MIC/BIO/BCH522

Spring 2006

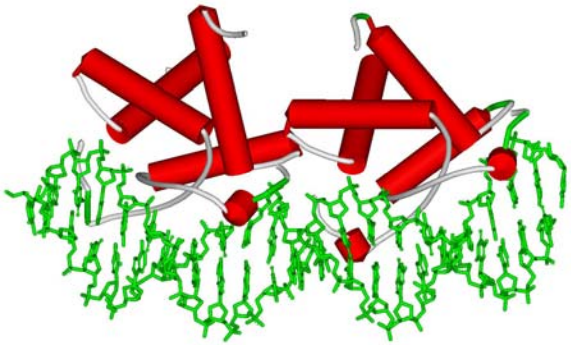
Interactions of proteins with nucleic acids
5. β-ribbon proteins

Beta ribbon group proteins

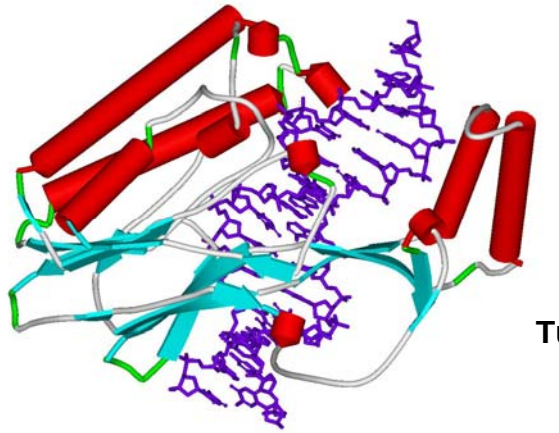
- The defining feature of these proteins is their DNA binding through β structures.
- The interaction with DNA involves the **major groove** in the **ribbon-helix-helix** family,
- and the **minor groove** in the **DNA benders and TATA box-binding proteins**.
- Ribbon-helix-helix domains are found in repressor proteins.
  - They are composed of two identical, intertwined monomers, each consisting of two α helices and one β strand.
  - The β strands from both subunits form an antiparallel β ribbon which protrudes from the core and enters into the major DNA groove.
  - Some proteins of this family have been shown to bind their target site cooperatively as dimers of dimers.

