



PART 2

Structure, Function, and Engineering

DNA Recognition in Procaryotes by Helix-Turn-Helix Motifs

DNA Recognition in Procaryotes by Helix-Turn-Helix Motifs

- Proteins that regulate transcription of DNA recognize specific DNA sequences through discrete **DNA-binding domains** within their polypeptide chain.
- These domains are in general relatively small, **less than 100 amino acid residues**.
- Many procaryotic DNA-binding domains contain a **helix-turn-helix motif** that recognizes and binds specific regulatory regions of DNA.
- We will discuss the functional properties of the helix-turn-helix motif and the way this motif is integrated into structurally different DNA-binding domains of procaryotic repressors and activators.

DNA Recognition in Procaryotes by Helix-Turn-Helix Motifs

- The mechanism of action of bacterial and bacteriophage repressors and activators is in principle very simple.
- Repressors bind tightly to the DNA at the promoter of a structural gene, preventing the RNA polymerase from gaining access and hence blocking the initiation of transcription.
- Activators work by binding next to the promoter and helping the RNA polymerase to bind to the adjacent promoter, thereby increasing the rate of transcription of the gene.
- However, the subtle regulation of these binding interactions can be quite complex.

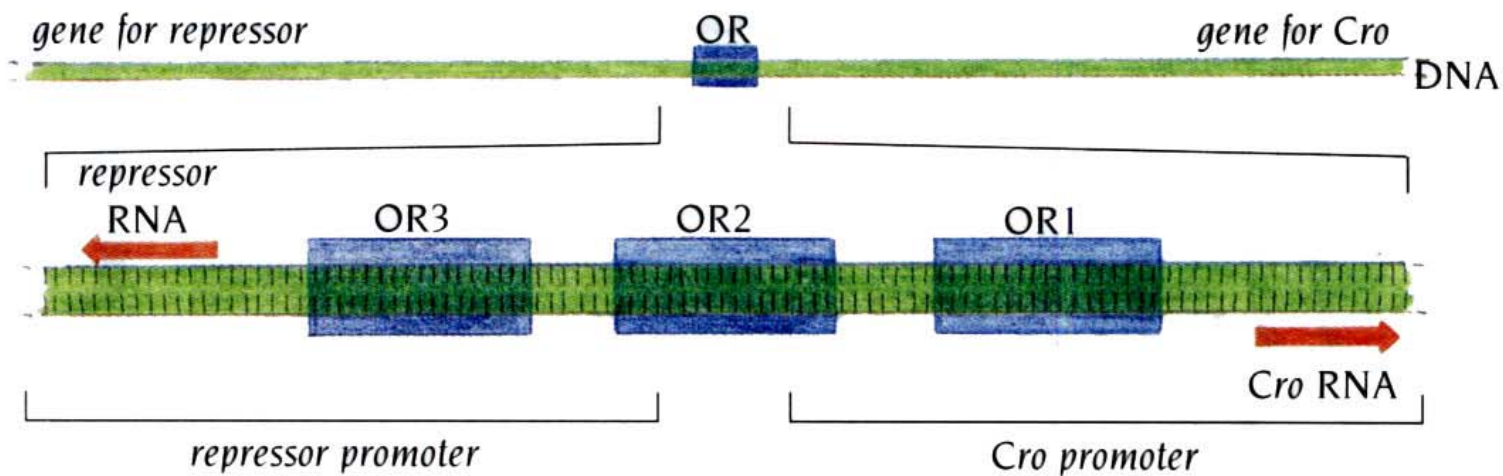
DNA Recognition in Procaryotes by Helix-Turn-Helix Motifs

- The most thoroughly studied procaryotic regulator proteins belong to bacteriophage λ and related phages.
- These phages produce two regulatory proteins, namely, repressor and Cro.
- Repressor and Cro proteins operate a switch between two states of λ phage replication.
- Both proteins contain the helix-turn-helix motif.
- The two bacteriophage proteins are amongst the best-understood DNA regulatory proteins.

A molecular mechanism for gene control

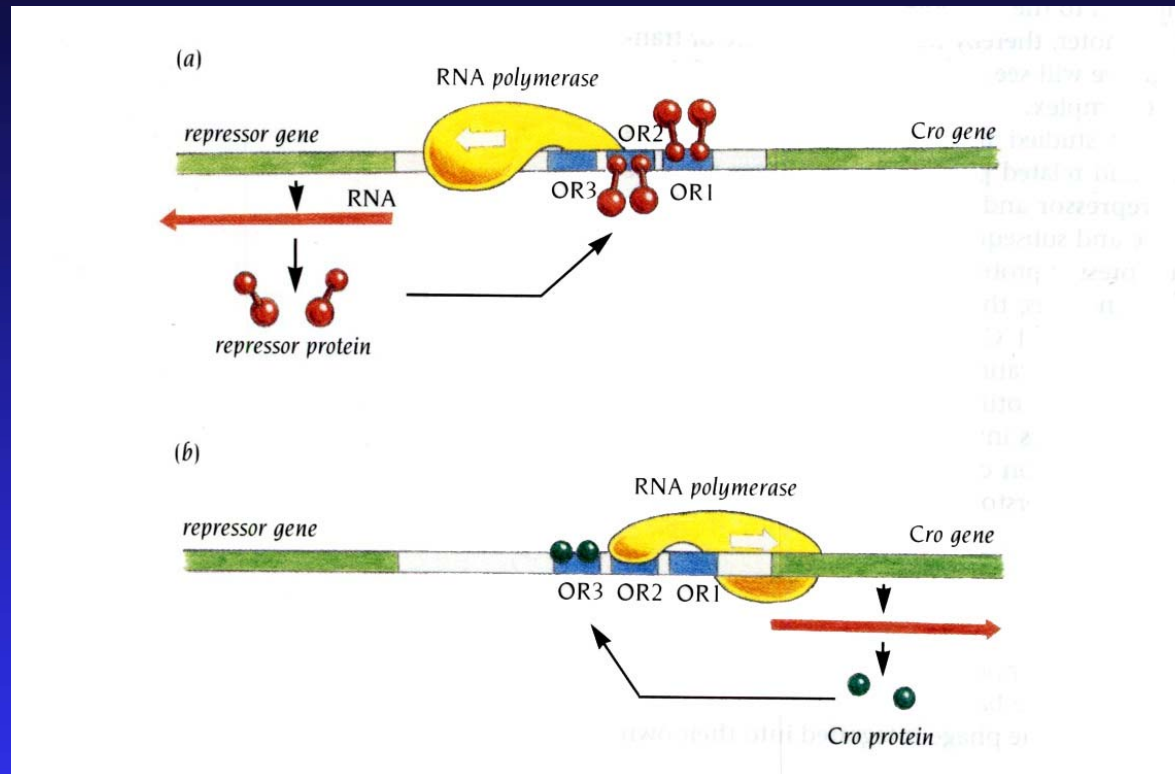
- Certain strains of *E. coli* can be stimulated by irradiation with a moderate dose of UV light to stop normal growth and start producing bacteriophages that eventually lyse the bacterium.
- Bacteria of these so called **lysogenic** strains carry the DNA of the phage integrated into their own chromosomes, where it is dormant during normal cell growth; the phage DNA is replicated as an integral part of the bacterial chromosome, but the phage genes are not expressed.
- UV light switches on the phage genes which then produce new phages and the cell eventually dies.
- The molecular mechanism of this genetic on-off switch can be now explained based on results obtained by a combination of genetic and x-ray structural studies.

A molecular mechanism for gene control



A region of DNA in the related bacteriophages λ , 434, P22 that controls the switch for synthesis of new phage particles is shown. Two structural genes are involved in this switch; one coding for a repressor protein and one coding for the Cro protein. Between these genes there is an operator region (OR) that contains three protein binding sites – OR1, OR2, and OR3.

Repressor and Cro proteins operate a procaryotic genetic switch region



- a) In a **lysogenic** strain of *E. coli*, repressor and RNA polymerase bind to the switch region. The repressor binds to OR1 and OR2, thereby turning off synthesis of Cro. The repressor also works as an activator for its own synthesis by facilitating RNA-polymerase binding to the repressor promoter through its binding to OR2.
- b) In the **lytic** phase, synthesis of Cro protein turns off synthesis of the repressor, since Cro binds to OR3 and blocks RNA-polymerase binding to the repressor promoter. Transcription of phage genes to the right can now occur

The sequences OR1, OR2, and OR3 in phage λ are palindromic and similar but not identical

