



## **Specific Transcription Factors Belong to a Few Families**

## **Specific Transcription Factors Belong to a Few Families**

- Many procaryotic transcription factors are dimers that recognize palindromic DNA sequences in such a way that each subunit binds one half-site.
- The POU transcription factors exhibit a variation of this theme by having two similar domains in one polypeptide chain.
- A similar arrangement is found in another important class of specific eucaryotic transcription factors – those that have zinc atoms in their DNA-binding motifs.
- The polypeptide chains of these proteins may contain several homologous domains, each capable of specific DNA binding and each containing zinc as an integral part of the DNA-binding domain.

## **Specific Transcription Factors Belong to a Few Families**

- These proteins belong to several different groups with different structures and different modes of DNA binding.
- They differ by ways in which zinc-containing motifs recognize specific binding sites arranged contiguously along DNA segments.
- One unifying concept to emerge from the study of different DNA-binding motifs is that these motifs provide 3-D scaffold that match the contours of DNA.
- These scaffolds ensure proper positioning of the interacting protein surfaces against the DNA, allowing both sequence-specific and nonspecific interactions to occur.

# Specific Transcription Factors Belong to a Few Families

- In some cases it is the amino acid sequences of the scaffolds, rather than the sequences of amino acids involved in DNA interactions, that make it possible to identify a DNA-binding motif in a protein sequence.
- This principle applies in particular to two important families of eucaryotic transcription factors that contain leucine zippers
  - the basic region/leucine zipper (b/zip) family
  - the basic region/helix-loop-helix/leucine zipper (b/HLH/zip) family.

# Zinc Finger Transcription Factors

- Classic zinc finger transcription factors  
(Cys-Cys-His-His family).
- Zinc fingers in the glucocorticoid receptors  
(Cys-Cys-Cys-Cys family).
- C<sub>6</sub>-zinc cluster transcription factors  
(Cys-Cys-Cys-Cys-Cys-Cys family).
- Retroviral zinc fingers  
(Cys-Cys-His-Cys family).

## The classic zinc finger motif

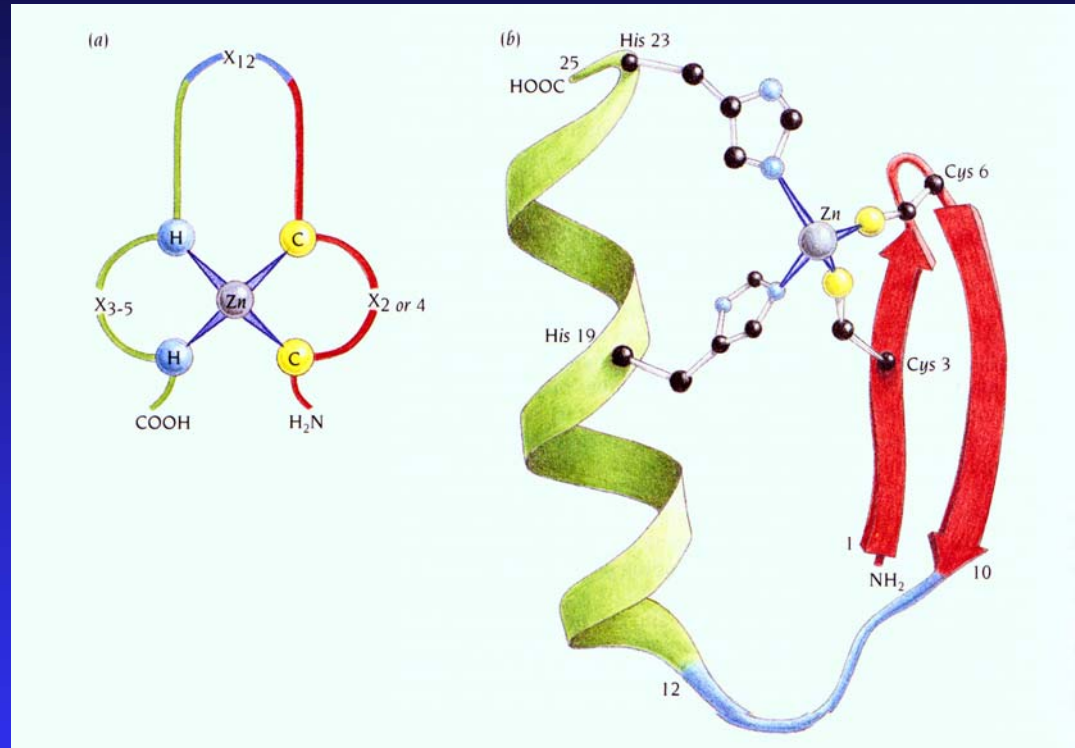
- The classic zinc finger motif was first identified in the transcription factor **TFIIIA** from *Xenopus laevis*.
- TFIIIA regulates **transcription of ribosomal 5S RNA** in frog oocytes.
- The amino acid sequence of the **344-residue polypeptide chain** of TFIIIA contains **9 repeated sequences** of about **30 residues each**.
- The repeats are not identical in sequence but each contains **two Cys residues at the amino end** and **two His residues at the carboxy end**.
- Since the protein contains intrinsic zinc atoms and its transcriptional activity is dependent on the presence of these zinc atoms, **it has been suggested that the Cys and His are ligands to a zinc atom** and that the loop between these residues forms the **DNA-binding region**.

## The classic zinc finger motif

- The classic zinc finger transcription factors contain several and sometimes over 30 concatenated zinc finger motifs.
- These motifs bind to a sequence of binding sites in the target DNA.
- Although each motif only recognizes three or four bases via a loop region at the tip of each finger, a zinc finger transcription factor containing multiple fingers, each with its own short recognition site, becomes a highly discriminating DNA-binding protein.

# The classic zinc fingers bind to DNA in tandem along the major groove

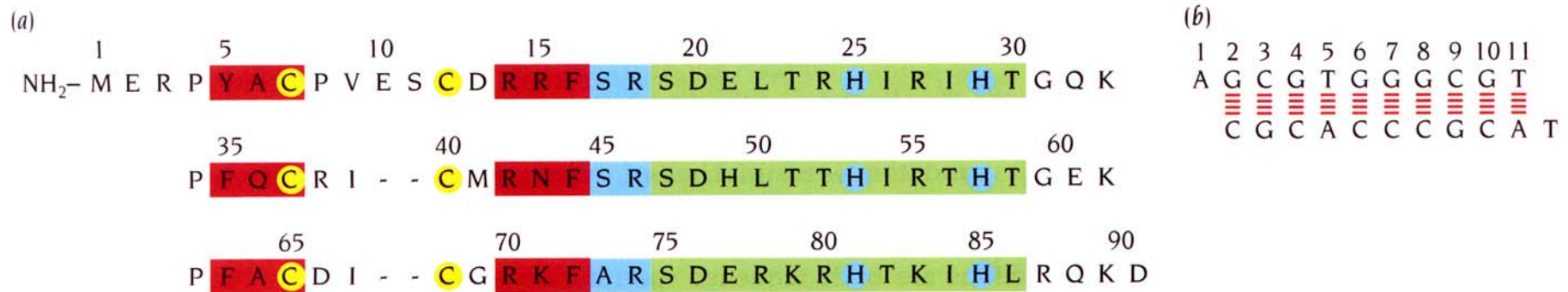
(a) The classic zinc finger comprises about 30 amino acid residues, with two cysteine and two histidine residues which bind to a zinc atom. The linker region between the last cysteine and the first histidine is 12 residues long and is called the finger region.



(b). The 3-D structure of a chemically synthesized 25-residue peptide with an amino acid sequence corresponding to one of the zinc fingers in an embryonic protein, Xfin, from *Xenopus laevis*. The structure is built up from an antiparallel  $\beta$  hairpin motif followed by an  $\alpha$  helix. The four zinc ligands Cys 3, Cys 6, His 19, and His 23 anchor one end of the helix to one end of the  $\beta$  strand. Models, quite similar to the observed structure, were predicted from amino acid sequences of members of this zinc finger family.



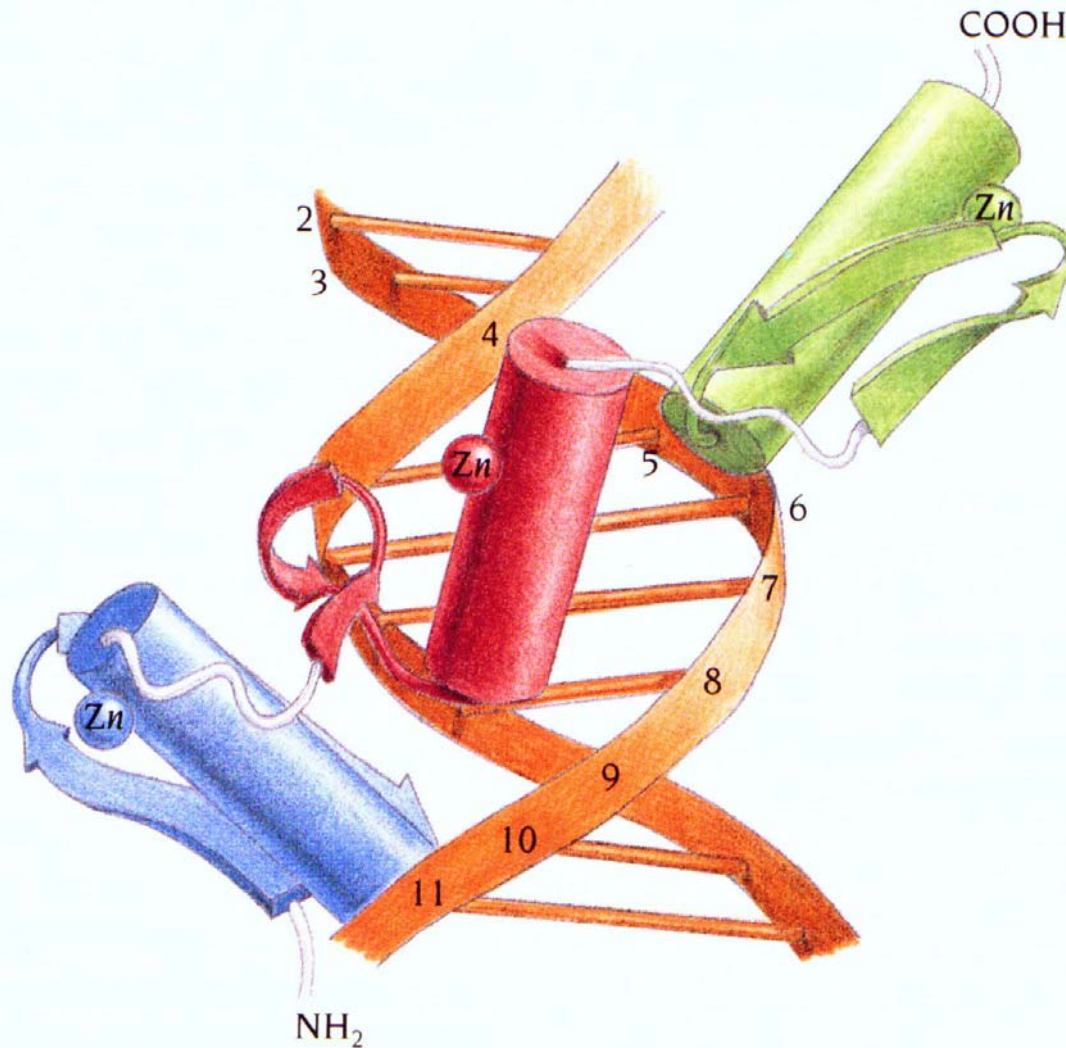
## The classic zinc fingers bind to DNA in tandem along the major groove



(a) Amino acid sequence of a fragment of the Zif 268 protein that contains three zinc fingers. Residues forming the  $\beta$  strands and  $\alpha$  helices are red and green, respectively, and those involved in the turn between the last  $\beta$  strand and the  $\alpha$  helix are blue.

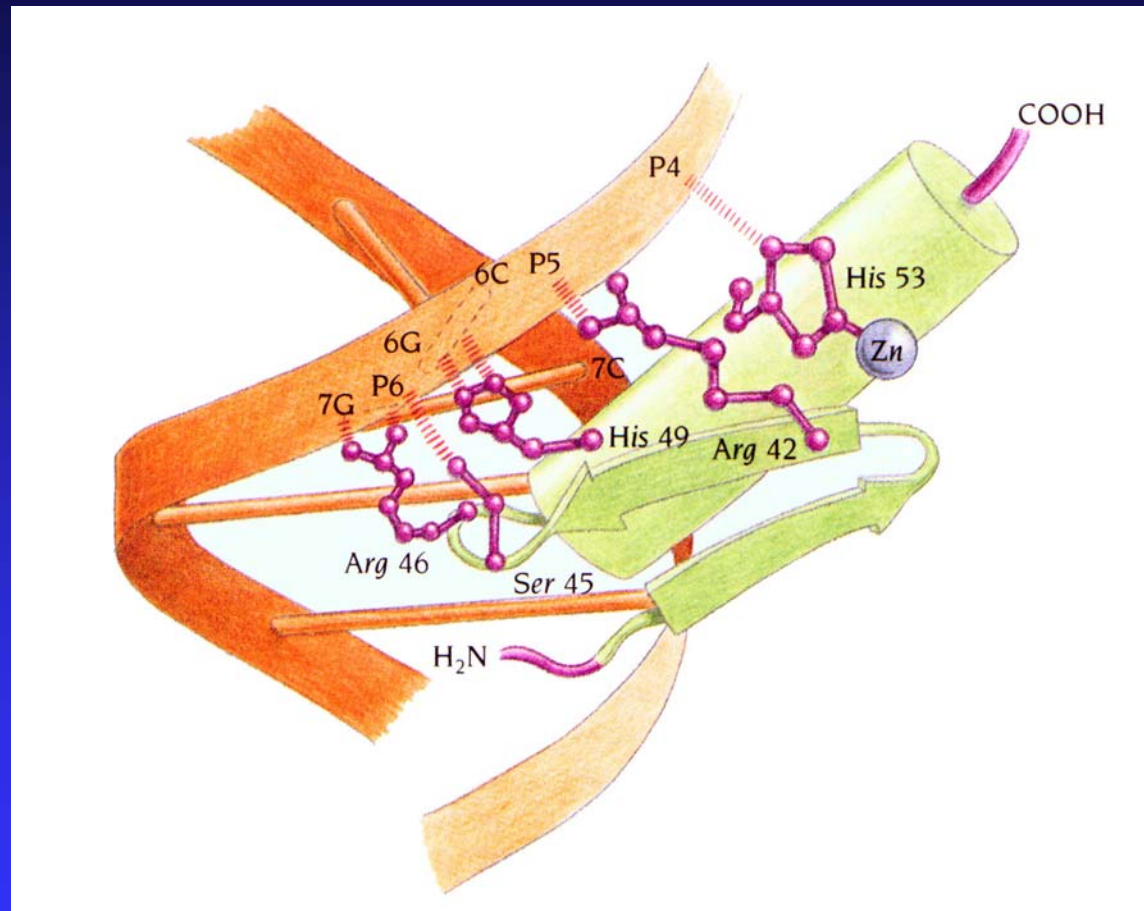
(b) The nucleotide sequence of the DNA fragment that was used in the x-ray structure determination of the Zif 268 fragment complexed with DNA.