Beta ribbon group proteins

1. **Ribbon-helix-helix**

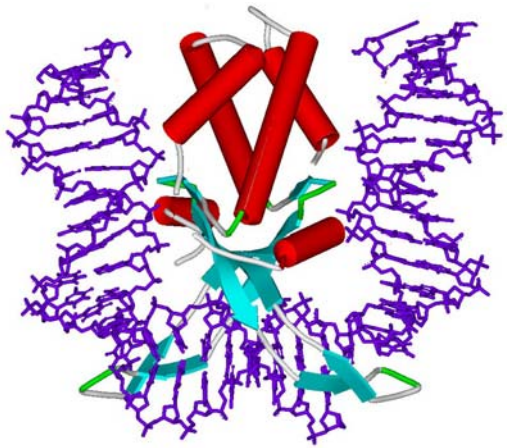


Arc repressor

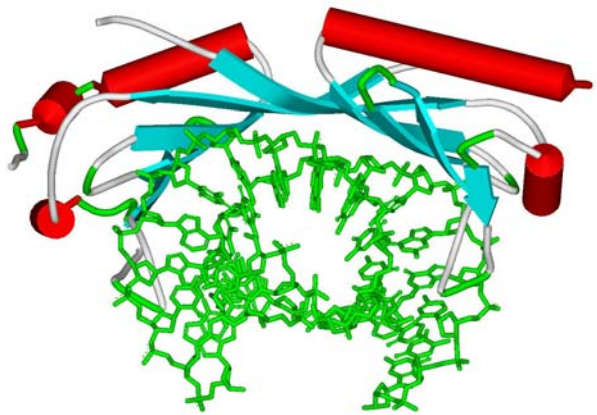


Tus

2. **DNA benders**



IHF

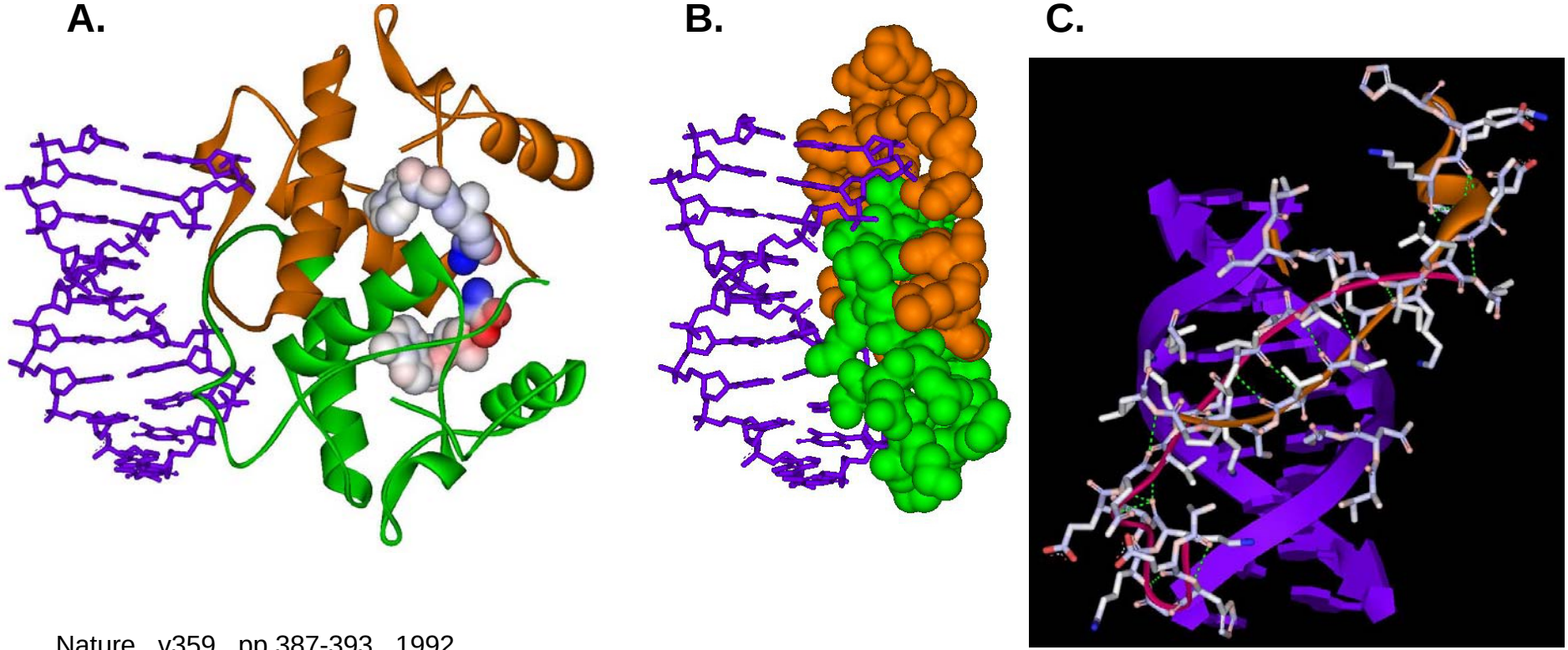


TATA box binding protein

met repressor

- Transcription regulator of structural genes in the methionine biosynthetic pathway. One of the products, S-adenosylmethionine, acts as a corepressor.
- A homodimeric protein of ~100 residue subunits, each containing a beta-strand followed by three alpha-helices.
- Helix 2 and the strand provide the dimer interface by aligning against their counterparts in an anti-parallel fashion. The beta-ribbon formed by the strands acts as the DNA-recognition element.
- Basic repressor binding site, referred to as the "met box" is 8 base-pairs long and is frequently found in two to five tandem repeats.
- Repressors bind cooperatively to adjacent sites to form a left-handed superhelix of proteins around the DNA.
- Base contacts are provided through insertion of the beta-ribbon in the major groove. Surrounding backbone contacts are made by helix 2 and the N-terminal loop in each subunit.
- Somers W S & Phillips S E V (1992). Crystal structure of the met repressor-operator complex at 2.8A resolution reveals DNA recognition by beta-strands. *Nature*, **359**, 387-393.

The met repressor homodimer



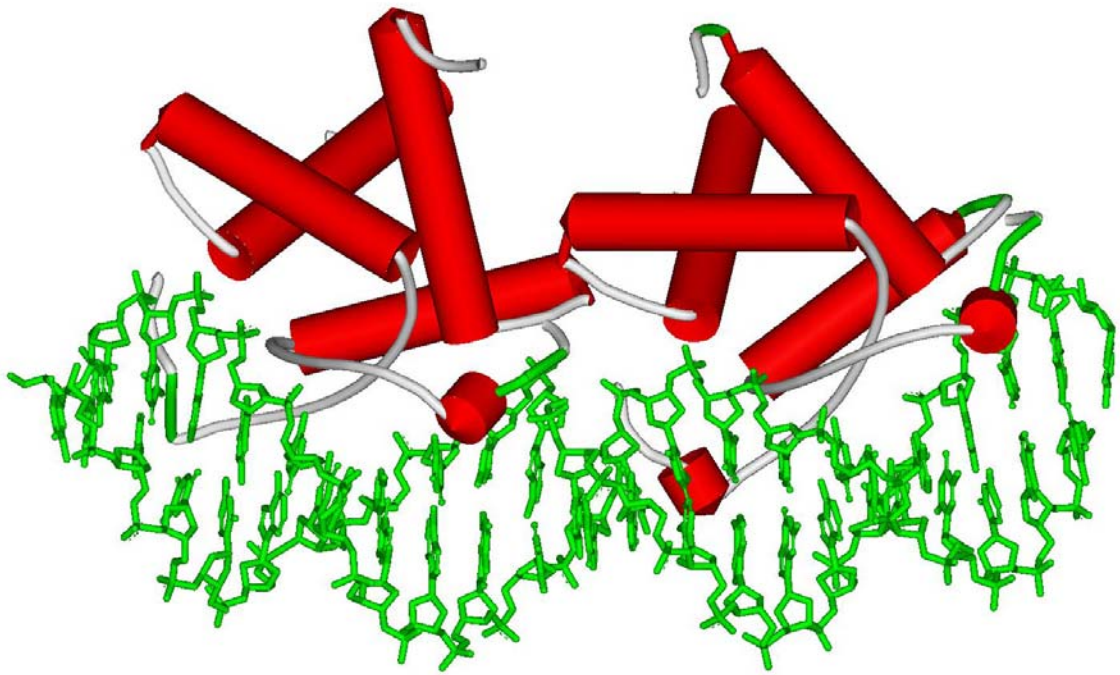
Nature v359 pp.387-393 , 1992

A. Ribbon diagram with SAM molecules shown. B. Surface representation of beta sheets only C. Beta sheets positioned in the major groove of DNA

- The crystal structure of the E. coli met repressor in complex with a synthetic 19-base pair oligonucleotide reveals two dimeric repressor molecules bound to adjacent sites on the DNA.
- The oligonucleotide contains two adjacent repeats of an 8-mer known as a met-box, which represents the consensus of the met operator sites.
- DNA binding takes place through the insertion of a beta-ribbon into the major groove of B-form DNA, representing a novel DNA binding motif.
- Sequence specificity arises from direct interactions between side chains of the beta-strands and the edges of the bases in the major groove.
- The repressor is activated through binding of S-adenosyl methionine (SAM), the corepressor, to the face opposite to that used for DNA binding. The lack of significant conformational change upon SAM binding, together with electrostatic calculations, suggests that DNA binding enhancement occurs through long-range electrostatic interactions.

