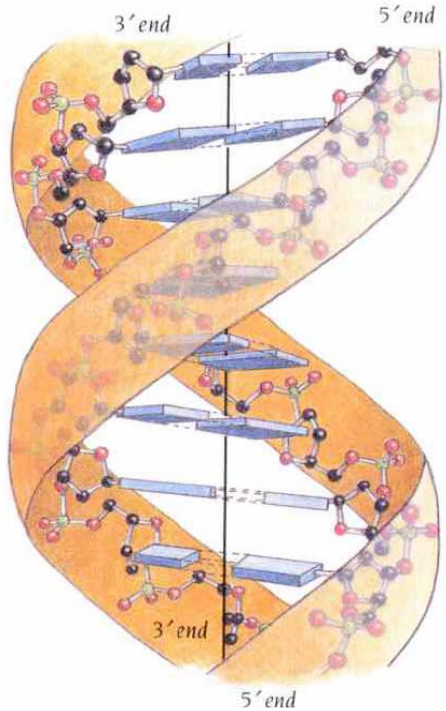
Interactions of proteins with nucleic acids

1. The helix-turn-helix motif
2. Homeodomain proteins
3. Nucleosomes

B-form dsDNA – quick review

1. B-form of dsDNA is the most common DNA conformation *in vivo*
   - it is the conformation to which the majority of proteins bind
2. Major groove of B-DNA is 12 Å wide, 6-8 Å deep.
3. Minor groove is deep and narrow (~5.7 Å). Too deep for side chains alone, too narrow for an α-helix.
4. DNA helix does not need to unwind for recognition by proteins
5. Proteins bind to outer surface, at edges of base-pairs
6. Nucleotide chemical groups are accessible, make unique patterns
7. Nitrogen and oxygen atoms at the edges of the base pairs and at the bottom of the major groove can hydrogen bond with proteins to make sequence specific contacts
8. H-bonding and van der Waals interactions between side chains and exposed edges of the base pairs



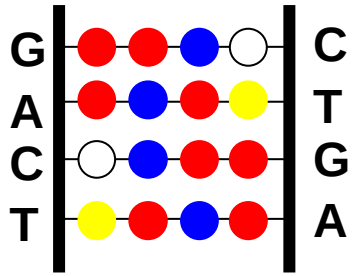


How are specific bases recognized in DNA?

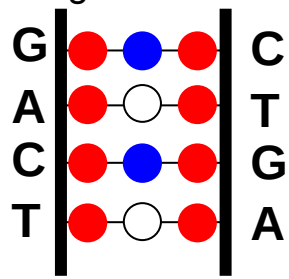
- As the DNA helix remains intact for recognition to occur, and the bases are unique,
- The edges are the only part of each base pair that can be recognized
- The edges available on each base pair in the major and minor groove are distinct

Color codes for recognition patterns

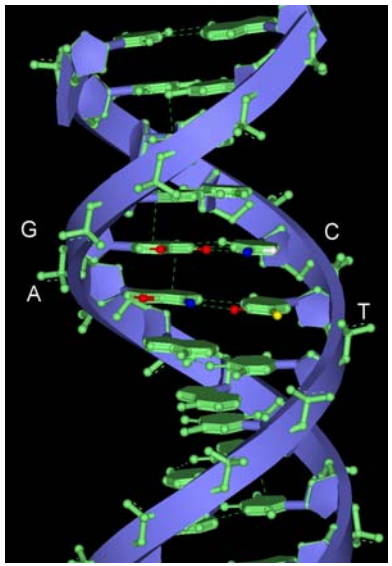
Major groove of B-DNA

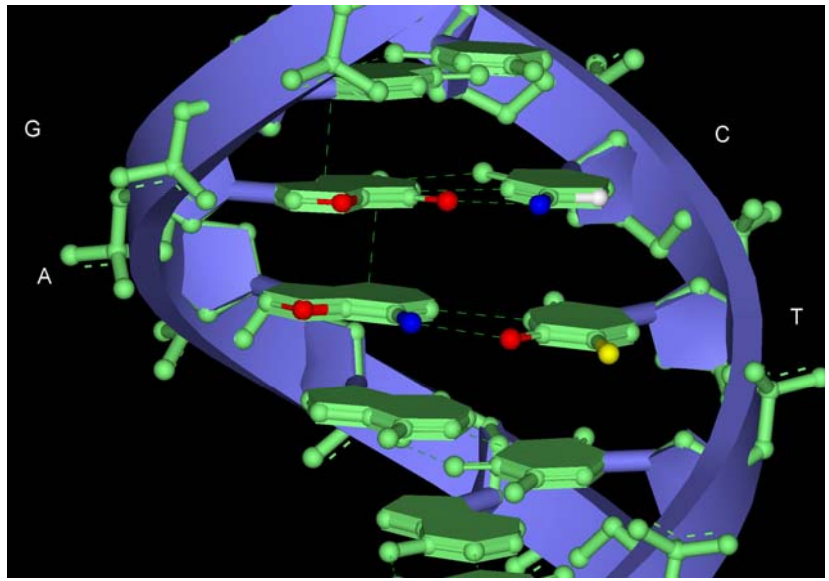


Minor groove of B-DNA



The edges of base pairs available in the major groove





Methyl group H-bond donor H-bond acceptor H-atom

What are the motifs commonly found in proteins used to bind to DNA?

dsDNA binding proteins

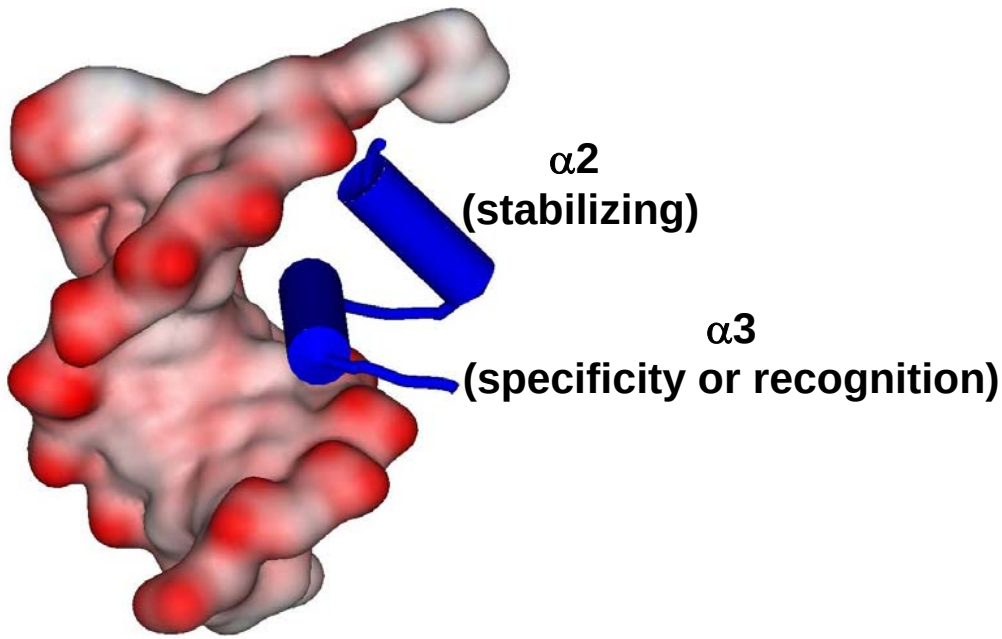
- Helix-turn-helix
  - Prokaryotic repressors (cro, met, λ)
- Homeodomain proteins (modified HTH)
- Nucleosomes
- Zinc finger
  - Eukaryotic transcription factors
- Leucine zipper
  - Eukaryotic transcription factors (oncogenes)
- Winged helix
  - Eukaryotic transcription factors
- Beta ribbon proteins
  - Met and arc repressor, TATA binding protein, IHF and HU

ssDNA binding proteins

- OB-fold
  - ssDNA binding proteins (SSB, RPA, BRCA2)

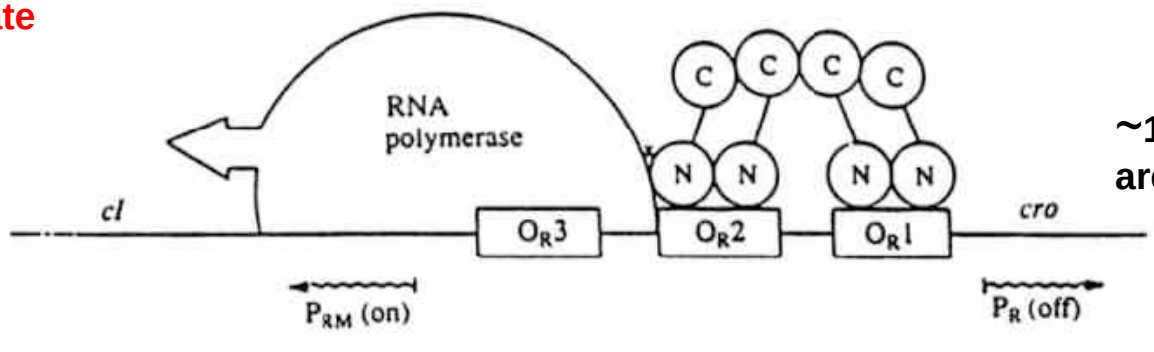
The Helix-Turn-Helix Motif

- Most common DNA-binding motif and is typically ~20 amino acids in length
- It consists of two α-helices and a short extended amino acid chain (non-helical segment) between them
- α-helices are at right angles to one another
- One α-helix fits into major groove of DNA and functions as the recognition helix
- The second α-helix lies at an angle across DNA and functions as a stabilizing helix.
- The stabilizing helix butts up against outside of DNA and ensures proper positioning of recognition helix
- The Helix-turn-helix motif is structurally stable as a component of a larger protein
- Note that the HTH is a motif, not a separate domain.



Repressors control lytic vs. lysogeny decision in phage

Prophage state



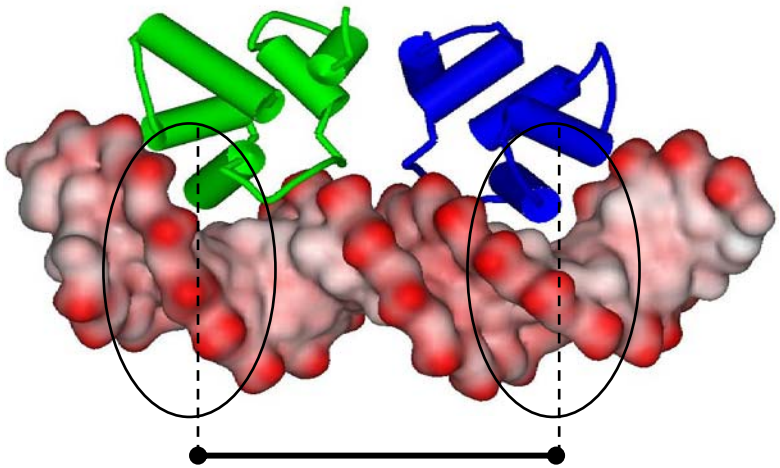
~17 bp operators are nearly palindromic

λ phage rightward operator (λPR) with two dimers of λ repressor binding to OR2 and OR1, turning off transcription of cro gene and stimulating transcription from PRM.

