Leucine Zippers

Advanced article

Toshio Hakoshima, *Nara Institute of Science and Technology, Nara, Japan*

The leucine zipper (ZIP) motif consists of a periodic repetition of a leucine residue at every seventh position and forms an α-helical conformation, which facilitates dimerization and in some cases higher oligomerization of proteins. In many eukaryotic gene regulatory proteins, the ZIP motif is flanked at its *N*-terminus by a basic region containing characteristic residues that facilitate DNA binding.

Article contents

- Introduction
- Structural Basis of ZIP
- Occurrence of ZIP and Coiled-coil Motifs
- Dimerization Specificity of ZIP
- DNA-binding Specificity of bZIP

doi: 10.1038/npg.els.0005049

Introduction

A structure referred to as the leucine zipper or simply as ZIP has been proposed to explain how a class of eukaryotic gene regulatory proteins works (Landschulz *et al.*, 1988). A segment of the mammalian CCAAT/enhancer binding protein (C/EBP) of 30 amino acids shares notable sequence similarity with a segment of the cellular Myc transforming protein. The segments have been found to contain a periodic repetition of a leucine residue at every seventh position. A periodic array of at least four leucines has also been noted in the sequences of the Fos and Jun transforming proteins – gene regulatory proteins encoded by the proto-oncogenes c-*fos* and c-*jun* that have a role in cell growth control – as well as in that of the yeast gene regulatory protein GCN4. (***See*** Enhancers; Oncogenes; Transcription Factors.)

These regulatory proteins also contain a segment of 20 amino acids that is rich in basic residues at the amino (*N*)-terminal flanking region of the ZIP motif. This basic region has been thought to facilitate DNA binding by interacting with negatively charged phosphate groups of DNA. The segment that comprises the basic region and the periodic arrays of leucine residues is referred to as the 'basic-region leucine zipper' or bZIP motif.

The polypeptide segments containing these periodic arrays of leucine residues are thought to exist in an α-helical conformation, and the leucine side chains extending from one α helix interdigitate with those displayed from a similar α helix of a second polypeptide, which facilitates dimerization. In the dimerized form, two basic regions stick out from the coiled coil of the ZIP segments to interact with DNA. This hypothetical structure has been proved by *in vitro* biochemical studies and by several crystal structural determinations that include protein–DNA complexes (Ellenberger *et al.*, 1992; Glover and Harrison, 1995; Fujii *et al.*, 2000).

The ZIP motif and the related ZIP-like motifs are frequently found in proteins other than transcription factors and are thought to be one of the general protein modules for protein–protein interactions. Knowing the structure and function of these motifs enables us to understand the molecular recognition system in several biological processes.

Structural Basis of ZIP

The α helix is a secondary structure element that occurs frequently in proteins. Alpha helices are stabilized in proteins by being packed into the hydrophobic core of a protein through hydrophobic side chains. The simple way to achieve such stabilization is to pack two helices together. The side-chain interactions are maximized if the two helices are wound around each other in a coiled-coil or supercoil arrangement. Thus, the α-helical coiled coil is one of the principal subunit dimerization motifs in proteins. Coiled-coil structures of polypeptide chains can be formed by two antiparallel α helices, as well as by two parallel α helices. In general, however, the ZIP motif forms a parallel coiled-coil structure.

The α helix is a right-handed helical structure with 3.6 residues per turn. This residue number indicates that two residues that are separated by seven residues in sequence are located at nearly the same molecular surface in an α helix. Francis Crick (1953) showed that a left-handed coiled coil of two α helices reduces the number of residues per turn in each helix from 3.6 to 3.5, so that the pattern of side-chain interactions between the helices repeats exactly every seven residues after two turns. This is reflected in the amino acid sequences of polypeptide chains that form an α-helical coiled coil. Such sequences are repetitive with a period of seven residues, which is known as the heptad repeat. The amino acid residues within one such heptad repeat are usually labeled ***a*–*g*** (**Figure 1**). The heptad repeats of coiled coils show that nonpolar residues occur preferentially in the ***a*** and ***d*** positions.