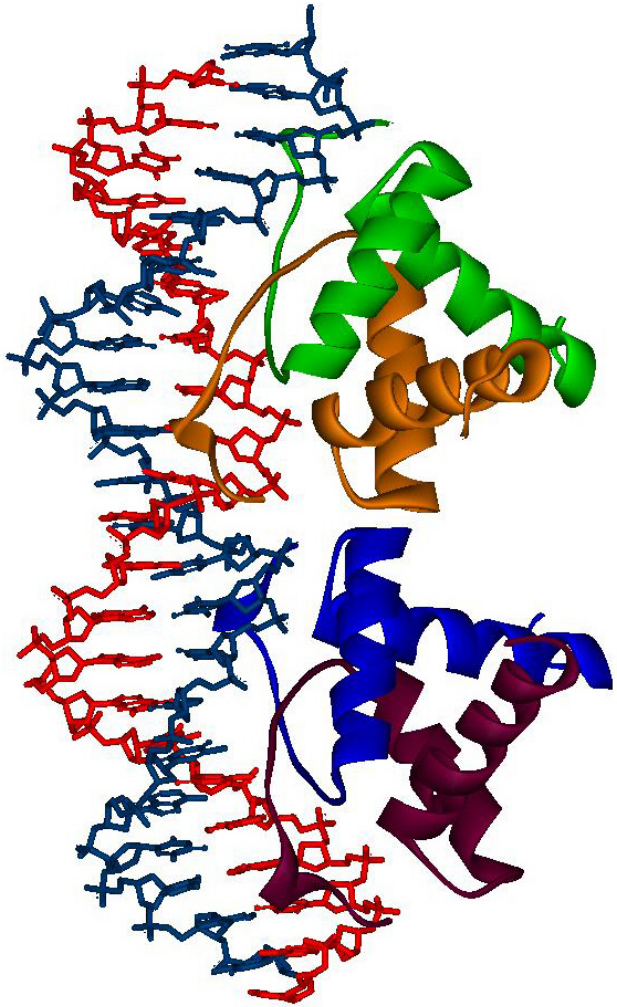
Arc repressor

- Transcription of the ant gene during lytic growth of bacteriophage P22 is regulated by the cooperative binding of two Arc repressor dimers to a 21-base- pair operator site.
- Arc is a small (~100 residue), homodimeric repressor of the ribbon-helix- helix family of transcription factors.
- Each monomer consists of a pair of helices connected by an anti-parallel beta-sheet.
- Each Arc dimer uses the beta-sheet to recognize bases in the major groove and the N-termini of the second helix in each pair contact the DNA backbone.

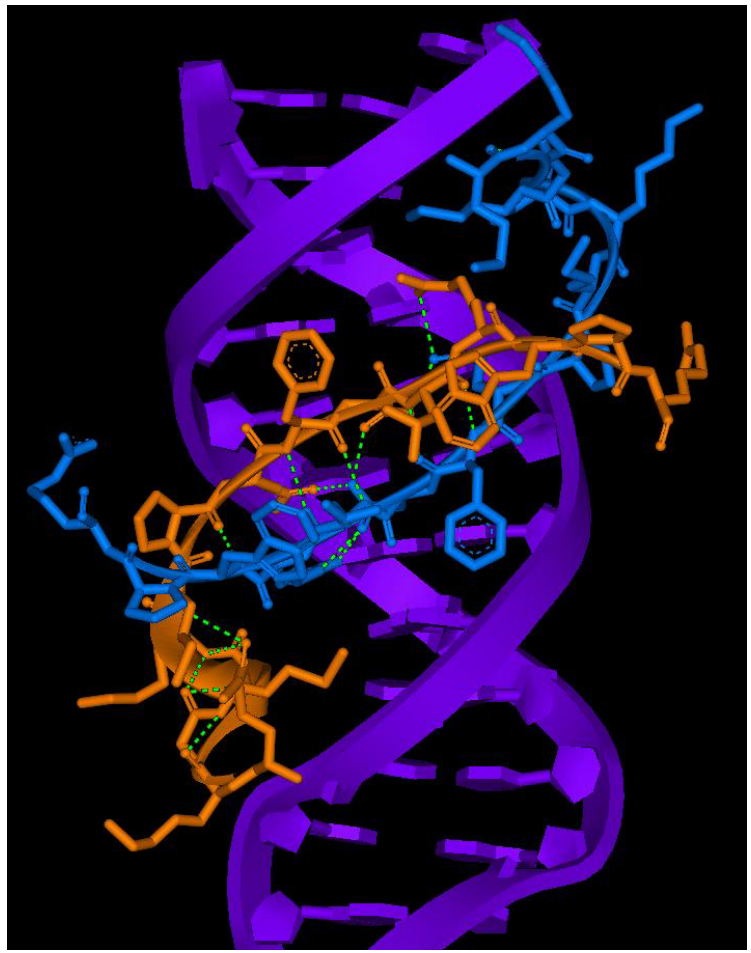


Arc repressor binds to DNA as a dimer of dimers

A.



B.



DNA recognition by beta-sheets in the Arc repressor-operator crystal structure. (Nature v367 pp.754-757 , 1994)

