

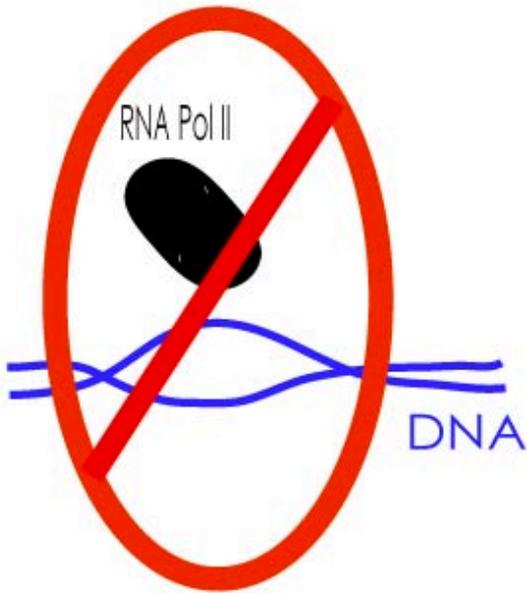
## Structure/function relationship in DNA-binding proteins

Devlin Chapter 8.8-9

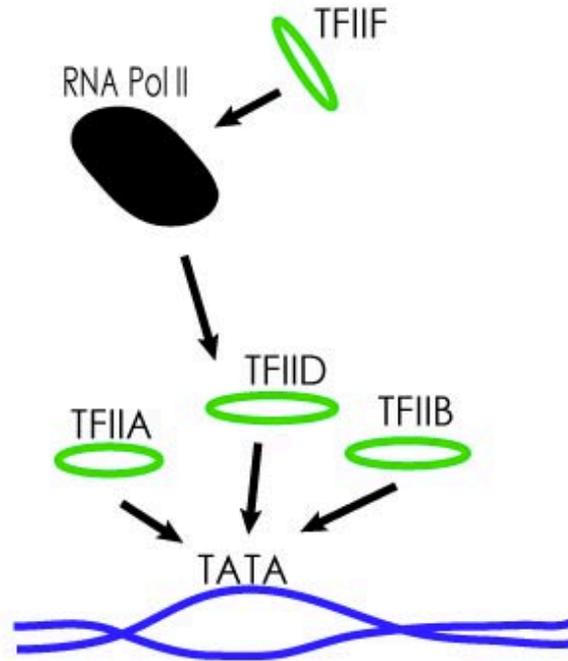
- ◆ General description of transcription factors (TFs)
  - ◆ Sequence-specific interactions between DNA and proteins
  - ◆ Structural motifs of TF DNA-binding domains (helix-turn-helix; zinc fingers; bZIP; helix-loop-helix; beta-scaffold)
  - ◆ DNA-protein interactions in transcription
    - ◆ RNA polymerase and preinitiation complex for transcription
    - ◆ Mechanism of gene activation by transcription factors
- 
- 

# Proteins are the processing units for transcription

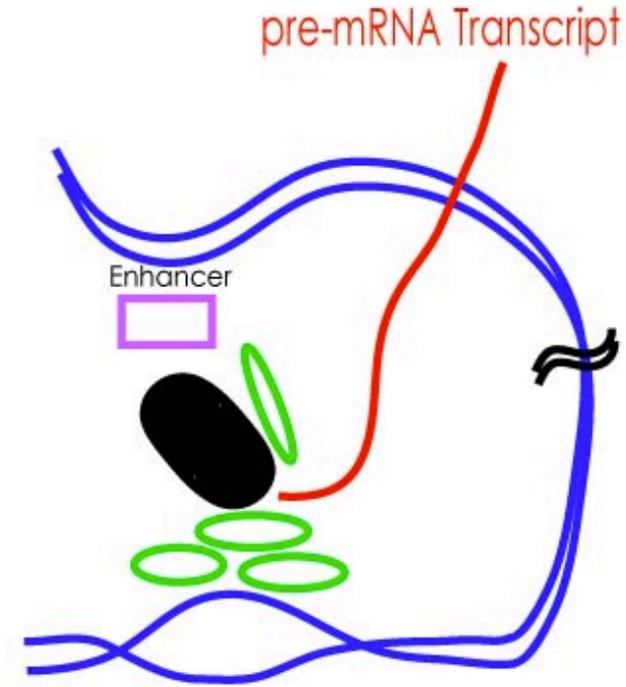
cnx.org



Eukaryotic RNA Polymerases cannot bind DNA alone!



Basal transcription factors are REQUIRED for initiation!



Distant enhancer sequences can also affect transcription!

# Protein-DNA interactions

- ▶ DNA-binding proteins
  - ▶ Transcription factors
  - ▶ Polymerases & nucleases
- ▶ Transcription factors utilize a wide range of DNA-binding structural motifs
- ▶ Interaction between dimeric proteins and palindromic sequences are common.
- ▶ Binding often leads to the conformational changes in the protein and DNA.



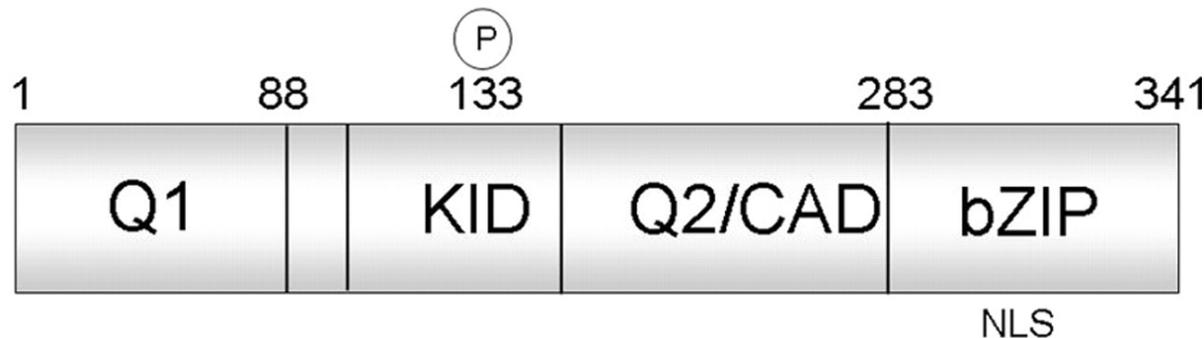
$\lambda$  repressor  
PDB 2ORI

# Transcription factors

- ▶ Sequence-specific DNA-binding factor that controls the rate of transcription by promoting (activator) or blocking (repressor) the recruitment of RNA polymerase
  - ▶ Other proteins critical for regulating transcription lack DNA-binding domains: coactivators, corepressors, deacetylases, methylases, chromatin remodelers
- ▶ Bind either promoter or enhancer regions of DNA
- ▶ Regulatory mechanism
  - ▶ Stabilize or block RNA polymerase association with DNA
  - ▶ Catalyze, directly or by recruitment, acetylation or deacetylation of histones
  - ▶ Recruit coactivator or corepressor proteins to the complex
- ▶ Function: basal level transcription, development, signaling, cell cycle control
- ▶ Activation of TFs:
  - ▶ may be activated or deactivated by ligand binding to a 'sensing' domain (hormone receptors),
  - ▶ chemical modification (e.g. STAT proteins must be phosphorylated to bind DNA)
- 4 ▶ Interaction coregulatory proteins or with other TFs (homo- or hetero-dimerization)