- Phage have two stages of their lifecycle:
  - (1) Prophage - phage genome is incorporated at attB and lytic gene expression is off
  - (2) Lytic - turn off repressor of lytic genes, turn on expression of the lytic genes
- Conversion between pstages is modulated by different repressor proteins **Cro and cI**
- Prophage state (as diagrammed):
  - cI (λ repressor) binds to OR1 and OR2, turns off PR by blocking RNA pol from transcribing lytic
- genes (e.g., cro). cI binding turns on PRM to make more cI (both a repressor and an activator).
- Lytic state:
  - cro binds to OR3, turns off PRM, turns on PR to transcribe the lytic genes.
- Cro and cI bind to same operators but with differing affinities:
  - cI: OR1 > OR2 > OR3
  - Cro: OR3 > OR2 > OR1

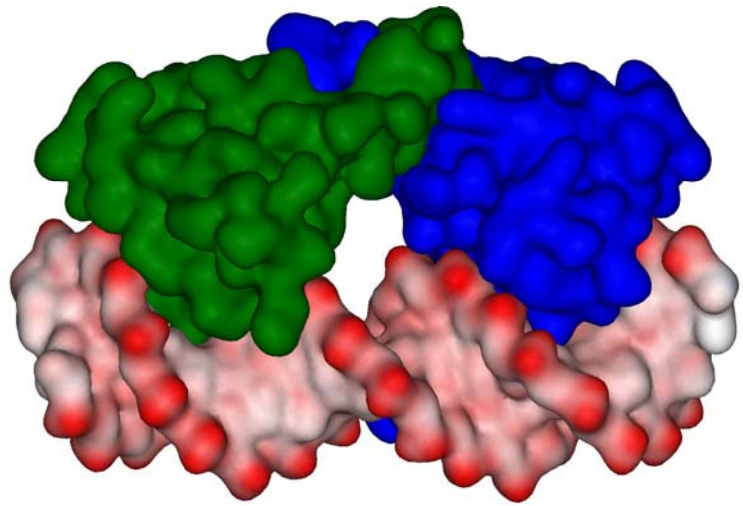
Features of the HTH motif's interaction with DNA

- The helix-turn-helix motif was originally identified as the DNA-binding domain of phage repressors.
- This motif has since been found in hundreds of DNA-binding proteins:
  - λ repressor, tryptophan repressor, catabolite activator protein (CAP), lac repressor
- Although the HTH is *monomeric in nature*, dimerization is required for full activity
- Dimerization is necessary for complete recognition of the palindromic nature of operator regions
- Within the palindrome are 2 identical recognition sites one for each recognition helix (ie one per subunit of the dimer)
- Sequences separated by distance equal to one turn of helix
- Two sites doubles contact zone between protein and DNA
- There is little/no contact in the minor groove
- Binding typically distorts the DNA (distortion varies from protein to protein)



5′-AAT**ACCA**CTGGCG**GTGA**TAT

3′-TTA**TGGT**GACCGC**CACT**ATA

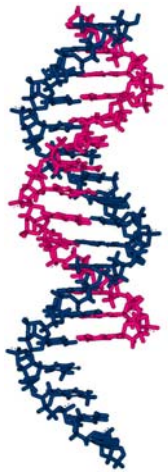


Lambda repressor (file 1LMB)

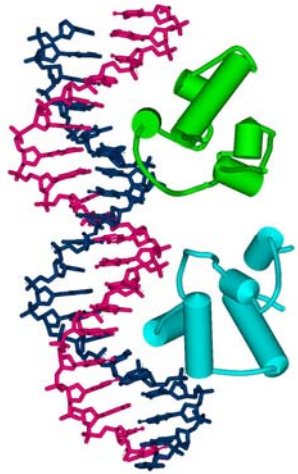
Proteins with the HTH-motif distort the DNA

1. B-form dsDNA is shown for comparison to five repressor molecules
2. Note the recognition α-helix positioned deep within the major groove
3. In each case the DNA is distorted but that the distortions are different
4. In lac repressor and CAP the DNA distortion is very large

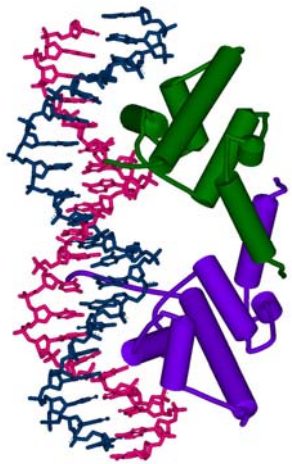
DNA only



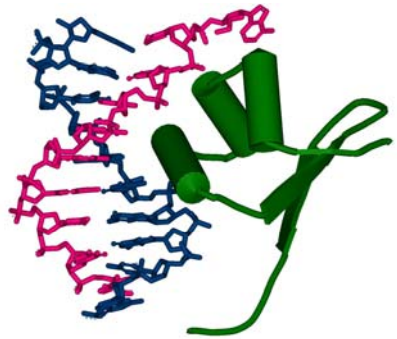
434 repressor

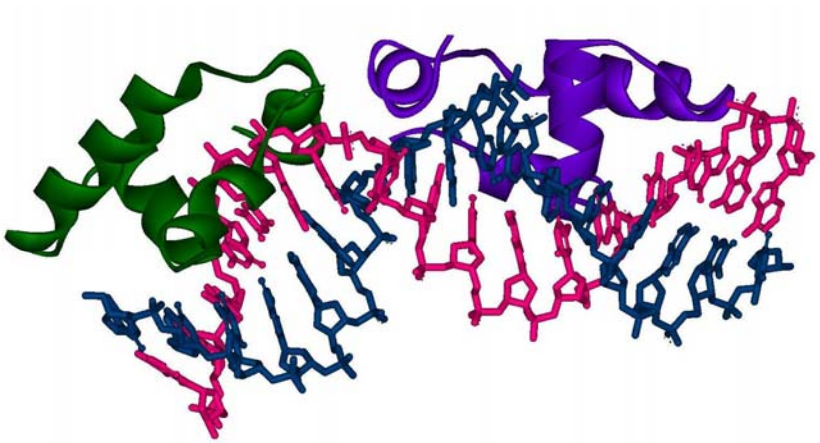


λ repressor

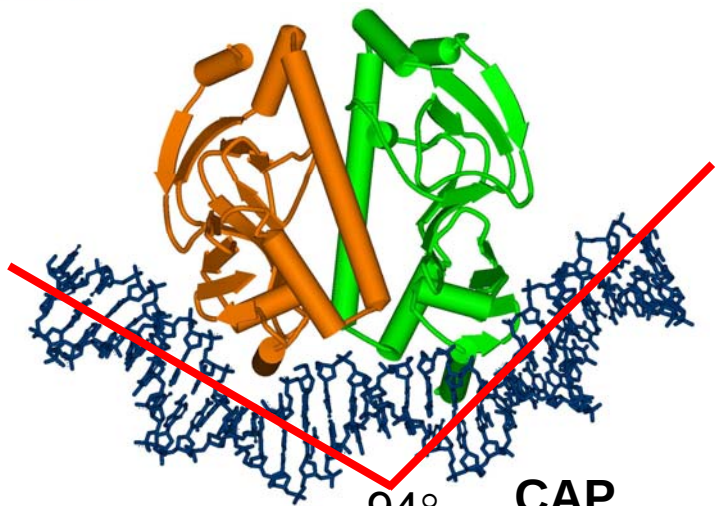


λ cro





lac repressor

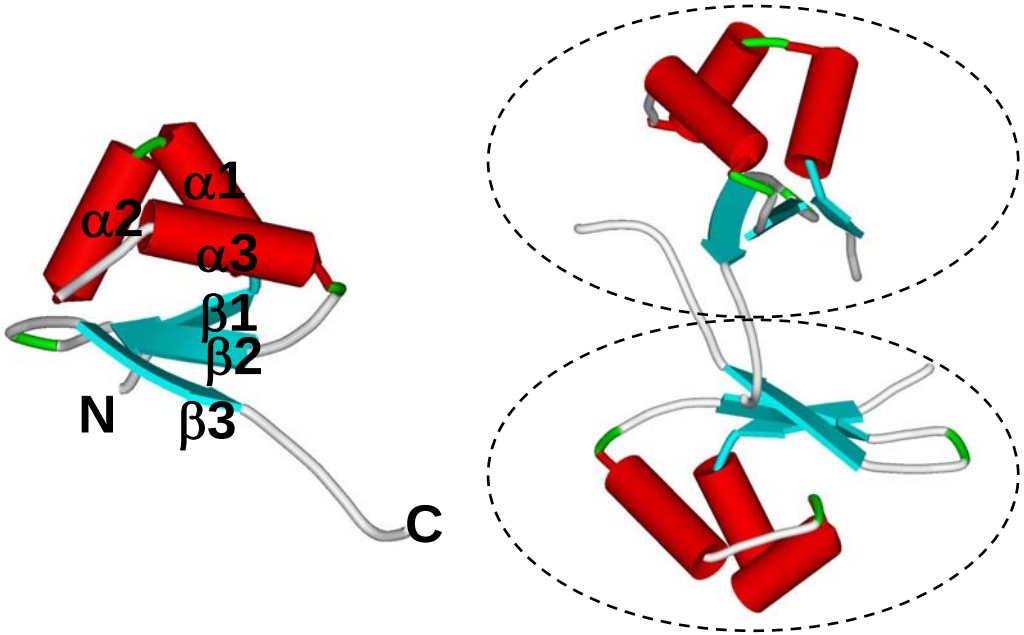


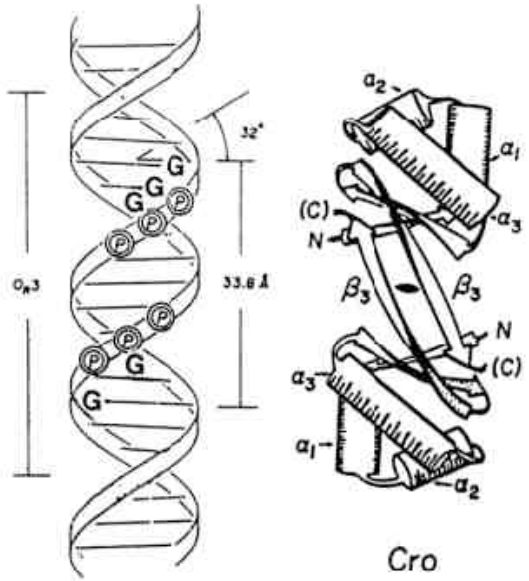
CAP

Crystal structures of HTH protein/DNA complexes (λ repressor, λ cro, phage 434 repressor, 434 cro, Trp repressor, CAP)

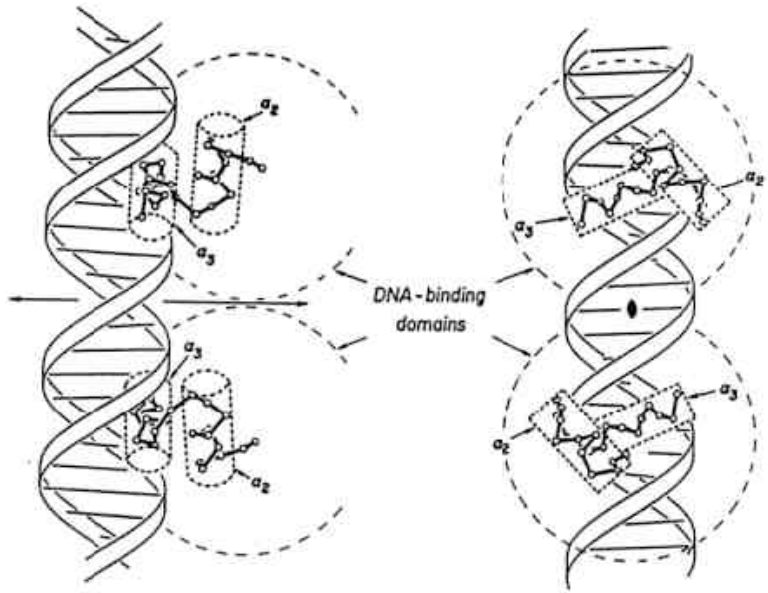
- Operator sites are B DNA, but with distortions in DNA that are different for each complex.
- Repressors bind as dimers with each monomer recognizes a half site.
- Approximate symmetry of DNA reflected in approximate symmetry of complexes.
- Conserved HTH unit contacts the DNA in each half of the operator site.
  - 1st helix of HTH motif above the major groove,
  - 2nd helix fits into major groove.
- N-terminus of recognition helix closest to edges of bps.
- Side chains from HTH units make site-specific contacts with groups in major groove
- Each complex has extensive network of H-bonds between protein and DNA backbone.
  - Protein groups contacting backbone are Arg, Lys, polar side chains, and/or NH groups.

Cro from lambda



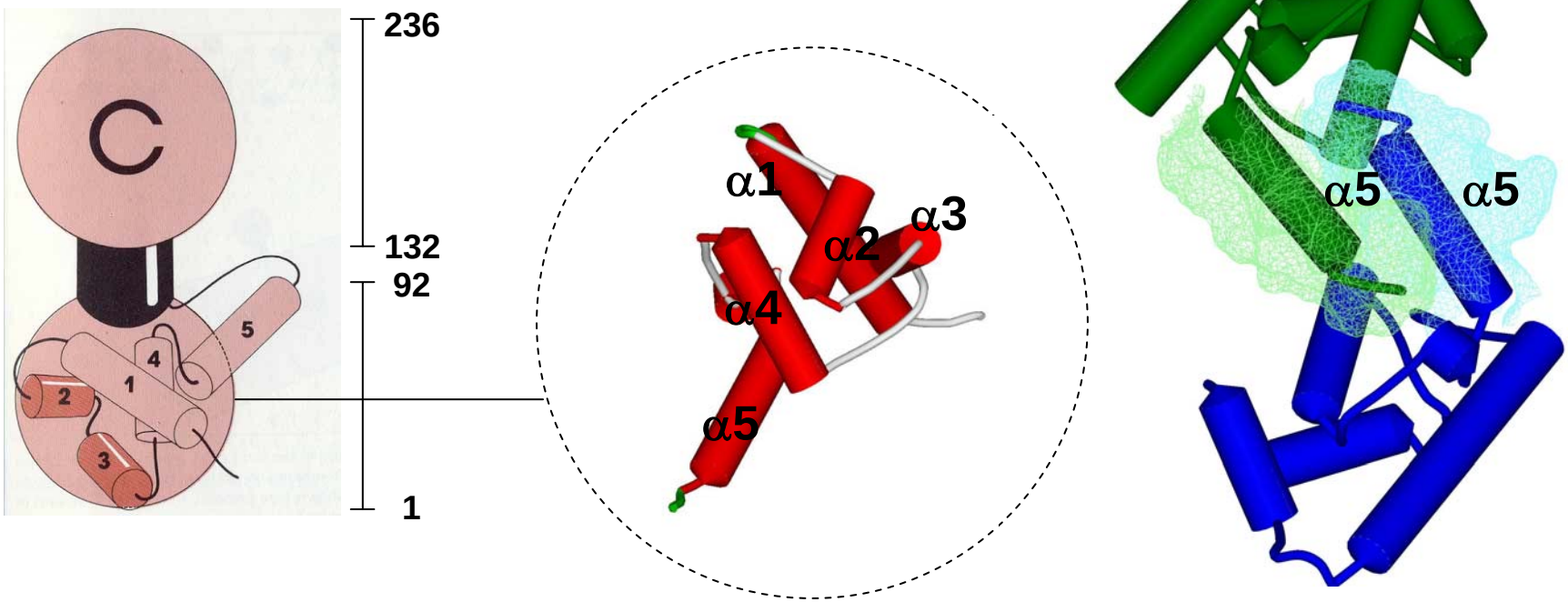


- **structure of Cro monomer (left) from the Matthews lab** consists of 66 amino acids and contains three alpha helices and three beta strands.
- HTH motif is 2nd and 3rd α-helices
- Helix 3 is the recognition helix.
- In the dimer there is an antiparallel alignment of the C-terminal strand of each subunit, giving a six-stranded antiparallel sheet.
- Binds as dimer to 17 bp pseudosymmetric operator
- Dimer structure suggested how cro binds to DNA:
  - two copies of the recognition helix (2nd helix of HTH motif) form ridges separated by 34 Å, .