Of the heptad repeats that form coiled coils, the ZIP motif is characterized by the occurrence of leucine at every ***d*** position. In the ZIP structure, leucines at the ***d*** positions pack against each other every second turn

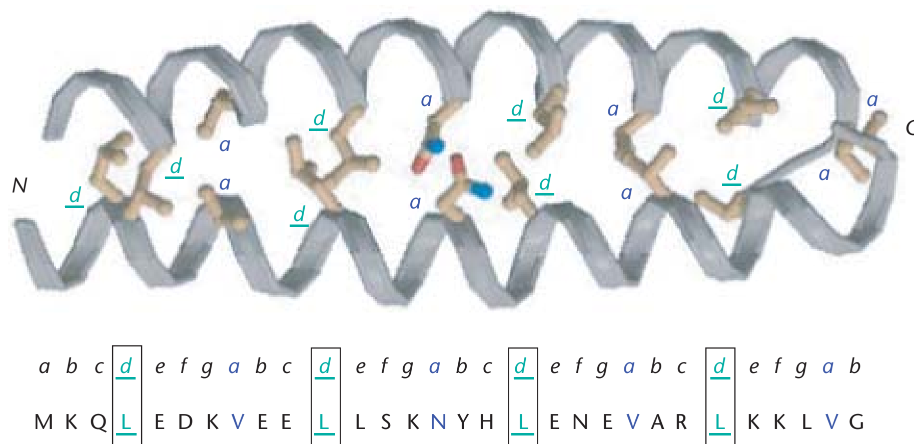

Figure 1 Parallel coiled-coil structure of GCN4 ZIP homodimer (Protein Data Bank (PDB) accession code 1gd2). The main chains of the two peptide chains are represented as ribbons in gray. The side chains participating in the dimer association are represented as stick models with carbon atoms in brown, nitrogen atoms in blue and oxygen atoms in red. The positions of the heptad repeat are labeled ***a***–***g***. The ***d***-positioned leucines are boxed and highlighted in green with underline. The ***a***-positioned residues are highlighted in blue. Unusually, this GCN4 ZIP motif has a polar asparagine residue in the ***a*** position located at the middle of the ZIP motif.

of the α helices. The hydrophobic region between two helices is completed by residues at the ***a*** positions, which are also frequently hydrophobic. The regularity of the well-known 'knobs-into-holes' packing of the coiled-coil residues is maintained in the dimerized ZIP motifs. Side chains of ***d***-positioned leucines ('knobs') pack into 'holes' formed by the side chains of one ***d***-, one ***e***- and two ***a***-positioned residues from the opposite helix. Side chains of ***a***-positioned residues pack into holes formed by the side chains of one ***a***-, one ***g***- and two ***d***-positioned residues on the opposite helix. Side chains of ***e***- or ***g***-positioned residues are often long charged residues, but the aliphatic portions of their side chains pack against the knob residues to form the solvent-exposed walls of the holes. Thus, the side chains of the ***d***-positioned leucines and the ***a***-positioned nonpolar residues extending from one α helix interdigitate with those displayed from a similar α helix of a second polypeptide, which promotes cooperative formation of a coiled-coil structure and facilitates dimerization.

The bZIP pattern is present in many gene regulatory proteins in eukaryotic cells. The typical ZIP motif contains a periodic array of at least five leucines, and the periodic repetition of a leucine residue at every seventh position occurs over a distance covering eight helical turns. Because the rise per turn of α helix is 5.4 Å along the helical axis, a typical ZIP motif is roughly 40 Å long. The two basic regions of the dimerized bZIP motifs constitute flanking α helices from the ZIP coiled coils and contact DNA with DNA sequence-specific interactions (**Figure 2**).