# Eukaryotic transcription factors have modular structure: CREB example

- ▶ DNA binding domain (bZIP)
  - ▶ Recognize specific sequences in DNA for sequence-specific binding. (e.g. HLH)
  - ▶ Basis of structural classification
- ▶ Dimerization domain (bZIP)
  - ▶ Promotes the formation of heterodimer or homodimers.
- ▶ Activation domain (Q1 and Q2)
  - ▶ Interact with general transcription factors, RNA polymerase II, or other regulators of transcription.
  - ▶ e.g. acidic domains, glutamine-rich domains, and proline-rich domains.
- ▶ Protein interaction domain (KID)
  - ▶ Associate with proteins like histone acetyltransferases or coactivators.

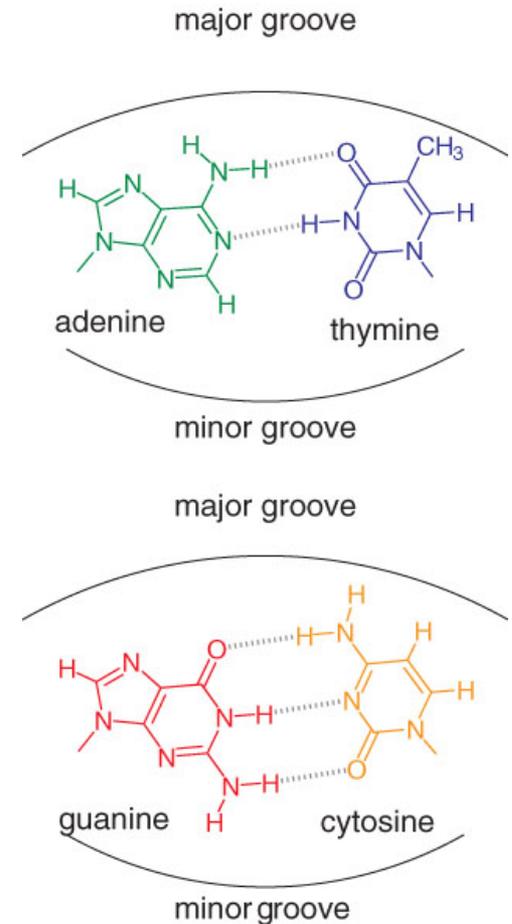


cAMP response element binding protein, CREB

Order varies

# DNA sequence-specificity of DNA-binding proteins

- ▶ **Sequence-specific interactions**
  - ▶ Frequently involve DNA major groove
  - ▶ Base-specific H-bond donor, acceptors, and nonpolar groups are recognized by DNA-binding proteins
  - ▶ DNA structure can deviate from classic B-form helix, and therefore be specifically recognized by a protein.
  - ▶ No simple recognition code between DNA and protein sequences.
- ▶ **Nonspecific interactions**
  - ▶ interactions with DNA phosphate backbones



# Helix-turn-helix (HTH) motif

- ▶ ~20 amino-acid long DNA-binding motif.
- ▶ Formed by two helices connected by a short turn.
- ▶ The second helix is the recognition helix that binds in a sequence-specific manner in the major groove.
- ▶ The first helix stabilizes the motif.
- ▶ Examples
  - ▶ *E. coli* lactose repressor
  - ▶ Homeobox domains: appear in developmentally important transcription factors
    - ▶ e.g. antennapedia

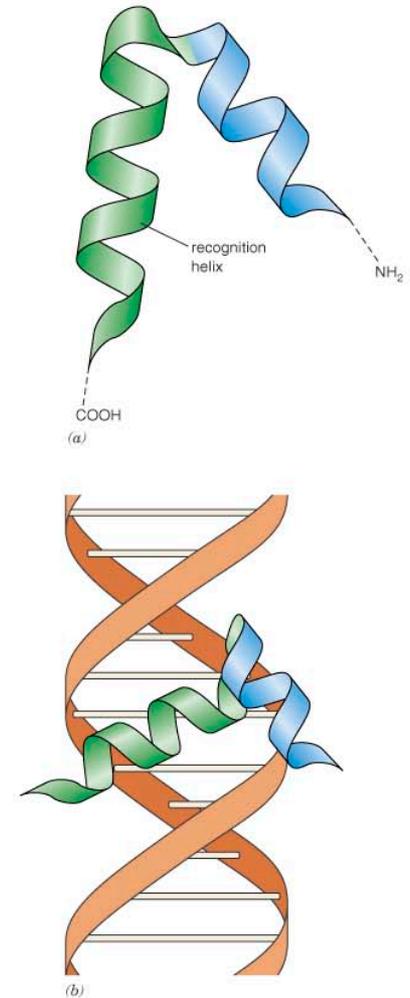
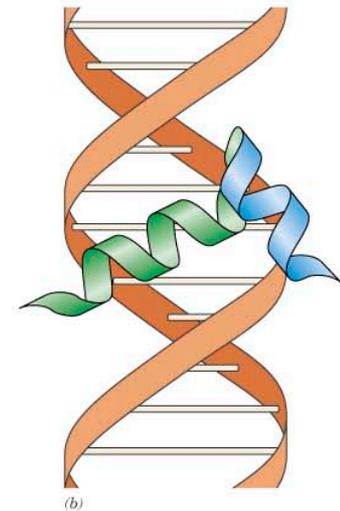
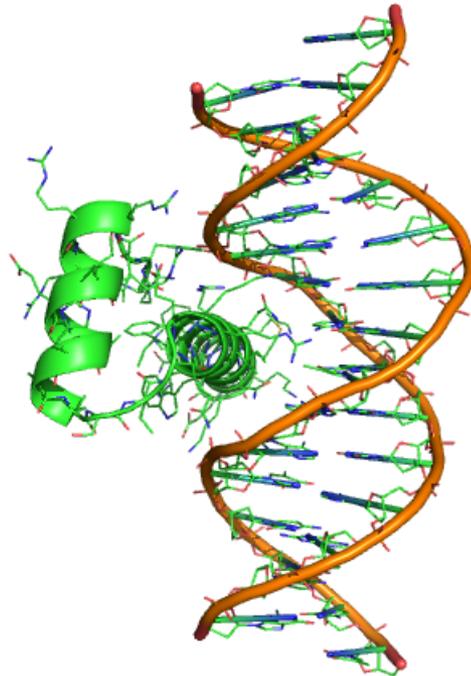
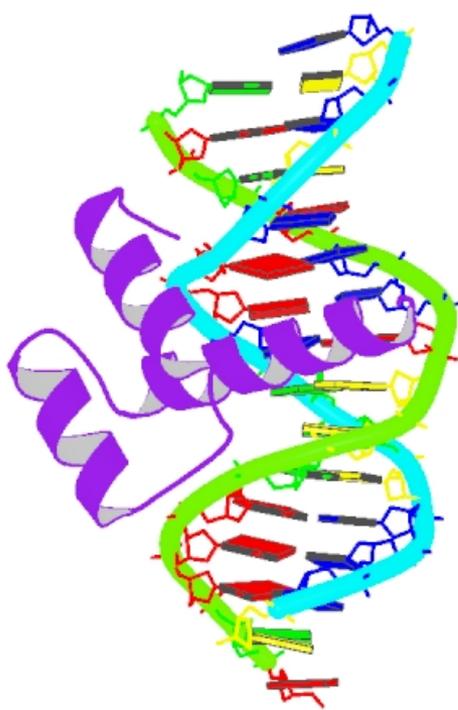