Structure of DNA-binding domain of cI (λ repressor)

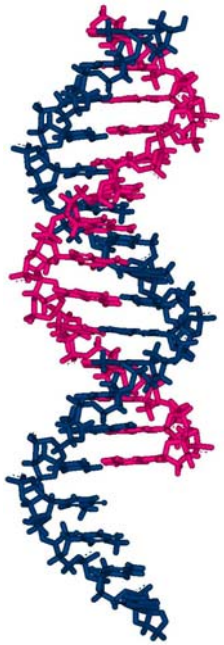
- a small, 236 amino acid, protein. It contains two domains:
  - an N-terminal DNA binding domain (amino acids 1-92) with a characteristic HTH DNA-binding motif.
  - a C-terminal oligomerization (dimerization) domain (amino acids 132-236)
- Structure of DNA-binding domain shows five α-helices.
- HTH motif is helices 2 and 3.
- Dimer contact is mediated by helix 5.



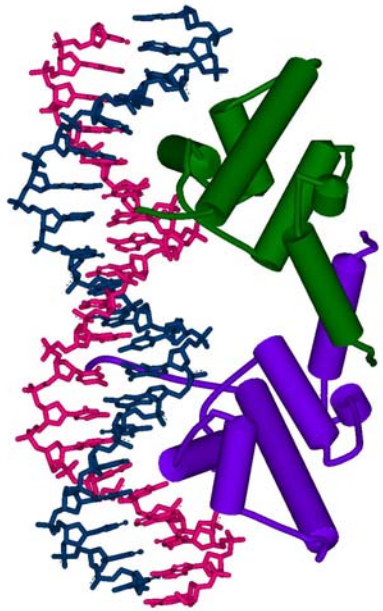
λ repressor and DNA contacts

1. Protein dimer symmetry axis coincides with approximate two-fold axis of DNA.
2. Recognition helices in dimer are on adjacent major grooves and separated by ~34 Å as seen in λ cro.
3. N-terminal arms reach around to back side of DNA to make contacts with the major grooves
4. DNA slightly distorted from B-DNA.

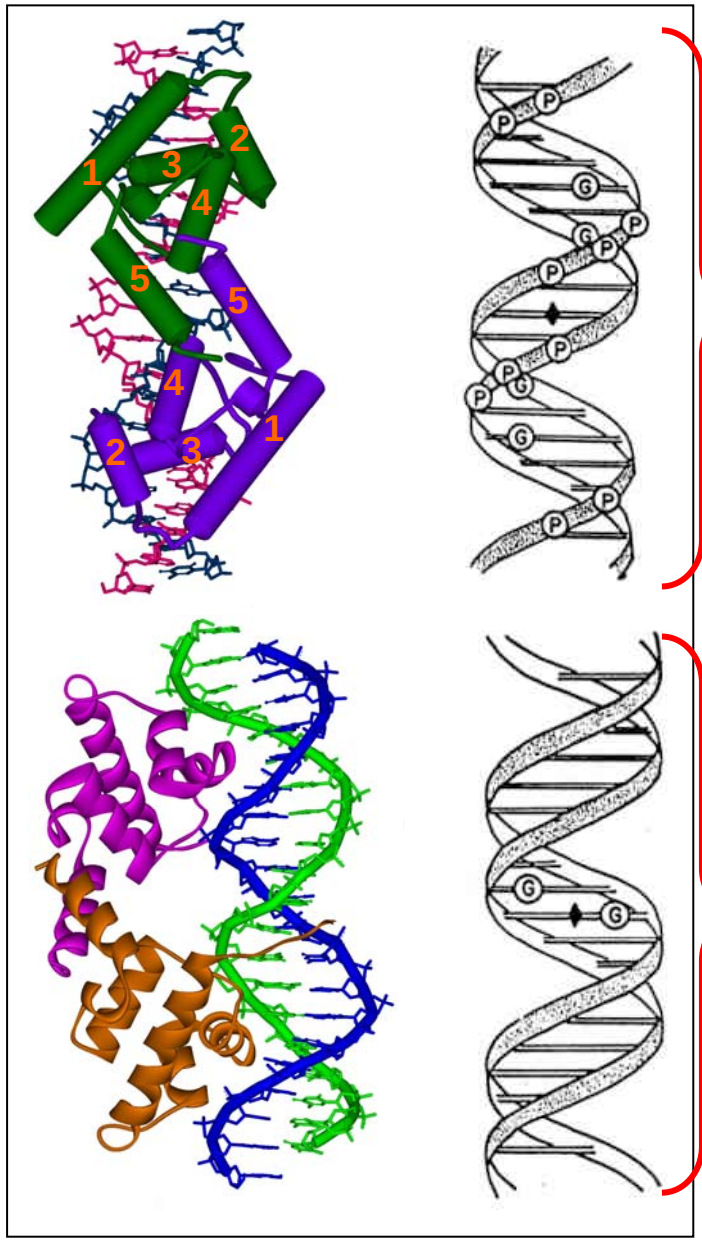
DNA only



λ repressor



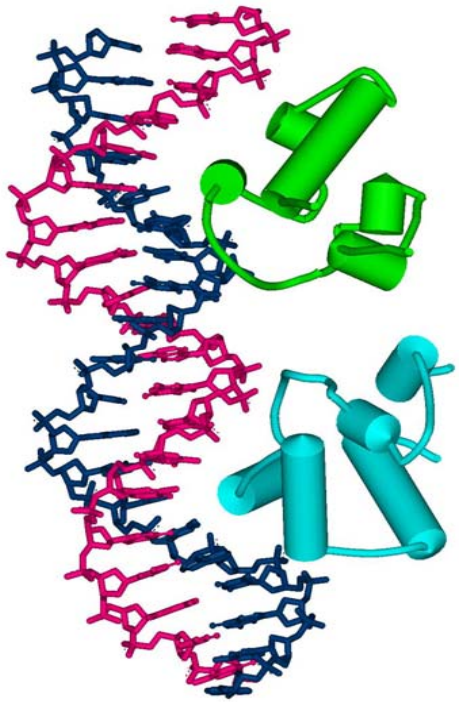
Beamer & Pabo, 1992,
J. Mol. Biol. 227: 177.



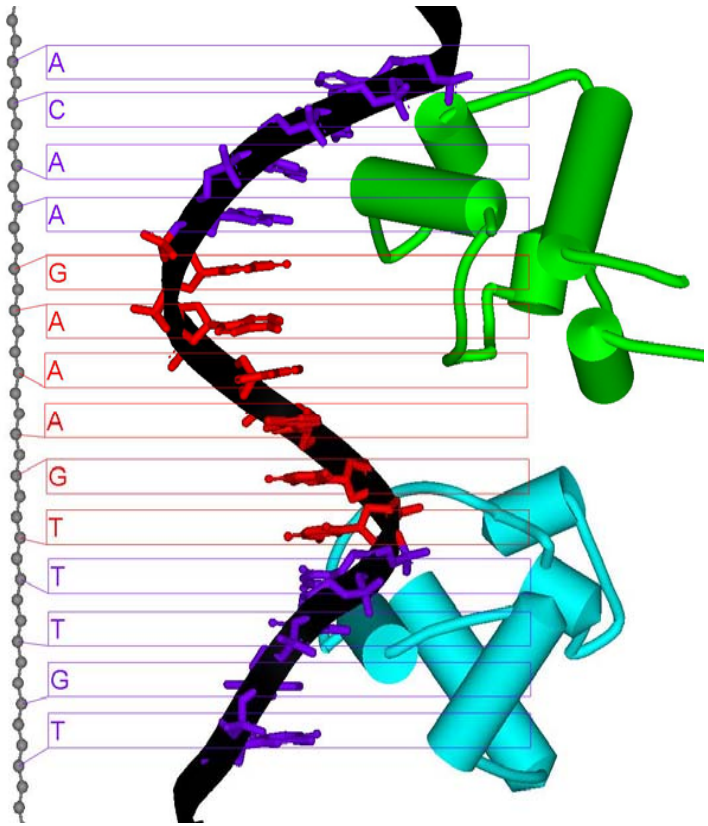
434 repressor

- 434 is a lysogenic phage and this is its repressor (i.e., cI)
- Each monomer within the dimer has a 4-helix cluster with helices 2 and 3 forming the HTH
- The HTH motifs are positioned at the N-termini of each subunit so that the recognition helices are
- positioned in successive turns of the double helix
- As for lambda repressor there are 10 bp or one turn between each recognition helix
- The B-form of DNA is twisted with the DNA bending around the protein in an arc (radius of ~ 65 A)
- The minor groove is compressed by 2.5 A near the center of the dimer

434 Cro



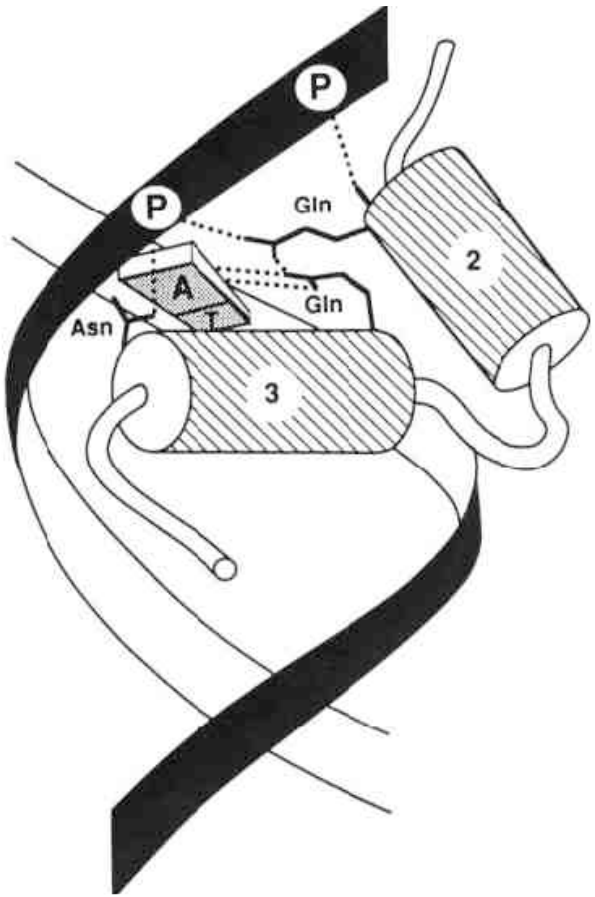
(HARRISON, PDB 2OR1)