A. Dimer of dimers bound to DNA. Each dimer is colored differently. The upper dimer is orange/green; lower blue/maroon.

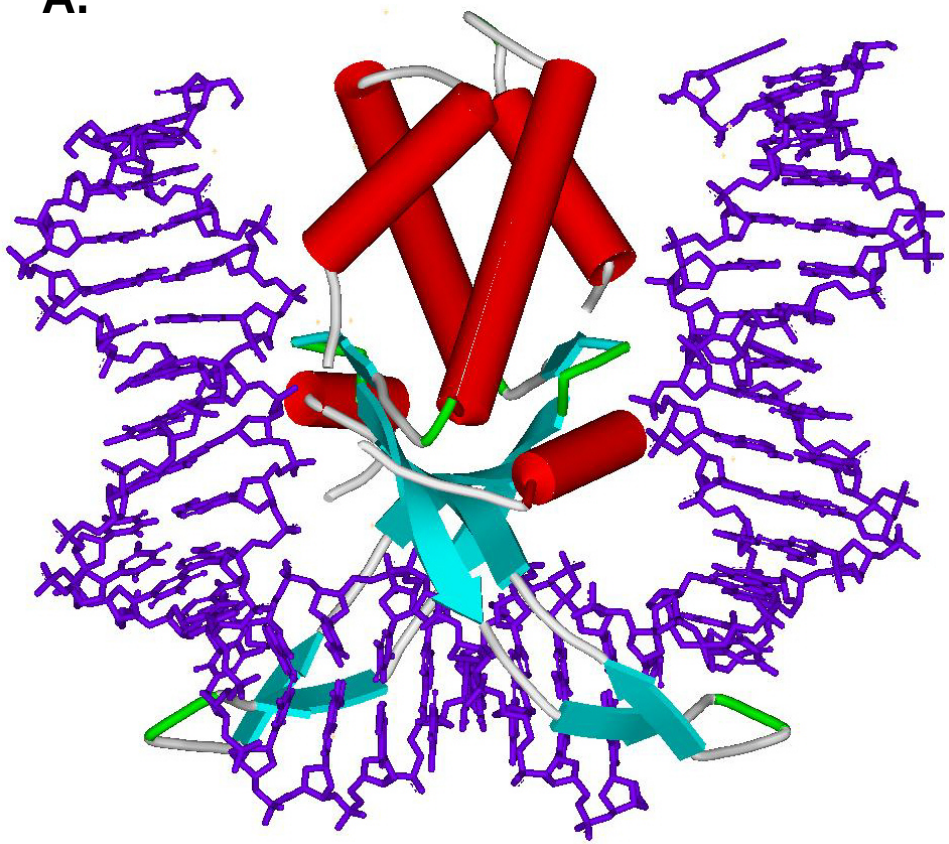
B. Positioning of the two strand of the beta sheet of one monomer positioned in the major groove.

Integration host factor

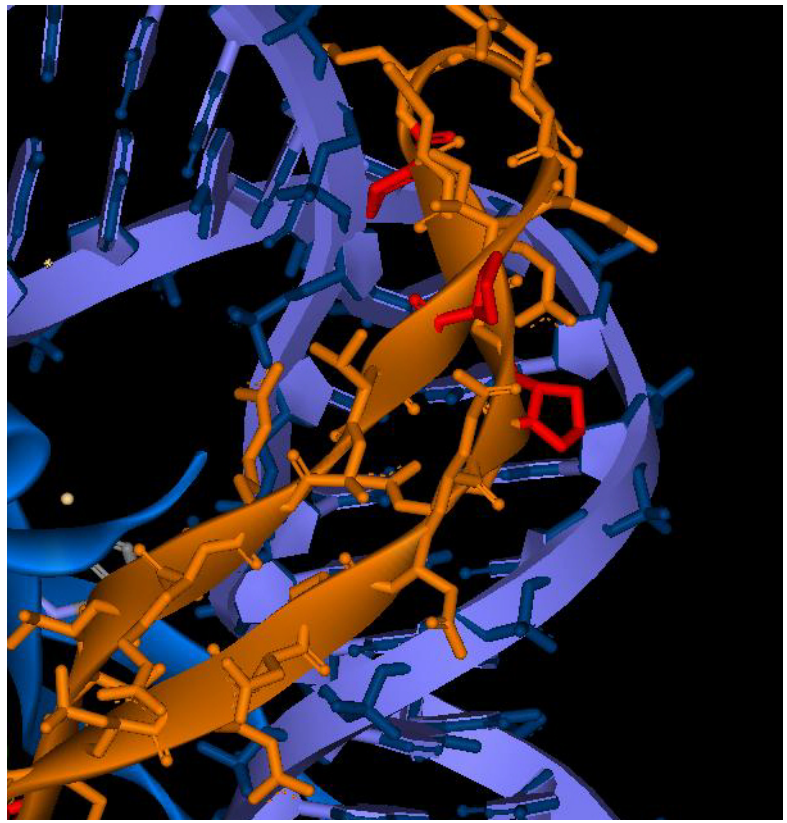
- **T**he genetic material in bacterial cells is organized in a structure called the nucleoid.
- In *Escherichia coli*, nucleoid consists of a 4.7 Megabases of DNA, RNA, and a large variety of bound proteins.
- Among these, are 10 “histone-like proteins”, including HU, integration host factor (IHF), and H-NS
- These proteins shape the short-scale structure of the nucleoid by bending DNA locally on binding.
- These proteins play an important role in compacting the DNA molecule,
- **IHF is** an architectural protein which assists formation of high order protein-DNA complexes such as those found in replication and long distance transcription regulation.
- A heterodimeric protein with ~90 residue subunits that have very similar three-dimensional structures.
- Each subunit consist of both alpha-helices and beta-strands which intertwine to form a compact body from which two beta-ribbons extend.
- Binds to DNA in a sequence specific manner to a 35 base-pair site.
  - Very little consensus is observed between the sites, however many contain a poly-A tract.
- Binding is pseudosymmetric and is found entirely in the **minor groove**.
- The 180 degree bend in the DNA structure is achieved by intercalation of conserved proline side chains from the tips of the beta-ribbon arms.
- The DNA strands exiting from the U-turn are "clamped" in place by the rest of the protein.

IHF

A.



B.

