

Structural basis of transcription factors

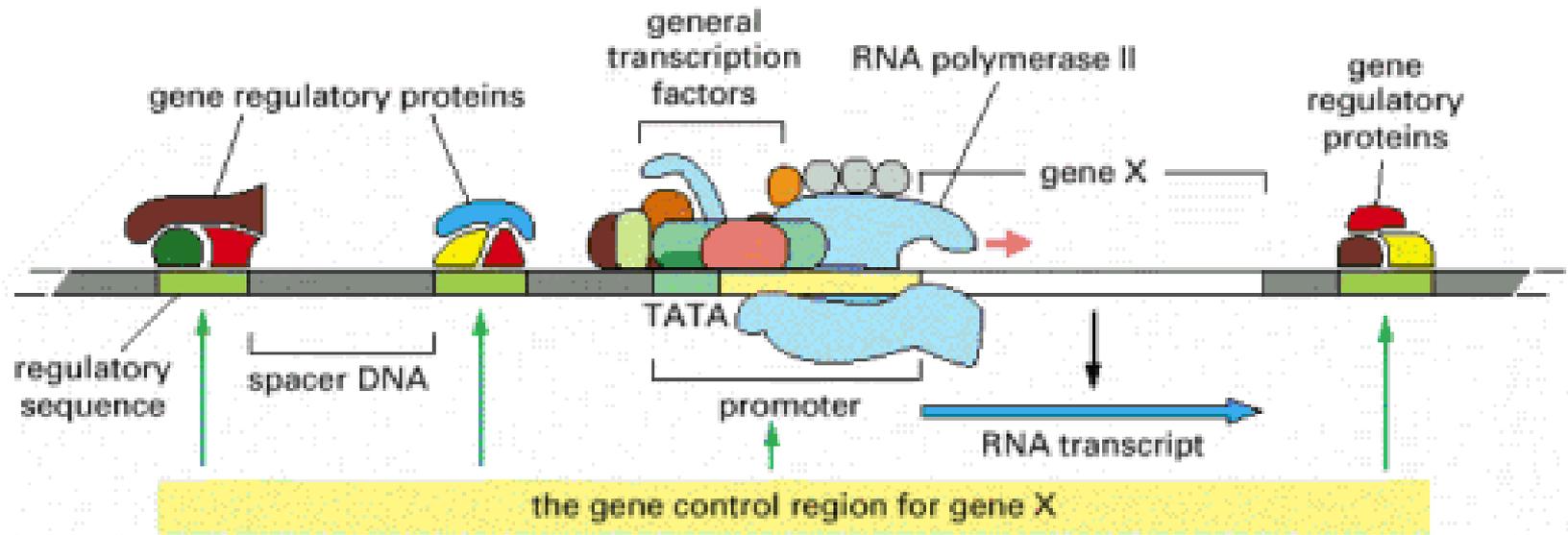
Luo Jingchu

30 Jan 2007

Outline

- General concepts of TFs
- Types of TFs
- Resources of 3D information of TFs
- DATF TF structure information
- Literature references

Eucaryotic gene and transcription factors

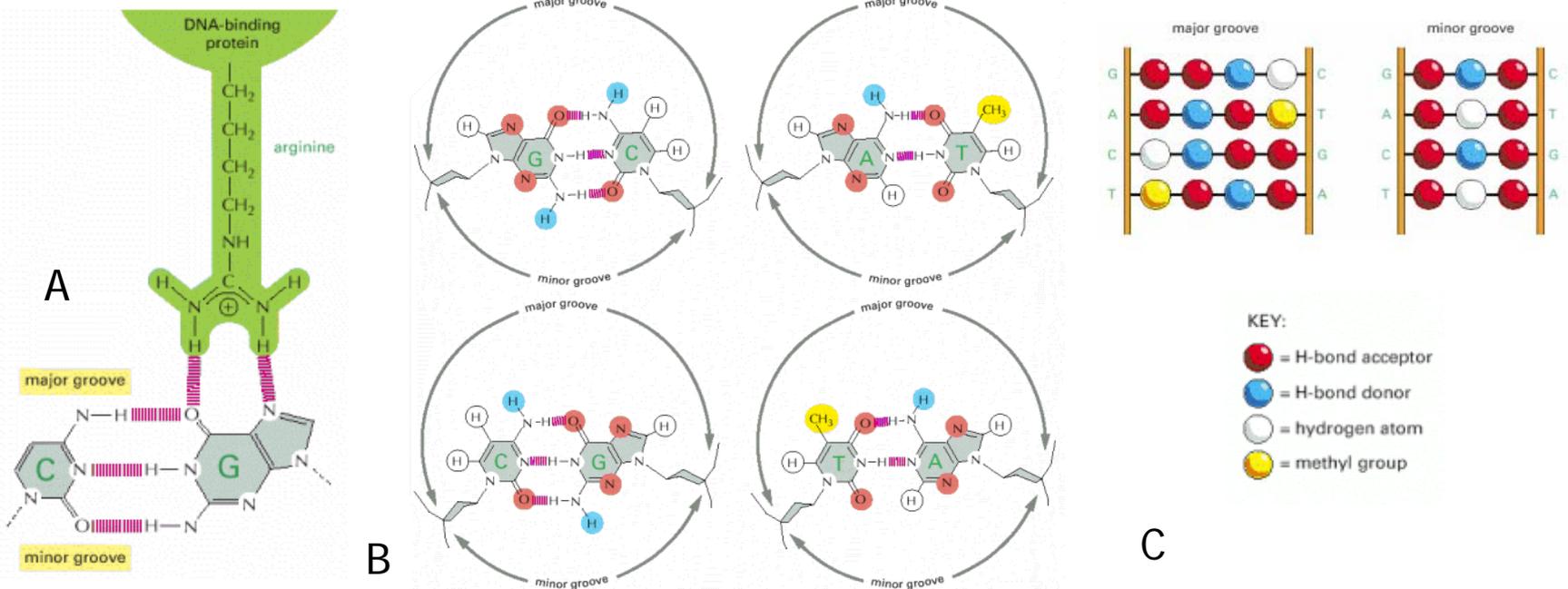


- RNA polymerase II: bind to promotor
- General transcription factors: bind to promotor
- Gene regulatory proteins (TFs) : bind to cis-element
(Alberts et al. MBC, 2002, 4th ed. Figure 7-41)

Types of TFs

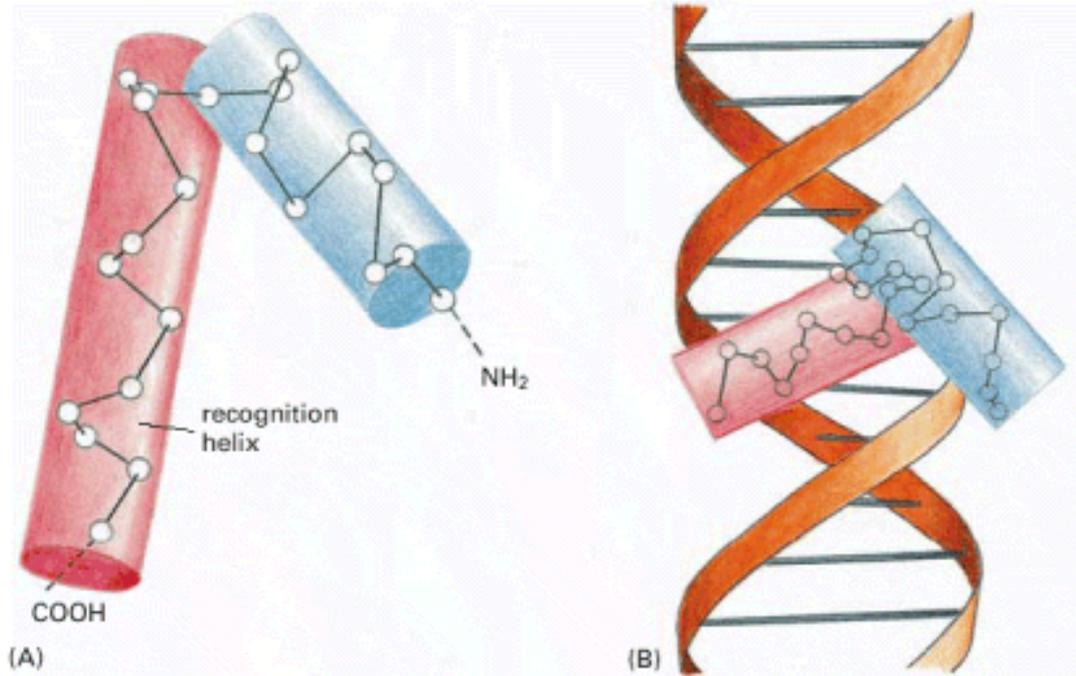
- Helix-turn-helix
- Helix-loop-helix
- Zinc Finger
- Leucine zipper
- Other types

Base pairs for protein-DNA interaction



- A. Hydrogen bonds formed between Arginine and Guanine - One of the most common protein-DNA interactions.
- B. The different base pairs in DNA can be recognized from their edges without the need to open the double helix.
- C. Each base-pair in the major groove has a unique pattern. The patterns in the minor groove are similar.
- C. Color code of base-pairs in the major and minor groove
- (Alberts et al. MBC, 2002, 4th ed. Figure 7-7 and 7-8, P. 381)