Figure 8.24

# Helix-turn-helix (HTH) motif

- ▶ Homeobox domains: appear in developmentally important transcription factors
  - ▶ e.g. antennapedia

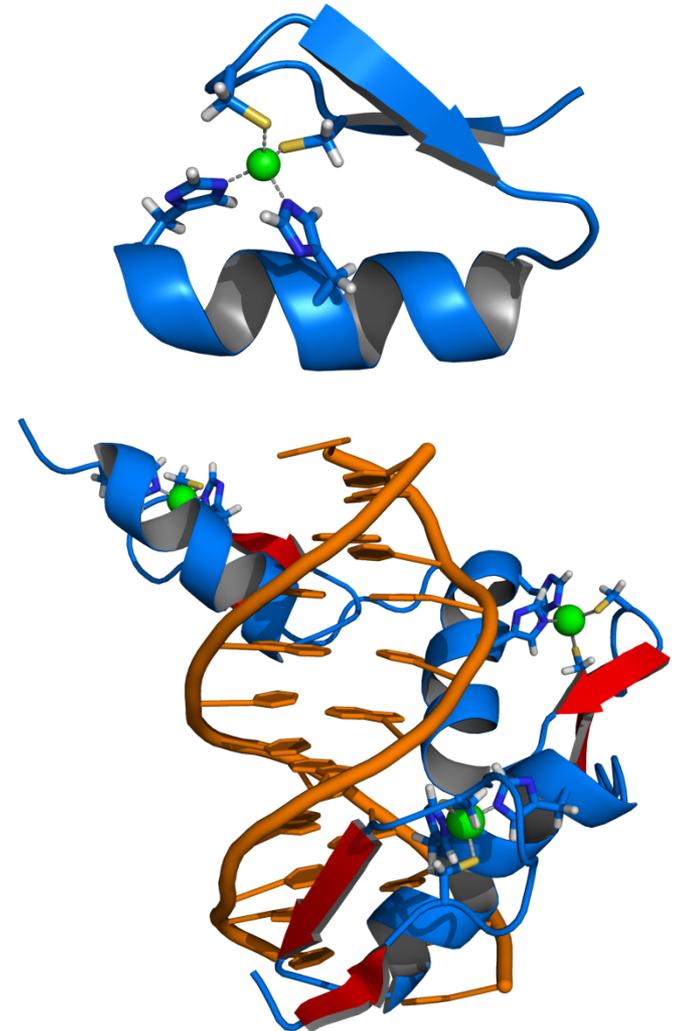


PDB entry 9ANT. Sequence specific interactions

# Zinc fingers (ZnFs)

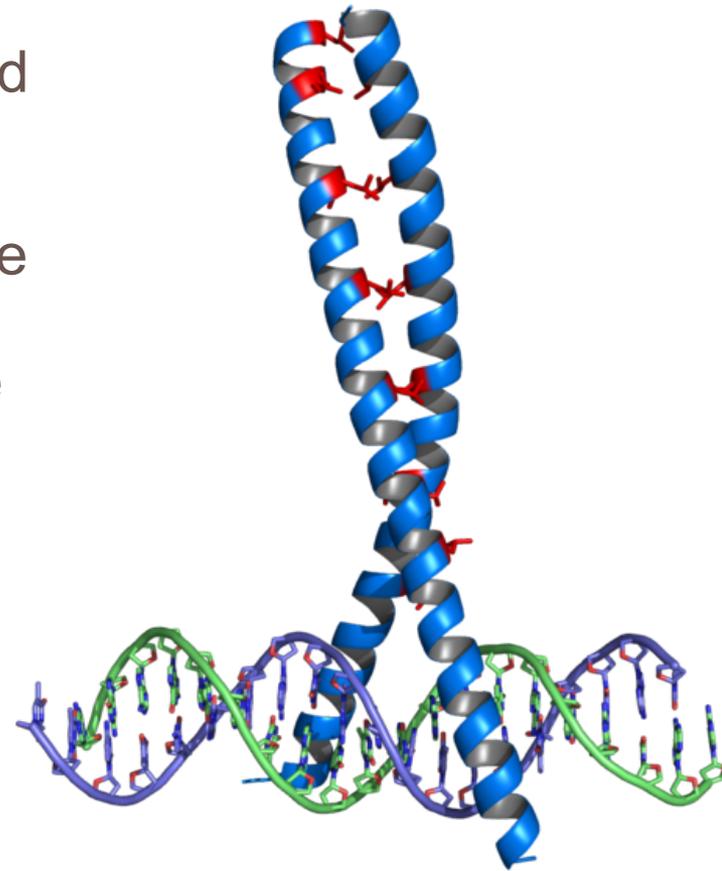
PDB-101 entry <http://www.rcsb.org/pdb/101/motm.do?momID=87>