Occurrence of ZIP and Coiled-coil Motifs

In a nonredundant database containing no identical pairs of sequences (Schultz *et al.*, 2000), 792 bZIP proteins are recorded. The numbers of bZIP proteins in the cells of selected organisms are as follows: 16 in yeast (*Saccharomyces cerevisiae*), 30 in nematode (*Caenorhabditis elegans*) and 110 in fruitfly (*Drosophila melanogaster*). At present, 118 and 114 proteins have been identified in cells from the plant *Arabidopsis thaliana* and from humans respectively.

The ZIP motif also occurs in a form fused with other DNA-binding motifs or domains. An example is the basic region helix–loop–helix (bHLH) motif, which also forms a superfamily of eukaryotic transcription factors. The bHLH motif comprises two amphipathic α helices joined by a variable-length linker region that can form a loop. The HLH region mediates protein dimerization for DNA binding and specifically binds to DNA through a short basic region adjacent to the *N*-terminus of each HLH region. A subset of the bHLH proteins has the ZIP motif linked to the carboxy (*C*)-terminus of the HLH motif. These proteins are referred to as the bHLH-ZIP proteins. The ZIP motifs in these proteins are thought to stabilize the dimerized state and form α helices immediately adjacent to the *C*-terminal α helix in the HLH motif. The bHLH-ZIP proteins include well-known transcription factors, such as Myc and Max, as oncogene products.

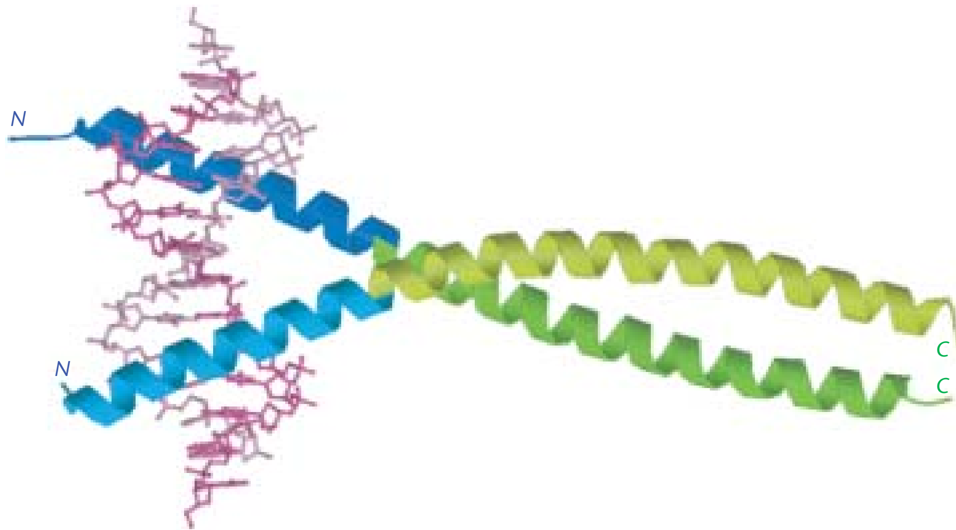

Figure 2 Pap1 bZIP homodimer bound to DNA (PDB accession code 1gd2). The basic region of the ZIP motif is shown in blue and the ZIP motif in green. The double-stranded DNA fragment is shown as a purple stick model. The basic regions that flank the ZIP coiled coil bind DNA with sequence-specific interactions.

Another example of a fused form of the ZIP motif is the homeodomain-leucine zipper (HD-ZIP) motif. The HD-ZIP proteins comprise a family of transcription factors that are apparently unique to plants. The homeodomain consists of 60 amino acid residues and forms a compact globular domain with three α helices, which contain the DNA-binding motif called the helix–turn–helix (HTH) motif. Homeodomain proteins are well known to participate in development, and they function as a monomer. The ZIP motif in the HD-ZIP proteins has relatively longer heptad repeats (six repeats) and facilitates dimerization of the proteins.