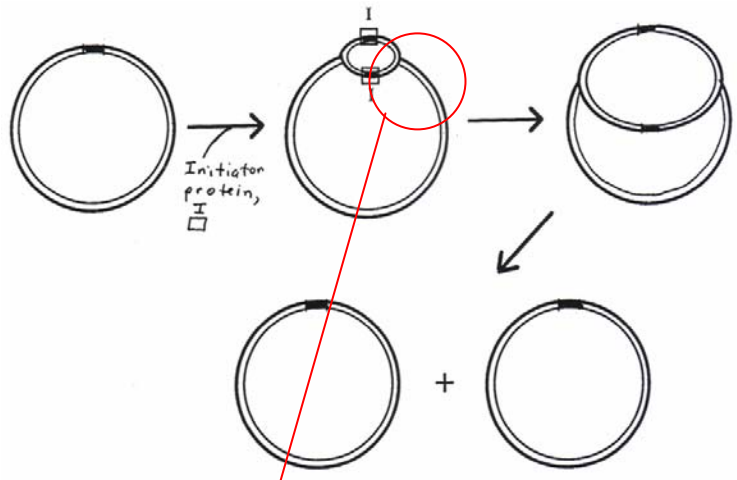


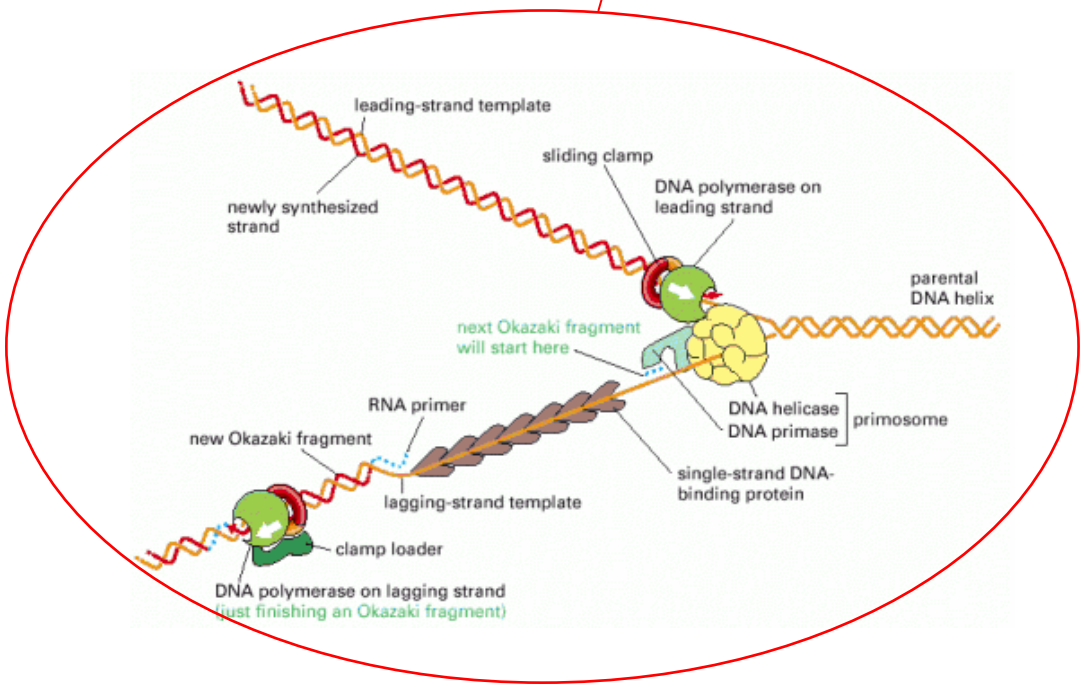
Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. Cell v87 pp.1295-1306 , 1996

A. IHF dimer bound to DNA inducing a 180° bend

B. B. Close-up of one beta sheet inserted into the minor groove. Conserved proline residues are highlighted in red

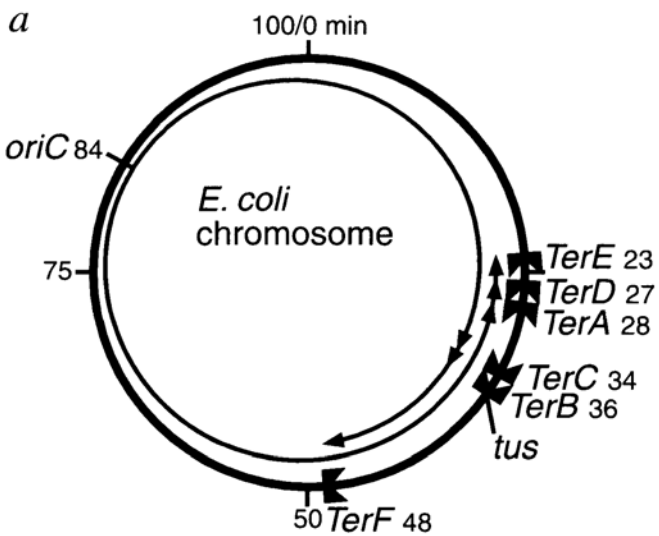
DNA replication termination in *E.coli*





The proteins at a bacterial DNA replication fork. The major types of proteins that act at a DNA replication fork are illustrated, showing their approximate positions on the DNA.