Comparison of λ and 434 repressor/DNA complexes

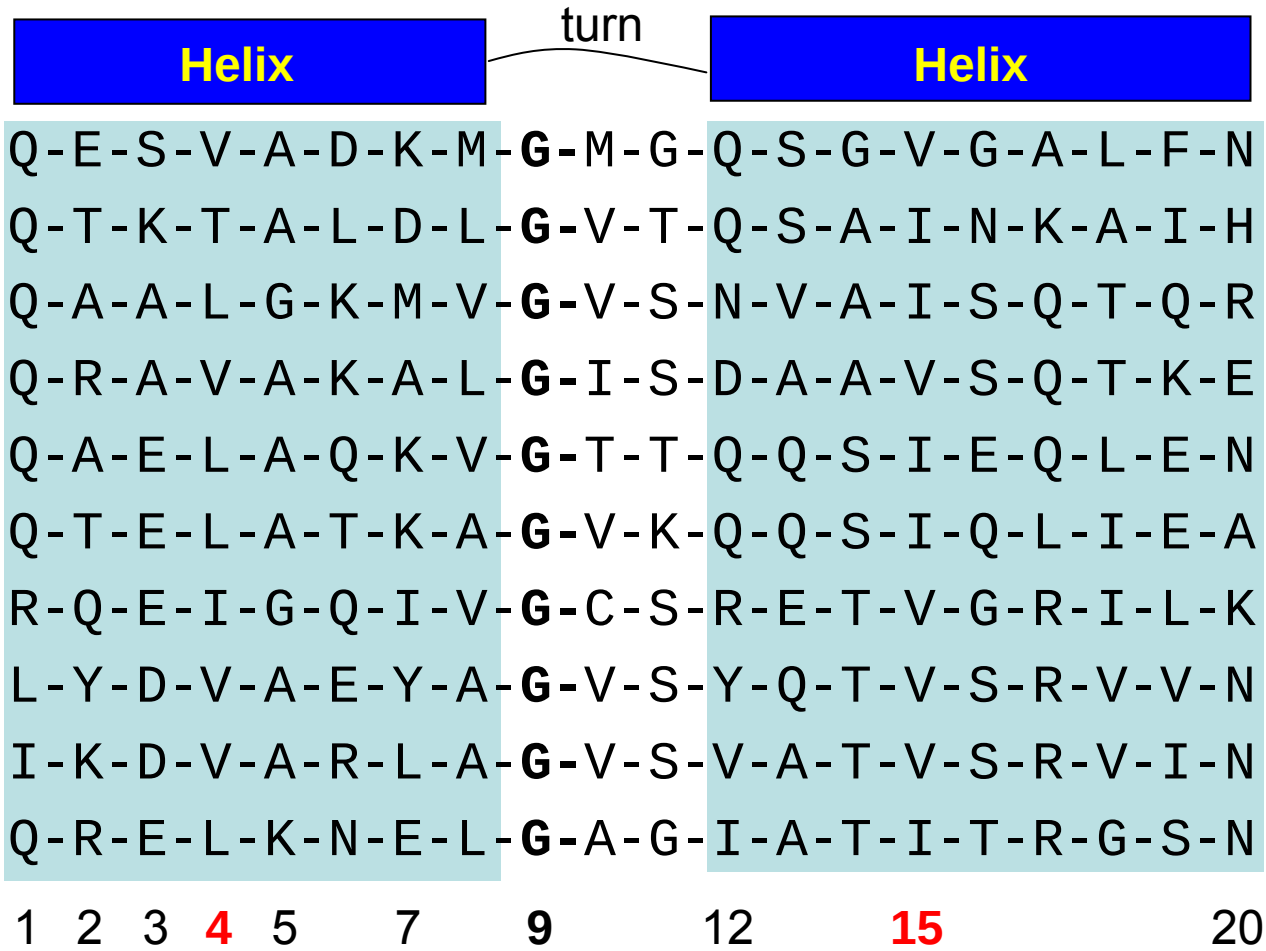
	33							41			44							52
Lambda	**Gln**	Glu	Ser	Val	**Ala**	Asp	**Lys**	Met	**Gly**	Met	Gly	**Gln**	Ser	Gly	Val	Gly	Ala	**Leu**	Phe	**Asn**
434	**Gln**	Ala	**Glu**	**Leu**	**Ala**	Gln	**Lys**	Val	**Gly**	Thr	Thr	**Gln**	Gln	Ser	Ile	Glu	Gln	**Leu**	Glu	**Asn**
		Helix							Turn					Helix						

- Binding at the recognition sequence occurs via interactions with both bases and the phosphodiester backbone of the DNA
- 1st residue of 1st helix (Gln) makes two H-bonds with DNA backbone on outer edge of operator.
- Aligns + helix dipole of 1st helix with phosphates.
- 1st residue of 2nd helix (Gln) contacts Adenine near end -- makes bidentate H-bonds as
- predicted. Gln is specific for A of A-T base pair.
- Glns at beginning of each helix H-bond to each other. Example of why there isn't a real code for protein/DNA recognition & why simple mutagenesis schemes to change specificity won't work.
- Asn at end of recognition helix H-bonds to same phosphodiester oxygen contacted by first Gln.



Conservation of HTH motifs

The structure of the HTH is important **NOT** the amino acid sequence although some important residues are:



9 must be gly; 5 cannot have a branched side chain; 4 and 15 are buried and cannot be charged

Hydrophobic bracing of HTH motifs

The helix-turn-helix motif showing the portions of conserved hydrophobic residues that form the contacts between the two helices and the conserved residues at the bend between the helices. The sequence and numbers are those of CAP. The most highly conserved residue (Gly177) is triply underlined, usually conserved doubly underlined and less conserved singly underlined. This structure is stabilized by the hydrophobic interactions between the side chains of the conserved Val183, Ile172, Val176 and Ile186. Stippled side chains are identical in CAP, gal and lac repressors while striped side chains are closely similar.

Conservation:

- C-α's superimpose within 0.7 to 1.0 Å rmsd
- 6/21 aa conserved in related sequences
- 4 residues make hydrophobic contacts between helices, preserving their orientation
- Conserved Gly important for bend between helices

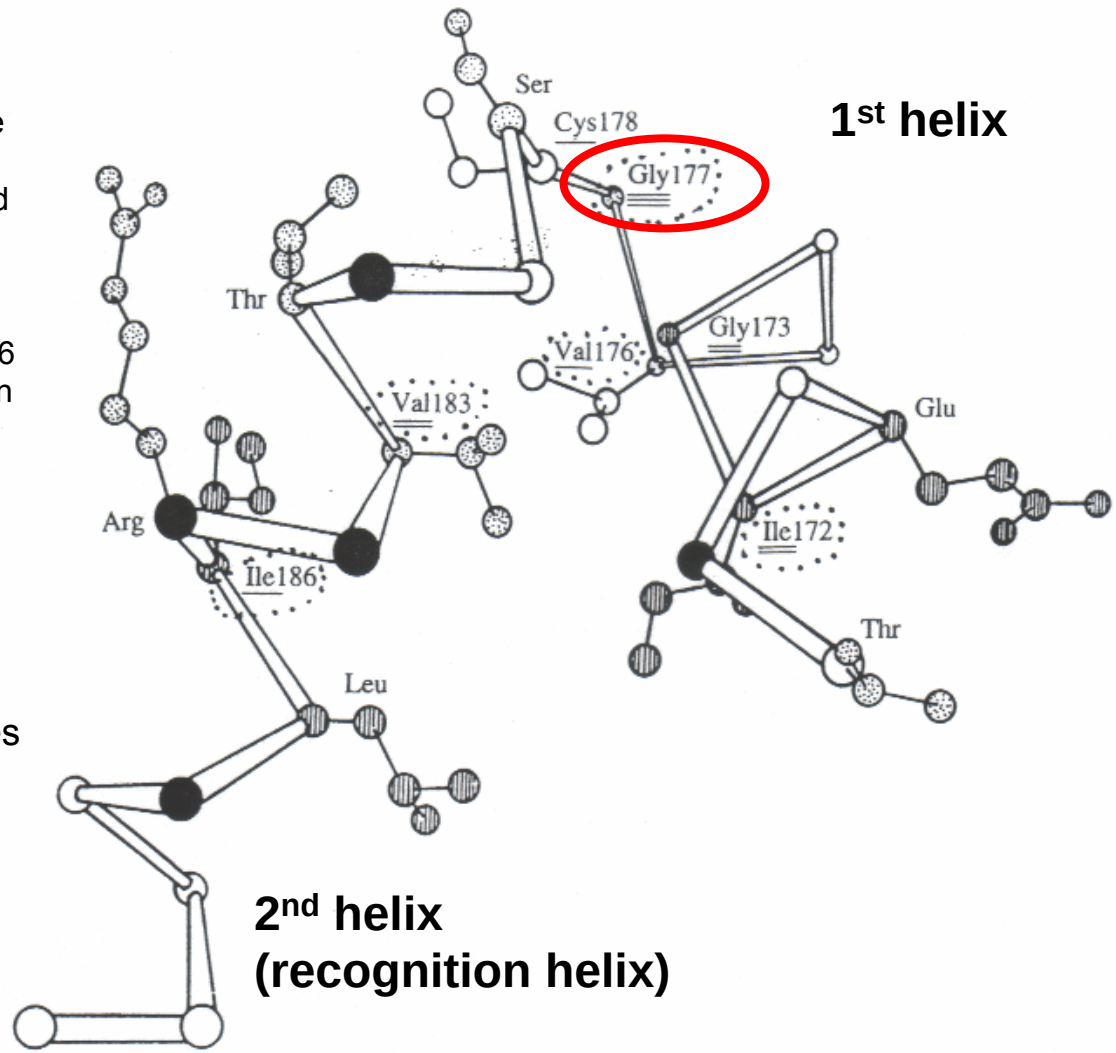
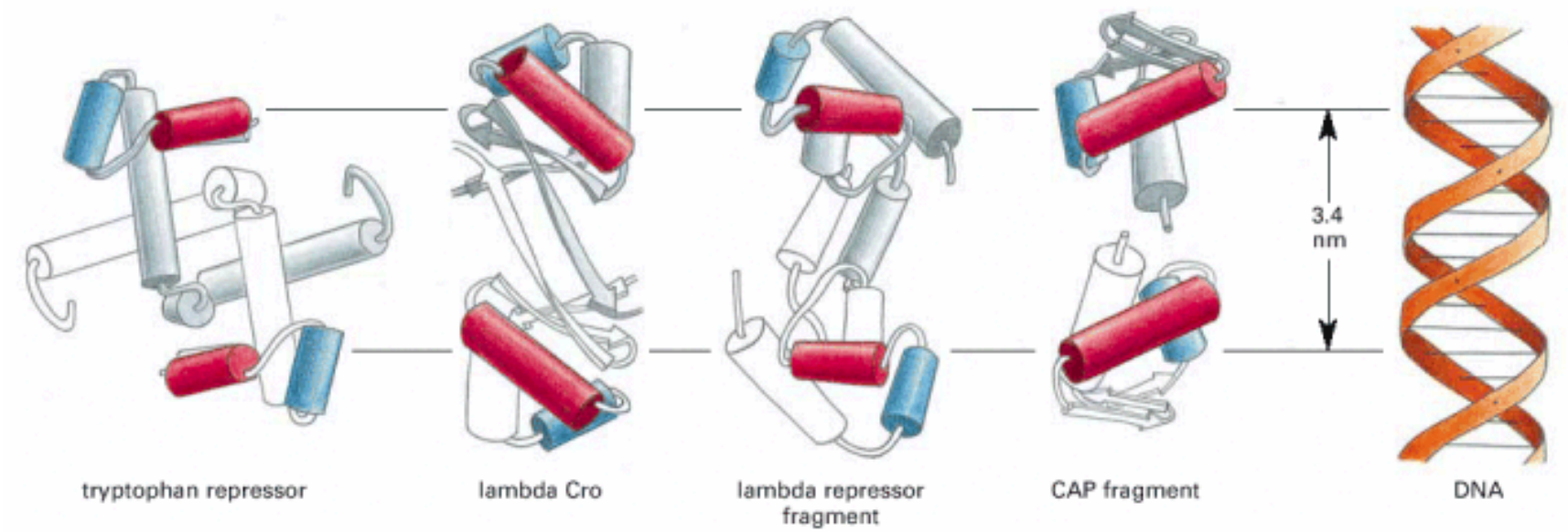


Figure from Steitz TA. Structural studies of protein-nucleic acid interaction: the sources of sequence-specific binding. Q Rev Biophys. 1990 Aug;23(3):205-80. Review.

HTH motif conserved, but relative orientation of motif and DNA is not conserved

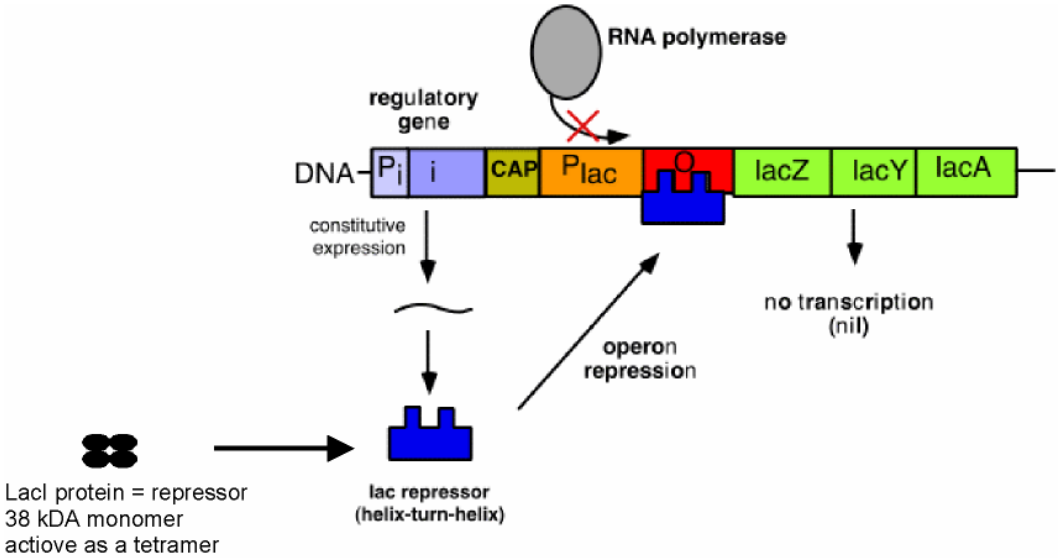
•Outside the helix-turn-helix region, the structure of the various proteins that contain this motif can vary enormously
•each protein "presents" its helix-turn-helix motif to the DNA in a unique way, a feature thought to enhance the versatility of the helix-turn-helix motif by increasing the number of DNA sequences that the motif can be used to recognize.
•parts of the polypeptide chain outside the helix-turn-helix domain also make important contacts with the DNA, helping to fine-tune the interaction.