Because the α-helical coiled coil is a highly general structural element in proteins, some motifs other than ZIP can form α-helical coiled-coil structures to facilitate dimerization. These motifs contain several leucine residues, but the heptad repeat – especially the periodic repetition of a leucine residue at every seventh position – is incomplete in sequence. For example, yeast transcription factors such as GAL4 and PPR1 have a zinc-finger type of DNA-binding domain and relatively short coiled coils of 14 and 19 residues respectively. The coiled-coil regions are separated from the zinc-finger domain by a nine-residue linker but function to catalyze dimerization of the proteins for full DNA-binding activity.

In some cases, these ZIP-related heptad repeats catalyze higher-order oligomerization. For example, the envelope glycoprotein complex gp120–gp41 of human immunodeficiency virus type 1 (HIV-1) oligomerizes through a heptad repeat located in the *N*-terminal region of the gp41 molecule. This heptad repeat forms a parallel three-stranded coiled coil. In addition, some other heptad repeats form antiparallel coiled coils. The hepatitis delta antigen is well known to form dimer and higher-order structures through the *C*-terminal heptad repeat, which forms an antiparallel coiled coil. The effector domain of the protein kinase PKN forms an antiparallel coiled coil, which binds the small G-protein RhoA (Maesaki *et al.*, 1999). It is unusual in that one strand of this coiled coil has an incomplete ZIP motif but the other strand has a complete ZIP motif.

Dimerization Specificity of ZIP

As originally proposed by Landschultz *et al.* (1988), bZIP proteins not only homodimerize but also heterodimerize by forming coiled-coil structures of the ZIP motifs. However, only specific bZIP protein pairs can heterodimerize. In addition, some bZIP proteins do not homodimerize but do heterodimerize. For example, Fos and Jun associate preferentially to form a heterodimer, and the Fos dimer is extremely unstable. These observations indicate that leucines and nonpolar residues at every ***d*** and ***a*** position, respectively, are the prerequisite for the ZIP dimerization because they provide the fundamental framework of the coiled-coil structure, whereas other residues are important for dimerization specificity.

Dimerization specificity between two ZIP motifs is determined primarily by electrostatic interactions between two helices. The side chains of residues at

e and ***g*** positions of the heptad repeat flank the dimerized coiled-coil interface. These residues are mostly long charged residues, as mentioned above. The charged termini of these side chains form interhelical salt bridges to stabilize the coiled-coil structure (**Figure 3**). Two possible types of interhelical interaction can occur between ***e***- and ***g***-positioned residues: one between the ***g*** position of one helix and the following ***e*** position on the opposite helix; and the other between the ***e*** position and the following ***g*** position of the opposite helix. These different structural configurations of interhelical salt bridges are referred to as $i+5$ and $i+2$ salt bridges respectively. The $i+5$ salt bridges are frequently observed in crystal structures. The $i+2$ salt bridges seem to be destabilized by steric repulsion with leucine side chains at the intervening ***d*** positions. A ZIP motif of typical length (30 residues) has eight possible pairs of ***g*** and ***e*** positions that can form interhelical $i+5$ salt bridges.

In the ZIP motif, some residues in ***a*** positions may be charged or polar. For example, Fos has four polar residues out of six in ***a*** positions. The aliphatic portions of these residues pack against ***a***- and ***g***-positioned residues of the opposite helix. Some long polar side chains of these residues form hydrogen bonds with ***g***-positioned residues of the opposite helix (**Figure 3**). Alternatively, a short polar side chain of an ***a***-positioned residue forms a hydrogen bond with the corresponding residue of the opposite helix. Thus, ***a***- and ***g***-positioned residues provide interhelical polar interactions as an additional determinant of dimerization specificity.

Similar to ***a***-positioned residues, a few residues in the ***d*** position are exceptionally charged or polar. In the case of the yeast AP1-like bZIP transcription factor Pap1, threonine at the ***d*** position forms water-mediated hydrogen bonds with histidine at the ***a*** position of the same helix and with these corresponding residues of the opposite helix (**Figure 4**).