- ▶ Classic ZnF is ~30 amino-acid domain
  - ▶ A two-stranded antiparallel  $\beta$ -sheet and short  $\alpha$ -helix
  - ▶ The  $\alpha$ -helix makes sequence-specific contacts along the major groove.
  - ▶ Initially known as sequence-specific DNA-binding motifs
  - ▶ Now known that ZnFs also recognize RNA and other proteins
- ▶ Small modules in which zinc plays a structural role
  - ▶ Structural diverse: ~8 fold groups
  - ▶ Present in ~1000 different proteins
- ▶ A  $\text{Zn}^{2+}$  ion coordinated by 4 Cys or 2 Cys and 2 His residues.
- ▶ Often occur as tandem repeats with two, three, or more fingers.
- ▶ ZnFs designed to bind targeted DNA sequences with ultimate goal of therapeutics



# Basic region-leucine zippers (bZIP)

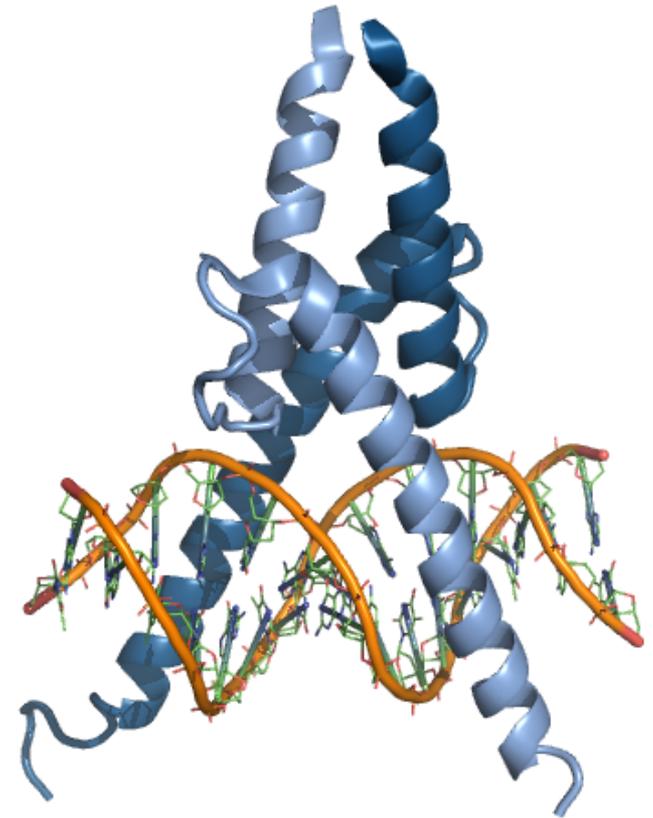
- ▶ Contain leucine residues every 7<sup>th</sup> position in an  $\alpha$ -helix.
- ▶ Form homo- or heterodimers with coiled coil structure (blue region)
- ▶ The basic region with arginine and lysine residues bind to the major groove of DNA
- ▶ The basic amino acids interact with the phosphate backbone of DNA through electrostatic interactions and also the DNA bases through hydrogen bonding.
- ▶ Examples
  - ▶ fos and jun
  - ▶ cAMP response element-binding protein (CREB)



[http://en.wikipedia.org/wiki/leucine\\_zipper](http://en.wikipedia.org/wiki/leucine_zipper)

# Helix-loop-helix (HLH) motif

- ▶ Two amphipathic  $\alpha$ -helices connected by a loop.
- ▶ Forms homo- or heterodimers.
- ▶ Dimerization domain has a four-helix bundle structure.
- ▶ Extension of one of the  $\alpha$ -helices from DNA binding domain binds the major groove of DNA.
- ▶ Examples of HLH transcription factors
  - ▶ myoD
  - ▶ myc
  - ▶ max

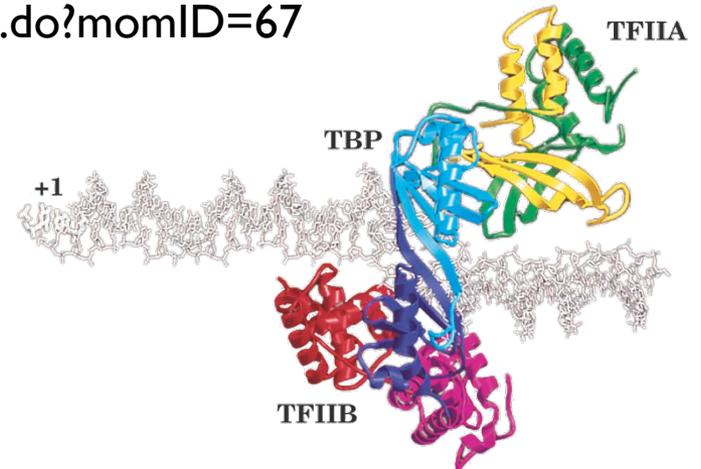


PDB entry 1MDY

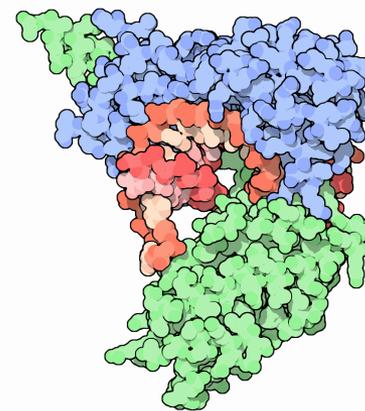
# Beta-scaffold with minor groove contacts: e.g. TATA binding protein (TBP)

PDB-101 entry <http://www.rcsb.org/pdb/101/motm.do?momID=67>

- ▶ Uses a large  $\beta$ -sheet surface to recognize DNA by binding in the minor groove.
  - ▶ Recognizes T-A-T-A-a/t-A-a/t and variations of it
  - ▶ Function=control of which gene gets transcribed.
  - ▶ TATA box is in front of start site of transcription and TBP binding creates a marks for transcription start
- ▶ Binding induces significant changes in the DNA structure
  - ▶ Enables good fit between the protein and DNA bases.
- ▶ The binding of TBP directs assembly of the initiation complex by ordered addition of several general transcription factors and RNA polymerase II.