Tus replication terminator



b

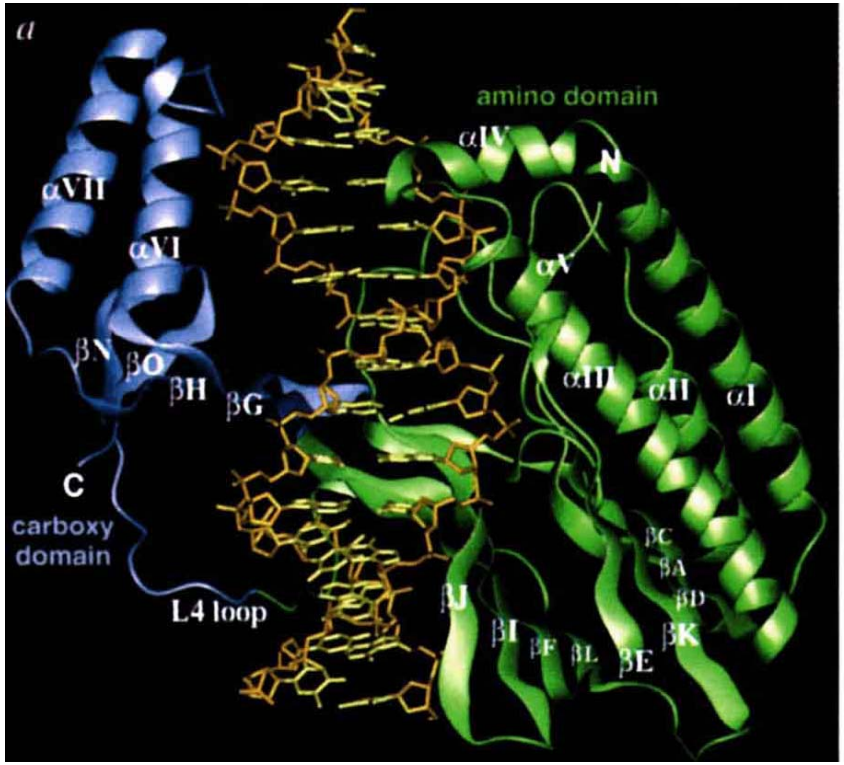
E. coli chromosome

TerA	TTTAGTTACAACATACTAATT
TerB	T----------TATT
TerC	A----------C--ATAT
TerD	T----------AATG
TerE	C----------TTAA
TerF	CG-C----------GAAGG

- A terminator protein which stops a replication fork travelling in one direction but not the other. The proposed mechanism involves direction dependent inhibition of DnaB helicase.
- The Tus protein binds to a set of six consensus binding sites named Ter. Each is 20 base-pairs long.
- The DNA is bound in the protein cleft, with the interdomain beta-strands contacting bases in the major groove.
- DNA backbone contacts are provided by all three regions of the protein.
- The directional nature of replication arrest may be explained by the assymetry of the Tus-Ter complex.
- The helical bundles are positioned on the helicase-blocking side, thus preventing dissociation of the central beta-strand region from the major groove.
- A helicase approaching the complex from the opposite direction may easily disrupt these interactions.
- Furthermore, phosphate backbone contacts in the interdomain region are restricted to one DNA strand.

- A ~300 residue monomeric protein which is divided into three regions which jointly produce a large central cleft.
  - C-terminal domain: smaller domain composed of a two helix bundle and two stranded beta-sheet.
  - N-terminal domain: larger domain consisting of a 5 helix bundle and beta-sandwich.
  - Central beta strands: four antiparallel beta-strands connecting the domains. They can be viewed as extensions of the beta-sheets from the two domains.

Tus



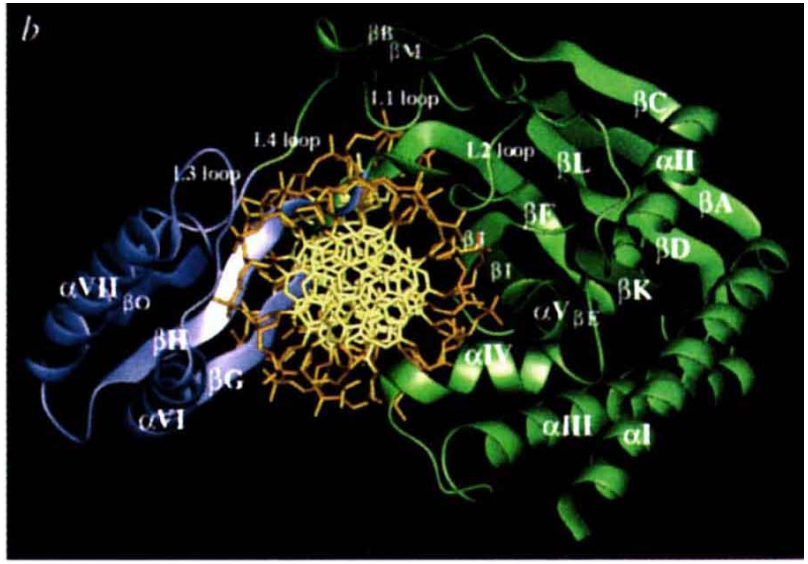
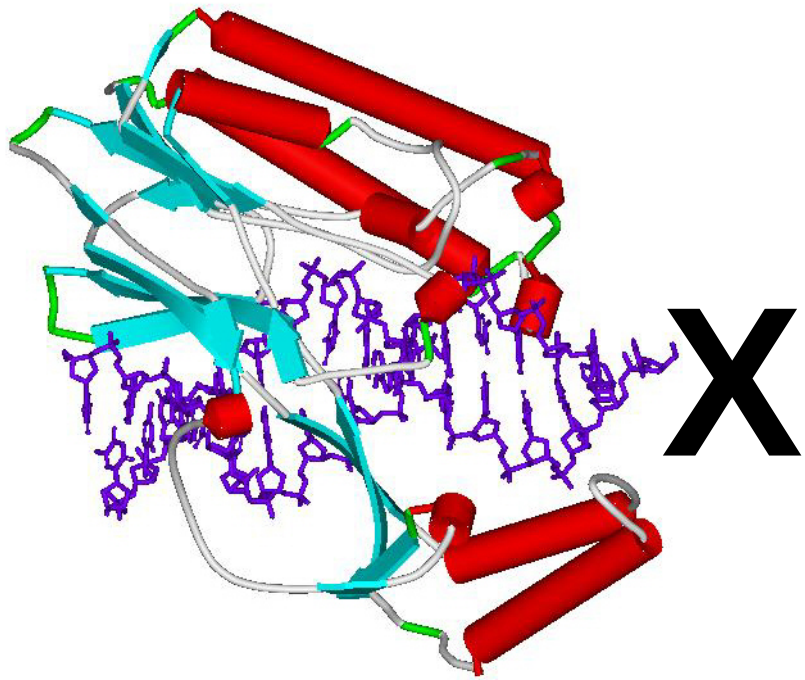