In conclusion, the dimerization specificity of the ZIP motifs is determined by polar interactions involving nonconsensus residues at the ***e*** and ***g*** positions, as well as by exceptionally polar residues at the ***a*** and ***d*** positions. These variant residues represent hallmarks that indicate homo and hetero dimerization specificity.

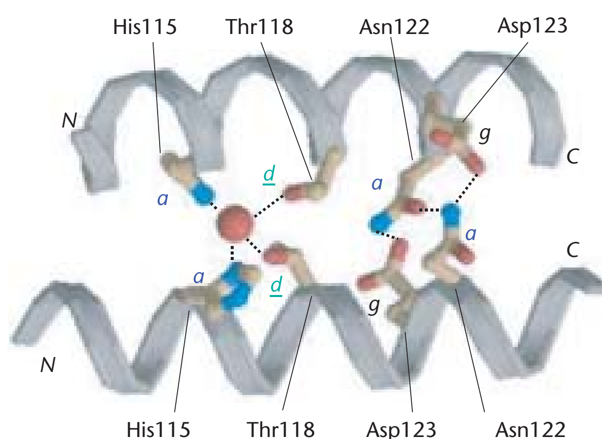

Figure 4 Internal interactions of polar residues in ***a*** and ***d*** positions in the coiled-coil structure of the Pap1 ZIP dimer. Part of the coiled-coil structure is presented to show that His115 in the ***a*** position and Thr118 in the ***d*** position form water-mediated hydrogen bonds at the coiled-coil interface. Polar residues Asn122 in the ***a*** position and Asp123 in the ***g*** position also form interchain hydrogen bonds.

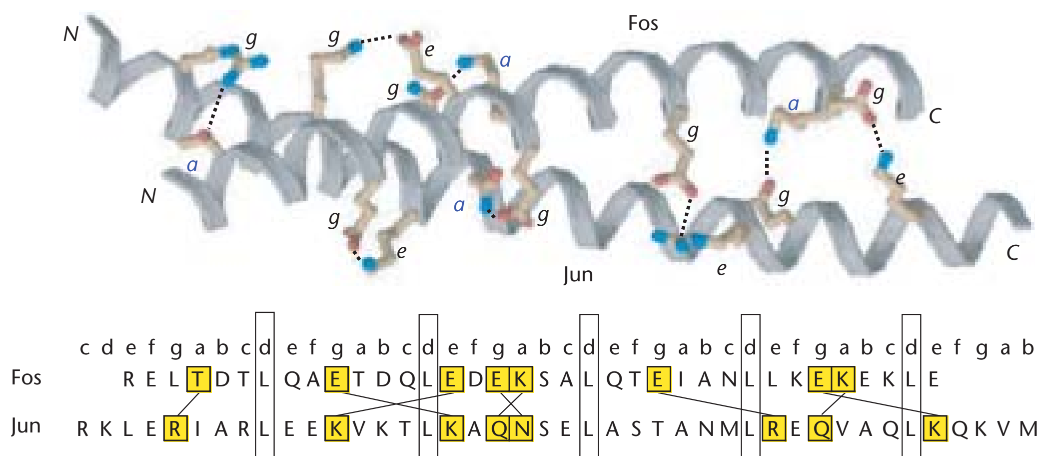

Figure 3 Interchain polar interactions found in the coiled-coil structure of the Fos–Jun ZIP heterodimer (PDB accession code 1fos). Hydrogen bonding interactions are indicated by dotted lines. The dimerization specificity is primary determined by polar interactions involving the side chains of residues at the ***e*** and ***g*** positions. A few polar residues at the ***a*** positions also contribute to the specificity.