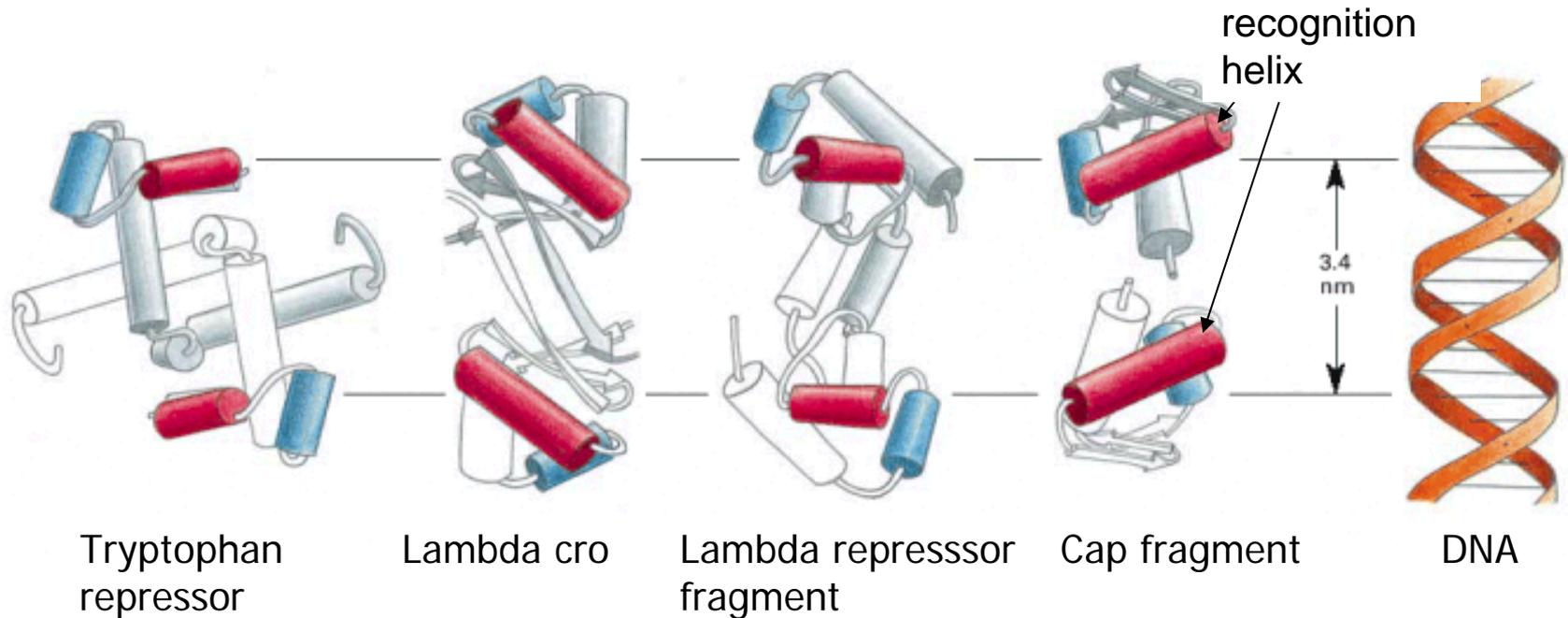
The DNA-binding helix-turn-helix motif



- (A) Each white circle denotes the central carbon of an amino acid. The C-terminal α helix is the recognition helix.
- (B) The C-terminal helix fits into the major groove of DNA, where it contacts the edges of the base pairs. The N-terminal helix helps to position the recognition helix.

(Alberts et al. MBC, 2002, 4th ed. Figure 7-13, P. 384)

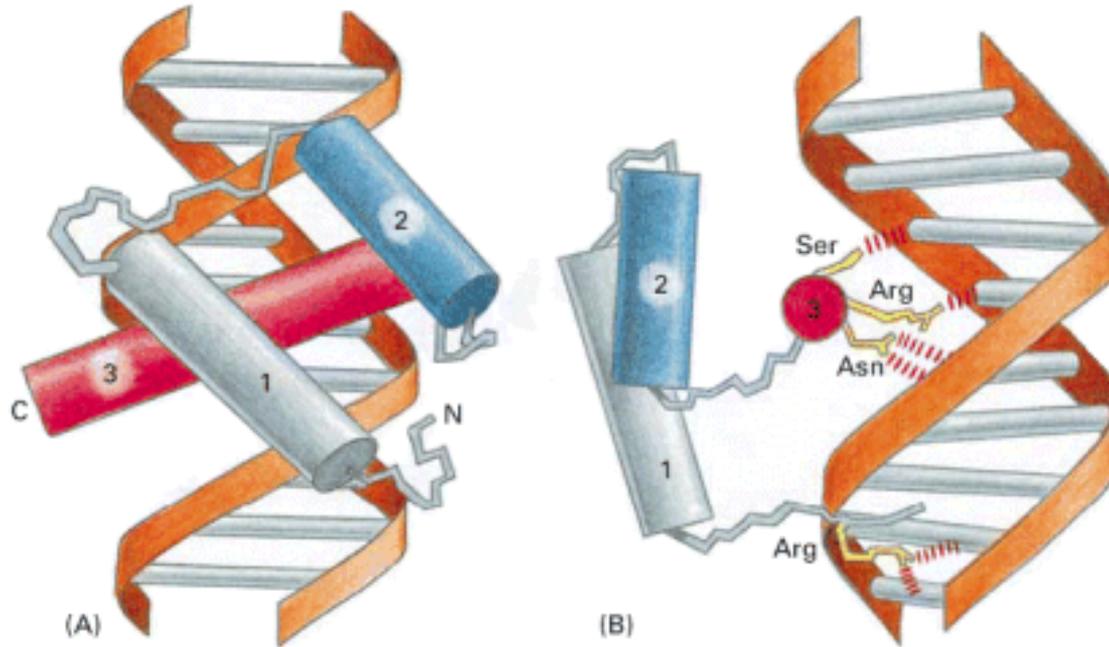
Helix-turn-helix DNA-binding proteins



- All of the proteins bind DNA as dimers in which the two copies of the recognition helix are separated by exactly one turn of the DNA helix (3.4 nm).
- 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.

(Alberts et al. *MBC*, 2002, 4th ed. Figure 7-14, P. 387)

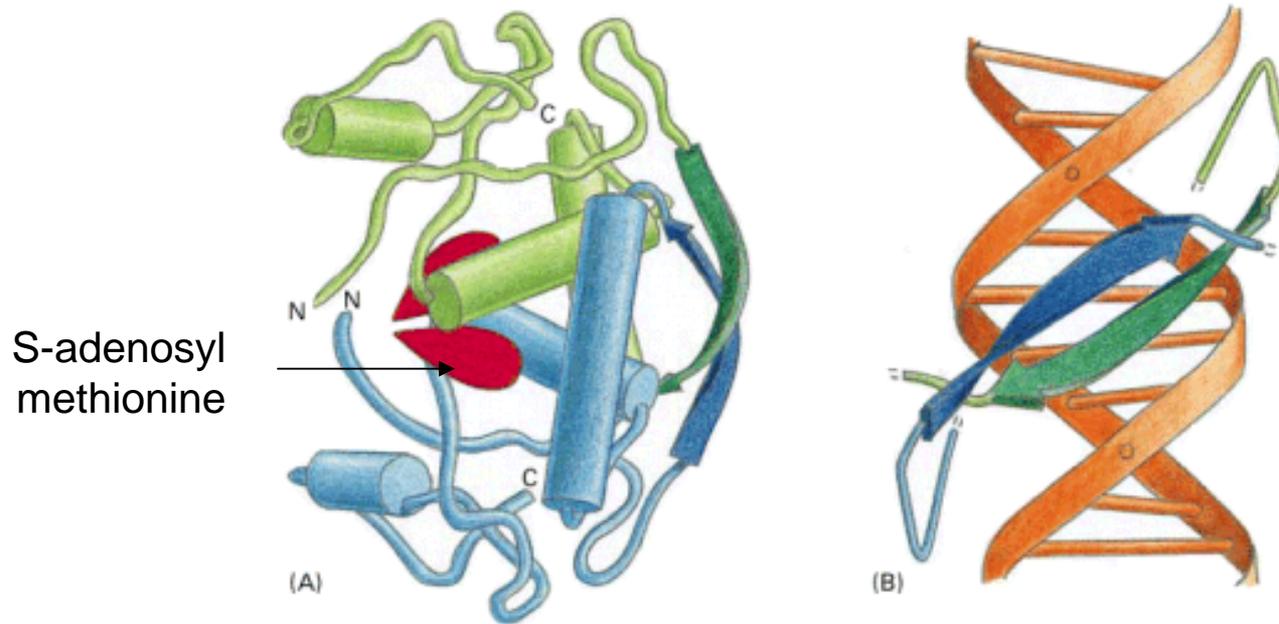
Homeodomain and DNA complex



- (A) The homeodomain from yeast 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 makes important contacts with the major groove of DNA. The Asn of helix 3 contacts an adenine. Nucleotide pairs are also contacted in the minor groove by a flexible arm attached to helix 1.

(Alberts et al. MBC, 2002, 4th ed. Figure 7-16, P. 386)

The bacterial met repressor protein

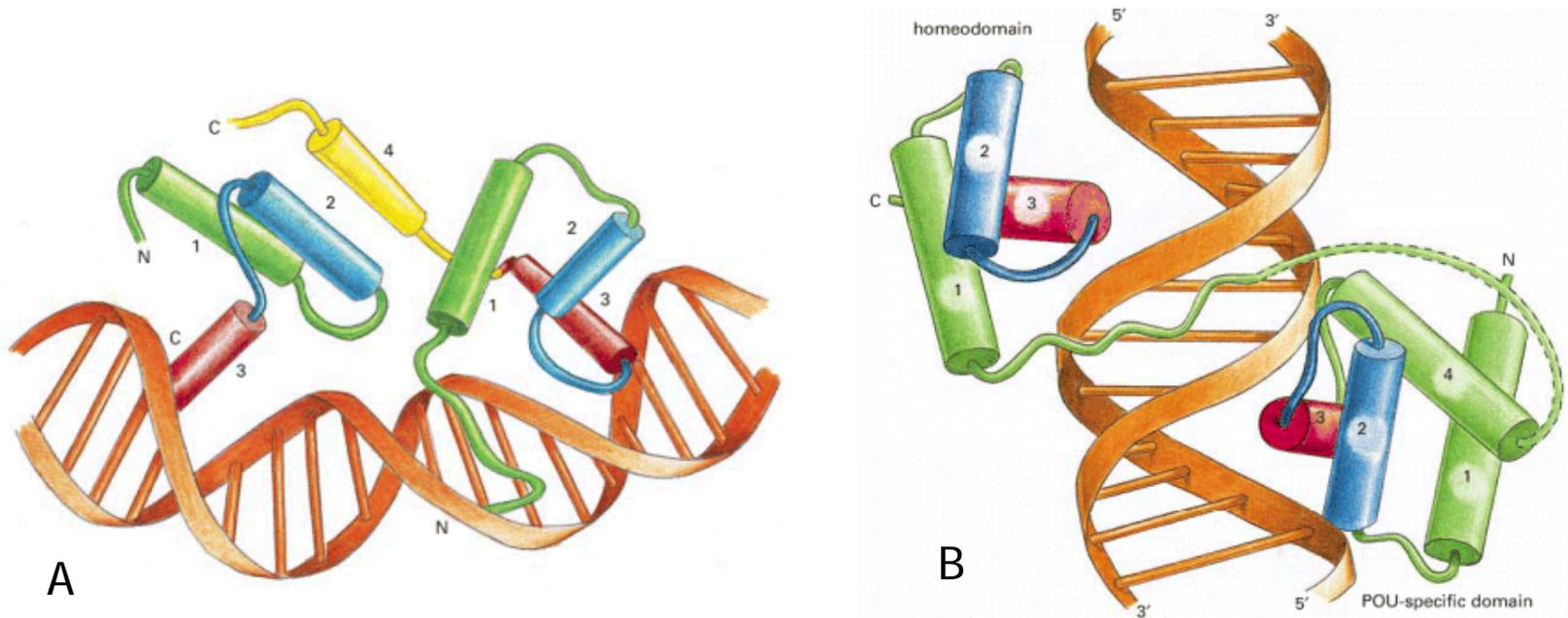


(A) In order to bind to DNA tightly, the met repressor must be complexed with S-adenosyl methionine. The two-stranded β sheet that binds to DNA is formed by one strand from each subunit and is shown in dark green and dark blue.