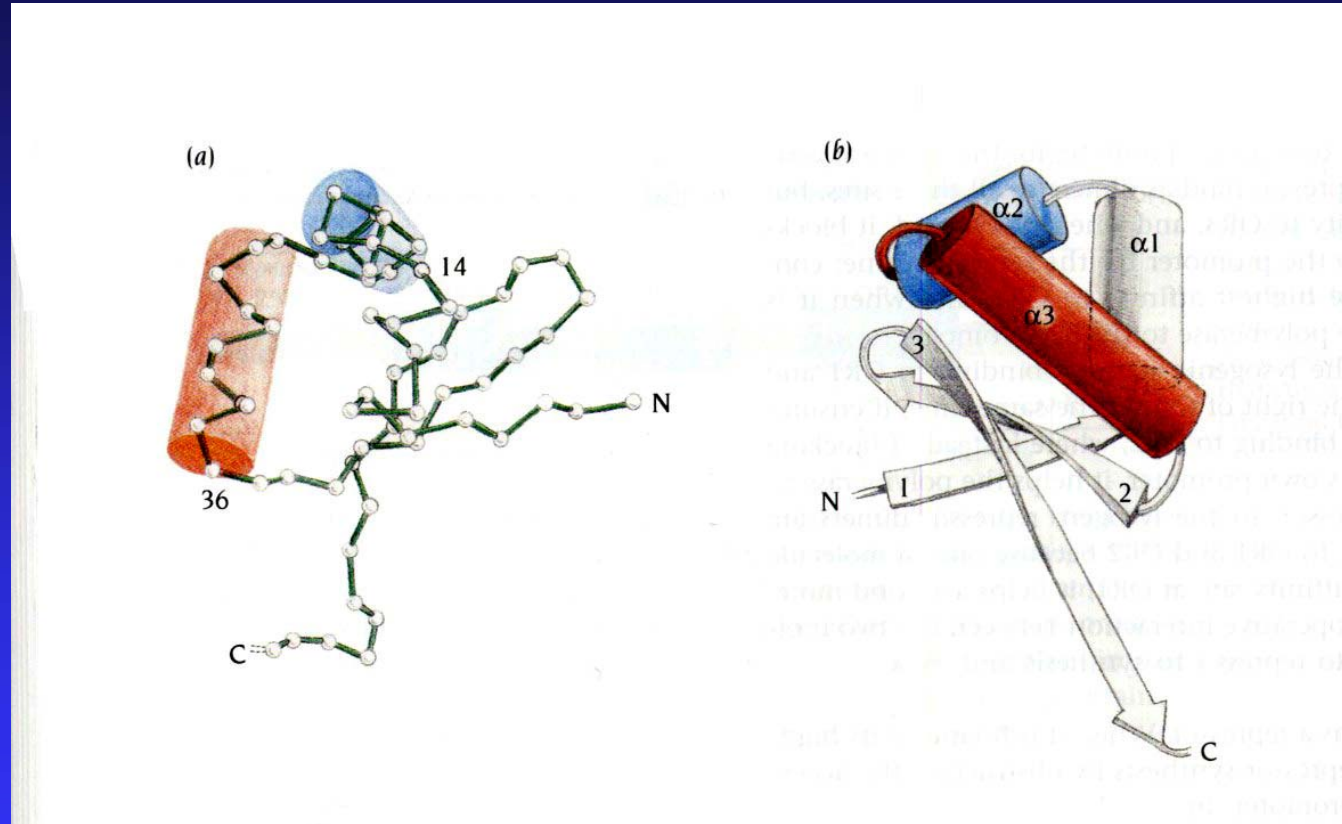
Table 8.1 The nucleotide sequences of the three protein-binding regions OR1, OR2, and OR3 of the operator of bacteriophage lambda

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
OR1	5'	T	A	T	C	A	C	C	G	C	C	A	G	T	G	G	T	A	3'
	3'	A	T	A	G	T	G	G	C	G	G	T	C	A	C	C	A	T	5'
OR2	5'	T	A	A	C	A	C	C	G	T	G	C	G	T	G	T	T	G	3'
	3'	A	T	T	G	T	G	G	C	A	C	G	C	A	C	A	A	C	5'
OR3	5'	T	A	T	C	A	C	C	G	C	A	A	G	G	G	A	T	A	3'
	3'	A	T	A	G	T	G	G	C	G	T	T	C	C	C	T	A	T	5'

Palindromic base pairs that are most frequent at the two ends are green, and the pseudo-twofold symmetry axis is indicated by a red dot.

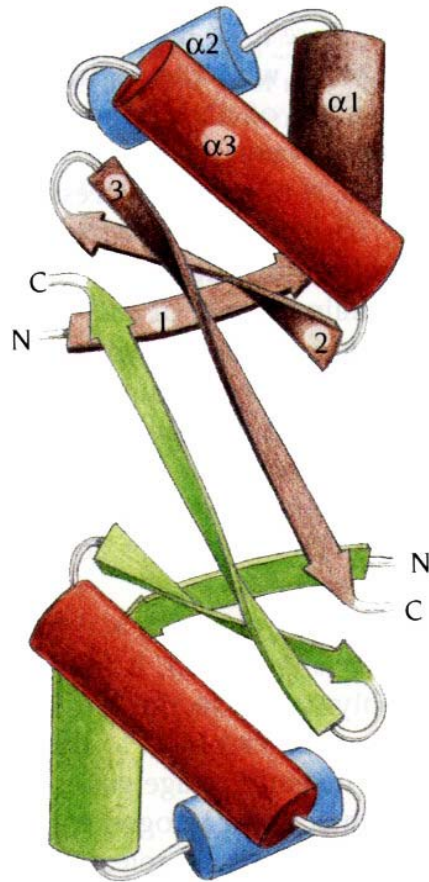
How do repressor and Cro recognize the specific operator regions and achieve the subtle differential binding to switch regions?

The x-ray structure of the complete λ Cro protein is known



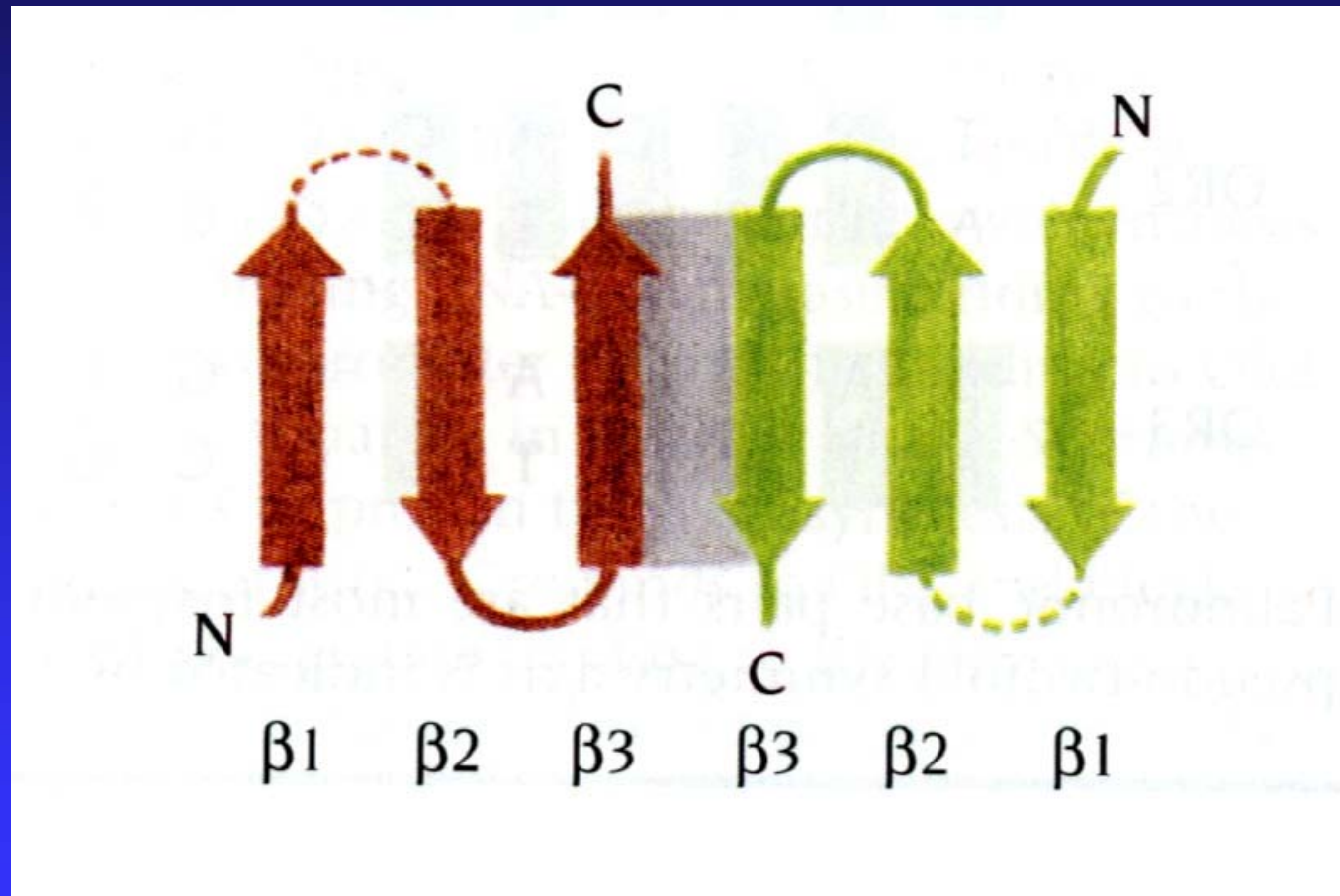
The DNA-binding protein Cro from bacteriophage λ contains 66 amino acid residues that fold into three α helices and three β strands. α Helices 2 and 3 that form the helix-turn-helix motif are colored blue and red, respectively.