DNA-binding Specificity of bZIP

Transcription factors of the bZIP superfamily can bind specifically to the short palindromic or pseudo palindromic target DNA sequences. Members of the adaptor protein complex 1 (AP-1) and cyclic-AMP-responsive (CRE) binding (CREB) transcriptor factor subfamilies bind TGACTCA (the AP-1 site) and TGACGTCA (the CREB site) respectively (the most strongly conserved palindromic bases are underlined). These binding sites are typical consensus sequences that are recognized by the members of the bZIP family. X-ray structures of the bZIP protein–DNA complexes show that the DNA bases of both the AP-1 and the CREB sites are recognized by direct interactions with only five conserved amino acid residues in the signature sequence of their bZIP basic regions, Asn-X-X-Ala-Ala-X-X-Cys-Arg (where X indicates a variable residue).

The other subfamily members have preferences for divergent DNA sequences. The Pap1 subfamily members prefer TTACGTAA and TTAGTAA. The CREB-2 and proline- and acidic amino acid-rich region (PAR) subfamily members also prefer AT-rich binding sites, TTACGTAA and GTTACGTAAC respectively. The C/EBP subfamily members prefer ATTGCGCAAT. The bZIP basic region of Pap1 has a signature sequence, Asn-X-X-Ala-Gln-X-X-Phe-Arg, in which the second alanine of the AP-1 signature sequence is replaced with glutamate and the cysteine is replaced with phenylalanine. These Pap1-specific residues recognize the target bases together. The side chain of the invariant asparagine in the bZIP motif adopts an alternative conformation in Pap1. This conformation, which is stabilized by a Pap1-specific residue and its associated water molecule, recognizes a different base in the target sequence from that recognized by other bZIP subfamilies (Fujii *et al.*, 2000).

See also

Transcription Factors

References

Crick FHC (1953) The packing of α-helices: simple coiled-coils. *Acta Crystallographica* **6**: 689–697.

Ellenberger TE, Brandl CJ, Struhl K and Harrison SC (1992) The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted α helices: crystal structure of the protein–DNA complex. *Cell* **71**: 1223–1237.

Fujii Y, Shimizu T, Toda T, Yanagida Y and Hakoshima T (2000) Structural basis for the diversity of DNA recognition by bZIP transcription factors. *Nature Structural Biology* **7**: 889–893.

Glover JNM and Harrison SC (1995) Crystal structure of heterodimeric bZIP transcription factor c-Fos–c-Jun bound to DNA. *Nature* **373**: 257–261.

Landschulz WH, Johnson PF and McKnight SL (1988) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**: 1759–1764.

Maesaki R, Ihara K, Shimizu T, *et al.* (1999) The structural basis of Rho effector recognition revealed by the crystal structure of human RhoA complexed with the effector domain of PKN/PRK1. *Molecular Cell* **4**: 793–803.

Schultz J, Copley RR, Doerks T, Ponting CP and Bork P (2000) SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Research* **28**: 231–234.

Further Reading

Baranger AM (1998) Accessory factor–bZIP–DNA interactions. *Current Opinion in Chemical Biology* **2**: 18–23.

Branden C and Tooze J (1998) Alpha-domain structures. *Introduction to Protein Structure*, 2nd edn, pp. 35–46. New York, NY: Garland.

Burkhard P, Strelkov SV and Stetefeld J (2001) Coiled coils: a highly versatile protein folding motif. *Trends in Cell Biology* **11**: 82–88.

Hurst HC (1994) Transcription factors. 1: bZIP proteins. *Protein Profile* **1**: 123–168.

O'Shea EK, Rutkowski R and Kim PS (1989) Evidence that the leucine zipper is a coiled coil. *Science* **243**: 538–542.

O'Shea EK, Rutkowski R and Kim PS (1992) Mechanism of specificity in the Fos–Jun oncoprotein heterodimer. *Cell* **68**: 699–708.

Vinson CR, Hai T and Boyd SM (1993) Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: prediction and rational design. *Genes and Development* **7**: 1047–1058.