(B) Simplified diagram of the met repressor bound to DNA, showing how the two-stranded β sheet of the repressor binds to the major groove of DNA.

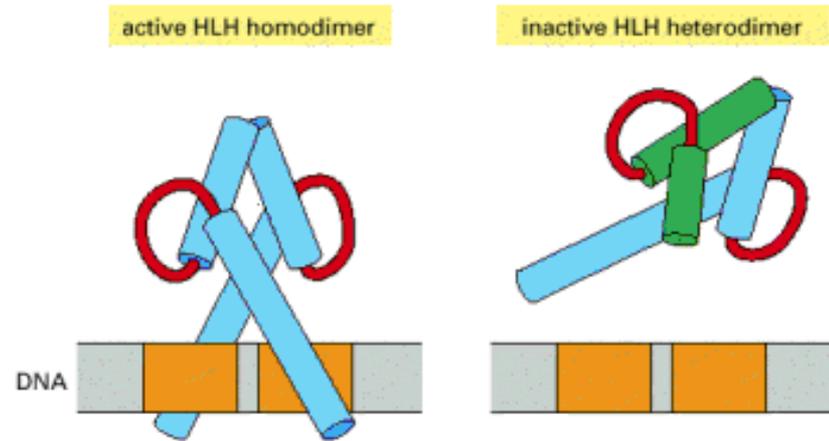
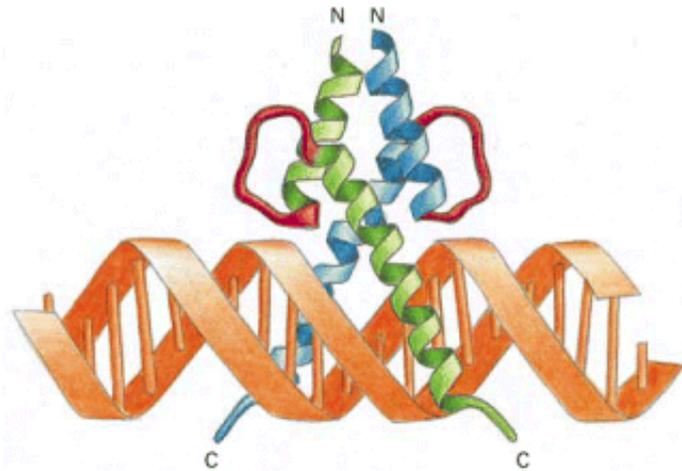
(Alberts, MCB, 2002, 4th ed. Figure 7-20, p. 388)

Heterodimer or different domains bound to its DNA recognition site



- A. The yellow helix 4 of the protein on the right is unstructured in the absence of the protein on the left, forming a helix only upon heterodimerization.
- B. Two DNA-binding domains covalently joined by a flexible polypeptide (POU-domain, both a homeodomain and a helix-turn-helix structure) (Alberts, MCB, 2002, 4th ed. Figure 7-23 and 7-24 , p. 391)

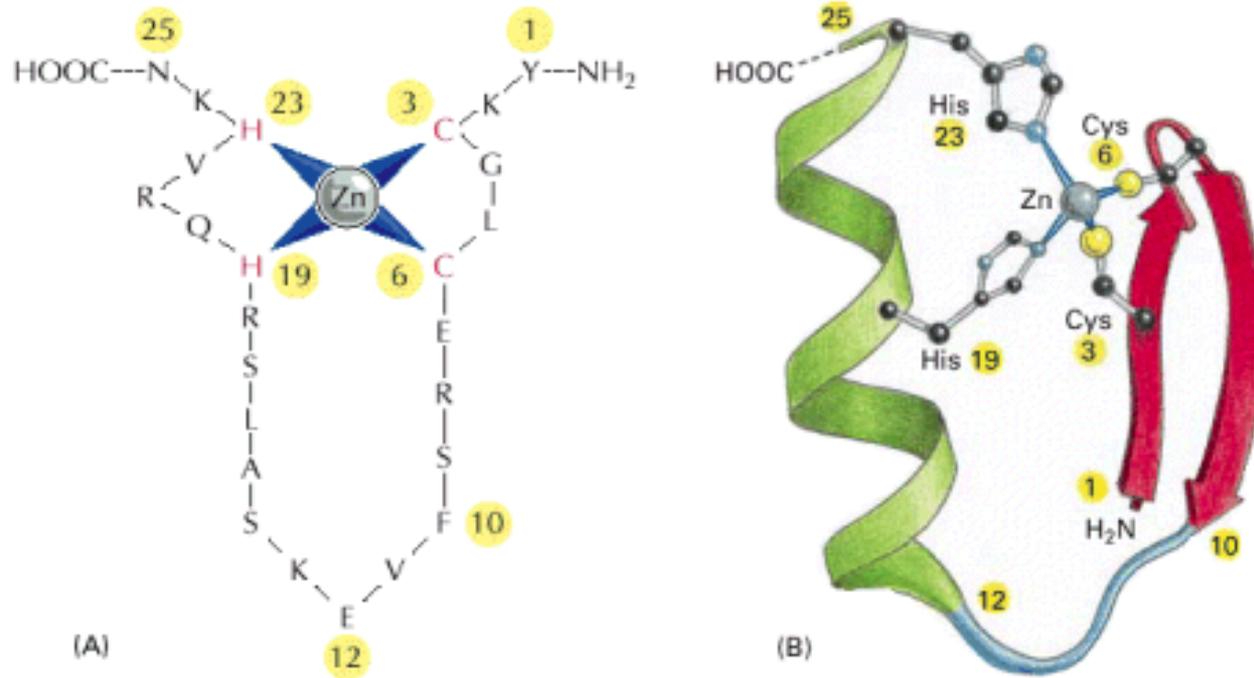
A helix-loop-helix dimer bound to DNA



- The two monomers are held together in a four-helix bundle: each monomer contributes two α helices connected by a flexible loop of protein.
- Inhibitory regulation by truncated HLH proteins. The binding of a full-length HLH protein to a truncated HLH protein that lacks the DNA-binding α helix generates a heterodimer that is unable to bind DNA tightly.

(Alberts, MCB, 2002, 4th ed. Figure 7-25 , p. 390)

C2H2 zinc finger protein

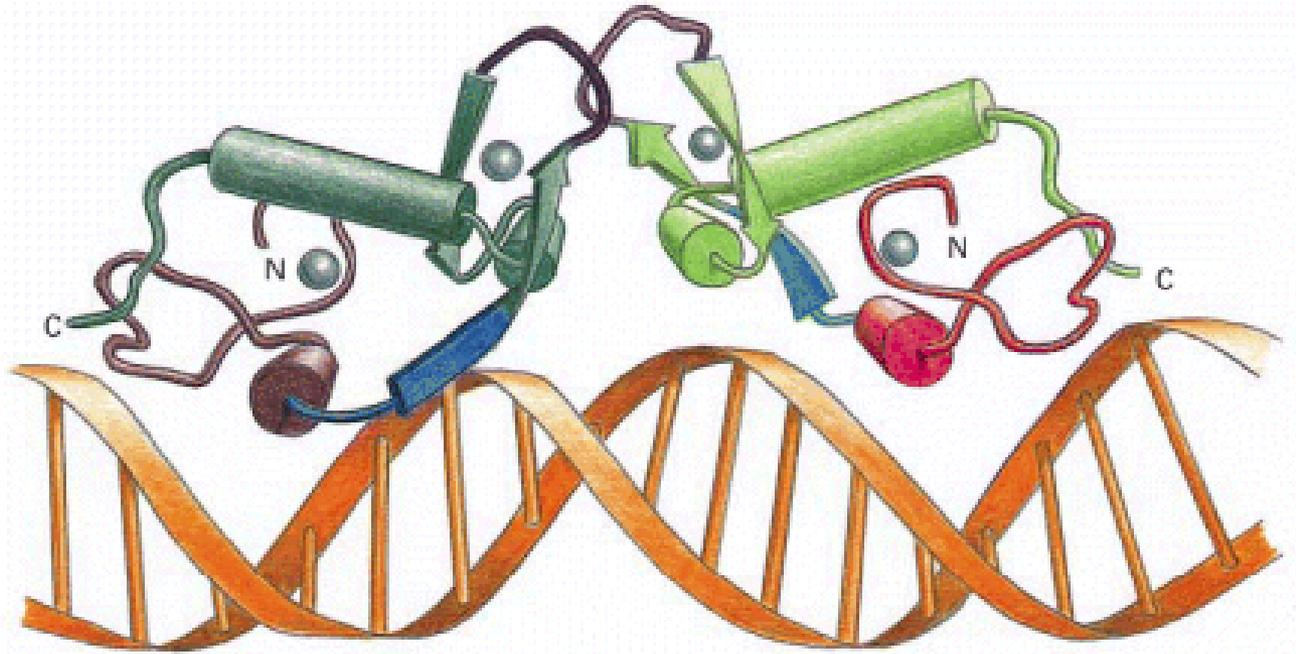


(A) Schematic drawing of the amino acid sequence of a zinc finger from a frog protein of this class.