Cro molecules from bacteriophage λ form dimers both in solution and in the crystal structure

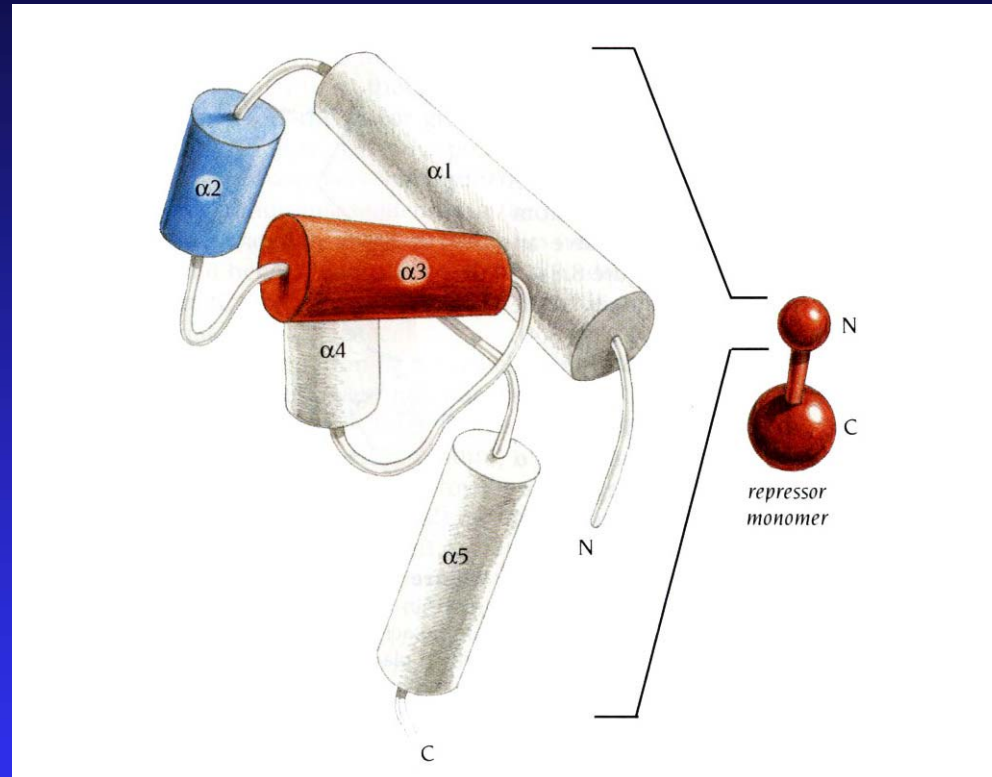


The main dimer interactions are between β strands 3 from each subunit (shown in green and brown).

The three β strands of each subunit in λ Cro are aligned in the dimer so that a six-stranded antiparallel β sheet is formed

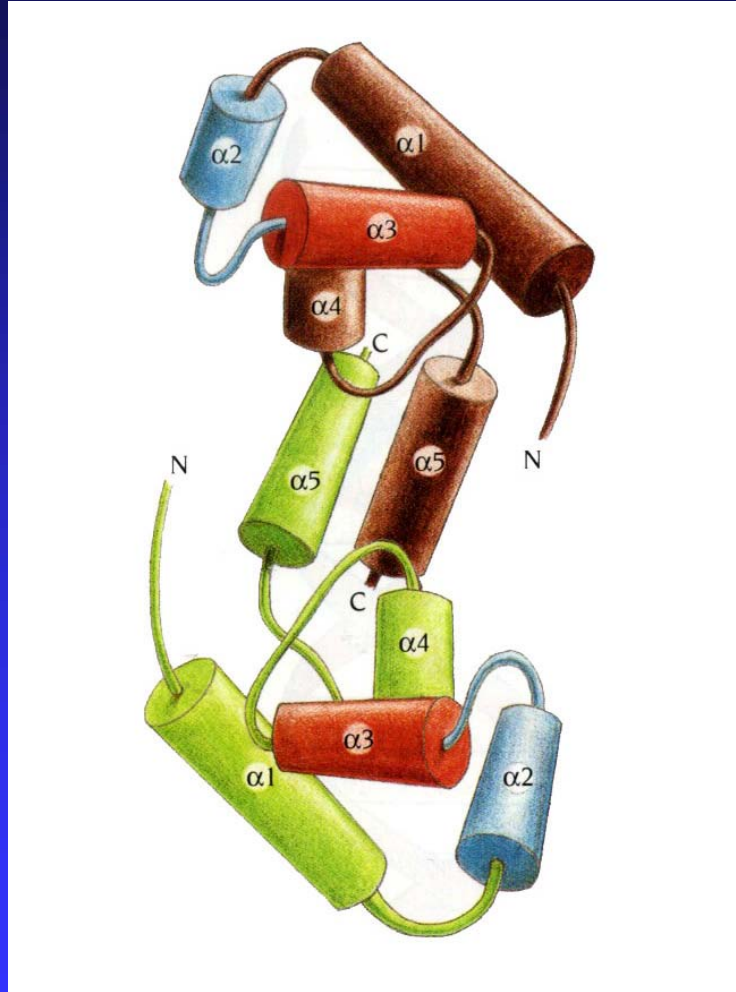


The x-ray structure of the DNA-binding domain of the λ repressor is known



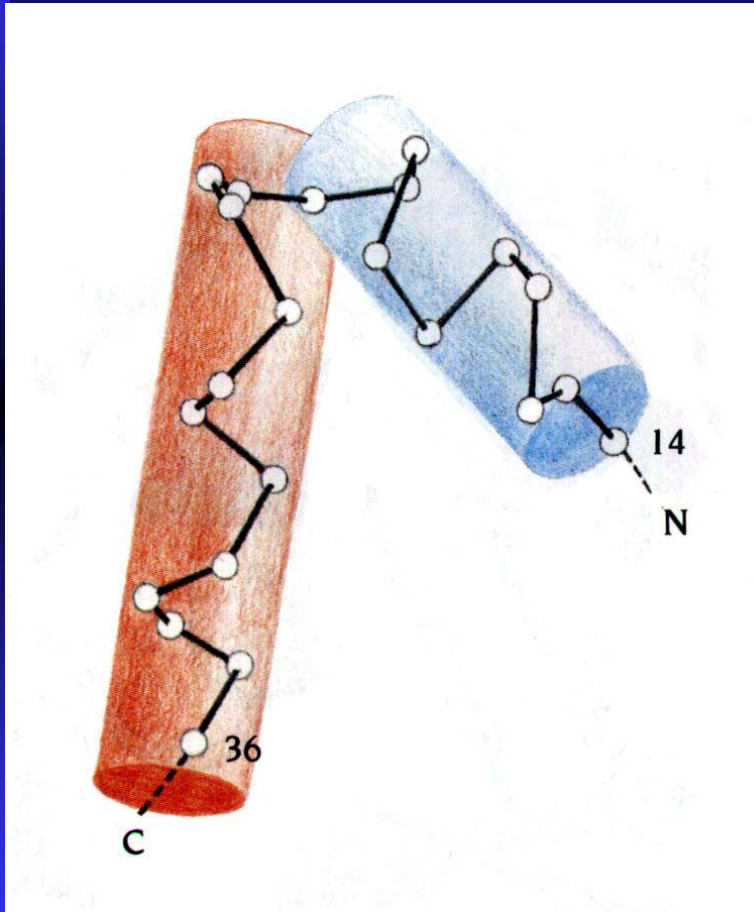
The N-terminal domain of λ repressor, which binds DNA, contains 92 amino acid residues folded into five α helices. Two of these, $\alpha 2$ and $\alpha 3$ form a helix-turn-helix motif with a very similar structure to that of λ Cro. The complete repressor monomer contains in addition a large C-terminal domain.

The N-terminal domains of λ repressor form dimers



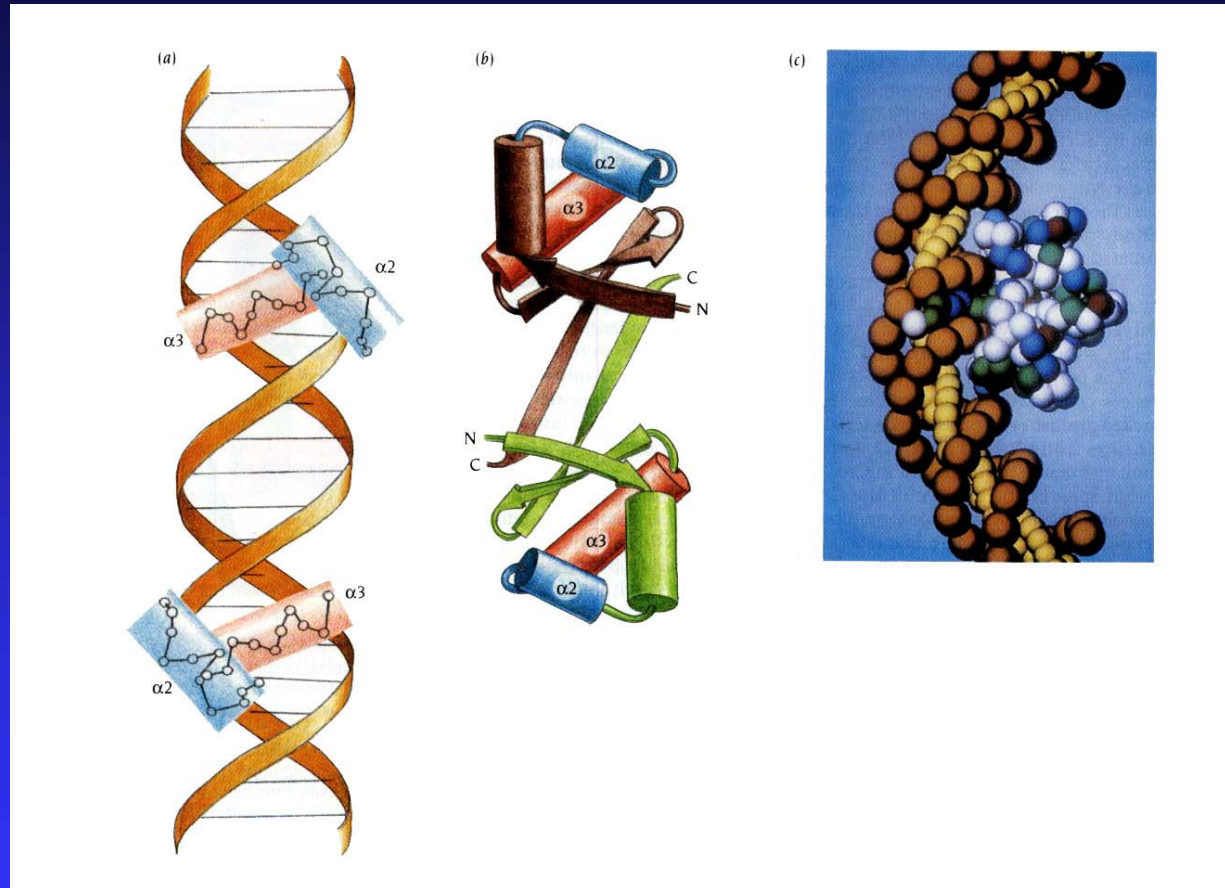
The dimers are formed by interactions between α helix 5 from each subunit. The different subunits are colored green and brown except for the helix-turn-helix motif. In the intact repressor, the C-terminal domains are mainly responsible for dimer formation.