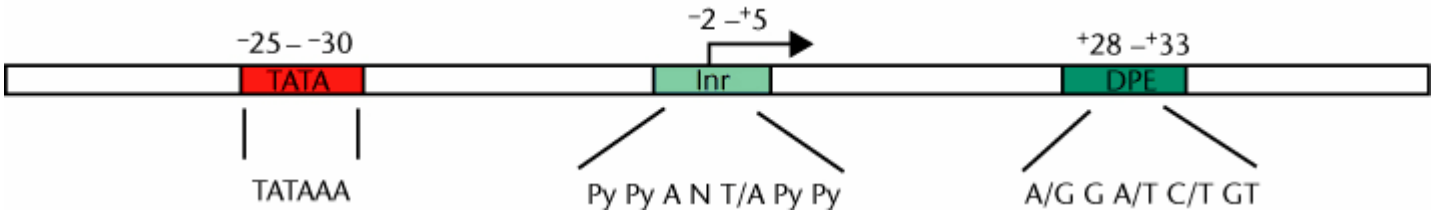
FIG. 2*a*, Ribbon drawing of the complex viewed perpendicular to the helical axis of the DNA. Amino and carboxy domains are coloured green and blue, respectively. *b*, Ribbon drawing of the complex displayed from the fork-blocking side. *c*, Sequence and secondary

(Nature v383 pp.598-603 , 1996)

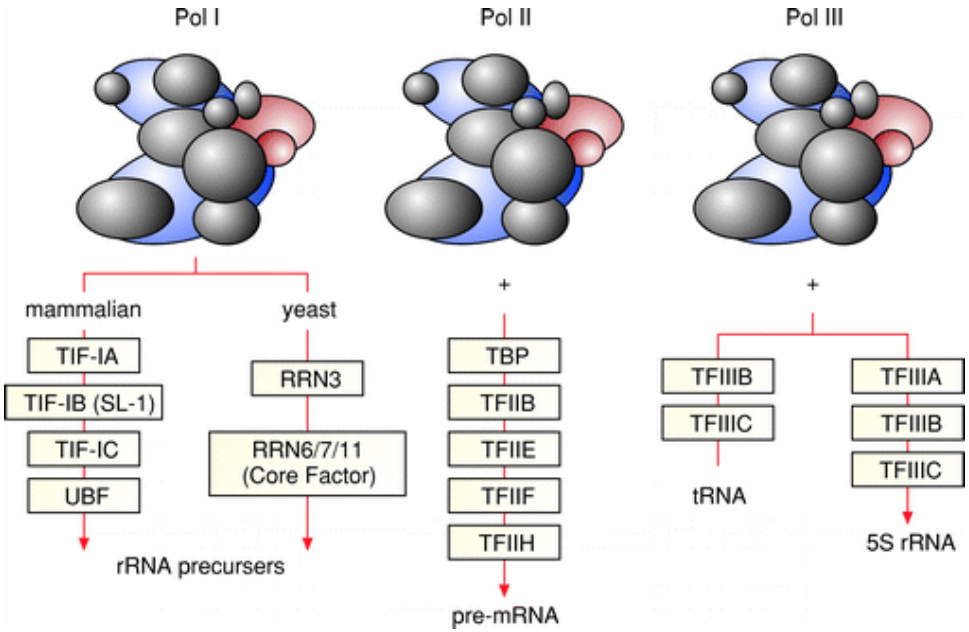


X

Transcription in eukaryotes

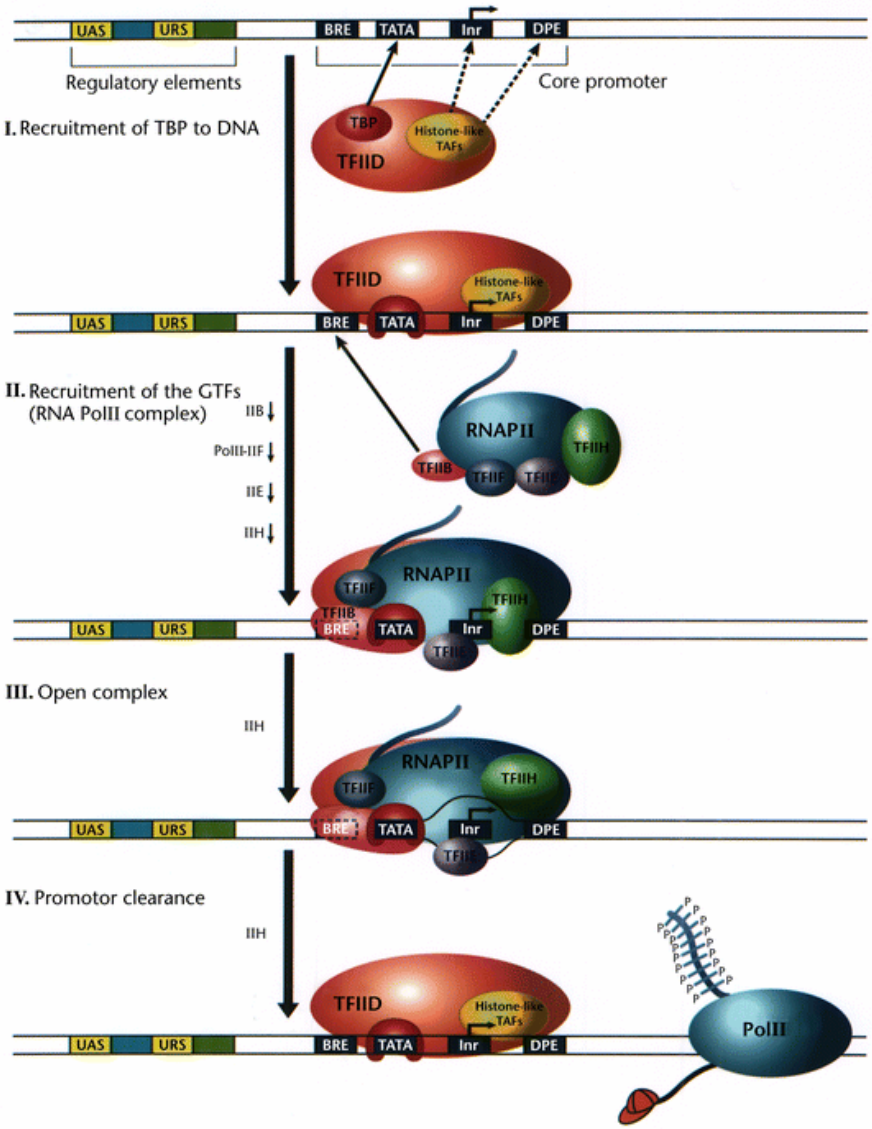


Core promoter elements. Three DNA elements that are commonly found in core promoters for mammalian genes are shown, along with the consensus sequence for each element. Inr: initiator element; DPE: downstream promoter element; Py: C or T; N: any nucleotide.