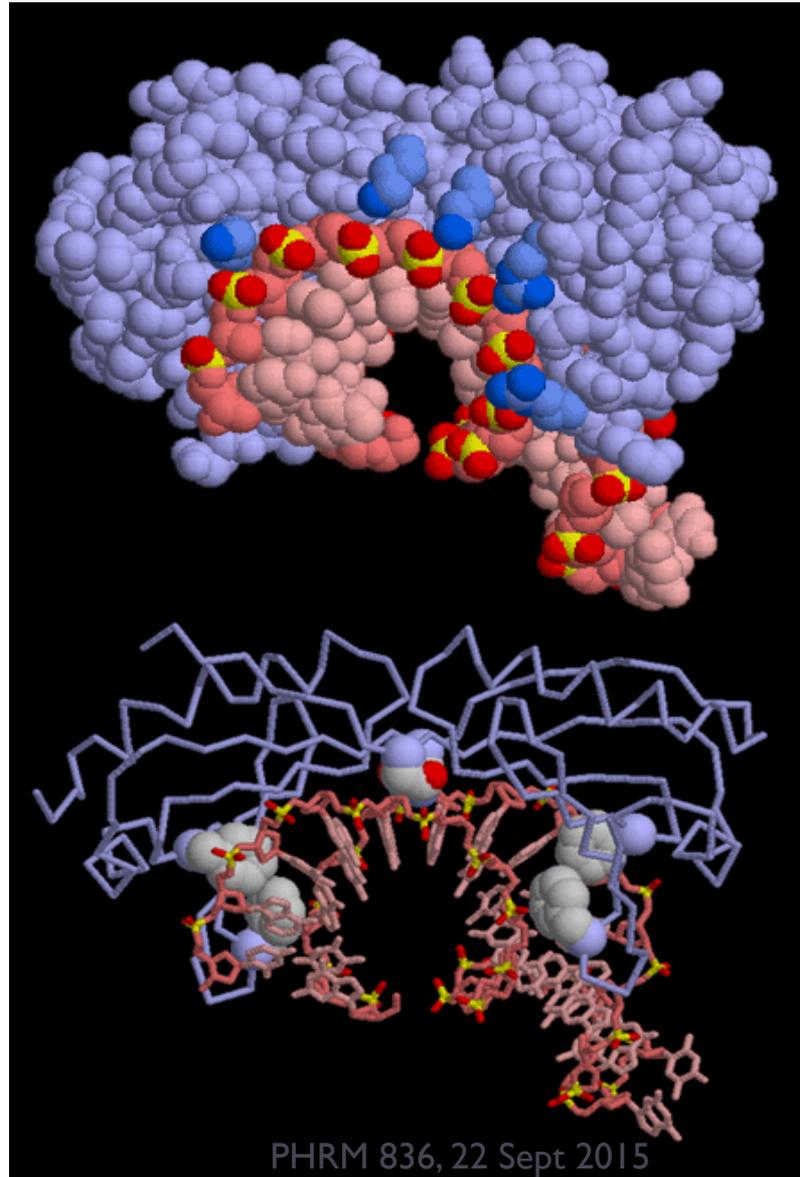


TFIIA, a transcription activator (pdb 1YTF)



TATA-binding protein  
DNA  
Transcription  
regulator (inhibitor)

# Beta-scaffold with minor groove contacts: e.g. TATA binding protein (TBP)



Lys and Arg interact with  
phosphate groups

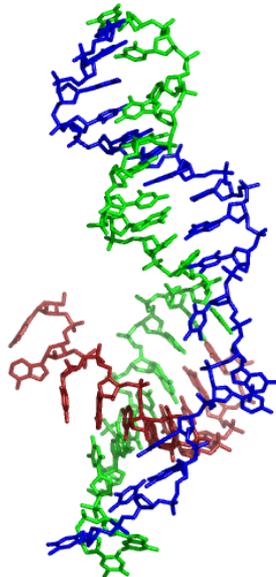
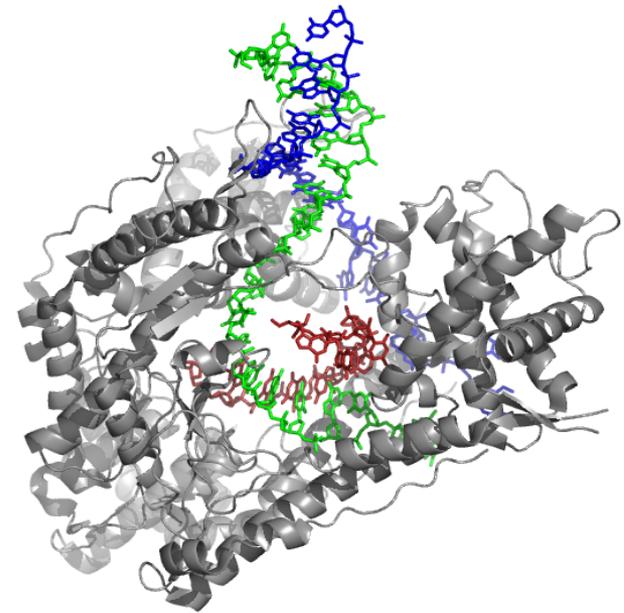
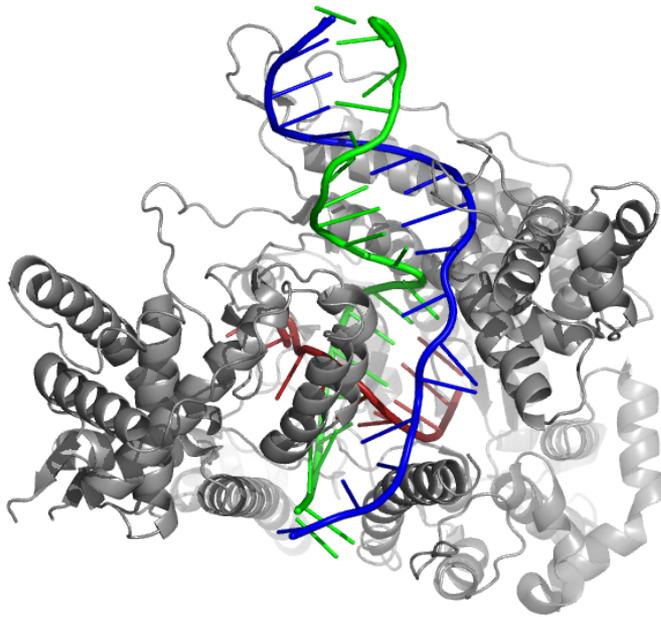
Phe groups jam into the  
DNA minor groove and  
kink the DNA; TATA is  
relatively flexible  
sequence

pdb 1YTB

# General transcription

- ▶ Eukaryotic RNA polymerases
  - ▶ RNA polymerase I transcribes the rRNA genes.
  - ▶ RNA polymerase II transcribes the protein-encoding genes to produce mRNA.
  - ▶ RNA polymerase III transcribes the genes for tRNAs and other small RNAs.
- ▶ Preinitiation complex: binding of RNA polymerase II to a promoter region of DNA requires the initial contact of the promoter with transcription factors (TFs).
  - ▶ TFIID : multi-subunit complex with TATA binding protein (TBP) and different TBP-associated factors (TAFs).
  - ▶ TATA box
    - ▶ Located ~27 bp upstream of the transcription start site.
    - ▶ Recognized by TATA binding protein.
    - ▶ Needs to be dissociated from histones for transcription.