The DNA-binding helix-turn-helix motif in λ Cro



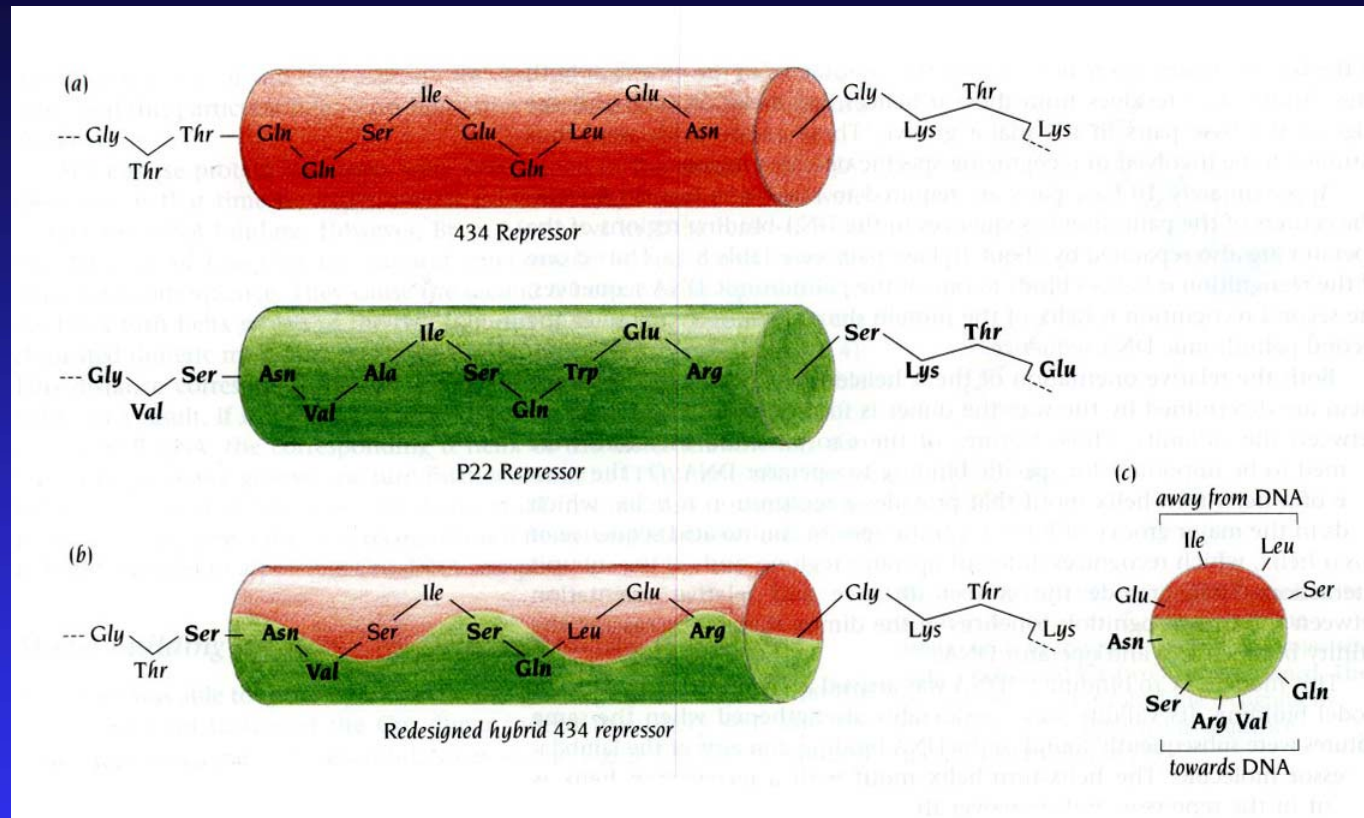
C_{α} positions of the amino acids in this motif have been projected onto a plane and the two helices outlined. The second helix (red) is called the recognition helix because it is involved in sequence-specific recognition of DNA. Similar helix-turn-helix motif is found in λ repressor, CAP, and some eucaryotic transcription factors.

Model building predicts Cro-DNA interactions



The helix-turn-helix motif in λ Cro is bound to DNA with the two recognition helices of the Cro dimer sitting in the major groove of DNA.

Genetic studies agree with the structural model



The proposed DNA-binding surface of the recognition helix of bacteriophage 434 repressor was redesigned genetically to that of P22 repressor by changing six amino acid residues. The redesigned 434 repressor acquired all the DNA-binding properties of the P22 repressor. The six amino acid residues that were changed in the 434 repressor are shown in boldface in (b) and (c).

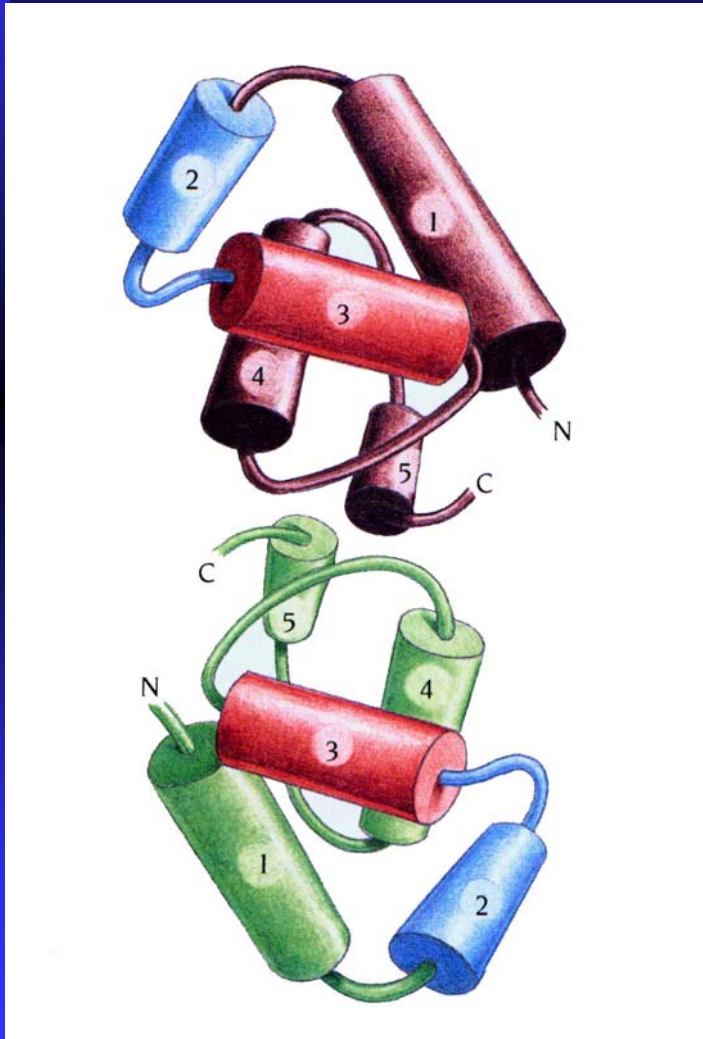
The x-ray structure of DNA complexes with 434 Cro and repressor revealed novel features of protein-DNA interactions