(B) The three-dimensional structure of this type of zinc finger is constructed from an antiparallel sheet (amino acids 1 to 10) followed by an helix (amino acids 12 to 24). The four amino acids that bind the zinc (Cys 3, Cys 6, His 19, and His 23) hold one end of the helix firmly to one end of the sheet.

(Alberts et al. MBC, 2002, 4th ed. Figure 7-17, P. 386)

Zinc finger dimer and DNA complex



- Each zinc finger domain contains two atoms of Zn, one stabilizes the DNA recognition helix, and one stabilizes a loop involved in dimer formation. Each Zn atom is coordinated by four appropriately spaced cysteine residues.
- The two recognition helices of the dimer are held apart by a distance corresponding to one turn of the DNA double helix. (Glucocorticoid receptor, 肾上腺皮质激素受体)
(Alberts, MCB, 2002, 4th ed. Figure 7-19 , p. 387)

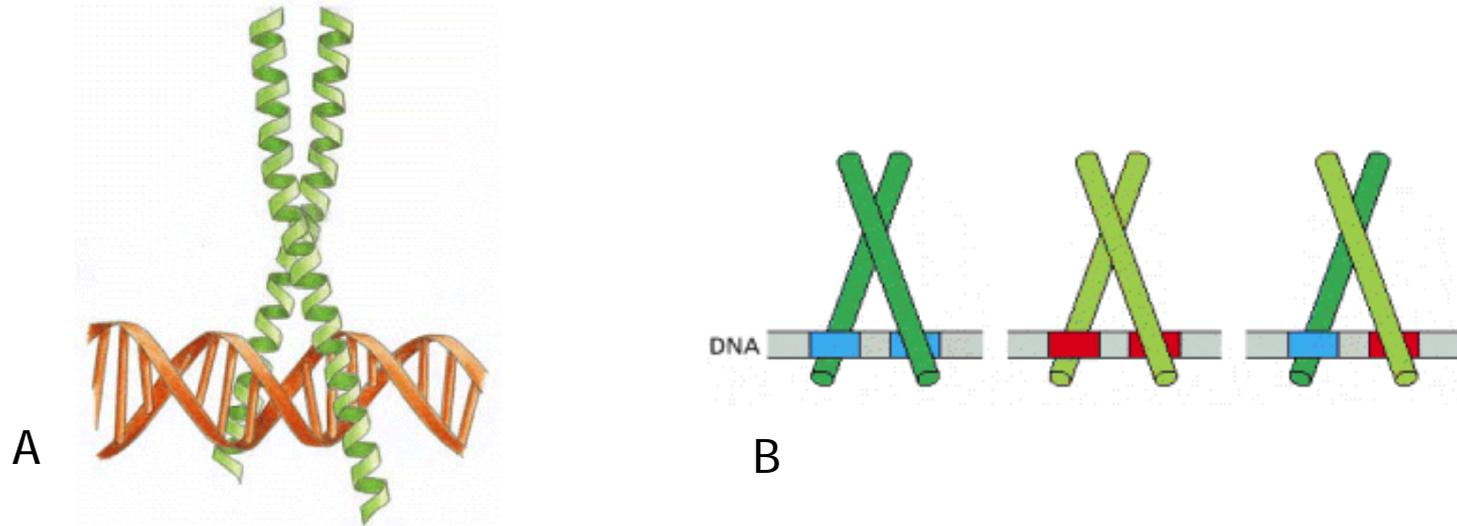
Leucine Zipper With DNA complex



Gcn4 Basic Domain, Leucine Zipper Complexed With Atf/Creb Site DNA (PDB id 2DGC)

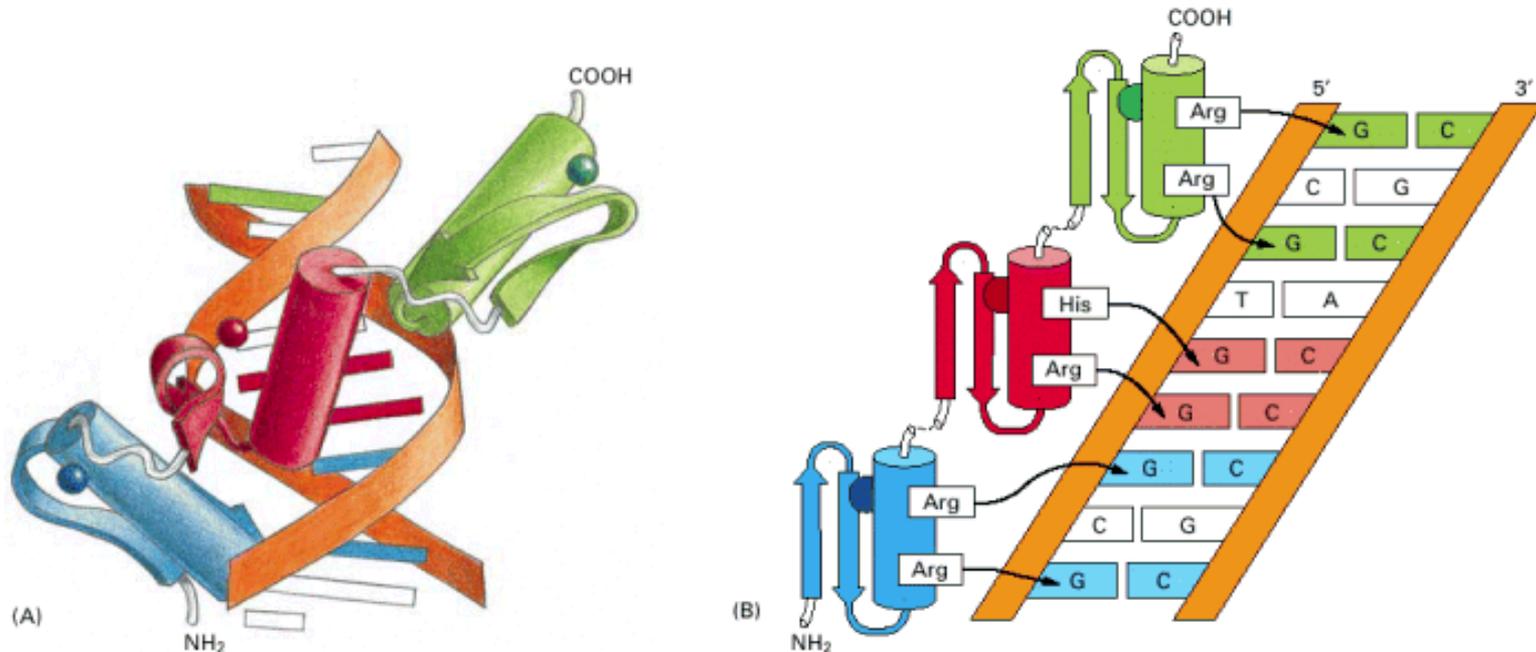
(Keller W, Konig P, Richmond TJ. Crystal structure of a bZIP/DNA complex at 2.2 Å: determinants of DNA specific recognition. *J Mol Biol.* 1995 Dec 8;254(4):657-67.)

Leucine zipper dimer bound to DNA



- A. Two α -helical DNA-binding domains dimerize through their α -helical leucine zipper region to form an inverted Y-shaped structure. Each arm of the Y is formed by a single α helix, one from each monomer, that mediates binding to a specific DNA sequence in the major groove of DNA. Each α helix binds to one-half of a symmetric DNA structure.
- B. Heterodimerization of leucine zipper proteins can alter their DNA-binding specificity. The two different monomers can combine to form a heterodimer, which now recognizes a hybrid DNA sequence
(Alberts, MCB, 2002, 4th ed. Figure 7-21 , p. 388)

DNA binding by a zinc finger protein

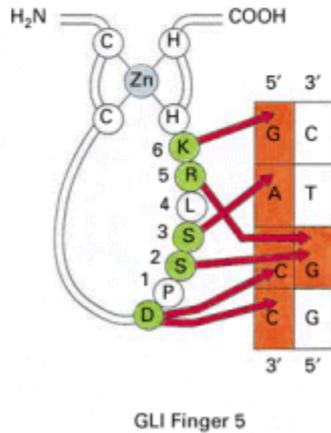
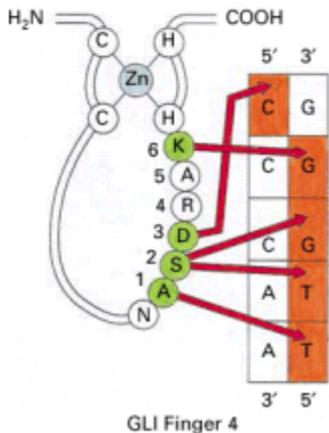
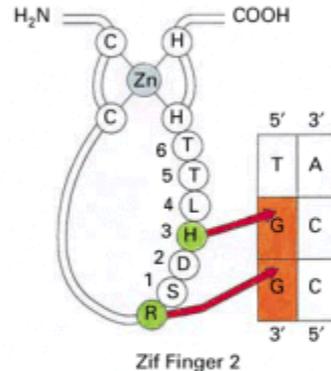
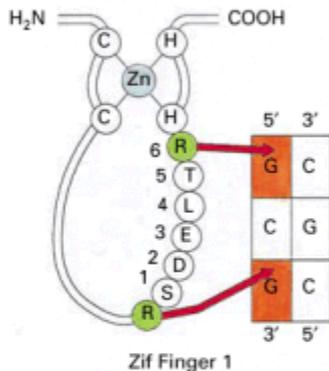
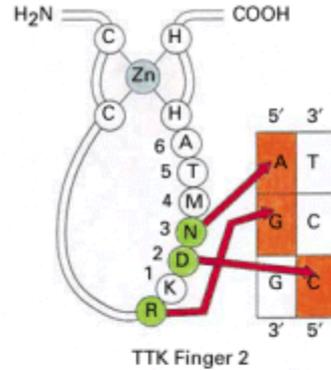
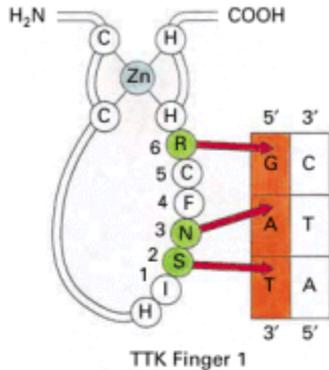


(A) The structure of a fragment of a mouse gene regulatory protein bound to a specific DNA site. This protein recognizes DNA using three zinc fingers of the Cys-Cys-His-His type arranged as direct repeats.