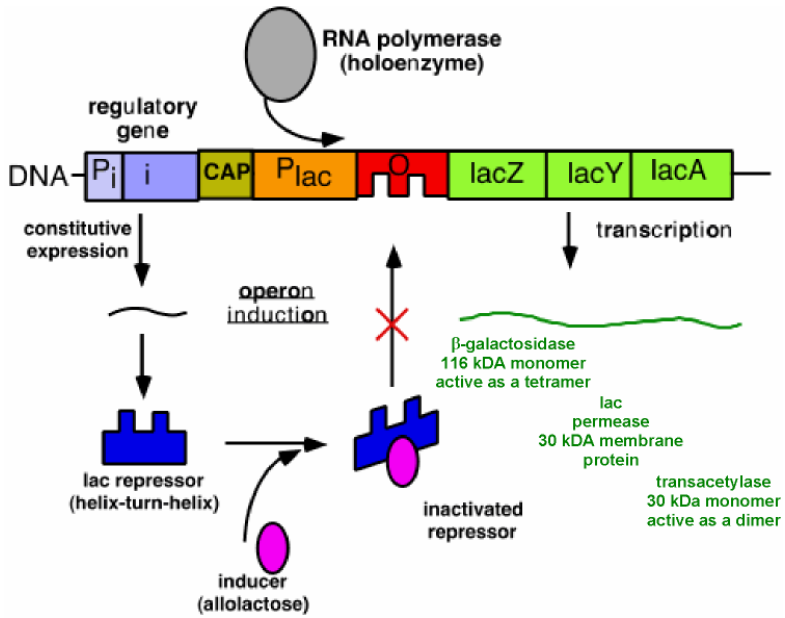
helix-turn-helix DNA-binding proteins. All of the proteins bind DNA as dimers in which the two copies of the recognition helix *(red cylinder)* are separated by exactly one turn of the DNA helix (3.4 nm). The other helix of the helix-turn-helix motif is colored *blue*. The lambda repressor and Cro proteins control bacteriophage lambda gene expression, and the tryptophan repressor and the catabolite activator protein (CAP) control the expression of sets of *E. coli* genes.

Expression from the lac operon

"off configuration"

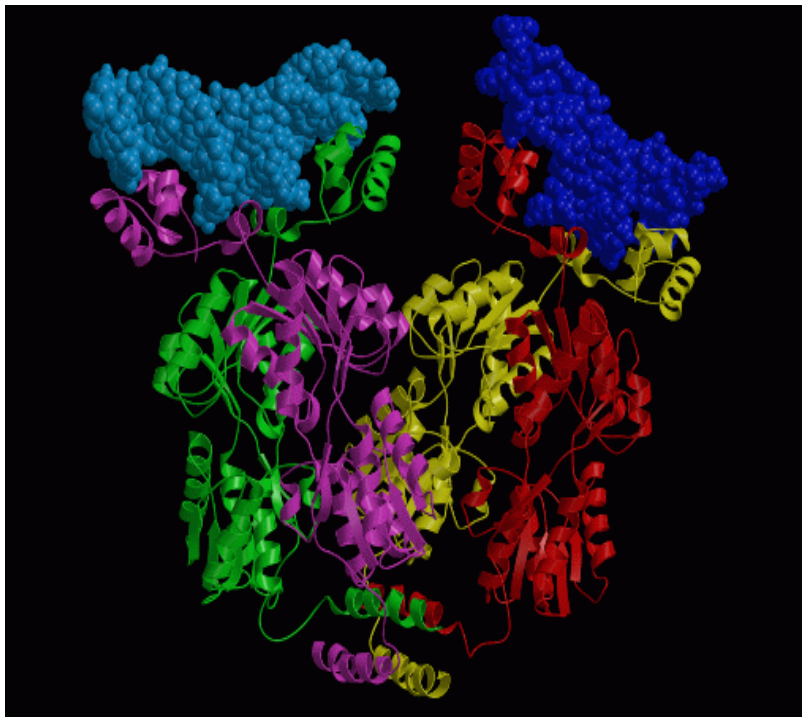


"on configuration"

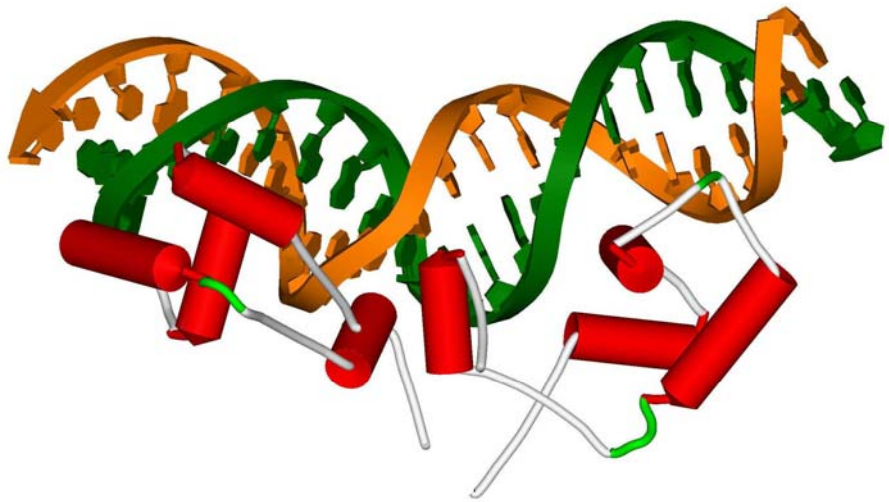


The lac repressor contains HTH motifs

- The lac repressor is a protein of 360 amino acids that forms a homotetramer
- It has five distinct fragments: four NH_2-terminal fragments and a COOH-terminal tetrameric core
- The NH_2-terminal fragments (each 60 aa), bind in a specific manner to the operator.
- The COOH-terminal tetrameric core binds the inducer