Table 8.2 The six operator regions (OR1–OR3 and OL1–OL3) in bacteriophage 434

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	
OR1	5'	A	C	A	A	G	A	A	A	G	T	T	T	G	T	3'
	3'	T	G	T	T	C	T	T	T	C	A	A	A	C	A	5'
OR2	5'	A	C	A	A	G	A	T	A	G	T	T	T	G	T	3'
	3'	T	G	T	T	C	T	A	T	C	A	A	A	C	A	5'
OR3	5'	A	C	A	A	G	A	A	A	A	A	C	T	G	T	3'
	3'	T	G	T	T	C	T	T	T	T	T	G	A	C	A	5'
OL1	5'	A	C	A	A	G	G	A	A	G	A	T	T	G	T	3'
	3'	T	G	T	T	C	C	T	T	C	T	A	A	C	A	5'
OL2	5'	A	C	A	A	T	A	A	A	T	A	T	T	G	T	3'
	3'	T	G	T	T	A	T	T	T	A	T	A	A	C	A	5'
OL3	5'	A	C	A	A	T	G	G	A	G	T	T	T	G	T	3'
	3'	T	G	T	T	A	C	C	T	C	A	A	A	C	A	5'
Synthetic DNA	5'	A	C	A	A	T	A	T	A	T	A	T	T	G	T	3'
	3'	T	G	T	T	A	T	A	T	A	T	A	A	C	A	5'
		14'	13'	12'	11'	10'	9'	8'	7'	6'	5'	4'	3'	2'	1'	

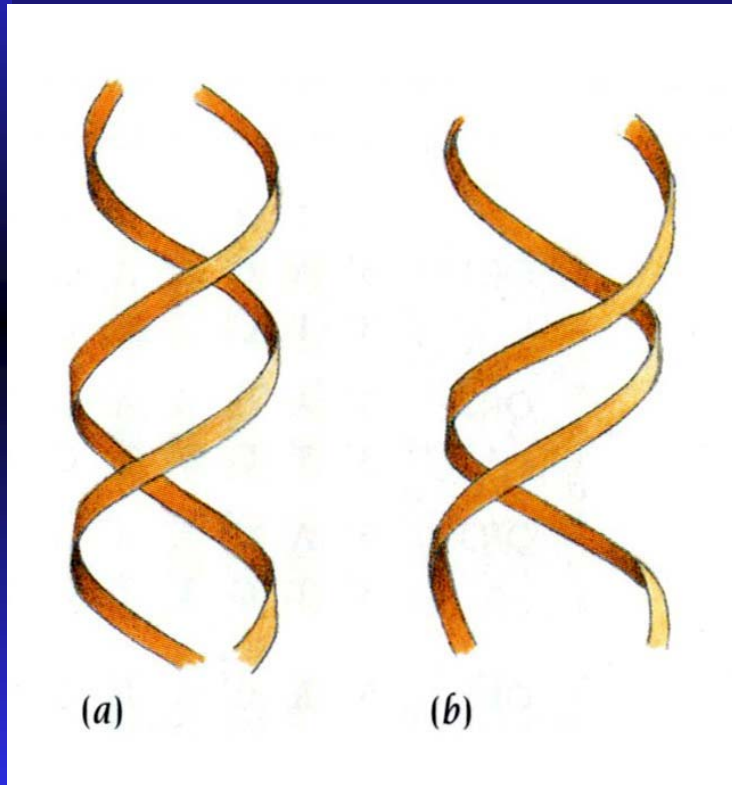
Crystal structures have been determined of complexes between both 434 Cro and the repressor fragment with synthetic DNA fragments – one 14 base pairs long (a 14 mer), which is completely palindromic, and one 20 base pairs long (a 20mer), which contains the sequence of OR1 in its middle region.

The x-ray structure of DNA complexes with 434 Cro and repressor revealed novel features of protein-DNA interactions



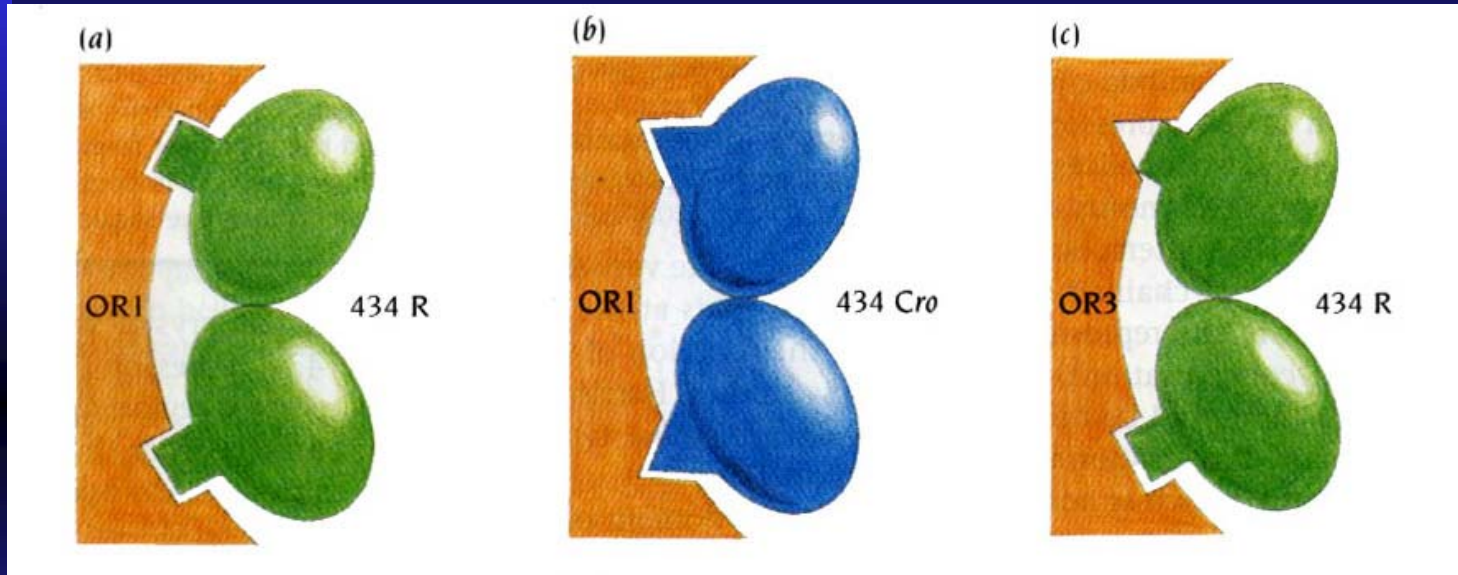
The DNA-binding domain of 434 repressor is a dimer in its complexes with DNA fragments. Each subunit (green and brown) folds into a bundle of four α helices (1-4) that have a structure similar to the corresponding region of λ repressor including the helix-turn-helix motif (blue and red). A fifth α helix is involved in the subunit interactions. The structure of the 434 Cro dimer is very similar to the 434 repressor. The corresponding regions of 434 Cro and repressor show 48 % sequence identity.

The 434 Cro and repressor impose precise distortions on the B-DNA in the complexes



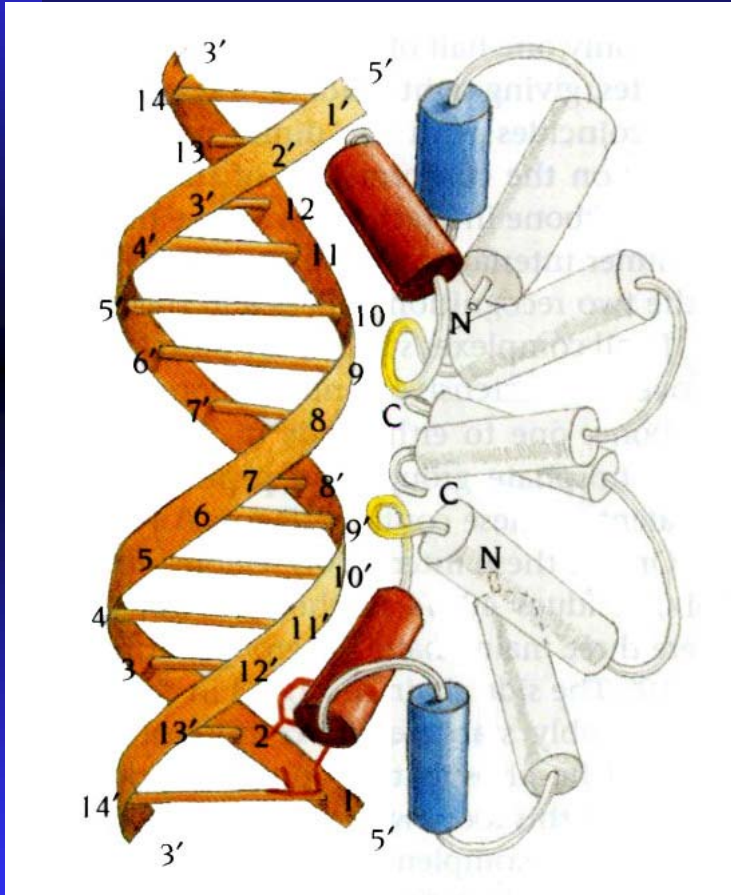
The proteins impose changes of DNA structure from regular B-DNA (a) to a distorted version (b) when bound to operator regions. The distortions essentially involve bending of DNA and over-winding of the middle regions. The diagram shows the sugar-phosphate backbones of DNA as orange ribbons viewed in the narrow groove in the middle region of the operator.

The 434 Cro and repressor impose precise distortions on the B-DNA in the complexes



Binding of 434 repressor fragment and 434 Cro to operator region OR1 induces different structural changes in the region of the DNA that binds the proteins. In the complex of OR3 with the 434 repressor fragment the two half sites of OR3 are different: one is similar to OR1, with bound repressor, whereas the other has a different nucleotide sequence that adopts the Cro-type binding conformation on binding repressor. The binding surfaces of the DNA and repressor fragment do not complement each other as they do in the OR1 complex; consequently, the repressor fragment binds more weakly to OR3 than OR1.