(B) The three fingers have similar amino acid sequences and contact the DNA in similar ways. In both (A) and (B) the zinc atom in each finger is represented by a small sphere.

(Albert et al. MBC, 2002, 4th ed. Figure 7-18.)

Sequence-specific zinc fingers



TTK: *Drosophila*, development
 Zif: Mouse, Zif 268
 GLI: Human, cancer related

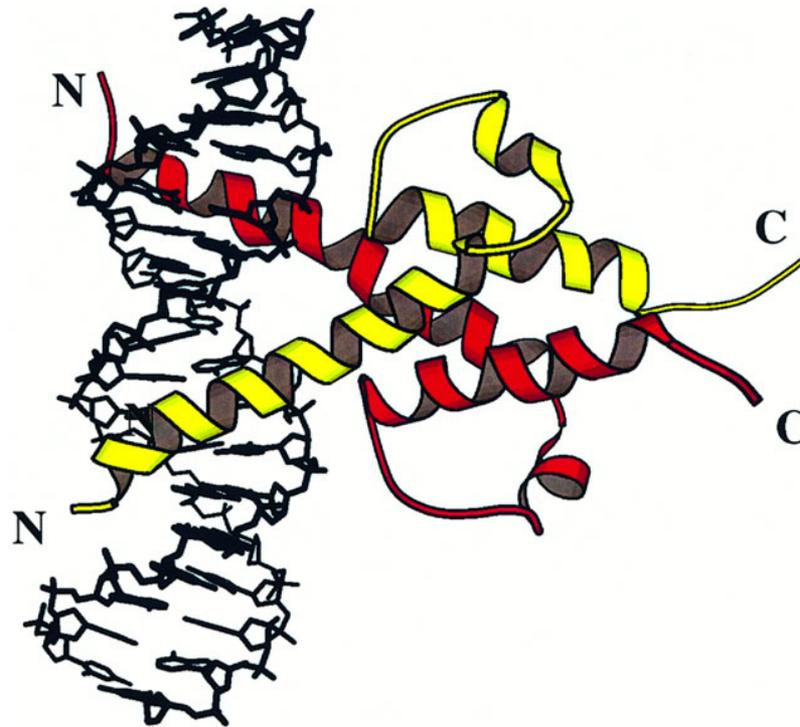
- Arginine-guanine contacts are common, guanine can also be recognized by serine, histidine, and lysine.
- The same amino acid (serine, in this example) can recognize more than one base.

(Alberts et al. MBC, 2002, 4th ed. Figure 7-28, P.392.)

PDB hits of in DATF

Family	Fold	Cis-element	Code	No	Source
AP2	ZF-like	AGCCGCC	1GCC	3	Arab idopsis
ARID	HTH	TATTGAT	1KQQ	5	Drosophila
bHLH	bHLH	CACGTG	1A0A	6	Yeast
bZIP	bZIP	ACGT	2DGC	15	Yeast
C2H2	ZF		1G2D	35	Mouse
C3H	ZF		2DAN*		
FHA	ZF		1D5V	21	Human
HB	HD-Zip	CAAT(A/T)ATTG	1LFU*	26	Mouse
HMG	HTH		1K99	9	Human
HSF	HTH	TTCnnGAA	1HTS	3	Drosophila
LIM	C2HCC4		1QLI	5	Quail
MADS	HLH	GC(A/T)6GG	1EGW	6	Human
MYB	HTH	GTAACGGTCTAC	1AJ5	18	Chicken
NAC	alpha+beta		1UT4	2	Arab idopsis
PcG	SET		1MVH	14	Yeast
PHD	C4HC3		1MM2	4	Human
SBP	C3HC4	TNCGTACAA	1UL4	2	Arab idopsis
TAZ	C3H		1RU8	2	Mouse

Helix-loop-helix With DNA complex

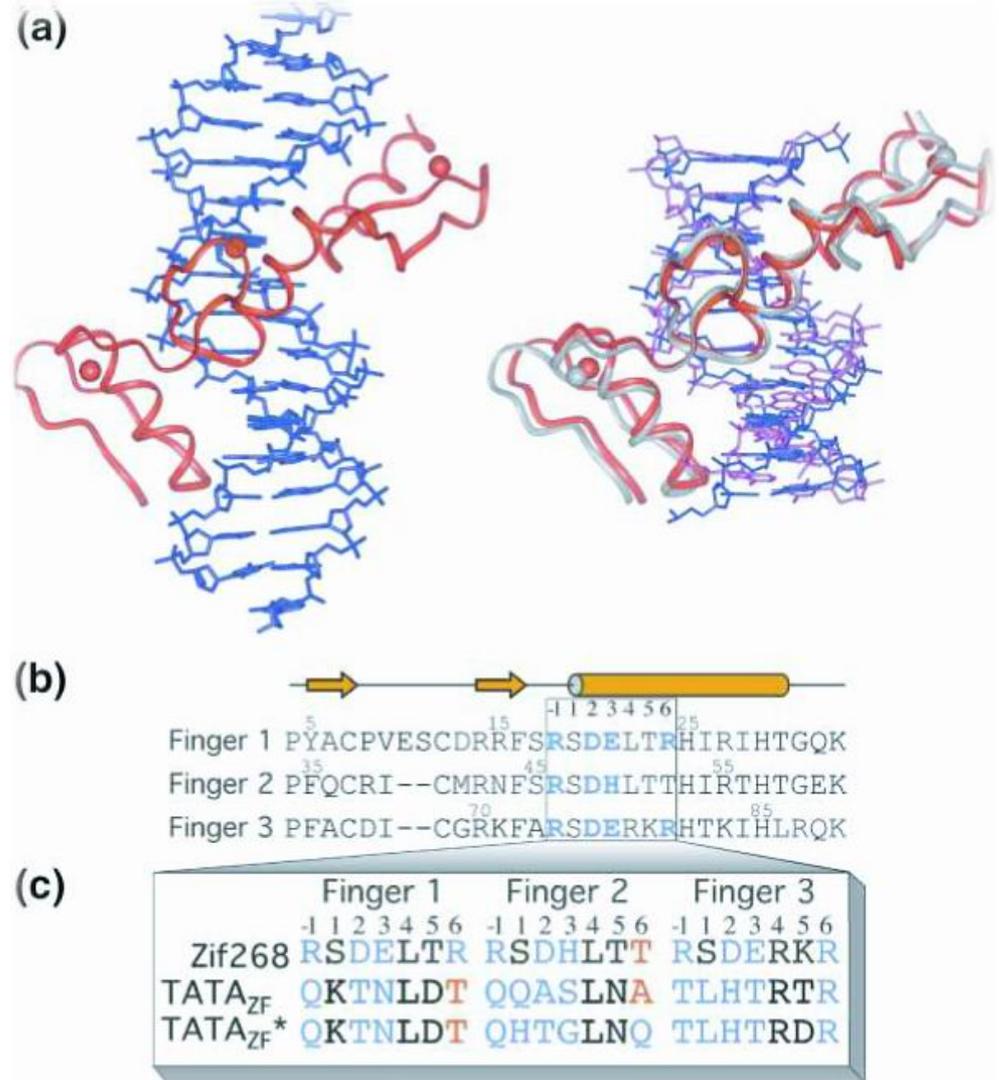


The structure of PHO4 protein bHLH domain-DNA complex. Molecules A and B are colored red and yellow, respectively. Helical structure is clearly seen in the loop region.

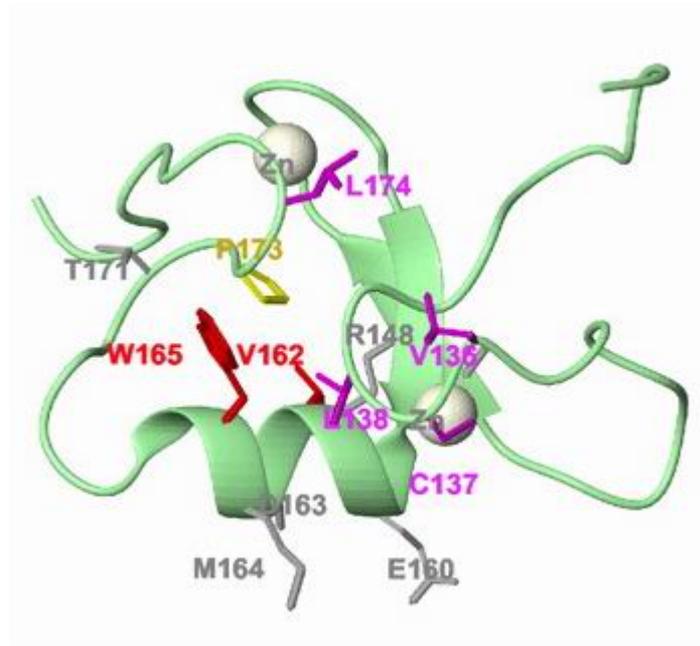
(Shimizu T, Toumoto A, Ihara K, Shimizu M, Kyogoku Y, Ogawa N, Oshima Y, Hakoshima T. Crystal structure of PHO4 bHLH domain-DNA complex: flanking base recognition. EMBO J. 1997 Aug 1;16(15):4689-97.)

C2H2 zinc finger

- a. Left: structure of TATAZF bound to DNA. Right: superposition of the TATAZF and Zif268 structures
- b. Aligned sequence of the three Cys2His2 zinc fingers from Zif268
- c. Sequences of the recognition helices from Zif268, TATAZF and TATAZF
(Wolfe SA, Grant RA, Elrod-Erickson M, Pabo CO. Beyond the "recognition code": structures of two Cys2His2 zinc finger/TATA box complexes. Structure (Camb). 2001 Aug;9(8):717-23.)