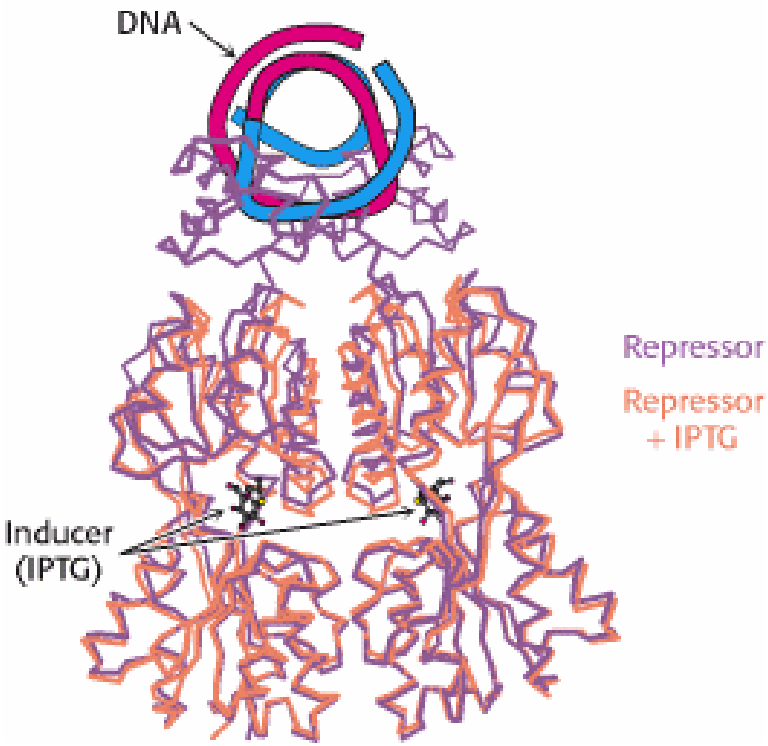


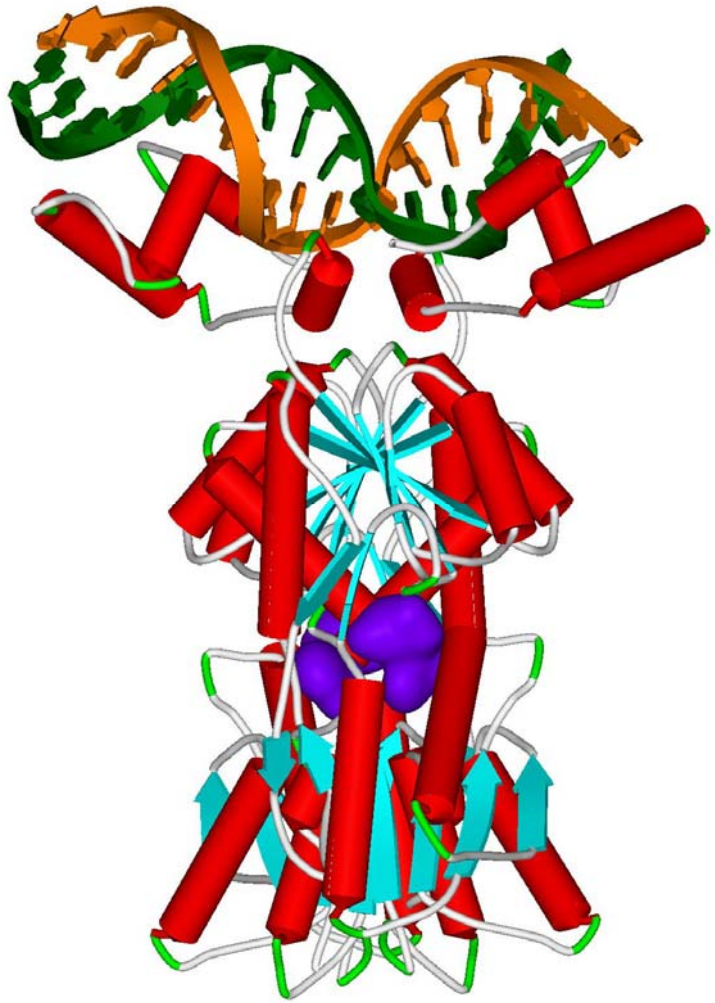
Lac repressor bound to operator DNA



Lac repressor contains HTH

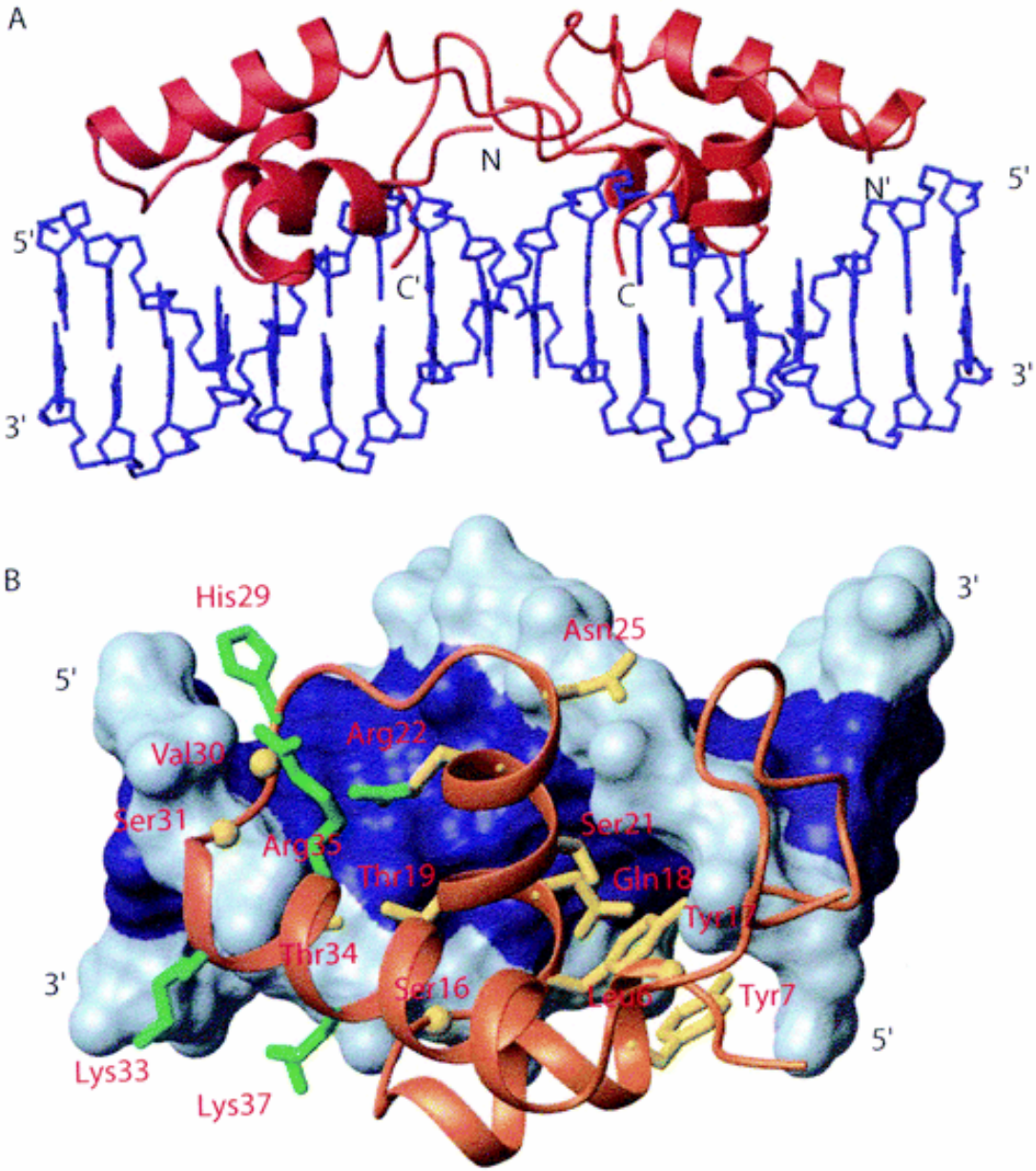
The lac repressor and allosteric regulation of DNA binding





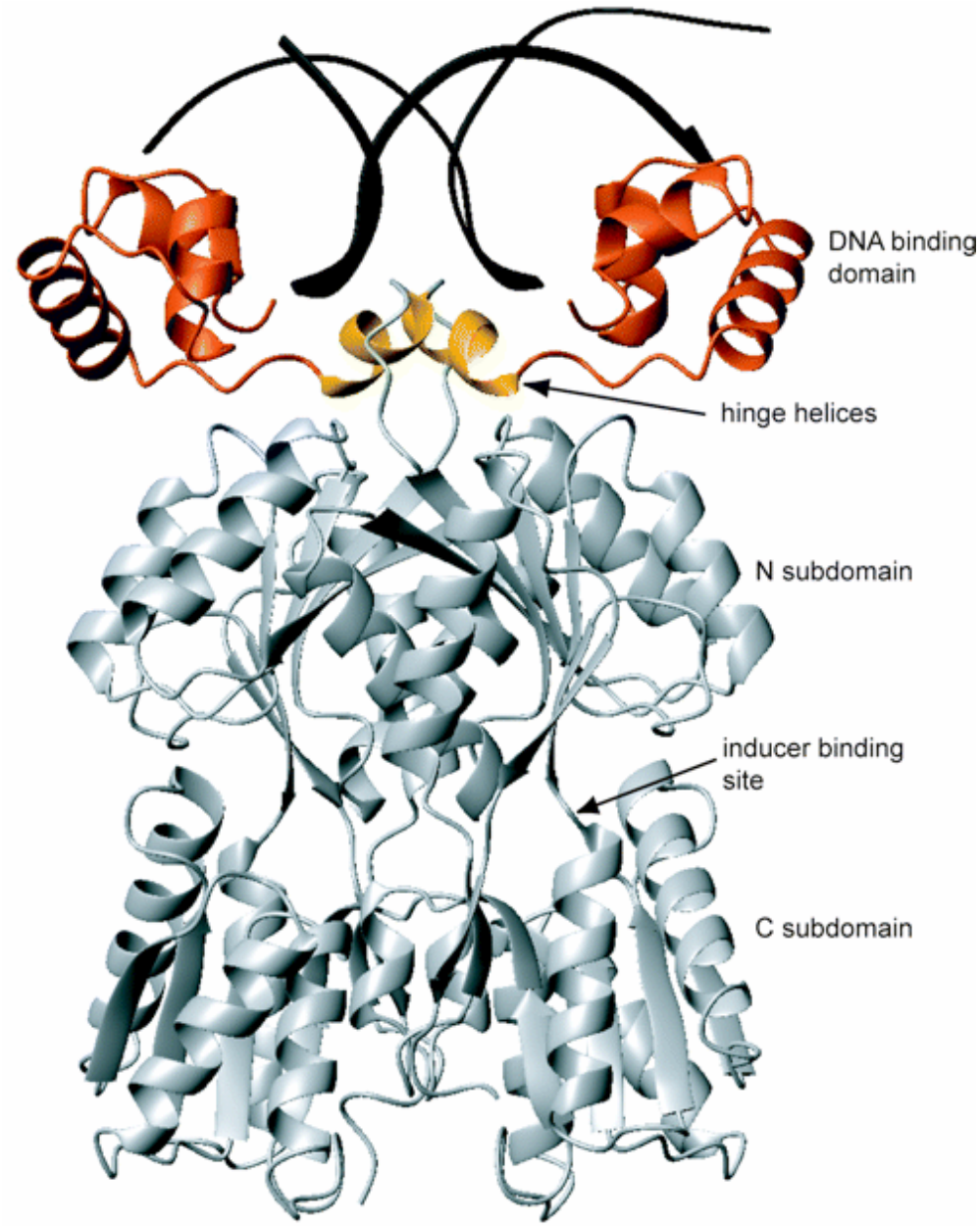
Effects of IPTG On *LAC* Repressor Structure. The structure of the *lac* repressor bound to the inducer isopropylthiogalactoside (IPTG), shown in orange, is superimposed on the structure of the *lac* repressor bound to DNA, shown in purple. The binding of IPTG induces structural changes that alter the relation between the two DNA-binding domains so that they cannot interact effectively with DNA. The DNA-binding domains of the *lac* repressor bound to IPTG are not shown, because these regions are not well ordered in the crystals studied

Lac repressor bound nonspecifically



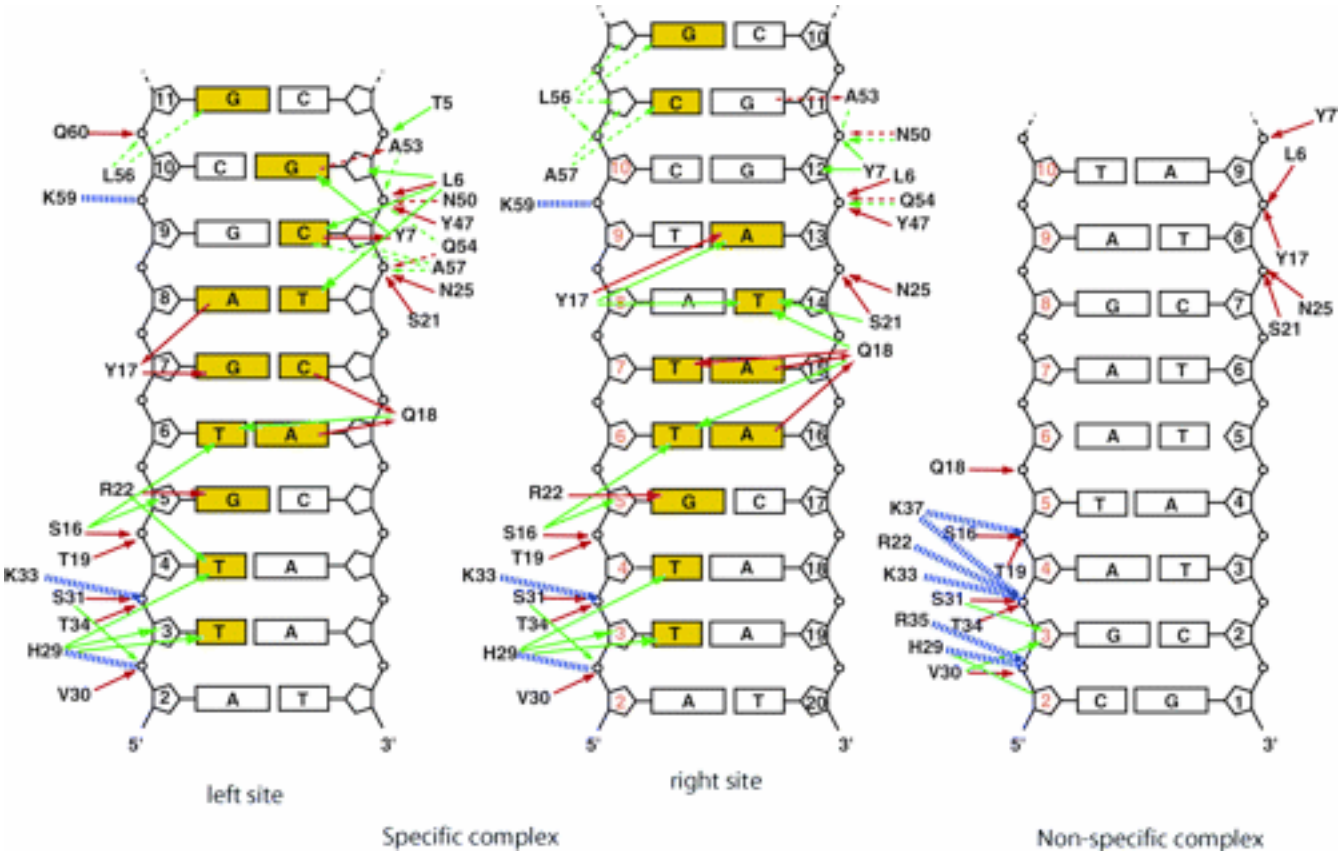
Structure of the dimeric lac HP62 complexed to nonspecific DNA. (A) The lowest-energy structural conformer. The C-terminus (residues 50-62) of each of the dimers is unstructured. The protein backbone is depicted in red whereas the DNA heavy atoms are depicted in blue. (B) Schematic diagram of the structure of the left site of the complex. A ribbon diagram of the protein is shown bound to the solvent-accessible surface of the DNA. The major and the minor grooves of the DNA are dark blue, and the ribose phosphate backbone is light blue. Residues involved in intermolecular hydrogen bonding and Coulombic interactions are shown in yellow and green, respectively. Backbone amides are indicated with spheres.

Lac repressor bound specifically



Crystal structure of a dimeric lac repressor bound to the SymL DNA operator. The intact repressor is composed of two such units that are assembled through the tetramerization domain (not shown). The DNA-binding domain and the hinge helices are red and orange, respectively. The inducer binding pocket is located at the junction of the N- and C-subdomains.

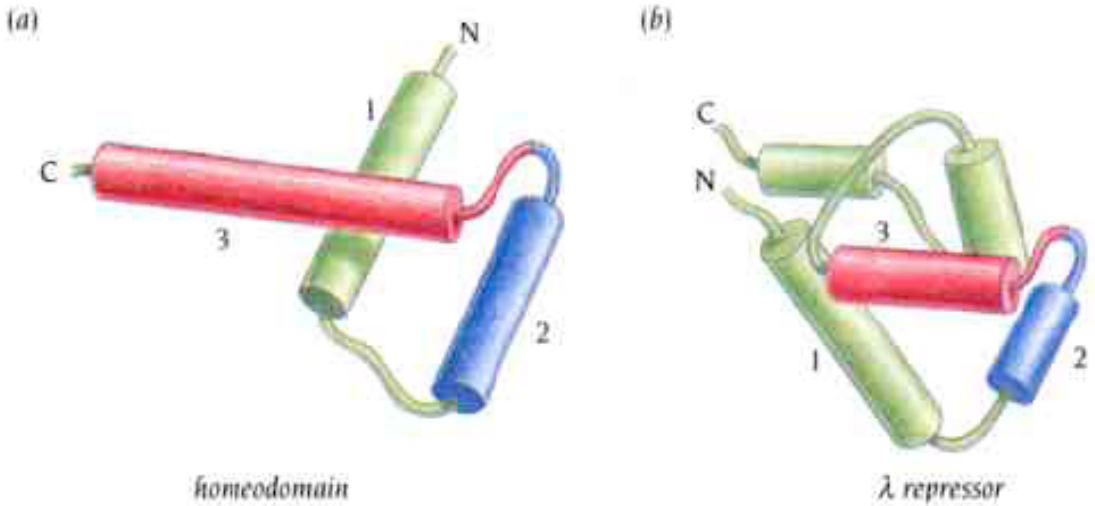
Comparison of specific and nonspecific binding of Lac repressor