# Bacteriophage T7 RNA polymerase



RNA polymerase, ~880 residues  
DNA, ~20 base pairs

transcribed strand

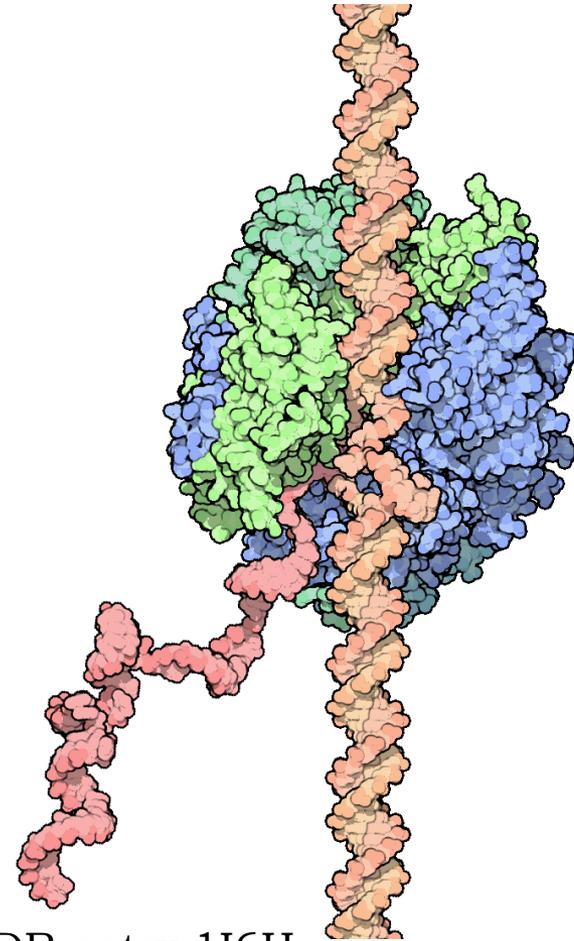
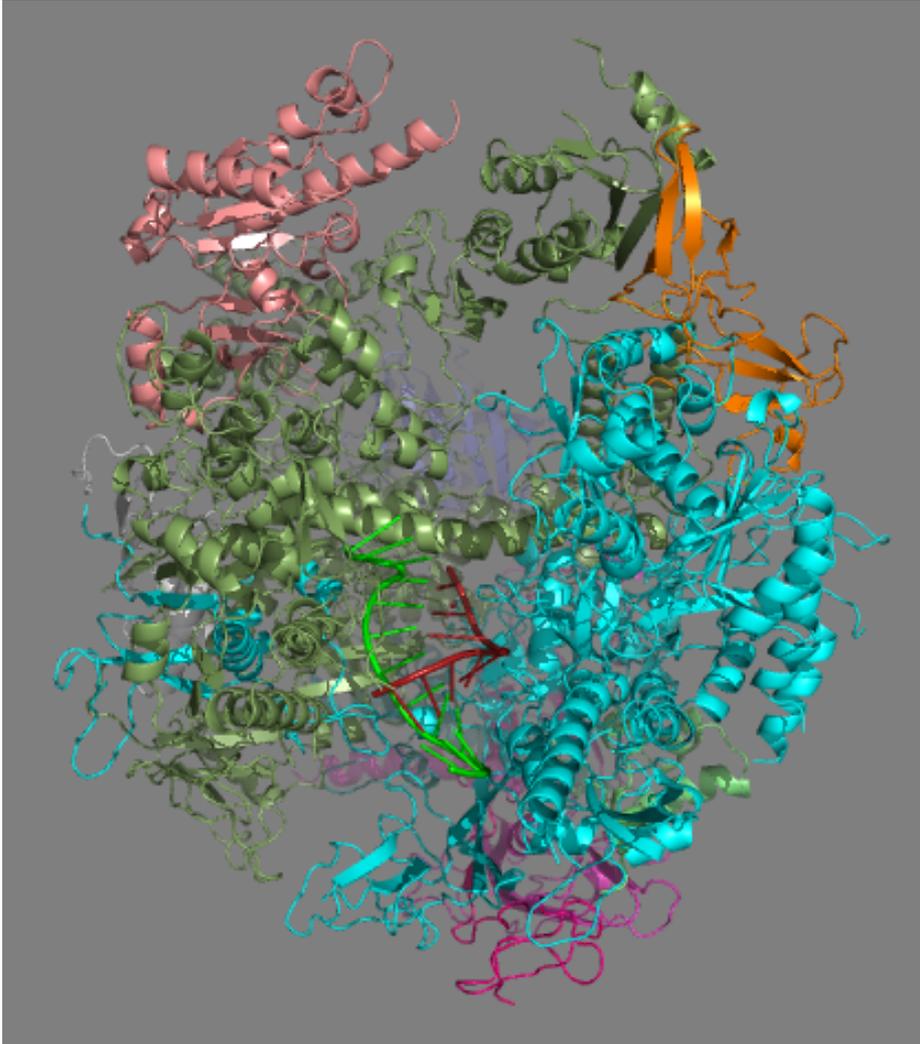
not transcribed strand