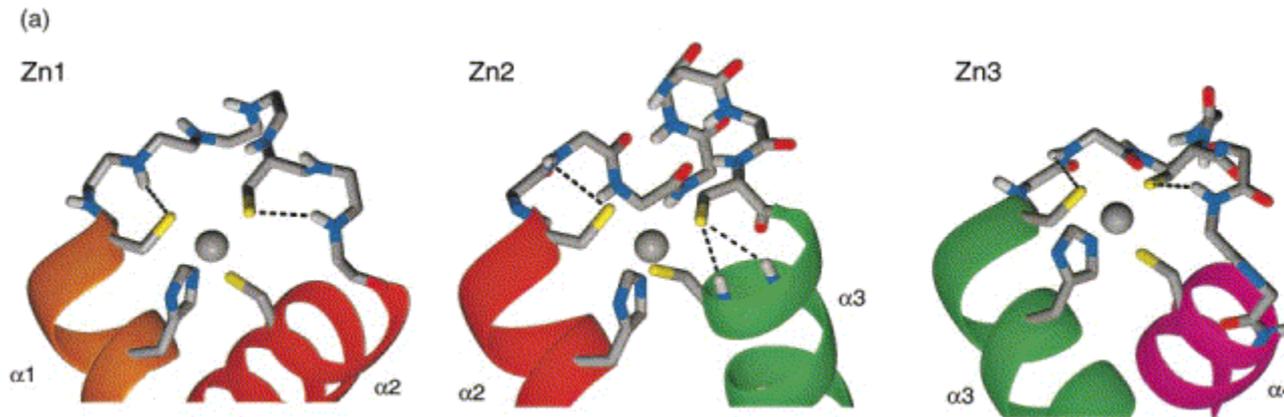


C3H zinc finger



EL5 RING finger domain. Residues that exhibited a high sensitive amide chemical shift in NMR titration experiments are highlighted. Residues that are well and poorly conserved between E2 binding RING finger domains (Group 1, 2, and 3) are shown in magenta and gray, respectively. Pro173, which cannot be detected in a ^{15}N HSQC spectrum, is shown in yellow. (Wolfe SA, Grant RA, Elrod-Erickson M, Pabo CO. Beyond the "recognition code": structures of two Cys2His2 zinc finger/TATA box complexes. *Structure* (Camb). 2001 Aug;9(8):717-23.)

Zinc-binding regions of the TAZ2 domain



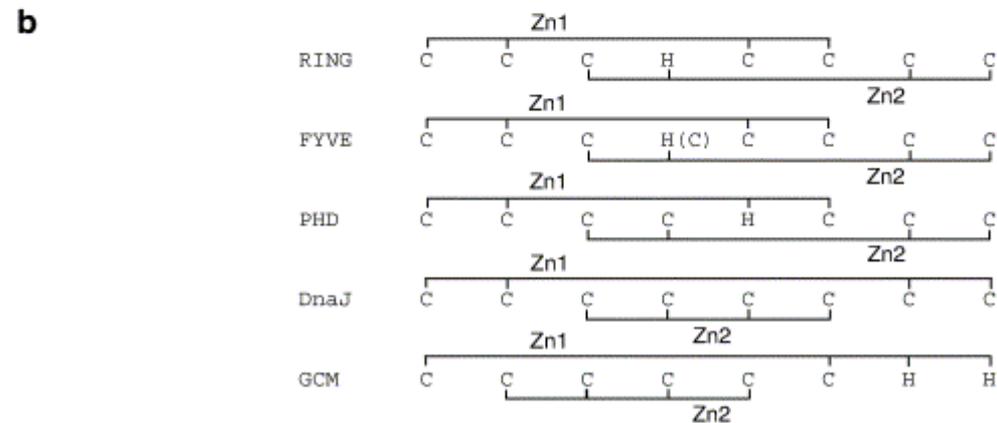
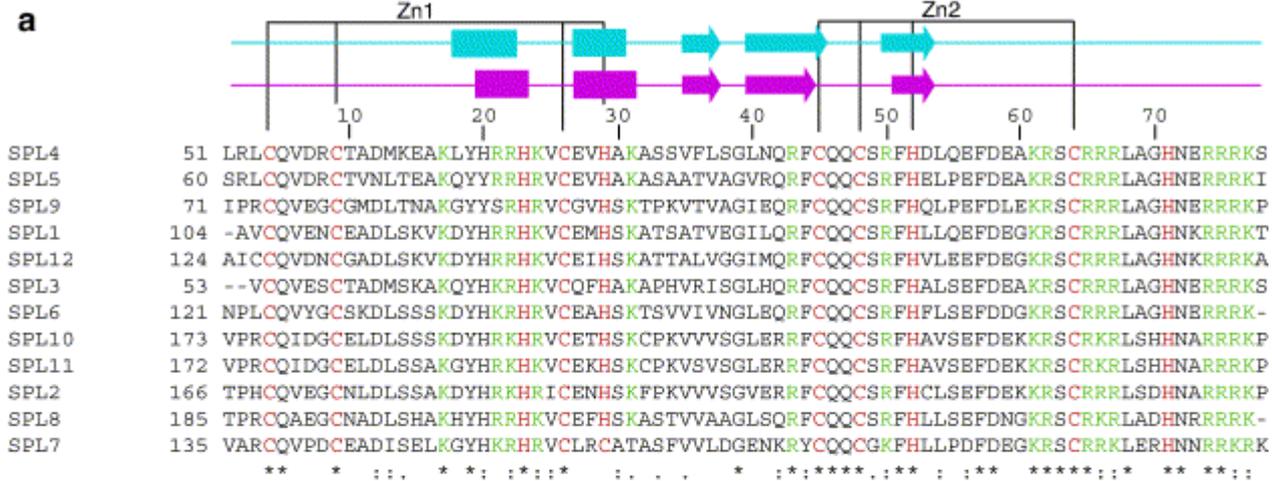
Each of the three HCCC-type zinc-binding domains of TAZ2 is formed from two helices and a loop. The structures of the Zn1, Zn2, and Zn3 binding sites are shown from the lowest-energy structure of TAZ2, together with the NH-S hydrogen bonds (broken lines) that stabilize the structures. Inhibitory regulation by truncated HLH proteins.

(De Guzman RN, Liu HY, Martinez-Yamout M, Dyson HJ, Wright PE.

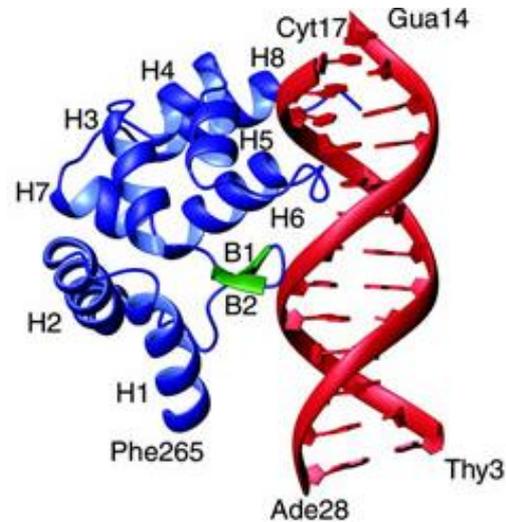
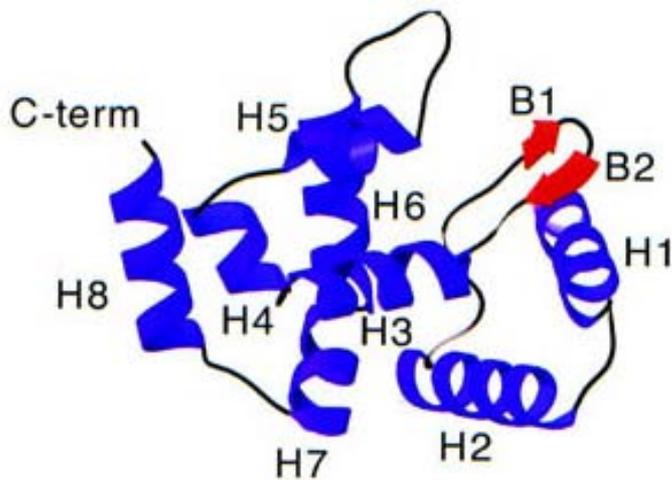
Solution structure of the TAZ2 (CH3) domain of the transcriptional adaptor protein CBP.

J Mol Biol. 2000 Oct 20;303(2):243-53.)

