**bZip Transcription factors: Picking up DNA with chopsticks**

It was said in the film ‘The Karate Kid’ that “Man who catch fly with chopsticks accomplish anything.” The cell that catches DNA with chopsticks however, accomplishes transcription.

Basic Leucine Zipper (bZip) domains form long α-helices which bind to DNA like a pair of chopsticks grasping a spring roll (view-1). The bZip family is one of the largest families of transcription factors, binding to promoter regions of genes to control their expression. As such, bZips are involved in numerous fundamental cellular processes and many are implicated in cancer. The activator protein 1 (AP-1) family for example, which contains the well known transcription factors c-Jun and c-Fos, is responsible for regulation of apoptosis and cell proliferation.

The bZip α-helix is composed of two parts. The N-terminal region, which binds to the DNA, contains many basic (positively charged) residues. The C-terminal region, on the other hand, is rich in hydrophobic leucine residues. Just as it’s difficult to eat with only one chopstick, bZip proteins pair up to grasp DNA. The leucine-rich region is responsible for dimerisation - there’s no hand holding these chopsticks together!

**Groovy interaction**

The iconic structure of the DNA double helix has two grooves of different widths as can be seen in view-1. The major groove is 22Å wide whilst the width of the minor groove is only about 12Å. The N-terminal region of a bZip helix binds in the major groove, which is approximately the same width as the diameter of the helix. The many basic amino acids (and some other polar amino acids) in this part of the bZip domain interact with the negatively charged phosphate backbone of the DNA. The yeast general control protein GCN4 (PDB entry 2dg), responsible for the transcription of many amino-acid biosynthetic genes and shown in view-1, contains nine basic residues in the DNA-binding region.

Most of the interactions between the protein and DNA are these non-specific salt links involving the phosphate backbone. Typically, there are only five residues involved in base-specific interactions. These are the residues positioned on the inward-facing side of the helix, in contact with the edges of the bases in the major groove. Of these base-specific residues, an asparagine and an arginine are invariant across the bZip family. The asparagine hydrogen bonds with the edges of two bases, one on each strand, while the arginine is often observed hydrogen-bonding to a guanine base (view-2).

The three other residues in the binding region of the bZip protein which interact with the DNA determine specificity and affinity. In GCN4 shown in view-2 a water-mediated hydrogen-bond interaction is formed between lysine and a guanidine base. Not all protein-DNA interactions are polar, there are some hydrophobic interactions too. GCN4 contains two alanines in its DNA-binding motif that both make Van der Waals contacts with the methyl group of a
thymidine.

A given bZip protein can recognise more than one DNA sequence (e.g., PDB entries 1gtw, 1gu4, 1gu5) and more than one bZip protein is able to recognise the same sequence of DNA. In this way, the cell's metabolic processes can be regulated with exquisite control.

**Zipping it up**

The C-terminal leucine-rich region of the Zip domain is responsible for dimerisation. This part of the protein is characterised by the presence of the amino acid leucine at every seventh position in the protein sequence. As there are 3.5 residues per turn of the bZip helix, this spacing makes that they are all exposed on the same side of the helix, at every second turn (view-3, PDB entry 1jnm). The dimer is formed by the leucines on one bZip monomer packing against the leucines on the other monomer. The name ‘leucine zipper’ comes from an early model in which the eponymous residues were thought to interdigitate like the teeth of a zip, but subsequent structure determination has revealed that they actually lie next to each other, forming the hydrophobic core of the interface. The two helices wind around each other in a coiled coil, which is a very common motif in protein-protein interactions.

The leucines are not the only residues forming the dimer interface. Between them, also at alternate turns of the helix, are other hydrophobic residues which pack together to form the core of the interface. In the center of the zipper region however, this ‘hydrophobic’ position is often occupied by the polar residue asparagine. This residue hydrogen-bonds to its equivalent in the other helix and helps maintain the register of the zipper, preventing it from becoming 'cross-zipped'.

bZip domains can bind DNA either as homodimers (two copies of the same protein) or by forming heterodimers with different bZip proteins, for example the c-Jun:c-Fos heterodimer AP-1 shown in view-4 (PDB entry 1fos). Their propensity to form homo- or heterodimers is determined mainly by residues around the sides of the helices. These are frequently residues with long sidechains such as arginine, lysine or glutamic acid that reach across the interface such that the hydrophobic part packs against the leucines of the zip but the polar end-group interacts with a similarly placed residue from the other monomer, ‘shaking hands’ across the interface (view-4). Other parameters which determine dimer formation include the relative concentrations of different bZip proteins, and indeed the DNA sequence to which they are binding (ref. 1). This ability to form both homo- and heterodimers combinatorially increases the number of DNA sequences that the bZip family can recognise.

**Zips in a complex picture**

Binding of a bZip protein to DNA is only part of a complicated pathway toward transcription, which typically involve many other proteins. For instance, AP-1 complex of c-Jun and c-Fos binds to DNA cooperatively with a further transcription factor, Nfat. On their own, neither transcription factor is sufficient to induce gene expression but together they induce expression of several immune-response genes. Nfat and AP-1 interact with each other (view-5) when bound to DNA (PDB entry 1a02). In turn, this complex promotes the binding of other transcription factors, and ultimately RNA polymerase, which begins transcribing the
gene.

**What makes the chopsticks pick up the DNA?**
Whilst bZip proteins control the expression of other genes, their own expression is controlled by other transcription factors. But that’s not the whole story.

In many cases, the bZip domain is only a quarter of the full-length transcription factor. The remaining region is frequently subject to phosphorylation protein kinases, and the phosphorylation state of the protein determines its stability. The cell can therefore regulate the amount of a bZip transcription factor by altering its phosphorylation state. To date, there are no structures of any full-length bZip transcription factors. The sequence of the remainder of the protein suggests that, at least in some cases, it is unlikely to have a stable structure. These chopsticks may have very long and floppy handles! But that’s a challenge to the structural biologists out there.

**Views**

**View 1: bZip domain binding to DNA.** The GCN4 bZip domain is shown as a dodgerblue cartoon representation and DNA is shown as spheres with carbon atoms coloured grey. Sidechains which interact with the DNA backbone are shown as cyan sticks and the interactions to the DNA backbone portrayed as white lines.
View 2: **Sequence-specific binding.** DNA is shown as spheres, changing to sticks, with grey coloured carbons. The protein is shown as dodgerblue cartoons. Sidechains which determine sequence specificity are shown, invariant residues in green and variable ones in mauve. Hydrogen-bonding side chains are shown as sticks with the hydrogen bonds as dotted white lines. Residues making Van der Waals contacts are shown as spheres, as are the atoms in the bases which they contact.

View 3: **The leucine-zipper dimerisation domain.** DNA is shown as spheres. The two c-Jun bZip domains are shown as dodgerblue cartoons. Hydrophobic residues in the interface are shown as spheres in shades of green for leucine and gray for other hydrophobic residues. The central asparagines are shown as deeppink sticks.
**View 4: Residues determining dimerisation specificity.** A c-Jun:c-Fos dimer is shown as cartoons. Residues around the sides of the helices which form inter-molecular hydrogen bonds are shown as sticks coloured violet for c-Jun and gold for c-Fos. Leucines in the zipper are shown as green sticks.

**View 5: bZip domains are a part of larger transcriptional complexes.** The complex of c-Jun and c-Fos (shades of blue) with Nfat (grey) is shown as cartoons bound to DNA (shown as spheres). The proteins are then shown as surfaces in the same colours.