RNA, ~10 bases

PDB entry 1MSW

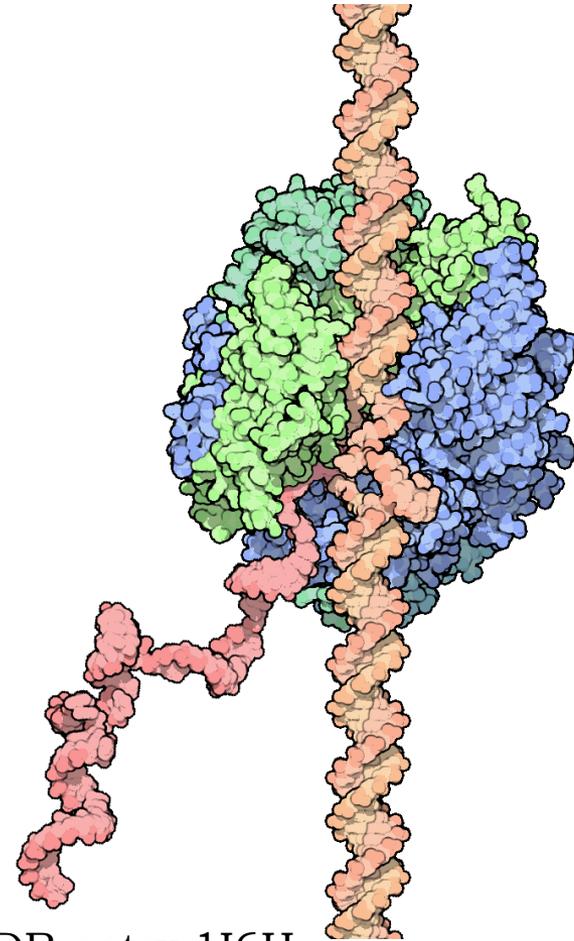
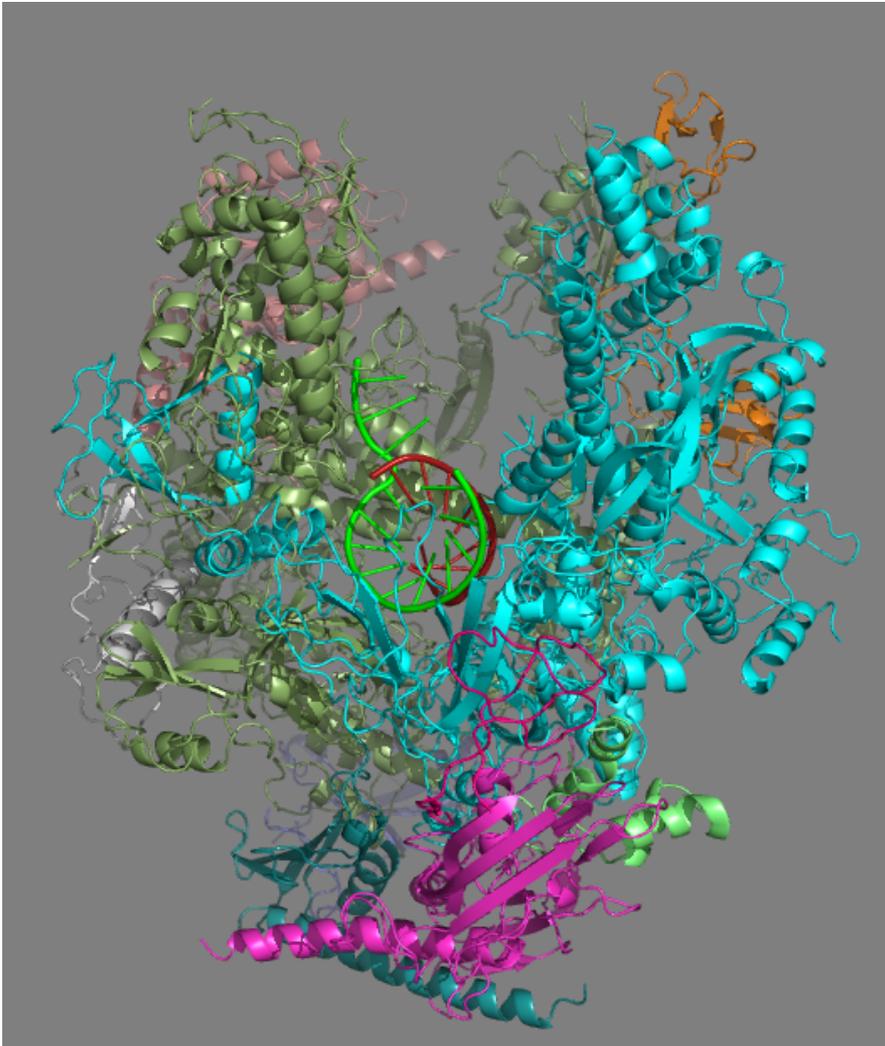
# RNA polymerase II elongation complex

PDB-101 entry <http://www.rcsb.org/pdb/101/motm.do?momID=40>



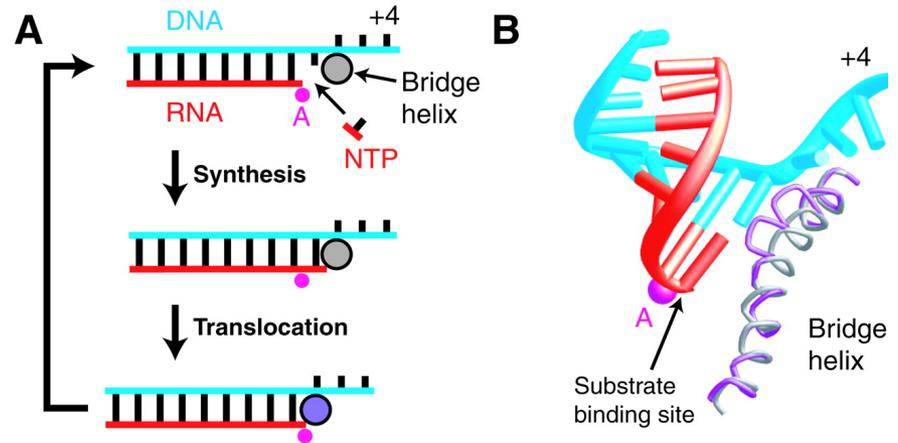
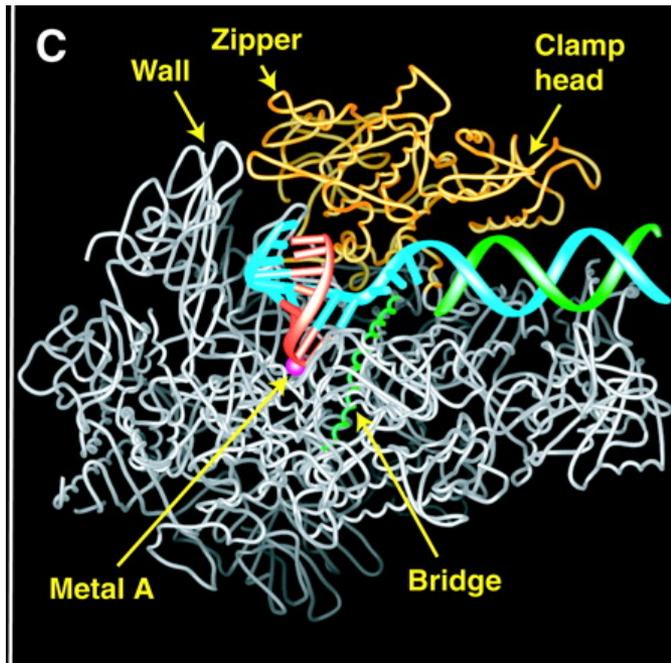
PDB entry 1I6H  
3521 amino acid residues (10 subunits),  
9 RNA bases, 13 DNA bases

# RNA polymerase II elongation complex



PDB entry 1I6H  
3521 amino acid residues (10 subunits),  
9 RNA bases, 13 DNA bases

# RNA polymerase II elongation complex



A mechanism for translocation was speculated from seeing the structure!