Sequence-specific protein-DNA interactions recognize operator regions



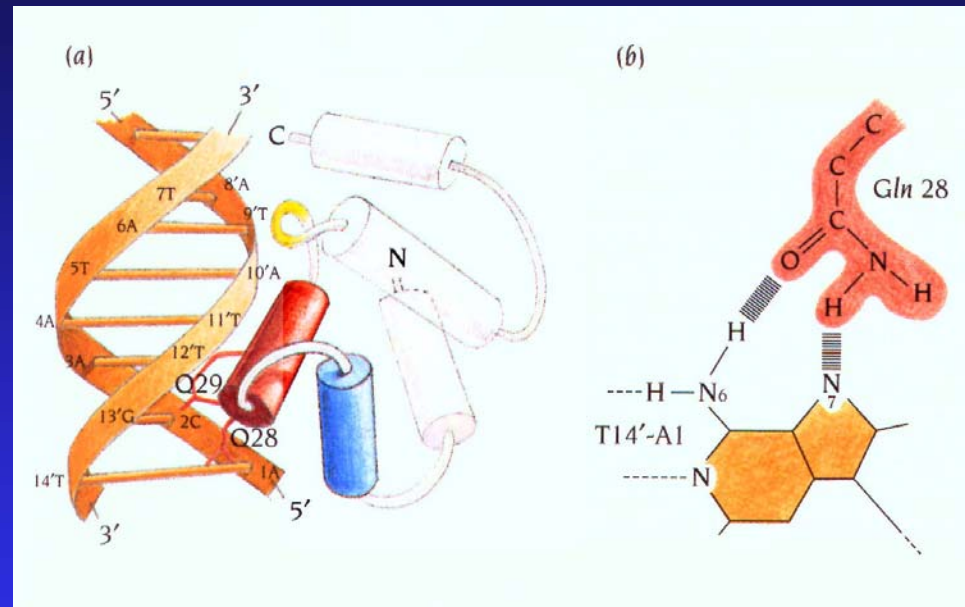
The protein dimer binds so that the recognition α helices at opposite ends of the protein molecule are in the major groove of the DNA as predicted, where they interact with base pairs at the end of the DNA molecule. Since these binding sites are separated by one turn of the DNA helix, it follows that at the center of the DNA molecule the narrow groove faces the protein. There are no interactions between the protein and the bases of the DNA in this middle region of the operator.

Sequence-specific protein-DNA interactions provide a general recognition signal for operator region in 434 bacteriophage

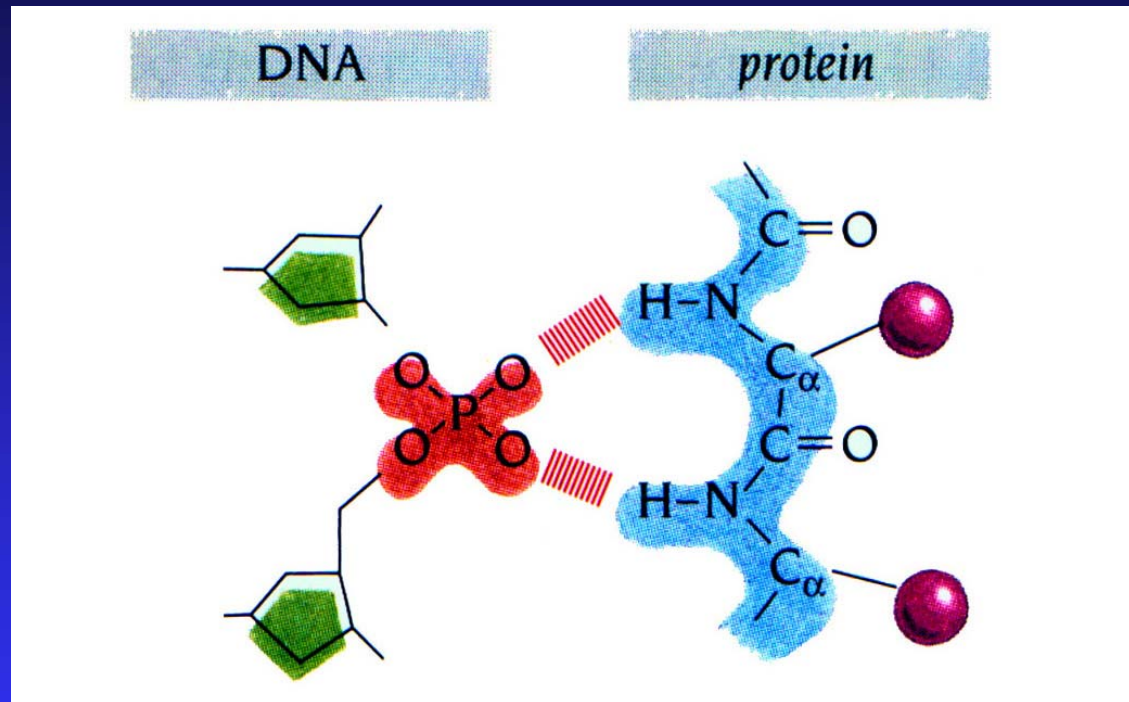
There are two glutamine residues (28 and 29) at the beginning of the recognition helix in the helix-turn-helix motif that provide the interactions with the first three base pairs of the operator regions. The first three base pairs are identical in all six operator regions recognized by phage 434 repressor. This means that the interactions cannot contribute to the discrimination between the six binding sites in the DNA.

This simple pattern of hydrogen bonds and hydrophobic interactions therefore accounts for the specificity of phage 434 Cro and repressor proteins for 434 operator regions.

Q28 forms two hydrogen bonds to A1 in (T14'-A1), and Q29 forms a hydrogen bond to O₆ of G13' in base pair 2 (G13'-C2). The methyl groups of T27 and Q29 form a hydrophobic pocket to receive the methyl group of T12' at base pair 3 (T12'-A3).

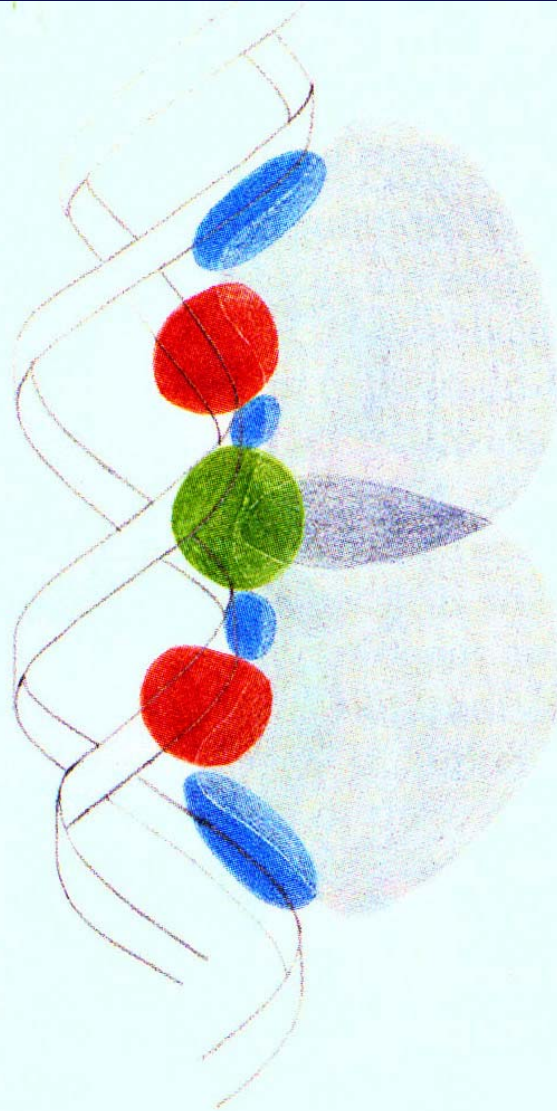


Protein-DNA backbone interactions determine DNA conformation



Non-specific protein-DNA interactions are frequently formed by hydrogen bonds between backbone phosphate oxygen atoms of DNA and main-chain NH groups of the protein.