## The classic zinc fingers bind to DNA in tandem along the major groove



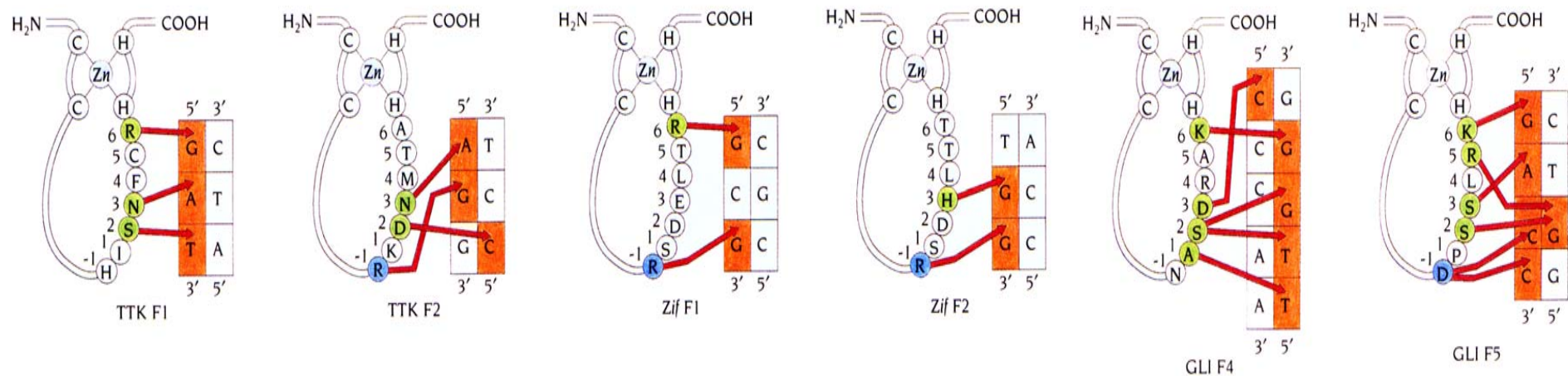
Schematic diagram shows the structure of three zinc fingers of Zif 268 bound to DNA. The three zinc fingers, which bind in tandem to the major groove of DNA, are colored blue, red and green from the N-terminus. The zinc fingers have the same structure and **bind** in a similar way **with the N-terminus of the  $\alpha$  helix pointing into the major groove.**

## The finger region of the classic zinc finger motif interacts with DNA



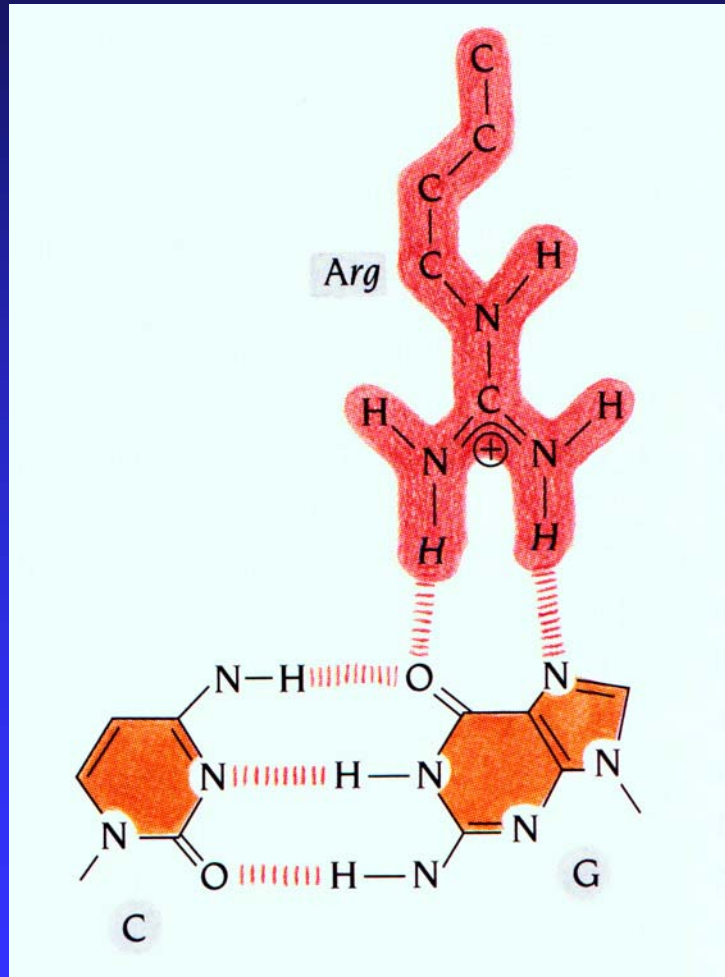
Detailed view of the binding of the second zinc finger of Zif 268 to DNA. Two side chains, **Arg 46** and **His 49**, form **sequence-specific interactions with DNA**. There are also **three nonspecific interactions** between phosphate groups of the DNA and the side chains of **Arg 42**, **Ser 45**, and **His 53**.

# Comparison of the sequence-specific binding to DNA of six different zinc fingers



Residues in the N-terminus of the  $\alpha$  helix in the finger regions are numbered 1 to 6. The residue immediately preceding the  $\alpha$  helix is numbered -1. In spite of the structural similarities between the zinc fingers and their overall mode of binding, there is **no simple rule that governs which bases the fingers contact**. It is therefore **not possible at present to predict from the amino acid sequence of a zinc finger the DNA sequence to which it will bind, or vice versa**.

## One sequence-specific interaction occurs more frequently than others in protein-DNA complexes



Two hydrogen bonds form between an arginine side chain of the protein and a guanine base of the DNA. The two strong bonds that can be formed between these two groups make an essential contribution to sequence-specific recognition, as in other families of DNA-binding proteins discussed previously.



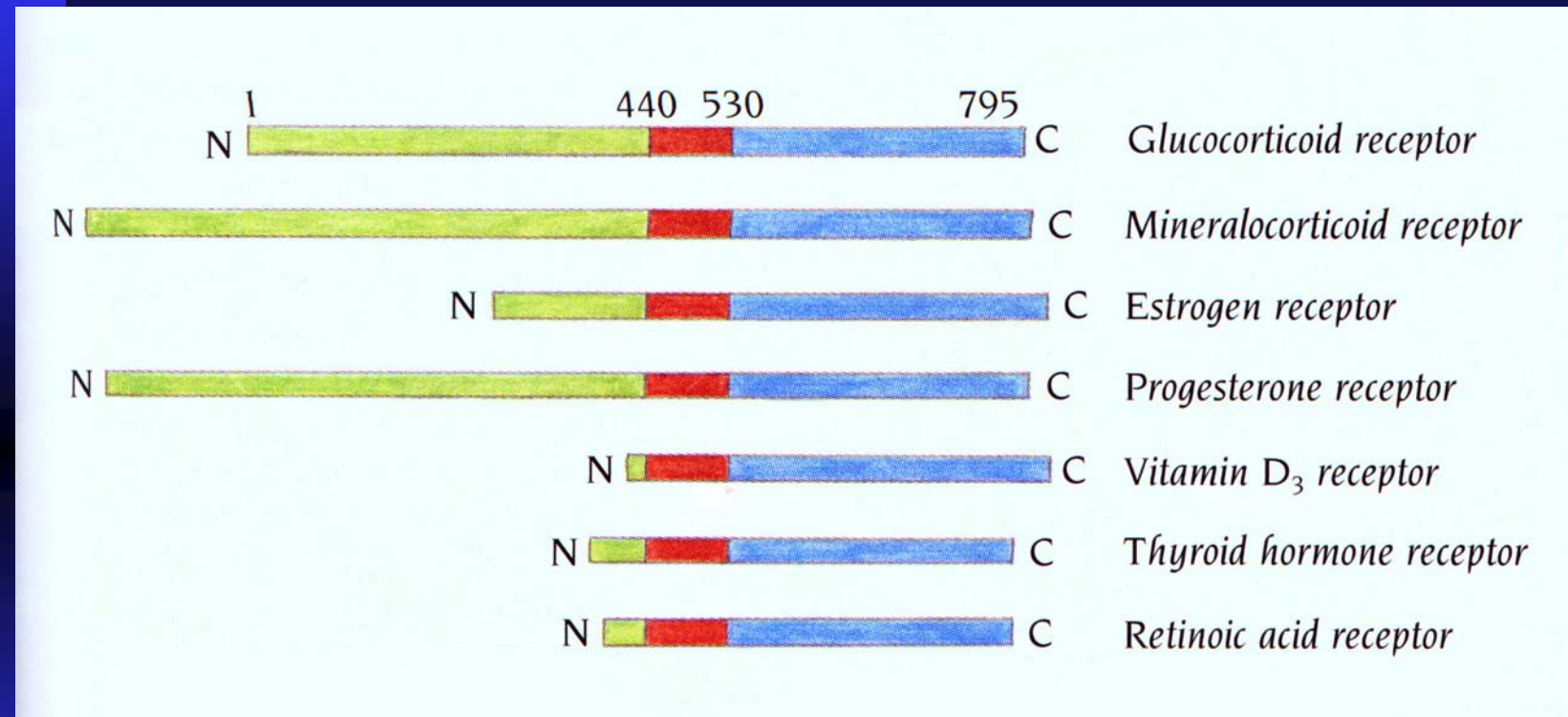
## Two zinc-containing motifs in the glucocorticoid receptor form one DNA-binding domain

- The nuclear receptors utilize pairs of zinc atoms to stabilize a DNA-binding domain.
- The steroid hormone receptors are homodimers with two DNA-binding sites that recognize an almost palindromic sequence in the DNA.
- The monomeric DNA-binding domains dimerize upon binding to DNA in such a way that the two recognition helices of the dimer can bind to target sequences of the response element with the appropriate spacer region.
- These dimers are symmetrical, like the dimers of the 434 repressor.

## Two zinc-containing motifs in the glucocorticoid receptor form one DNA-binding domain

- Other nuclear receptors form heterodimers that are nonsymmetric and bind to response elements in which the target sequences are organized as direct repeats.
- These heterodimers discriminate response elements on the basis of the different lengths of their spacer region.
- A DNA supported asymmetric dimerization interface located within the DNA-binding domains allows these heterodimers to distinguish between closely related response elements.

## Two zinc-containing motifs in the glucocorticoid receptor form one DNA-binding domain



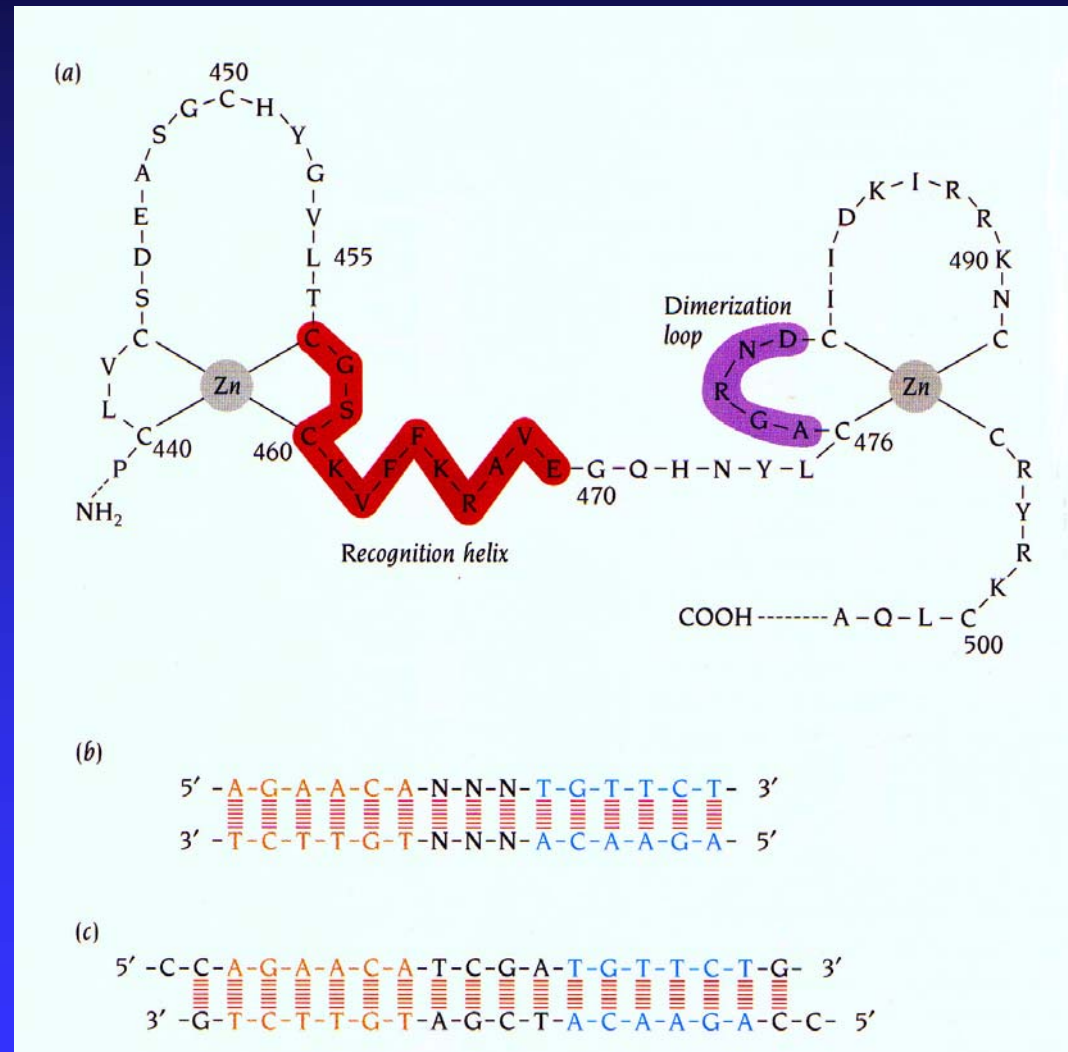
Evolutionary related members of the receptor family of transcription factors that have four cysteine residues bound to zinc in each of two regions of the sequence. The **DNA-binding domains (red)** have highly homologous amino acid sequences, whereas the **ligand-binding domains (blue)** are more variable. Residue numbers of the domain boundaries are given for the glucocorticoid receptor. **Exchange of individual domains between different members of the family suggests that they function as interchangeable modules.**