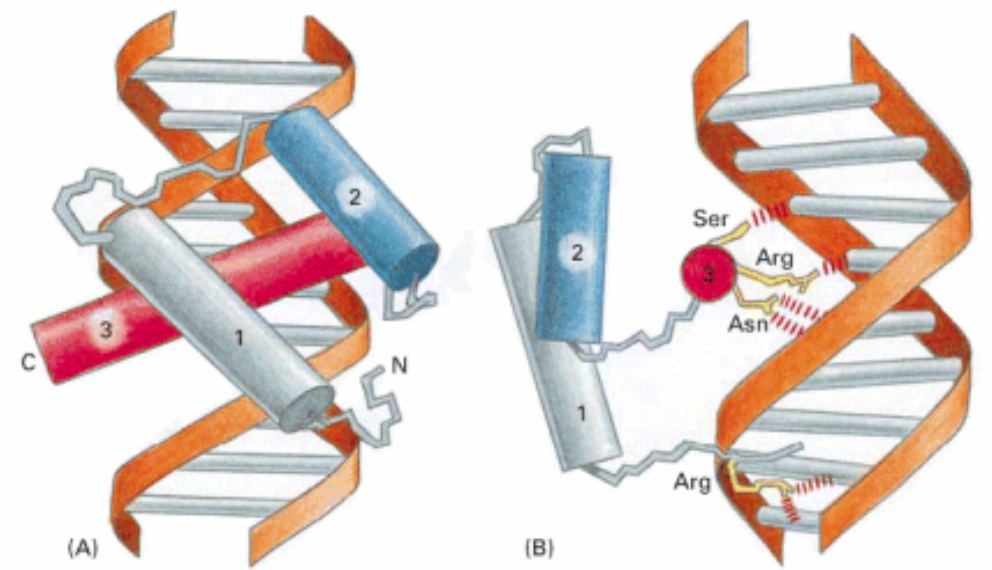
Comparison of the protein-DNA contacts in the two half-sites of the dimeric lac HP62-O1 operator complex and the nonspecific complex. The bases that are specifically recognized by the lac repressor are yellow. The solid and dashed lines indicate interactions in the major and minor grooves, respectively. Red, green, and dashed blue lines indicate hydrogen bonding, hydrophobic, and electrostatic contacts, respectively. In the right site of the specific complex and in the nonspecific complex, the numbering displayed in orange denotes the symmetry related base pairs in the left half-site of the specific complex.

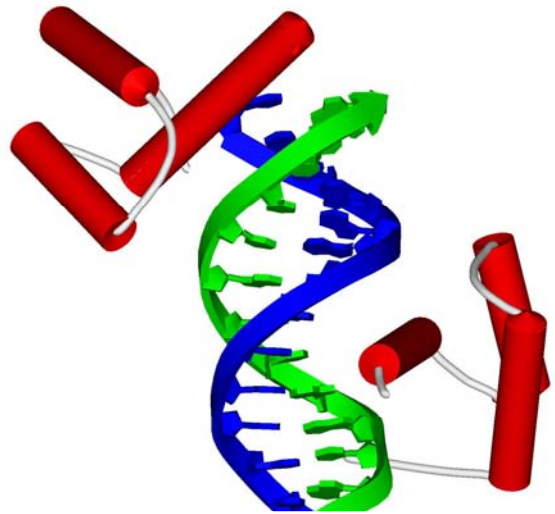
Homeodomains - eukaryotic HTH motifs

- Special class of helix-turn-helix motif
- Discovered in homeotic mutants of Drosophila - Mutations alter how body parts are assembled
- Now found in large family of proteins that regulate transcription in eukaryotes.
- Primary sequences highly conserved.
- Homeodomain proteins provide a system with a highly conserved sequence of 60 amino acids spanning across most eukaryotic organisms that fulfill a basic regulatory function.
- This high degree of conservation makes them an ideal model system for studies attempting to elucidate specific protein-DNA interactions.
- Proteins have nearly identical sequence of 60 amino acids
- DNA sequence to which they bind is called homeodomain
- The general structure of a homeodomain can be described as Helix-Loop-Helix-Turn-Helix, although in Antennapedia the third helix is generally considered to be composed of two helices.
- One might correctly identify these as Helix-Turn-Helix proteins, although the structure of a homeodomain is more stable and retains its function when isolated.



How does a homeodomain bind to its specific DNA sequence





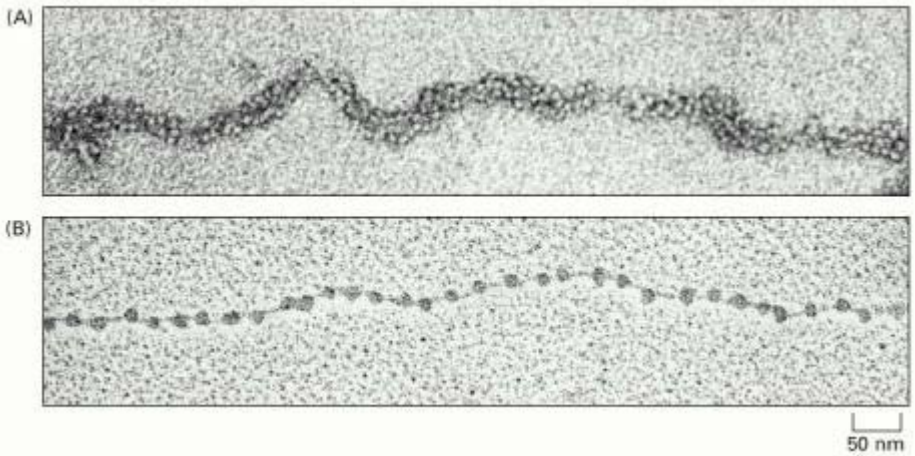
Two different views of the same structure are shown. (A) The homeodomain is folded into three α helices, which are packed tightly together by hydrophobic interactions. The part containing helix 2 and 3 closely resembles the helix-turn-helix motif. (B) The recognition helix (helix 3, *red)* makes important contacts with the major groove of DNA. The asparagine (Asn) of helix 3, for example, contacts an adenine. Nucleotide pairs are also contacted in the minor groove by a flexible arm attached to helix 1. The homeodomain shown here is from a yeast gene regulatory protein, but it closely resembles homeodomains from many eucaryotic organisms. (Adapted from C. Wolberger et al., *Cell* 67:517–528, 1991.)

3HDD engrailed homeodomain protein

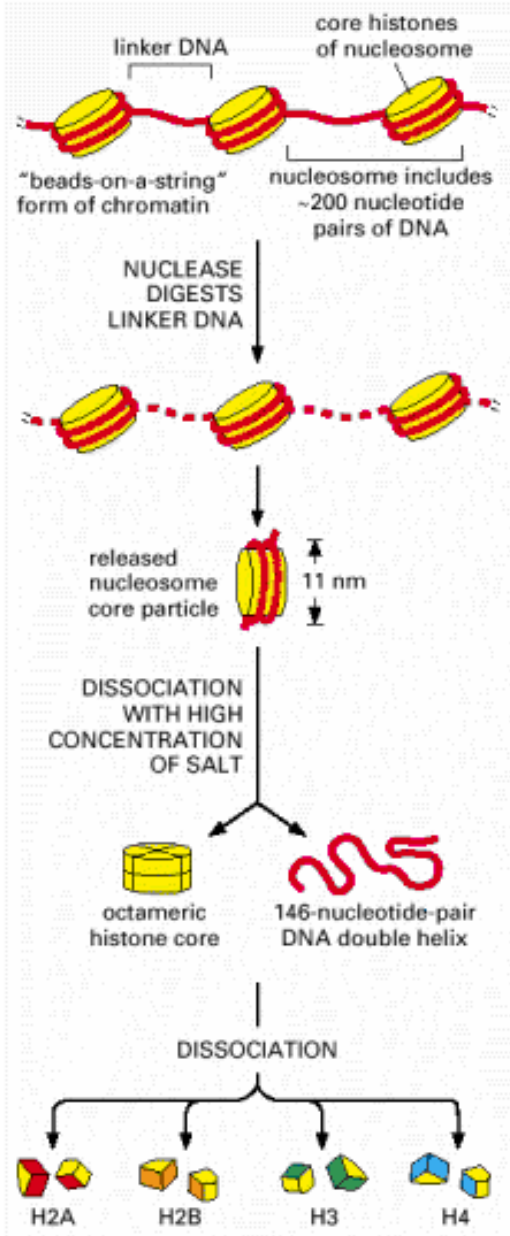
- Protein binds as two monomers to two sites on oligo (one specific, one non-specific).
- No protein-protein contacts.
- HTH motif: superimposes on prokaryotic repressors to 0.84 Å rmsd.
- N-terminal arm (residues 3-9) fits in minor groove.
- Three helices. Helix 1 and 2 are antiparallel, no
- contacts with DNA.
- Helix 3 is almost perpendicular to first two helices. Fits in major groove.

Structural organization of the nucleosome.

Nucleosomes as seen in the electron microscope. (A) Chromatin isolated directly from an interphase nucleus appears in the electron microscope as a thread 30 nm thick. (B) This electron micrograph shows a length of chromatin that has been experimentally unpacked, or decondensed, after isolation to show the nucleosomes.

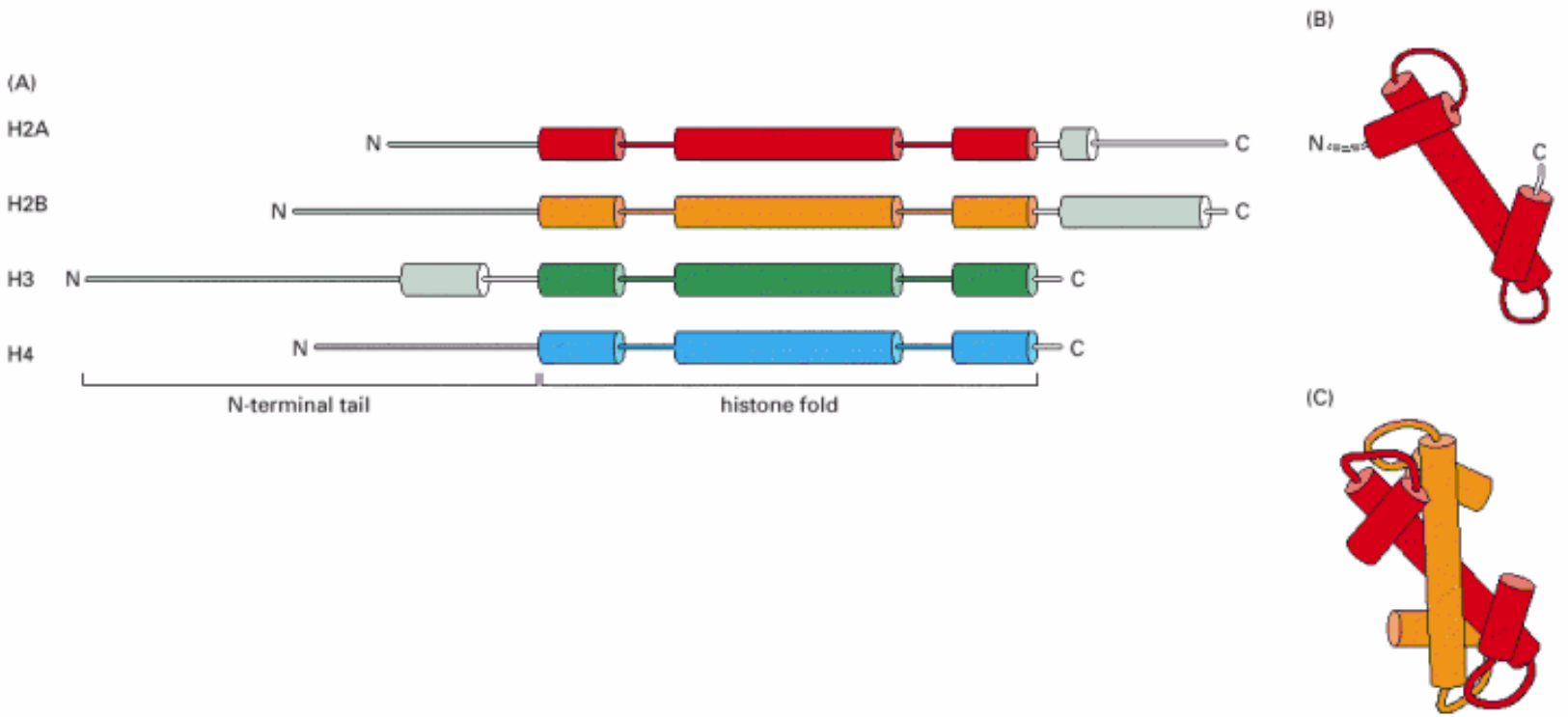


A nucleosome contains a protein core made of eight histone molecules. As indicated on the right, the nucleosome core particle is released from chromatin by digestion of the linker DNA with a nuclease, e.g., micrococcal nuclease. (The nuclease can degrade the exposed linker DNA but cannot attack the DNA wound tightly around the nucleosome core.) After dissociation of the isolated nucleosome into its protein core and DNA, the length of the DNA that was wound around the core can be determined. This length of 146 nucleotide pairs is sufficient to wrap 1.65 times around the histone core.