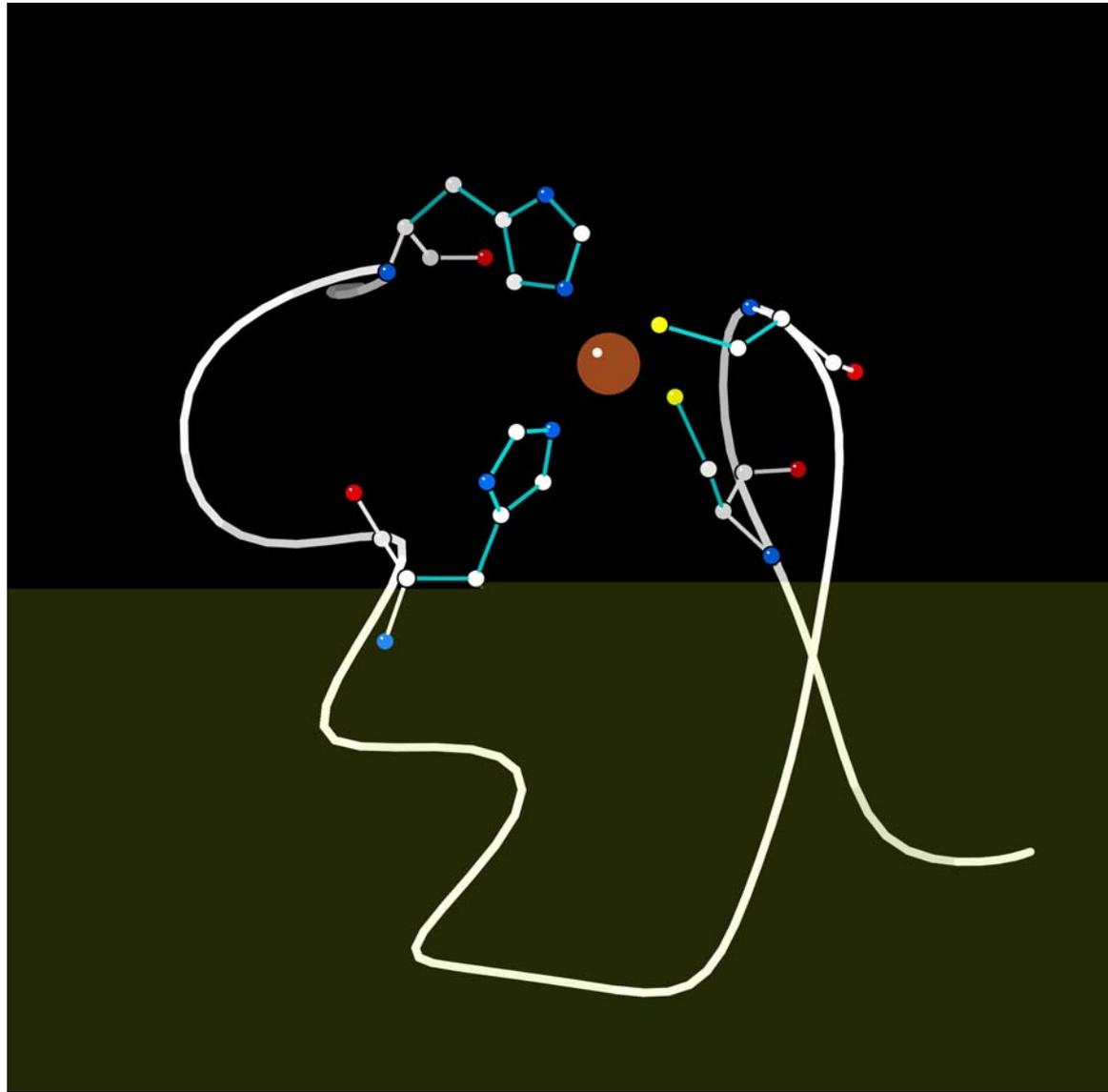
Sequence and Zinc finger in SBP TFs

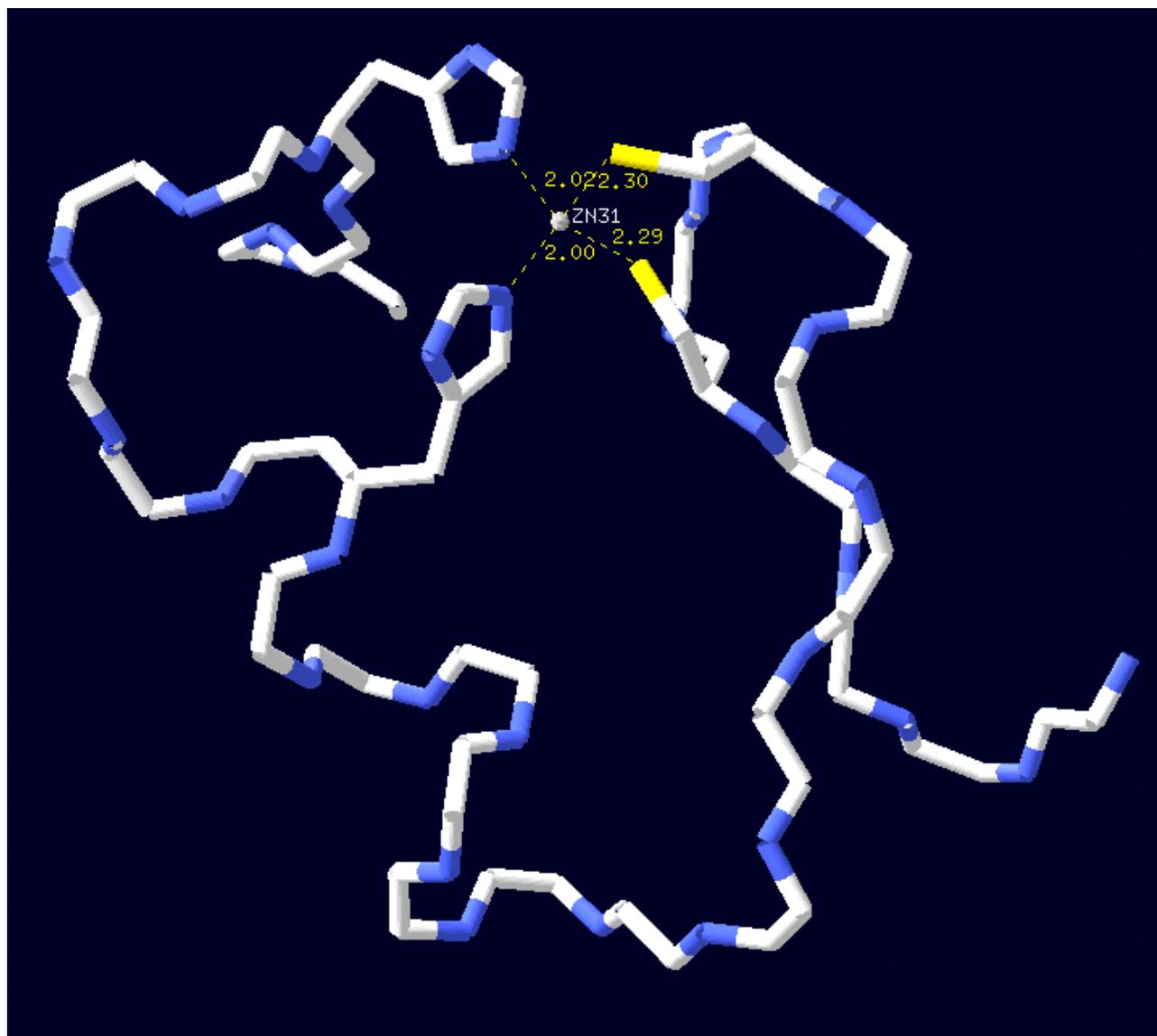


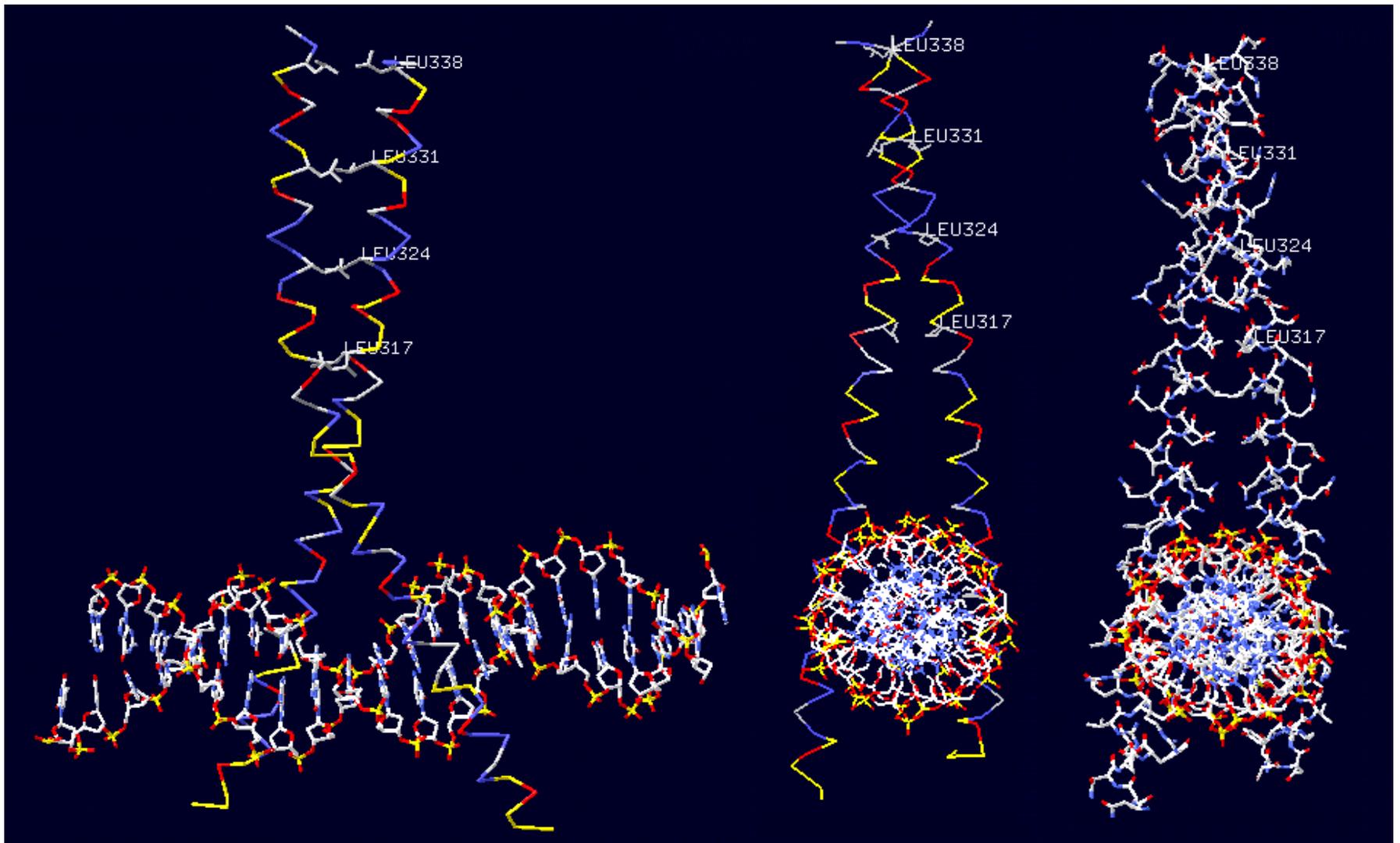
Model of Dead ringer-DNA complex (ARID)