# Two zinc-containing motifs in the glucocorticoid receptor form one DNA-binding domain

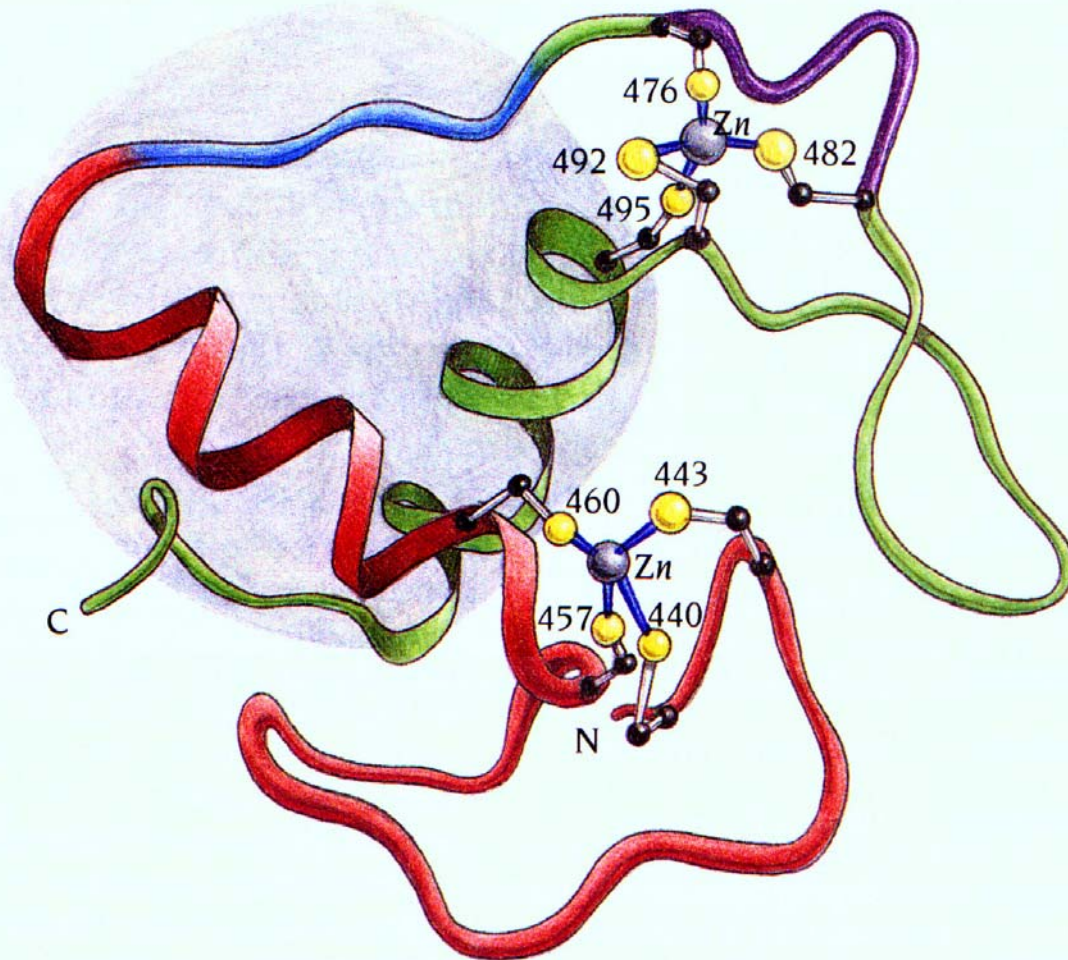
(a) Amino acid sequence of the zinc-containing DNA-binding domain of the glucocorticoid receptor. There are two zinc atoms within this domain, each bound to four cysteine residues. One of these stabilizes the recognition helix (red) that provides sequence-specific binding to DNA and the other zinc region contains a loop (purple) that is involved in formation of the dimeric molecule.

(b) Nucleotide sequence of the region of DNA that binds to the glucocorticoid receptor, the glucocorticoid receptor element, GRE. The two palindromic half-sites (blue and orange) are separated by a spacer region of three base pairs.



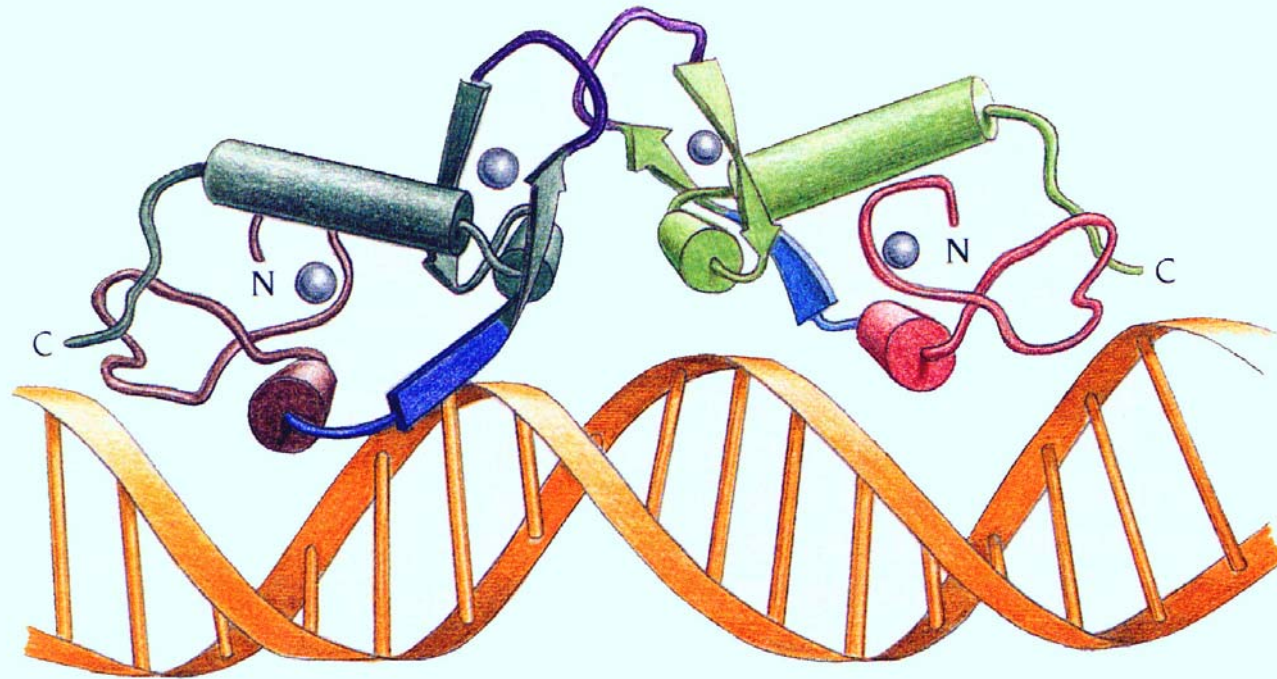
(c) The DNA fragment that was used in the crystal structure determination of the complex with the DNA-binding domain of the glucocorticoid receptor.

## Two zinc-containing motifs in the glucocorticoid receptor form one DNA-binding domain



Schematic diagram of the 3-D structure of the DNA-binding domain of the glucocorticoid receptor. The two zinc-binding regions, defined from the amino acid sequence are colored red and green, respectively, and the region that joins them is blue. Each has one  $\alpha$  helix and contains a zinc atom bound to four cysteine residues. The two  $\alpha$  helices are part of a compact globular core, shown in gray. The dimerization loop is shown in purple. The zinc atoms and the residues between the zinc ligands form protrusions from this globular core.

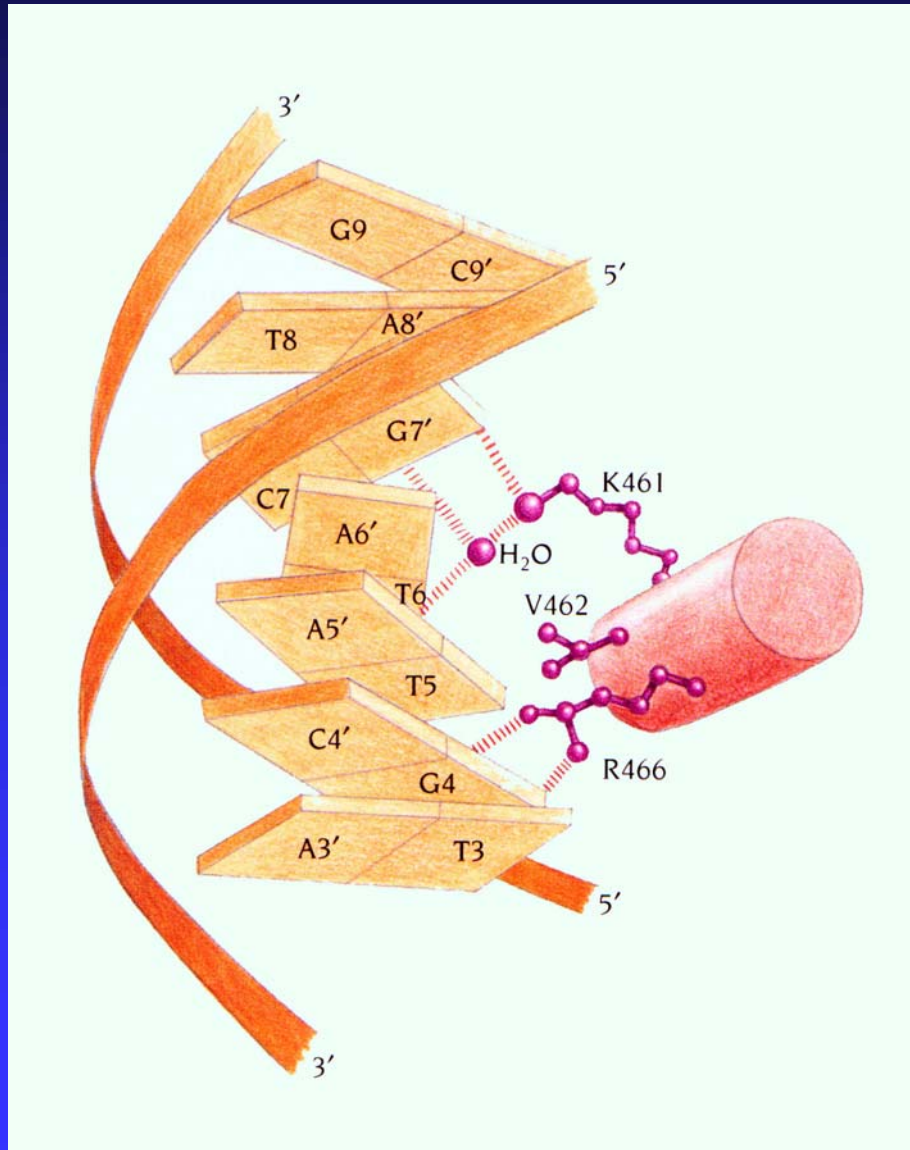
## Two zinc-containing motifs in the glucocorticoid receptor form one DNA-binding domain



Structure of the complex between the dimeric glucocorticoid receptor molecule and a DNA fragment. The two zinc-binding regions of each subunit have different colors; brown and dark green in one subunit and red and light green in the second. The linker region is blue. The recognition helices of the dimer (red and brown) are positioned in the major groove. The distance between them, which corresponds to one turn of the DNA helix, is fixed by the dimerization loop (purple). **This region undergoes a significant conformational change when the dimer binds to DNA.**



## Sequence-specific interactions between DNA and the recognition helix of the glucocorticoid receptor



Three residues, Lys 461, Val 462 and Arg 466 make specific contacts with the edges of the bases in the major groove.

## The retinoid X receptor forms heterodimers that recognize tandem repeats with variable length

- Generally, the family of nuclear receptors has low target affinity, high amino acid sequence homology and similar DNA response elements.
- How then are they targeted to their appropriate genes?
- The steroid hormone receptors bind exclusively as homodimers to response elements where the half-sites are organized in a palindromic orientation.
- Other nuclear receptors, such as those for **vitamin D (VDR)**, **thyroid hormone (TR)** or ***trans*-retinoic acid (RAR)**, form heterodimers with the ***cis*-retinoic acid receptor (RXR)** that bind to response elements with half-sites organized as direct repeats.
- Different heterodimers recognize response elements with different lengths of the spacer region.

## The retinoid X receptor forms heterodimers that recognize tandem repeats with variable length

-The individual domains of the two receptors both have structures similar to that of the glucocorticoid receptor, and they bind to DNA in a similar way, with their recognition helices in the major groove.

The dimer contacts are, however, totally different.

-In the RXR-TR heterodimer, which binds to direct DNA repeats like the classical zinc finger transcription factors, the domains bind to each other in a head-to-tail fashion: residues in one domain interact with totally different residues in the other domain.

-As a consequence, the dimer is polar, as is the response element, and in the complex the RXR domain binds exclusively to the 5' half-site of the response element even though the two direct repeats have the same nucleotide sequence.

## The retinoid X receptor forms heterodimers that recognize tandem repeats with variable length

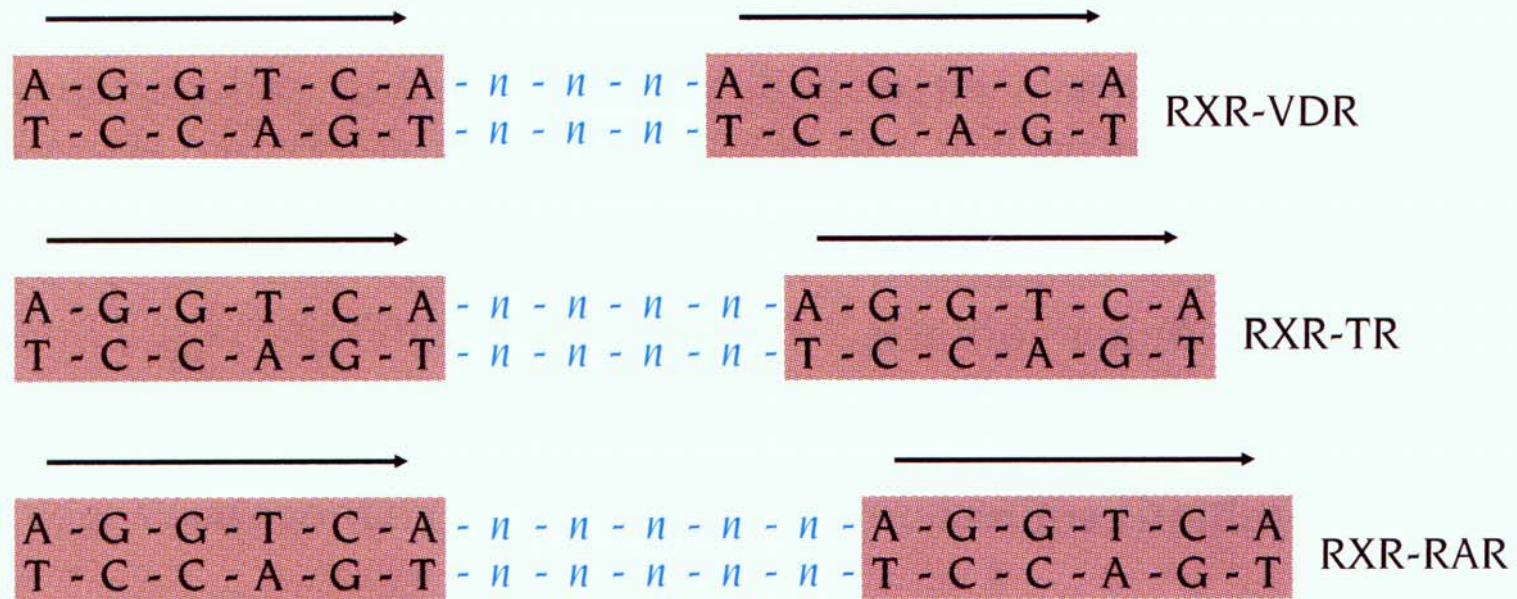
- The two domains are oriented in the dimer in such a way that each recognition helix binds to its half-site in the major groove of B-DNA with a spacer of four base pairs between the half-sites. The dimer interface recognition straddles the minor groove of the spacer region.
- Residues from the second zinc motif of RXR interact with residues in the first zinc motif of TR. These residues are also involved in nonsequence-specific interactions with the DNA backbone.