Reconstitution of selective transcription assays with RNA polymerases I, II, and III. The three types of RNA polymerases are shown with the respective protein fractions required to achieve promoter-selective transcription *in vitro*. Many of the activities represented in the protein fractions contain multiple polypeptides. Alternate names are shown in parentheses.

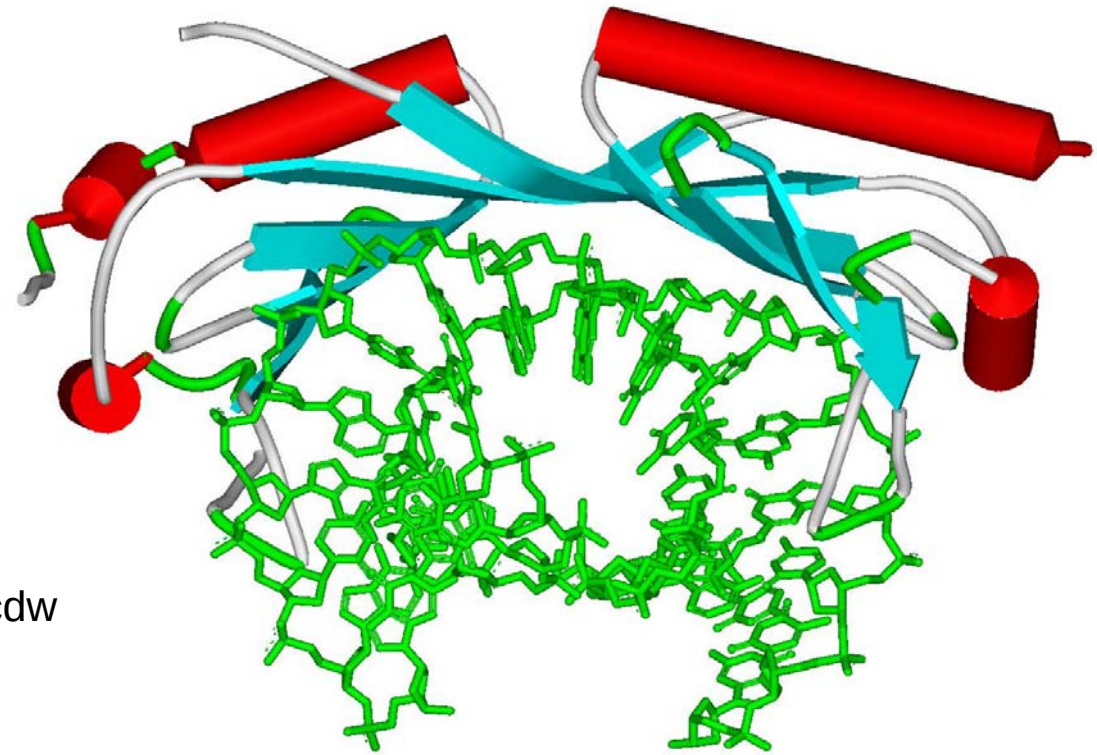
The RNA polymerase II (RNAPII) transcription initiation machinery.



A typical RNAPII promoter includes core elements that define the start site of transcription and regulatory elements that affect transcription either positively or negatively. Core elements include the TATA box, which binds the TATA-binding protein (TBP); the TFIIB-responsive element (BRE); the initiation region (Inr), which encompasses the transcription start site; and downstream elements, which bind TBP-associated factor (TAF) components of TFIID. Regulatory elements are typically located 50–1500 bp upstream of the core promoter and bind sequence-specific factors that interact, either directly or indirectly, with the core machinery to stimulate (UAS) or repress (URS) transcription. The core machinery shown here includes RNAPII and the general transcription factors TFIID, composed of TBP and TAFs, TFIIB, TFIIE, TFIIF and TFIIH. TBP, presumably as a subunit of TFIID, binds the TATA box to nucleate assembly of the initiation complex, followed by association of the other factors, either stepwise or as components of an RNAPII holoenzyme complex. TFIIH catalyses open complex formation (promoter melting) and phosphorylation of the C-terminal repeat domain (CTD) of RNAPII, followed by promoter clearance.

TATA binding protein

- TATA box-binding proteins play a key role in eukaryotic activator-dependent transcription.
- They specifically recognize AT-rich DNA sequences extending over 8 bp,
- with consensus TATA(A/T)A(A/T)N (/ denotes ‘or’; N denotes any base), corresponding to the best documented transcription promoter sequences in eukaryotes.
- DNA binding is mediated by a saddle-shaped, eight-stranded anti-parallel β sheet, which forms a large concave surface for DNA interactions.
- Upon DNA recognition, hydrophobic side-chains are partially intercalated into the minor groove
- DNA is kinked and locally unwound so that the minor groove edges can make contact with the protein.



- 1cdw

TB protein structure

- A monomeric protein that may be separated into two domain: the variable N-terminal domain and the highly conserved C-terminal domain.
- ~180 residue C-terminal domain contains two almost identical symmetrical folds
- Each half consists of five beta-strands and two alpha-helices.
- The strands jointly form a wide saddle-shaped beta-sheet with a pair of short helices flanking either end.
- The long helices span across the top of the sheet almost perpendicular to the direction of the strands. binding is to a 8 base-pair site with the consensus sequence TATAXAXN, where X is an A-T base-pair and N is any base.
- The central beta-sheet approaches the DNA from the minor groove side and mediates all base and backbone contacts.
- The DNA is bent by 180 degrees in the central region, allowing a close fit to the concave beta-sheet surface.
- The DNA is sharply kinked through intercalation of two phenylalanine residues between the first and last base steps of the binding site.
- The residues are situated in loops at either end of the sheet.