The essence of phage repressor and Cro



- 1 H-bonds between sugar-phosphate backbone and protein help anchor protein to DNA
- 2 Sequence-specific interaction between DNA and recognition helix allows recognition of OR regions
- 3 DNA distortion allows close interactions with other regions of Cro and repressor and accounts for differential affinities

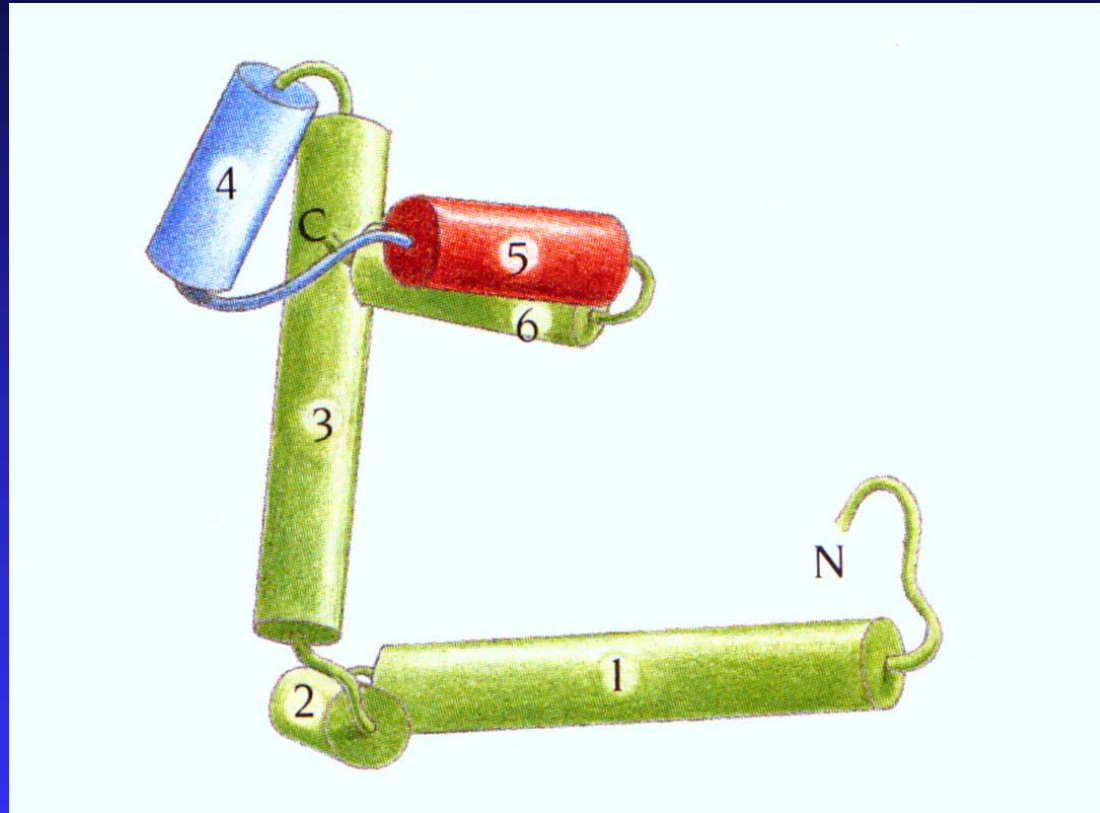
The essence of phage repressor and Cro

A few residues on the recognition helix of the helix-turn-helix motif form the recognition signal of Cro and repressor proteins for the operator regions by specific interaction with the edge of the base pairs in the major groove of DNA.

H-bonds between sugar-phosphate backbone and protein help anchor protein to DNA.

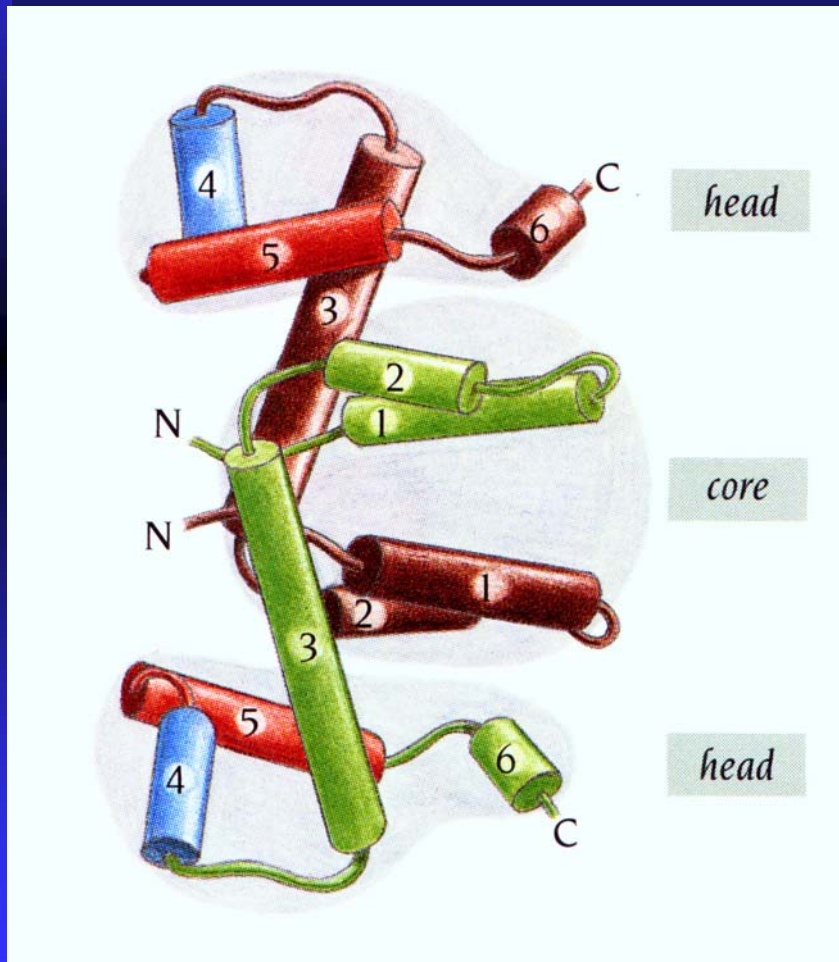
The differential affinities of Cro and repressor for the operator regions are determined by the ability of DNA to undergo specific structural changes. At least two factors are important to the ability of the complex to achieve the proper DNA conformation. Interactions between the sugar-phosphate backbone of DNA and regions of the protein form a large interaction area of complementary surfaces that stabilizes the structural changes in DNA. These protein regions are not restricted to the helix-turn-helix motif but are spread over almost the entire polypeptide chain. The ability of these interactions to accomplish a structural change can be modulated by the sequence of DNA.

The *trp* repressor forms a helix-turn-helix motif



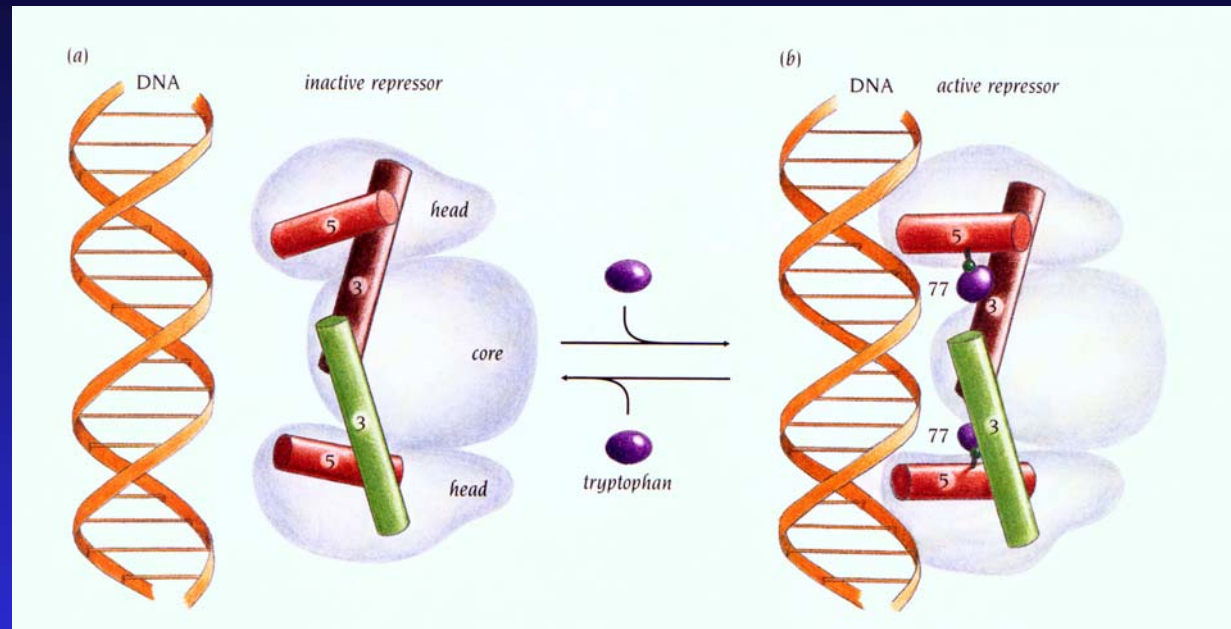
The subunit of the *trp* repressor contains 107 amino acid residues that are folded into six α helices. Helices 4 and 5 form the DNA-binding helix-turn-helix motif.

trp repressor stability is conferred by dimerization that results in a functional compact globular molecule



The α helices of the N-terminal region of the *trp* repressor are involved in subunit interactions and form a stable core in the middle of the dimer. α Helix 3 connects the core to the head in both subunits.

A conformational change operates a functional switch



In the absence of tryptophan the recognition helices are tilted inward toward the core. This makes the distance between the two DNA binding sites too short by 5-6 Å to allow them to fit into the major groove. Tryptophan causes a conformational change in the repressor that alters the orientations of recognition helices. They are properly poised 34 Å apart for binding in major groove, as they are in the 434 repressor.

A77V mutant is constitutively active as the bulkier valine side chain at position 77 maintains the heads in an active conformation even in the absence of bound tryptophan.