## The retinoid X receptor forms heterodimers that recognize tandem repeats with variable length

-Thus the DNA contacts support the dimer interface, which in turn supports both the specific and nonspecific contacts with DNA. These mutually supporting interactions between the protein domains and the DNA backbone reinforce a lesson learned from the glucocorticoid receptor-DNA complex concerning the importance of spacing between the half-sites: the DNA-binding domains of the nuclear receptors, which are monomers in the absence of DNA and which generally bind weakly as monomers, to a single half-site, bind with high affinity and strong cooperativity to response elements in which the appropriate half-sites are spaced correctly.



## The retinoid X receptor forms heterodimers that recognize tandem repeats with variable length



Response elements for heterodimers of the nuclear receptor for *cis*-retinoic acid (RXR) with the receptors for vitamin D (VDR), thyroid hormone (TR) and *trans*-retinoic acid (RAR). The half-sites of these response elements have identical nucleotide sequences and are organized as direct repeats. They differ in the number of base pairs in the spacer region between the half-sites. This difference forms the basis for the ability of the heterodimers to discriminate between the different response elements.

## The retinoid X receptor forms heterodimers that recognize tandem repeats with variable length

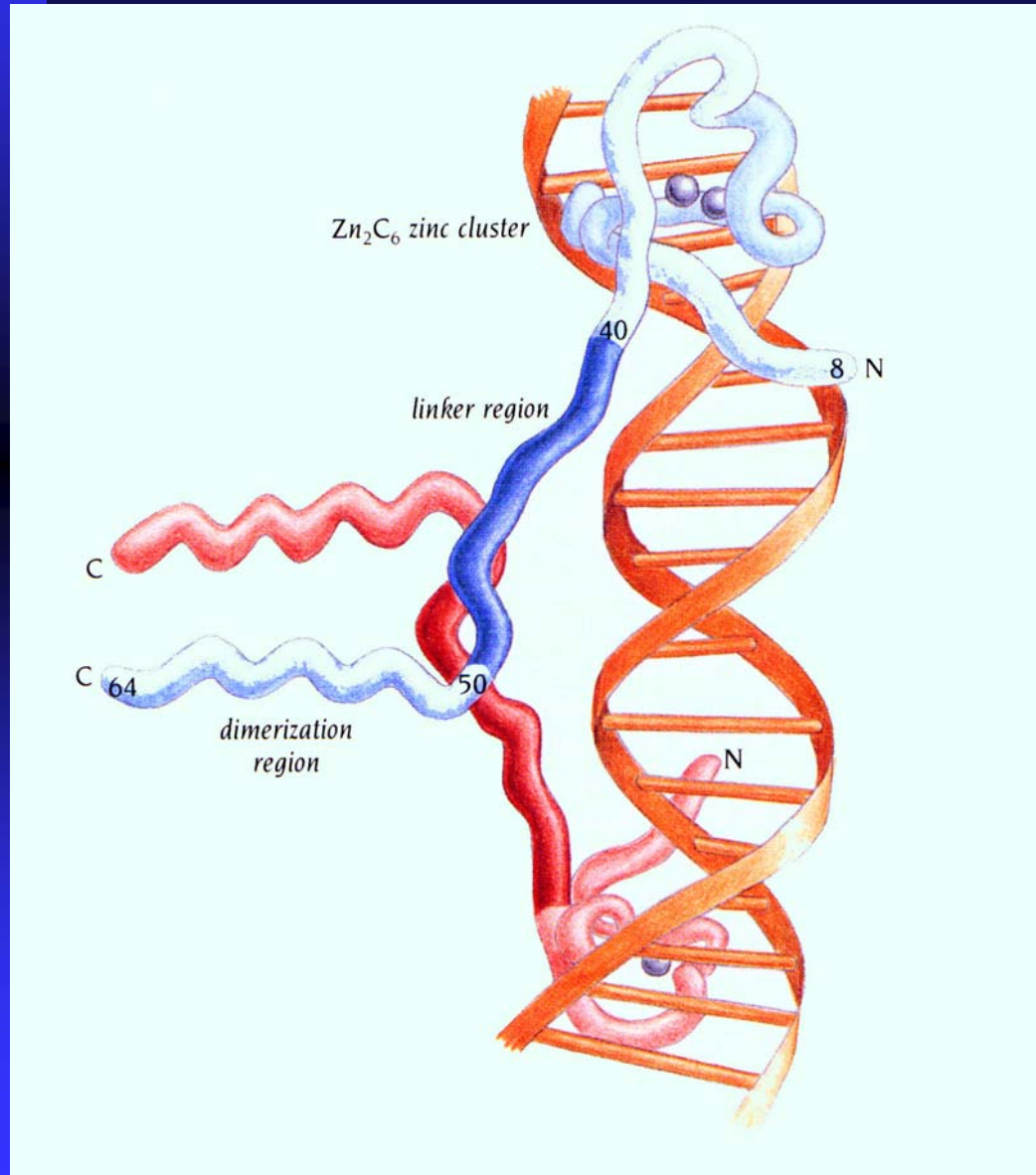
-In summary, a DNA supported asymmetric interface located within the DNA-binding domains of these nuclear receptors provides the molecular basis for receptor heterodimers to distinguish between closely related response elements.

-RXR can provide a repertoire of different dimerization surfaces, each one unique for a specific partner, allowing dimers to form that are adapted to the length of the spacer region in their corresponding response elements.

## Yeast transcription factor GAL4 contains a binuclear zinc cluster in its DNA-binding domain

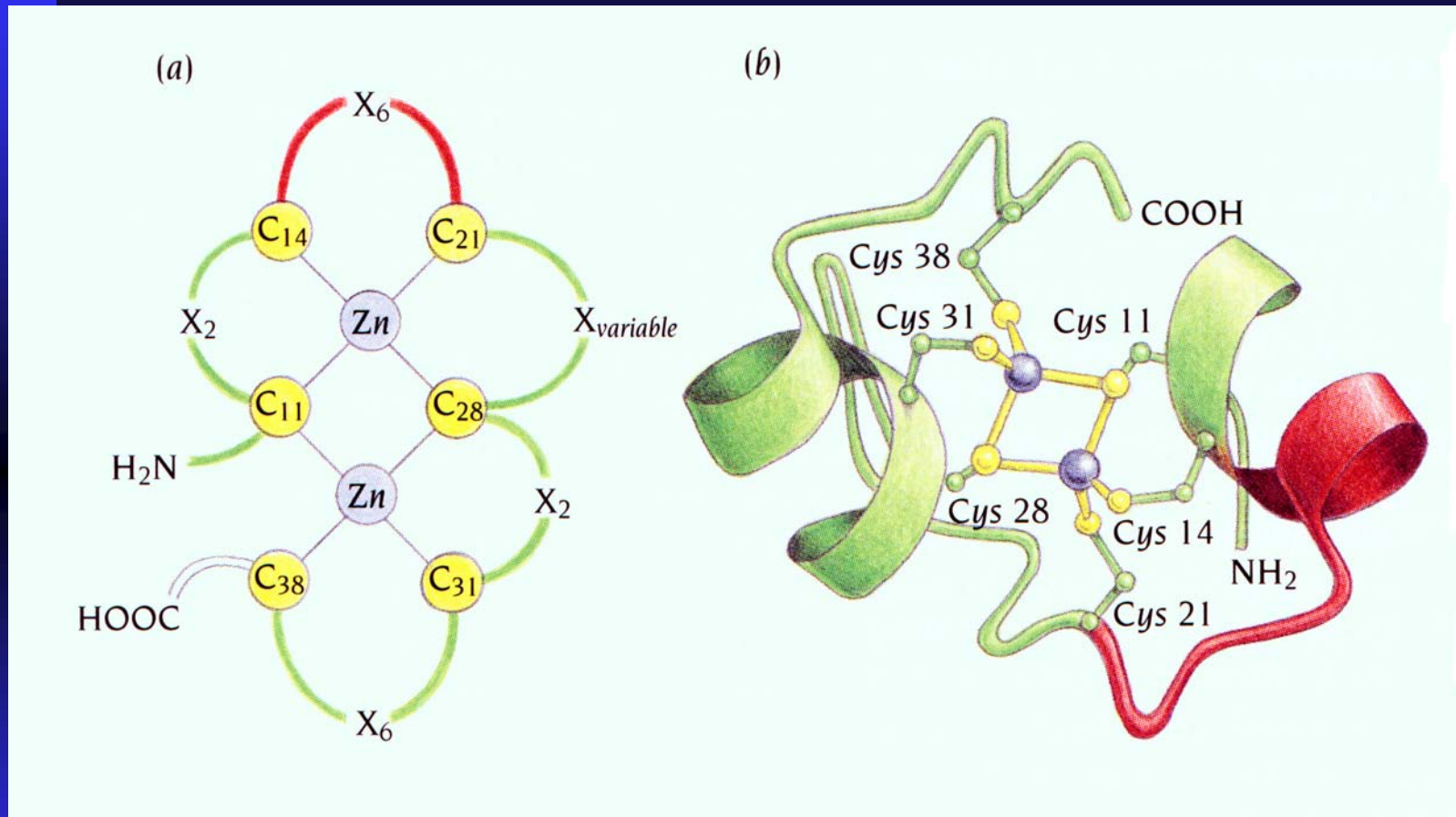
- A third subfamily of zinc-containing motifs has been found in several transcription factors from yeast and fungi.
- The most thoroughly studied member of this family, GAL4, activates transcription of genes required for the break-down of the sugars galactose and melibiose.
- GAL4 binds as a dimer to DNA recognition sequences that are 17 base pairs in length, and some of the genes that are activated by GAL4 have four of these upstream-activating sequences (UAS).

# Yeast transcription factor GAL4 contains a binuclear zinc cluster in its DNA-binding domain



**GAL4** binds to DNA as a **dimer**. The structure of each subunit is divided into three distinct regions: a **dimerization region** joined by a **linker region** to a **DNA-binding  $Zn_2Cys_6$  zinc cluster region**. The two dimerization  $\alpha$  helices are packed into a **coiled coil like a leucine zipper**, and the length of the helix regions determines the distance between the DNA-binding sites.

## DNA-binding zinc cluster region of the GAL 4 subunit

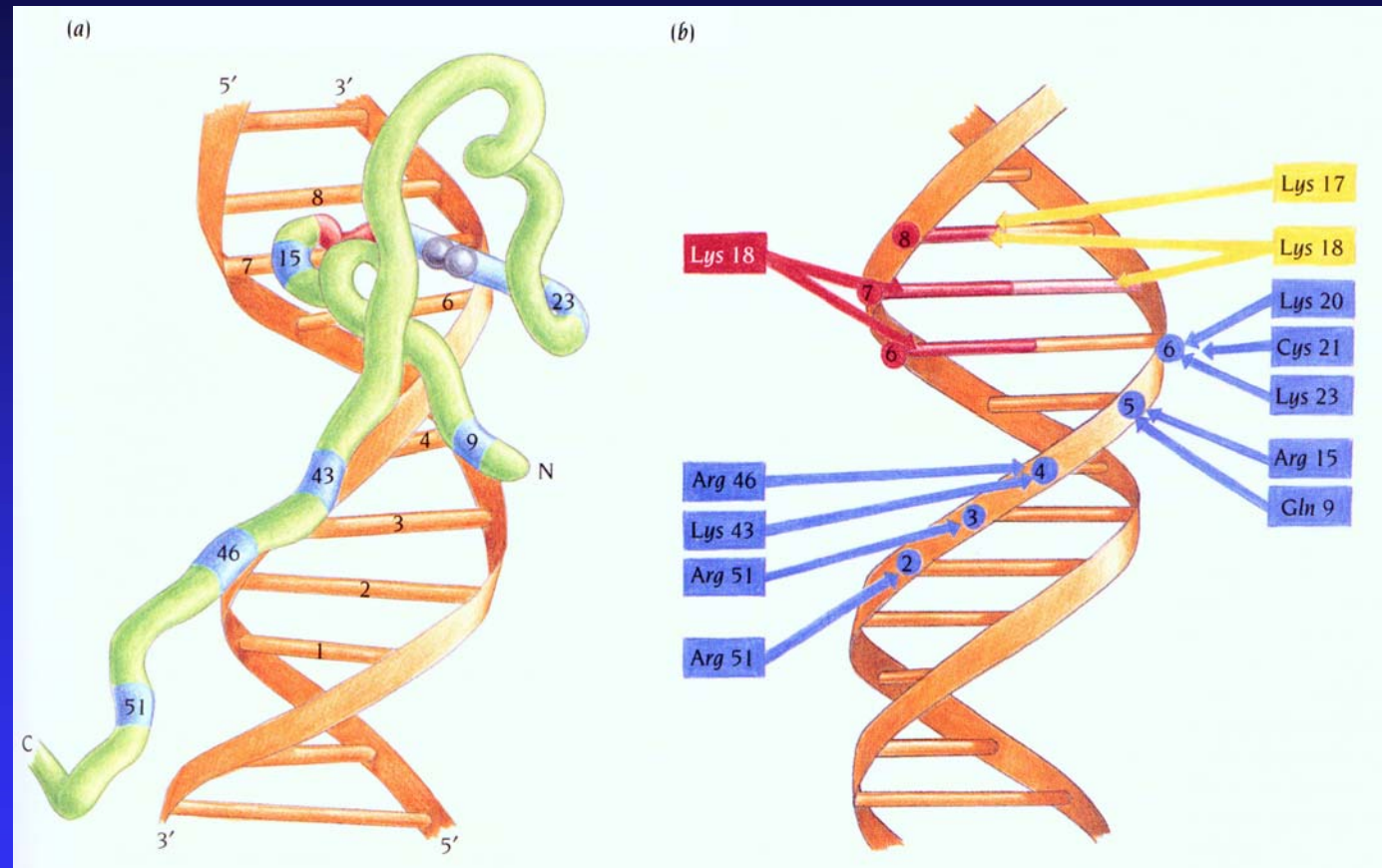


(a) The zinc cluster contains two zinc atoms each bound to four cysteine residues, two of which bridge the zinc atoms. The diagram illustrates the number of amino acids in the loop regions between the cysteine ligands. (b) Richardson-type diagram of the DNA-binding region. **The red region provides the sequence-specific DNA interactions.** The zinc cluster stabilizes the structure to give the proper fold for DNA binding.