Gnatt, Cramer, Fu, Bushnell and Roger Kornberg (2006 Nobel Prize)

Science 8 June 2001:

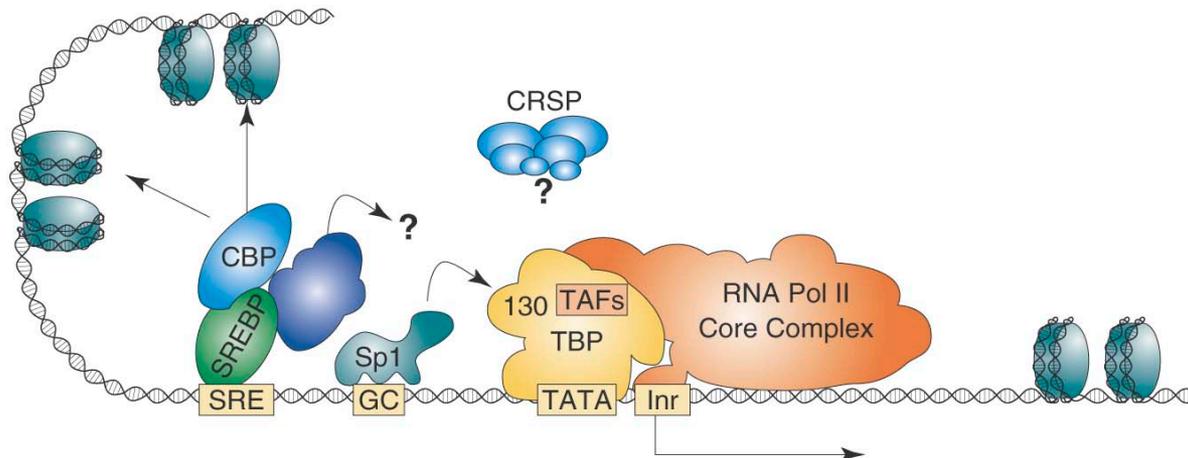
Vol. 292 no. 5523 pp. 1876-1882

This yeast complex has common features with the bacterial complex

# Multiple transcription factor-binding sites regulate eukaryotic transcription

## Promoter regulatory regions on DNA (example of LDL receptor gene)

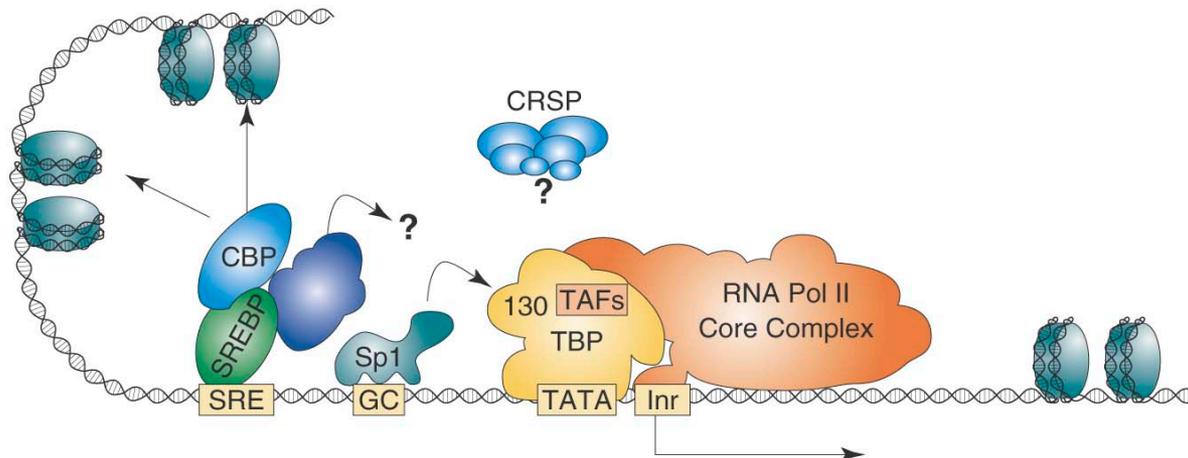
- ▶ In addition to preinitiation-complex assembly site, other TF-binding sites close to the TATA box (e.g. CAAT box and GB box) play a role in transcription
  - ▶ TF Sp1 binds GC boxes via ZnFs; a glu-rich activation domain thought to recruit TFIID with help from CRSP
- ▶ Some TFs are regulated by signaling molecules (e.g. hormones) or chemical modification (e.g. phosphorylation)
  - ▶ Bind to DNA response elements such as SRE (sterol response element) or CRE (cAMP response element)
  - ▶ Examples: SREBP-1A moves to nucleus when cholesterol levels are low; binds SRE via HLH motifs; recruits CBP, which couples chromatin remodeling to transcription factor recognition



# Multiple transcription factor-binding sites regulate eukaryotic transcription

## Promoter regulatory regions on DNA (example of LDL receptor gene)

- ▶ Other proteins can bridge between the TF and the preinitiation complex (rather than TF direct interaction).
- ▶ Recruitment of histone remodeling enzymes, e.g. CBP-p300 complex has histone acetylase activity, which modifies chromatin structure.
- ▶ Enhancer elements: bind TFs that activate transcription by increasing the rate of assembly of the preinitiation complex.
  - ▶ Unlike promoter sequences, can be located many thousands of bp from initiation assembly site (TATA-like site); brought close to preinitiation complex by DNA loop



# *Summary: DNA binding proteins, structure and function*

- ▶ A number of proteins recognize DNA using a variety of structural motifs
  - ▶ Dimeric interactions; palindromes
  - ▶ Flexibility of structures, both DNA and protein
  - ▶ Sequence specific or not
  - ▶ Motifs include HTH, zinc-fingers, leucine zipper, TATA binding protein
- ▶ Transcription factors regulate transcription through binding certain DNA regions and involve interactions with other proteins
- ▶ RNA polymerase recognition of DNA during transcription involves a promoter region and a multi-subunit complex (machine) to conduct a complex process (unwinding DNA, RNA elongation, accurate translation of DNA via proofreading)
- ▶ Numerous proteins involved in regulation of transcription including CBP, Sp1, etc