The lac repressor subunit is arranged in four domains



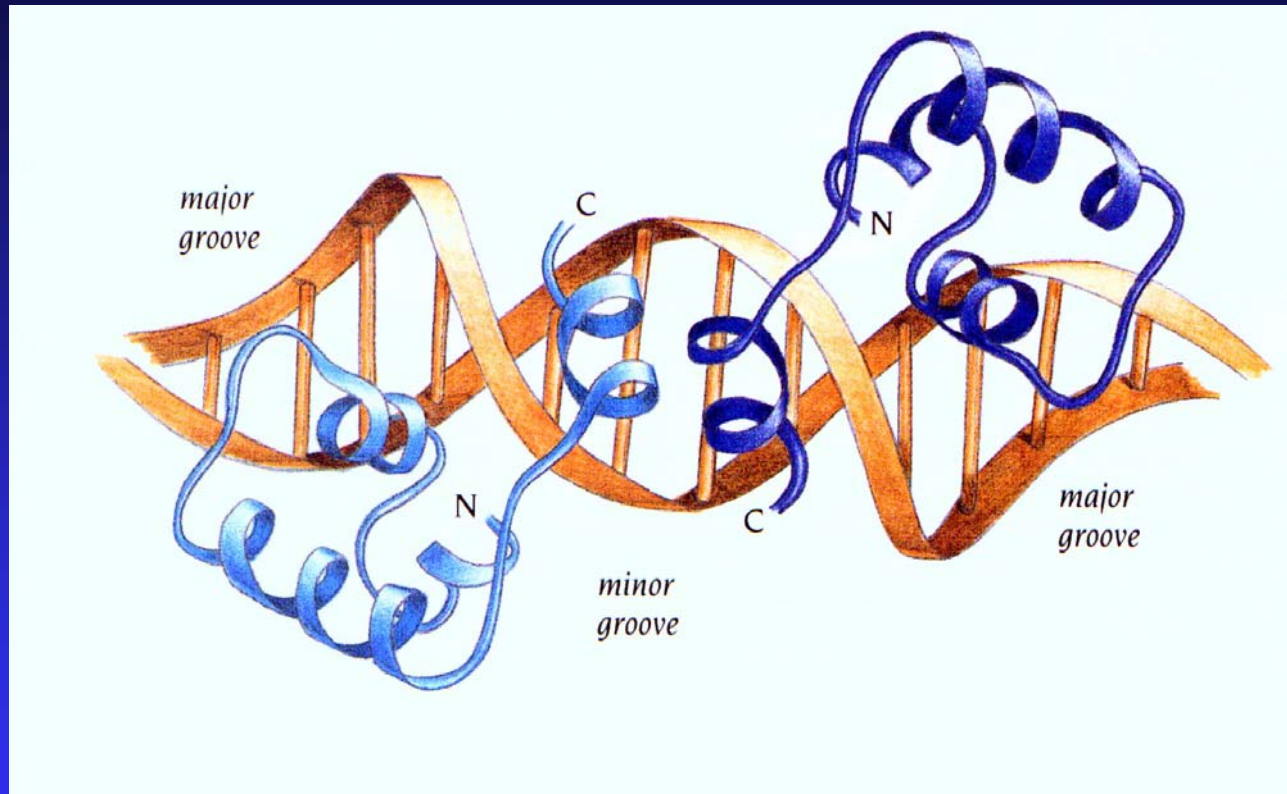
The polypeptide chain (360 amino acids) is arranged in four domains, an amino terminal DNA-binding domain (red), with a helix-turn-helix motif, a hinge helix (purple), a large core domain which has two sub-domains (green and blue), and a C-terminal α helix (yellow).

Lac repressor binds to both the major and the minor grooves inducing a sharp bend in the DNA



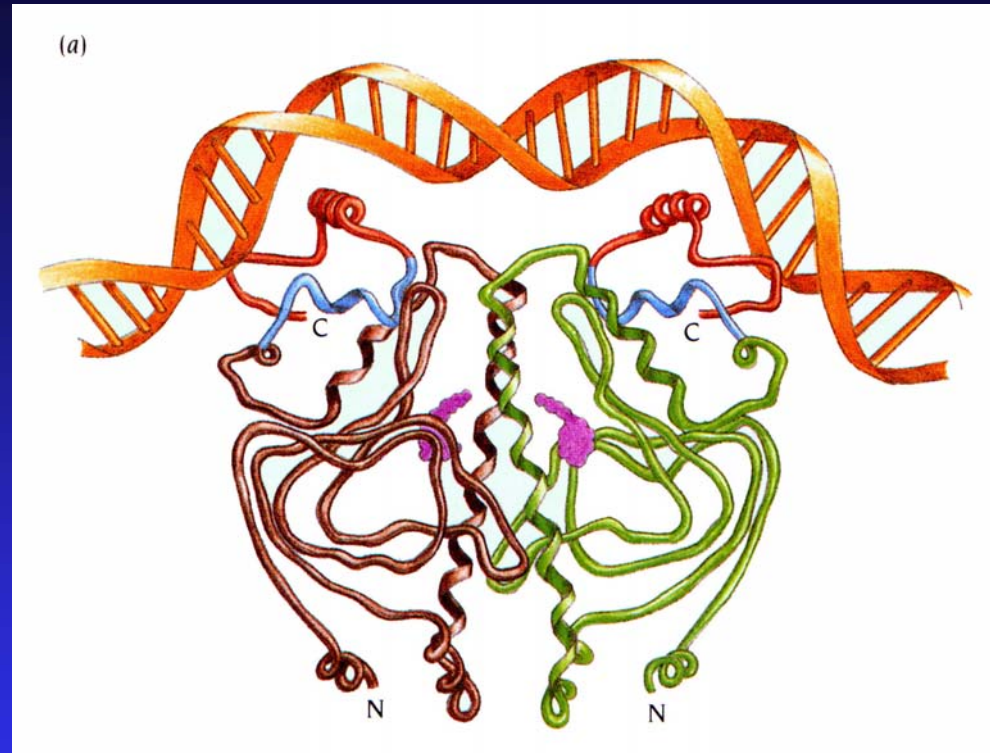
The lac repressor is a V-shaped tetramer in which each arm is a dimer containing a DNA-binding site. The helix-turn-helix motifs of each dimer bind in two successive major grooves and the hinge helices bind adjacent to each other in the minor groove between the two major groove binding sites. The four subunits of the tetramer are held together by the four C-terminal helices. The bound DNA fragments are bent.

Hinge helices of the lac repressor interact with the minor groove



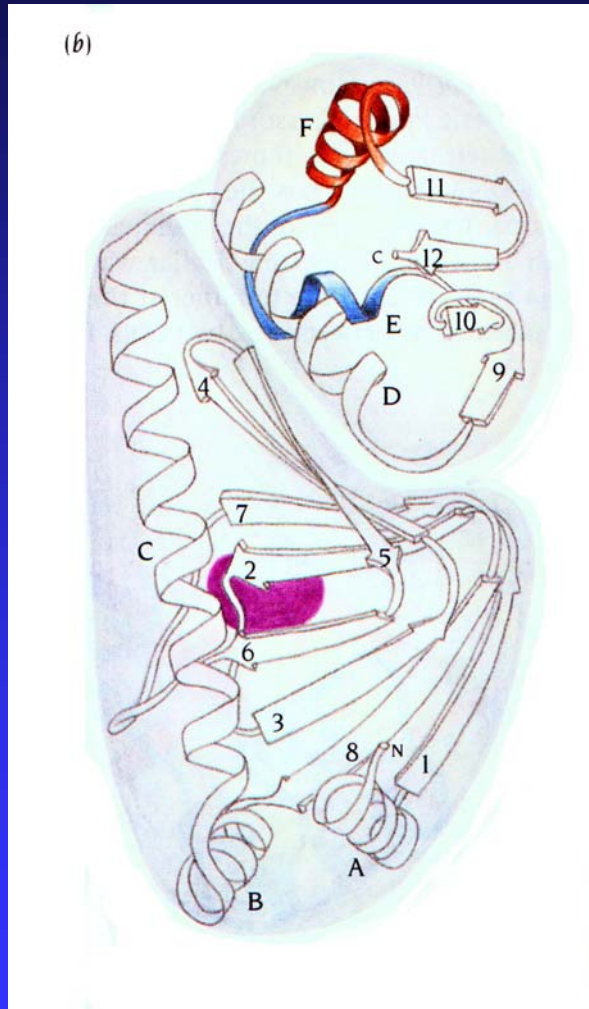
The helix-turn-helix motifs of lac repressor subunits bind to the major groove of DNA with the N-terminus of the second helix, the recognition helix, pointing into the groove. The two hinge helices of each arm of the V-shaped tetramer bind adjacent to each other in the minor groove of DNA, which is wide and shallow due to distortion of the B-DNA structure.

CAP-induced bending could activate transcription



The CAP dimer interacts directly with 27 of the 30 base pairs of the DNA fragment present in the crystal. The amino ends of the recognition helices of the helix-turn-helix motifs are bound in the major groove of DNA. These interactions provide sequence specificity between protein side chains and the exposed edges of base pairs. In addition, 22 of the phosphate groups of DNA interact directly with protein side chains or main chain amide groups providing a large number of contacts.

Structure of one subunit of CAP



The CAP molecule comprises two identical polypeptide chains of 209 amino acid residues. Each chain is folded into two domains that have separate functions. The larger N-terminal domain binds the allosteric effector molecule, cAMP, and provides all the subunit interactions that form the dimer. The C-terminal domain contains the helix-turn-helix motif that binds DNA.