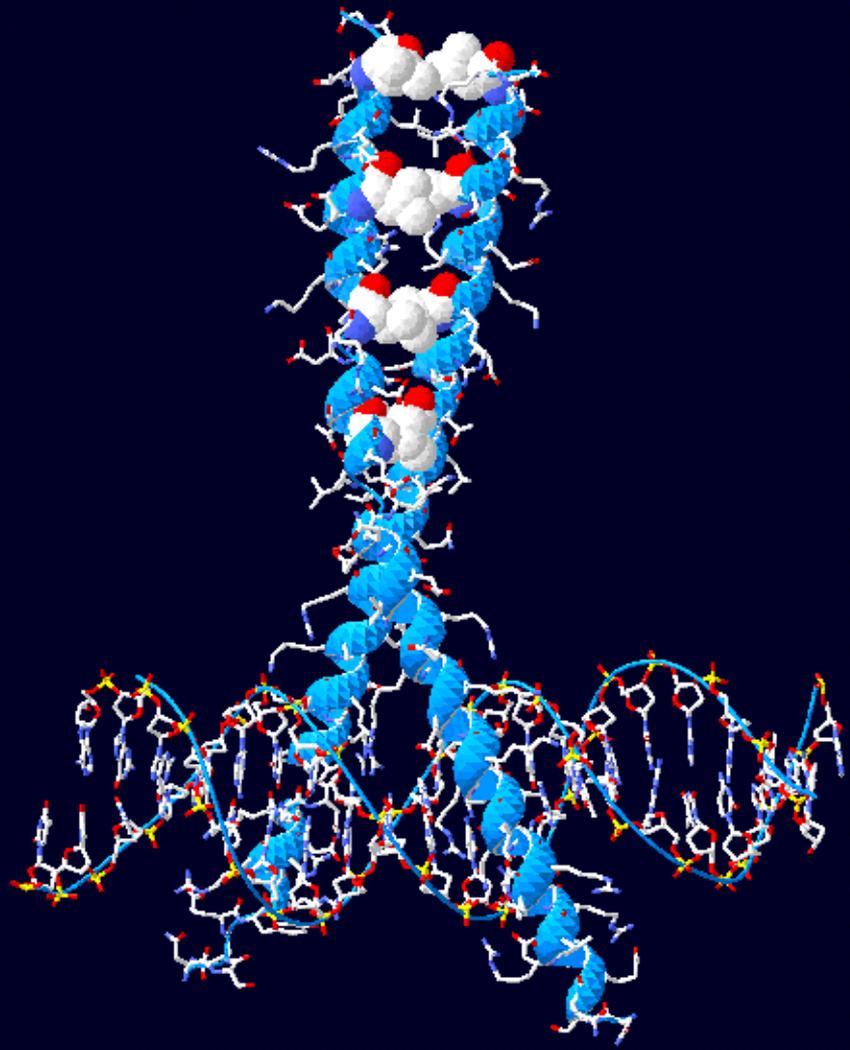
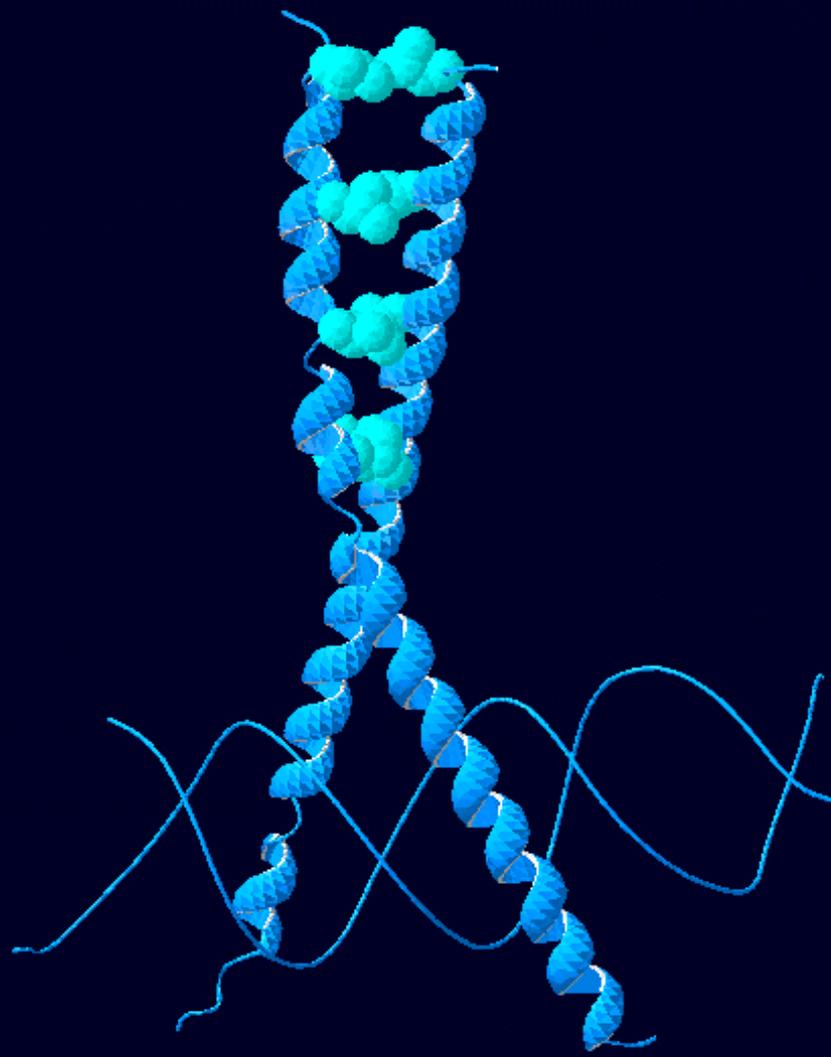
1. Iwahara J, Clubb RT. Solution structure of the DNA binding domain from Dead ringer, a sequence-specific AT-rich interaction domain (ARID). EMBO J. 1999 Nov 1;18(21):6084-94.
2. Iwahara J, Iwahara M, Daughdrill GW, Ford J, Clubb RT. The structure of the Dead ringer-DNA complex reveals how AT-rich interaction domains (ARIDs) recognize DNA. EMBO J. 2002 Mar 1;21(5):1197-209.

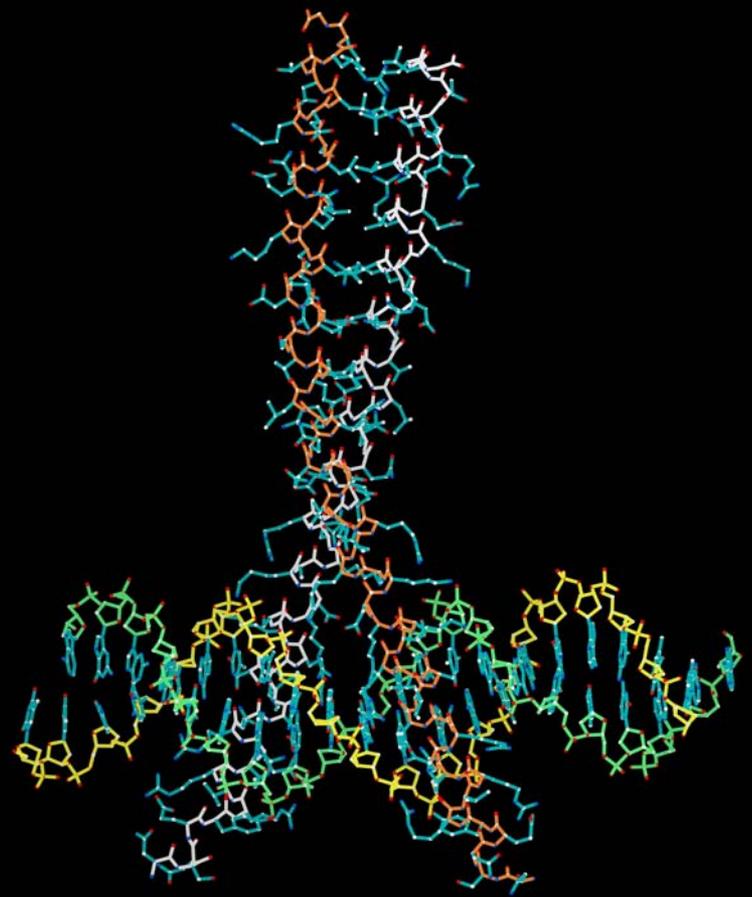
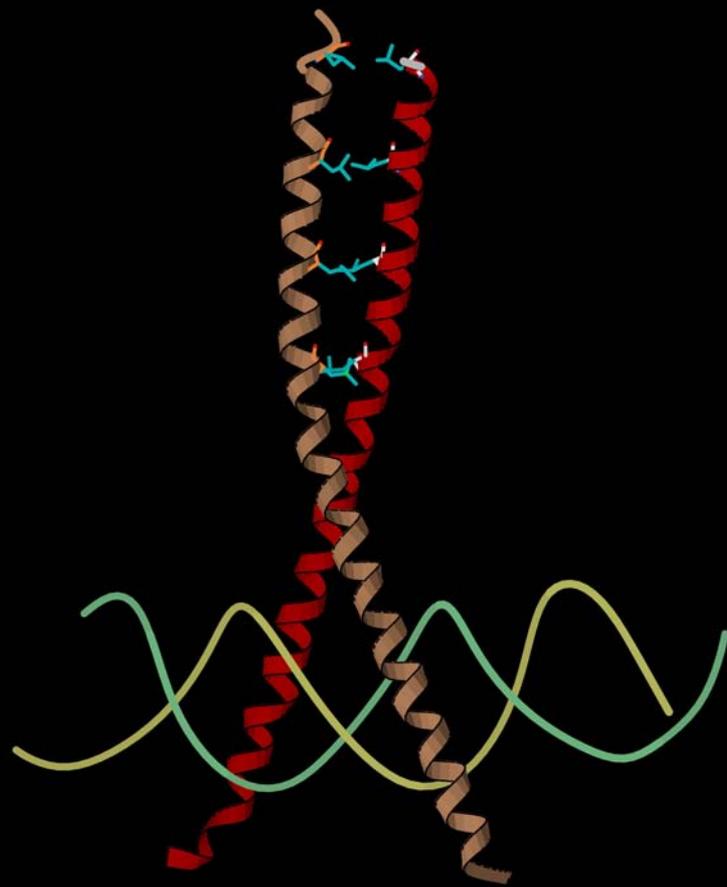






Leucine zipper complexed with DNA (1NWO)





Leucine zipper complexed with DNA (1NWO)

References on 3D structure of TFs

- Search strategy
 - “transcription factor” [TITLE] AND “structure” [TITLE]
 - “helix-loop-helix” [TITLE] AND “structure” [TITLE]
 - “zinc finger” [TITLE] AND “structure” [TITLE]
 - “leucine zipper” [TITLE] AND “structure” [TITLE]
- Manually removed hits mainly refer to gene structure of transcription factors
- Resulted in 182 papers, 14 reviews
- Represents a part of the papers in this field, not inclusive
- Color code: green – plant TFs, blue – modeling study