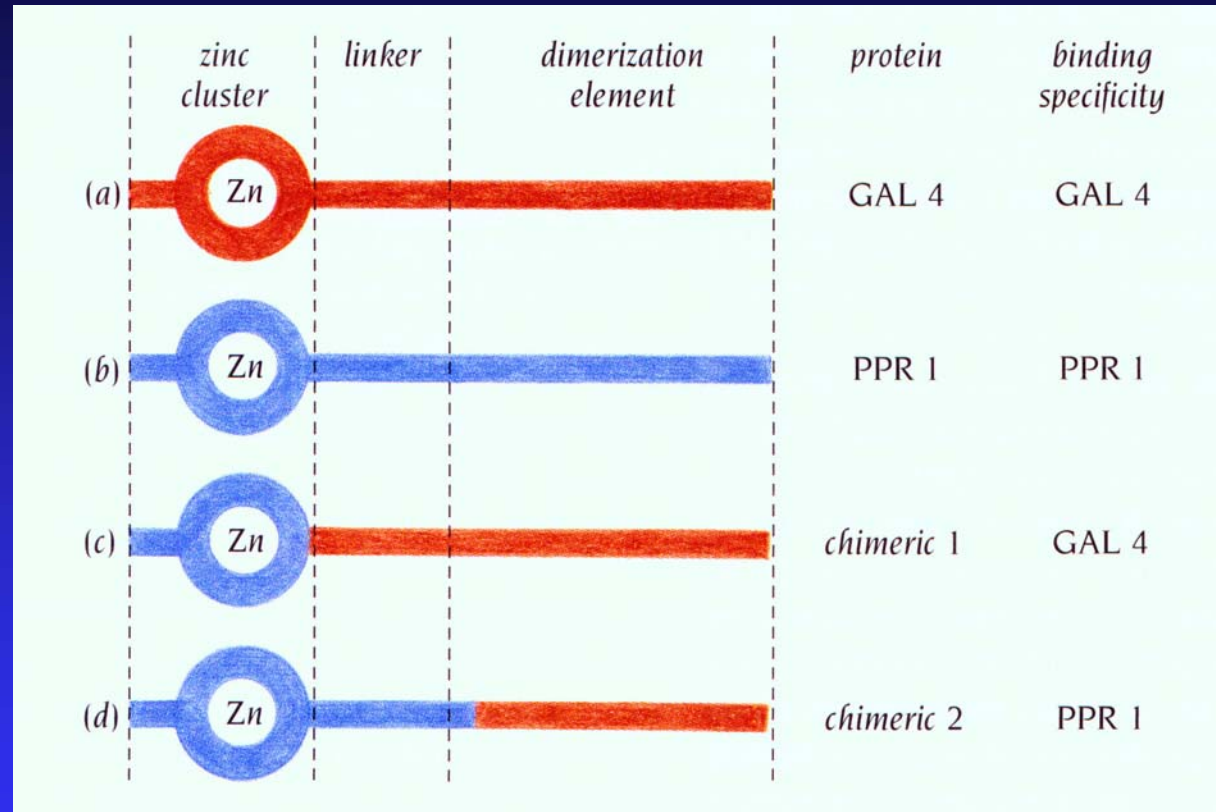
## Binding of one subunit of GAL4 to DNA

(a) Shows the structure of the DNA-binding domain in complex with the DNA; and (b) shows the interactions between amino acid residues and the DNA. The zinc cluster region binds in the major groove of DNA and is anchored to the sugar-phosphate backbone by nonsequence-specific interactions (blue).



The C-terminus of the first  $\alpha$  helix points into the major groove, and two main-chain C=O groups from residues 17 and 18 (yellow) form hydrogen bonds to the edge of the bases. There is **only one sequence-specific interaction with an amino acid side chain, which is provided by Lys 18 (red)**. The linker region is an extended chain that follows one strand of the DNA and provides several nonspecific contacts (blue) to the DNA. The numbering of the base pairs starts from the center of the DNA fragment.

## DNA-binding site specificity among the C<sub>6</sub>-zinc cluster family of transcription factors is achieved by the linker regions



Domain-swapping experiments with GAL4 (a) and a related transcription factor PPR1 (b) have shown that the specificity of DNA binding depends on the size of the linker region. A chimeric protein with the PPR1 zinc cluster combined with GAL4 linker and dimerization regions (c) has GAL4 DNA-binding specificity, whereas a chimeric protein with PPR1 zinc cluster and linker regions combined with a GAL4 dimerization region (d) has PPR1 specificity.

## C<sub>6</sub>-zinc cluster transcription factor summary

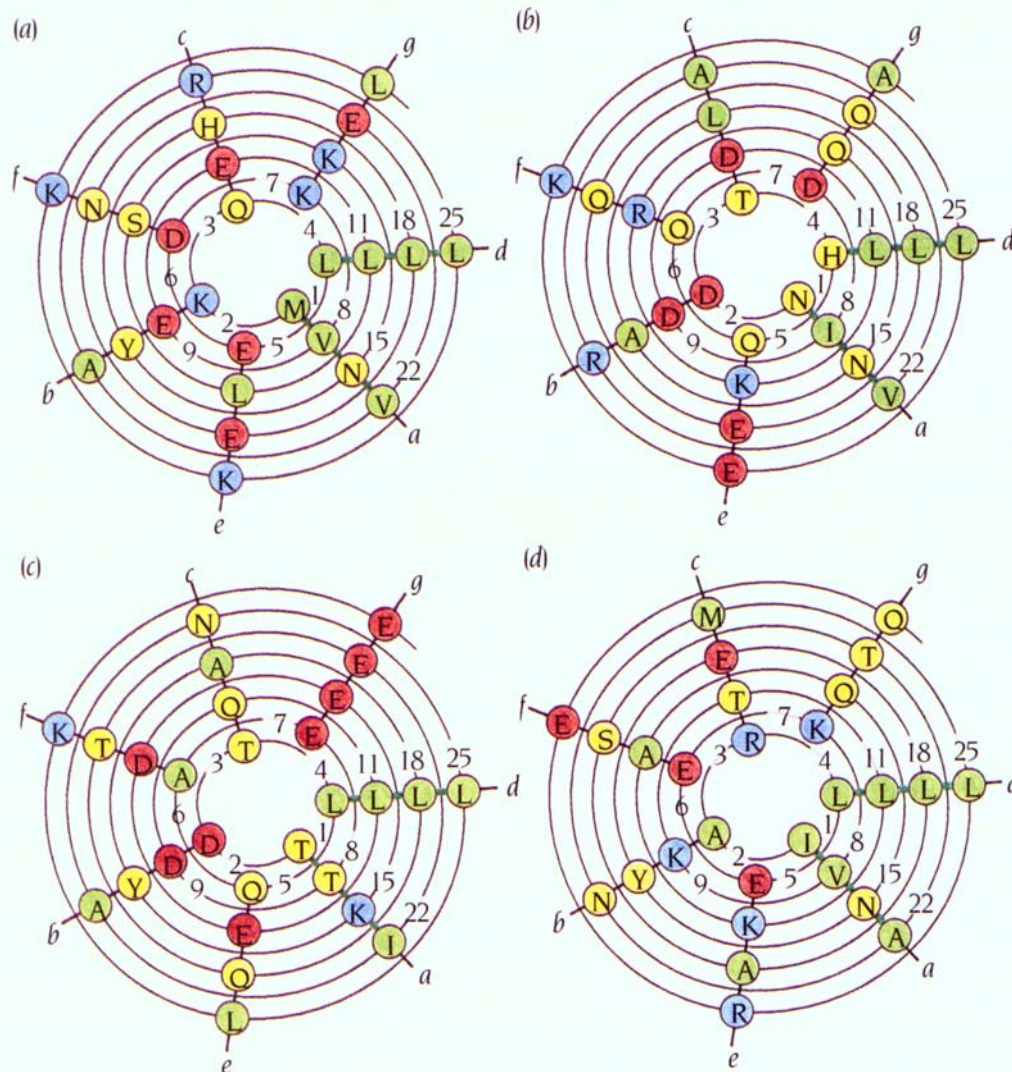
- The C<sub>6</sub>-zinc cluster transcription factors utilize **two zinc atoms bound by six cysteine residues** to stabilize the scaffold of the monomeric subunit.
- The active factor is a **dimer** with a distinct dimerization region. The dimer resembles the letter T: the dimerization region is in the upright stem and the two DNA binding motifs are at the ends of the crossbar.
- All members of this family recognize the same base sequence, the **triplet CCG**.
- Specificity** is achieved by the length of the crossbar, in other words, the **separation of the CCG triplet along the DNA**.



## Leucine zippers provide dimerization interactions for some eucaryotic transcription factors

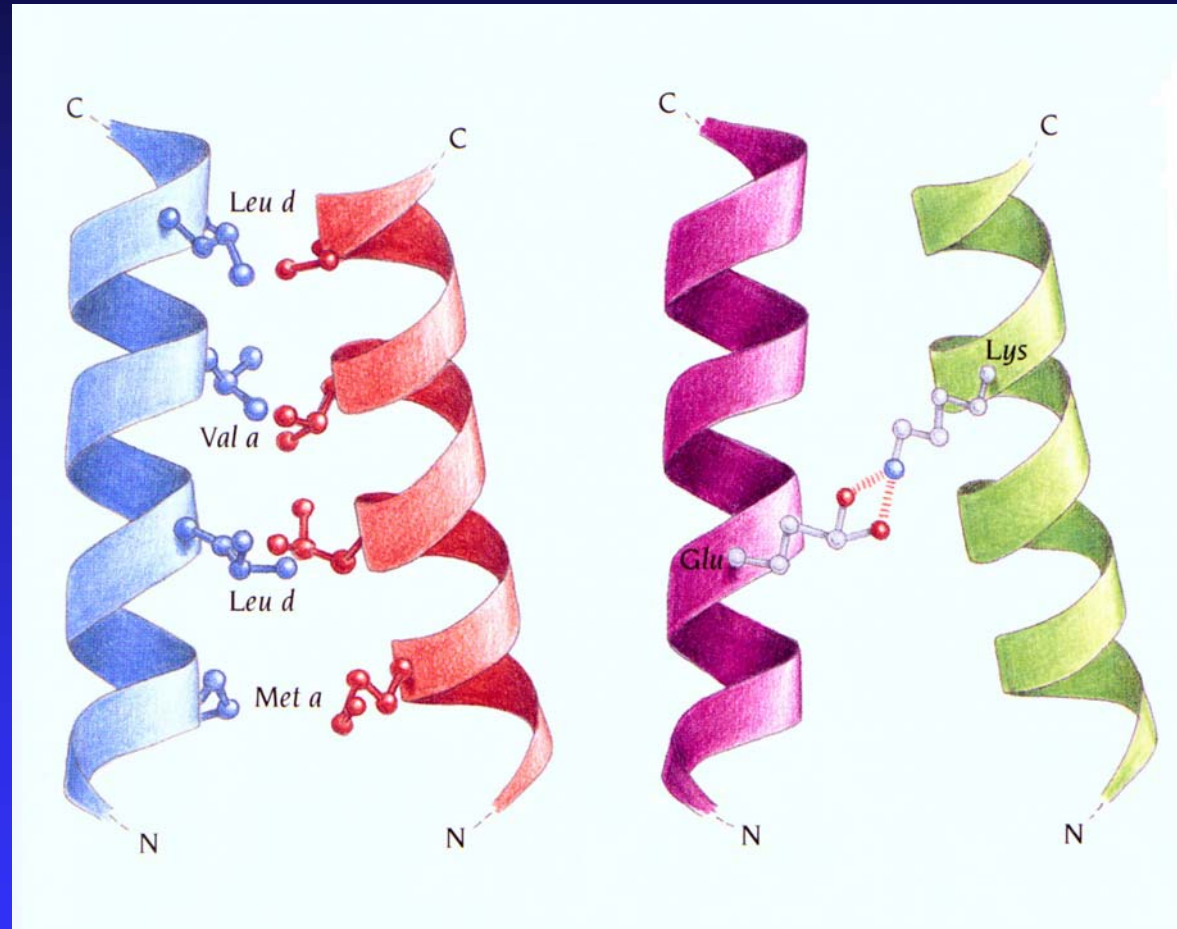
- The leucine zipper motif was first recognized in the amino acid sequences of a yeast transcription factor **GCN4**, the mammalian transcription factor **C/EBP**, and three oncogene products, **Fos**, **Jun** and **Myc**, which also act as transcription factors.
- When the sequences of these proteins are plotted on a helical wheel, a remarkable pattern of leucine residues emerges.
- In all of them there is a region of about 30 residues where the sequence can be arranged in modules of seven residues, and in almost all of these modules the fourth residue is leucine hence the name leucine zipper. In addition, the first residue of each module is frequently hydrophobic.

## Leucine zippers provide dimerization interactions for some eucaryotic transcription factors



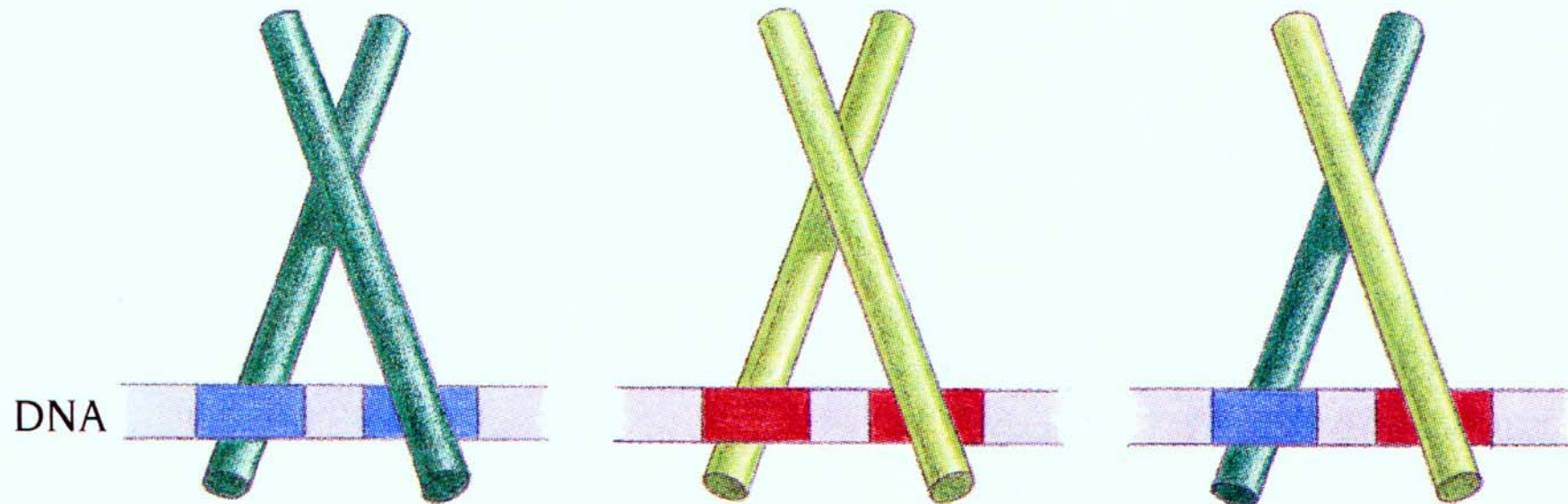
Amino acid sequences, represented as a helical wheels with **3.5 residues per turn**, of the transcription factors (a) GCN4, (b) Max, (c) Fos and (d) Jun. Residues with spacing of seven residues are connected into spikes of the wheel which are labeled a to g. Side chains from each residue in a spike form a ridge on the outside of the  $\alpha$  helix which is parallel to the helical axis. Spike d in all four sequences forms a heptad pattern of leucine residues that is characteristic of a leucine zipper in coiled-coil  $\alpha$ -helical structure.