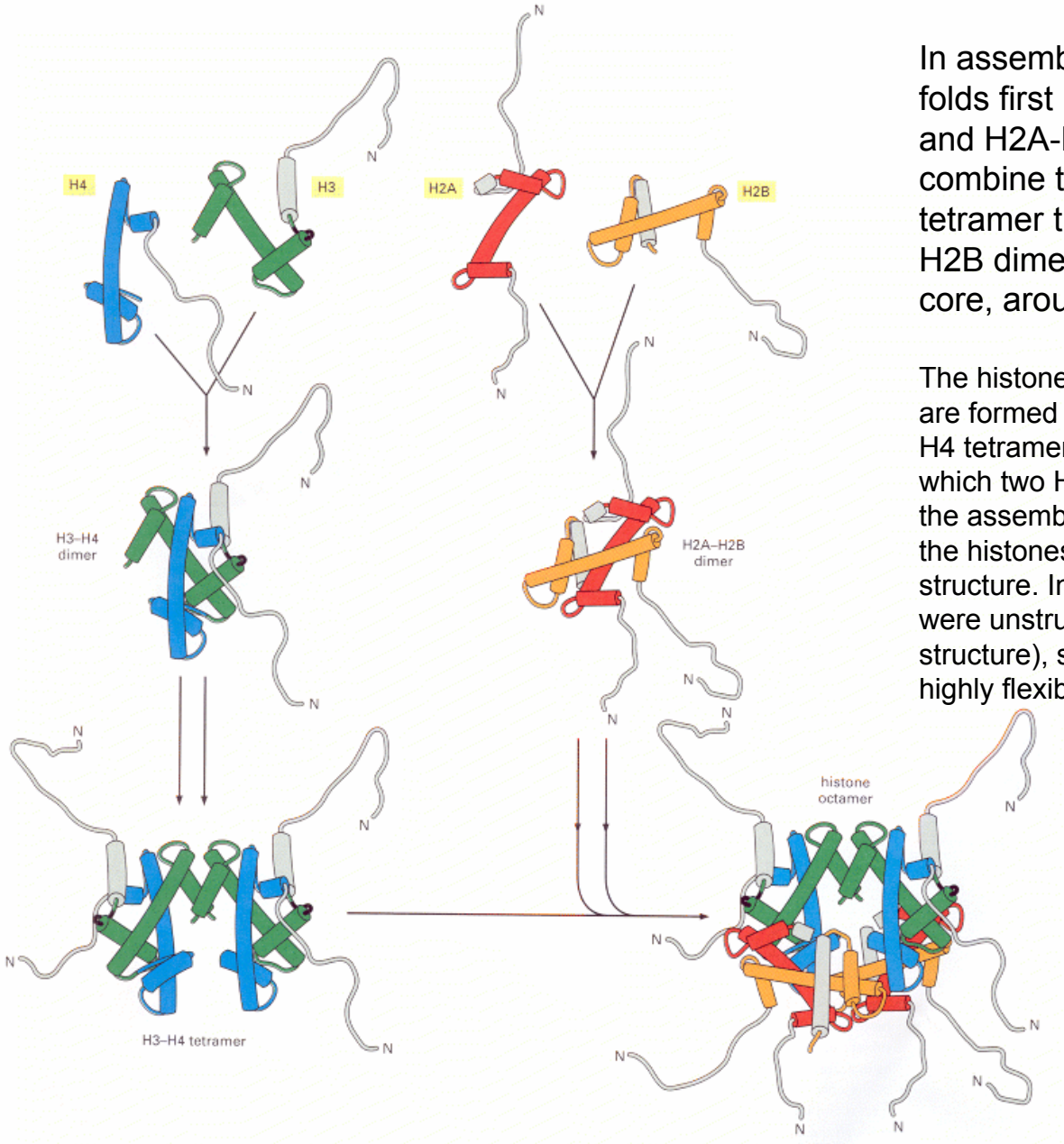
Histone proteins

There are 4 histone proteins that make up the core – H2A, H2B, H3 and H4
these are relatively small proteins (102–135 amino acids),
they share a structural motif, known as the *histone fold*, formed from three α helices connected by two loops.



The overall structural organization of the core histones. (A) Each of the core histones contains an N-terminal tail, which is subject to several forms of covalent modification, and a histone fold region, as indicated. (B) The structure of the histone fold, which is formed by all four of the core histones. (C) Histones 2A and 2B form a dimer through an interaction known as the "handshake." Histones H3 and H4 form a dimer through the same type of interaction.

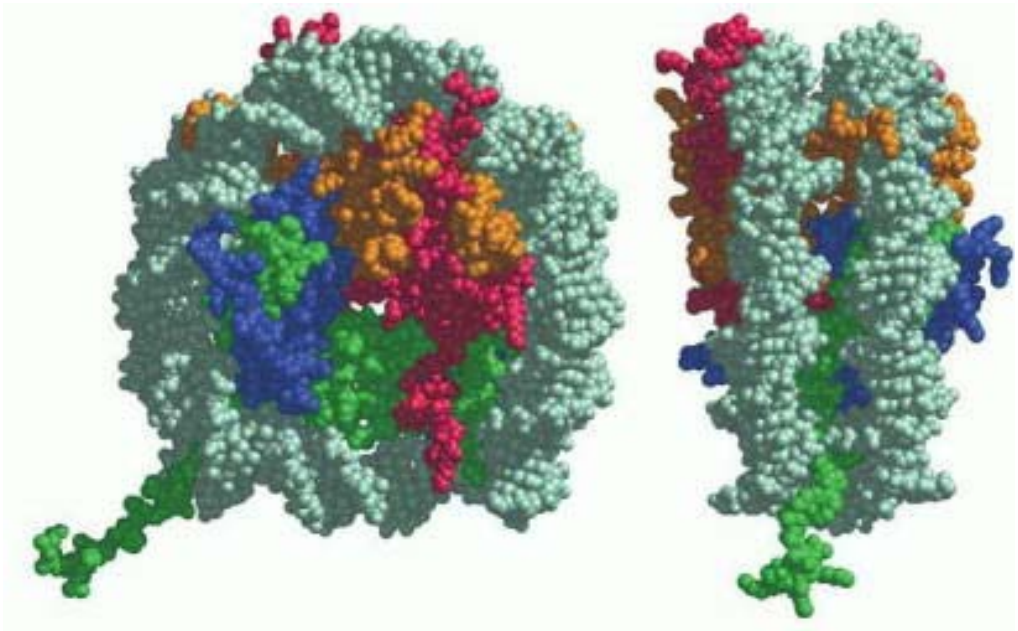
The assembly of a histone octamer

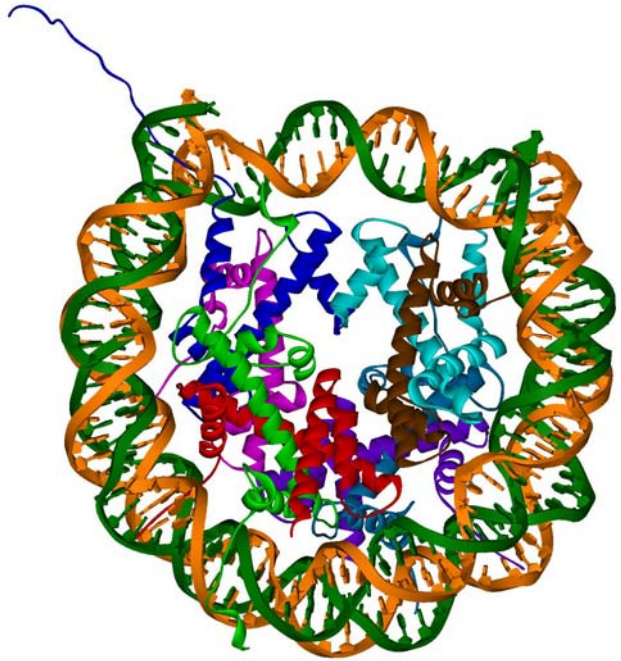


In assembling a nucleosome, the histone folds first bind to each other to form H3–H4 and H2A-H2B dimers, and the H3–H4 dimers combine to form tetramers. An H3–H4 tetramer then further combines with two H2A-H2B dimers to form the compact octamer core, around which the DNA is wound.

The histone H3–H4 dimer and the H2A-H2B dimer are formed from the handshake interaction. An H3-H4 tetramer forms the scaffold of the octamer onto which two H2A-H2B dimers are added, to complete the assembly. Note that all eight N-terminal tails of the histones protrude from the disc-shaped core structure. In the x-ray crystal most of the histone tails were unstructured (and therefore not visible in the structure), suggesting that their conformations are highly flexible.

The Structure of the Nucleosome Core Particle Reveals How DNA Is Packaged





•The structure contains a disc-shaped histone core around which the DNA was tightly wrapped 1.65 turns in a left-handed coil

•The interface between DNA and histone is extensive: 142 hydrogen bonds are formed between DNA and the histone core in each nucleosome.

•Nearly half of these bonds form between the amino acid backbone of the histones and the phosphodiester backbone of the DNA. Numerous hydrophobic interactions and salt linkages also hold DNA and protein together in the nucleosome.

• all the core histones are rich in lysine and arginine and their positive charges can effectively neutralize the negatively charged DNA backbone. These numerous interactions explain in part why DNA of virtually any sequence can be bound on a histone octamer core. The path of the DNA around the histone core is not smooth; rather, several kinks are seen in the DNA, as expected from the non-uniform surface of the core.

•In addition to its histone fold, each of the core histones has a long N-terminal amino acid “tail”, which extends out from the DNA-histone core These histone tails are subject to several different types of covalent modifications, which control many aspects of chromatin structure.

The Positioning of Nucleosomes on DNA Is Determined by Both DNA Flexibility and Other DNA-bound Proteins

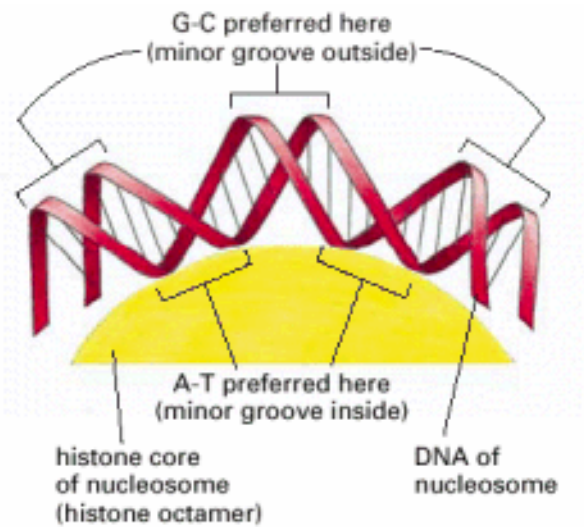
Two main influences determine where nucleosomes form in the DNA:

- the **difficulty of bending the DNA double helix** requires substantial compression of the minor groove of the DNA helix.

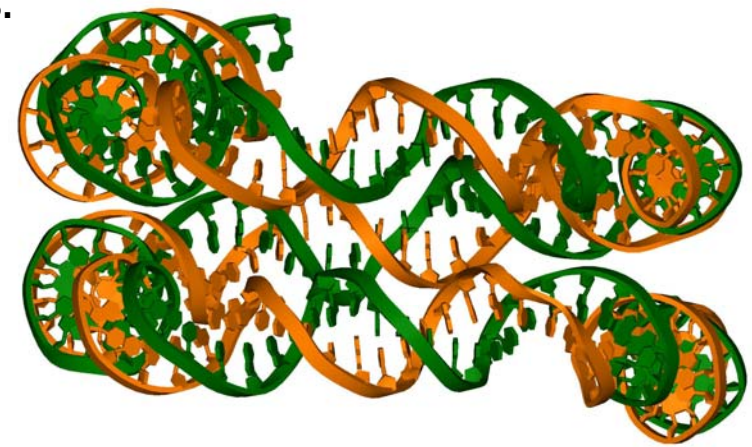
Because A-T-rich sequences in the minor groove are easier to compress than G-C-rich sequences, each histone octamer tends to position itself on the DNA so as to maximize A-T-rich minor grooves on the inside of the DNA coil

- the **presence of other tightly bound proteins on the DNA**.
  i. Some bound proteins favor the formation of a nucleosome adjacent to them.
  ii. Others force the nucleosomes to assemble at positions between them.
  iii. some proteins can bind tightly to DNA even when their DNA-binding site is part of a nucleosome.

A.



B.



The bending of DNA in a nucleosome. (A). The DNA helix makes 1.65 tight turns around the histone octamer. This diagram is drawn approximately to scale, illustrating how the minor groove is compressed on the inside of the turn. Owing to certain structural features of the DNA molecule, A-T base pairs are preferentially accommodated in such a narrow minor groove. (B). The DNA from a nucleosome