## Leucine zippers provide dimerization interactions for some eucaryotic transcription factors



Side-chain interactions in the leucine zipper structure. (a) The hydrophobic side chains in spikes *a* and *d* form a hydrophobic core between the two coiled  $\alpha$  helices. (b) Charged side chains in spikes *e* and *g* can promote dimer formation by forming complementary charge interactions between the two  $\alpha$  helices.

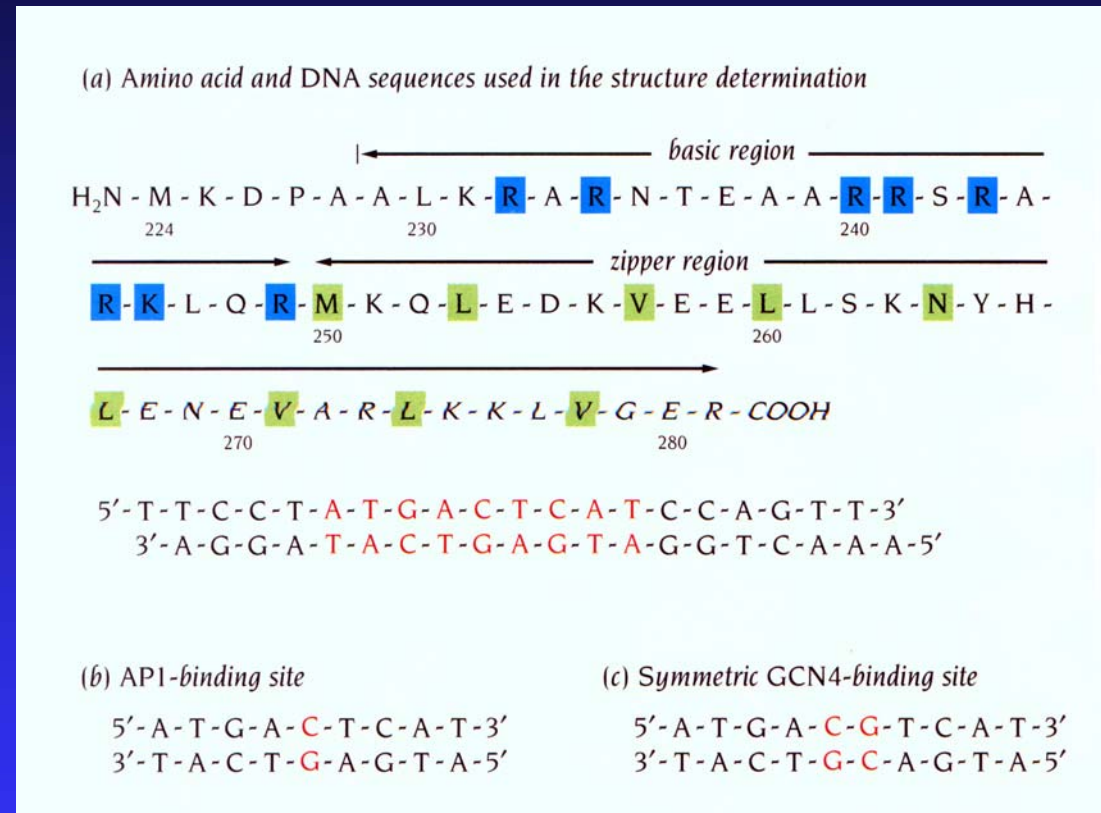
## Heterodimerization of leucine zipper proteins can alter their DNA-binding specificity – the basis of combinatorial control of cellular processes



Leucine zipper homodimers bind to symmetric DNA sequences, as shown in the left-hand and center drawings. These two proteins recognize different DNA sequences, as indicated by the red and blue regions in the DNA. The two different monomers can combine to form a heterodimer that recognizes a hybrid DNA sequence, composed of one red and one blue region.

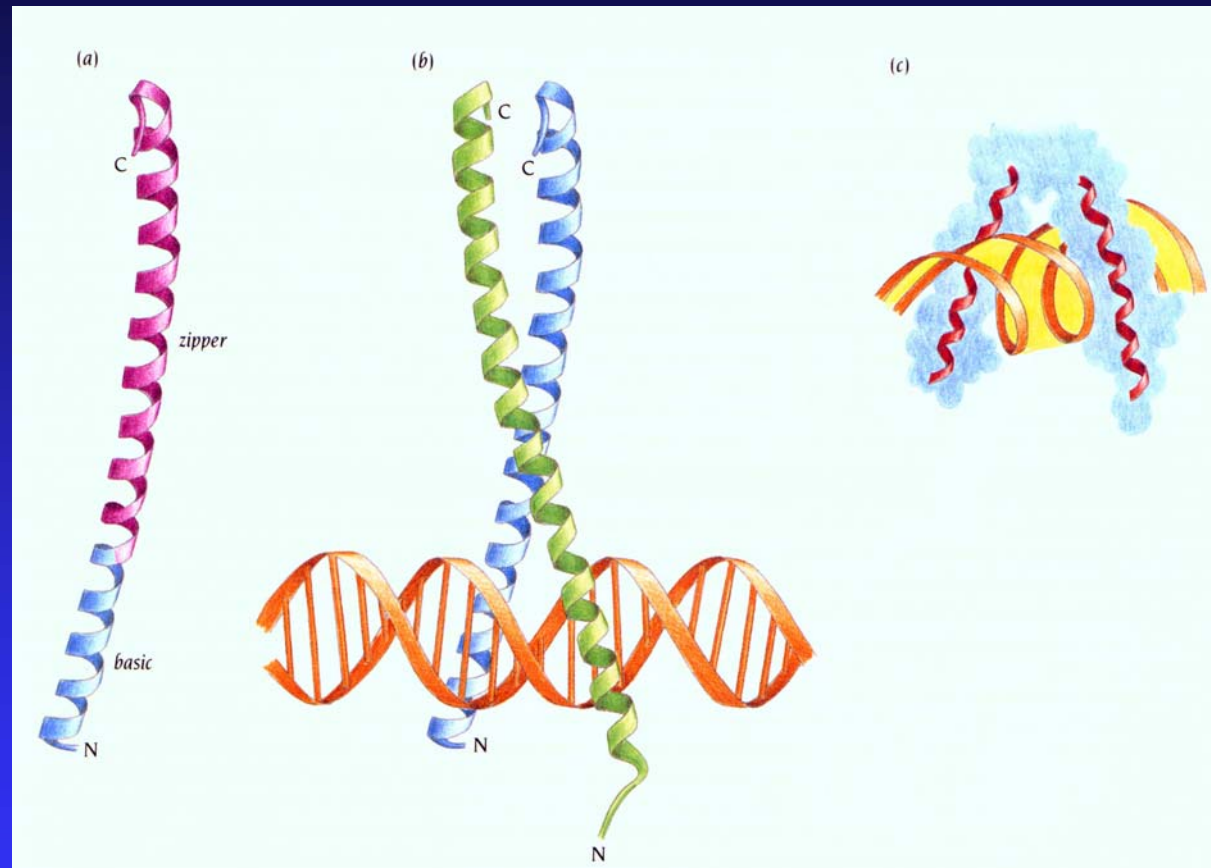


# Amino acid sequence of the DNA-binding domain of the transcription factor GCN4 and nucleotide sequence of the DNA used for the x-ray structure determination of the complex



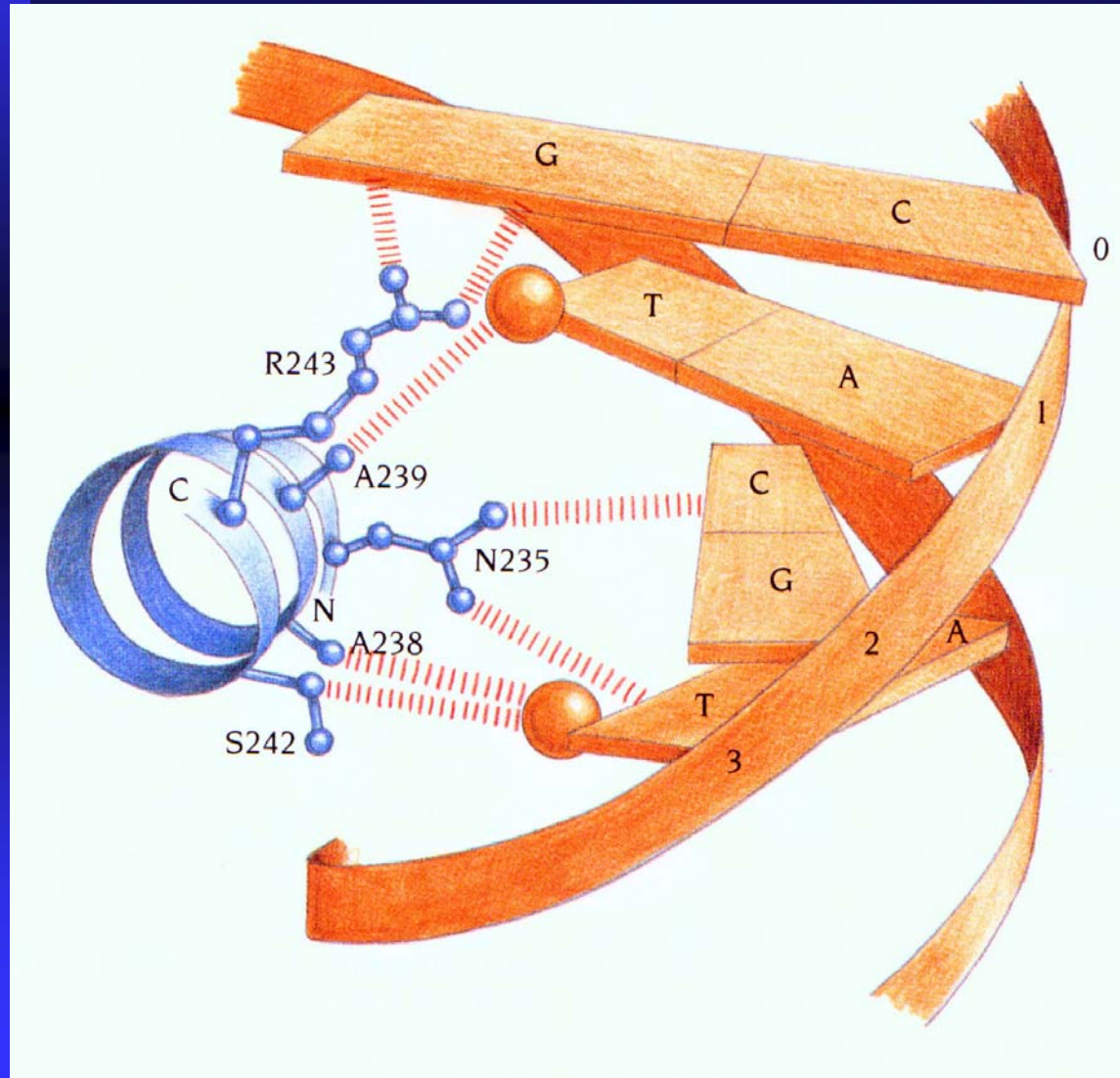
The positively charged residues **arginine** and **lysine** in the basic region that are involved in **DNA binding** are colored blue. The residues in *a* and *d* positions that form the hydrophobic core of the leucine zipper region are green. The pseudo-symmetric nucleotide sequence to which GCN4 binds is colored red (b). The consensus nucleotide sequence of the *in vivo* AP1 pseudo-symmetric DNA recognition sites of GCN4. (c) A symmetric nucleotide sequence to which GCN4 can also bind.

## The GCN4 basic region leucine zipper binds DNA as a dimer of two uninterrupted $\alpha$ helices



(a) Each monomer of the GCN4 domain forms a smoothly curved continuous  $\alpha$  helix comprising both the basic and the leucine zipper regions. (b) The monomers are held together in a dimer in the zipper region. They diverge from each other in the basic regions, which are bound to DNA in the major groove on opposite sides of the B-DNA fragment (c), like helical forceps gripping the DNA.

## GCN4 binds to DNA with both specific and nonspecific contacts



Sequence-specific interactions between one of the  $\alpha$ -helical basic regions of GCN4 (blue) and bases in the DNA fragment (orange). The methyl groups of the thymine bases are shown as spheres.

## The HLH motif is involved in homodimer and heterodimer associations

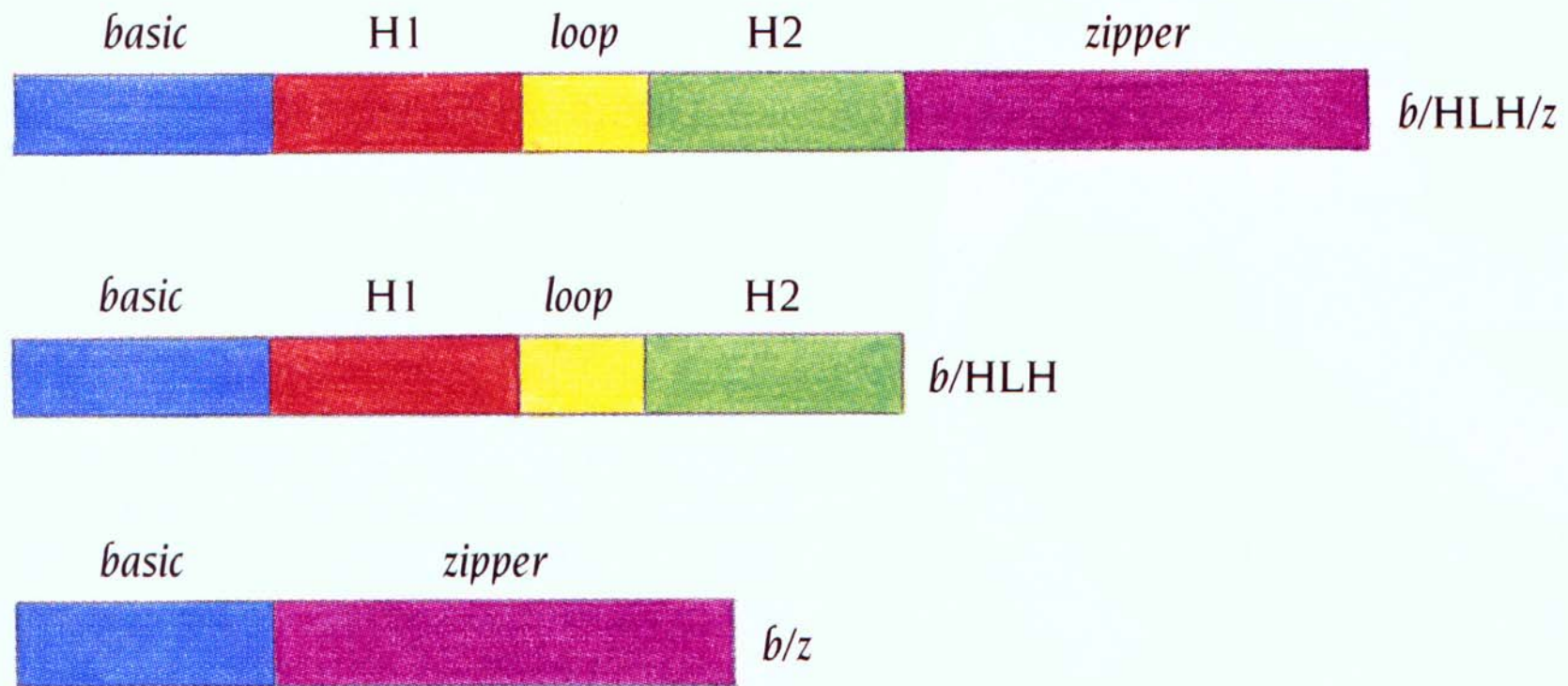
- The helix-loop-helix (HLH) family of transcription factors dimerize through formation of four-helix bundle structure.
- In these proteins, the helix-loop-helix region is preceded by a sequence of basic amino acids that provide the DNA-binding site, and therefore these proteins are called the **b/HLH** transcription factors, just as members of the leucine zipper family are called **b/zip** factors.
- Members of the **b/HLH family** have substantial amino acid sequence identity and they bind to the consensus DNA sequence **5'-CANNTG-3'**, where N is any nucleotide.



## The b/HLH transcription factors

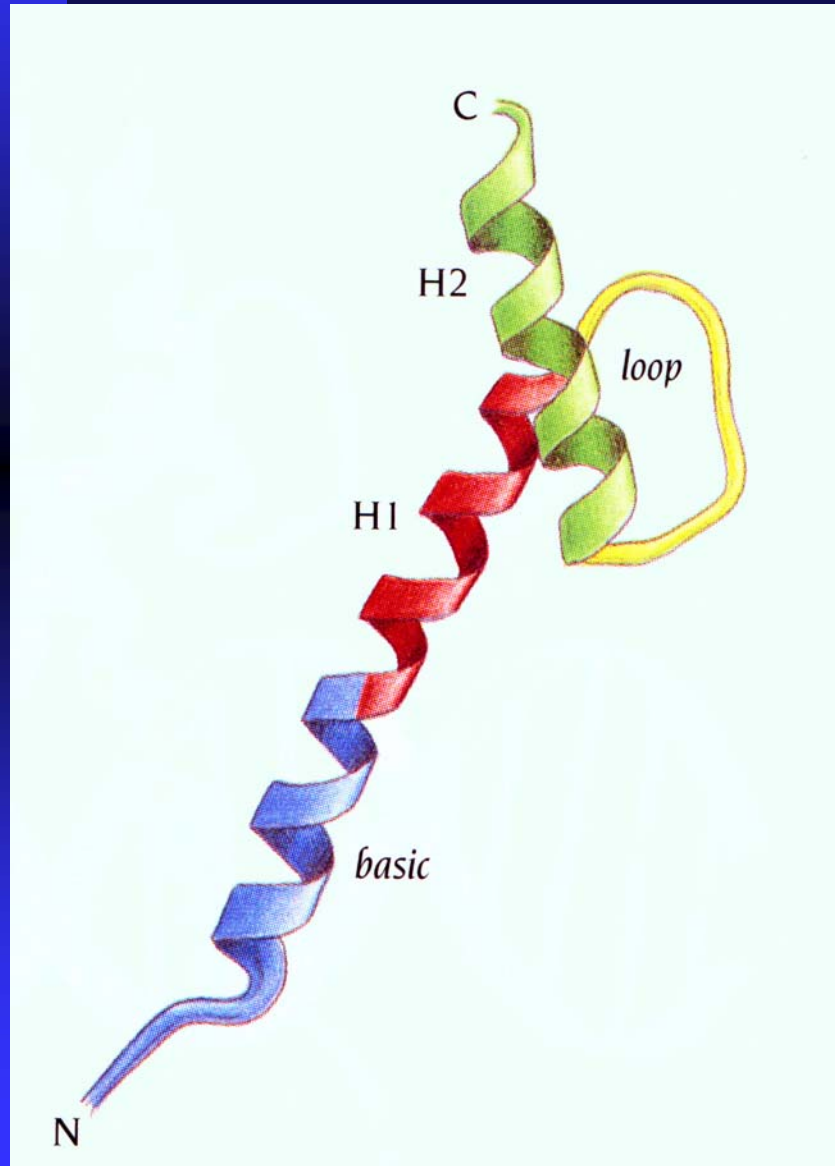
- The **myogenic proteins** are an important class of b/HLH transcription factors that are crucially involved in the **development of muscle cells**.
- A **mature muscle** is distinguished from the other cell types by a large number of characteristic proteins, including **specific types of actin and myosin**.
- The entire program of muscle differentiation can be triggered *in vitro* in cultured skin fibroblast cells by introducing any one of the myogenic proteins. **Introduction of one myogenic protein activates the production of all the others, which in combination with already present present but inactive HLH proteins then activate the muscle-specific genes.**
- Thus, muscle differentiation is determined by specific combinations of these transcription factors, including heterodimers.

## Domain arrangement along the polypeptide chains of three families of transcription factors: b/z, b/HLH and b/HLH/z



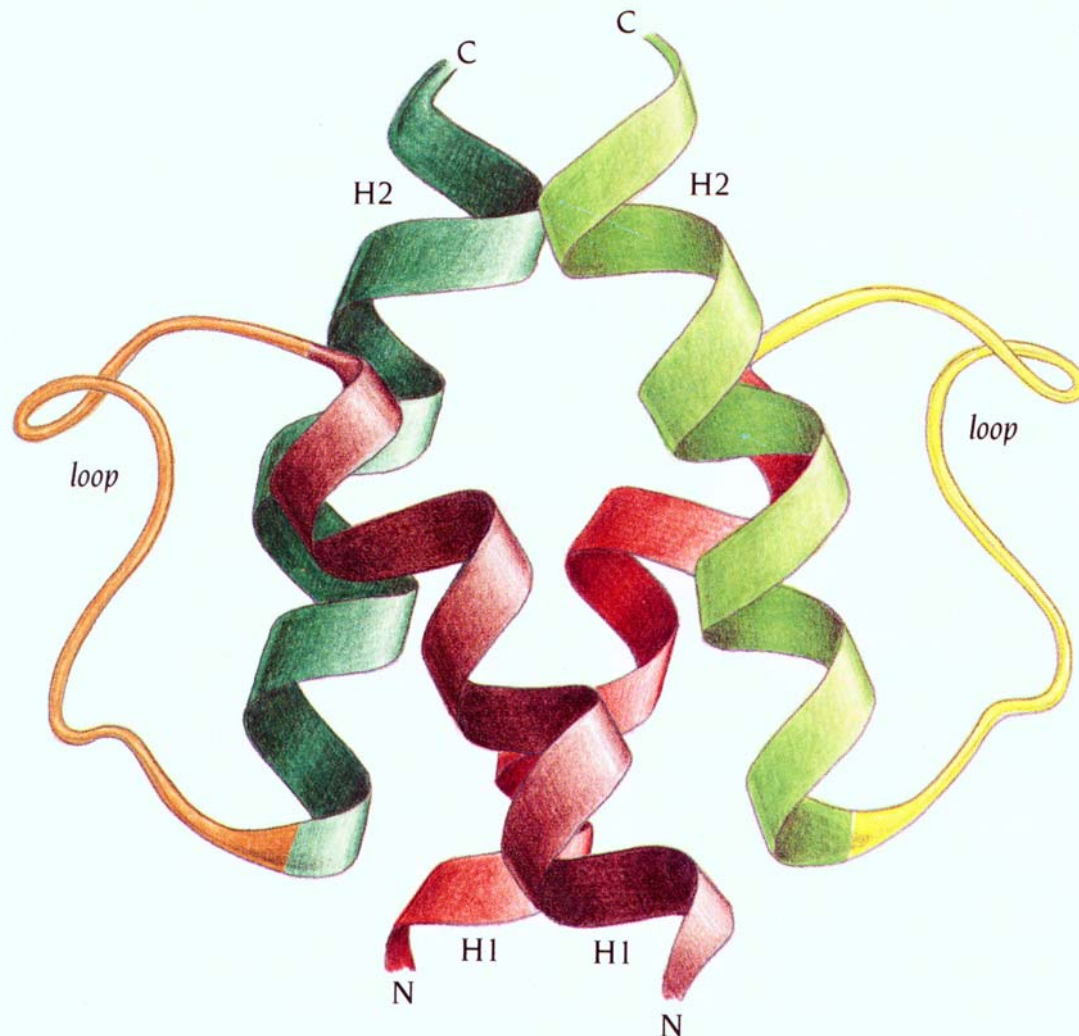
All three have a basic region (blue) that binds DNA. Dimerization is achieved by the zipper region (purple) in the b/z family, by the H1-loop-H2 region (red-yellow-green) in the b/HLH family and by a combination of both the zipper and the HLH regions in the b/HLH/z family.

## Structure of a monomer of the DNA-binding domain of the transcription factor MyoD



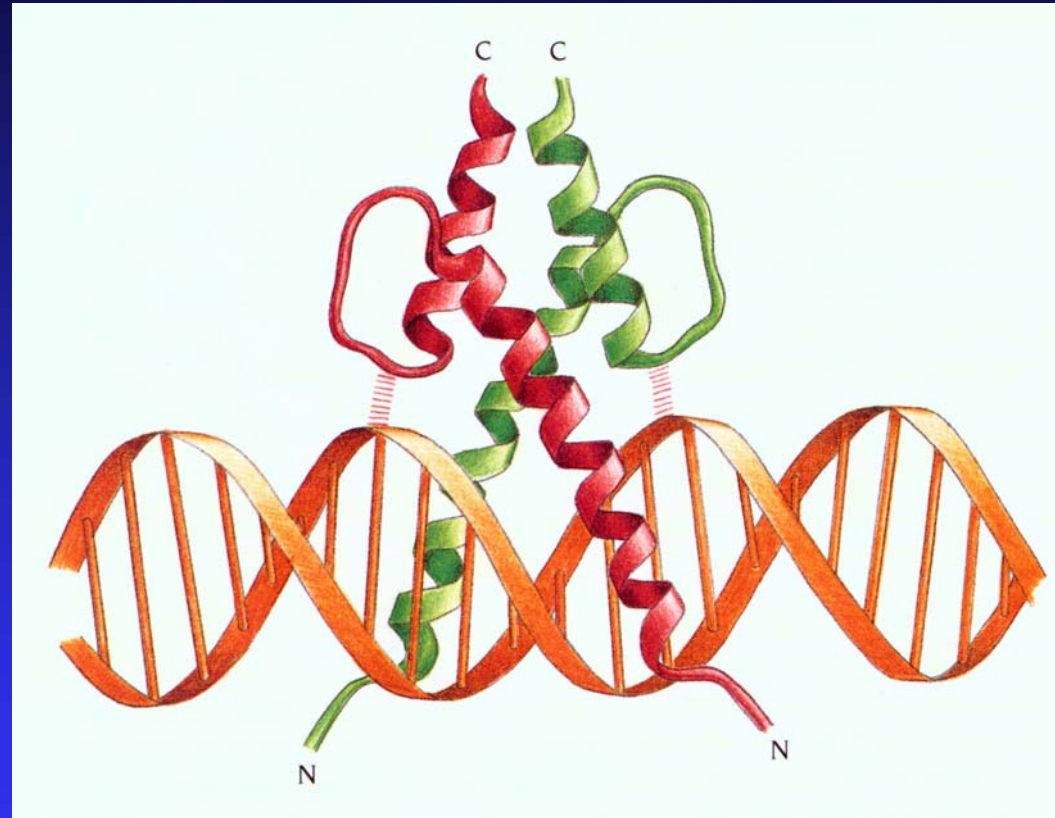
The domain, which belongs to the b/HLH family, comprises two  $\alpha$  helices joined by a loop region. The basic region and the first helix H1 of the helix-loop-helix region form one continuous  $\alpha$  helix.

## Structure of the dimerization region of MyoD



The  $\alpha$  helices H1 and H2 of the two monomers form a four-helix bundle that keeps the dimer together. The loops are on the outside of the four-helix bundle.

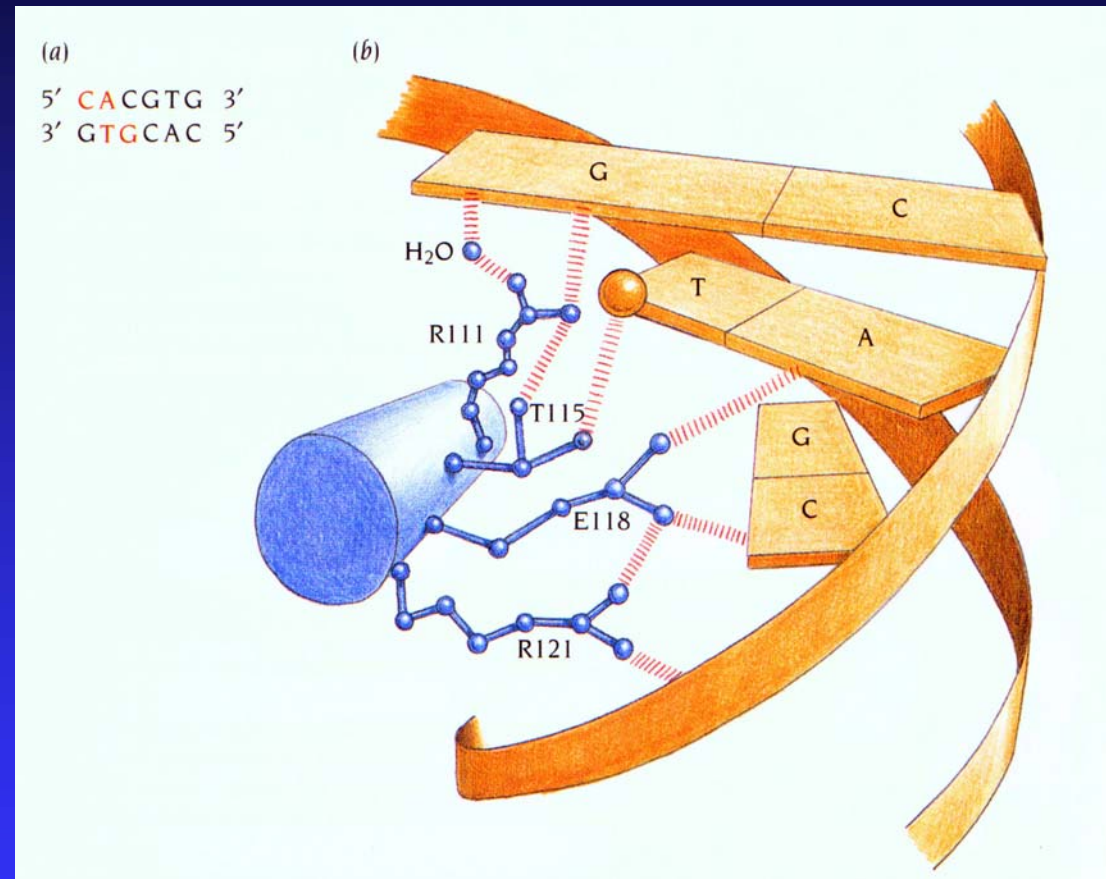
## The $\alpha$ -helical basic region of the b/HLH motif binds the major groove of DNA



MyoD binds to DNA as a dimer with the N-terminal basic regions interacting with the major groove of DNA. The monomers diverge from each other after the H1 region and bind on opposite sites of DNA, as for GCN4. The **four-helix bundle**, together with contacts with phosphates of the DNA backbone, rigidify the fork and allow MyoD only to **bind to enhancer elements with a specific spacing between the half-sites**. The DNA structure is essentially B-DNA.



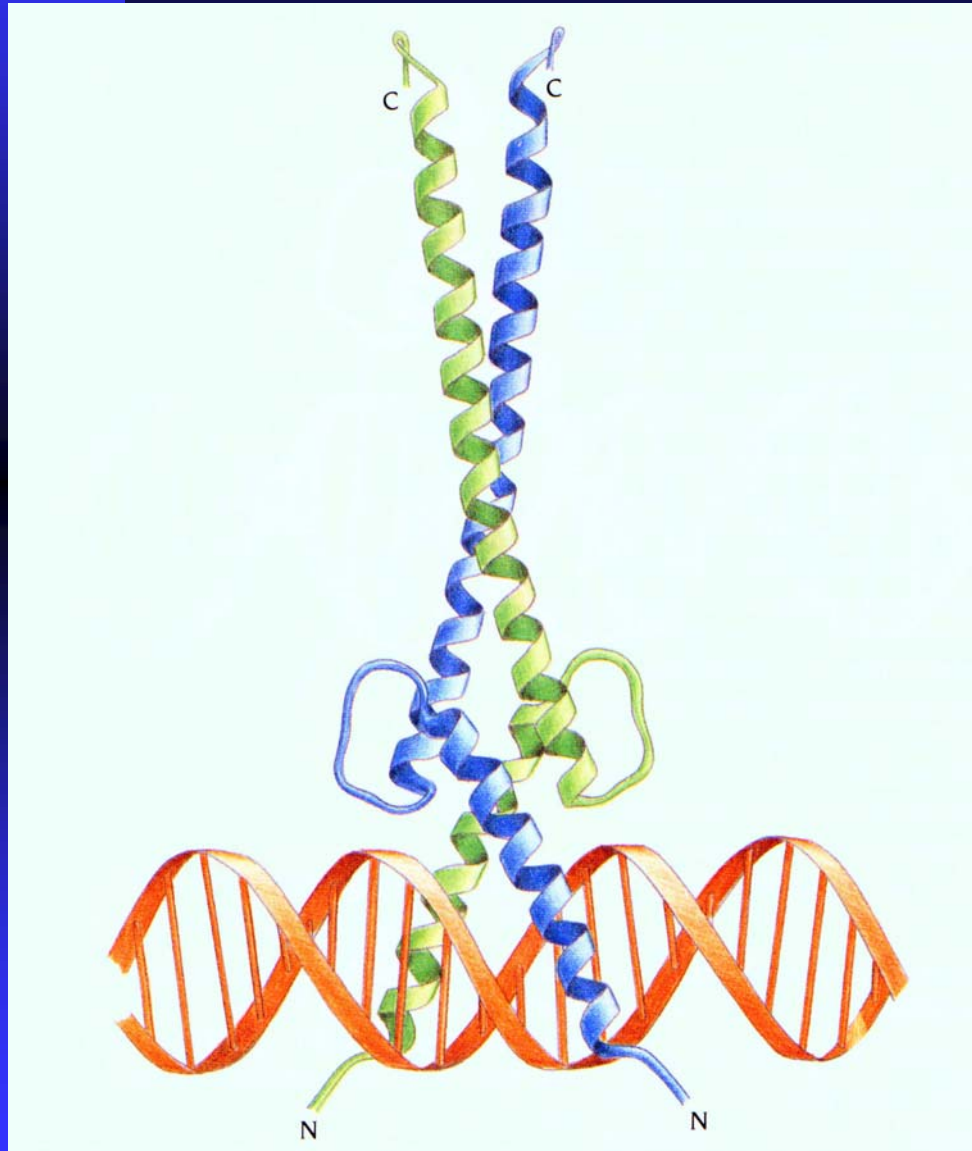
## Sequence-specific contacts between DNA and one monomer of MyoD



(a) The palindromic MyoD recognition sequence. Bases that contact one monomer of MyoD are shown in red. (b) Three residues, Glu118, Thr115 and Arg111 from the basic region of MyoD, bind to the edges of the bases in the major groove. Glu118 recognizes the first two bases, C and A, in the MyoD consensus recognition sequence.

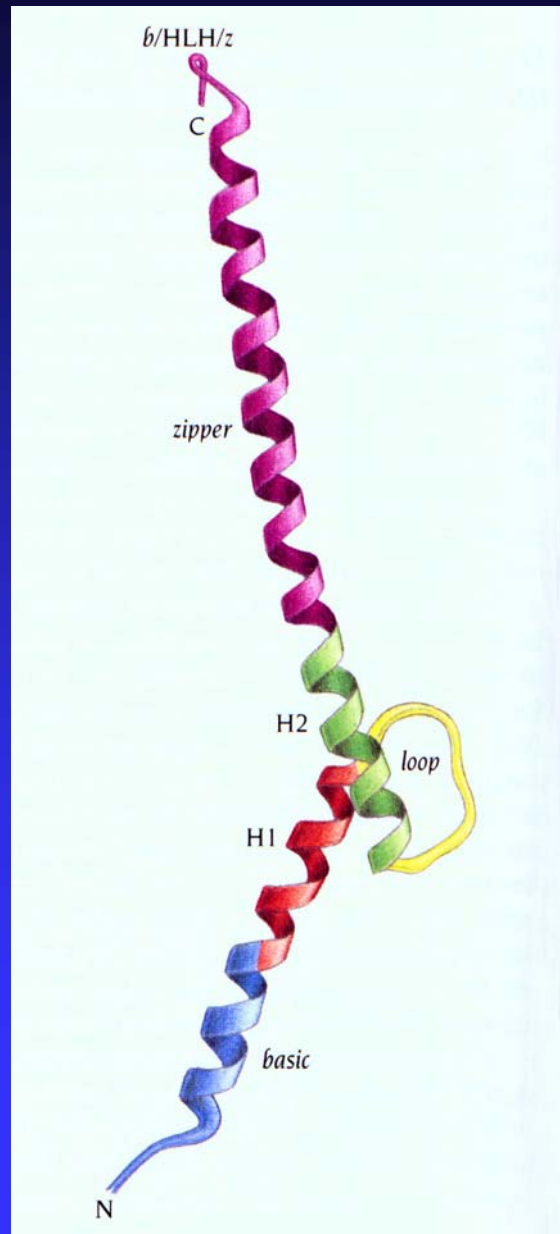


## The b/HLH/zip family of transcription factors have both HLH and leucine zipper dimerization motifs



The binding of the transcription factor **Max** to DNA. The two monomers of Max form a dimer through both the helix-loop-helix regions which form a four-helix bundle like MyoD, and the zipper regions, which are arranged in a coiled coil. The N-terminal basic regions bind to DNA in a way similar to GCN4 and MyoD.

## H2 and zip form one long $\alpha$ helix in the Max monomer



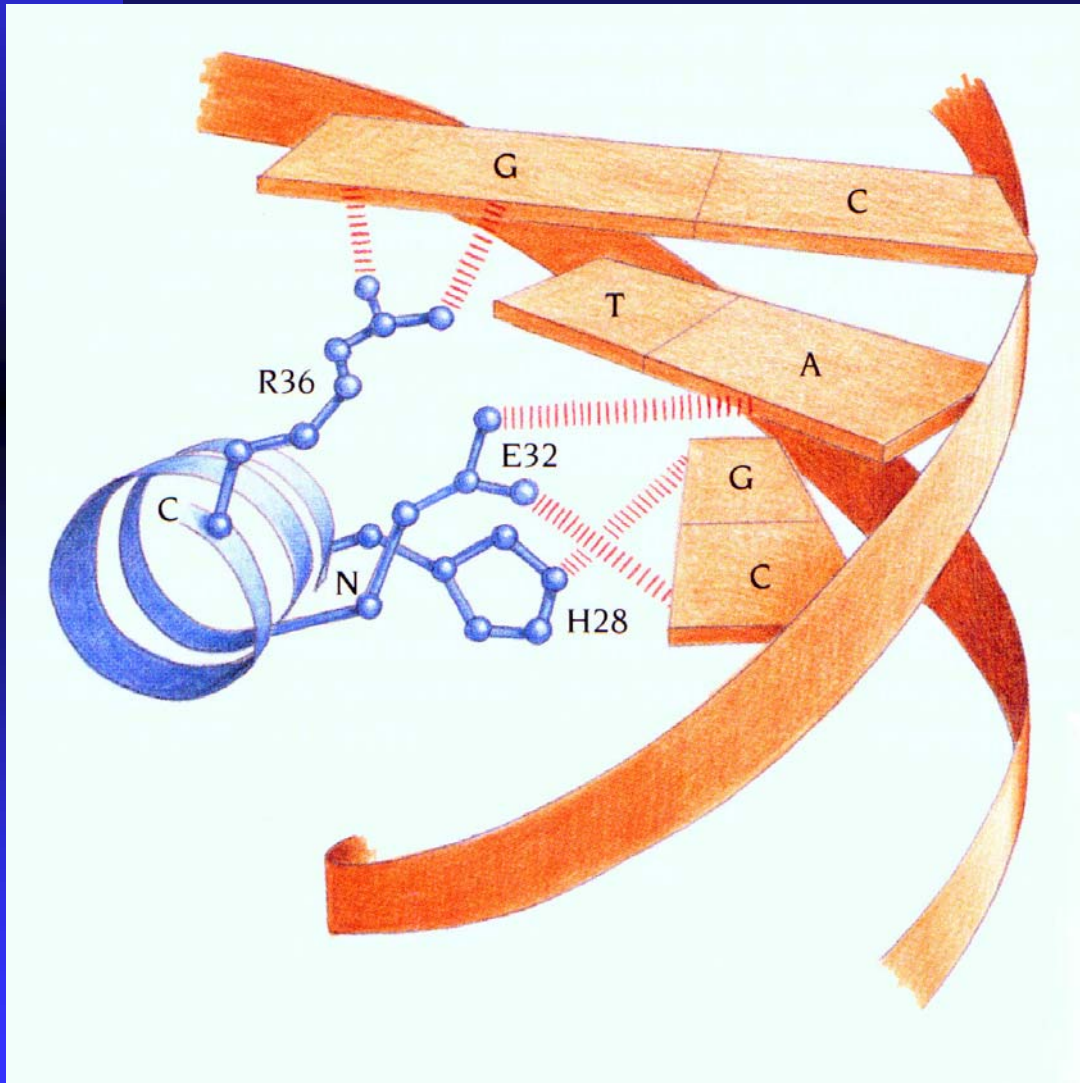
The structure of the Max monomer is essentially built up from two long  $\alpha$  helices joined by a loop region. The basic region and H1 of the helix-turn-helix region form one continuous  $\alpha$  helix, and H2 and the zipper region form a second continuous  $\alpha$  helix.

## Amino acid sequences of the helix-turn-helix region of some members of the b/HLH and b/HLH/zip families of transcription factors

		H1											loop											H2																									
Max	NH <sub>2</sub>	-	D	H	I	K	D	S	F	H	S	L	R	D	S	V	P	-	-	-	-	-	-	S	L	Q	G	E	K	A	S	-	R	A	Q	I	L	D	K	A	T	E	Y	I	Q	Y	M	-	COOH
C-Myc		-	N	E	L	K	R	S	F	F	A	L	R	D	Q	I	P	-	-	-	-	-	-	E	L	E	N	N	E	K	A	P	-	K	V	V	I	L	K	K	A	T	A	Y	I	L	S	V	-
MyoD		-	S	K	V	N	E	A	F	E	T	L	K	R	C	T	S	-	-	-	-	-	-	S	N	P	N	Q	R	L	P	-	K	V	E	I	L	R	N	A	I	R	Y	I	E	G	L	-	
CBF1		-	E	N	I	N	T	A	I	N	V	L	S	D	L	L	P	-	-	-	-	-	-	-	V	R	E	S	S	-	K	A	A	I	L	A	R	A	A	E	Y	I	Q	K	L	-			
Pho4		-	N	R	L	A	V	A	L	H	E	L	A	S	L	I	P	-	A	E	W	K	Q	Q	N	V	S	A	A	P	S	-	K	A	T	T	V	E	A	A	C	R	Y	I	R	H	L	-	

Residues that form the hydrophobic core of the four-helix bundle are colored green and a conserved lysine residue is blue. The loop region between H1 and H2 is highly variable in length but must be at least four or five residues long.

## Sequence-specific interactions between DNA and one monomer of Max



Three residues, His 28, Glu 32 and Arg 36, form specific interactions with the edges of the bases in the major groove of DNA. Like MyoD, a Glu residue recognizes the first two bases, C and A, of the recognition sequence.

## Summary

- Homodimerization, the common feature of procaryotic transcription factors, is in eucaryotes extended by heterodimerization (as in the leucine zippers) and by concatenation (as in zinc fingers and POU).
- These devices increase the range of DNA sequences that can be recognized and the degree of binding specificity.
- In evolutionary terms this increase in range and specificity became necessary as genomes became larger and more complex.
- The real complexity of regulation of gene expression in eucaryotes has so far barely been touched upon by structural biologists, because it lies in the assembly of aggregates of general and specific transcription factors at the promoter region.
- The biochemical data are strong enough to conclude that the various transcription factors bound to distinct sites at a complex promoter have specific 3-D relationships, and so form a macromolecular assembly analogous to a ribosome, proteasome or spliceosome.